Math. 396. Notes I: Parametric curves

Let I denote an interval in $\mathbb{R}$. It could be an open interval (a,b), or a closed interval $[a,b]$, or semi-closed $[a,b),(a,b]$. Here a,b are real numbers or infinity symbols $-\infty,\infty$. For example, $\mathbb{R}$ is the interval $(-\infty,\infty)$.

Recall that a map of sets $f:A\to B$ is called *injective* if $f(a)=f(a')$ implies $a=a'$ (or, equivalently, the pre-image of any element of B is either empty or consists of one element). The map f is called *surjective* if the image set $f(A)$ coincides with B. A map which is both injective and surjective is called *bijective* or *one-to-one*.

Intuitively we imagine that the image of a map $\mathbf{f}:I\to\mathbb{R}^n$ is a curve in $\mathbb{R}^n$. When t moves along $I=[a,b]$ from a to b, the value $\mathbf{f}(t)$ moves in $\mathbb{R}^n$ from $\mathbf{f}(a)$ to $\mathbf{f}(b)$.

Example 1. Take $\mathbf{f}(t)=(\cos t,\sin t)$ and $[a,b]=[0,2\pi]$. We know that $\cos^2 t+\sin^2 t=1$, hence $\mathbf{f}(t)\in S^1(0;1)$ for any t. Conversely, if $(a,b)\in S^1=\{(x,y)\in\mathbb{R}^2: x^2+y^2=1\}$, then the complex number $z=a+bi\in S^1$ can be written in the parametric form $z=\cos t+i\sin t$ for a unique $t\in[0,2\pi)$. This shows that $\mathbf{f}([0,2\pi])=S^1$.

Example 2. Take $\mathbf{f}(t)=(\frac{t^2-1}{t^2+1},\frac{2t}{t^2+1}):\mathbb{R}\to\mathbb{R}^2$. We have

$$(\frac{t^2-1}{t^2+1})^2+(\frac{2t}{t^2+1})^2=\frac{t^4-2t^2+1+4t^2}{(t^2+1)^2}=\frac{(t^2+1)^2}{(t^2+1)^2}=1.$$

Thus the image of our map is contained in the unit circle S^1. Do we get all points? Since the first coordinate $\frac{t^2-1}{t^2+1}<1$ we will never get a point $(1,0)$. Assume that $a^2+b^2=1$ and $(a,b)\neq(1,0)$. We have

$$\frac{(\frac{b}{a-1})^2-1}{(\frac{b}{a-1})^2+1}=\frac{b^2-(a^2-2a+1)}{b^2+a^2-2a+1}=\frac{1-a^2-a^2+2a-1}{1-2a+1}=\frac{-2a^2+2a}{2-2a}=a,$$

$$\frac{\frac{2b}{a-1}}{(\frac{b}{a-1})^2+1}=\frac{2b(1-a)}{b^2+(1-a)^2}=\frac{2b(1-a)}{b^2+a^2+1-2a}=\frac{2b(1-a)}{2-2a}=b.$$

This shows that, if $t=\frac{b}{a-1}$ we get $\mathbf{f}(t)=(a,b)$. So all points $(a,b)\neq(1,0)$ in S^1 are covered. Observe that

$$(1,0)=\lim_{t\to\infty}\mathbf{f}(t)=\lim_{t\to-\infty}\mathbf{f}(t).$$

So, if we add ∞ and $-\infty$ to $\mathbb{R}$ and "join" them we get a bijective map from the obtained set to the unit circle.

Definition. A *parametric curve* or a *path* in $\mathbb{R}^n$ is a non-constant continuous map $\mathbf{f}:I\to\mathbb{R}^n$ where I is an interval in $\mathbb{R}$. A subset of $\mathbb{R}^n$ is called a *curve* if it is equal to the image of a parametric curve. If $I=[a,b]$ is a finite closed interval, the parametric curve is called a *path* with the origin at $\mathbf{f}(a)$ and the end at $\mathbf{f}(b)$. If $\mathbf{f}(a)=\mathbf{f}(b)$, we say that the path is *closed*. A curve C is closed if it is equal to the image of a closed path. A parametric curve $\mathbf{f}:I\to\mathbb{R}^n$ is called differentiable if the function $\mathbf{f}$ is differentiable. More general, we say that a path belongs to class C^k if $\mathbf{f}$ belongs to class C^k, i.e. the k-derivative of $\mathbf{f}$ exists and continuous at each $t\in I$. A path of class C^∞ is called a *smooth path*.

Warning. If we don't assume anything on $\mathbf{f}$ then its image could be the whole space $\mathbb{R}^n$. An example is the Peano curve.

Example 3. Let $v = (a_1, \dots, a_n), w = (b_1, \dots, b_n) \in \mathbb{R}^n$. Define $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ by the formula

$$\mathbf{f}(t) = v + tw = (a_1 + tb_1, \dots, a_n + tb_n).$$

For example if v is the origin $0 = (0, \dots, 0)$, the image of this map consists of all vectors which are proportional to w. In the case $n = 2$, it will be the line through the origin in the direction of the vector w. If $n = 2$ and $v \neq 0$ we can view the image as the line parallel to the previous line which passes through v. For any n we call the image of the map $\mathbf{f}$ the *line* through the vector v in the direction of the vector w. Since for any scalars λ and $\mu \neq 0$, we can always write

$$v + tw = (v + \lambda w) + (t - \lambda)w = v + (t/\mu)\mu w$$

we obtain that the same line can be defined by replacing v by any vector in it, and replacing w by any non-zero vector proportional to w.

Example 4. Assume $n = 2$. Let us view $\mathbb{R}^2$ as the field of complex numbers $\mathbb{C}$. Then, using the trigonometric form of complex numbers, we can define a parametric curve $\mathbf{f} : \mathbb{R} \to \mathbb{C}$ by the formula

$$\mathbf{f}(t) = r(t)(\cos\phi(t) + i \sin\phi(t)) = r(t)e^{i\phi(t)}.$$

Here, we don't necessary assume that $\phi(t) \in [0, 2\pi)$. The old formula $\mathbf{f}(t) = (a(t), b(t))$ and the new formula (called sometimes *polar parametric curve*) are related to each other by the formulas: $\mathbf{f}(t) = (a(t), b(t))$ where

$$a(t) = r(t)\cos\phi(t), \quad b(t) = r(t)\sin\phi(t).$$

$$r(t) = (a(t)^2 + b(t)^2)^{1/2}, \phi(t) = \arctan(a(t)/b(t)).$$

Very often, the polar parametric curve is given in the form $r = f(\phi), \phi \in [a, b]$. This means $\mathbf{f}(t) = r(t)e^{it}$. For example one can divide a circle of radius r by $r(\phi) = r$ (a constant).

Let $\mathbf{f} : I \to \mathbb{R}^n$ be a differentiable path. For any interior point $t_0 \in I$ we can define the derivative vector $\mathbf{f}'(t_0)$. It is called the *velocity vector* of the parametric curve. Its length $||\mathbf{f}'(t_0)||$ is called the *speed* of the parametric curve at t_0. To explain the name consider the case when $n = 1$. The set of vectors

$$T(\mathbf{f})_{t_0} = \{\mathbf{f}(t_0) + \lambda \mathbf{f}'(t_0) \mid \lambda \in \mathbb{R}\}$$

is called the *tangent line* of the parametric curve at t_0. Here we shall assume that $\mathbf{f}'(t_0) \neq 0$. Otherwise the tangent line is not defined.

Definition A differentiable path $\mathbf{f} : I \to \mathbb{R}^n$ is called an *immersion* if $\mathbf{f}'(t) \neq 0$ for any $t \in I$. It is called a *piecewise immersion* if I is the union of intervals such that the restriction of $\mathbf{f}$ tothe interior of each subinterval is an immersion. The image of an immersion is called an *immersed curve.*

Thus for an imemrsed path the tangent line defined at each point $t \in I$.

Example 5 Let $I = \mathbb{R}$ and $\mathbf{f}(t) = (t^2, t^3)$. The image is the curve C given by the equation $y^2 - x^3 = 0$. The derivative is equal to $\mathbf{f}'(t) = (2t, 3t^2)$. The curve is not immersed at $t = 0$. However, it is piecewise immersed. If we restrict $\mathbf{f}$ to $(-\infty, 0)$ and $(0, \infty)$, the map is an immersion.

Example 6. Let $f : [a, b] \to \mathbb{R}$ be a continuous function and $\Gamma_f \subset \mathbb{R}^2$ be its graph. Consider the parametric curve

$$\gamma_f : [a, b] \to \mathbb{R}^2, \quad \gamma_f(x) = (x, f(x)).$$

Here we deliberately changed the letter t to x. Its image is the graph Γ_f of f. This curve is differentiable if and only if f is differentiable. Its derivative at x_0 is equal to $(1, f'(x_0))$. The tangent line is the set of vectors

$$(x, y) = (x_0, f(x_0)) + \lambda(1, f'(x_0)) = (x_0 + \lambda, f(x_0) + \lambda f'(x_0))$$

Obviously they satisfy

$$y - f(x_0) = f'(x_0)(x - x_0).$$

This is the usual equation of the tangent line of the graph Γ_f at the point $(x_0, f(x_0))$.

Example 7. Let $\mathbf{f}(t) = (r\cos t, r\sin t), t \in [0, 2\pi]$. The image is the circle $S^1(0; r)$ of radius r with center at the origin. The velocity vector at t_0 is

$$\mathbf{f}'(t_0) = (-r\sin t_0, r\cos t_0).$$

We have for any $t \in [0, 2\pi]$,

$$\mathbf{f}'(t_0) \cdot \mathbf{f}(t_0) = -r^2 \cos t_0 \sin t_0 + r^2 \sin t_0 \cos t_0 = 0.$$

Thus the line passing through the origin in the direction of $\mathbf{f}'(t_0)$ is perpendicular to the "radius vector" $\mathbf{f}(t_0)$. So the tangent line $T(\mathbf{f})_{t_0}$ is perpendicular to the radius vector $\mathbf{f}(t_0)$.

Example 8. Let $\mathbf{f}(t) = (t^2 - 1, t^3 - t) : (-\infty, \infty) \to \mathbb{R}^2$. For any (x, y) in the image we have

$$y^2 = t^2(t^2 - 1)^2 = x^2(x + 1).$$

Conversely, if $(x, y) \in \mathbb{R}^2$ satisfies

$$y^2 = x^2(x + 1),$$

and $x \neq 0$ we get $(y/x)^2 = x + 1$, so setting $t = y/x$ we get $x = t^2 - 1, y = tx = t^3 - t$. Clearly, $x = t^2 - 1 \geq -1$. If $x = 0$ we have $y = 0$. The point $(0, 0)$ is the value of $\mathbf{f}$ at $t = 1$ and $t = -1$. This shows that all points of the curve $y^2 = x^2(x + 1)$ are covered. We have $\mathbf{f}'(t) = (2t, 3t^2 - 1)$. So

$$\mathbf{f}'(-1) = (-2, 2), \quad \mathbf{f}'(1) = (2, 2),$$

however, as we have noticed before

$$\mathbf{f}(-1) = \mathbf{f}(1) = (0, 0).$$

So at the point $(0, 0)$ we have two different velocity vectors attached. This can be explained as follows. When t runs from -1 to ∞, the point $\mathbf{f}(t)$ passes twice through the point $(0, 0)$. So one cannot attach one tangent line to the image curve at the point $(0, 0)$. There are two tangent lines $T(\mathbf{f})_1$ and $T(\mathbf{f})_{-1}$.

Warning. The velocity vector and the speed obviously depend on the parametrization of the image curve. For example, if we replace $\mathbf{f} : I \to \mathbb{R}^n$ by the composition $\mathbf{g} = \mathbf{f} \circ \phi : J \to \mathbb{R}^n$, where $\phi : J \to \mathbb{R}$ is a differentiable function with image equal to I, then $\mathbf{g}(J) = \mathbf{f}(g(J)) = \mathbf{f}(I)$. So the two parametric curves $\mathbf{g} : J \to \mathbb{R}^n$ and $\mathbf{f} : I \to \mathbb{R}^n$ have the same image curve. Write $\mathbf{f}(t) = (f_1(t), \ldots, f_n(t))$, then

$$\mathbf{g}(t) = (g_1(t), \ldots, g_n(t)) = (f_1(\phi(t)), \ldots, f_n(\phi(t)),$$

and, by the chain rule,

$$\mathbf{g}(t)' = (g_1'(t), \ldots, g_n'(t)) = (f_1(\phi(t))'\phi'(t), \ldots, f_n'(\phi(t)\phi'(t)) = \phi'(t)\mathbf{f}'(\phi(t)).$$

Obviously, $\mathbf{g}(t) = \mathbf{f}(\phi(t))$, the velocity vectors are proportional but not equal if $\phi'(t) \neq 1$. For example, if we take $\phi : [-2, \infty) \to [-1, \infty)$ defined by the formula $\phi(t) = t/2$, and consider $\mathbf{f}$ from Example 6, we get two parametrizations of the same curve $y^2 = x^2(x+1)$ given by $\mathbf{f} : [-1, \infty) \to \mathbb{R}^2$ and by $\mathbf{g} = \mathbf{f} \circ \phi : [-2, \infty) \to \mathbb{R}^2$. We have $\phi'(t) = 1/2$ and the velocity vectors satisfy $\mathbf{f}'(t/2)) = 2\mathbf{g}(t)$. For example, at $t = 2$ we get $\mathbf{g}(2) = \mathbf{f}(1) = (0,0)$ but $\mathbf{g}'(2) = (1,1)$, $\mathbf{f}'(1) = (2,2)$.

A path $\mathbf{f} : [a,b] \to \mathbb{R}^n$ is called *simple* if $\mathbf{f}$ is injective on $[a,b)$ and on $(a,b]$ (i.e. $\mathbf{f}(t_1) \neq \mathbf{f}(t_2)$ for $t_1 < t_2$ unless $t_1 = a, t_2 = b$). A curve is *simple* if it is equal to the image of a simple path. For example, a circle is a simple curve, but the curve from Example 8 is not simple. A simple path which is an immersion is called an *embedding.* A curve C is called *embedded* if it is the image of an embedding.

For example, the curve from Example 8 is immersed but not embedded. A half-circle is an embedded curve. The curve from example 5 is simple and smooth but not embedded.

Remark. Let C be the image of a simple closed path $\mathbf{f} : [a,b] \to C$. Any simple parametrization $\mathbf{g} : [c,d] \to C$ of C is closed. In fact, assume $\mathbf{g}(c) \neq \mathbf{g}(d)$. Then $\mathbf{g}^{-1} \circ \mathbf{f} : [a,b] \to [c,d]$ is a continuous map which defines a bijective continuous map $[a,b) \to [c,d]$. But this is impossible (prove it!).

Let $\mathbf{f} : [a,b] \to C \subset \mathbb{R}^n$ be a simple path with $\mathbf{f}(a) \neq \mathbf{f}(b)$. For any $x \in C$ we can find a unique $t \in [a,b]$ such that $x = \mathbf{f}(t)$. Let $x = \mathbf{f}(t), y = \mathbf{f}(t') \in C$ be points on C. We say that $v \leq w$ if $t \leq t'$. Let us see that this definition of order does not depend on the choice of injective parametrization $\mathbf{f}$ up to reversing the order.

