THE TEACHING OF MATHEMATICS

2015, Vol. XVIII, 2, pp. 61–80

FROM DIFFERENTIATION IN AFFINE SPACES TO CONNECTIONS

Jovana Đuretić

Abstract. Connections and covariant derivatives are usually taught as a basic concept of differential geometry, or more precisely, of differential calculus on smooth manifolds. In this article we show that the need for covariant derivatives may arise, or at lest be motivated, even in a linear situation. We show how a generalization of the notion of a derivative of a function to a derivative of a map between affine spaces naturally leads to the notion of a connection. Covariant derivative is defined in the framework of vector bundles and connections in a way which preserves standard properties of derivatives. A special attention is paid on the role played by zero–sets of a first derivative in several contexts.

MathEduc Subject Classification: I 95, G 95

MSC Subject Classification: 97 I 99, 97 G 99, 53–01

Key words and phrases: Affine space; second derivative; connection; vector bundle.

1. Introduction

Definition 1. We say that a real valued function $f : (a, b) \to \mathbb{R}$ is *differentiable* at a point $x_0 \in (a, b) \subset \mathbb{R}$ if a limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists. We denote this limit by $f'(x_0)$ and call it *a derivative* of a function f at a point x_0.

We can write this limit in a different form, as

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{1}$$

This expression makes sense if the *codomain* of a function is $\mathbb{R}^n$, or more general, if the codomain is a normed vector space. Expression on the right-hand side is well defined, dividing with h is just a multiplication with a scalar $\frac{1}{h}$ in a vector space. If we change the *domain* of a function to be a normed vector space we cannot use the equation (1) as a definition of a derivative. Dividing with h (which is now a

This work is partially supported by Ministry of Education and Science of Republic of Serbia Project $\sharp ON174034$

vector) does not make sense. Let us take two normed vector spaces V and W over the same field $\mathbb{K}$ and let $U \subset V$ be an open subset.

DEFINITION 2. We say that a map $f : U \to W$ is *differentiable* at a point $x_0 \in U$ if there exists a *continuous* linear map $L : V \to W$ such that

$$f(x_0 + h) = f(x_0) + L(h) + o(h), \text{ as } h \to 0. \tag{2}$$

We denote this linear map L by $Df(x_0)$ and call it a *derivative* of a map f between normed vector spaces U and W.

REMARK 3. Every linear map between finite dimensional vector spaces is continuous. So, if we work with finite dimensional vector spaces we can omit the condition of L being continuous.

REMARK 4. One can show, as a corollary of Definition 2, that a map f is continuous at a point x_0. If we change a definition and ask for a map f to be continuous (instead of L) then we can prove that L is also going to be continuous. Therefore continuity of a map f and a map L are equivalent conditions (see [4] for details).

REMARK 5. It is clear that Definition 2 is a generalization of Definition 1, since for $U = (a, b) \subset V = \mathbb{R}$ the linear map L is a multiplication with a constant $f'(x_0)$,

$$L(h) = f'(x_0) \cdot h.$$

We can write the equation (2) as

$$f(x) - f(x_0) = Df(x_0)(x - x_0) + o(x - x_0), \text{ as } x \to x_0. \tag{3}$$

Seemingly equivalent, equations (2) and (3) are essentially different in one very important sense. Equation (2) requires a vector structure while equation (3) requires only an affine structure. Formulation (3) gives us an idea on how to define a derivative of a map between normed affine spaces.

2. Derivative in affine spaces

Let us recall the definition of a normed affine space.

DEFINITION 6. An *affine space* is a triple $(A, V, +)$ where A is a set, V is a vector space and $+$ is a map

$$+ : A \times V \to A$$

such that

(Af1) $a + 0 = a$ for all $a \in A$,

(Af2) $a + (u + v) = (a + u) + v$ for all $a \in A$, $u, v \in V$,

(Af3) for every $a_1, a_2 \in A$ there exists a unique $v_a \in V$ such that $a_2 = a_1 + v_a$.

If V is a normed vector space we say that $(A, V, +)$ is a *normed affine space.*

We can define a metric on the set A as

$$d(a_1, a_2) := \|v_a\|,$$

where v_a is the unique vector provided by (Af3). (Af3) also allows us to define a "subtraction" on A:

$$\begin{aligned} - : A \times A &\to V, \\ a_2 - a_1 &:= v_a. \end{aligned} \tag{4}$$

The result of this subtraction is not in A but in V.

We can associate a vector space to an affine space if we pick one point $a \in A$ to be the zero vector. This process is called a *vectorization* of A with respect to a. The vector space we obtain is isomorphic to V and will be denoted by T_aA. Generalization of Definition 2 gives us the definition of a derivative of a map defined on affine spaces.

DEFINITION 7. Let $f : A \to B$ be a map, where A and B are normed affine spaces. We say that f is *differentiable* at a point $a \in A$ if there exists a *continuous* linear map $L : T_aA \to T_{f(a)}B$ such that

$$f(x) - f(a) = L(x - a) + o(x - a), \text{ as } x - a \to 0.$$

Again, we denote L as $Df(a)$ and call it a derivative of a map f. The minus sign that appears on the left–hand side is the map we defined in (4).

Let us see some standard properties of a derivative. The following proposition shows its linearity.

