The Pantheon of Derivatives

Contents

1 Differentiation in a Nutshell

There are many terms and special cases, which deal with the process of differentiation. The basic idea, however, is the same in all cases: something non-linear, as for instance a multiplication, is approximated by something linear which means, by something we can add. This reveals already two major consequences: addition is easier than multiplication, that's why we consider it at all, and as an approximation, it is necessarily a local property around the point of consideration.

Thus the result of our differentiation should always be a linear function like the straight lines we draw in graphs and call them tangents. And our approximation will get worse the farther away we are from the point we considered. That's the reason why these ominous infinitesimals come into play. They are nothing obscure, but merely an attempt to quantify and get a hand on the small deviations of our linear approximation to what is really going on.
To begin with, let's clarify the language:

differentiation –	certain process to achieve a linear approximation
to differentiate –	to proceed a differentiation
differential –	infinitesimal linear change of the function value
differentiability –	condition that allows the process of differentiation
derivative –	result of a differentiation
derivation –	linear mapping that obeys the Leibniz rule
to derivate –	to deduce a statement by logical means

All these terms are context sensitive and their meanings change, if they are used, e.g. in chemistry, mechanical engineering or common language. But even within mathematics, the terms may vary among different authors. E.g. differential has two meanings, as adjective or as notation for df, i.e. the infinitesimal linear change on the function values. Differentials are used in various applications with varying meanings and even with different mathematical rigor. This is essentially true in calculus where $\int f(x)dx$ and $\frac{df(x)}{dx}$ is only of notational value. The most precise meaning of the term can be found in differential geometry as an exact $1-$form. As a thumb rule might serve: diff... refers to the process, derivative to the result.

As differentiability is a local property, it is defined on a domain U which is open, not empty and connected, at a point x_0 or z_0 in U. I will not mentioned these requirements every time I use them. They are helpful, as one doesn't have to deal with isolated points or the behavior of a function on boundaries and one always has a way to approach x_0 from all sides. So with respect to the approximation which is intended, they come in naturally. I also won't distinguish between approximations from the left or from the right, since this article is only an overview. So it is always meant as identical from both sides. Moreover, a function is said to be in $C(U) = C^0(U)$ if it is continuous, in $C^n(U)$ if it is $n-$times continuously differentiable, and in $C^\infty(U) = \bigcap_{n\in\mathbb{N}} C^n(U)$ if it is infinitely many times continuously differentiable. The latter functions are also called **smooth**.

1.1 Real Functions in one Variable: $\mathbb{R}$

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at x_0, if the limit

$$\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{v\to 0} \frac{f(x_0 + v) - f(x_0)}{v} = \lim_{\Delta x\to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

exists, which is then called the derivative of f at x_0 and denoted by

$$f'(x_0) = \left.\frac{d}{dx}\right|_{x=x_0} f(x)$$

This is the definition we learn at school. But I think it hides the crucial point. There is another way to define it, which describes much better the purpose and geometry of the concept, **Weierstraß' decomposition formula**:

f is differentiable at x_0 if there is a linear map J, such that

$$\mathbf{f(x_0 + v) = f(x_0) + J(v) + r(v)} \tag{1}$$

where the error or remainder function r has the property, that it converges faster to zero than linear, which means

$$\lim_{v\to 0} \frac{r(v)}{v} = 0$$

The **derivative** is now the linear function $J(.) : \mathbb{R} \to \mathbb{R}$ which approximates f at x_0 with a (more than linear) error function r. This function is e.g. quadratic as in the Taylor series. Both functions may depend on x_0 which plays the role of a constant parameter for them.

1.2 Real Functions in many Variables: $\mathbb{R}^n$

Here another advantage of the last definition becomes obvious. If our function is defined on real vector spaces, say $f : \mathbb{R}^n \to \mathbb{R}^m$, how could we divide vectors? Well, we can't. So the definition

$$f(x_0 + v) = f(x_0) + J(v) + r(v)$$

comes in handy, because it has none. We can directly use the same formula without any adjustments. We only have to specify, what "the remainder $r(v)$ converges faster to zero than linear" means. Since we're not especially interested in it except it's general behavior at zero, we simply require

$$\lim_{v\to 0} \frac{r(v)}{||v||} = 0$$

which is more out of practicability to quantify "faster than linear" than it is an essential property. The essential part is the linear approximation $J(.) : \mathbb{R}^n \to \mathbb{R}^m$ which is why differentiation is done for. Again both functions may depend on the constant x_0.

We can also see, that now the direction of v comes into play, which makes sense, since the tangents are now tangent spaces, planes for example. And a plane has many different slopes, generated by two coordinates. It makes a lot of a difference whether we walk on a hill surrounding it, or climbing it. Therefore the process above is called total differentiation and $J(.)$ the **total derivative**, because it includes all possible directions. In standard coordinates it is the **Jacobian matrix** of $f(x)$ in $\mathbb{M}_{m\times n}(\mathbb{R})$.

If we are only interested in one special direction, then we get the **directional derivative**. We can take the same definition, only that v is now a specified vector pointing in a certain direction. This means that we approximate f only in one direction. Our directional derivative as linear approximation therefore depends only on one vector v and it can be written as $J(v)(f(x)) = v \cdot f(x) = \vec{v}^\tau \cdot \overrightarrow{f(x)} =< v, f(x) >$ which maps a vector $f(x)$ to its part of the slope in direction of v. In this case, the directional derivative is also written as:

$$\mathbf{J}(\mathbf{v}) = \mathbf{D_v f}(\mathbf{x}) = \nabla_\mathbf{v}\mathbf{f}(\mathbf{x}) = \partial_\mathbf{v}\mathbf{f}(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{v}} = \mathbf{f}'_\mathbf{v}(\mathbf{x}) \tag{2}$$

If f is also totally differentiable, then additional notations are in use:

$$\mathbf{J}(\mathbf{v}) = \mathbf{Df}(\mathbf{x})\mathbf{v} = \mathbf{Df_x}\,\mathbf{v} = \text{grad }\ \mathbf{f}(\mathbf{x}) \cdot \mathbf{v} = \nabla\mathbf{f}(\mathbf{x}) \cdot \mathbf{v} = (\mathbf{v} \cdot \nabla)\mathbf{f}(\mathbf{x}) \tag{3}$$

A directional derivative is often defined for scalar functions $f : \mathbb{R}^n \to \mathbb{R}$, i.e. $m = 1$. This isn't really a restriction, because one could always simply take all components of $f = (f_1, \ldots, f_m)$. Furthermore they are also often defined for unit vectors v_0 as direction and then as the limit of $\frac{1}{t}(f(x_0 + t \cdot v_0) - f(x_0)$ for $t \to 0$. However, there is no need to do this. It's a matter of taste and only means, that we have to divide by $||v||$ if scales like coordinates are involved.

From here **partial derivatives** are obviously simply the directional derivatives in the various variables $x_1, \ldots, x_n$ of f, the coordinates of $\mathbb{R}^n$.
The dependencies among these differentiability conditions are as follows:

$$\begin{gathered}\text{continuous partially differentiable,}\\ \text{i.e. all partial derivatives are continuous}\\ \Downarrow\\ \text{totally differentiable or differentiable for short}\\ \Downarrow\\ \text{differentiable in any direction}\\ \Downarrow\\ \text{partially differentiable}\end{gathered}$$

All implications are proper implications. (Counter-) Examples are (from Wikipedia):

$$f(x,y) = \begin{cases} (x^2+y^2)\cdot \sin\frac{1}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

is totally differentiable but not continuous partially.

$$f(x,y) = \begin{cases} \frac{3x^2y-y^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

is differentiable in all directions, but they don't define a linear function J.

$$f(x,y) = \begin{cases} \frac{xy^3}{x^2+y^4} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

is differentiable in all directions, and they define a linear function J, but it is not totally differentiable, because the remainder term doesn't converge to zero.

$$f(x,y) = \begin{cases} \frac{2xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

is partially differentiable but not all directional derivatives exist.

1.3 Complex Functions: $\mathbb{C}$

Complex functions $f : U \to \mathbb{C}$ are somehow special and entire textbooks deal with the complex part of analysis. So I will restrict myself to a brief listing of terminology and dependencies. What appears to be more complicated at first glance is to some extend even easier than in the real case. To begin with, I like to mention, that we haven't used any specifically properties of $\mathbb{R}$ in the previous sections apart the Euclidean norm and directions. However, both is given over $\mathbb{C}$ as well, and all we have to think about is, that linearity in our definition

$$\mathbf{f}(\mathbf{x_0}+\mathbf{v}) = \mathbf{f}(\mathbf{x_0}) + \mathbf{J}(\mathbf{v}) + \mathbf{r}(\mathbf{v}) \tag{4}$$

now means $\mathbb{C}-$linearity of J. A complex differential function is called **holomorphic function** and in older literature sometimes **regular function**. As regularity is widely used in various areas of mathematics, it should be avoided here. A function which is holomorphic in the entire complex plane $U = \mathbb{C}$ is called an **entire function** or an **integral function**. These are strong requirements, which means that we sometimes need a weaker condition, namely one that allows us to consider poles. Poles are isolated points, at which functions are not defined. Therefore a function, which is holomorphic on U except at its poles, is called a **meromorphic function**.

It might be due the many different terms in complex analysis, which sometimes leads to the impression, that the complex case is more difficult than the real case. I think this is mainly for historical reasons and the need to have useful adjectives for certain properties. Until now I've neglected the representation of functions by series, which have - beside their practical advantages - often been the historically first approach to deal with the various concepts. Their names are:

- Power series
$$\sum_{n=0}^{\infty} a_n(x-x_0)^n$$
- Laurent series
$$\sum_{n=-\infty}^{\infty} a_n(x-x_0)^n$$
- Taylor series
$$\sum_{n=0}^{\infty} \frac{1}{n!} \cdot f^{(n)}(x_0) \cdot (x-x_0)^n$$

- Maclaurin series
$$\sum_{n=0}^{\infty} \frac{1}{n!} \cdot f^{(n)}(0) \cdot x^n$$

Now it has to be considered, as some functions in complex analysis are called analytic. Although often used in the context of complex valued functions, analytic can equally be defined for real valued functions.

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ be either the real or complex numbers and $f : U \to \mathbb{K}$. Then f is called **analytic** at x_0, if there is a power series
$$\sum_{n=0}^{\infty} a_n (x - x_0)^n \tag{5}$$
that converges to $f(x)$ in a neighborhood of x_0. If f is analytic in every point of U, then f is called analytic without the emphasis on any points. Analytic functions are smooth, i.e. in $C^\infty(U)$. This implication is proper, as the real function
$$f(x, y) = \begin{cases} \exp(-x^{-2}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$
is smooth everywhere, but not analytic at zero.

How do all these definitions relate to each other? $\mathbb{C}$ is a two dimensional real vector space and the defining equations are the same. The only difference is the $\mathbb{C}-$ linearity of J. However, this is a quite powerful difference:
A function $f : U \to \mathbb{C}$ with $f(x + iy) = u(x, y) + iv(x, y)$ is holomorphic (differentiable) at $z_0 = x_0 + iy_0$ if f is totally differentiable as function on $\mathbb{R}^2$ and the derivative J is a $\mathbb{C}-$linear mapping. This means that
$$J = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}$$
is represented by a skew-symmetric matrix w.r.t. the basis $\{1, i\}$, i.e. that for f the **Cauchy-Riemann (differential) equations** hold
$$u_x = \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = v_y \text{ and } u_y = \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} = v_x \tag{6}$$

In a neighborhood U of $z_0 \in \mathbb{C}$ a function $f(x + iy) = u(x, y) + iv(x, y)$ is holomorphic at z_0

$$\Longleftrightarrow$$

once complex differentiable at z_0, i.e. $f \in C^1(U)$

$$\Longleftrightarrow$$

infinite many times complex differentiable at z_0, i.e. $f \in C^\infty(U)$

$$\Longleftrightarrow$$

analytic at z_0 (locally)

$$\Longleftrightarrow$$

u and v are at least once real totally differentiable at (x_0, y_0) and satisfy the Cauchy-Riemann differential equations (6)

$$\Longleftrightarrow$$

f is real totally differentiable at (x_0, y_0) and $\frac{\partial f}{\partial \zeta} = 0$

with the **Cauchy-Riemann operator** $\partial\zeta = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$

$$\Longleftrightarrow$$

f is continuous at z_0 and its path integral over any closed, simple connected (0-homotopic), rectifiable curve γ in U is identically zero: $\oint_\gamma f(z)d(z) = 0$

Cauchy's integral or **Cauchy-Goursat theorem** (7)

$$\Longleftrightarrow$$

if U_0 is a circular disc in U with center z_0 then for $z \in U_0$ holds

$$\textbf{Cauchy's integral formula } f(z) = \frac{1}{2\pi\, i} \oint_{\partial U_0} \frac{f(\zeta)}{\zeta - z}\, d\zeta \qquad (8)$$

2 Generalizations Beyond $\mathbb{R}$ and $\mathbb{C}$

As mentioned in the section of complex functions, the main parts of defining a differentiation process are a norm and a direction. So to extend the differentiation concepts on normed vector spaces seems to be the obvious thing to do.

