Introducing the tangent space
Mathematics of Machine Learning

Evgenii Samutichev

January 24, 2024

Definition 3.14. Let $\mathcal{M}$ be a **subset** of $\mathcal{E}$. For all $x \in \mathcal{M}$, define

$$\mathrm{T}_x\mathcal{M} = \{c'(0) \mid c : I \to \mathcal{M} \text{ is smooth and } c(0) = x\}$$

where I is any open interval containing $t = 0$. That is, $v \in \mathrm{T}_x\mathcal{M}$ if and only if there exists a smooth curve on $\mathcal{M}$ passing through x with velocity v.

Discussion The introduced set of vectors $\mathrm{T}_x\mathcal{M}$ will later be called the **tangent space** to $\mathcal{M}$ at x. And if we draw this for a sphere, we see that it makes a lot of sense. In particular, latitude and longtitude lines may be viewed as curves providing basis vectors for this space
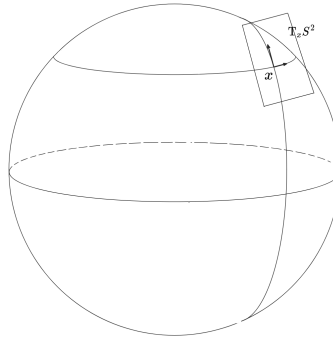
Notice how it's defined for a set $\mathcal{M}$, which might not be a manifold. This is similar to how in real analysis we are comfortable with having piecewise-smooth curves s.t. they don't have tangent line everywhere. Ok, but you might reasonably ask:

Haven't we defined the tangent space before through the kernel of the differential?

The answer is: we sort of defined it, but this definition is much better since it is **coordinate-free** (and there is this whole program in differential geometry to replace coordinate dependent definitions). What that means, is that if you want to define the tangent space through the kernel of the differential for $n = d-k$-dimensional manifold, you'll need this local defining function $h : U \to \mathbb{R}^k$ at x. And what if it so happens, that when you take some other local defining function $\tilde{h}$, then the resulting tangent space will be different? This is not obvious at all, and I'll give you an example in a minute, but we still have some more things to discuss here.

So, the second question you might want to ask:

Is it really a vector space?

The thing is, we can't really "add" curves. If there are two smooth curves providing us vectors v_1 and v_2, can we always construct a smooth curve passing through x with

velocity $v_1 + v_2$? When $\mathcal{M}$ is not a manifold, this problem will arise. Consider the following set $\mathcal{M}$
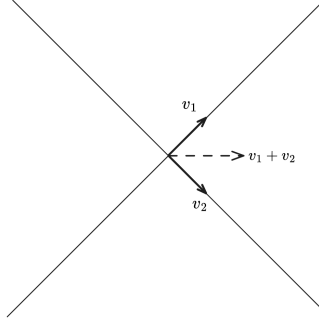
Notice that there is a singularity at the origin, and so for two vectors that are clearly in $\mathrm{T}_x\mathcal{M}$ their sum is not.

Now, I want to show you by providing an example that the uniqueness of the tangent space is not obvious.

Example Let's consider

$$S^1 = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\} \tag{1}$$

and notice that for $x \neq 0$

$$x^2 + y^2 - 1 = 0 \Leftrightarrow x^3 + xy^2 - x = 0$$

then

$$\begin{aligned} h(x,y) &= x^2 + y^2 - 1 \\ \tilde{h}(x,y) &= x^3 + xy^2 - x \end{aligned} \tag{2}$$

both are local defining for S^1 everywhere except $(0, \pm 1)$.

We can compute that

$$\mathrm{D}h(x,y)[v] = 2xv_x + 2yv_y \tag{3}$$

and

$$\mathrm{D}\tilde{h}(x,y)[v] = (3x^2 + y^2 - 1)v_x + 2xyv_y \tag{4}$$

Q: Just by looking at this, are you immediately convinced that the kernels of these two linear transformations are the same?