Lemma 1. *Let $\mathbf{f} : [a,b] \to \mathbb{R}^n, \mathbf{g} : [c,d] \to \mathbb{R}^n$ be two injective continuous maps with the same image C, Let $v \leq_{\mathbf{f}} w$ be the relation of order in C defined by $\mathbf{f}$ and $v \leq_{\mathbf{g}} w$ be the relation of order in C defined by $\mathbf{g}$. Then either these order relations coincide or*

$$v \leq_{\mathbf{f}} w \Longleftrightarrow w \leq_{\mathbf{g}} v \quad \textit{for all } v, w \in C.$$

Proof. Consider the maps $\mathbf{f}$ and $\mathbf{g}$ as bijective maps onto their image C. Consider the composition map $\phi : \mathbf{g}^{-1} \circ \mathbf{f} : [a,b] \to [c,d]$. Suppose we prove that this map is continuous. Since ϕ is bijective, ϕ is either strictly monotonous increasing or strictly monotonous decreasing. In the first case

$$\mathbf{f}(t) <_{\mathbf{f}} \mathbf{f}(t') \Longleftrightarrow t < t' \Longleftrightarrow \phi(t) < \phi(t') \Longleftrightarrow \mathbf{g}(\phi(t)) <_{\mathbf{g}} \mathbf{g}(\phi(t') \Longleftrightarrow \mathbf{f}(t) <_{\mathbf{g}} \mathbf{f}(t').$$

Here in the last implication we used that $\mathbf{g} \circ \phi = \mathbf{f}$. In the second case we similarly obtain

$$\mathbf{f}(t) <_{\mathbf{f}} \mathbf{f}(t') \Longleftrightarrow t < t' \Longleftrightarrow \phi(t) > \phi(t') \Longleftrightarrow \mathbf{g}(\phi(t)) >_{\mathbf{g}} \mathbf{g}(\phi(t') \Longleftrightarrow \mathbf{f}(t) >_{\mathbf{g}} \mathbf{f}(t').$$

So let us prove that ϕ is continuous.

Claim: If $\mathbf{h} : [\alpha, \beta] \to \mathbb{R}^n$ is any continuous injective map, then for any $\delta > 0$ there exists $r > 0$ such that for any $x, x' \in [\alpha, \beta]$,

$$||\mathbf{h}(x) - \mathbf{h}(x')|| < r \Longrightarrow |x - x'| < \delta.$$

The proof is by contradiction. Assume it is false. Then there exists such δ that for any $r > 0$ one can find $t, t' \in [a, b]$ with $||\mathbf{h}(t) - \mathbf{h}(t')|| < r$ but $|t - t'| > \delta$. Taking $r = 1/n$ we find two sequences $\{t_n\}, \{t'_n\}$ such that $|t_n - t'_n| > \delta$ but $||\mathbf{h}(t) - \mathbf{h}(t')|| < 1/n$. Since $[a, b]$ is compact, we can replace these sequences by converging subsequences to assume that $\lim_{n\to\infty}\{t_n\} = c$ and $\lim_{n\to\infty}\{t'_n\} = c'$ for some $c, c' \in [a, b]$. Obviously $|c - c'| \geq \delta$ so that $c \neq c'$. Since $\mathbf{h}(t)$ is continuous, $\lim_{n\to\infty} \mathbf{h}(t_n) = \mathbf{h}(c), \lim_{n\to\infty} \mathbf{h}(t'_n) = \mathbf{h}(c')$. Now I can make $||\mathbf{h}(t_n) - \mathbf{h}(c)||, ||\mathbf{h}(t'_n) - \mathbf{h}(c')||, ||\mathbf{h}(t_n) - \mathbf{h}(t'_n)||$ arbitrary small when n is large enough, so, by the triangle inequality, I can make $||\mathbf{h}(c) - \mathbf{h}(c')|| = ||(\mathbf{h}(c) - \mathbf{h}(t_n)) + (\mathbf{h}(t_n) - \mathbf{h}(t'_n)) + (\mathbf{h}(t'_n) - \mathbf{h}(c'))||$ arbitrary small. This of course implies that $||\mathbf{h}(c) - \mathbf{h}(c')|| = 0$, hence $\mathbf{h}(c) = \mathbf{h}(c')$. This contradicts the injectivity of the map $\mathbf{h}$.

Now let us use this claim to prove the uniform continuity of ϕ. Recall that this means that for any $\epsilon > 0$ there exists $\delta > 0$ such that $|t - t'| < \delta \Longrightarrow |\phi(t) - \phi(t')| < \epsilon$. Apply the claim to the function $\mathbf{g}$ and take $x = \phi(t), x' = \phi(t')$. Then we find r such that

$$||\mathbf{f}(t) - \mathbf{f}(t')|| = ||\mathbf{g}(\phi(t)) - \mathbf{g}(\phi(t'))|| < r \Longrightarrow |\phi(t) - \phi(t')| < \epsilon.$$

Since $\mathbf{f}$ is continuous on a compact interval $[a, b]$, it is uniformly continuous. Thus there exists $\delta > 0$ such that $|t - t'| < \delta \Longrightarrow ||\mathbf{f}(t) - \mathbf{f}(t')|| < r$, hence

$$|t - t'| < \delta \Longrightarrow |\phi(t) - \phi(t')| < \epsilon.$$

This proves the uniform continuity of ϕ, hence the continuity of ϕ.

Definition Let C be any simple curve. A choice of order on C is called an *orientation*. A curve together with its orientaion is called an *oriented curve* or a *directed curve*.

It follows from Lemma 1 that any two simple parametrizations $\mathbf{f} : I \to C$ and $\mathbf{g} : J \to C$ of a simple not closed curve C differ by a continuous bijective map $\phi : I \to J$ (so that $\mathbf{f} = \mathbf{g} \circ \phi$). We say that two simple parametrizations are *equivalent* if they define the same orientation. There are two equivalence classes of simple orientations. Two parametrizations $\mathbf{f}$ and $\mathbf{g}$ belong to the same equivalence class if only if the map ϕ is strictly monotonously increasing. Thus we can say that an orientation of a simple curve C is a choice of an equivalence class of a simple parametrization.

Definition. A path $\mathbf{f} : [a,] \to C$ is called *piecewise simple* if there exist $a = t_0 < t_1 < \ldots < t_{k-1} < t_k = b$ such that the restriction of $\mathbf{f}$ to each $[t_{i-1}, t_i]$ is a simple path.

Observe that each simple path is piecewise simple. Any subdivision of $[a, b]$ will satisfy the definition above.

Suppose we have two piecewise simple paths $\mathbf{f} : [a, b] \to C$ and $\mathbf{g} : [c, d] \to C$ with the same image C. Let $[t_{i-1}, t_i]$ and $[t'_{j-1}, t'_j]$ be the corresponding subintervals. By subdividing further, we can assume that $\mathbf{g}([t'_{j-1}, t'_j]) = \mathbf{f}([t_{i-1}, t_i])$. Then, using Lemma 1, we find bijective continuous maps $\phi_i : [t_{i-1}, t_i] \to [t'_{j-1}, t'_j]$ such that $\mathbf{g} = \mathbf{f} \circ \phi$, where $\phi : [a, b] \to [c, d]$ is the unique map such that its restriction to each interval $[t_{i-1}, t_i]$ coincides with ϕ_i. It is easy to see that ϕ is a continuous map. Also, if ϕ_1 is strictly increasing on $[t_0, t_1]$, then ϕ is strictly increasing on the whole $[a, b]$. We say that two piecewise simple parametrizations are equivalent if they obtained in the above way with ϕ strictly increasing. We define an orientation of a piecewise simple path, as a choice of one of the two equivalence classes of piecewise simple parametrizations. In this way we can define an orientation of a simple closed curve, since it is obviously piecewise simple.

Now let $C \subset \mathbb{R}^n$ be an embedded curve. Choose an injective parametrization $\mathbf{f} : I \to C$ such that $\mathbf{f}'(t)$ exists and non-zero at each interior point t in I. Then we can define the *tangent line*

$T(C)_x$ at a point $x = \mathbf{f}(t)$ of C as the line passing through x parallel to the velocity vector $\mathbf{f}'(t)$. Note that the parameter t is determined by the point x uniquely since the map $\mathbf{f}$ is injective. Let us show the tangent line $T(C)_x$ does not depend on the choice of the parametrization.

This fact follows from the following:

Lemma 2. *Let us keep the notation of Lemma 1 and its proof. Assume additionally that* $\mathbf{f}$ *is differentiable at* $t_0 \in (a,b)$, $\mathbf{g}$ *is differentiable at* $\phi(t_0)$, *and* $\mathbf{f}'(t_0) \neq 0, \mathbf{g}'(\phi(t_0) \neq 0$. *Then* ϕ *is differenetiable at* t_0 *and*

$$\mathbf{f}'(t_0) = \mathbf{g}'(\phi(t_0))\phi'(t_0).$$

Proof. Of course the last assertion is just the chain rule. In its proof we have to assume that $\phi'(t_0)$ exists. So, we have to prove that the latter exists under our assumptions on $\mathbf{f}$ and $\mathbf{g}$. Since $\mathbf{f}$ and $\mathbf{g}$ are injective, $\phi = \mathbf{g}^{-1} \circ \mathbf{f}$ is injective too. So $\phi(t) - \phi(t_0) \neq 0$ for any $t \in [a,b]$. Write

$$\frac{\mathbf{g}(\phi(t)) - \mathbf{g}(\phi(t_0))}{\phi(t) - \phi(t_0)} \cdot \frac{\phi(t) - \phi(t_0)}{t - t_0} = \frac{\mathbf{f}(\phi(t)) - \mathbf{f}(\phi(t_0))}{t - t_0}.$$

Now we can go to the limit when $t \to t_0$. By Lemma 1, ϕ is a continuous function, so $\lim_{t\to t_0} \phi(t) = \phi(t_0)$. Since $\mathbf{f}$ and $\mathbf{g}$ are differentiable at t_0 and $\phi(t_0)$, respectively, we get

$$\mathbf{g}'(\phi(t_0)) = \lim_{t\to t_0} \frac{\mathbf{g}(\phi(t)) - \mathbf{g}(\phi(t_0))}{\phi(t) - \phi(t_0)}, \quad \mathbf{f}(t_0) = \lim_{t\to t_0} \frac{\mathbf{f}(\phi(t)) - \mathbf{f}(\phi(t_0))}{t - t_0}.$$

By assumption, both of these limits are not equal to 0. Then the algebra of limits tells us that $\lim_{t\to t_0} \frac{\phi(t)-\phi(t_0)}{t-t_0}$ exists. Of course this limits is the derivative of ϕ at t_0. This proves the assertion.

Let us see that the previous lemma shows that the velocity vector changes to a proportional vector when we change a parametrization of a curve. This will show that the tangent space is independent of a parametrizations of a curve.

If we have two parametrizations $\mathbf{f}$ and $\mathbf{g}$, then velocity vector at $x = \mathbf{f}(t_0)$ computed via $\mathbf{f}$ is equal to $\mathbf{f}'(t_0)$. The velocity vector at the same point computed via $\mathbf{g}$ is equal to $\mathbf{g}'(\phi(t_0))$, where $\phi = \mathbf{g}^{-1} \circ \mathbf{f}$. So, we may apply Lemma 2 which tells us that these vectors are proportional.

Each embedded curve is simple. The class of simple curves is very restrictive. It is usually very hard to find an injective parametrization of a simple curve.A possible solution for this problem is to break the curve into the union of simple curves.

Example 7. Let

$$f_i : D \to \mathbb{R}, \quad i = 1, \dots, n-1$$

be $n-1$ twicely differentiable functions on $\mathbb{R}^n$. Assume that the matrix of partial derivatives

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_{n-1}} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_{n-1}}{\partial x_1} & \cdots & \frac{\partial f_{n-1}}{\partial x_{n-1}} \end{pmatrix}$$

is of rank $n-1$ at some point $(a_1, \dots, a_n)$. The the Implicit Function Theorem yields that there exists a number $\epsilon > 0$ such that the set of points

$$C = \{(x_1, \dots, x_n) \in \mathbb{R}^n : f_1(x_1, \dots, x_n) = \ldots = f_{n-1}(x_1, \dots, x_n) = 0, |x_1 - a_1| < \epsilon\}$$

is an embedded curve. If the matrix is of rank $n-1$ at at any point $(a_1, \ldots, a_n) \in \mathbb{R}^n$, then for any interval $I \subset \mathbb{R}$ the set

$$C = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : f_1(x_1, \ldots, x_n) = \ldots = f_{n-1}(x_1, \ldots, x_n) = 0, x_1 \in I$$

is an immersed curve but not necessarily embedded.

Now let $C \subset \mathbb{R}^n$ be an embedded curve. Choose an injective parametrization $\mathbf{f} : I \to C$ such that $\mathbf{f}'(t)$ exists and non-zero at each interior point t in I. Then we can define the tangent line $T(C)_x$ at a point $x = \mathbf{f}(t)$ of C as the line passing through x parallel to the velocity vector $\mathbf{f}'(t)$. Note that the parameter t is determined by the point x uniquely since the map $\mathbf{f}$ is injective. Let us show the tangent line $T(C)_x$ does not depend on the choice of the parametrization.

This fact follows from the following:

Lemma 2. *Let us keep the notation of Lemma 1 and its proof. Assume additionally that* $\mathbf{f}$ *is differentiable at* $t_0 \in (a, b)$*,* $\mathbf{g}$ *is differentiable at* $\phi(t_0)$*, and* $\mathbf{f}'(t_0) \neq 0, \mathbf{g}'(\phi(t_0)) \neq 0$*. Then* ϕ *is differentiable at* t_0 *and*

$$\mathbf{f}'(t_0) = \mathbf{g}'(\phi(t_0))\phi'(t_0).$$

Proof. Of course the last assertion is just the chain rule. In its proof we have to assume that $\phi'(t_0)$ exists. So, we have to prove that the latter exists under our assumptions on $\mathbf{f}$ and $\mathbf{g}$. Since $\mathbf{f}$ and $\mathbf{g}$ are injective, $\phi = \mathbf{g}^{-1} \circ \mathbf{f}$ is injective too. So $\phi(t) - \phi(t_0) \neq 0$ for any $t \in [a, b]$. Write

$$\frac{\mathbf{g}(\phi(t)) - \mathbf{g}(\phi(t_0))}{\phi(t) - \phi(t_0)} \cdot \frac{\phi(t) - \phi(t_0)}{t - t_0} = \frac{\mathbf{f}(\phi(t)) - \mathbf{f}(\phi(t_0))}{t - t_0}.$$

Now we can go to limit when $t \to t_0$. By Lemma 1, ϕ is a continuous function, so $\lim_{t \to t_0} \phi(t) = \phi(t_0)$. Since $\mathbf{f}$ and $\mathbf{g}$ are differentiable at t_0 and $\phi(t_0)$, respectively, we get

$$\mathbf{g}'(\phi(t_0)) = \lim_{t \to t_0} \frac{\mathbf{g}(\phi(t)) - \mathbf{g}(\phi(t_0))}{\phi(t) - \phi(t_0)}, \quad \mathbf{f}(t_0) = \lim_{t \to t_0} \frac{\mathbf{f}(\phi(t)) - \mathbf{f}(\phi(t_0))}{t - t_0}.$$

By assumption, both of these limits are not equal to 0. Then the algebra of limits tells us that $\lim_{t \to t_0} \frac{\phi(t) - \phi(t_0)}{t - t_0}$ exists. Of course this limits is the derivative of ϕ at t_0. This proves the assertion.

Let us see that the previous lemma shows that the velocity vector change to a proportional vector when we change a parametrization of a curve. This will show that the tangent space is independent of a parametrizations of a curve.