PROPOSITION 8. *Let A be a normed affine space, W a normed vector space (both over the same field $\mathbb{K}$) and let $f, g : A \to W$ be maps that are differentiable at $a \in A$. Then $\lambda f + \mu g$ is differentiable at a for all $\lambda, \mu \in \mathbb{K}$ and it holds*

$$D(\lambda f + \mu g)(a) = \lambda \cdot Df(a) + \mu \cdot Dg(a).$$

Proof. We make a difference of a map $\lambda f + \mu g$ at two nearby points x and a:

$$\begin{aligned} (\lambda f + \mu g)(x) - (\lambda f + \mu g)(a) &= \lambda f(x) + \mu g(x) - \lambda f(a) - \mu g(a) \\ &= \lambda(f(x) - f(a)) + \mu(g(x) - g(a)) \\ &= \lambda\,(Df(a)(x - a) + o(x - a)) + \mu(Dg(a)(x - a) + o(x - a)) \\ &= \lambda \cdot Df(a)(x - a) + \mu \cdot Dg(a)(x - a) + o(x - a),\ x - a \to 0, \end{aligned}$$

(third equality holds because f and g are differentiable at a point a). This difference is a continuous linear map

$$\lambda \cdot Df(a) + \mu \cdot Dg(a) : T_aA \to W$$

(up to $o(x - a)$). We conclude that $\lambda f + \mu g$ is differentiable at a point a and proposition holds. ■

REMARK 9. For maps $f, g : A \to B$ defined on affine spaces we cannot take arbitrary linear combination $\lambda f + \mu g$, but we can take linear combinations of the form $\lambda f + (1-\lambda)g$ and the conclusion of the previous proposition holds

$$D(\lambda f + (1-\lambda)g)(a) = \lambda \cdot Df(a) + (1-\lambda) \cdot Dg(a).$$

The product rule is also satisfied.

PROPOSITION 10. *Let us take functions $f, g : A \to \mathbb{R}$ defined on a normed affine space A. If f and g are differentiable at $a \in A$ then fg is differentiable function and it holds*

$$D(fg)(a) = f(a) \cdot Dg(a) + g(a) \cdot Df(a).$$

The following proposition states that the derivative of a composition is a composition of derivatives.

PROPOSITION 11. *Let A_1, A_2 and A_3 be normed affine spaces and let $g : A_1 \to A_2$, $f : A_2 \to A_3$ be maps between them. If g if differentiable at $a \in A_1$ and f is differentiable at $g(a)$ then $f \circ g$ is differentiable at a and it holds*

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

Proofs of Proposition 10 and Proposition 11 are standard and well known.

From Definition 7 it follows that a derivative of a map f (which is differentiable at every point of an open subset $U \subset A$) is a map

$$Df : U \to \bigcup_{a \in U} \mathcal{L}(T_a A; T_{f(a)} B), \tag{5}$$

such that $Df(a) \in \mathcal{L}(T_a A; T_{f(a)} B)$ for all $a \in U$. In other words first derivative of a map is a section of a vector bundle[1]

$$\pi : \widetilde{\mathcal{L}} = \bigcup_{a \in U} \mathcal{L}(T_a A; T_{f(a)} B) \to U, \tag{6}$$

where the fiber over $a \in U$ is $\pi^{-1}(a) = \mathcal{L}(T_a A; T_{f(a)} B)$.

3. Second derivative

Our next goal is to define the second derivative of a map f. Let us see what is the second derivative in our motivating example of a map defined on normed vector spaces. The first derivative is

$$Df : U \to \mathcal{L}(V; W),$$

[1]Informally, a vector bundle is a family of vector spaces that are attached to a manifold in a smoothly varying way. It combines topology with linear algebra. A section assigns to every point of a manifold a vector from the vector space attached to this point. See Section 4 for the precise definitions of a vector bundle (Definition 14) and of a section of a vector bundle (Definition 16).

and the codomain of this map is again a normed vector space (the space of linear continuous maps with standard operator norm). From Definition 2 we conclude that the second derivative, i.e. the differential of Df, is

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

The space $\mathcal{L}(V; \mathcal{L}(V; W))$ is canonically identified with the space of continuous bilinear maps $V \times V \to W$, which is again a normed vector space denoted by $\mathcal{L}^2(V; W)$. In a similar manner we can define the derivative of order n,

$$D^n f : U \to \mathcal{L}^n(V; W),$$

where $\mathcal{L}^n(V; W)$ is a space of n–linear maps from V to W.

In the case of maps between affine spaces the situation is more complicated. First derivative in (5) is not a map defined on affine spaces since $\widetilde{\mathcal{L}}$ defined in (6) does not have an affine structure. If we want to find the second derivative of a map f defined on affine spaces we actually have to differentiate a section of a vector bundle (6). We could overcome this problem if we worked in the framework of smooth manifolds. However this way we would not preserve the specific structure of a map. When we work with affine spaces we cannot interpret n-th derivative as a n-linear map. But we will preserve some other properties.

Let us see some more examples of differentiation of some suitable sections.

Example 12. (*Critical points of a function*) Let us take a smooth n-dimensional manifold M and a smooth function

$$f : M \to \mathbb{R}.$$

The first derivative of a function f at any point $p \in M$ is a linear map

$$df(p) : T_p M \to \mathbb{R}.$$

In other words $df(p)$ is an element of a dual space $T_p^* M$ (we know that the elements of the cotangent space are called cotangent vectors or tangent *covectors*). We can see a derivative of f as a map[2]

$$df : M \to T^* M,$$

such that $df(p) \in T_p^* M$ for all $p \in M$. This means that the differential of a function defined on a manifold is a section of the cotangent bundle

$$\pi : T^* M \to M.$$

Differentiation of this section df (which we have not yet defined) gives us the second derivative of a function f. However we can find the second derivative (at some points) even if do not know how to differentiate sections.

Take a critical point q of a function f. Let us recall that a critical point is a point where the differential vanishes (using the critical points of a Morse function

[2] In [19] one can find a nice exposition on manifolds and differential forms on manifolds.

we can describe the topology of a manifold, see [12] for more details). We can define a bilinear form of the second derivative

$$d^2 f(q) : T_q M \times T_q M \to \mathbb{R},$$

as

$$d^2 f(q)(X_q, Y_q) = X_q(Y(f)). \tag{7}$$

In the previous equality Y is a vector field that is an expansion of a vector Y_q on some neighbourhood of q. The first derivative of a function defined on $\mathbb{R}$ is again a function defined on $\mathbb{R}$. So it is clear what is the second derivative of such function, it is again the same object, a derivative of the function. On the other side, the first derivative of a function defined on a manifold is not a function. But its derivative in the Y-direction,

$$Y(f) = df(Y),$$

is a function (we can see that that this is the place where the first derivative of f appears in (7)). Now, differentiation of this function $Y(f)$ in a direction X_q makes sense.