2.1 Fréchet Derivative

Definition: Let X and Y be two Banach spaces, i.e. normed real or complex vector spaces, which are complete as normed topological spaces, and $U \subseteq X$

an open subset. A function

$$f : (U, ||.||_X) \longrightarrow (Y, ||.||_Y)$$

is Fréchet differentiable at $x_0 \in U$ if there is a continuous linear operator $J : X \to Y$ such that

$$\lim_{v \to 0} \frac{||f(x_0 + v) - f(x) - J(v)||_Y}{||v|||_X} = 0 \tag{9}$$

The operator J is called the **Fréchet derivative** of f at x_0 and is written $J = Df_{x_0} = Df(x_0)$ indicating the dependence of the linear approximation at x_0.
Sometimes it is only required that X , Y are normed vector spaces, but as limits are involved, it is more convenient to require Banach spaces, i.e. complete spaces. Also the continuity requirement is new here, as it is not automatically the case and continuous functions are the natural (homo-)morphisms in the category of topological spaces. Taking a closer look on this limit reveals, that the similar (equivalent) change to the definition can be made as in the real case. Therefore we consider (as always)

$$\mathbf{f}(\mathbf{x_0} + \mathbf{v}) = \mathbf{f}(\mathbf{x_0}) + \mathbf{J}(\mathbf{v}) + \mathbf{r}(\mathbf{v}) \tag{10}$$

with a faster than linear vanishing term $r(v)$ and note, that the Fréchet derivative is unique, if it exists. This also means that the Fréchet derivative coincides with the usual derivative in finite dimensional spaces, where the linear operator J can be represented by the Jacobian matrix. Whereas in the finite dimensional case all linear operators are Fréchet differentiable, in the infinte dimensional case only and exactly the bounded linear operators are Fréchet differentiable, unbounded are not.

2.2 Gâteaux Derivative

2.2.1 The Directional Derivative

The Gâteaux derivative is a generalization to normed vector spaces, too, the directional derivative. Let $f : (X, ||.||_X) \longrightarrow (Y, ||.||_Y)$ be a function on Banach spaces, x_0 a point in an open neighborhood $U \subseteq X$ and v a directional vector in $(X, ||.||_X)$.

Unfortunately this is were the easy part gets to a hold. I chose X, Y to be Banach spaces for the sake of simplicity. Usually they are only required to be locally convex, normed vector spaces. This is already an indicator of the difficulty we will face: the additivity of Gâteaux derivatives.

The English Wikipedia [20] defines (remember that df is the differential, J the derivative)

"*At each point* $x_0 \in U$*, the Gâteaux differential defines a function* $df(x_0, .) = J_{x_0} : X \to Y$ *which is homogeneous, i.e.* $J_{x_0}(\alpha \cdot v) = \alpha \cdot J_{x_0}(v)$*. However, this function need not be additive, so that the Gâteaux derivative may fail to be linear, unlike the Fréchet derivative. Even if linear, it may fail to depend continuously on* v *if* X *and* Y *are infinite dimensional. Furthermore, for Gâteaux derivatives that are linear and continuous in* v*, there are several inequivalent ways to formulate their continuous differentiability.*"

The German version [19] defines
"*If* $df(x_0, .)$ *is a continuous, linear functional, i.e. the function* $v \mapsto J_{x_0}(v)$ *is homogeneous, additive and continuous, then it is called a Gâteaux derivative at* x_0*.*"

Well, René Gâteaux has been a French mathematician, so let's have a look on the French Wikipedia [21]

"*The Gâteaux derivative of* f *at* x_0 *in the direction of* v *is the limit in* Y *(so it exists)*"

$$J_{x_0}(v) = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(x_0+tv)-f(x_0)}{t} = \left.\frac{d}{dt}\right|_{t=0} f(x_0 + tv) \qquad (11)$$

where the variable t *is taken real ... The function* f *is Gâteaux differentiable at* x_0 *if there is a linear, continuous operator* $J_{x_0} : X \to Y$ *such that* $v \mapsto J_{x_0}(v)$ *exists for all* $v \in X$"

Maybe it's best to handle it like nlab [18] which links directly to their definition of directional derivatives or as in a paper from Texas Tech [13] in which they don't bother linearity within the definition either. What's all in common is the fact that the Gâteaux derivative definition isn't unique, but always generalizes the concept of a directional derivative to infinite dimensional normed vector spaces of some kind, local convexity as minimal requirement.

2.2.2 Definitions and Examples

We assume the same conditions as in the previous section. Let $f : X \to Y$ be a function on Banach spaces, $x_0 \in U \subseteq X$ a point at which we differentiate and $v \in X$ a direction in which we differentiate.

Definition [Weierstraß]: A linear function $J : X \to Y$ such that

$$\mathbf{f}(\mathbf{x_0} + \mathbf{v}) - \mathbf{f}(\mathbf{x_0}) = \mathbf{J}(\mathbf{v}) + \mathbf{r}(||\mathbf{v}||_{\mathbf{X}}) \tag{12}$$

with $r(t)$ vanishing faster than linear is called Gâteaux derivative J_{x_0} at x_0.

Definition [Variational Derivative]: The Gâteaux derivative of f at x_0 in the direction of v is the limit in Y with real t

$$J_{x_0}(f)(v) = \lim_{\substack{t\to 0\\ t\neq 0}} \frac{f(x_0 + tv) - f(x_0)}{t} = \left.\frac{d}{dt}\right|_{t=0} f(x_0 + tv) \tag{13}$$

Second Variation

$$d^2 f(x_0; v) = \left.\frac{d^2}{dt^2}\right|_{t=0} f(x_0 + t \cdot v) \tag{14}$$

Derivatives of higher orders are defined accordingly. Also derivatives from the left or from the right are sometimes distinguished when dealing with Gâteaux derivatives.

Linearity and Continuity

Let's consider the function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = \begin{cases} \frac{x^3}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

Then f is Gâteaux differentiable with the derivative

$$J_{(0,0)}((a,b)) = \begin{cases} \frac{a^3}{a^2+b^2} & \text{if } (a,b) \neq (0,0) \\ 0 & \text{if } (a,b) = (0,0) \end{cases}$$

at $(0,0)$ in the direction (a,b) according to the variational definition. $J_{(0,0)}$ is even continuous, however, f is not Gâteaux differentiable in the Weierstraß' sense, because $J_{(0,0)}$ is not linear. Note that it is still homogeneous, i.e. $J_{(0,0)}(\alpha v) = \alpha J_{(0,0)}(v)$.

Next consider the space of real, smooth functions on $[0,1] \subseteq \mathbb{R}$. That is

$$X = C^\infty_{\mathbb{R}}([0,1])$$

equipped with the uniform norm, the supremum norm

$$||f|| = \sup_{x\in[0,1]} \{|f(x)|\}$$

and $Y = (\mathbb{R}, |.|)$. Then the derivative-at-zero operator $T(f) := f'(0)$ is linear, closed, but not continuous. (Consider the sequence $f_n(x) = \frac{\sin(n^2 x)}{n}$ which converges uniformly to $f \equiv 0$ but $(T(f_n))_{n\in\mathbb{N}}$ does not.)

As the completeness condition of the normed vector spaces play an important role here, too, the general advice when using the Gâteaux derivative has to be: Make sure which definition you use and what the exact nature of the normed vector spaces are (locally convex, complete, Banach, etc.)

Connection to the Fréchet Derivative

If f is Fréchet differentiable at x_0 with Fréchet derivative $J_{x_0,F}$ then f is also Gâteaux differentiable in all directions v and for the Gâteaux differential $df(x_0; v)$ holds

$$df(x_0; v) = J_{x_0,G}(f)(v) = J_G(f)(v) = J_{x_0,F}(f)(v) \tag{15}$$

Especially $J_F(f) = J_G(f)$. In general, the opposite direction does not hold, i.e. from Gâteaux differentiability cannot be concluded Fréchet differentiability. Since the finite dimensional real case is a special case of both concepts, this was to be expected.

If f is Gâteaux differentiable in a neighborhood U of x_0, such that $J_{x,G}(v)$ is continuous (in x) and linear, and the operator

$$J_G(f) : U \to \mathcal{L}(X,Y)$$

$$J_G(f) : x \mapsto (v \mapsto J_{x,G}(f)(v)) \tag{16}$$

is continuous with respect to the operator norm on the space of linear functions $\mathcal{L}(X, Y)$, then f is also Fréchet differentiable. It is not a necessary condition, so we have the real finite dimensional case again as a special case.

Lagrange Formalism

Let us define the function

$$f(\varepsilon) := \int dt\, L(q(t) + \varepsilon\delta q\,,\, \dot{q}(t) + \varepsilon\delta\dot{q}\,,\, t) \tag{17}$$

For the Gâteaux differential we get as first order approximation (*) and by partial integration (with constant endpoints of integration t_i) and thus a vanishing $\delta q(t_i)$ term in the anti-derivative (**)

$$\begin{aligned}
\delta f &= J_{\delta q}(f) \\
&= \int dt\, J_{\delta q}(L) \\
&= \int dt\, \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(L(q(t) + \varepsilon\delta q\,,\, \dot{q}(t) + \varepsilon\delta\dot{q}\,,\, t) - L(q(t)\,,\, \dot{q}(t)\,,\, t) \right) \\
&\overset{(*)}{=} \int dt\, \left(\frac{\partial L}{\partial q}\delta q + \frac{\partial L}{\partial \dot{q}}\delta\dot{q} \right) \\
&\overset{(**)}{=} \int dt\, \frac{\partial L}{\partial q}\delta q \,-\, \int dt\, \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \right) \delta q
\end{aligned} \tag{18}$$

By the variation principle this means

$$J_{\delta q}(L) = \frac{\partial L}{\partial q}\delta q \,-\, \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \right) \delta q$$

or

$$\frac{\delta L}{\delta q} = \frac{\partial L}{\partial q} \,-\, \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \right) \tag{19}$$

3 Lie Derivative - Preliminaries

In this section we go even further with our generalizations. The main reason to consider differentiation processes is to calculate a linear approximation of non-linear objects. So far we regarded functions of normed linear spaces, i.e. (non-linear) equations which described curves and other analytic varieties. What they all had in common was, that they took place in an outer frame, normed vector spaces like $\mathbb{R}^n$ or Banach spaces. The frame brought with it the coordinates in which points and directions have been expressed. It was one of the greatest achievements in differential geometry to abandon this restriction: what if there is no outer frame like in General Relativity? Carl Friedrich Gauß had been obliged as land surveyor of the Kingdom of Hanover. The earth isn't flat either nor is it naturally placed in an outer Euclidean frame. So mathematicians started to consider the analytic varieties which they called manifolds by themselves. Coordinates became local properties of the manifold, which is sufficient as we deal with local phenomena in differential geometry anyway. Outer frames were no longer needed to solve the problems within or on the manifolds.