If we have two parametrizations $\mathbf{f}$ and $\mathbf{g}$, then velocity vector at $x = \mathbf{f}(t_0)$ computed via $\mathbf{f}$ is equal to $\mathbf{f}'(t_0)$. The velocity vector at the same point computed via $\mathbf{g}$ is equal to $\mathbf{g}'(\phi(t_0))$, where $\phi = \mathbf{g}^{-1} \circ \mathbf{f}$. So, we may apply Lemma 2 which tells us that these vectors are proportional.

Take any points $v_1 < \ldots < v_n$ and set

$$L(C, v_1, \ldots, v_n) = \rho(v_0, v_1) + \ldots + \rho(t_{n-1}, v_n) = ||v_1 - v_0|| + \ldots + ||v_n - v_{n-1}||.$$

Definition. A simple curve C is called *rectifiable* if all possible sums $L(C, v_1, \ldots, v_n)$ are bounded from above. The supremum $L(C)$ of these sums is called the *length* of C.

Clearly, this definition does not depend on the choice of order in C. Because each sum $L(C, v_1, \ldots, v_n) = L(C, v_n, \ldots, v_1)$.

Example 8. Let $\mathbf{f}(t) = v + t(w - v) : [0, 1] \to \mathbb{R}^n$, where $v \neq w$ are two fixed vectors in $\mathbb{R}^n$. Comparing it with Example 3 we see that the image of $\mathbf{f}$ is a "segment" of the line through v in the direction of $w - v$ between the point $v = \mathbf{f}(0)$ and $w = \mathbf{f}(1)$. It is called the straight path between v and w. Obviously the map is injective, so the segment is a simple curve. Let $v = v_0 < v_1, \ldots < v_n = w$ be $n + 1$ ordered points on the line, $v_i = \mathbf{f}(t_i)$ for some partition $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ of $[a, b]$. We have

$$L(\mathbf{f}, v_0, \ldots, v_n) = \sum_{i=1}^{n} ||v_i - v_{i-1}|| = \sum_{i=1}^{n} ||(v + t_i(v - w)) - (v + t_{i-1}(v - w)|| =$$

$$\sum_{i=1}^{n} ||(t_i - t_{i-1})(v - w)|| = \sum_{i=1}^{n}(t_i - t_{i-1})||(v - w)|| = ||v - w|| \sum_{i=1}^{n}(t_i - t_{i-1}) = ||v - w||.$$

So the length of the curve $\mathbf{f}([0, 1])$ exists and is equal to the distance $\rho(v, w) = ||v - w||$ between the two ends of the segment (as it should be!).

Lemma 2. *Let $\mathbf{f} : [a, b] \to \mathbb{R}^n$ be a continuous vector-function. Then*

$$\left|\left| \int_a^b \mathbf{f} dt \right|\right| \leq \int_a^b ||\mathbf{f}(t)|| dt.$$

Proof. Note that on the left we have the length of the integral of the vector function, and on the right we have the integral of the scalar function $t \to ||\mathbf{f}(t)||$. Let $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ be a partition of $[a, b]$ and $S(\mathbf{f}, P, \xi)$ be a Riemann sum of $\mathbf{f}$. By the triangle inequality, we have

$$||S(\mathbf{f}, P, \xi)|| = || \sum_{i=1}^{n} \mathbf{f}(\xi_i)(t_i - t_{i-1})|| \leq \sum_{i=1}^{n} ||\mathbf{f}(\xi_i)||(t_i - t_{i-1}) = S(||\mathbf{f}||, P, \xi).$$

Assume we have the opposite inequality. Choose ϵ less than the difference between the left-hand-side and the right-hand-side. Then for sufficiently "thin" partitions P we have, using the triangle inequality,

$$\left|\left| \int_a^b \mathbf{f} dt \right|\right| \leq ||S(\mathbf{f}, P, \xi)|| + \left|\left| \int_a^b \mathbf{f} dt - S(\mathbf{f}, P, \xi) \right|\right| < \epsilon/2 + ||S(\mathbf{f}, P, \xi)|| \leq \epsilon/2 + S(||\mathbf{f}||, P, \xi) \leq$$

$$\epsilon/2 + \left| \int_a^b ||\mathbf{f}(t)|| dt \right| + \left| S(||\mathbf{f}||, P, \xi) - \int_a^b ||\mathbf{f}(t)|| dt \right| < \epsilon/2 + \epsilon/2 + \int_a^b ||\mathbf{f}(t)|| dt = \epsilon + \int_a^b ||\mathbf{f}(t)|| dt.$$

Here we dropped the absolute value because $||\mathbf{f}(t)|| = (f_1(t)^2 + \ldots + f_n(t)^2)^{1/2}$ is continuous and takes non-negative values. Hence we get

$$\left|\left| \int_a^b \mathbf{f} dt \right|\right| - \int_a^b ||\mathbf{f}(t)|| dt \leq \epsilon$$

contradicting the choice of ϵ.

Lemma 3. *Let C be a simple curve in $\mathbb{R}^n$ and $\mathbf{f} : [a, b] \to \mathbb{R}^n$ be its injective parametrization. Let $c \in (a, b)$,and $C_1 = \mathbf{f}([a, c]), C_2 = \mathbf{f}([c, b])$. Then each C_i is rectifiable and*

$$L(C) = L(C_1) + L(C_2).$$

Proof. The first assertion is obvious since any sum $L(C_i, v_1, \ldots, v_n)$ can be considered as a sum $L(C, v_1, \ldots, v_n)$. Let $v_1 < \ldots < v_n$ be a strictly ordered set of points on C. It is obvious that adding an additional point only increases $L(C, v_1, \ldots, v_n)$ (this follows from the triangle inequality). So we may assume that $f(c) = v_m$ for some $m \leq n$. For any $\epsilon > 0$ we can find a $v_1 < \ldots < v_m$ in C_1 and $v_{m+1} < \ldots < v_n$ in C_2 such that

$$|L(C_1) - L(C_1, v_1, \ldots, v_m)| < \epsilon/2, \quad |L(C_2) - L(C_2, v_m, \ldots, v_n)| < \epsilon/2.$$

Then $L(C, v_1, \ldots, v_n) = L(C_1, v_1, \ldots, v_m) + L(C_2, v_m, \ldots, v_n) \leq L(C_1) + L(C_2)$, and

$$|L(C_1)+L(C_2)-L(v_1, \ldots, v_n)| \leq |L(C_1)-L(v_1, \ldots, v_m)|+|L(C_2)-L(v_m, \ldots, v_n)| < \epsilon/2+\epsilon/2 = \epsilon.$$

This shows that $L(C_1) + L(C_2) = L(C)$.

Theorem 2 (Length of a simple curve). *Let $\mathbf{f} : [a, b] \to \mathbb{R}^n$ be a parametric curve. Assume*
(i) $\mathbf{f}'(t)$ exists and continuous on (a, b);
(ii) $\lim_{t\to a+} \mathbf{f}'(t)$ and $\lim_{t\to b-} \mathbf{f}'(t)$ exist;
(iii) the map $\mathbf{f}$ is injective.

Then the curve $C = \mathbf{f}([a, b])$ is rectifiable and

$$L(C) = \int_a^b ||\mathbf{f}'(t)||dt = \int_a^b (f_1'(t)^2 + \ldots + f_n'(t)^2)^{1/2}dt.$$

Proof. By assumption (iii) the curve C is simple. and comes with an order on it. Let $v_1 < \ldots < v_n$ be some points on C. Since we want to find the supremum of the numbers $L(C, v_1, \ldots, v_n)$ we can always add the points $\mathbf{f}(a) \leq v_1$ and $\mathbf{f}(b) \geq v_n$ to assume that $v_i = \mathbf{f}(t_i)$ where $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ is a partition of $[a, b]$. We have, by Theorem 1 and the previous Lemma,

$$L(\mathbf{f}, v_0, \ldots, v_n) = \sum_{i=1}^n ||\mathbf{f}(t_i)-\mathbf{f}(t_{i-1})|| = \sum_{i=1}^n || \int_{x_{i-1}}^{x_i} \mathbf{f}'(t)dt|| \leq \sum_{i=1}^n || \int_{x_{i-1}}^{x_i} ||\mathbf{f}'(t)||dt = \int_a^b ||\mathbf{f}'(t)||dt.$$

Here we also used the additivity property of integrals of vector-functions. This already shows that all sums $L(\mathbf{f}, v_0, \ldots, v_n)$ are bounded by $\int_a^b ||\mathbf{f}'(t)||dt$ and hence C is rectifiable and

$$L(C) \leq \int_a^b ||\mathbf{f}'(t)||dt.$$

For any $t \in [a, b]$ let $C_t = \mathbf{f}([a, t])$. By Lemma 2, for any $t_0 < t$,

$$L(C_t) - L(C_{t_0}) = L(C_{t_0,t}),$$

where $C_{t_0,t} = \mathbf{f}([t_0, t])$. Using the above estimate of the length, we get

$$L(C, \mathbf{f}(t_0), \mathbf{f}(t)) = ||\mathbf{f}(t) - \mathbf{f}(t_0)|| < L(C_{t_0,t}) \leq \int_{t_0}^t ||\mathbf{f}'(t)||dt.$$

Dividing by $t - t_0$, we obtain

$$\left|\left|\frac{\mathbf{f}(t) - \mathbf{f}(t_0)}{t - t_0}\right|\right| \leq \frac{L(C_t) - L(C_{t_0})}{t - t_0} \leq \frac{1}{t - t_0}\int_{t_0}^{t} ||\mathbf{f}'(t)||dt = \frac{g(t) - g(t_0)}{t - t_0},$$

where $g(t)$ is a primitive function of $||\mathbf{f}'(t)||$. When t tends to t_0 the left expression converges to $||\mathbf{f}'(t_0)||$, and the right expression converges to $g'(t_0)$ which is the same number. Hence the expression in the middle converges to the same number $||\mathbf{f}'(t_0)||$. This shows that the right one-sided limit $\lim_{t\to t_0+} \frac{L(C_t)-L(C_{t_0})}{t-t_0}$ exists and is equal to $||\mathbf{f}'(t_0)||$. Similarly we show that the left one-sided limit exists and is equal to the same number. Hence the function $t \to L(C_t)$ is differentiable and its derivative at t_0 is equal to $||\mathbf{f}'(t_0)||$. Since $L(C_a) = 0$, by Theorem 1,

$$L(C) = L(C_b) = \int_a^b ||\mathbf{f}'(t)||dt.$$

Remark The assumptions of the theorem are essential. Let C be the graph of the function $f : [0,1] \to \mathbb{R}$ defined as follows:

$$f(x) = \begin{cases} -(2k+1)x + 1 & \text{if } x \in [\frac{1}{2k+2}, \frac{1}{2k+1} \\ (2k+1)x - 1 & \text{if } x \in [\frac{1}{2k+1}, \frac{1}{2k}] \\ 0 & \text{if } x = 0. \end{cases}$$

Take $v_1 = 1, v_1 = 1/2, \ldots, v_{2k} = 1/2k$. Then

$$L(C, v_1, \ldots, v_{2k}) \leq f(1/2) + f(1/4) + \ldots + f(1/2k) = 1/2 + 1/4 + \ldots + 1/2k = 1/2(S_k)$$

where S_k is the k-th partial sum of the harmonic series. Since the latter diverges we obtain that the sums $L(C, v_1, \ldots, v_{2k})$ are unbounded. Thus our curve is not rectifiable. Of course the reason why we cannot apply our theorem is that the function $f(x)$ is not differentiable. See another example of not rectifiable curve in Problem 10.

Definition. A curve C is called a *piecewise simple* closed curve if there exists a parametrization $\mathbf{f} : [a,b] \to \mathbb{R}^n$ of C and a partition $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ of $[a,b]$ such that $\mathbf{f}$ is is injective on each $[t_{i-1}, t_i]$ and

$$C = C_1 \cup \ldots \cup C_n$$

where $C_i = \mathbf{f}([t_{i-1}, t_i])$ and $C_i \cap C_{i+1} = \{\mathbf{f}(t_i)\}$. In this case we define the length of C by

$$L(C) = L(C_1) + \ldots + L(C_n).$$

One can show this definition is independent of the parametrization of C and the choice of a partition.

Example 9. Let us compute the length of circle $S^1(v_0, R)$. Since for any $v \in S^1(v_0; R) \iff v - v_0 \in S^1(0; R)$ we can use the parametrization $\mathbf{f}(t) = (R\cos t, R\sin t) + v_0$. Let $C(\phi) = \mathbf{f}([0, \phi])$, where $\phi \in [0, 2\pi)$. Then the length of $C(\phi)$ is equal to

$$L(\phi) = \int_0^\phi (R^2\cos^2 t + R^2\sin^2 t)^{1/2}dt = \int_0^\phi (R^2)^{1/2}dt = \int_0^\phi Rdt = R\phi.$$

We can't take $\phi = 2\pi$ because this invalidate our assumption that the map $\mathbf{f}$ is injective. However we can compute the length of the full circle $C(2\pi)$ by writing $C = C_1 \cup C_2$ where $C_1 = \mathbf{f}([0,\pi])$ (the upper-semicircle) and $C_2 = \mathbf{f}([\pi, 2\pi])$ (the lower-semicircle). We have

$$L(S^1(v_0; R)) = \int_0^{\pi} Rdt + \int_{\pi}^{2\pi} Rdt = 2\pi R$$

as it should be.

Now we see the usual definition of the functions $\cos x$ and $\sin x$. If we take $x \in [0, 2\pi]$ we can identify x with the length of the circle arc of the unit circle $S^1(0; 1)$ which starts at $(1, 0)$ and ends at the point $(\cos x, \sin x)$. Then we set $\cos(x + 2\pi) = \cos x, \sin(x + 2\pi) = \cos x$.

Example 10 (Archimedian spiral). Let us consider the parametric curve $\mathbf{f}(t) = (t\cos t, t\sin t) : [0, 2\pi k] \to \mathbb{R}^n$. Let $C(k)$ be its image. Since $f_1(t)^2 + f_2(t)^2 = t^2$ we can reconstruct uniquely $t \geq 0$ from $\mathbf{f}(t)$. So our map is injective. The remaining assumptions of Theorem 1 are obviously satisfied. Thus we have

$$L(C(k)) = \int_0^{2\pi k} ((\cos t - t\sin t)^2 + (\sin t + t\cos t)^2)^{1/2} dt =$$

$$\int_0^{2\pi k} (\cos^2 t - 2t\cos t\sin t + t^2\sin^2 t + \sin^2 t + 2t\sin t\cos t + t^2\cos^2 t)^{1/2} dt = \int_0^{2\pi k} (1 + t^2)^{1/2} dt =$$

$$\frac{1}{2} x \sqrt[2]{1 + x^2}\Big|_0^{\pi k} + \frac{1}{2}\ln(x + \sqrt[2]{1 + x^2})\Big|_0^{\pi k} = \frac{1}{2} k\pi \sqrt[2]{1 + \pi^2 k^2} + \frac{1}{2}\ln(k\pi + \sqrt[2]{1 + k^2\pi^2}).$$

Example 11 (Ellipse). The ellipse is defined as the set of points $(x, y) \in \mathbb{R}^2$ satisfying

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

for some nonzero fixed real numbers a, b. We can parametrize ellipse using the function $\mathbf{f}(t) = (a\cos t, b\cos t) : [0, 2\pi] \to \mathbb{R}^2$. Obviously

$$L(C) = \int_0^{2\pi} (a^2\sin^2 t + b^2\cos^2 t)^{1/2} dt = \int_0^{2\pi} (a^2 - (a^2 - b^2)\cos^2 t)^{1/2} dt = 4a \int_0^{\pi/2} (1 - \varepsilon\cos^2 t)^{1/2} dt$$

where $\varepsilon = (\frac{a^2 - b^2}{a})^{1/2}$ is called the eccentricity of the ellipse. If we set $x = \cos t$ then we can transform the last integral to

$$\int_0^{\pi/2} (1 - \varepsilon\cos^2 t)^{1/2} dt = \int_0^1 \frac{1 - \varepsilon x^2}{\sqrt[2]{(1 - x^2)(1 - \varepsilon x^2)}} dx.$$

The corresponding indefinite integral is called *elliptic integral of the second kind.* If $\varepsilon \neq 1$, it can not be expressed via elementary functions.