We can easily prove that the definition (7) does not depend on the expansion Y but only on its value at q. Let us take some expansion X of a vector X_q. Then

$$X_q(Y(f)) - Y_q(X(f)) = [X, Y]_q(f) = df(q)([X, Y]_q) = 0.$$

The last equality follows from the fact that q is a critical point. In the previous relation $[\cdot, \cdot]$ denotes a commutator of vector fields. We conclude

$$X_q(Y(f)) = Y_q(X(f)). \tag{8}$$

The right-hand side depends only on the value of a vector field Y at a point q, not on the expansion Y. The left-hand side depends only on X_q, not on its expansion. We conclude that both sides depend only on X_q, Y_q. Therefore, the relation (7) gives us well defined bilinear form and from (8) it follows that this form is symmetric.

Thus, we can define the second derivative of a function at its critical points without any additional structure or definitions. This is not the case if q is not a critical point. We will see later, in a more general framework, why this is the case.

Similar to the case of normed vector spaces, we defined the second derivative, $d^2 f$, to be a bilinear map on TM, but *only at critical points.*

If we fix a vector field X then,

$$d^2 f(q)(X_q, \cdot) : T_q M \to \mathbb{R},$$

is the linear map on a tangent space $T_q M$. We interpret the second derivative of a function “along a vector field” as a covector, i.e. it is of the same type as the differential of the function $df(q)$.

If we take $M = \mathbb{R}^n$ then $d^2 f(q)$ is given by the well known Hessian matrix of a function f,

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Here, $(x_1, x_2, \ldots, x_n)$ are coordinates in $\mathbb{R}^n$. $\triangle$

The following example also uses manifolds but now we differentiate a section of the tangent bundle.

Example 13. (*Linearization of dynamical systems*) As in the previous example, let us take a smooth manifold M and let TM denotes its tangent bundle. It is well known that sections of this bundle are vector fields

$$X : M \to TM,$$

such that $X(p) \in T_pM$ for all $p \in M$. Differentiation of this section (i.e. vector field) gives us a linearization of a vector field. Let $q \in M$ be a point where X vanishes, $X(q) = 0$. These zeroes of a vector field are also called equilibrium points. The condition of being an equilibrium point is an analogue of being a critical point in Example 12.

The flow of a vector field, ϕ^t, is the solution of the differential equation

$$\frac{d}{dt}\phi^t(x) = X(\phi^t(x)),$$

with an initial condition

$$\phi^0(x) = x.$$

The dynamical system generated by a vector field X is described by this differential equation (see [1] for an introduction to dynamical systems and application to the three-body problem). Zeroes of a vector field, i.e. equilibrium points of a dynamical system, are fixed points of this system. For those points it holds

$$\phi^t(q) = q$$

for all t (this follows from the existence and uniqueness of the solution of a differential equation).

For $V \in T_qM$ we define "derivative of X along V" as:

$$\nabla_V X := \frac{d}{dt} |_{t=0} \, \phi_*^t V.$$

From the definition of a push–forward we know that

$$\phi_*^t V \in T_{\phi^t(q)}M = T_qM,$$

since q is an equilibrium and $\phi^t(q) = q$. It follows that $\nabla_V X \in T_q M$. So "differentiation" of a vector field gives us a vector field.

We saw in Example 12 that we can find the second derivative of a function only at its critical points. Similarly, described linearization is possible *only at equilibrium points* of a dynamical system. Without a property of q being an equilibrium point we cannot achieve a condition that a derivative of X along V at q belongs to $T_q M$. In general, $\phi^t_* V$ belongs to $T_{\phi^t(q)} M$ and if q is not an equilibrium $T_{\phi^t(q)} M$ is not equal $T_q M$. $\triangle$

The idea in previous examples is to define the derivative of a map in a way that the new map we obtain is of the same type as the one we differentiate. For example, derivative of a differential (along a vector field) is again a linear form. Derivative of a vector field (along a vector field) is a vector field, again. On the other hand, we want to generalize well known differentiation in Euclidean spaces at any point, not only at a zero sets as in Examples 12 and 13. Modifying the first derivative so that we get the object of the same type as the one we differentiate leads to the notion of connections and covariant derivatives[3]. We introduce them in the next section, and then come back to the role of the zero sets of the first derivative in a more general context.

4. Connections

In Euclidean space $\mathbb{R}^n$ we have an identification of vectors by translation. We can translate any vector to the origin and then, for example, sum two vectors. We cannot translate tangent vectors if we are on the sphere or on a manifold. In differential geometry we introduce a connection, an object that connects nearby tangent spaces and allows us to "translate" tangent vectors.

Christoffel was the first to study connections from an infinitesimal perspective in Riemannian geometry. Ricci and Levi–Civita continued this and gave a geometric interpretation by means of associated *parallel*[4] transport. This was later generalized by Koszul, Cartan and Ehresmann.

Let us define some objects we will use in this section.

Definition 14. A *smooth vector bundle* is a triple (E, M, π) where the total space E and the *base space* M are smooth *manifolds*, the projection $\pi : E \to M$ is a smooth map and the following holds:

- for all $p \in M$ the fiber $E_p := \pi^{-1}(p)$ has a structure of a vector space of dimension r over the field $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$,

[3] These two notions are often used to name the same object, since they are closely related. But it is useful to have in mind that connections are geometrical, and covariant derivatives analytical, or algebraic, objects, as we will see later in the text.

[4] We mention here that Euclidean way of introducing the notion of parallelism was one of the most controversial points in his "Elements", that gave a birth to several non-Euclidean geometries, developed by Lobachevsky, Bolyai, Gauss, Riemann and others. We will come back to this point later in this article.

- for all $p_0 \in M$ there exists an open subset $U \subset M$, $p_0 \in U$, and a diffeomorphism ψ (called *local trivialization*)

$$\psi : \pi^{-1}(U) \to U \times \mathbb{K}^r,$$

such that for all $p \in M$ restriction

$$\psi : E_p \to \{p\} \times \mathbb{K}^r$$

is a linear isomorphism of vector spaces.

The integer $r \geq 0$ is called the *dimension* of a vector bundle.