3.1 Manifolds

Definition [5]: An **m-dimensional mainfold** (sometimes shortly **m-manifold**) is a set M, together with a countable collection of subsets $U_i \subseteq M$, called the **coordinate charts**, and $1:1$ functions $\chi : U_i \to V_i$ onto connected open subsets V_i of $\mathbb{R}^m$, called local coordinate maps, which satisfy the following properties:

The coordinate charts cover M.

$$\bigcup_i U_i = M \tag{20}$$

On the overlap of any pair of coordinate charts $U_i \cap U_j$ the composite map

$$\chi_j \circ \chi_i^{-1} \; : \; \chi_i(U_i \cap U_j) \to \chi_j(U_i \cap U_j) \tag{21}$$

is a smooth (infinitely differentiable) function.

(c) If $x_i \in U_i \, , x_j \in U_j$ are distinct points of M, then there exist open subsets W_i of $\chi_i(x_i)$ in V_i and W_j of $\chi_j(x_j)$ in V_j such that

$$\chi_i^{-1}(W_i) \cap \chi_j^{-1}(W_j) = \emptyset \tag{22}$$

The coordinate charts endow the manifold M with the structure of a topological space. Equation (22) is basically a restatement of the Hausdorff separation axiom.

The overlapping functions $\chi_j \circ \chi_i^{-1}$ determine the degree of differentiability of the manifold. If they are smooth (C^∞) diffeomorphisms on open subsets of the corresponding Euclidean space $\mathbb{R}^m$ then the manifold is called **smooth**, if they are real analytic functions, then the manifold is called **analytic**. Similar is true for the other differentiation classes C^k.

It is important to note, that the manifold M isn't part of $\mathbb{R}^m$. As a set it is defined without any reference to a Euclidean space in which it might or might not be embedded. With the usual standard examples: $M = \mathbb{R}^m$, $M = S^n$ or M a torus, there might be some surrounding Euclidean space in our imagination, but things change if we consider **Lie Groups** as manifolds instead, i.e. manifolds which carry an analytic group structure, means inversion and multiplication are analytic functions. Or if you like, the universe. The role of $\mathbb{R}^m$ in the definition is therefore not to characterize the manifold globally, but locally instead. A manifold behaves locally in an open neighborhood of a point like an open set in the Euclidean space $\mathbb{R}^m$ where we can use charts of M as we use flat roadmaps to find our routes through a mountainous countryside.

3.2 Vector Fields

Definition [3]: A **vector field** X on a m-dimensional manifold M is a mapping, that assigns to each point $p \in M$ a vector $X(p) = X_p$. If the X_p are tangent to M, then X is called a **tangent vector field**, if the X_p are perpendicular to M, then X is called a **normal vector field**. Usually if not defined otherwise, the term vector field always refers to the tangent field.

In physics we often distinguish between vector fields and scalar fields. The difference is, that in case of a **scalar field**, there is a scalar (number) assigned at each point of the manifold. Temperatures on earth are a standard example for a scalar field, whereas the meteorologic wind chart represents a vector field, because at each point on earth there is a wind vector with a direction and a magnitude attached. Well, at least almost everywhere according to the Hairy-Ball-Theorem. This leads us directly to some important vector fields.

The gradient of a real valued function $f : U \to \mathbb{R}\,,\, U \subseteq \mathbb{R}^n$

$$\nabla f(x_0) = \operatorname{grad}(f)(a) = \left(\frac{\partial}{\partial x_1} f(x_0), \ldots, \frac{\partial}{\partial x_n} f(x_0)\right) \tag{23}$$

defines a **gradient (vector) field** $F = \nabla f$. The mapping $x_0 \mapsto \nabla f(x_0)$ determines the linear function

$$f'(x_0)(v) = \sum_{i=1}^{n} v_i f'(x_0)(e_i) = \sum_{i=1}^{n} v_i \delta_i f(x_0) = \delta_v f(x_0) = \langle \nabla f(x_0), v \rangle \tag{24}$$

The gradient field is a special case of a tangent field. Basically all derivatives introduced so far have been tangent fields: $x_0 \mapsto J_{x_0}(v)$.

Another important vector field in the Euclidean three-dimensional space is given by the **curl operator** or **rotation**. Let $F : U \to \mathbb{R}^3$, $U \subseteq \mathbb{R}^3$ be a partially differentiable vector field. Then the curl defines a new vector field

$$\operatorname{curl} F = \operatorname{rot} F = \nabla \times F \tag{25}$$

An example of a scalar field in this context is the **divergence** of a vector field F defined by scalar or dot product

$$\operatorname{div} F = \nabla \cdot F \tag{26}$$

Combined, i.e.

$$\Delta f = \nabla^2 f = \nabla \cdot \nabla f \tag{27}$$

they define the **Laplace operator**: the divergence of the gradient field.

To define a Lie derivative, we could shortly say: it's the multiplication in a Lie algebra. No manifolds, no vector fields. Of course this wouldn't meet the requirements to actually understand what it is, because it meant to define Lie derivatives by one aspect of the resulting function, rather than by it's motivation. Therefore we need some more terminology.

3.3 Flows

A curve $\gamma : [a, b] \to X$ on a vector field V of the set X is defined by the property $\frac{d}{dt}\big|_{t=t_0} \gamma(t) = V(\gamma(t_0))$. If V is Lipschitz continuous, i.e. the overlapping charts are, then for each point $x \in X$ there is a unique differentiable curve γ_x such that for some $\varepsilon > 0$

$$\gamma_x(0) = x \text{ and } \left.\frac{d}{dt}\right|_{t=t_0} \gamma_x(t) = V(\gamma_x(t_0))\,,\ t \in (-\varepsilon, +\varepsilon) \subseteq \mathbb{R} \tag{28}$$

These curves γ_x are called **integral curves** or **trajectories** or **flow lines** of the vector field V and they partition X into equivalence classes.

We speak of a **flow on a vector field** as the set of all these curves, and

$$\gamma_{\gamma_x(t)}(s) = \gamma_x(s + t) \text{ or more convenient } \gamma(\gamma(x, t), s) = \gamma(x, s + t) \tag{29}$$

holds, i.e. it doesn't matter, whether we first move by t and then by s along the curve or vice versa. Flows are usually required to be compatible with structures endowed on X, which means in our case, that the curves

$\gamma_x(t)$ must be continuous (in both arguments). If X is equipped with a differentiable structure, then they are required to be differentiable as well. In these cases the flow forms a one parameter subgroup of homeomorphisms and diffeomorphisms, respectively. **Local flows** are the curves in an open neighborhood of a certain point x_0.

4 Some Topology

Whereas the terminology of vector fields, trajectories and flows almost by itself suggests its origins and physical relevance, the general treatment of vector fields, however, require some abstractions. The following might appear to be purely mathematical constructions, and I will restrict myself to a minimum, but they actually occur in modern physics: from the daily need to solve differential equations on various (non Euclidean) geometric objects like in general relativity or quantum field theory, to the front end research in cosmology.

4.1 Vector Bundles

The tangents on a manifold M define a vector field in a natural way. That is, at each point $x \in M$ there are the tangents to all possible curves through x and they span the tangent (vector) space $TM|_x$ at this point. If M is an m-dimensional manifold, then $TM|_x$ is an m-dimensional vector space with the local coordinates $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^m}$. Now we consider the collection of all these tangent spaces, i.e. for all points of M. This gives us a collection

$$TM = \bigcup_{x \in M} TM|_x \tag{30}$$

which we call **tangent bundle** of M. This can be generalized to an arbitrary vector field, in which case it is called a **vector bundle**. Note that these objects are actually **tangent space bundles**, resp. **vector space bundles**.

Definition: A real (complex) **vector bundle**, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, is a triple (E, X, π) which consists of

- two topological spaces E, the **total space**, and X, the **base space**,
- a continuous projection $\pi : E \to X$, the **bundle projection**,
- and $E_x = \pi^{-1}(x)$ is endowed with the structure of a finite-dimensional real (complex) vector space for every $x \in X$, the **fiber** of x,

together with the following local compatibility conditions:
For every point in $x \in X$, there is an open neighborhood U of x, a natural number n, and a homeomorphism

$$\eta : U \times \mathbb{K}^n \to E|_U = \pi^{-1}(U) \subseteq E \tag{31}$$

such that for all $x \in U$,

$$(\pi \circ \eta)(x, v) = x \tag{32}$$

for all $v \in \mathbb{K}^n$ and the map

$$v \mapsto \eta(y, v) \tag{33}$$

establishes a linear isomorphism between the vector spaces $\mathbb{K}^n$ and $E_y = \pi^{-1}(y)$ for each $y \in U$.

The open neighborhood U together with the homeomorphism η is called a **local trivialization** (U, η) **of the vector bundle**. It means, that the bundle projection π behaves locally like the projection of $U \times \mathbb{K}^n$ onto U. This guarantees, that the dimension of all fibers of a connection component of a point $x \in X$ is the same. If it is the same number n for the entire topological space X, then the vector bundle $E = (E, X, \pi)$ is of **rank n**.

Note that the notation in (30) is also the notation of the tangent bundle of a manifold M by its total space TM, whereas T_xM denotes a single fiber at $x \in M$. The total space is not denoted as a pair $(M, T_{(.)}M)$ to avoid the false expression of a Cartesian product. The tangent bundle (field) on a manifold is often simply referred to as vector bundle (field) on M.

Vector bundles of rank 1 are called **line bundles**. A vector bundle of the form $(X \times \mathbb{K}^n, X, proj_1)$ is called a **trivial vector bundle**. I think the easiest non trivial vector bundle is the line bundle of a Möbius strip, which is locally homeomorph to $U \times \mathbb{R}$ with an open set $U \subseteq S^1$ of a circle. The twist guarantees, that it is nontrivial, because a global structure $(S^1 \times \mathbb{R}, S^1, proj_1)$ would define a cylinder.

The most important vector bundles when dealing with derivatives are tangent bundles $TM = (TM, M, \pi)$ of (smooth) manifolds M. As their fibers are the tangent spaces at points $x \in M$ they are vector spaces $V = T_xM$, which have dual spaces $V^* = \operatorname{Hom}_{\mathbb{K}}(V, \mathbb{K})$, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, of all linear functions to their underlying scalar field.
If we assign each point x its dual tangent space, the **cotangent space** T^*_xM we will get the **dual vector bundle**, here the **cotangent bundle**

$$T^*M = (T^*M, M, \pi) = \bigcup_{x \in M} T^*_xM = \bigcup_{x \in M} \operatorname{Hom}_{\mathbb{K}}(T_xM, \mathbb{K}) \tag{34}$$

If we drop the requirement on the fibers $E_x = \pi^{-1}(x)$ to be a vector space, which was motivated by tangent spaces, and substitute it by the requirement, that the fibers are all homeomorphic to another topological space F, then (E, X, π, F) is called a **fiber bundle**. Vector bundles are a special case of fiber bundles. One can express fiber bundles by the short exact sequence

$$F \overset{\iota}{\rightarrowtail} E \overset{\pi}{\twoheadrightarrow} X \tag{35}$$

If we have in addition a continuous operation $(E, G) \to E$ of a topological group, e.g. a Lie group, on the total space E of a fibre bundle, then (E, X, π, F, G) is called a **principal bundle**, if the group operation maps each fiber E_x on itself, i.e. $\pi(xg) = \pi(x)$ for all $x \in E$, $g \in G$, the group operates freely (only $g = 1 \in G$ leaves points in a fiber invariant) and transitive (all points $y \in E_x$ in a fiber can be reached by some $g \in G$). G is called **structure group** of the principal bundle.

4.2 Sections

One can think of a fiber bundle $E = (E, X, \pi, F)$ as a topological base space X to which at each of its points x a fiber E_x is attached and the bundle projection π maps each fiber to its base point. Now we consider the opposite direction: a pairing of a base point with some point of its fiber. This generalizes somehow the concept of the graphs of functions f we draw, which are also a pairing $(x, f(x))$ of base points and certain points in another dimension. In the case of functions, the graphs are part of a Cartesian product, which we don't have here, except for trivial bundles. If we think of a fiber bundle of something similar as the spines of a hedgehog (line bundle), and the bundle projection along the spines to the point where they grow out of the hedgehog, then we are now interested in a cut through all spines.