Example 12.

Problems.

1. State and prove a version of the Mean Value Theorem for integrals of vector-functions.

2. Let $\mathbf{f}, \mathbf{g}$ be two differentiable vector-functions on an interval I. Consider the scalar function $\phi(t) = \langle \mathbf{f}(t), \mathbf{g}(t) \rangle$. Prove the formula:

$$\phi(t)' = \langle \mathbf{f}'(t), \mathbf{g}(t) \rangle + \langle \mathbf{f}(t), \mathbf{g}'(t) \rangle.$$

3. Show that among all simple paths from a point v to a point b, the straight path is the shortest.
4. Let $v = (1,2,3,4)$ and $w = (4,3,2,1)$. Let L be the line passing through v with direction $(1,1,1,1)$. Find the unique point u_0 on the line such that for any other point $u \in L$, one has $\rho(w,u) > \rho(w,u_0)$. Show that $w - u_0$ is perpendicular to $(1,1,1,1)$.
5. Let $\mathbf{f} : I \to \mathbb{R}^n$ be a parametric curve. Suppose $\mathbf{f}''(t) = (\mathbf{f}')'$ is defined and is equal to zero for all t. Show that the image of $\mathbf{f}$ is a line.
6. Suppose that $\mathbf{f} : [a,b] \to \mathbb{R}^2$ satisfies the property that $\langle \mathbf{f}'(t_0), \mathbf{f}(t_0) \rangle = 0$ for all $t \in (a,b)$. Show that the image of the map is contained in a circle.
7. Compute the lengths of the following curves given in parametric form:

(a) $\mathbf{f}(t) = (t^{-1}\cos t, t^{-1}\sin t) : [1,c] \to \mathbb{R}^2$. Draw the picture of the curve. What happens with $L(C)$ when c goes to infinity? [Use

$$\int \frac{\sqrt[2]{1+x^2}}{x^2} dx = -\frac{\sqrt[2]{x^2+1}}{x} + \operatorname{arcsinh}(x),$$

where $\operatorname{arcsinh}(x)$ is the inverse function for $\sinh(x) = (e^x + e^{-x})/2$].

(b) Same as in the previous problem but for the curve $\mathbf{f}(t) = (t^{-2}\cos t, t^{-2}\sin t)$.

(c) $\mathbf{f}(t) = (\cos t, \sin t, t) : [0, 2k\pi] \to \mathbb{R}^3$. Draw the picture.

8. Let $\mathbf{f}(\phi) = r(\phi)e^{i\phi} : [a,b] \to \mathbb{R}^2$.

(a) Prove the formula for the length of the curve image:

$$L(C) = \int_a^b (r(\phi)^2 + r'(\phi)^2)^{1/2} d\phi.$$

(b) Compute the length of one loop of the curve $\mathbf{f}([0,\pi/4]))$ where $r(\phi) = 1 + \cos\phi$. Draw the picture of $\mathbf{f}([0,2\pi])$.

9. Prove that the length of the ellipse $x^2/a^2 + y^2/b^2 = 1$ is equal to the length of one wave ($x \in [0,2\pi]$) of the sinusoid $y = c\sin(x/b)$, where $c = \sqrt[2]{a^2 - b^2}$.
10. Show that the graph of the function $f : [0,1] \to \mathbb{R}$ defined by the formula

$$f(x) = \begin{cases} x^2 \sin(1/x^2) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$$

is a simple but not rectifiable curve. Which assumptions of Theorem 2 does it fail?

Math. 396. Notes II: Length of a curve

Let C be a simple finite curve. Recall that an orientation on C defines an order on C. Take any points $v_1 < \ldots < v_n$ and set

$$L(C, v_1, \ldots, v_n) = \rho(v_0, v_1) + \ldots + \rho(t_{n-1}, v_n) = ||v_1 - v_0|| + \ldots + ||v_n - v_{n-1}||.$$

Here $\rho : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is the distance function.

Definition. A simple finite curve C is called *rectifiable* if all possible sums $L(C, v_1, \ldots, v_n)$ are bounded from above. The supremum $L(C)$ of these sums is called the *length* of C.

Since $L(C, v_1, \ldots, v_n) = L(C, v_n, \ldots, v_1)$, this definition does not depend on the choice of orientation of C.

Example 1. Let $\mathbf{f}(t) = v + t(w - v) : [0,1] \to \mathbb{R}^n$, where $v \neq w$ are two fixed vectors in $\mathbb{R}^n$. The image of $\mathbf{f}$ is a "segment" of the line through v in the direction of $w - v$ between the point $v = \mathbf{f}(0)$ and $w = \mathbf{f}(1)$. It is called the *straight path* between v and w. Obviously the map is injective, so the segment is a simple curve. Let $v_1, \ldots < v_k$ be k ordered points on the segment, $v_i = \mathbf{f}(t_i)$ for some partition $t_i \in [0,1]$. We have

$$L(\mathbf{f}, v_1, \ldots, v_k) = \sum_{i=1}^{k-1} ||v_i - v_{i-1}|| = \sum_{i=1}^{k-1} ||(v + t_i(v - w)) - (v + t_{i-1}(v - w)|| =$$

$$\sum_{i=1}^{k-1} ||(t_i - t_{i-1})(v - w)|| = \sum_{i=1}^{n} (t_i - t_{i-1})||(v - w)|| =$$

$$||v - w|| \sum_{i=1}^{k-1} (t_i - t_{i-1}) = (t_k - t_1)||v - w|| \leq ||v - w||.$$

So the length of the curve $\mathbf{f}([0,1])$ exists. If we take $v_1 = v, v_2 = w$ we obtain that $L(C, v_1, v_2) = ||v - w||$. Thus the length is equal to $\rho(v, w) = ||v - w||$.

In the following we shall use the notion of definite integral of a vector function $\mathbf{f} : [a,b] \to \mathbb{R}^n$. Recall its definition. For any partition $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ of $[a,b]$ and $\xi = (\xi_1, \ldots, \xi_n) \in [x_0, x_1] \times \ldots \times [x_{n-1}, x_n]$, we define the Riemann sum $S(\mathbf{f}, P, \xi)$ as

$$S(\mathbf{f}, P, \xi) = \sum_{i=1}^{n-1} \mathbf{f}(\xi_i)(t_i - t_{i-1}) \in \mathbb{R}^n.$$

Definition. The *Riemann integral* of a vector-function $\mathbf{f} : [a,b] \to \mathbb{R}^n$ is a vector

$$\int_a^b \mathbf{f}(t)dt$$

in $\mathbb{R}^n$ such that for any $\epsilon > 0$ there exists $\delta > 0$ such that for any partition P of thickness $\leq \delta$ and any Riemann sum $S(\mathbf{f}, P, \xi)$,

$$|| \int_a^b \mathbf{f}(t)dt - S(\mathbf{f}, P, \xi)|| < \epsilon.$$

If it exists the function $\mathbf{f}$ is called Riemann-integrable over $[a,b]$.

If $\mathbf{f}$ is given in terms of its coordinate functions

$$\mathbf{f}(t) = (f_1(t), \ldots, f_n(t),$$

then $\int_a^b \mathbf{f}(t)dt$ exists if and only if $\int_a^b f_i(t)dt$ exist for each $i = 1, \ldots, n$ and is equal to

$$\int_a^b \mathbf{f}(t)dt = (\int_a^b f_1(t)dt, \ldots, \int_a^b f_n(t)dt).$$

This easily implies

Theorem 1 (Fundamental Theorem of Calculus). *Let* $\mathbf{f} = (f_1(t), \ldots, f_n(t)) : [a, b] \to \mathbb{R}^n$ *where each* $f_i(t)$ *is piecewise continuous. Let* $\mathbf{g}(t) = (g_1(t), \ldots, g_n(t))$ *where each* $g_i(t)$ *is a primitive function for* $f_i(t)$*. Then*

$$\int_a^b \mathbf{f} dt = \mathbf{g}(b) - \mathbf{g}(a).$$

Lemma 1. *Let* $\mathbf{f} : [a, b] \to \mathbb{R}^n$ *be a continuous function. Then*

$$\left|\left| \int_a^b \mathbf{f} dt \right|\right| \leq \int_a^b ||\mathbf{f}(t)|| dt. \qquad (1))$$

Proof. Note that on the left we have the length of the integral of the vector function, and on the right side we have the integral of the scalar function $t \to ||\mathbf{f}(t)||$. Let $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ be a partition of $[a, b]$ and $S(\mathbf{f}, P, \xi)$ be a Riemann sum of $\mathbf{f}$. By the triangle inequality, we have

$$||S(\mathbf{f}, P, \xi)|| = ||\sum_{i=1}^{n} \mathbf{f}(\xi_i)(t_i - t_{i-1})|| \leq \sum_{i=1}^{n} ||\mathbf{f}(\xi_i)||(t_i - t_{i-1}) = S(||\mathbf{f}||, P, \xi).$$

Assume we have the opposite inequality. Choose ϵ less than the difference between the left-hand-side and the right-hand-side. Then for sufficiently "thin" partitions P we have, using the triangle inequality,

$$\left|\left| \int_a^b \mathbf{f} dt \right|\right| \leq ||S(\mathbf{f}, P, \xi)|| + \left|\left| \int_a^b \mathbf{f} dt - S(\mathbf{f}, P, \xi) \right|\right| < \epsilon/2 + ||S(\mathbf{f}, P, \xi)|| \leq \epsilon/2 + S(||\mathbf{f}||, P, \xi) \leq$$

$$\epsilon/2 + \left| \int_a^b ||\mathbf{f}(t)|| dt \right| + \left| S(||\mathbf{f}||, P, \xi) - \int_a^b ||\mathbf{f}(t)|| dt \right| < \epsilon/2 + \epsilon/2 + \int_a^b ||\mathbf{f}(t)|| dt = \epsilon + \int_a^b ||\mathbf{f}(t)|| dt.$$

Here we dropped the absolute value because $||\mathbf{f}(t)|| = (f_1(t)^2 + \ldots + f_n(t)^2)^{1/2}$ is continuous and takes non-negative values. Hence we get

$$\left|\left| \int_a^b \mathbf{f} dt \right|\right| - \int_a^b ||\mathbf{f}(t)|| dt \leq \epsilon$$

contradicting the choice of ϵ.

Lemma 2. *Let* C *be a simple curve in* $\mathbb{R}^n$ *and* $\mathbf{f} : [a, b] \to \mathbb{R}^n$ *be its injective parametrization. Let* $c \in (a, b)$*, and* $C_1 = \mathbf{f}([a, c]), C_2 = \mathbf{f}([c, b])$*. Then* C *is rectifiable if and only if* C_1 *and* C_2 *are rectifiable. Moreover, in this case*

$$L(C) = L(C_1) + L(C_2).$$

Proof. Suppose C is rectifiable. Since any sum $L(C_i, v_1, \ldots, v_k)$ can be considered as a sum $L(C, v_1, \ldots, v_k)$, we obtain $L(C_i, v_1, \ldots, v_k) \leq L(C)$. Thuis each C_i is rectifiable. Let $v_1 < \ldots < v_k$ be any strictly ordered set of points on C. Obviously, there exists m such that $v_i \in C_1, i \leq m, v_j \in C_2, j > m$. Thus

$$L(C, v_1, \ldots, v_k) = L(C_1, v_1, \ldots, v_m, \mathbf{f}(c)) + L(C_2, \mathbf{f}(c), v_{m+1}, \ldots, v_k) \leq L(C_1) + L(C_2).$$

Here, if $\mho(c) = v_m$, we repalce $L(C_1, v_1, \ldots, v_m, \mathbf{f}(c))$ with $L(C_1, v_1, \ldots, v_m)$. This shows that C is rectifiable. It remain to prove the equality for the lengths.

Let us prove that $L(C_1)+L(C_2)$ is the upper bound for the sums $L(C, v_1, \ldots, v_k)$. By definition of upper bound, for any $\epsilon > 0$ we can find a $v_1 < \ldots < v_m$ in C_1 and $v_{m+1} < \ldots < v_n$ in C_2 such that

$$L(C_1) - L(C_1, v_1, \ldots, v_m) < \epsilon/2, \quad L(C_2) - L(C_2, v_{m+1}, \ldots, v_k) < \epsilon/2.$$

Then, using the triangle inequality, we obtain

$$L(C, v_1, \ldots, v_k) = L(C_1, v_1, \ldots, v_m) + L(C_2, v_{m+1}, \ldots, v_n) + ||v_m, v_{m+1}|| \leq$$

$$L(C_1, v_1, \ldots, v_m) + ||v_m, \mathbf{f}(c)|| + ||\mathbf{f}(c), v_{m+1}|| + L(C_2, v_m, \ldots, v_k)$$
$$= L(C_1, v_1, \ldots, v_m, \mathbf{f}(c)) + L(C_2, \mathbf{f}(c), v_m, \ldots, v_k).$$

Since adding a new point only increases the sum $L(C, v_1, \ldots, v_n)$, we get

$$L(C_1) - L(C_1, v_1, \ldots, v_m, \mathbf{f}(c)) \leq L(C_1) - L(C_1, v_1, \ldots, v_m) < \epsilon/2,$$

$$L(C_2) - L(C_2, \mathbf{f}(c), v_{m+1}, \ldots, v_n) \leq L(C_2) - L(C_2, v_{m+1}, \ldots, v_k) < \epsilon/2,$$

and, after adding up,

$$L(C_1) + L(C_2) - L(v_1, \ldots, v_n) = L(C_1) - L(C_1, v_1, \ldots, v_m) + L(C_2) - L(C_2, v_m, \ldots, v_n) <= \epsilon.$$

This shows that $L(C_1) + L(C_2)$ is the upper bound of the sums $L(C, v_1, \ldots, v_k)$, and hence $L(C) = L(C_1) + L(C_2)$.