It is obvious that $\dim E = r + \dim M$.

Remark 15. We could define a vector bundle in the category of the topological spaces. In that case we ask for E and M to be topological spaces, π to be a continuous map and a local trivialization to be a homeomorphism (see [13] for more details). Here, we are dealing with the smooth structure so when we say vector bundle we mean a vector bundle in the category of smooth manifolds.

We can now define a section of a vector bundle.

Definition 16. A *section* of a vector bundle (E, M, π) is a smooth map $s : M \to E$ such that

$$\pi \circ s = \mathrm{Id}_M .$$

In other words, $s(p)$ belongs to E_p for all $p \in M$.

The most important example of a section is the *zero section*, denoted by s_0, which satisfies

$$s_0(p) = 0_p \in E_p,$$

for all $p \in M$.

Definition 17. A *distribution* Δ of dimension k on a manifold E is a smooth collection of k-dimensional subspaces

$$\Delta_p \subset T_pE.$$

This means that every point on a manifold E has a neighbourhood U and smooth vector fields

$$X_1, X_2, \ldots, X_k : U \to TE$$

such that $X_1(p), X_2(p), \ldots, X_k(p)$ generate Δ_p for all $p \in U$.

One way to define k-distribution on n-dimensional manifold is to see it as a kernel of $n-k$ 1-forms (see [9]).

It is interesting for us to observe a distribution on a special class of manifolds, namely the total space of a vector bundle.

So, let us take a vector bundle (E, M, π) of range r. Projection π has a derivative at a point $e \in E$ (this is the differential of a smooth map defined on smooth manifolds):

$$\pi_*(e) : T_eE \to T_{\pi(e)}M.$$

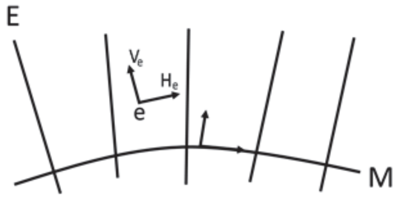

Figure 1. Vector bundle

Its kernel,

$$V_e := \ker \pi_*(e),$$

defines a distribution $\{V_e\}_{e\in E}$ on E. This distribution is called a *vertical distribution*. We can see that this is a r–distribution, since

$$V_e \cong \mathbb{K}^r.$$

This vertical distribution is defined canonically, without any auxiliary structure. On the other hand, we cannot define its complement canonically. More precisely, we cannot choose in a unique way a family of subspaces

$$H_e \leq T_e E,$$

such that

$$T_e E = V_e \oplus H_e,$$

for all $e \in E$. One choice of such a family of subspaces H_e is a distribution of dimension $\dim M$, and we call it a horizontal distribution.

Definition 18. A *connection* on a vector bundle is a smooth distribution of *horizontal subspaces* $\{H_e\}_{e\in E}$ such that

$$T_e E = V_e \oplus H_e \tag{9}$$

for all $e \in E$.

We can define a connection (at least locally) as a kernel of r 1-forms or as a kernel of a smooth 1–form

$$\omega : T_e E \to \mathbb{K}^r$$

with values in a vector space $\mathbb{K}^r$. We can define a connection on every vector bundle however it is not unique, in general.

Let us take a section

$$s : M \to E.$$

Its derivative is

$$ds(p) : T_p M \to T_{s(p)} E.$$

We will denote $s(p)$ by e. For a chosen connection on a vector bundle, equation (9) defines a smooth family of projections

$$\mathrm{pr}_V(e) : T_e E \to V_e.$$

We can now define a new type of derivative of a section.

DEFINITION 19. Composition

$$\mathrm{pr}_V(e) \circ ds(p) : T_pM \to V_e,$$

is called a *covariant derivative* of a section s and is denoted by ∇s.

If we are given a vector field on M, $X : M \to TM$, we can define a derivative of a section s along X

$$\nabla_{X_p} s := \mathrm{pr}_V(s(p)) \circ ds(p) X_p.$$

This map can be interpreted as a section of the same vector bundle

$$\nabla_X s : M \to E,$$

since

$$\nabla_{X_p} s(p) = \mathrm{pr}_V(s(p)) \circ ds(p) X_p \in V_e \cong E_p,$$

(E_p is a fiber). Therefore, we accomplished one very important thing, $\nabla_X s$ is of the same type as the section s we differentiate.

A covariant derivative satisfies standard properties.

PROPOSITION 20. *For sections $s, \sigma : M \to E$, vector fields $X, Y : M \to TM$ and a smooth function $f : M \to \mathbb{K}$ it holds*

- $\nabla_{X+Y} s = \nabla_X s + \nabla_Y s$,
- $\nabla_{fX} s = f \nabla_X s$,
- $\nabla_X (s + \sigma) = \nabla_X s + \nabla_X \sigma$,
- $\nabla_X (fs) = df(X)s + f \nabla_X s$.

If we are given a connection we can define a covariant derivative. Vice versa is also true. If we are given an operator that satisfies properties from Proposition 20 then we can find, in a unique way, a connection which defines that covariant derivative (see [16] for details).

EXAMPLE 21. (*Second derivatives in affine spaces*) Let us, again, consider the vector bundle

$$\pi : \widetilde{\mathcal{L}} \to U,$$

which we defined in (6). If we are given a connection on this bundle, we can define a covariant derivative of the specific section given by the derivative of f

$$Df : U \to \widetilde{\mathcal{L}}.$$

In other words, connections allow us to define the second derivative of a function between affine spaces. $\triangle$

Example 22. (*Affine and Levi-Civita connections*) If we are given a smooth manifold M, a smooth connection on its tangent bundle is called an affine connection. As we said, after Proposition 20, the choice of an affine connection is equivalent to prescribing a way of differentiating vector fields on a manifold.