Definition: Let $E = (E, X, \pi, F)$ be a fiber bundle. A **global section** in E is a continuous function $\sigma : X \to E$ such that for all $x \in X$

$$(\pi \circ \sigma)(x) = \pi(\sigma(x)) = x \tag{36}$$

Thus σ is a right inverse to the bundle projection π. We denote the set of all global sections of E by

$$\Gamma(X, E) = \Gamma(E) \tag{37}$$

If $U \subseteq X$ is an open set, then a continuous function $\sigma : U \to E$ is called a **local section** of E, if it satisfies equation (36) on U, i.e. is a *global* section on (E, U, π, F),.

In case E is a smooth fiber bundle on a smooth manifold X and $\sigma : X \to E$ a smooth function, then it is called a **smooth section** and the set of all smooth sections of E is sometimes denoted by $\Gamma^\infty(E)$.

If $C \subseteq X$ is a compact set, and $\sigma \in \Gamma(X, E)$ a section of a vector bundle E such that $\sigma(x) = (x, 0)$ whenever $x \notin C$, then σ is called a **section with compact support** and we denote the set of all sections with a compact support by $\Gamma_C(X, E) = \Gamma_C(E)$ or $\Gamma_0(X, E) = \Gamma_0(E)$.

4.3 Pullbacks

Assume we have some kind of mapping $m : X \to Y$. This could be literally everything: an operator, a morphism in any category, tensor fields, connections, Lie derivatives on fiber bundles etc. Now if there is an object somehow related to Y, it is natural to ask, whether there is something similar related to X which respects the ways along m and back. It is easy if m is a bijective morphism, but what can be said in general? This leads us to the concepts of pullbacks.

Definition: Let X and Y be topological spaces, $\varphi : X \to Y$ a continuous function and $E = (E, Y, \pi)$ a fiber bundle over Y. Then the fiber bundle defined by

$$\varphi^*E = \{(x, e) \in X \times E \,|\, \varphi(x) = \pi(e)\} \subseteq X \times E \tag{38}$$

equipped with the subspace topology and the projection map

$$\pi^* : \varphi^*E \to X \,,\, \pi^*(x, e) = x \tag{39}$$

is called the **pullback bundle** (φ^*E, X, π^*). Let $\psi : \varphi^*E \to E$ be the projection onto the second factor, then the following diagram commutes:

$$\begin{array}{ccc} \varphi^*E & \xrightarrow{\psi} & E \\ \pi^* \downarrow & & \downarrow \pi \\ X & \xrightarrow{\varphi} & Y \end{array} \tag{40}$$

φ^*E is now a fiber bundle over X. The bundle φ^*E is called the **pullback** of E by φ or **the bundle induced by** φ. The pair (ψ, φ) is a (case of a) **fiber bundle morphism** - ψ is called a **cover** of φ - which are generally defined by the equation (commutative diagramm) (40)

$$\pi \circ \psi = \varphi \circ \pi^* \tag{41}$$

If (U, η) is a local trivialization of E then $(\varphi^{-1}(U), \eta^*)$ is a local trivialization of $\varphi^* E$ where $\eta^*(x, e) = (x, proj_2(\eta(e)))$.

Given a section $\sigma \in \Gamma(Y, E)$ then

$$\varphi^* \sigma := \sigma \circ \varphi \in \Gamma(X, \varphi^* E) \tag{42}$$

is called **pullback section** of σ by φ.

Topological spaces in differential geometry are usually smooth manifolds. In this case one requires φ and the vector bundles also to be smooth.

4.3.1 Smooth Manifolds and Scalar Functions

Let $\varphi : M \to N$ be a smooth function between smooth manifolds and $\nu : N \to \mathbb{R}$ a smooth function on N. Then the pullback $\varphi^* \nu$ of ν by φ is defined by

$$(\varphi^* \nu)(x) = \nu(\varphi(x)) \tag{43}$$

The set $C^\infty(M)$ of smooth scalar functions $\psi : M \to \mathbb{R}$ can be naturally identified with the vector space of smooth sections $\Gamma^\infty(M, M \times \mathbb{R})$ on the vector bundle $\coprod_{x \in M} \mathbb{R} \cong M \times \mathbb{R}$. Thus the **pullback of the smooth scalar function** ψ

$$(\varphi^* \psi)(x) = \psi(\varphi(x)) \tag{44}$$

is the pullback of a smooth section on the smooth vector bundle $M \times \mathbb{R}$.

4.3.2 Multilinear forms

Let $\varphi : V \to W$ be a linear map and F a multilinear form $F : W \times \ldots \times W \to \mathbb{R}$. Then the pullback $\varphi^* F$ of F by φ is defined by

$$(\varphi^* F)(v_1, \ldots, v_n) = F(\varphi(v_1), \ldots, \varphi(v_n)) \tag{45}$$

4.3.3 Cotangents and 1-Forms

To imagine the following example, it is helpful to think of $M = \mathbb{R}^m$, $N = \mathbb{R}^n$ and $J_x(f)$ as the Jacobi-matrix of f at a point $x \in M$.

Let Let $f : M \to N$ be a smooth function between smooth manifolds. Then the differential $J_x(f) = f_* = df = Df$ of f is a function that transforms tangent vectors of M to the tangent vectors of N. It can be viewed as a bundle morphism over M of the tangent bundle TM of M to the pullback bundle f^*TN: for a tangent vector v_x of M it is $J_x(f)(v_x) = w_{f(x)}$ a tangent vector of N, that still *remembers* x.

$$\begin{array}{ccc} TM & \stackrel{J_x(f)=f_*}{\longrightarrow} & f^*TN \\ \pi_M \downarrow & & \downarrow \pi' \\ M & \stackrel{=}{\longrightarrow} & M \end{array} \tag{46}$$

Now the transpose of $f_* = J_x(f) = Df$ maps the corresponding dual spaces, which are the **cotangent bundles** in the opposite direction:

$$J_x(f)^\tau = f_*^\tau \; : \; f^*T^*N \longrightarrow T^*M \tag{47}$$

The sections $\sigma \in \Gamma(N, T^*N)$ are the $1-$**forms** or **Pfaffian forms** on N. Then $(f^*\sigma) = \sigma \circ f \in \Gamma(M, f^*T^*N)$ is the pullback section by f (cp. (42)). A $1-$form on M is then achieved with the help of the above bundle morphism (47) at each point $x \in M$ defined by

$$(f^*\sigma)_x(v_x) := \sigma_{f(x)}(J_x(f)(v_x)) = \sigma_{f(x)}(f_*(v_x)) = \sigma_{f(x)}(D_x f(v_x)) \tag{48}$$

for $v_x \in T_xM$. Tangent vectors as part of a vector field are usually written X instead of v_x. However, one has to be careful not to confuse single tangent vectors with the entire vector field, which often is also written as X. In the latter case, one denotes the points by $p \in M$ instead of $x \in M$ and the tangent vectors as X_p.

4.3.4 Differential forms

Let $f : M \to N$ be a smooth function between smooth manifolds and $\sigma \in \Gamma(N, \wedge^k(T^*N))$ a $k-$form on N, i.e. a section of N to the $k-$fold **outer bundle** of its cotangent bundle. As in the case of $1-$forms (eq. (48)) we define for tangent vectors $X_1, \ldots, X_k \in T_pM$ at a point $p \in M$ a $k-$**fold pullback differential form** on M by

$$(f^*\sigma)_p(X_1, \ldots, X_k) = \sigma_{f(p)}(f_*(X_1 \wedge \ldots \wedge X_k) = \sigma_{f(p)}(d_pf(X_1), \ldots, d_pf(X_k)) \tag{49}$$

The pullback f^* of a differential form has two important compatibility properties

$$f^*(\omega_1 \wedge \omega_2) = f^*(\omega_1) \wedge f^*(\omega_2) \tag{50}$$

$$f^*(d\omega) = df^*(\omega) \tag{51}$$

which makes them an important tool in differential geometry.

4.3.5 Covariant Tensor Fields

Let $f : M \to N$ be a smooth function between smooth manifolds and $\mathcal{F}$ a **covariant tensor field** of rank $(0,k)$ on N, which is a section of the **tensor bundle** on N whose fiber at $y \in N$ is the vector space of multinlinear $k-$forms

$$F \,:\, T_y(N) \times \ldots \times T_y N \longrightarrow \mathbb{K} \tag{52}$$

Now the pullback of $\mathcal{F}$ by f is the $(0,k)-$tensor field $f^*\mathcal{F}$ on M defined by $(p \in M \ ,\ X_i \in T_pM)$

$$(f^*\mathcal{F})_p(X_1,\ldots,X_k) = \mathcal{F}_{f(p)}(f_*(X_1 \otimes \ldots \otimes X_k) = \mathcal{F}_{f(p)}(d_pf(X_1),\ldots,d_pf(X_k)) \tag{53}$$

4.3.6 Diffeomorphisms

Let $f : M \to N$ be a diffeomorphism between smooth manifolds, i.e. f has an inverse function f^{-1}. Then the linear map

$$J_p(f) = f_* = d_pf \in GL(T_pM, T_{f(p)}N) \tag{54}$$

can be inverted by

$$J_p^{-1}(f) = f_*^{-1} = (d_pf)^{-1} \in GL(T_{f(p)}N, T_pM)) \tag{55}$$

4.3.7 Connections - Covariant Derivatives

Let $f : M \to N$ be a smooth function between smooth manifolds and ∇ a **connection** on a vector bundle E over N, then there is a **pullback connection** $f^*\nabla$ on f^*E over M defined by

$$(f^*\nabla)_X(f^*\sigma) = f^*(\nabla_{df(X)}\sigma) \tag{56}$$

I will return to them in section 6.2.

4.3.8 Dual Operators

Let (E, M, π_M) and (F, N, π_N) be two vector bundles and $\varphi : M \to N$ a continuous function with pullback $\varphi^* : F \to E$. Then the **dual operator** $\varphi_* : E \to F$ is the **pushforward** of φ. An example is the tangent bundle morphism.

4.4 Pushforwards

4.4.1 Tangent Bundles

Let $\varphi : M \to N$ be a smooth map on smooth manifolds. Then the differential $\varphi_* = D\varphi$ defines a **pushforward** (and a cover) of φ of the corresponding tangent bundles:

$$\begin{array}{ccc} TM & \stackrel{\varphi_*=D\varphi}{\longrightarrow} & TN \\ \pi_M \downarrow & & \downarrow \pi_N \\ M & \stackrel{\varphi}{\longrightarrow} & N \end{array} \tag{57}$$

The tangent vectors can be seen as directional derivatives. The maps are given by $\varphi_* : TM \to TN$

$$\varphi_* : T_pM \to T_{\varphi(p)}N \tag{58}$$

$$(\varphi_*(v))(f) = v(f \circ \varphi) \tag{59}$$

with $p \in M, v \in T_pM, f \in C^\infty(N)$. Unfortunately there isn't a general convention of how to write this pushforward. It depends on the context (bundles, category theory, tangent bundels, tangent spaces, differential geometry, physics etc.) and emphasis (linearity, locality, functions and curves). It varies from author to author. Other notations which are frequently used (for what I started with $J_p(\varphi)v$):

$$\varphi_*(p, v) = \varphi'_p(v) = \varphi'(p)v = D\varphi_p(v) = D_p\varphi(v) = d\varphi_p(v) = d_p\varphi(v) = T_p\varphi(v) \tag{60}$$