For points satisfying $x^2 + y^2 = 1$ the second linear map can be written as

$$\mathrm{D}\tilde{h}(x,y)[v] = 2x^2 v_x + 2xyv_y \tag{5}$$

so

$$\mathrm{D}\tilde{h}(x,y) = x\mathrm{D}h(x,y) \tag{6}$$

Although clearly

$$\mathrm{D}h(x,y) \neq \mathrm{D}\tilde{h}(x,y) \tag{7}$$

but

$$\ker \mathrm{D}h(x,y) = \ker \mathrm{D}\tilde{h}(x,y) \tag{8}$$

which wasn't initially obvious, since we had to perform certain transformations to achieve this. For $(x,y) \notin S^1$ this isn't even true in general.

Now look back at the Definition 3.14. We clearly don't use any local defining function h in it, it only involves the concept of a smooth path on the manifold, which is intrinsic to its geometry, and not to a particular choice of its coordinate representation. And keep in mind that we call it *the* tangent space, which implies that it's unique. But, again, are we sure about that?

The following theorem provides the connection between our intuitive definition and the actual definition of the tangent space, while answering questions asked in the beginning of the lecture.

Theorem 3.15. Let $\mathcal{M}$ be an embedded submanifold of $\mathcal{E}$. Consider $x \in \mathcal{M}$ and the set $T_x\mathcal{M}$. Then

1. If $\mathcal{M}$ is an open submanifold, then $T_x\mathcal{M} = \mathcal{E}$

2. Otherwise, $T_x\mathcal{M} = \ker Dh(x)$ with h **any** local defining function at x

Discussion But before we get to the proof, I would like to briefly remind you the geometric side of **Theorem 3.12** from the previous lecture.

As an example, let's look at a sphere again. Consider $x = (x_1, x_2, x_3)$ and around
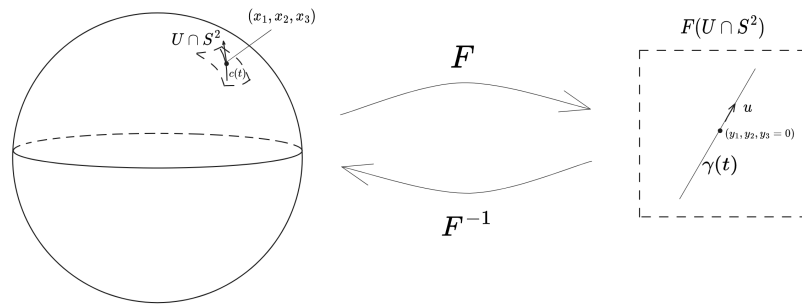
that point, meaning in some neighborhood U we have a diffeomorphism F s.t.

$$F(U \cap S^2) = V \cap E \tag{9}$$

where $E = \{y \in \mathbb{R}^3 \mid y_3 = 0\}$ and V is some neighborhood of $F(x)$. So it locally flattens the sphere. But this can also be interpreted as a change of coordinates

$$(x_1, x_2, x_3) \to F(x) = (y_1, y_2, y_3) \tag{10}$$

In these new coordinates we can easily define straight path, which I denoted as $\gamma(t)$ with a given velocity vector u. In fact, it's a 3-dimensional vector, but its third component is zero. Then we return back to the original coordinates and get the path $c(t)$ on a sphere. Of course the velocity vector will transform somehow, but this is not important as long as all such vectors form a vector space (in this case 2-dimensional). In the proof of **Theorem 3.15** we will make this statement more formal, so let's finally get to the proof, which should now be fairly easy to understand.

Proof of theorem 3.15.
Part 1 For open submanifolds the central idea, is that with each $x \in \mathcal{M}$ there is an open ball $B(x; \varepsilon) \subseteq \mathcal{M}$, so for every $0 \neq v \in \mathcal{E}$ we can always take $I = \left(-\frac{\varepsilon}{\|v\|}, \frac{\varepsilon}{\|v\|}\right)$ and

$$c : I \to \mathcal{M}, \quad c(t) = x + tv \tag{11}$$

then $c(0) = x$ and $c'(0) = v$. Thus, $\mathcal{E} \subseteq \mathrm{T}_x\mathcal{M}$ and so $\mathrm{T}_x\mathcal{M} = \mathcal{E}$. For $v = 0$ we can just take the "path" remaining at the same point x for all $t \in I$.

Part 2 Consider the case of $\mathcal{M}$ an embedded submanifold of dimension $n = d - k$ with $k \geq 1$. Let $h : U \to \mathbb{R}^k$ be a local defining function of $\mathcal{M}$ around x. We will prove the result in two steps.