If a manifold is moreover endowed with a Riemannian metric g then it is useful to consider a special connection, the Levi-Civita connection. It is a connection on the tangent bundle TM that is symmetric

$$\nabla_X Y - \nabla_Y X = [X, Y],$$

and is compatible with the metric

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$

We ask for previous relations to hold for all vector fields $X, Y, Z : M \to TM$. The existence and uniqueness of the Levi-Civita connection follows from the fundamental theorem of Riemannian geometry (one can find a proof of this theorem in [3]). Since we have a metric on a manifold (on the tangent bundle, to be more precise) we can speak about orthogonality of vectors in TM. Now a notion of an orthogonal complement makes sense. Levi-Civita connection is defined by a horizontal distribution which is an orthogonal complement of a vertical distribution.

A geometric interpretation of connections was given by Levi-Civita who introduced the notion of parallel transport on surfaces. If we are given a curve on a surface and a tangent vector at the starting point, the vector can be transported along the curve by requiring the moving vector to remain parallel to the original one, to belong to the tangent bundle of the surface and to remain of the same length. It is hard to understand what is a notion of parallel on a surface since it is not usually possible to identify all the tangent planes of the surface. The requirement of vector field $X(t)$ to be parallel along curve $\gamma(t)$ can be stated in terms of the covariant derivative as

$$\nabla_{\dot{\gamma}} X = 0.$$

In a local trivialization this is a first–order system of differential equations which has a unique solution if we are given an initial condition (i.e. the vector at the starting point). $\triangle$

Example 23. (*Hermitian connections*) The most common case of a vector bundle that we observe is the case when fibers are vector spaces over $\mathbb{R}$. Sometimes, to point out a difference, we call these bundles real vector bundles. If we ask for a fiber to be a vector space over $\mathbb{C}$ then we have defined a complex vector bundle (see [13] for a precise definition). A Hermitian metric on a complex vector bundle is a Euclidean metric on the underlying real vector bundle, which satisfies the identity

$$\|iv\| = \|v\|.$$

Hermitian connection is the one which is compatible with the Hermitian metric. $\triangle$

5. Examples of moving coordinate systems

Coordinates on a base space $(x_1, x_2, \dots, x_n)$ (or in some local chart U) give us the basis of a tangent space $\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_n}$. In the classical theory we usually use these vector fields given by this coordinate system. When we work with an operator ∇ calculation is more complicated and sometimes it is convenient to express a result of covariant differentiation in terms of arbitrary, linear independent vector fields. If we are given vector fields that are linearly independent at each point of a manifold then we actually have a frame of a tangent space which varies from point to point. We emphasize this notion with the following definition.

Definition 24. Let, for each $p \in M$, an ordered basis

$$F_p = (X_1(p), X_2(p), \dots, X_n(p))$$

of a tangent spaces T_pM be given. The family $\{F_p\}_{p\in M}$ is called a *moving frame* on M.

A covariant derivative of natural frame, $\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_n}$, can be written as

$$\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = \sum_{k=1}^{n} \Gamma_{ij}^k \frac{\partial}{\partial x_k},$$

where Γ_{ij}^k are Christoffel symbols. In Cartan's theory, the basic idea is to express everything in terms of a moving frame F_p, not just in terms of the natural frame (see [16]). In the next example we work with polar coordinate system, the two-dimensional moving frame in which each point in the plane is determined by a distance from the origin and an angle from the x-axis.

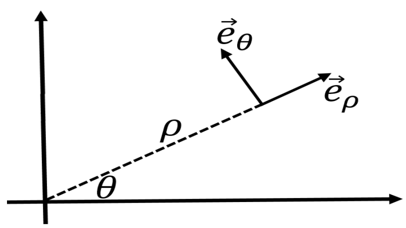

Figure 2. Polar coordinate system

Example 25. (*Polar coordinate system*) Let us find Christoffel symbols in polar coordinates (ρ, θ) in $\mathbb{R}^2$. We know that relations between Cartesian coordinates and the polar coordinates are (see Figure 2)

$$x = \rho \cos\theta,$$
$$y = \rho \sin\theta.$$

We change the basis using the equations

$$\boldsymbol{e}_\rho = \frac{\partial x}{\partial \rho}\boldsymbol{e}_x + \frac{\partial y}{\partial \rho}\boldsymbol{e}_y,$$
$$\boldsymbol{e}_\theta = \frac{\partial x}{\partial \theta}\boldsymbol{e}_x + \frac{\partial y}{\partial \theta}\boldsymbol{e}_y.$$

Thus

$$\begin{aligned}\boldsymbol{e}_\rho &= \cos\theta\,\boldsymbol{e}_x + \sin\theta\,\boldsymbol{e}_y,\\ \boldsymbol{e}_\theta &= -\rho\sin\theta\,\boldsymbol{e}_x + \rho\cos\theta\,\boldsymbol{e}_y.\end{aligned}$$

It follows that

$$\begin{aligned}\frac{\partial\boldsymbol{e}_\rho}{\partial\rho} &= 0,\\ \frac{\partial\boldsymbol{e}_\rho}{\partial\theta} &= -\sin\theta\,\boldsymbol{e}_x + \cos\theta\,\boldsymbol{e}_y = \frac{1}{\rho}\boldsymbol{e}_\theta,\\ \frac{\partial\boldsymbol{e}_\theta}{\partial\rho} &= -\sin\theta\,\boldsymbol{e}_x + \cos\theta\,\boldsymbol{e}_y = \frac{1}{\rho}\boldsymbol{e}_\theta,\\ \frac{\partial\boldsymbol{e}_\theta}{\partial\theta} &= -\rho\cos\theta\,\boldsymbol{e}_x - \rho\sin\theta\,\boldsymbol{e}_y = -\rho\,\boldsymbol{e}_\rho.\end{aligned}$$

We conclude that Christoffel symbols are

$$\begin{aligned}&\Gamma^\rho_{\rho\rho} = \Gamma^\theta_{\rho\rho} = 0,\\ &\Gamma^\rho_{\rho\theta} = 0, \quad \Gamma^\theta_{\rho\theta} = \frac{1}{\rho},\\ &\Gamma^\rho_{\theta\rho} = 0, \quad \Gamma^\theta_{\theta\rho} = \frac{1}{\rho},\\ &\Gamma^\rho_{\theta\theta} = -\rho, \quad \Gamma^\theta_{\theta\theta} = 0.\end{aligned}$$