Theorem 2 (Length of a simple curve). *Let $\mathbf{f} : [a, b] \to \mathbb{R}^n$ be a simple non closed path of class C^1. Then the curve $C = \mathbf{f}([a, b])$ is rectifiable and*

$$L(C) = \int_a^b ||\mathbf{f}'(t)||dt = \int_a^b (f_1'(t)^2 + \ldots + f_n'(t)^2)^{1/2} dt.$$

Proof. Let $v_1 < \ldots < v_k$ be some points on C. Since we want to find the supremum of the numbers $L(C, v_1, \ldots, v_k)$ we can always add the points $\mathbf{f}(a) \leq v_1$ and $\mathbf{f}(b) \geq v_k$ to assume that $v_i = \mathbf{f}(t_i)$ where $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ is a partition of $[a, b]$. We have, by the Fundamental Theorem of Calculus and Lemma 1,

$$L(C, v_0, \ldots, v_n) = \sum_{i=1}^n ||\mathbf{f}(t_i) - \mathbf{f}(t_{i-1})|| = \sum_{i=1}^n || \int_{t_{i-1}}^{t_i} \mathbf{f}'(t)dt|| \leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} ||\mathbf{f}'(t)||dt = \int_a^b ||\mathbf{f}'(t)||dt.$$

Here all the integrals are defined since $||\mathbf{f}'(t)||$ is continuous. We also used the additivity property of integrals of vector-functions. This already shows that all sums $L(\mathbf{f}, v_0, \ldots, v_n)$ are bounded by $\int_a^b ||\mathbf{f}'(t)||dt$ and hence C is rectifiable and

$$L(C) \leq \int_a^b ||\mathbf{f}'(t)||dt.$$

For any $t \in [a, b]$, set $C_t = \mathbf{f}([a, t])$. By Lemma 2, for any $t' < t$,

$$L(C_t) - L(C_{t'}) = L(C_{t',t}),$$

where $C_{t',t} = \mathbf{f}([t', t])$. Using the above estimate of the length, we get

$$L(C, \mathbf{f}(t'), \mathbf{f}(t)) = ||\mathbf{f}(t) - \mathbf{f}(t')|| < L(C_{t',t}) \leq \int_{t'}^{t} ||\mathbf{f}'(x)||dx.$$

Dividing by $t - t'$, we obtain

$$\left|\left|\frac{\mathbf{f}(t) - \mathbf{f}(t')}{t - t'}\right|\right| \leq \frac{L(C_t) - L(C_{t'})}{t - t'} \leq \frac{1}{t - t'} \int_{t'}^{t} ||\mathbf{f}'(x)||dx = \frac{g(t) - g(t')}{t - t'},$$

where $g(t)$ is a primitive function of $||\mathbf{f}'(t)||$. When t tends to t' the first expression converges to $||\mathbf{f}'(t')||$, and the last expression converges to $g'(t')$ which is the same number. Hence the expression in the middle converges to the same number $||\mathbf{f}'(t')||$. This shows that the right one-sided limit $\lim_{t \to t'+} \frac{L(C_t) - L(C_{t'})}{t - t'}$ exists and is equal to $||\mathbf{f}'(t')||$. Similarly we show that the left one-sided limit exists and is equal to the same number. Hence the function $t \to L(C_t)$ is differentiable and its derivative at t' is equal to $||\mathbf{f}'(t')||$. Since $L(C_a) = 0$, by the Fundamental Theorem of Calculus,

$$L(C) = L(C_b) = \int_{a}^{b} ||\mathbf{f}'(t)||dt.$$

Remark. The assumptions of the theorem are essential. Let C be the graph of the function $f : [0, 1] \to \mathbb{R}$ defined as follows:

$$f(x) = \begin{cases} -(2k+1)x + 1 & \text{if } x \in [\frac{1}{2k+2}, \frac{1}{2k+1}], \\ (2k+1)x - 1 & \text{if } x \in [\frac{1}{2k+1}, \frac{1}{2k}], \\ 0 & \text{if } x = 0. \end{cases}$$

Note that $f(\frac{1}{2k}) = \frac{1}{2k}$ and $f(\frac{1}{2k+1}) = 0$ so that $\mathbf{f}$ is continuous. It is obviously simple.

Take $v_i = (1/i, f(1/i)), i = 1, 2, \ldots, n = 2k$. Then, $||v_{2k-1} - v_{2k}|| \geq f(1/2k) = 1/2k$, and hence

$$L(C, v_1, \ldots, v_{2k}) \geq f(1/2) + f(1/4) + \ldots + f(1/2k) = 1/2 + 1/4 + \ldots + 1/2k = 1/2(S_k)$$

where S_k is the k-th partial sum of the harmonic series. Since the latter diverges, we obtain that the sums $L(C, v_1, \ldots, v_{2k})$ are unbounded. Thus our curve is not rectifiable. Of course the reason why we cannot apply our theorem is that the function $f(x)$ is not differentiable.

Let $\mho : [a, b] \to C$ be a piecewise simple path of class C^1. Let $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ be such that the restriction of $\mathbf{f}$ to each interval $[t_{i-1}, t_i]$ is injective. Let $C_i = \mathbf{f}([t_{i-1}, t_i]$. Then we define

$$L(C) = \sum_{i=1}^{k} L(C_i).$$

To show that this definition is independent from a choice of a partition of $[a, b]$, we argue as follows. For any two partitions $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ and $[a = t'_0 < \ldots < t_{m-1} < t_m = b$ we consider the third partition formed by the points t_i, t'_j which we put in order. Then using Lemma 2, we find that both $L(C)$'s computed with respect to the first and the second partition coincide with the $L(C)$ computed with respect to the third partition. This proves the assertionan.

Example 2. Let us compute the length of the circle $S^1(v_0, R)$ with center at v_0 and radius R. Since for any $v \in S^1(v_0; R) \iff v - v_0 \in S^1(0; R)$ we can use the parametrization $\mathbf{f}(t) = (R\cos t, R\sin t) + v_0$. Let $C(\phi) = \mathbf{f}([0, \phi])$, where $\phi \in [0, 2\pi)$. Then the length of $C(\phi)$ is equal to

$$L(\phi) = \int_0^{\phi} (R^2\cos^2 t + R^2\sin^2 t)^{1/2} dt = \int_0^{\phi} (R^2)^{1/2} dt = \int_0^{\phi} R dt = R\phi.$$

We can't take $\phi = 2\pi$ because this invalidate our assumption that the map $\mathbf{f}$ is injective. However we can compute the length of the full circle $C(2\pi)$ by writing $C = C_1 \cup C_2$ where $C_1 = \mathbf{f}([0, \pi])$ (the upper-semicircle) and $C_2 = \mathbf{f}([\pi, 2\pi])$ (the lower-semicircle). We have

$$L(S^1(v_0; R)) = \int_0^{\pi} R dt + \int_{\pi}^{2\pi} R dt = 2\pi R$$

as it should be.

Example 3 (Archimedian spiral). Let us consider the parametric curve $\mathbf{f}(t) = (t\cos t, t\sin t) : [0, 2\pi k] \to \mathbb{R}^n$. Let $C(k)$ be its image. Since $f_1(t)^2 + f_2(t)^2 = t^2$ we can reconstruct uniquely $t \geq 0$ from $\mathbf{f}(t)$. So our map is injective. The remaining assumptions of Theorem 1 are obviously satisfied. Thus we have

$$L(C(k)) = \int_0^{2\pi k} ((\cos t - t\sin t)^2 + (\sin t + t\cos t)^2)^{1/2} dt =$$

$$\int_0^{2\pi k} (\cos^2 t - 2t\cos t\sin t + t^2\sin^2 t + \sin^2 t + 2t\sin t\cos t + t^2\cos^2 t)^{1/2} dt = \int_0^{2\pi k} (1 + t^2)^{1/2} dt =$$

$$\frac{1}{2} x \sqrt[2]{1 + x^2}\Big|_0^{\pi k} + \frac{1}{2}\ln(x + \sqrt[2]{1 + x^2})\Big|_0^{\pi k} = \frac{1}{2} k\pi \sqrt[2]{1 + \pi^2 k^2} + \frac{1}{2}\ln(k\pi + \sqrt[2]{1 + k^2\pi^2}).$$

Example 4 (Ellipse). The ellipse is defined as the set of points $(x, y) \in \mathbb{R}^2$ satisfying

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

for some nonzero fixed real numbers a, b. We assume that $b \leq a$. We can parametrize the ellipse using the function $\mathbf{f}(t) = (a \cos t, b \cos t) : [0, 2\pi] \to \mathbb{R}^2$. This parametrization defines a simple smooth closed curve if we restrict it to any proper closed interval of $[0, 2\pi]$. Let $C_t = \mathbf{f}([0, t]), t < 2\pi$. Obviously

$$L(C_t) = \int_0^t (a^2 \sin^2 \tau + b^2 \cos^2 \tau)^{1/2} d\tau = \int_0^t (a^2 - (a^2 - b^2) \cos^2 \tau)^{1/2} d\tau = a \int_0^t (1 - \varepsilon^2 \cos^2 \tau)^{1/2} d\tau$$

where $\varepsilon = (\frac{a^2-b^2}{a})^{1/2}$ is called the *eccentricity* of the ellipse. If we set $x = \cos \tau$ then we can transform the last integral to

$$\int_0^{\cos^{-1}(x)} \frac{1 - \varepsilon^2 x^2}{\sqrt[2]{(1 - x^2)(1 - \varepsilon^2 x^2)}} dx.$$

The corresponding indefinite integral is called the *elliptic integral of the second kind.* If $\varepsilon \neq 1$, it can not be expressed via elementary functions (like trigonometric, exponential, or polynomial functions) using operations of addition, multiplication, division, composition and inverse. As a function in the upper limit the corresponding improper integral is a new function not expressable via elementary functions). It was first discovered by L. Euler.

We have seen that the velocity vector $\mathbf{f}'(t)$ depends on the parametrization. However, there is a special parametrizaton which is distinguished by the property that $||\mathbf{f}'(t)|| = 1$ for all t.

Theorem 3. *Assume that* $\mathbf{f} : [0, t] \to C$ *is an embedding of class* C^1. *Then there exists a parametrization* $\mathbf{s} : [0, L(C)] \to C$ *such that for any* $t \in [0, L(C)]$,

$$||\mathbf{s}'(t)|| = 1.$$

Moreover, if $\mathbf{g} : [0, c] \to C$ *is any other parametrization of* C *which defines the same orientation of* C *and satisfying the previous property then* $c = L(C)$ *and* $\mathbf{g} = \mathbf{s}$.

Proof. We know from the proof of Theorem 2 that the function $\ell : [a, b] \to [0, L(C)]$ defined by the formula

$$\ell(t) = L(C_t)$$

has derivative equal to $||\mathbf{f}'(t)||$. Since this number is always strictly positive (we use the assumption of the theorem) the function $\ell(t)$ is strictly monotonous increasing. Hence its inverse is a continuous function $\ell^{-1} : [0, L(C)] \to [a, b]$. If we compose it with the parametrization $\mathbf{f}$, we obtain a continuous parametrization $\mathbf{s} = \mathbf{f} \circ \ell^{-1} : [0, L(C)] \to C$ of C. By definition, $\mathbf{s}(l)$ is the unique point x on C such that the length of the curve $C(x) = \{x' \in C : x' \leq x\}$ is equal to l. It is clear that this parametrization does not depend on $\mathbf{f}$.

Since $\ell'(t) \neq 0$, the function $\ell^{-1}(l)$ is differentiable and hence $\mathbf{s}$ is differentiable. By the Chain Rule

$$\mathbf{s}'(l) = (\mathbf{f} \circ \ell^{-1})'(l) = \mathbf{f}'(\ell^{-1}(l))(\ell^{-1})'(l) = \frac{\mathbf{f}'(\ell^{-1}(l))}{||\mathbf{f}'(\ell^{-1}(l)||}. \tag{4}$$

In particular,

$$||\mathbf{s}'(l)|| = ||\frac{\mathbf{f}'(\ell^{-1}(l))}{||\mathbf{f}'(\ell^{-1}(l))||}|| = 1.$$

It remains to show the uniqueness. But this is easy. If $\mathbf{g} : [c, d] \to C$ is another parametrization satisfying the assertion of the theorem, then we know from Lemma 2 in Part 1 that $\mathbf{s} = \mathbf{g} \circ \phi$, where $\phi : [0, L(C)] \to [c, d]$ is a differentiable function, and

$$\mathbf{s}'(l) = \mathbf{g}'(\phi(l))\phi'(l).$$

Taking the norm of both sides, we obtain that $\phi'(l) = 1$ for all l. here we assume that two parametrizations induce the same orientation of the curve. This shows that $\phi(l) = l + c$ for some contant c. Since we requre additionally that $\mathbf{s}(0) = \mathbf{g}(0)$, we obtain $c = 0$ and $\mathbf{s} = \mathbf{g}$.

The parametrization defined in the previous theorem is called the *natural parametrization.* If we use the natural parametrization, the formula for the length of a curve becomes a tautology:

$$L(C) = \int_0^{L(C)} ||\mathbf{s}'(l)||dl = \int_0^{L(C)} dl = L(C).$$

Of course, it is predictable; finding a natural parametrization is equivalent to finding the lenght of each portion of the curve.

Although the first derivative of $\mathbf{s}$ is always of lenghth 1, the second derivative is not necessarily of length 1 (we shall assume that it exists). Its length has a nice interpretation.

Definition. Let $\mathbf{s} : [0, L(C)] \to C$ be the natural parametrization of a curve C. Let $x \in X$ and $l = L(C(x))$. The number

$$\kappa(x) = ||\mathbf{s}''(l)||$$

(if it is define) is called the *curvature* of C at the point x.

Example 5. Let $\mathbf{f}(t) = (R \cos t, R \sin t) : [0, \pi] \to C$ be the usual parametrization of the upper half-circle of radius 1 with center at the origin. We have $||\mathbf{f}'(t)|| = R$, so it is not a natural parametrization. However, if we change the formula to $(R \cos(t/R), R \sin(t/R))$ this will be the natural parametrization defined on $[0, \pi R]$. Clearly, $||\mathbf{s}''(t)|| = 1/R$. Thus at each point of the circle the curvature is constant and equal to $1/R$. On the other hand, a straight line is obviously of curvature 0 (all its parametrizations are linear functions) at any point. When the radius of the circle goes to infinity, it becomes straighter and approaches the straight line.

Theorem 4 (Formula for the curvature). *Let $\mathbf{f} : [a, b] \to \mathbb{R}$ be a simple parametric curve in $\mathbb{R}^3$. Assume that $\mathbf{f}$ is of class C^2 and $\mathbf{f}'(t) \neq 0$ for any t. Then the curvature is defined at any point of $C = \mathbf{f}[a, b]$ different from the end points and is equal to*

$$\kappa(\mathbf{f}(t)) = \frac{||(\mathbf{f}'(t) \cdot \mathbf{f}'(t))\mathbf{f}'(t) - (\mathbf{f}''(t) \cdot \mathbf{f}'(t))\mathbf{f}'(t)||}{||\mathbf{f}'(t)||^4}.$$

Proof. Without loss of generality we may assume that the orientation of C is defined in such a way that $f(a)$ is the origin of the curve. Then the vectors $\mathbf{s}'(l)$ and the vector $\mathbf{g}(t) = \mathbf{f}'(t)/||\mathbf{f}'(t)||$ are both of length 1 and look in the same direction, hence equal. Here $l = \ell(t) = L(C_t)$. Now

$$\frac{d\mathbf{s}'}{dl} = \frac{d\mathbf{g}(t))}{dt} \cdot \frac{dt}{dl} = \frac{1}{||\mathbf{f}'(t)||} \cdot \Big(\frac{\mathbf{f}''(t)}{||\mathbf{f}'(t)||} - \frac{\mathbf{f}'(t)||\mathbf{f}'(t)||'}{||\mathbf{f}'(t)||^2}\Big).$$

To transform it further, notice that for any vector function $\mathbf{h}(t)$ we have

$$||\mathbf{h}(t)||' = \frac{d(\mathbf{h}(t) \cdot \mathbf{h}(t))^{1/2}}{dt} = \frac{1}{2}\frac{d(\mathbf{h}(t) \cdot \mathbf{h}(t))}{dt}(\mathbf{h}(t) \cdot \mathbf{h}(t))^{-1/2} = \frac{\mathbf{h}'(t) \cdot \mathbf{h}(t)}{||\mathbf{h}(t)||}. \tag{5}$$

Using this formula in the above we obtain

$$\frac{d\mathbf{s}'}{dl} = \frac{1}{||\mathbf{f}'(t)||} \cdot \Big(\frac{\mathbf{f}''(t)}{||\mathbf{f}'(t)||} - \frac{\mathbf{f}'(t)(\mathbf{f}''(t)\cdot\mathbf{f}'(t))}{||\mathbf{f}'(t)||^3}\Big)$$

$$= \frac{(\mathbf{f}'(t)\cdot\mathbf{f}'(t))\mathbf{f}'(t) - (\mathbf{f}''(t)\cdot\mathbf{f}'(t))\mathbf{f}'(t)}{||\mathbf{f}'(t)||^4}. \quad (5).$$

This gives us the asserted formula

$$\kappa(\mathbf{f}(t)) = ||\frac{d\mathbf{s}'}{dl}|| = \frac{||(\mathbf{f}'(t)\cdot\mathbf{f}'(t))\mathbf{f}''(t) - (\mathbf{f}''(t)\cdot\mathbf{f}'(t))\mathbf{f}'(t)||}{||\mathbf{f}'(t)||^4}. \quad (6)$$

Math. 396. Notes 3: Calculus of variations

Let $\mathbf{f} : [a,b] \to \mathbb{R}^n$ be a parametric curve of class C^k. It defines the map $(\mathbf{f},\mathbf{f}') : [a,b] \to \mathbb{R}^{2n}$ given by

$$(\mathbf{f},\mathbf{f}')(t) = (f_1(t),\ldots,f_n(t),f_1'(t),\ldots,f_n'(t)).$$

Suppose we have a scalar function $L : \mathbb{R}^{2n} \to \mathbb{R}$ of class C^k. We denote the first n variables by $(x_1,\ldots,x_n)$ and the last n variables by $\xi_1,\ldots,\xi_n$. Composing L with $(\mathbf{f},\mathbf{f}')$ we plug $f_i(t)$ in x_i and $f_i'(t)$ in ξ_i. Let us denote the composition by $L(\mathbf{f},\mathbf{f}')$. This is a function $[a,b] \to \mathbb{R}$. Then we integrate it over $[a,b]$ and get the number

$$S_L(\mathbf{f}) = \int_a^b L(\mathbf{f},\mathbf{f}')dt \quad (1)$$

Thus to each parametric curve $\mathbf{f}$ we associate a number $S_L(\mathbf{f})$. This defines a map from the set of parametric curves defined on $[a,b]$ to the set of real numbers. It is called the *action map*. Of course, it depends on the choice of the function L (called a *Lagrangian*).