The first step is straightforward, using the chain rule, we show that every $\mathrm{T}_x\mathcal{M} \subseteq \ker \mathrm{D}h(x)$.

If $v \in \mathcal{M}$, then there exists $c : I \to \mathcal{M}$, smooth, s.t. $c(0) = x$ and $c'(0) = v$. Since $c(t)$ is in $\mathcal{M}$ we have

$$h(c(t)) = 0 \tag{12}$$

both function are smooth, so we can apply the chain rule

$$\mathrm{D}h(c(t))[c'(t)] = 0 \tag{13}$$

thus at $t = 0$

$$\mathrm{D}h(x)[v] = 0 \tag{14}$$

and $v \in \ker \mathrm{D}h(x)$.

For the second step, using **Theorem 3.12** we will perform the coordinate transformation, construct the neccessary path for each $u \in \mathbb{R}^{d-k}$ similar to how we did it with S^1, which in turn implies that $\mathrm{T}_x\mathcal{M}$ contains a linear subspace of the same dimension as $\ker \mathrm{D}h(x)$. Then immediately $T_x\mathcal{M} = \ker \mathrm{D}h(x)$.

Q: Is this obvious to everyone here? (if not, just refer to the basis vectors and step 1)

We get diffeomorphism $F : U \to V$ from **Theorem 3.12** s.t. U is a neighborhood of x and

$$F(\mathcal{M} \cap U) = E \cap V \tag{15}$$

where $E = \{y \in \mathbb{R}^d \mid y_{n+1} = 0, ..., y_d = 0\}$, so $\dim E = \dim M$. For $u \in \mathbb{R}^{d-k}$ Define

$$\gamma : I \to \mathbb{R}^d, \quad \gamma(t) = F(x) + t \begin{bmatrix} u \\ 0_k \end{bmatrix} \tag{16}$$

where 0_k is a zero vector of dimension k, and I is picked so γ remains in $E \cap V$ (open set).

Now, let's define

$$c(t) = F^{-1}(\gamma(t)) \tag{17}$$

then $c(t) \in \mathcal{M}$ from $F^{-1}(E \cap V) = \mathcal{M} \cap U$ and our choice of I. Notice that

$$c(0) = F^{-1}(\gamma(0)) = F^{-1}(F(x)) = x \tag{18}$$

Moreover, F is a diffeomorphism, so F^{-1} is smooth, and we can apply the chain rule again

$$c'(t) = \mathrm{D}F^{-1}(\gamma(t))[\gamma'(t)] \tag{19}$$

in particular for $t = 0$

$$c'(0) = [\mathrm{D}F^{-1}(F(x))] \begin{bmatrix} u \\ 0_k \end{bmatrix} \tag{20}$$

and $[\mathrm{D}F^{-1}(F(x))] = (\mathrm{D}F(x))^{-1}$ is invertible from the inverse function theorem (which was used to prove **Theorem 3.12**). What's crucial, is that it only depends on x and linear, so it's an isomorphism between $\mathbb{R}^n$ (where u lives) and some subset of $\mathrm{T}_x\mathcal{M}$. This is exactly what we wanted to achieve step 2, and ends the proof.

$\square$

Now we are confident in saying that

Definition 3.15. the previously defined $\mathrm{T}_x\mathcal{M}$ we call the tangent space to $\mathcal{M}$ at x. Vectors in $\mathrm{T}_x\mathcal{M}$ are called **tangent vectors**.

In addition, it follows from the proof of the theorem

$$\dim \mathrm{T}_x\mathcal{M} = \dim \mathcal{M} \tag{21}$$

so the dimension of the tangent space is independent of x.

Example 3.17. The linear space $\mathcal{M} = \mathbb{R}^d$ is a linear manifold of dimension d, because it's open in $\mathbb{R}^d$. It's tangent spaces $\mathrm{T}_x\mathcal{M} = \mathbb{R}^d$ are given by **Theorem 3.15 (1)**.

The affine space $\mathcal{A} = \{x \in \mathbb{R}^d \mid Ax = b\}$ defined by a matrix $A \in \mathbb{R}^{k \times d}$ or rank k and arbitrary vector $b \in \mathbb{R}^k$ is a manifold of dimension $n = d - k$ embedded in $\mathbb{R}^d$.