Once we know how to differentiate in polar coordinates we can easily deduce the famous Kepler's laws of planetary motion. It is the motion in a central field[5] and here we only sketch the ideas of Arnold (see [2] for details). If we are given a planet of a mass m in a central field (in our case the Sun of a mass M is in the origin of a field) then the angular momentum

$$\boldsymbol{L} := \vec{\rho}\times\dot{\vec{\rho}} = \rho\dot{\theta}\,\boldsymbol{e}_\rho\times\boldsymbol{e}_\theta$$

does not change with time. Thus, the motion of the planet always remains in the plane. The quantity

$$L = \rho^2\dot{\theta}$$

is also preserved. Preservation of quantity L has a geometric meaning also known as Kepler's second law. It states that *a line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time.* Using the Newton's Second Law

$$m\ddot{\vec{\rho}} = \boldsymbol{F}\left(= -G\frac{mM}{\rho^2}\boldsymbol{e}_\rho\right)$$

(G is a constant) we reduce the problem of the motion of the planet to the problem of resolving a differential equation of second order

$$\ddot{\rho} = \frac{L^2}{\rho^3} - \frac{GM}{\rho^2}.$$

[5] A vector field in the plane $\mathbb{R}^2$ is called central with center at 0 if it is invariant with respect to the group of motions of the plane which fix 0.

Resolving this equation one easily obtains Kepler's first law: *the planet describes ellipse with the Sun at one of the two foci.* Kepler's third law says that *the period of revolution around an elliptical orbit depends only on the size of the major axes.* One can prove Kepler's third law by comparing well-known formula for the area of an ellipse, $P = ab\pi$ (a and b are major and minor axes), with the one that depends on the period of the revolution. $\triangle$

If we are given some (local) basis for a fiber $e_1, e_2, \ldots, e_r$ then every section s can be written locally as

$$s = h_1 e_1 + h_2 e_2 + \ldots + h_r e_r,$$

for some locally defined functions $h_1, h_2, \ldots, h_r$. Covariant derivative of this section is

$$\nabla s = \sum_{i=1}^{r} (dh_i \otimes e_i + h_i \nabla e_i).$$

Thus, when we know covariant derivative of e_i then we can find derivative of any section s. Covariant derivative of e_i can be expressed in the form

$$\nabla e_i = \sum_{j=1}^{r} \theta_{ij} \otimes e_j,$$

where θ_{ij} are 1–forms. Form $\theta = (\theta_{ij})_{i,j=1}^{n}$, with matrix values, is called the *connection form* in the basis $e_1, e_2, \ldots, e_r$.

Example 26. (*Frenet coordinate system*) Let us find the connection form in the Frenet orthonormal frame of a curve $\gamma : I \to \mathbb{R}^3$, where I is an interval. Here we consider the bundle of a rank 3 over the one–dimensional base[6].

We know that a curve $\gamma = \gamma(s)$, parameterized by the arc length, defines the Frenet frame in which basis vectors are

$$\begin{aligned} \boldsymbol{T}(s) &= \gamma'(s), \\ \boldsymbol{N}(s) &= \frac{\boldsymbol{T}'(s)}{\|\boldsymbol{T}'(s)\|}, \\ \boldsymbol{B}(s) &= \boldsymbol{T}(s) \times \boldsymbol{N}(s). \end{aligned}$$

Family $\{\boldsymbol{T}(s)\}$ defines a horizontal distribution in 3–dimensional vector bundle over γ. We will find a covariant derivative of any section with respect to this distribution. From the Frenet-Serret formulas follows

$$\begin{bmatrix} \boldsymbol{T}' \\ \boldsymbol{N}' \\ \boldsymbol{B}' \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{T} \\ \boldsymbol{N} \\ \boldsymbol{B} \end{bmatrix}$$

where κ is the curvature and τ is the torsion of the curve. Any section σ can be written as

$$\sigma(s) = \sigma_T(s)\boldsymbol{T} + \sigma_N(s)\boldsymbol{N} + \sigma_B(s)\boldsymbol{B},$$

[6] In a formal language of vector bundles, it is a vector bundle over I denoted by $\gamma^* T\mathbb{R}^3$, but we do not intend to discuss that formal language here.

for some functions $\sigma_T, \sigma_N, \sigma_B$. Then

$$\begin{aligned}\nabla\sigma &= pr_{\boldsymbol{N},\boldsymbol{B}} \circ d\sigma \\ &= pr_{\boldsymbol{N},\boldsymbol{B}} \circ (d\sigma_T \boldsymbol{T} + d\sigma_N \boldsymbol{N} + d\sigma_B \boldsymbol{B} + \sigma_T \boldsymbol{T}' + \sigma_N \boldsymbol{N}' + \sigma_B \boldsymbol{B}').\end{aligned}$$

The Frenet-Serret formulas give us

$$\nabla\sigma = (d\sigma_N + \kappa\sigma_T - \tau\sigma_B) \otimes \boldsymbol{N} + (d\sigma_B + \tau\sigma_N) \otimes \boldsymbol{B}. \quad \triangle$$

6. The case of zero section

In Section 3 we started the discussion on the special role of the zero sets of the first derivative. Now we come back to this point from the more general viewpoint.

As we saw in Section 4 every vector bundle (E, M, π) has at least one section, namely the zero section s_0. It maps every point $p \in M$ to the zero element of the vector space E_p. The zero section gives a natural embedding $M \hookrightarrow E$.

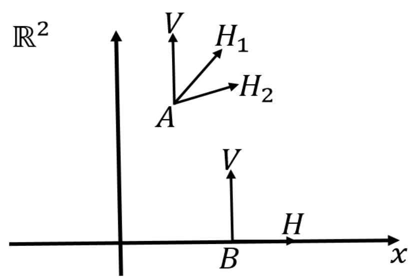

Figure 3. Multiple choice on horizontal connection

It was mentioned before that we cannot choose horizontal spaces in the tangent bundle of a total space in a canonical way[7]. There is more then one possibility for choosing subspaces H_e such that the relation (9) holds. Let us take, for example, $\mathbb{R}^2$ to be a vector bundle over a base space $\mathbb{R}$ (x-axis in Figure 3). Then, at point A of $\mathbb{R}^2$ we have a vertical subspace (direction V), and we can take for H_e a subspace defined with the direction H_1 or with the direction H_2 (see Figure 3). We do not have one direction (or subspace) that steps-up among the other. But if we take a point B, that lies in the base space, and look for a horizontal subspace, then the situation does not require making a choice. We can simply take the x-axis for the horizontal subspace. Since we have an identification of tangent spaces when we work with $\mathbb{R}^n$, we can identify x-axis with the tangent space of the base space $\mathbb{R}$.