Let $\gamma(t) : I \to M$ be a (smooth) curve through p on the manifold M and $v = \dot{\gamma}(t_0)$ its tangent vector at $p = \gamma(t_0)$. Then $\overline{\gamma}(t) = (\varphi \circ \gamma)(t)$ defines a corresponding curve on the codomain manifold N through $\varphi(p)$ and we have

$$\varphi_*(v) = d\varphi_p(v) = \dot{\overline{\gamma}}(t_0) = \frac{d}{dt}\bigg|_{t=t_0} (\varphi \circ \gamma)(t) \tag{61}$$

or in case the tangent vectors are defined by derivations (cp. 6.4)

$$\varphi_*(X)(f) = d\varphi_p(X)(f) = X(f \circ \varphi) \tag{62}$$

For a representation in coordinates (by charts, cp. 3.1) let us consider local coordinates $(x_1, \ldots, x_m)$ on M in an open chart U around $p \in M$ and $(y_1, \ldots, y_n)$ on N in an open chart V around $\varphi(p) \in N$. For the vectors $v \in T_pM$ and $w = \varphi_*(v) \in T_{\varphi(p)}N$ we thus get

$$v = \sum_{i=1}^{m} v^i \frac{\partial}{\partial x^i} \, , \; w = \sum_{j=1}^{n} w^j \frac{\partial}{\partial y^j} \tag{63}$$

$$w^j = \sum_{i=1}^{m} \frac{\partial \widehat{\varphi}_j}{\partial x^i} v^i \, , \, \varphi = (\widehat{\varphi}^1, \ldots, \widehat{\varphi}^n) \text{ w.r.t. charts} \tag{64}$$

$$d\varphi_p \left(\frac{\partial}{\partial x^i} \right) = \frac{\partial \widehat{\varphi}^j}{\partial x^i} \frac{\partial}{\partial y^j} \tag{65}$$

4.4.2 Vector Fields

Let $\varphi : M \to N$ be a smooth map on smooth manifolds and $\sigma \in \Gamma(M, TM)$ a section of the vector bundle (E, M, π_M, TM) which is a vector field $X = \sigma$ on M. Then the pullback along φ is a vector field on M, i.e. a section in $\Gamma(M, \varphi^*TN)$ with $(\varphi^*(Y))_p = (Y \circ \varphi)_p$ (cp. (42)). If we apply the differential $D_p(\varphi)(X) = \varphi_*(p, X)$ pointwise (cp. (58) to (60)) we get the pushforward along φ by $\varphi_*(p, X) = (D_p\varphi)X$, i.e. $\varphi_*(X) \in \Gamma(M, \varphi^*TN)$.

Any vector field $Y \in \Gamma(N, TN)$ defines a pullback section $\varphi^*Y \in \Gamma(M, \varphi^*TN)$ with $(\varphi^*Y)_p = (Y \circ \varphi)(p) = Y_{\varphi(p)}$. A vector field $X \in \Gamma(M, TM)$ and a vector field $Y \in \Gamma(N, TN)$ are $\varphi-$**related**, if

$$d\varphi(X) = D\varphi(X) = \varphi_*(X) = \varphi^*(Y) = Y \circ \varphi \tag{66}$$

that is for all points $p \in M$ holds $D_p(\varphi)(X) = Y_{\varphi(p)}$. In case φ is a diffeomorphism, we even get

$$Y_q = \varphi_*(X_{\varphi^{-1}(q)}) \tag{67}$$

Assume X, Y are vector fields such that $\varphi_*(X) = d\varphi(X)$ and $\varphi_*(Y) = d\varphi(Y)$ are $\varphi-$related, $f : N \to \mathbb{R}$ and $q = \varphi(p) \in N$. By multiple application of (59) we get

$$\begin{aligned}
\varphi_*([X, Y])(f(q)) &= \varphi_*(X \circ Y - Y \circ X) f(q) \\
&= (X \circ Y - Y \circ X)(f \circ \varphi(p)) \\
&= X(Y(f \circ \varphi)(p)) - Y(X(f \circ \varphi)(p)) \\
&= X(\varphi_*(Y)(f(\varphi(p)))) - Y(\varphi_*(X)(f(\varphi(p)))) \\
&= \varphi_*(X)(\varphi_*(Y)(f(q))) - \varphi_*(Y)(\varphi_*(X)(f(q))) \\
&= [\varphi_*(X), \varphi_*(Y)](f(q))
\end{aligned} \tag{68}$$

and the commutator (Lie bracket) of $\varphi-$related vector fields is again a $\varphi-$related vector field.

For two smooth functions $\varphi : M \to N$, $\psi : N \to P$ the chain rule holds for the pushforwards of $\psi \circ \varphi : M \to P$

$$(\psi \circ \varphi)_* = \psi_* \circ \varphi_* \tag{69}$$

5 Lie Derivatives

A Lie derivative is in general the differentiation of a tensor field along a vector field. This allows several applications, since a tensor field includes a variety of instances, e.g. vectors, functions or differential forms. In case of vector fields we additionally get a Lie algebra structure. This is, although formulated in a modern language, the actual reason why Lie algebras have been considered in the first place: as the tangent bundle of Lie groups which are themselves the invariants which appear as symmetry groups in the standard model of particle physics or more generally in the famous theorem of Emmy Noether, which is actually a theorem about invariants of differential equations (see [9],[10]). The Jacobi identity, e.g., which together with anti-commutativity defines a Lie algebra is simply a manifestation of the Leibniz rule of differentiation.

5.1 Definitions

"*Let X be a vector field on a manifold M. We are often interested in how certain geometric objects on M, such as functions, differential forms and other vector fields, vary under the flow* $\exp(\varepsilon X)$ *induced by X. The Lie derivative of such an object will in effect tell us its infinitesimal change when acted on by the flow. ... More generally, let σ be a differential form or vector field defined over M. Given a point $p \in M$, after 'time' ε it has moved to* $\exp(\varepsilon X)$ *with its original value at p. However, $\sigma|_{\exp(\varepsilon X)p}$ and $\sigma|_p$, as they stand are, strictly speaking, incomparable as they belong to different vector spaces, e.g. $TM|_{\exp(\varepsilon X)p}$ and $TM|_p$ in the case of a vector field. To effect any comparison, we need to 'transport' $\sigma|_{\exp(\varepsilon X)p}$ back to p in some natural way, and then make our comparison. For vector fields, this natural transport is the inverse differential*

$$\Phi_\varepsilon^* \equiv d\exp(-\varepsilon X) : TM|_{\exp(\varepsilon X)p} \to TM|_p \tag{70}$$

whereas for differential forms we use the pullback map

$$\Phi_\varepsilon^* \equiv \exp(\varepsilon X)^* : \wedge^k T^*M|_{\exp(\varepsilon X)p} \to \wedge^k T^*M|_p \tag{71}$$

This allows us to make the general definition of a Lie derivative." [5]

The exponential function comes into play here, because [5] is about the theory of Lie Groups G, and the exponential map is the natural function that transports objects in the tangent bundle of a Lie group G to those on the manifold G, a form of integration if you like. It is the same reason as we often use an Ansatz with the exponential function to solve differential equations, that are also statements on tangent bundles.

Definition [5]: The **Lie derivative** along a vector field X of a vector field or differential form ω at a point $p \in M$ is given by

$$\mathcal{L}_X(\omega)_p = X(\omega)|_p = \lim_{t \to 0} \frac{1}{t} \left(\Phi_t^*(\omega|_{\exp(tX)_p}) - \omega|_p \right) = \left. \frac{d}{dt} \right|_{t=0} \Phi_t^*(\omega|_{\exp(tX)_p}) \tag{72}$$

In this form it is obvious that the Lie derivative is a directional derivative and another form of the equation (1) which is the leitmotif of all the concepts presented here and which is the crucial part of all differentiation processes: the infinitesimal change in an object's behavior along a certain direction (or all). The equivalence of this definition to those given below must be shown and can be found, e.g. in [5].

Lie Derivative of a Function. Let $f : M \to \mathbb{R}$ be a smooth map on a smooth manifolds M. The **Lie derivative** of f along the smooth vector field X is the directional derivative at a point $p \in M$:

$$\mathcal{L}_X f(p) = X(p)(f) = d_p f(X(p)) \tag{73}$$

In local coordinates $(x_1, \ldots, x_n) : U \subseteq M \to \mathbb{R}^n$ with $X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}$ it's

$$\mathcal{L}_X f(p) = \sum_{i=1}^n X_i(p) \frac{\partial f}{\partial x_i}(p) \tag{74}$$

Lie Derivative of a Vector Field. Let X, Y be two vector fields on a smooth manifold M and $X(\gamma(t))$ a flow of X (see (28)). The **Lie derivative** from Y along X is defined by

$$\mathcal{L}_X Y = \left. \frac{d}{dt} \right|_{t=0} (X^*(\gamma(t))Y) \tag{75}$$

Furthermore

$$\mathcal{L}_X Y = [X, Y] = X \circ Y - Y \circ X \tag{76}$$

Lie Derivative of a Tensor Field. Let T be a tensor field and X a vector field with a local flow $\xi(\gamma(t))$. Then the **Lie derivative** of T along X is given by

$$\mathcal{L}_X T = \left. \frac{d}{dt} \right|_{t=0} (\xi^*(\gamma(t))T) \tag{77}$$

There is also an axiomatic and algebraic approach ([20], (73), (81), (82),(6.3))

$(A1)$ $\mathcal{L}_X f = X(f)$

$(A2)$ $\mathcal{L}_X(S \otimes T) = \mathcal{L}_X S \otimes T + S \otimes \mathcal{L}_X T$

$(A3)$ $\mathcal{L}_X(T(Y_1, \ldots, T_k)) = (\mathcal{L}_X T)(Y_1, \ldots, Y_k) + \sum T(Y_1, \ldots, (\mathcal{L}_X Y_i), \ldots, Y_k)$

$(A4)$ $\mathcal{L}_X \circ d = d \circ \mathcal{L}_X$

Lie Derivative of a Differential Form. To define a directional derivative of differential forms more detailed than in (71) we have to take the arguments into consideration. Let M be a manifold, X a vector field on M and $\omega \in \wedge^k(M)$ a $k-$form, i.e. for all $p \in M$ we have an alternating $k-$linear map $\omega(p) : (T_pM)^k \to \mathbb{R}$. We first define an interior product of X and ω as the $(k-1)-$form $\iota_X\omega$ given by

$$(\iota_X\omega)(X_1, \ldots, X_{k-1}) = \omega(X, X_1, \ldots, X_{k-1}) \tag{78}$$

which is also called the **contraction** of ω with X. The multilinear map $\iota_X : \wedge^k(M) \to \wedge^{k-1}(M)$ has the property that for two differential forms ω, σ

$$\iota_X(\omega \wedge \sigma) = (\iota_X\omega) \wedge \sigma + (-1)^{k-1}\omega \wedge (i_X\sigma) \tag{79}$$

which is the graded form of the Leibniz rule on Graßmann algebras, i.e. the version that takes alternating into account. For a differentiable function on M we get

$$\iota_{fX}\omega = f\iota_X\omega \text{ and } \mathcal{L}_X f = \iota_X df \tag{80}$$

For a general differential form ω we define the **Lie derivative** as

$$\mathcal{L}_X\omega = (\iota_X \circ d + d \circ \iota_X)\,\omega \tag{81}$$

Two important properties of the Lie derivative are

$$d\mathcal{L}_X\omega = \mathcal{L}_X d\omega \tag{82}$$

$$\mathcal{L}_{fX}\omega = f\mathcal{L}_X\omega + df \wedge \iota_X\omega \tag{83}$$

5.2 Left-Invariant Vector Fields and GL_n

Let us now consider an analytic group G, a **Lie group**, i.e. an analytic topological manifold G endowed with analytic multiplication and inversion. This group acts differentiable on itself via left multiplication $L_g : G \to G$ by $L_g(h) = gh\,,\ (g, h \in G)$.

Definition: A vector field X on G is called **left-invariant** if for all $g, h \in G$

$$dL_g(X_h) = X_{L_g(h)} = X_{gh} \tag{84}$$

The vector space of all left-invariant vector fields of G is called the **Lie algebra** $\mathfrak{g}$ of G. Since

$$X_g = dL_g(X_e) \tag{85}$$

each vector field is already determined by its value at the identity element $e \in G$. If we conversely have a vector field at e that satisfies (85) we get by the chain rule

$$dL_g(X_h) = dL_g(dL_h(X_e)) = d(L_g \circ L_h)(X_e) = dL_{gh}(X_e) = X_{gh} \tag{86}$$

the left-invariance of X. Thus $\mathfrak{g} \cong T_eG$ and $\dim \mathfrak{g} = \dim G$. By (68) we have $dL_g[X,Y] = [dL_gX, dL_gY] = [X,Y]$ for left-invariant vector fields.