Let us consider the set of parametric curves with fixed end points $\mathbf{f}(a) = P, \mathbf{f}(b) = Q$. Let us look for parametric curves which minimize (or maximize) the value $S_L(\mathbf{f})$.

Example 1 Let us take

$$L(x_1,\ldots,x_n,\xi_1,\ldots,\xi_n) = (\xi_1^2 + \ldots + \xi_n^2)^{1/2}.$$

Then

$$S_L(\mathbf{f}) = \int_a^b (f_1'^2 + \ldots + f_n'^2)^{1/2} dt$$

is equal to the length of the image curve $C = \mathbf{f}([a,b])$. So, we want to find $\mathbf{f}$ such that the image curve is a curve of minimal length which joins two points in $\mathbb{R}^n$. We know the answer: the image curve $\mathbf{f}([a,b])$ must be a straight line, so the functions $f_i(t)$ must define a parametrization of a straight line.

Theorem (Euler-Lagrange Equations). *Assume that $S_L(\mathbf{f}) \le S_L(\mathbf{h})$ for any $\mathbf{h} : [a,b] \to \mathbb{R}^n$ with $\mathbf{h}(a) = \mathbf{f}(a), \mathbf{h}(b) = \mathbf{f}(b)$. Suppose L is of class C^2. Then $\mathbf{f}$ satisfies a system of n differential equations:*

$$\frac{\partial L}{\partial x_i}(\mathbf{f},\mathbf{f}') = \frac{d}{dt}\frac{\partial L}{\partial \xi_i}(\mathbf{f},\mathbf{f}'), \quad i = 1,\ldots,n \quad (2)$$

*Proof.*Let $\mathbf{g} = \mathbf{f} - \mathbf{h}$. Then $\mathbf{g}(a) = \mathbf{g}(b) = 0$. Conversely, if $\mathbf{g}(a) = \mathbf{g}(b) = 0$, then $\mathbf{h} = \mathbf{f} + \mathbf{g}$ is another funtion with $\mathbf{h}(a) = \mathbf{f}(a), \mathbf{h}(b) = \mathbf{f}(b)$. Take a number ϵ and consider the function

$$\phi(\epsilon) = \frac{S_L(\mathbf{f} + \epsilon\mathbf{g}) - S_L(\mathbf{f})}{\epsilon}.$$

Here $\mathbf{g}$ is a fixed function with $\mathbf{g}(a) = \mathbf{g}(b) = 0$. Suppose this function has a limit when ϵ goes to 0. Since the numerator is always non-negative (by the assumption on $\mathbf{f}$), we have

$$\lim_{\epsilon \to 0+} \phi(\epsilon) \leq 0, \quad \lim_{\epsilon \to 0-} \phi(\epsilon) \geq 0.$$

This implies that

$$\lim_{\epsilon \to 0} \phi(\epsilon) = 0.$$

Let us try to compute this limit explicitly. By definition of the derivative of L we have for any $(a, b) = (a_1, \ldots, a_n, b_1, \ldots, b_n)$

$$L(x_1, \ldots, x_n, \xi_1, \ldots, \xi_n) = L(a_1, \ldots, a_n, b_1, \ldots, b_n)+$$

$$\sum_{i=1}^{n} [\frac{\partial L}{\partial x_i}(a, b)(x_i - a_i) + \frac{\partial L}{\partial \xi_i}(a, b)(\xi - b_i)] + ||(x, \xi) - (a, b)||\alpha(x, \xi),$$

where $\alpha(x - a, \xi - b)$ goes to zero when $||(x, \xi) - (a, b)||$ goes to zero. Take

$$a_i = f_i(t), \quad b_i = f_i'(t), \quad x_i = f_i(t) + \epsilon g_i(t), \quad \xi_i = f_i'(t) + \epsilon g_i'(t).$$

Then we get

$$\phi(\epsilon) = \frac{1}{\epsilon} \int_a^b [L(\mathbf{f} + \epsilon\mathbf{g}, \mathbf{f}' + \epsilon\mathbf{g}) - L(\mathbf{f}, \mathbf{f}')]dt =$$

$$\int_a^b \Big[\frac{1}{\epsilon} \sum_{i=1}^{n} [\frac{\partial L}{\partial x_i}(\mathbf{f}, \mathbf{f}')\epsilon g_i(t) + \frac{\partial L}{\partial \xi_i}(\mathbf{f}, \mathbf{f}')\epsilon g_i'(t))] + \frac{1}{\epsilon}||\epsilon(\mathbf{g}(t), \mathbf{g}'(t)||\alpha(\epsilon\mathbf{g}, \epsilon\mathbf{g}'(t))\Big]dt =$$

$$\int_a^b \Big[\sum_{i=1}^{n} [\frac{\partial L}{\partial x_i}(\mathbf{f}, \mathbf{f}')g_i(t) + \frac{\partial L}{\partial \xi_i}(\mathbf{f}, \mathbf{f}')g_i'(t))] + ||(\mathbf{g}(t), \mathbf{g}'(t)||\alpha(\epsilon\mathbf{g}, \epsilon\mathbf{g}'(t))\Big]dt.$$

Obviously, when ϵ tends to zero $\alpha(\epsilon\mathbf{g}, \epsilon\mathbf{g}'(t))$ goes to zero. This gives

$$\lim_{\epsilon \to 0} \phi(\epsilon) = \int_a^b \sum_{i=1}^{n} [\frac{\partial L}{\partial x_i}(\mathbf{f}, \mathbf{f}')g_i(t) + \frac{\partial L}{\partial \xi_i}(\mathbf{f}, \mathbf{f}')g_i'(t)]dt. \tag{3}$$

So, if $\mathbf{f}$ minimizes S_L, the expression (3) must be qual to zero for all $\mathbf{g}$. Applying integration by parts we get

$$0 = \int_a^b \sum_{i=1}^{n} [\frac{\partial L}{\partial x_i}(\mathbf{f}, \mathbf{f}')g_i(t) + \frac{\partial L}{\partial \xi_i}(\mathbf{f}, \mathbf{f}')g_i'(t)]dt =$$

$$\int_a^b \sum_{i=1}^{n} \frac{\partial L}{\partial x_i}(\mathbf{f}, \mathbf{f}')g_i(t)dt + \int_a^b \sum_{i=1}^{n} \frac{\partial L}{\partial \xi_i}(\mathbf{f}, \mathbf{f}')g_i'(t))dt =$$

$$\int_a^b \sum_{i=1}^n \frac{\partial L}{\partial x_i}(\mathbf{f},\mathbf{f}')g_i(t)dt + \sum_{i=1}^n \frac{\partial L}{\partial \xi_i}(\mathbf{f},\mathbf{f}')g_i(t)\Big|_a^b - \int_a^b \sum_{i=1}^n \frac{d}{dt}\frac{\partial L}{\partial \xi_i}(\mathbf{f},\mathbf{f}')g_i(t)dt.$$

Since $\mathbf{g}(a) = \mathbf{g}(b) = 0$, the middle term is zero. Therefore we get

$$\int_a^b \sum_{i=1}^n \Big(\frac{\partial L}{\partial x_i}(\mathbf{f},\mathbf{f}') - \frac{d}{dt}\frac{\partial L}{\partial \xi_i}(\mathbf{f},\mathbf{f}')\Big)g_i(t)dt = 0$$

This must be true for all possible functions g_i satisfying $g_i(a) = g_i(b) = 0$. It is not difficult to show that this implies that all the brackets must be zero, i.e.

$$\frac{\partial L}{\partial x_i}(\mathbf{f},\mathbf{f}') - \frac{d}{dt}\frac{\partial L}{\partial \xi_i}(\mathbf{f},\mathbf{f}') = 0, \quad i = 1,\ldots,n.$$

These are our Euler-Lagrange equations.

Example 2. Let us reconsider Example 1. We have

$$\frac{\partial L}{\partial x_i} = 0, \quad i = 1,\ldots,n,$$

$$\frac{\partial L}{\partial \xi_i} = \xi_i\Big(\sum_{i=1}^n \xi_i^2\Big)^{-1/2}, i = 1,\ldots,n.$$

Thus the Euler-Lagrange equations become

$$\frac{d}{dt}(f_i'(t)\Big(\sum_{i=1}^n f_i'(t)^2\Big)^{-1/2}) = 0, \quad i = 1,\ldots,n.$$

At this point let us assume that we are using the natural parametrization of $\mathbf{f}([a,b])$. Then $\Big(\sum_{i=1}^n f_i'(t)^2\Big)^{-1/2}) = 1$ and we obtain $f_i''(t) = 0$ for $i = 1,\ldots,n$. This tells us that each function $f_i(t)$ is linear, hence the curve must be the straight line.

The following result often helps to find a solution of the Euler-Lagrange equations:

Corolary. *Let $\mathbf{f}(t)$ be a solution of the Euler-Lagrange equations. Then*

$$L(\mathbf{f}(t),\mathbf{f}'(t)) - \sum_{i=1}^n f_i'(t)\frac{\partial L}{\partial \xi_i}(\mathbf{f}(t),\mathbf{f}'(t)) \quad \textit{is a constant function.}$$

*Proof.*Let us differentiate the function. We have

$$\frac{d}{dt}\Big(L(\mathbf{f}(t),\mathbf{f}'(t)) - \sum_{i=1}^n f_i'(t)\frac{\partial L}{\partial \xi_i}(\mathbf{f}(t),\mathbf{f}'(t))\Big) =$$

$$\frac{d}{dt}L(\mathbf{f}(t),\mathbf{f}'(t)) - \sum_{i=1}^n \frac{d}{dt}\Big(f_i'(t)\frac{\partial L}{\partial \xi_i}(\mathbf{f}(t),\mathbf{f}'(t))\Big) =$$

$$\sum_{i=1}^{n} \frac{\partial L}{\partial x_i}(\mathbf{f}(t), \mathbf{f}'(t)) f_i'(t) + \sum_{i=1}^{n} \frac{\partial L}{\partial \xi_i}(\mathbf{f}(t), \mathbf{f}'(t)) f_i''(t) -$$

$$- \sum_{i=1}^{n} f_i''(t) \frac{\partial L}{\partial \xi_i}(\mathbf{f}(t), \mathbf{f}'(t)) - \sum_{i=1}^{n} f_i'(t) \frac{d}{dt} \frac{\partial L}{\partial \xi_i}(\mathbf{f}(t), \mathbf{f}'(t)) =$$

$$\sum_{i=1}^{n} \Big(\frac{\partial L}{\partial x_i}(\mathbf{f}(t), \mathbf{f}'(t)) - \frac{d}{dt} \frac{\partial L}{\partial \xi_i}(\mathbf{f}(t), \mathbf{f}'(t)) \Big) f_i'(t) = 0$$

Since $\mathbf{f}$ satisfies the Euler-Lagrange equation, the expression in each bracket must be equal to zero. Thus the derivative of our function is zero, hence it is a constant.

Example 3 Take

$$L(x, \xi) = \frac{1}{2} m \Big(\sum_{i=1}^{n} \xi_i^2 \Big) - U(x)$$

for some function $U(x) = U(x_1, \ldots, x_n)$. We have

$$\frac{\partial L}{\partial x_i} = -\frac{\partial U}{\partial x_i}, \quad \frac{\partial L}{\partial \xi_i} = m\xi_i.$$

Thus the Euler-Lagrange equation gives

$$m f_i''(t) = -\frac{\partial U}{\partial x_i}(f_1(t), \ldots, f_n(t)).$$

This is the Newton Law of motion under the potential force

$$\mathbf{F} = -(\frac{\partial U}{\partial x_1}, \ldots, \frac{\partial U}{\partial x_n}).$$

The function U is called the *potential energy.* By Corollary,

$$L(\mathbf{f}, \mathbf{f}') - m \sum_{i=1}^{n} f_i'(t)^2 = -\Big(\frac{m}{2} \sum_{i=1}^{n} f_i'(t)^2 + U(f_1(t), \ldots, f_n(t)\Big)$$

is constant. The function

$$\frac{m}{2} \sum_{i=1}^{n} f_i'(t)^2 = \frac{m}{2} ||\mathbf{f}'(t)||^2$$

is called the *kinetic energy.* The sum of the kinetic energy and the potential energy is called the *total energy.* Thus Corollary says that the total energy of the particle moving along the parametric curve $\mathbf{f} : [a, b] \to \mathbb{R}^n$ is conserved.

Example 4 Consider the following problem. Let $y = f(x)$ be a function of class C^2 in one variable defined on the interval $[a, b]$. Suppose $f(a) = c, f(b) = d$. Consider the area of the surface obtained by rotating the graph of f about the x-axis. What is the function $y = f(x)$ which gives the minimal area?

The surface area is given by the integral

$$A = \int_a^b y(1 + y'^2)^{1/2} dx.$$

We are dealing with parametric curves $x \to (x, f(x))$. Thus our Lagrangian L is

$$L = x_2(1 + \xi_2^2)^{1/2}.$$

By Corollary,

$$y(1 + y'^2)^{1/2} - y'^2 y(1 + y'^2)^{-1/2} = C$$

for some constant C. Redicing to the common denominator, we get

$$y = C(1 + y'^2)^{1/2}.$$

After squaring we get

$$y'^2 = \frac{y^2}{C^2} - 1.$$

This is easy to integrate. We find

$$\frac{dy}{\frac{y^2}{C^2} - 1} = dx$$

hence

$$x = \int \frac{dy}{\frac{y^2}{c^2} - 1} + C' = C' + \frac{1}{C} \ln(\frac{y}{C} + \sqrt{(y/C)^2 - 1})$$

After some easy calculation we deduce from this that

$$y = \frac{C}{2}[e^{\frac{x-C'}{C}} + e^{-\frac{x-C'}{C}}] = \frac{C}{2} \cosh \frac{x - C'}{C}.$$

It remains to find the constants C and C'. We know that $y(a) = c, y(b) = d$. This gives two equations on C and C'

$$\frac{C}{2} \cosh \frac{a - C'}{C} = c, \quad \frac{C}{2} \cosh \frac{b - C'}{C} = d.$$

It is easy to solve them. So the answer is the *catenary curve* which is the graph of hyperbolic cosine.