Now, suppose a section $s : M \to E$ has a zero at $p \in M$. Then at a point p we have a preferred horizontal space. This space is defined as an image

$$ds_0(TM),$$

[7]To construct (to prove or to define) something canonically means to construct it using only the objects we have defined until that moment. For example, we know that any two vector spaces of the same dimension are isomorphic. However, this isomorphism is not canonical. In order to construct the isomorphism one has to define some new structure on these vector spaces. For example, we can construct the isomorphism by choosing the basis for the vector spaces.

where s_0 is already mentioned zero section. This is the reason why we could define the second derivative of a function at its critical points without introducing a new structure (see Example 12). These points are exactly intersection points of a differential df and the zero section in T^*M. Similarly, in Example 13 we could canonically linearize a vector field at its zeroes.

7. Connections and geometry

There were several approaches that generalized Levi-Civita's and Christoffel's study of connections. *Koszul* gave an algebraic framework for describing connections on vector bundles as differential operators. He defined a connection as a map

$$\nabla : \chi(M) \times \Gamma(M, E) \to \Gamma(M, E)$$

which satisfies properties from Proposition 20. Here, $\chi(M)$ denotes the set of vector fields on M and $\Gamma(M, E)$ denotes the set of sections of a vector bundle. *Cartan* generalized a notion of an affine connection and defined another thread in the theory of connections, by presenting connections as differential forms. *Ehresmann* defined a connection as a family of horizontal and vertical subspaces on a total space of a principal bundle. A principal G-bundle, where G is a topological group, is a fiber bundle $\pi : P \to X$ with a continuous right action $P \times G \to P$ such that G preserves the fibers of P, acts freely and transitively on them[8].

Euclid in his Elements founded geometry by logical deduction from his famous axioms. For centuries it was not clear if the Axiom of parallels follows from the other axioms. This led to a construction of non-Euclidean geometry. Klein in his Erlangen Program gave a very general viewpoint for studying geometry. He regarded geometry as the study of invariants under a group of transformations. This idea is, in a certain sense, present already in Elements, where the group of symmetries is the group of isometric transformations.

Theory of connections brings together these two important concepts: the notion of parallelism and the symmetry with respect to a group action (as we said above in Ehresmann connections). As we mentioned in Example 22, notion of parallelism and covariant differentiation are tightly connected. Originally algebraic and topological in nature, differentiation can be geometrized, if we remove an algebraic structure from an ambient and consider it to be just an affine, instead of a vector space. It still possesses enough structure to carry the definition of the first derivative, but the higher order derivatives naturally lead to the notion of connections and the investigation of the symmetries needed in order to avoid getting more complicated objects after each differentiation. As we have seen in Section 6, at the zero section there is a natural way to do it in a canonical way. In general, additional structures or symmetries need to be imposed, even in affine spaces.

The case of smooth manifolds is the beginning of the connection theory in classical differential geometry, and, more generally, theory of Cartan connections.

[8] A principal bundle generalize a vector bundle since any vector bundle can be obtained from a principal bundle using the appropriate group representation.

Discussion on role of covariant differentiation in Cartan's generalization of Klein's Erlangen Program is something that is beyond the aim of this article. In this generalization a group acts not only on a homogeneous space but on fibers of a principal bundle. Further details can be found in [14] and [15].

8. Examples from recent research

Theory of connections appears in topology, especially in the theory of 4-manifolds, theoretical physics, gauge theory, symplectic topology, etc. It has been known for a long time that the methods of classical algebraic topology in studying manifolds of dimensions greater than 4 cannot be applied in lower dimensional cases. In particular, famous Smale's proof of Poincaré Conjecture in dimensions 5 or more was proved in a completely different ways in dimensions 3 and 4, due to certain nontrivial linking phenomena in lower dimensions.

In a topology of four dimensional manifolds Donaldson applied the method of elliptic PDEs to the geometry of connections on principal $SU(2)$ bundles over the smooth 4-manifold M. He showed that the set $\mathcal{N}$ of so called *anti self dual connections* is a finite dimensional smooth manifold inside the space of all connections. Furthermore, he showed that the intersection theory on $\mathcal{N}$ can be used to construct important topological invariants of an original manifold M (see [5, 6]).

Many analytical difficulties of Donaldson approach were simplified by Seiberg-Witten's approach, where the $SU(2)$ bundles were replaced by $\mathbb{S}^1$ bundles, and connections on them paired with the Dirac operator (see [11, 17]). The fact that the group $\mathbb{S}^1$ is commutative, unlike $SU(2)$, greatly simplified the analysis in this case. In a similar spirit, Gromov developed his theory of (pseudo)holomorphic curves (see famous Gromov's paper [7]). At the end, we mention that Gromov's theory, in special (four dimensional) case is related to Seiberg-Witten's invariants (see, for example [18]).

We will end this article by an example that illustrates the above mentioned ideas (selecting topologically useful finite dimensional *moduli space* within certain infinite dimensional manifold), and also demonstrates the use of connections in a more complex, infinite dimensional context.