Definition: A **Lie algebra** $\mathfrak{g}$ on G is a vector space with a bilinear multiplication

$$[.,.] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \tag{87}$$

that is anti-commutative (cp. (6.3))

$$[X,X] = 0 \tag{88}$$

and satisfy the **Jacobi-Identity** (cp. (A2), **Leibniz rule** and (81) in section 5.1)

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0 \tag{89}$$

Example. Let us consider the **general linear group** $G = GL(n,\mathbb{R})$ of all regular real $(n,n)-$matrices and $A = (a_{ij})$ be such a matrix. Then the tangent space at the identity element $e = I \in G$ is given by

$$V_A|_I = \sum_{i,j} a_{ij} \left.\frac{\partial}{\partial x_{ij}}\right|_I \tag{90}$$

Here we denote the tangent bundle by V to avoid confusion with group elements $X = (x_{ij}), Y = (y_{ij})$. The matrix entries of $L_Y(X) = YX$ are

$$\sum_{k=1}^{n} y_{ik}x_{kj} = Y^i{}_k X^k{}_j \tag{91}$$

and with (85)

$$
\begin{aligned}
V_A|_Y &= dL_Y(V_A|_I) \\
&= dL_Y \sum_{i,j} a_{ij} \frac{\partial}{\partial x_{ij}} \\
&= \sum_{l,m} \sum_{i,j} a_{ij} \, (\sum_{k} \frac{\partial}{\partial x_{ij}} \, y_{lk} x_{km}) \, \frac{\partial}{\partial x_{lm}} \\
&= \sum_{l,m} \sum_{i,j} a_{ij} \, \delta_{mj} \, y_{li} \, \frac{\partial}{\partial x_{lm}} \\
&= \sum_{l,m} \sum_{i} y_{li} a_{im} \frac{\partial}{\partial x_{lm}}
\end{aligned}
\tag{92}
$$

It can now be shown by direct matrix multiplications, that

$$
[V_A, V_B] = V_A \circ V_B - V_B \circ V_A = V_{[AB-BA]} = V_{[A,B]} \tag{93}
$$

This means that

$$
\mathfrak{gl}(n, \mathbb{R}) = \mathfrak{g} = T_e G = T_e GL(n, \mathbb{R}) = \{ \, \mathbb{M}(n, \mathbb{R}) \, | \, [A, B] = AB - BA \, \} \tag{94}
$$

Traditionally the basis point $e \in G$ of the tangent bundle $T_e G$ is denoted by the neutral group element, even in the case of matrix (Lie) groups G, where $e = I$ the identity map. So as a basis point it's e and in coordinates it's I. For the same reason (tradition) are Lie algebras denoted by letters in fraktur types.

Let us assume $G = SL(n, \mathbb{R})$ the **special linear group** of all regular real $(n, n)-$matrices with determinant 1. If we evaluate the differential of the determinant at $e = I$ we get

$$
\begin{aligned}
d_e \det X &= \sum_{\pi} \operatorname{sgn}(\pi) d_e (\prod_{k} x_{k\pi(k)}) \\
&= \sum_{\pi} \operatorname{sgn}(\pi) \sum_{i,j} \left. \frac{\partial}{\partial x_{ij}} \right|_I (\prod_{k} x_{k\pi(k)}) \\
&= \sum_{k} \left. \frac{\partial}{\partial x_{kk}} \right|_I
\end{aligned}
\tag{95}
$$

because evaluation at $e = I = (\delta_{ij})$ leaves only the diagonal $(x_{11}, \ldots, x_{nn})$ different from 0 which evaluates to 1 after applying the product rule. In the general case we calculate the push forward $\det_*$ of $\det : SL(n,\mathbb{R}) \to \mathbb{R}$. Sine it is a constant function, we get by the previous result

$$\begin{aligned} 0 &= \det_*(V_A|_I)(X) \\ &= d_e \det(V_A)(X) \\ &= \sum_k a_{kk} \left.\frac{\partial}{\partial x_{kk}}\right|_I (X) \\ &= \operatorname{tr} A \end{aligned} \tag{96}$$

This means

$$\mathfrak{sl}(n,\mathbb{R}) = \mathfrak{g} = T_eG = T_eSL(n,\mathbb{R}) = \{\, \mathfrak{gl}(n,\mathbb{R}) \mid \operatorname{tr} A = 0 \,\} \tag{97}$$

Unitary matrices ($U^\dagger = U^{-1}$) have a Lie algebra determined by

$$U^\dagger = d_eU^\dagger = d_eU^{-1} = -U$$

i.e. skew-Hermitian matrices, so

$$\mathfrak{su}(n,\mathbb{C}) = \mathfrak{g} = T_eG = T_eSU(n,\mathbb{C}) = \{\, \mathfrak{sl}(n,\mathbb{R}) \,|\, U = -U^\dagger \,\} \tag{98}$$

6 Derivatives in Other Contexts

6.1 Material Derivative

The material derivative is a special case of a derivative in order to describe the flow of fluids or gases. It is more of a special tool for these currents rather than a special concept of a differentiation process. The **material derivative** of a scalar or vectorial field $\Phi(x,t)$ is defined by

$$D_v\Phi = \frac{d_v}{dt}\Phi = \frac{\partial\Phi}{\partial t} + (v\cdot\nabla)(\Phi) = \left(\frac{\partial}{\partial t} + v_x\frac{\partial}{\partial x} + v_x\frac{\partial}{\partial x} + v_x\frac{\partial}{\partial x}\right)(\Phi) \tag{99}$$

where v represents the velocity of the flow at point x and time t. The first summand is the **local behavior** in time at a fixed point, the second is the **convective change** of a particle in the flow. Wikipedia [20] names various other names for the material derivative:

- **Euler operator**
- **advective derivative**

- **convective derivative**
- **derivative following the motion**
- **hydrodynamic derivative**
- **Lagrangian derivative**
- **substantial derivative**
- **substantive derivative**
- **Stokes derivative**
- **total derivative**

Assume a temperature field $\Phi(x, y, t)$ on the two dimensional surface of a lake, that gets warmer in a certain direction (SW to NE), e.g. by warming inflows, and which is additionally warming up by sunshine (form ([19])).

$$\Phi(x, y, t) = 300\,K \;+\; (1\,K/m)\,x \;+\; (2\,K/m)\,y \;+\; (3\,K/s)\,t$$

If we assume a current $v = (3, 1)\,\mathrm{m/s}$, then

$$D_v\Phi = \frac{d_v\Phi}{dt} = (\frac{\partial}{\partial t} + v \cdot \nabla)(\Phi) = 3\,\mathrm{K/s} + \begin{bmatrix}3\\1\end{bmatrix} m/s \cdot \begin{bmatrix}1\\2\end{bmatrix} K/m = 8\,K/m$$

An observer in a boat, floating with the current experiences an additional convective decrease in temperature by $5\,K/s$.

Important versions of this derivative are (among many more):

Navier-Stokes equations. Compressible fluids.

$$\rho \cdot \vec{v} = \rho \left(\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} \right) = -\nabla p + \mu \Delta v + (\lambda + \mu)\nabla(\nabla \cdot \vec{v}) + \vec{f} \tag{100}$$

Navier-Stokes equations. Incompressible fluids.

$$\rho \cdot \vec{v} = \rho \left(\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} \right) = -\nabla p + \mu \Delta v + \vec{f} \tag{101}$$

Euler Equation (Current in Frictionless Elastic Fluids).

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} + \frac{1}{\rho}\,\mathrm{grad}(\vec{p}) = \vec{k} \tag{102}$$

Cauchy-Euler Law of Movement.

$$\rho \vec{a} = \rho \vec{k} + \mathrm{div}\,\boldsymbol{\sigma} \tag{103}$$

6.2 Connection - Covariant Derivative

In this section I will essentially restrict myself to real affine connections, as the subjects as a whole would lead too far. The basic idea is to get a hold on curvatures of differentiable manifolds. In an ordinary Euclidean space one would simply calculate the second derivative. On manifolds, however, the second derivative would mean a limit of a difference quotient of tangent vectors $\gamma'(t_1)$ and $\gamma'(t_2)$ of a curve $\gamma(t)$ which at different points belong to different tangent fibers. So to succeed, one has to *connect* the fibers somehow. This leads to the concept (and name) of connections and parallel transport of tangent vectors. Formally we consider a smooth real manifold M, its tangent bundle TM and a vector bundle (E, M, π) on M and sections of TM and E (see 4.2) for we want to define a directional derivative of a vector field along a tangent vector.

Definition: A **connection** on the bundle (E, M, π) is a function

$$\begin{aligned} \nabla : \ \Gamma(TM) \times \Gamma(E) &\longrightarrow \Gamma(E) \\ (X, \sigma) &\longmapsto \nabla_X(\sigma) \end{aligned} \tag{104}$$

such that the following conditions hold with
$X, Y \in \Gamma(TM)$, $f, g \in C^\infty(M)$, $\mu, \nu \in \mathbb{R}$, $\sigma, \tau \in \Gamma(E)$

$\mathbb{C}^\infty(M)-$**linearity in the first argument.**

$$\nabla_{fX+gY}(\sigma) = f \cdot \nabla_X(\sigma) + g \cdot \nabla_Y(\sigma) \tag{105}$$

$\mathbb{R}-$**linearity in the second argument.**

$$\nabla_X(\mu\sigma + \nu\tau) = \mu\nabla_X(\sigma) + \nu\nabla_X(\tau) \tag{106}$$

Leibniz rule.

$$\nabla_X(f\sigma) = D_X f \cdot \sigma + f \cdot \nabla_X \sigma \tag{107}$$

The connection ∇ is called a **linear connection or affine connection**, if $(E, M, \pi) = TM$. A positive definite and symmetric bilinear form g on TM is called a **Riemann metric** or a **metric tensor**, if

$$g : TM \times TM \longrightarrow \mathbb{R} \tag{108}$$

depends smoothly on $p \in M$, i.e. $p \mapsto g_p(X_p, Y_p)$ is a smooth function. A **Riemannian connection** is an affine connection which additionally satisfies the following conditions:

Lie multiplication.

$$\nabla_X Y - \nabla_Y X = [X, Y] \tag{109}$$

Metric compatibility.

$$X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \tag{110}$$

If (M, g) is a Riemannian manifold, then there is exactly one affine connection on M, which is a Riemannian connection. Are M a three dimensional orientable Riemannian manifold and X, Y, Z vector fields, then

$$\nabla_Z(X \times Y) = (\nabla_Z X) \times Y + X \times (\nabla_Z Y) \tag{111}$$

for the corresponding Riemannian connection, i.e. the product rule holds for cross products. Furthermore there is a unique $\mathbb{R}$-linear mapping

$$d_\nabla : \wedge^n(M, TM) \longrightarrow \wedge^{n+1}(M, TM) \ \ , \ \ n \in \mathbb{N} \tag{112}$$

with the properties

$$d_\nabla X(Y) = D_Y(X) \tag{113}$$

and d_∇ is a derivation on $\wedge(M, TM)$

$$d_\nabla(\omega_1 \wedge \omega_2) = d_\nabla\omega_1 \wedge \omega_2 + (-1)^n \omega_1 \wedge d_\nabla\omega_2 \tag{114}$$

d_∇ is called the corresponding **exterior derivative**. [4]

6.3 Exterior Derivative or Cartan Derivative

We've already met exterior derivatives (5.1 (79) and 6.2 (114)). I personally like the term boundary operator for it - **coboundary operator** to be exact - because it emphasizes the topological nature of exterior derivatives as the homomorphism in **cochain complexes**. It represents the approach to calculate topological objects by algebraic means and results in, e.g. the **de Rahm cohomology**. Exterior derivatives can also be seen geometrically by their close relationship to Lie multiplication (113) and the Riemannian metric.