Example 5 (Brachistochrone). This probelm gave rise to Calculus of Variation. It was first solved by brothers Johann and Jacob Bernoulli in 1696.

It consists in finding which curve joining two given points A and B, not lying on the same vertical line, has the property that a massive particle sliding down along this curve from A to B reaches B in the shortest possible time.

If we choose the coordinates in the plane $\mathbb{R}^2$ by putting A at the origin and taking $B = (a, b)$ for some $a < 0, b > 0$ we obtain the equation for the motion of the particle in the gravitation field. The force $\mathbf{F}$ acting on the particle is equal to the projection of the gravitation force $F = (0, -g)$ to the velocity vector $(1, y')$. Its length is equal to

$$F = \frac{(0, -g) \cdot (1, y')}{||(1, y')||} = -gy'/\frac{ds}{dx} = -gy'\frac{dy}{ds} = -g\frac{dy}{dt}/\frac{ds}{dt},$$

where s is the natural parameter on the curve.

On the other hand, by the Newton Law we have

$$F = -g\frac{dy}{dt}/\frac{ds}{dt} = \frac{d^2 s}{dt^2}.$$

This gives

$$-g\frac{dy}{dt} = -g\frac{ds}{dt}\frac{d^2s}{dt^2} = -\frac{1}{2}\frac{d}{dt}(\frac{ds}{dt})^2.$$

After integrating, we get

$$v = \frac{ds}{dt} = \sqrt{2gy}.$$

This gives

$$\frac{dt}{ds} = \frac{dt}{dx}\frac{dx}{ds} = \frac{1}{\sqrt{2gy}},$$

$$\frac{dt}{dx} = \frac{ds}{dx}\frac{dt}{ds} = \frac{ds}{dx}\frac{1}{\sqrt{2gy}} = \frac{\sqrt{1+y'^2}}{\sqrt{2gy}}.$$

After integrating, we obtain the expression for the time needed for the particle to slide down to the origin:

$$T = \frac{1}{\sqrt{2g}}\int_a^0 \frac{\sqrt{1+y'^2}dt}{\sqrt{y}}.$$

Now we are in business. We want to find $y = f(x)$ which minimizes T. The Lagrangian here is

$$L = \frac{\sqrt{1+\xi_2^2}}{\sqrt{x_2}}.$$

Our parametric curves are the graphs $(x, y(x))$. By Corollary the expression

$$\frac{\sqrt{1+y'^2}}{\sqrt{y}} - y'\frac{\partial L}{\xi_2}(x,y) =$$

$$\frac{\sqrt{1+y'^2}}{\sqrt{y}} - \frac{y'^2}{\sqrt{y}\sqrt{1+y'^2}} = C.$$

After easy calculations we simplify it to obtain

$$1 + y'^2 = C/y.$$

This gives

$$x = \int \frac{dy}{\sqrt{\frac{C}{y} - 1}}.$$

Using the substitution $y = \frac{C}{1+u^2}$ we arrive at

$$x = c(\frac{u}{1+u^2} + \tan^{-1} u) + C'.$$

Let $t = tan^{-1}u$. Then

$$x = C(\frac{\tan t}{1+\tan^2 t} + t) + C' = \frac{C}{2}(\sin 2t + 2t) + C',$$

$$y = \frac{C}{1+u^2} = C\cos^2 t = \frac{C}{2}(1 + \cos 2t).$$

When we replace t with $\frac{t}{2} - \frac{\pi}{2}$ we get the parametric curve

$$x = \frac{C}{2}(t - \sin t) + C', \quad y = \frac{C}{2}(1 - \cos t).$$

This a familiar parametrization of a *cycloid curve.*

Math. 396. Notes 4: Manifolds

Let S be a subset of $\mathbb{R}^n$. We say that a subset U of S is open in S if $U = U' \cap S$ for some open subset U' of $\mathbb{R}^n$. We say that U is closed in S if $S \setminus U$ is open in S. If $S' \subset \mathbb{R}^m$ and $f : S \to S'$ is a map of sets, we say that f is continuous if the pre-image of any open subset of S' is open in S. This is equivalent to the following. For any $y \in S'$ and any $\epsilon > 0$ and any $x \in S$ with $f(x) = y$ there exists $\delta > 0$ such that $||x' - x|| < \delta$ and $x' \in S$ implies $||f(x') - y|| < \epsilon$.

An open subset of S containing a point $x \in S$ is called an open neighborhood of x.

Example 1. If $S = (a, b)$, then a subset of S is open in S if and only if it is open in $\mathbb{R}$. On the other hand, if $S = [a, b]$ then any interval $[a, c), c \leq b$ is an open neighborhood of a in S but it is not an open subset in $\mathbb{R}$.

A metric space X defines the notion of an open subset. The set of open subsets is a subset $\mathcal{T}$ of the boolean $\mathcal{P}(X)$ (= the set of subsests of X) satisfying teh following properties:

(i) $\emptyset \in \mathcal{T}$;
(ii) $X \in \mathcal{T}$;
(iii) $\cap_{i \in I} U_i \in \mathcal{T}$ if $U_i \in \mathcal{T}$, I is a finite set of indices;
(iv) $\cup_{i \in I} U_i \in \mathcal{T}$ if $U_i \in \mathcal{T}$, I is an arbitrary set of indices.

In general a *topology* on a set X is a subset $\mathcal{T}$ of $\mathcal{P}(X)$ satisfying the properties (i)-(iv) as above. Elements of $\mathcal{T}$ are called *open sets* of the topology. A set with a topology is called a *toplogical space.* Thus any metric space carries a natural structure of a topogical space. If S is a subset of a topological space, then the sets $S \cap U$ where U is an open subset in X, define a topology on S. It is called the *induced topology.* For example, a subset of a metric space carries a natural toplogy, the induced topology of the topology of the metric space.

Note that a subset of a metric space is a metric space, the metric is defined by the restriction of the metric of the ambient space. The induced topology is the topology of the corresponding metric space.

The notion of a continuous map of metric spaces extends to a map of topological spaces. A map $f : X \to Y$ is called continuous if the pre-image of an open subset of Y is an open subset of X. A bijective continuous map is called a *homeomorphism* if its inverse is continuous. In other words, a homeomorphism is a continuous bijective map such that the image of an open subset is an open subset. One can rephrase it in the sense of limits of sequences. A sequence $\{x_n\}$ of points in a topological space X converges to a point x (we write $\lim\{x_n\} = x$) if for any open neighborhood of U x (i.e. an open subset containing x) all members of $\{x_n\}$ except finitely many belong to U. A map $f : X \to Y$ is continuous at a point $x \in X$ if $\lim\{f(x_n)\} = f(x)$ for any sequence $\{x_n\}$ convergent to x. If X is a compact topological space (i.e. any sequence contains a convergent subsequence), then a bijective continuous map is a homeomorphism (prove it!). the following example shows that in general a bijective continuous map is not a homeomorphism.

Example 2 $X = \mathbb{Q}$ equipped with a discrete topology (i.e. each point is an open subset), and $Y = \mathbb{Q}$ equipped with the induced topology of $\mathbb{R}$. the map is the identity map.

Definition. A subset M of $\mathbb{R}^n$ is called a d-dimensional differential manifold of class $C^k, k \geq 1$, if for any point $x \in M$ there exists an open neighborhood V of x in M, an open subset U of $\mathbb{R}^d$ and a bijective map $\alpha : U \to V$ satisfying the following properties:

(a) α is of class C^k considered as the map $\alpha : U \to \mathbb{R}^n$;
(b) $\alpha^{-1} : V \to U$ is continuous (in other words α is a homeomorphism);
(c) $D\alpha(t)$ is of rank d for any $t \in U$.

The map α is called a *coordinate patch* or a *local parametrization* at the point x. Clearly it can serve as a coordinate patch at any other point in V. A manifold of class C^∞ is called a *smooth manifold.*

Example 4. Let $C \subset \mathbb{R}^n$ be an embedded curve. By definition, there exists $\mathbf{f} : [a, b] \to \mathbb{R}^n$ with $\mathbf{f}([a, b]) = C$ such that $\mathbf{f}$ is injective and $\mathbf{f}'(t) \neq 0$ for any $t \in (a, b)$. Let $M = C \setminus \{\mathbf{f}(a), \mathbf{f}(b)\}$. We have proved in Lemma 2 in Notes 1 that for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$||\mathbf{f}(t) - \mathbf{f}(t')|| < \delta \Longrightarrow |t - t'| < \epsilon.$$

Let $\mathbf{f}^{-1} : C \to [a, b]$. We can write $t = \mathbf{f}^{-1}(x), t' = \mathbf{f}^{-1}(x')$ for unique $x, x' \in C$. From this we obtain that

$$||\mathbf{f}(\mathbf{f}^{-1}(x)) - \mathbf{f}(\mathbf{f}^{-1}(x'))|| = ||x - x'|| < \delta \Longrightarrow |t - t'| = |\mathbf{f}^{-1}(x) - \mathbf{f}^{-1}(x')| < \epsilon.$$

Since ϵ was arbitrary, we obtain that the map $\mathbf{f}^{-1} : C \to [a, b]$ is continuous. Thus if we take $\alpha : (a, b) \to M$ to be the restriction of $\mathbf{f}$ to the open interval (a, b) we obtain that M is a 1-dimensional manifold of class equal to the class of the parametrization $\mathbf{f}$.

Example 5. A set M of points is a 0-dimensional manifold if and only if for any point $x \in S$ there exists a number r such that the ball $B_r(x)$ of radius r does not contain any points from S except x. In particular, S must be a countable set (since each such a ball contains a point with rational coordinates).

Example 6. The sphere

$$S = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : ||x||^2 = 1\}.$$

is a smooth $(n-1)$-dimensional manifold. It is easy to give explicit local coordinate parametrization. For simplicity we assume that $p = (x_1, \ldots, x_n) \in S$ has strictly positive coordinates. The set of such points is an open subset V in S. We choose

$$U = \{(\phi_1, \ldots, \phi_{n-1}) : \pi/2 > \phi_i > 0, i = 1, \ldots, n-1.\}$$

and use the local parametrizations α given by the formula:

$$x_1 = \cos\phi_1 \cos\phi_2 \ldots \cos\phi_{n-1}, \quad x_2 = \sin\phi_1 \cos\phi_2 \ldots \cos\phi_{n-1},$$

$$x_3 = \sin\phi_2 \cos\phi_3 \ldots \cos\phi_{n-1}, \ldots, x_n = \sin\phi_{n-1}.$$

I leave it to you to check the properties of the local parametrization.

Definition Let A and B be open subsets in $\mathbb{R}^n$. A map $f : A \to B$ is called a *diffeomorphism of class* C^k if it is of class C^k and it has the inverse $f^{-1} : B \to A$ which is also of class C^k. A diffeomorphism of class C^∞ is called a *smooth diffeomorphism.*

By the Inverse Function theorem and the Chain rule this is equivalent to f be bijective and $\det Df(x) \neq 0$ for all $x \in U$.

Theorem 1. *Let M be a subset of $\mathbb{R}^n$ and $p \in M$. The following properties are equivalent:*

(i) There exists an open neighborhood V of p in $\mathbb{R}^n$, open neighborhood U of $0 \in \mathbb{R}^d$, and a homeomorphism $\alpha : U \to V \cap M$ of class C^k with rank $D\alpha(0) = d, \alpha(0) = p$.

(ii) There exists an open neighborhood V of p in $\mathbb{R}^n$ and a map $f : V \to \mathbb{R}^{n-d}$ of class C^k such that $\{x \in V : f(x) = 0\} = V \cap M$ and $Df(p)$ is of rank $n - d$.

(iii) There exists an open neighborhood V of p in $\mathbb{R}^n$ and a diffeomorphism ϕ of class C^k from V onto an open neighborhood W of 0 in $\mathbb{R}^n$ such that $\phi(p) = 0$ and

$$\phi(V \cap M) = \{y = (y_1, \ldots, y_n) \in W : y_{d+1} = \ldots = y_n = 0\}.$$

Proof.(i) $\Rightarrow$ (ii) Since $D\alpha(0)$ is of maximal rank d, there exists an open neighborhood $U' \subset U$ of 0 such that $D\alpha(t)$ is of maximal rank d for all $t \in U'$. Let $\alpha(t) = (\alpha_1(t), \ldots, \alpha_n(t))$. Without loss of generality we may assume that the Jacobian matrix $(\frac{\partial \alpha_i}{\partial t_j})_{i=1,\ldots,d, j=1,\ldots,d}$ is of rank d at each point $t \in U'$. Consider the map $U' \to \mathbb{R}^d$ given by the formulas $t \to (\alpha_1(t), \ldots, \alpha_d(t))$. By the Inverse Function Theorem, it defines a diffeomorphism from U' onto an open subset V_1 of $\mathbb{R}^d$ containing the first d coordinates $(a_1, \ldots, a_d)$ of the point p. Let

$$t_i = \beta_i(x_1, \ldots, x_d), i = 1, \ldots, d$$

define the inverse map $\beta : V_1 \to U'$. Choose $V_2 \subset \mathbb{R}^{n-d}$ such that $V_1 \times V_2 \subset V$. Then $(V_1 \times V_2) \cap M$ is the set of solutions of the system of equations

$$f_j(x_1, \ldots, x_n) = x_j - \alpha_j(\beta_1(x_1, \ldots, x_d), \ldots, \beta_d(x_1, \ldots, x_d)) = 0, \quad j = d+1, \ldots, n$$

in $V_1 \times V_2$. The last $n - d$ columns of the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial f_{d+1}}{\partial x_1} & \ldots & \frac{\partial f_{d+1}}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \ldots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

form the identity matrix. Thus J is of rank $n - d$ at each point in $V_1 \times V_2 \cap M$. So, if we take V any subset of $V_1 \times V_2$ we obtain (ii).

(ii)$\Rightarrow$ (iii) We can find $n - d$ coordinates in $\mathbb{R}^n$ such that the minor of the Jacobian matrix corresponding to these coordinates is not equal to zero at any point in a neighborhood of p. Without loss of generality we may assume that these coordinates are $x_d, \ldots, x_n$. By the Implicit Function Theorem, we find $n - d$ functions $g_j(x_1, \ldots, x_d)$ such that in an open neighborhood $V' \subset V$ of p the solutions of the system of equations $f(x) = 0$ is equal to the set of solutions of the system

$$x_j - g_j(x_1, \ldots, x_d) = 0, j = d+1, \ldots, n.$$

Consider the map $\phi : V \to \mathbb{R}^n$ given by the formulas

$$\phi(x) = (x_1, \ldots, x_d, x_{d+1} - g_{d+1}(x_1, \ldots, x_d), \ldots, x_n - g_n(x_1, \ldots, x_d)).$$

Its derivative is defined by the block-matrix

$$\begin{pmatrix} I_d & 0_{n-d} \\ Dg & I_{n-d} \end{pmatrix},$$

where $g = (g_{d+1}, \ldots, g_n)$. Obviously its determinant is not equal to 0 at any point $x \in V$. Thus ϕ defines a diffeomorphism of V onto an open subset W of $\mathbb{R}^n$. The image of $M \cap V$ is the set where the last $n-d$ coordinates are equal to zero. The image of p is the origin. This checks (iii).