Example 27. (*Non-linear Cauchy-Riemann operator*) Let us take a symplectic manifold (P,ω) (smooth manifold P with a closed nondegenerate differential 2-form ω) with an almost complex structure J. From Floer theoretical viewpoint it is interesting to describe a smooth structure on a set of perturbed pseudo holomorphic maps. These are the maps

$$u : \mathbb{R} \times [0,1] \to P,\ u = u(s,t),$$

that satisfy the Cauchy-Riemann equation

$$\frac{\partial u}{\partial s} + J\left(\frac{\partial u}{\partial t} - X_H(u)\right) = 0. \tag{10}$$

We ask for some additional boundary condition, namely the image of $\mathbb{R}\times\{0\}$ and $\mathbb{R}\times\{1\}$ should be on appropriate Lagrangian submanifolds of P. The equation (10) is perturbed with Hamiltonian vector field X_H. We can denote the left hand–side of equation (10) by

$$\overline{\partial}_J(u) = \frac{\partial u}{\partial s} + J\frac{\partial u}{\partial t} - JX_H(u).$$

We are interested in the zeroes of this operator. Standard procedure is to look at $\overline{\partial}_J$ as a section of a suitable Banach bundles (see [8] for details). We define $\mathcal{E}$ to be the set of $W^{1,p}_{loc}$-maps that decay exponential at the positive and negative end of the strip $\mathbb{R} \times [0,1]$. The main idea is to prove that the map (or better to say the section)

$$\overline{\partial}_J : \mathcal{E} \to \bigcup_{u \in \mathcal{E}} L^p(u^*TP),$$

is a Fredholm map, i.e. its first derivative is a Fredholm operator[9] between two Banach spaces—the tangent spaces of its domain and the fibre of a bundle $\mathcal{E}$. In order to prove that, we have to find a linearization of $\overline{\partial}_J$. Let ∇ be Levi-Civita connection on P (which we defined in Example 22). It easily follows that the linearization of this section at its zeroes is

$$D\overline{\partial}_J(u)\xi = \nabla_{\frac{\partial u}{\partial s}}\xi + J(u)\nabla_{\frac{\partial u}{\partial t}}\xi + (\nabla_\xi J(u))\frac{\partial u}{\partial t} - \nabla_\xi(JX_H(u)),$$

and, as we already mentioned, it does not require any additional structure. Note that D is a covariant derivative on infinite dimensional bundle $\mathcal{E}$, constructed with the help of Levi-Civita connection on a finite dimensional bundle P. Hence, $D\overline{\partial}$ is the covariant derivative of a section $\overline{\partial}$ on $\mathcal{E}$, i.e. its linearization in a sense of Example 13, but in an infinite dimensional situation. This linearized operator is an elliptic operator and from elliptic regularity follows that all solutions of the Cauchy-Riemann equation are smooth.

Due to the fact that Fredholm operators have a finite dimensional kernel, the Fredholmness of $\overline{\partial}_J$ has a consequence that the set

$$\mathcal{M} := \{u \mid \overline{\partial}_J(u) = 0\}$$

of pseudo holomorphic maps is a finite dimensional submanifold of an infinite dimensional manifold of all smooth maps u. The finite dimensional manifold $\mathcal{M}$ gives rise to several interesting topological invariants of the original manifold P. For example, by counting (in the sense of enumerative Algebraic Geometry) the number of maps $u \in \mathcal{M}$ that intersect given homology classes in P, one can construct a new algebraic structure in a cohomology ring $H^*(P;\mathbb{Z})$. This new product is called *quantum product* and can be viewed as a deformation (and in a certain sense a generalization) of a standard cup product in cohomology (see [10]). $\triangle$

Acknowledgments. The author thanks Jelena Katić, Darko Milinković and Vukašin Stojisavljević for useful comments during the preparation of this article.

[9] A Fredholm operator is a bounded linear operator between two Banach spaces, with finite-dimensional kernel and cokernel, and with closed range.

REFERENCES

[1] R. Abraham, J. E. Marsden, *Foundations of Mechanics*, Second Edition, Addison-Wesley Publishing Company, 1987.

[2] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer Graduate Text in Mathematics 60, Springer, New York, 1978.

[3] M. P. do Carmo, *Riemannian Geometry*, Birkhauser, Boatan (1992).

[4] J. Dieudonné, *Treatise on analysis*, Volume 1.

[5] S. K Donaldson, *Polynomial invariants for smooth four–manifolds*, Topology **29** (1990), 257–315.

[6] S. K. Donaldson, P.B. Kronheimer, *The Geometry of Four–Manifolds*, Oxford Mathematical Monographs, 1997.

[7] M. Gromov, *Pseudo holomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985), 307–347.

[8] J. Katić, *Modulski prostori kombinovanog tipa u Mors-Florovoj teoriji*, PhD Thesis (in Serbian).

[9] J. M. Lee, *Introduction to Smooth Manifolds*, Springer 2003.

[10] D. McDuff, D. Salamon, *J-holomorphic Curves and Quantum Cohomology*, Amer. Math. Society, 1994.

[11] J. W. Morgan, *The Selberg-Witten Equations and Applications to the Topology of Smooth Four-Manifolds*, Mathematical Notes, Vol. 44, 1995.

[12] J. Milnor, *Morse theory*, Ann. Math. Studies 51, Princeton Univ. Press, 1963.

[13] J. Milnor, J. Stasheff, *Characteristic Classes*, Princeton University Press, Princeton, New Jersey, 1975.

[14] R. W. Sharpe, *Differential Geometry, Cartan's Generalization of Klein's Erlanger Programm*, Graduate Texts in Mathematics 166, Springer-Verlag, New York, 1997.

[15] R. W. Sharpe, *An introduction to Cartan geometries*, in: Proceedings of the 21st Winter School "Geometry and Physics" (Srní, 2001), Rend. Circ. Mat. Palermo (2) Suppl. **69** (2002), 61–75.

[16] M. Spivak, *A comprehesive Introduction to Differential Geometry*, Publish or Perish Inc, 1979.

[17] E. Witten, *Monopoles and four-manifolds*, Math. Research Letters **1** (1994), 769–706.

[18] C. H. Taubes, R. Wentworth, *Selberg-Witten and Gromov invariants for symplectic 4-manifolds*, International Press of Boston, 2010.

[19] V. A. Zorich, *Mathematical Analysis II*, Springer-Verlag.

Faculty of Mathematics, University of Belgrade, Belgrade, Serbia

E-mail: `jovanadj@matf.bg.ac.rs`