Definition: Let $U \subseteq M$ be a open set in an $n-$dimensional smooth manifold and $\wedge(M) = \wedge(M, TM)$ the algebra of differential forms on M. The (existing and unique) function $d : \wedge^k(U) \longrightarrow \wedge^{k+1}(U)$, $n \in \mathbb{N}_0$, is called **exterior derivative** or **Cartan derivative** if it has the following properties

$$d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^k \omega_1 \wedge d\omega_2 \text{ for all } \omega_1 \in \wedge^k(U)\,,\, \omega_2 \in \wedge^l(U) \tag{115}$$

$$df = Df \text{ the total differential of } f \in C^\infty(U) \tag{116}$$

$$d^2 = 0 \tag{117}$$

$$d(\omega|_U) = (d\omega)|_U \text{ for all } \omega \in \wedge^k(V) \text{ and } U \subseteq V \subseteq M \text{ open} \tag{118}$$

The exterior derivative can be expressed by the explicit formula

$$\begin{aligned} d\omega(X_0,\ldots,X_k) = \sum_{i=0}^{k}(-1)^i X_i(\omega(X_0,\ldots,\hat{X}_i,\ldots,X_k)) + \\ \sum_{i<j}(-1)^{i+j}\omega([X_i,X_j],X_0,\ldots,\hat{X}_i,\ldots,\hat{X}_j,\ldots,X_k) \end{aligned} \tag{119}$$

For the pullback of a smooth function $f : M \to N$ we remind on the important equations (sec. 4.3.4, eq. (50),(51))

$$f^* : \Gamma(N, T^*N) = \wedge N \longrightarrow \wedge M = \Gamma(M, T^*M) \tag{120}$$

$$f^*(\omega_1 \wedge \omega_2) = f^*(\omega_1) \wedge f^*(\omega_2) \tag{121}$$

$$f^*(d\omega) = df^*(\omega) \tag{122}$$

For the corresponding dual operator, the **boundary operator** on the corresponding chain complex or **coderivative** $\delta : \wedge^k M \to \wedge^{k-1} M$ on a Riemannian manifold (M, g) we have for $\omega_1, \omega_2 \in \wedge(M)$ the adjoint equations

$$\delta^2 = 0 \tag{123}$$

$$g(d\omega_1, \omega_2) = g(\omega_1, \delta\omega_2)) \tag{124}$$

δ is sometimes also denoted by d^* to stress the duality to d. The operator $d\delta + \delta d$ is called **Laplace-Beltrami differential operator** [4],[20] or **Hodge-Laplace-Operator** [19].

There are many formulas which connects these two operators to other differential operators and Riemannian metric. And it is not by accident that the equation (122) already looks like Stokes' theorem. For an example on how to actually calculate with exterior products and derivations, see [17].

6.4 Derivations

Derivations are essentially linear functions, which obey the Leibniz rule. This already shows, we need an algebra $\mathcal{A}$ to define them and anyone will do.

Definition: A **derivation** on an algebra $\mathcal{A}$ over a field $\mathbb{F}$ is a $\mathbb{F}-$linear map d that satisfies for all $a, b \in \mathcal{A}$

$$d(a \circ b) = d(a) \circ b + a \circ d(b) \tag{125}$$

If M is a smooth manifold, then the (tangent) vector field on M are the $\mathbb{R}-$derivations of $C^\infty(M)$. [4].

We often deal with Lie algebras in our context, so a derivation of a Lie algebra $\mathfrak{g}$ becomes

$$d([X,Y]) = [dX,Y] + [X,dY] \tag{126}$$

For the derivations $\operatorname{ad} Z : \mathfrak{g} \to \mathfrak{g}$ defined by the left multiplication in $\mathfrak{g}$, i.e. $\operatorname{ad} Z(X) := [Z,X]$ equation (126) becomes the **Jacobi identity**

$$\begin{aligned} [Z,[X,Y]] &= \operatorname{ad} Z([X,Y]) \\ &= [\operatorname{ad} Z(X),Y] + [X,\operatorname{ad} Z(Y)] \\ &= [[Z,X],Y] + [X,[Z,Y]] \end{aligned} \tag{127}$$

Derivations play a central role in the cohomology theory of Lie algebras (cp. (119)). $\operatorname{ad} Z$ are called the **inner derivations** of $\mathfrak{g}$ which build an **ideal** in the algebra $\operatorname{Der}(\mathfrak{g})$ of all derivations of $\mathfrak{g}$ in the Lie algebra $\mathfrak{gl}(\mathfrak{g})$ of the general linear group of $\mathfrak{g}$ as

$$[d,\operatorname{ad} Z](X) = d[Z,X] - [Z,dX] = [dZ,X] = \operatorname{ad}(dZ(X)) \tag{128}$$

The mapping $X \mapsto \operatorname{ad} X$ (again by the Jacobi identity) is a **Lie algebra homomorphism**

$$\mathfrak{g} \longrightarrow \mathfrak{gl}(\mathfrak{g}) \tag{129}$$

and therefore defines a representation of $\mathfrak{g}$, so ad is called the **adjoint representation** of $\mathfrak{g}$. This refers to the corresponding names given in the Lie groups: $\iota_z(x) = zxz^{-1}$ is called an **inner automorphism** of a group G. If G is a Lie group this induces a natural automorphism of the Lie algebra $\mathfrak{g}$ of smooth (analytic) vector fields of G by $\operatorname{Ad}(y)(X) = yXy^{-1}$, the **adjoint representation** of G. Both are related by the equation [7]

$$\exp \circ \operatorname{ad} = \operatorname{Ad} \circ \exp \tag{130}$$

For commuting $X, Y \in \mathfrak{g}$ we get

$$\exp X \cdot \exp Y = \exp(X+Y) \tag{131}$$

and for non commuting the **Campbell-Baker-Hausdorff formula**

$$\begin{aligned}
\exp X \cdot \exp Y = \exp(&X + Y \\
&+ \frac{1}{2}[X, Y] \\
&+ \frac{1}{12}\left([X, [X, Y]] + [Y, [Y, X]]\right) \\
&- \frac{1}{24}[Y, [X, [X, Y]]] \\
&- \frac{1}{720}\left([Y, [Y, [Y, [Y, X]]]] + [X, [X, [X, [X, Y]]]]\right) \\
&+ \frac{1}{360}\left([X, [Y, [Y, [Y, X]]]] + [Y, [X, [X, [X, Y]]]]\right) \\
&+ \frac{1}{120}\left([Y, [X, [Y, [X, Y]]]] + [X, [Y, [X, [Y, X]]]]\right) \\
&+ \cdots)
\end{aligned} \tag{132}$$

In other cases, exterior derivatives (cp.(115)), differential forms (cp.(79)), we have seen another form of derivations , derivations of **graded algebras**, e.g. the $\mathbb{Z}_2-$graded Lie superalgebras.

Definition: A **anti-derivation** on a graded algebra $\mathcal{A}$ is a linear map

$$d(a \circ b) = d(a) \circ b + (-1)^{|a|)} a \circ d(b) \quad (\, a, b \in \mathcal{A}\,) \tag{133}$$

where $|a|$ denotes the grade of $a \in \mathcal{A}$.
The multiplication in a **Lie superalgebra** is given by the equations

Super skew-symmetry.

$$[X, Y] = -(-1)^{|X||Y|}[Y, X] \tag{134}$$

Super Jacobi identity

$$(-1)^{|X||Z|}[X, [Y, Z]] + (-1)^{|Y||X|}[Y, [Z, X]] + (-1)^{|Z||Y|}[Z, [X, Y]] = 0 \tag{135}$$

7 Important Theorems - biased, of course

7.1 Implicit Function Theorem [1]

7.1.1 Jacobi Matrix (Chain Rule)

Let $(x_0, y_0) \in U_1 \times U_2 = \{x \in \mathbb{R}^k \,|\, ||x - x_0|| < \varepsilon_1\} \times \{y \in \mathbb{R}^m \,|\, ||y - y_0|| < \varepsilon_2\}$ and $f : U_1 \times U_2 \to \mathbb{R}^m$ a function with $f(x_0, y_0) = 0$ which is totally differentiable at (x_0, y_0) such that the $(m \times m)-$matrix $\frac{\partial f}{\partial y}(x_0, y_0)$ is invertible.

In addition let $g : U_1 \to \mathbb{R}^m$ be a continuous function with $g(x_0) = y_0$ and $g(U_1) \subseteq U_2$ and $f(x, g(x)) = 0$ for all $x \in U_1$

Then g is differentiable at x_0 and for the Jacobi matrices

$$\frac{\partial g}{\partial x} = \left(\frac{\partial g_i}{\partial x_j}\right)_{\substack{1 \le i \le m \\ 1 \le j \le k}}, \quad \frac{\partial f}{\partial x} = \left(\frac{\partial f_i}{\partial x_j}\right)_{\substack{1 \le i \le m \\ 1 \le j \le k}}, \quad \frac{\partial f}{\partial y} = \left(\frac{\partial f_i}{\partial y_j}\right)_{\substack{1 \le i \le m \\ 1 \le j \le m}} \tag{136}$$

the following equation holds:

$$\frac{\partial g}{\partial x}(x_0) = -\left[\frac{\partial f}{\partial y}(x_0, y_0)\right]^{-1} \cdot \frac{\partial f}{\partial x}(x_0, y_0) \quad \text{or} \quad \frac{\partial f}{\partial x}(x_0, y_0) + \frac{\partial f}{\partial y}(x_0, y_0) \cdot \frac{\partial g}{\partial x}(x_0) = 0 \tag{137}$$

7.1.2 Implicit Function

Let $f : U_1 \times U_2 \to \mathbb{R}^m$ be continuous differentiable on open sets $U_1 \subseteq \mathbb{R}^k$, $U_2 \subseteq \mathbb{R}^m$, i.e. $f \in C^1(U_1 \times U_2)$, and $f(x_0, y_0) = 0$ such that the $(m \times m)-$matrix $\frac{\partial f}{\partial y}(x_0, y_0)$ is invertible.

Then there are open neighborhoods $V_i \subseteq U_i$ of (x_0, y_0) and a continuous function $g : V_1 \to V_2$, i.e. $g \in C^0(V_1 \times V_2)$, such that $f(x, g(x)) = 0$ for all $x \in V_1$. For a point $(x, y) \in V_1 \times V_2$ with $f(x, y) = 0$, it follows that $y = g(x)$.

7.2 Cauchy's Integral Formula [2]

Let $f : U \to \mathbb{C}$ be a holomorphic function on the open set U and $A \subseteq U$ compact with a smooth boundary and z_0 a inner point of A. Then

$$f(z_0) = \frac{1}{2\pi i} \int_{\partial A} \frac{f(z)}{z - z_0} \, dz \tag{138}$$

7.3 Cauchy-Goursat Theorem [2],[22],[25]

7.3.1 Simply Connected Domain

Let U be a simply connected domain, e.g. a star domain, where each holomorphic function has an anti-derivative. Then

$$\oint_{\Gamma} f(z) \, dz = 0 \tag{139}$$

for all closed paths $\Gamma : [0, 1] \to U$.