Proof. Define $h : \mathbb{R}^d \to \mathbb{R}^k$ as

$$h(x) = Ax - b \tag{22}$$

then $x \in \mathcal{A} \Leftrightarrow h(x) = 0$, and

$$\mathrm{D}h[x] = A \tag{23}$$

so $\operatorname{rank} \mathrm{D}h(x) = \operatorname{rank} A = k$, which means that h is a defining function for $\mathcal{A}$ for all x, and by definition $\mathcal{A}$ is a manifold. As for it's tangent space, we have by **Theorem 3.15 (2)**

$$\mathrm{T}_x\mathcal{A} = \ker \mathrm{D}h(x) = \ker A = \{x \in \mathbb{R}^d \mid Ax = 0\} \tag{24}$$

$\square$

Example 3.18. The sphere $S^{d-1} = \{x \in \mathbb{R}^d \mid x^T x = 1\}$ is the zero level set of

$$h : \mathbb{R}^d \to \mathbb{R}, \quad h(x) = x^T x - 1 \tag{25}$$

which is smooth. As we discussed

$$\mathrm{D}h(x)[v] = 2x^T v \tag{26}$$

and for all $x \in S^d$ at least one coordinate is non-zero, so

$$\operatorname{rank} \mathrm{D}h(x) = 1 \tag{27}$$

as a result S^{d-1} is an embedded submanifold of $\mathbb{R}^d$ of dimension $n = d - 1$. While it's tangent space is (again, by **Theorem 3.15 (2)**)

$$\mathrm{T}_x S^{d-1} = \ker \mathrm{D}h(x) = \{v \in \mathbb{R}^d \mid x^T v = 0\} \tag{28}$$

Example 3.19. Consider

$$\mathrm{Sym}(2)_1 = \left\{ X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \;\middle|\; \operatorname{rank} X = 1 \right\} \tag{29}$$

this is a subset of $\mathrm{Sym}(2)$, the linear space of symmetric matrices of size two, that's diffeomorphic to $\mathbb{R}^3$. rank is not a smooth function, so it can't serve as a local defining

function. However, it's possible to construct local defining function as follows. Let $U = \mathrm{Sym}(2) \setminus \{0\}$, which is the open subset of $\mathrm{Sym}(2)$. And consider

$$h : U \to \mathbb{R}, \quad h(X) = x_1 x_3 - x_2^2 \tag{30}$$

clearly h is smooth, and $h^{-1}(0) = \mathrm{Sym}(2)_1$. Furthermore

$$\mathrm{D}h(X)[\dot{X}] = x_3 \dot{x}_1 + x_1 \dot{x}_3 - 2x_2 \dot{x}_2 = \begin{bmatrix} x_3 & -2x_2 & x_1 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} \tag{31}$$

where $\dot{X}$ is a matrix in $\mathrm{Sym}(2)$. This linear map has rank one provided $X \neq 0$, which is always the case in U. Thus, h is a defining function for $\mathrm{Sym}(2)_1$ everywhere. Which confirms that $\mathrm{Sym}(2)_1$ is a 2-dimensional submanifold of $\mathrm{Sym}(2)$.

   Q: What's the geometric representation of this manifold?

$$x_1 x_3 - x_2^2 = \frac{1}{2}\left((x_1 + x_3)^2 - (x_1 - x_3)^2\right) - x_2^2 \tag{32}$$

thus

$$x_1 x_3 - x_2^2 = 0 \Leftrightarrow (x_1 + x_3)^2 = (2x_2)^2 + (x_1 - x_3)^2$$

after the change of variables

$$\begin{cases} z_1 = x_1 - x_3 \\ z_2 = 2x_2 \\ z_3 = x_1 + x_3 \end{cases} \tag{33}$$

we get

$$z_1^2 + z_2^2 = z_3^2 \tag{34}$$

so it's an elliptic cone, but with excluded origin. It's not connected, neither open, nor closed. So it's a very non-trivial manifold in this sense.

   However, we have enough machinery to study it. For example, it's tangent space at X can be computed by **Theorem 3.15**.

$$\mathrm{T}_X \mathrm{Sym}(2)_1 = \{\dot{X} \mid x_3 \dot{x}_1 + x_1 \dot{x}_3 - 2x_2 \dot{x}_2 = 0\} \tag{35}$$