7.3.2 Homotopy Version

If $U \subseteq \mathbb{C}$ is an open set and $\gamma \sim \eta : [0,1] \to U$ two homotopic paths, then

$$\int_\gamma f(z)\,dz = \int_\eta f(z)\,dz \tag{140}$$

7.3.3 Homology Version

For a (open and connected) domain $U \subseteq \mathbb{C}$ and a closed path Γ in U

$$\int_\Gamma f(z)\,dz = 0 \text{ for holomorphic functions } f \iff \Gamma \text{ is homologous to zero in } U \tag{141}$$

7.3.4 Isolated Singularities

For a (non-empty, open, connected) domain $U \subseteq \mathbb{C}$, an inner point $z_0 \in U$, a holomorphic function $f : U \backslash \{z_0\} \to \mathbb{C}$ and a closed path Γ which surrounds positively oriented the isolated point $\{z_0\}$ $\mathrm{ind}_\Gamma(z_0)-$times in U,

there is an open punctured disc $U_\varepsilon \backslash \{z_0\} \subseteq U$, which is relatively compact in U, and $f|_{U_\varepsilon}$ is holomorphic, such that

$$\oint_\Gamma f(z)\,dz = \mathrm{ind}_\Gamma(z_0) \cdot \oint_{\partial U_\varepsilon} f(z)\,dz \tag{142}$$

$\mathrm{ind}_\Gamma(z_0)$ is called the **winding number** of Γ. With the definition of **residues**, i.e.

$$\mathrm{res}_{z_0}(f) := \frac{1}{2\pi i} \oint_{\partial U_\varepsilon} f(z)\,dz \tag{143}$$

this gets

$$\frac{1}{2\pi i} \oint_\Gamma f(z)\,dz = \mathrm{ind}_\Gamma(z_0) \cdot \mathrm{res}_{z_0}(f) \tag{144}$$

7.4 (Cauchy's) Residue Theorem [2],[23],[24]

Let $U \subseteq \mathbb{C}$ be a non-empty, simply connected, open domain and $U_d \subseteq U$ a discrete set, i.e. a set of isolated points. For a holomorphic function $f : U \backslash U_d \to \mathbb{C}$ and a closed path $\Gamma : I \to U \backslash U_d$, where $I \subseteq \mathbb{R}$ is a closed real interval, then

$$\frac{1}{2\pi i} \cdot \int_\Gamma f(z)\,dz = \sum_{z_0 \in U_d} \mathrm{ind}_\Gamma(z_0) \cdot \mathrm{res}_{z_0}(f) \tag{145}$$

7.5 Liouville's Theorem [26],[27],[28]

Let $f : \mathbb{C} \to \mathbb{C}$ a bounded entire function, i.e. f is holomorphic on the entire complex plane and there is a constant $C \in \mathbb{R}$ such that $||f(z)|| < C$ for all points $z \in \mathbb{C}$, then f is constant.

7.6 Riemann's Removable Singularity Theorem [29]

Let U be a non-empty, open, connected domain, $z_0 \in U$ and $f : U \backslash \{z_0\} \to \mathbb{C}$ a holomorphic function. If there is a punctured neighborhood U_0 of $\{z_0\}$, i.e. z_0 is a inner point of the closure of U_0 but isn't an element of U_0, such that $||f(z)|| < C$ for a constant $C \in \mathbb{R}$ and all points $z \in U_0$, then there is a holomorphic function $\overline{f} : U \to \mathbb{C}$ that extends f on the entire U, i.e. $\overline{f}|_{U \backslash \{z_0\}} = f$.

In this case z_0 is called a **removable singularity** of f.
Singularities in general of a function f are points, at which f isn't defined.

A singularity z_0 is called a **pole** of f if it isn't a removable singularity, but there is a natural number $k \in \mathbb{N}$ such that $(z - z_0)^k \cdot f(z)$ is a removable singularity at z_0. The minimal number k is then called the **order of the pole** z_0.

Singularities which are neither removable nor poles are called **essential singularities**.

7.7 Little Picard's Theorem [30]

The image $\mathrm{img}(f) = \{f(z) \,|\, z \in \mathbb{C}\}$ of a non-constant entire function f is the entire complex plane without at most one point, i.e. $\mathrm{img}(f) = \mathbb{C}$ or $\mathrm{img}(f) = \mathbb{C} \backslash \{z_0\}$ with a single point $z_0 \in \mathbb{C}$.

7.8 Great Picard's Theorem [31]

Let f be a holomorphic function with an essential singularity at a point z_0, then on any punctured neighborhood of z_0, $f(z)$ takes on all possible complex values, with at most a single exception, infinitely often.

7.9 Picard-Lindelöf Theorem [32]

7.9.1 Local Version

Let N be a Banach space and $[a,b] \times \overline{B}_{y_0}(r) \subseteq C \subseteq \mathbb{R} \times N$ on which a function

$$f(x,y) : C \longrightarrow N \tag{146}$$

is defined, that is continuous in the first real variable x and locally Lipschitz continuous in the second. $\overline{B}_{y_0}(r) \subseteq M$ is the closed ball with radius r and center $y_0 \in N$. Then there is exactly one solution to the initial value problem

$$\begin{aligned} y' &= f(x,y) \\ y(a) &= y_0 \end{aligned} \tag{147}$$

on the intervall $[a, a+\eta]$ with values in $\overline{B}_{y_0}(r)$ where $\eta = \min\{b-a, \frac{r}{r_0}\}$ and $r_0 = \max\{\, ||f(x,y)|| \mid (x,y) \in [a,b] \times \overline{B}_{y_0}(r)\}$.

7.9.2 Global Version

Let N be a Banach space and $f : [a,b] \times N \longrightarrow N$ a continuous function, which is Lipschitz continuous in the second variable. Then for every $y_0 \in N$ there is a unique global solution $y : [a,b] \to N$ (without any further local solutions) to the initial value problem

$$\begin{aligned} y' &= f(x,y) \\ y(a) &= y_0 \end{aligned} \tag{148}$$

7.10 Stokes Theorem [2],[5]

Let M be a orientable, compact smooth manifold with a piecewise smooth boundary ∂M and $\omega \in \Gamma(\wedge^{n-1}M)$ a smooth alternating differential form (exterior derivative, cp.(6.3)) of grade $n-1$. Then

$$\int_M d\omega = \int_{\partial M} \omega \tag{149}$$

7.11 Gauß Theorem - Divergence Theorem [2],[5]

Let $V \subseteq \mathbb{R}^n$ be a compact volume with a piecewise smooth boundary ∂V and F a smooth vector field defined in an open neighborhood of V and N its unit normal field, then

$$\iiint_V (\nabla \cdot F)\, dV = \iint_{\partial V} (F \cdot N)\, d(\partial V) \tag{150}$$

7.12 Frobenius Theorem [5]

A subbundle F of a tangent bundle TM of a smooth manifold M is integrable if and only if the vector fields with values in F build a Lie subalgebra of the Lie algebra of TM.

7.13 Hairy Ball Theorem [4]

On a sphere S^n exists a continuous tangent vector field which is nowhere zero, if and only if n is odd.

7.14 Noether's Theorem [5],[9,[10]

I will quote the two versions in Olver's book [5] without any explanations, since these would lead too far. For those who are interested in the original publications by Emmy Noether, I have linked the websites where they can be found, see [9],[10].

7.14.1 Common Version

Suppose G is a (local) one-parameter group of symmetries of the variational problem $\mathcal{L}[u] = \int L(x, u^{(n)})\, dx$. Let

$$v = \sum_{i=1}^{p} \xi^i(x,u) \frac{\partial}{\partial x^i} + \sum_{\alpha=1}^{q} \Phi_\alpha(x,u) \frac{\partial}{\partial u^\alpha} \tag{151}$$

be the infinitesimal generator of G, and

$$Q_\alpha(x,u) = \Phi_\alpha - \sum_{i=1}^{p} \xi^i u_i^\alpha\ , \quad u_i^\alpha = \frac{\partial u^\alpha}{\partial x^i}\ , \tag{152}$$

the corresponding characteristic of v. Then $Q = (Q_1, \ldots, Q_q)$ is also a characteristic of a conservation law for the Euler-Lagrange equations $E(L) = 0$; in other words, there is a $p-$tuple $P(x, u^n) = (P_1, \ldots, P_p)$ such that

$$\mathrm{Div}\, P = Q \cdot E(L) = \sum_{\nu=1}^{q} Q_\nu E_\nu L \tag{153}$$

is a conservation law in characteristic form for the Euler-Lagrange equations $E(L) = 0$.

7.14.2 General Version

A generalized vector field v determines a variational symmetry group of th functional $\mathcal{L}[u] = \int L\,dx$ if and only if its characteristic $Q \in \mathcal{A}^q$ is the characteristic of a conservation law $\operatorname{Div} P = 0$ for the corresponding Euler-Lagrange equations $E(L) = 0$. In particular, if $\mathcal{L}$ is a nondegenerate variational problem, there is a one-to-one correspondence between equivalence classes of nontrivial conservation laws of the Euler-Lagrange equations and equivalence classes of variational symmetries of the functional.

8 Epilog

In retrospect I might have chosen Stokes' theorem as central concept or simply Leibniz' product rule, as they somehow connect everything in this essay and can be compared with Paris in the French railway metric. Unfortunately one has to have already all concepts and definitions at hand for such an undertaking. This has been the reason to write this essay: to provide a place where the world of derivatives come together at one place, a quick reference guide, if you like. As I began to write it, I didn't imagine it would take so many pages, although there could be and have been written entire textbooks on each of the sections or even subsections above. For detailed studies I have to refer to them as I for sure have missed some aspects others might consider essential.

In the end, the question of how all these abstract analytic, algebraic and topological concepts relate to actual calculations and physics is not a matter of tools or obviousness, it is a matter of abstraction and the decision of how far one wants to go on the venture to explore the beauty of nature.

9 Sources

1. https://www.amazon.com/Analysis-Differentialrechnung-gewöhnliche-Differentialgleichungen-Mathematik/dp/3658023562

2. https://www.amazon.com/Analysis-Integrationstheorie-Integralsätze - Anwendungen-Aufbaukurs/dp/3658167459

3. https://www.amazon.com/Elementary-Differential-Geometry-Undergraduate-Mathematics/dp/0387903577

4. https://www.amazon.com/Alternierende-Differentialformen-German-Holmann/dp/3860258613
5. https://www.amazon.com/Applications-Differential-Equations-Graduate-Mathematics/dp/0387950001
6. https://www.amazon.com/Lie-Groups-Algebras-Cohomology-MN-34/dp/069108498X/
7. https://www.amazon.com/Groups-Algebras-Representation-Graduate-Mathematics/dp/0387909699/
8. Jean Dieudonné, Geschichte der Mathematik 1700-1900, Vieweg Verlag 1985
9. http://gdz.sub.uni-goettingen.de/en/dms/loader/img/?PID=GDZPPN002504979
10. http://gdz.sub.uni-goettingen.de/en/dms/loader/img/?PID=GDZPPN00250510X
11. http://www.math.nyu.edu/faculty/childres/fluidsbook.pdf
12. http://www.chem.mtu.edu/~fmorriso/cm4650/2012SubstantialDerivative.pdf
13. http://www.math.ttu.edu/~klong/5311-spr09/diff.pdf
14. https://www.math.uni-sb.de/ag/fuchs/Ana2/Paragraph18.pdf
15. https://www.mathematik.uni-marburg.de/~eckert/scripts/diffgeometrie.pdf
16. http://mitschriebwiki.nomeata.de/RieGeo.pdf.5.pdf
17. https://www.physicsforums.com/threads/why-the-terms-exterior-closed-exact.871875/#post-5474443
18. https://ncatlab.org/nlab/show/HomePage
19. https://de.wikipedia.org/wiki/Wikipedia:Hauptseite
20. https://en.wikipedia.org/wiki/Main_Page
21. https://fr.wikipedia.org/wiki/Wikipédia:Accueil_principal
22. http://web.math.ucsb.edu/~helena/teaching/math122b/cauchygoursat.pdf

23. https://homepage.tudelft.nl/11r49/documents/wi4006/cauchy.pdf

24. http://www.math.harvard.edu/∼knill/teaching/residues_1996/residue.pdf

25. https://www.math.ust.hk/ maykwok/courses/ma304/06_07/Complex_6.pdf

26. https://www.math.uni-hamburg.de/home/latschev/lehre/ws13/funk/5_Liouville.pdf

27. http://acc.physik.uni-frankfurt.de/vorlesung/WS1011/hl_liouville.pdf

28. http://www.mathematik.tu-darmstadt.de/Math-Net/Lehrveranstaltungen/Lehrmaterial/W
2006/AnaIII/elz/_Mitschrift/_DruckDocs/AnaIII_FT.pdf

29. https://www.math.uci.edu/∼brusso/220C040908.pdf

30. https://www.math.uh.edu/∼minru/6322-11/picard.pdf

31. https://projecteuclid.org/download/pdf_1/euclid.bbms/1086969317

32. https://embedded.eecs.berkeley.edu/eecsx44/lectures/Spring2013/Picard.pdf