(iii)$\Rightarrow$ (i) Let

$$U = \{t = (t_1, \ldots, t_d) \in \mathbb{R}^d : (t_1, \ldots, t_d, 0, \ldots, 0) \in W\}.$$

It is clearly an open subset of $\mathbb{R}^d$. The composition of $\phi^{-1} : W \to V$ with the map $U \to W$ defined by sending $(t_1, \ldots, t_d)$ to $(t_1, \ldots, t_d, 0, \ldots, 0)$ is the coordinate patch from property (i).

Corollary 1. *A subset $M \subset \mathbb{R}^n$ is a d-dimensional manifold of class C^k if and only if one of the equivalent properties (i), (ii), (iii) is satisfied for any point x of M.*

*Proof.*First of all, if (i) is satisfied at a point $x \in M$, then, by continuity, $D\alpha(t)$ is of rank d in some open neighborhood $U' \subset U$ of 0. So restricting α to U' we get a coordinate patch at x. Conversely, if $\alpha : U \to V \cap M$ is a coordinate patch and $\alpha(a) = x$, then composing α with the map $t \to t - a$ we may assume that $a = 0$. Then we see that (i) is satisfied.

Let M and N be two subsets of $\mathbb{R}^n$, and let $p \in M, q \in N$. We say that (M,p) is locally diffeomorphic (of class C^k) to (N,q) if there exists an open neighborhood U of x in $\mathbb{R}^n$, an open neighborhood V of q in $\mathbb{R}^n$ and a diffeomorphism $f : U \to V$ of class C^k such that $f(U \cap M) = V \cap N$.

Corollary 2. *A subset $M \subset \mathbb{R}^n$ is a d-dimensional manifold of class C^k if and only if for each point $p \in M$, the pair (M,p) is locally diffeomorphic of class C^k to the pair $(N,0)$, where*

$$N = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_{d+1} = \ldots = x_n = 0\}.$$

Example 5. Let $M = \{(x,y) \in \mathbb{R}^2 : xy = 0\}$ (the "coordinate cross"). Let $N = \{(x,y) \in \mathbb{R}^2 : y^2 = x^2(x+1)\}$ (the "nodal cubic"). I claim that $(M,0)$ is smoothly (i.e of class C^∞) diffeomorphic to $(N,0)$. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by the formula $f(x,y) = (y + x(x+1)^{1/2}, y - x(x+1)^{1/2})$. It is clear that $f(N) \subset M$. However this map is not invertible. Computing its derivative we have

$$\det Df = \det \begin{pmatrix} \frac{x}{2(x+1)^{1/2}} + (x+1)^{1/2} & 1 \\ \frac{-x}{2(x+1)^{1/2}} - (x+1)^{1/2} & 1 \end{pmatrix} = \frac{3x+2}{(x+1)^{1/2}}.$$

We see that f is a smooth map at each point where $x \neq -1$. Also $\det Df \neq 0$ if $x \neq -2/3$. Thus we can find a small neighborhood U of $(0,0)$ such that $\det Df(x,y) \neq 0$ for any $(x,y) \in U$ and f defines a diffeomorphism from U onto an open neighborhood V of 0 (the Inverse Function Theorem). The image $f(M \cap U)$ is contained in $N \cap V$ and the induced map $f : M \cap U \to N \cap V$ is bijective. Thus $(M,0)$ is locally diffeomorphic to $(N,0)$. Of course neither M nor N is a manifold.

Corollary 3. *Let $\alpha_1 : U_1 \to V_1$ and $\alpha_2 : U_2 \to V_2$ be two coordinate patches of a manifold M at a point p. Let $U_1' = \alpha_1^{-1}(V_1 \cap V_2), U_2' = \alpha_2^{-1}(V_1 \cap V_2)$. Then the composition map*

$$\beta = \alpha_1^{-1} \circ \alpha_2 : U_2' \to V_1 \cap V_2 \to U_1'$$

is a diffeomorphism of class C^k.

*Proof.*Obviously the map β is bijective. We have to check that β is of class C^k at each point $t \in U_1 \cap U_2'$ and $\det D\beta(t) \neq 0$. Let $p' = \alpha_1(t)$. In the notation of the proof of Theorem 1 we can find an open neighborhood V of p and a diffeomorphism $\phi : V \to W \subset \mathbb{R}^n$ such that $\phi(V \cap M) = W'$ where $W' = \{(y_1, \ldots, y_n) : y_{d+1} = \ldots = y_n = 0\} \cap W$. By composing ϕ with the projection map to the first d coordnates we get a map $\bar{\phi} : V \to U \subset \mathbb{R}^d$ of class C^k with non-degenerate derivative at each point of V. Replacing V by a smaller open subset we may assume that $V \subset V_1 \cap V_2$. Let $U_1'' = \alpha_1^{-1}(V), U_2'' = \alpha_2^{-1}(V)$. The compositions $\bar{\phi} \circ \alpha_1 : U_1'' \to U, \bar{\phi} \circ \alpha_2 : U_2'' \to U$ are bijective maps of class C^k. By the Chain Rule, their derivatives are non-degenerate at each point. This implies that each map has the inverse of class C^k. Thus $(\bar{\phi} \circ \alpha_1)^{-1} \circ (\bar{\phi} \circ \alpha_2)$ is of class C^k at $t \in U_1''$. But obviously this map coincides with $\beta = \alpha_1^{-1} \circ \alpha_2 : U_1'' \to U_2''$.

An open interval is obviously a 1-dimensional manifold but a closed one is not. Let us now extend the notion of a manifold to be able to consider closed intervals as well.

Let

$$\mathbb{H}^n = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_n \geq 0\}, \quad \mathbb{H}_+^n = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_n > 0\},$$

$$\mathbb{H}_0^n = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_n = 0\},$$

When $n = 1$, $\mathbb{H}^1 = [0, \infty), \mathbb{H}_+^1 = (0, \infty), , \mathbb{H}_0^1 = \{0\}$.

Definition Let $M \subset \mathbb{R}^n$ is called a *d-dimensional manifold of class C^k with a boundary* if there exists a non-empty closed subset ∂M of M such that $M \setminus \partial M$ is a d-dimensional manifold of class C^k, and for any point $p \in \partial M$ there exists a d-dimensional manifold $\tilde{M}$ of class C^k, such that

(i) for some open neighborhood V of p in $\tilde{M}$ the set $(M \setminus \partial M) \cap V$ is an open subset in $\tilde{M}$;

(ii) there exists an open neighborhood U of 0 in $\mathbb{R}^d$ and a coordinate patch $\alpha : U \to V$ of $\tilde{M}$ at p with $\alpha(0) = p$;

(ii) $\alpha(U \cap \mathbb{H}^d) = V \cap M$, $\alpha(U \cap \mathbb{H}_+^d) = V \cap (M \setminus \partial M)$;

(iii) $\alpha(U \cap \mathbb{H}_0) = V \cap \partial M$.

The subset ∂M is called the *boundary* of M, the subset $M \setminus \partial M$ is called the *interior* of M.

Example 6 Let $\mathbf{f} : [a, b] \to C \subset \mathbb{R}^n$ be a simple curve of class C^k. Assume that the one-sided derivatives up to order k of $\mathbf{f}$ at the points of $[a, b]$ exist and continuous (from one side). Also assume that $\mathbf{f}'(t) \neq 0$ for any $t \in [a, b]$. I claim that C is a manifold of class C^k with a boundary. Of course the boundary is the set of two end points $A = \mathbf{f}(a)$ and $B = \mathbf{f}(b)$. To show this let us first extend $\mathbf{f}$ to a function $\tilde{\mathbf{f}}$ of class C^k defined on a larger interval $[a - \epsilon, b + \epsilon]$ such that $\tilde{\mathbf{f}}$ coincides with $\mathbf{f}$ on $[a, b]$ and

$$\tilde{\mathbf{f}}^{(i)}(t) = \lim_{t \to a+} \tilde{\mathbf{f}}(i)(t), \tilde{\mathbf{f}}^{(i)}(t) = \lim_{t \to b-} \tilde{\mathbf{f}}^{(i)}(t) \quad i = 0, \ldots, k. \qquad (*)$$

Let

$$\mathbf{f}^{(i)}(a) = (c_{1i}, \ldots, c_{ni}), \quad \mathbf{f}^{(i)}(b) = (d_{1i}, \ldots, d_{ni}), \quad i = 0, \ldots, k.$$

Set

$$\mathbf{g}(t) = (\sum_{i=0}^{k} \frac{c_{1i}}{i!}(t-a)^i, \ldots, \sum_{i=0}^{k} \frac{c_{ni}}{i!}(t-a)^i),$$

$$\mathbf{h}(t) = (\sum_{i=0}^{k} \frac{d_{1i}}{i!}(t-b)^i, \ldots, \sum_{i=0}^{k} \frac{d_{ni}}{i!}(t-a)^i),$$

Clearly,

$$\mathbf{g}^{(i)}(a) = (c_{1i}, \ldots, c_{ni}), \quad \mathbf{h}^{(i)}(b) = (d_{1i}, \ldots, d_{ni}), \quad i = 0, \ldots, k.$$

Now set

$$\tilde{\mathbf{f}}(t) = \begin{cases} \mathbf{g}(t) & \text{if } t \le a, \\ \mathbf{f}(t) & \text{if } a \le t \le b, \\ \mathbf{h}(t) & \text{if } t \ge b. \end{cases}$$

Obviously $\tilde{\mathbf{f}}$ satisfies $(*)$. Since $\tilde{\mathbf{f}}'(a) = \lim_{t\to a+} \mathbf{f}'(a) \neq 0$, $\quad \tilde{\mathbf{f}}'(b) = \lim_{t\to b-} \mathbf{f}'(a) \neq 0$ and $\tilde{\mathbf{f}}'$ is continuous we can find a small ϵ such that $\tilde{\mathbf{f}}'(t) \neq 0$ in $[a-\epsilon, b+\epsilon]$ and $\tilde{\mathbf{f}}$ is injective in $[a-\epsilon, a+\epsilon]$ and in $[b-\epsilon, b+\epsilon]$. We already know that $C \setminus \{A, B\}$ is a manifold of class C^k. Consider the map $\tilde{\mathbf{f}} : [a-\epsilon, a+\epsilon] \to \mathbb{R}^n$. The image of $[a-\epsilon, a+\epsilon]$ is a simple curve C' and its part $\tilde{\mathbf{f}}([a, a+\epsilon])$ coincides with the part $\mathbf{f}([a, a+\epsilon])$ of our curve C. Let $U = (-\epsilon, \epsilon)$ and let $\theta : \mathbb{R} \to \mathbb{R}$ be the map $t \to t + a$. It defines a diffeomorphism from $\mathbb{H}^1 = [0, \infty)$ to $[a, \infty)$. The map $\alpha = \tilde{\mathbf{f}} \circ \theta$ sends U onto the part $\tilde{M} = \tilde{\mathbf{f}}((a-\epsilon, a+\epsilon))$ of the curve C'. Clearly it is a manifold which contains $\tilde{M} \cap (C \setminus \{A\})$ as an open subset. The map $\alpha : U \to V = \tilde{M}$ is a coordinate patch of the manifold $\tilde{M}$ at the point $A = \mathbf{f}(a) = \tilde{\mathbf{f}}(a)$. It sends $U \cap \mathbb{H}^1 = [0, \epsilon)$ in one-to-one fashion onto $V \cap C$ and sends 0 to the end A of C. Similarly we deal with the other end B of C.

Theorem 2. *Let M be a subset of $\mathbb{R}^n$ and N be its closed non-empty subset. For any point $p \in N$ the following properties are equivalent:*

(i) There exists an open neighborhood V of p in $\mathbb{R}^n$, a manifold $\tilde{M}$ of class C^k containing p, open neighborhood U of $0 \in \mathbb{R}^d$, and a coordinate patch $\alpha : U \to V \cap \tilde{M}$ of $\tilde{M}$ at $\alpha(0) = p$ such that $\alpha(\mathbb{H}^d \cap U) = M \cap V$, $\alpha(\mathbb{H}^d_0) = N$;

(ii) There exists an open neighborhood V of p in $\mathbb{R}^n$ and a map $f = (f_0, f_1, \ldots, f_{n-d}) : V \to \mathbb{R}^{n-d+1}$ of class C^k such that $\{x \in V : f_0(x) \ge 0, f_1 = \ldots = f_{n-d} = 0\} = V \cap M$ with derivative Df of rank $n - d + 1$ at p.

(iii) There exists an open neighborhood V of p in $\mathbb{R}^n$ and a diffeomorphism ϕ of class C^k from V onto an open neighborhood W of 0 in $\mathbb{R}^n$ such that $\phi(p) = 0$ and

$$\phi(V \cap M) = \{y = (y_1, \ldots, y_n) \in W : y_{d+1} = \ldots = y_n = 0, y_d \ge 0\},$$

$$\phi(V \cap N) = \{y = (y_1, \ldots, y_n) \in W : y_{d+1} = \ldots = y_n = 0, y_d = 0\}.$$

Proof.(i) $\Rightarrow$ (ii) As in the proof of Theorem 1 (i) $\Rightarrow$ (ii) we can find an open neighborhood U' of 0 in $\mathbb{R}^d$ such that $\alpha(U') \subset \tilde{M} \cap V$ is given by a system of equations $f_i = x_{d+i} - \psi_i(x_1, \ldots, x_d) = 0, i = 1, \ldots, n-d$. By construction, $\psi_i(x_1, \ldots, x_d) = \alpha_i(t_1, \ldots, t_d)$ after we plug in $t_i = \beta_i(x_1, \ldots, x_d)$ for some functions β_i. The image of the subset $U' \cap \mathbb{H}^d$ is equal to the subset of $\alpha(U')$ where $t_d = \beta_d(x_1, \ldots, x_d) \ge 0$. If we take $f_0 = \beta_d(x_1, \ldots, x_d)$ we obtain that

$$\alpha(U' \cap \mathbb{H}^d) = V \cap M = \{x \in \mathbb{R}^n : f_0(x) \ge 0, f_1(x) = \ldots = f_{n-d}(x) = 0\}.$$

The Jacobian matrix $(\frac{\partial f_i}{\partial x_j})$ of the functions $f_0, \ldots, f_{n-d}$ has $n - d + 1$ rows and n columns. The last $n - d$ columns equal to the unit vectors $(0, 1, 0, \ldots, 0), \ldots (0, \ldots, 0, 0, 1)$. The first row is equal to $(\frac{\partial \beta_d}{\partial x_1}, \ldots, \frac{\partial \beta_d}{\partial x_d}, 0, \ldots, 0)$. Since one of the partials $\frac{\partial \beta_d}{\partial x_i}$ is not equal to zero (why?) we easily see that the rank of the matrix is equal to $n - d + 1$. This checks (ii).

(ii)$\Rightarrow$ (iii) Let

$$N = \{f_0(x) = f_1(x) = \ldots = f_{n-d}(x) = 0\} \cap V,$$

$$\tilde{M} = \{f_1(x) = \ldots = f_{n-d}(x) = 0\} \cap V,$$

By the proof of Theorem 1 (ii)$\Rightarrow$ (iii) (applied to $N \cap V$) we can assume that

$$f_i = x_{d+i} - \phi_i(x_1, \ldots, x_{d-1}), i = 0, \ldots, n-d.$$

Again as in the proof of Theorem 1 we define a diffeomorphism $\phi : V \to W \subset \mathbb{R}^n$ by sending x to $(y_1, \ldots, y_n) = (x_1, \ldots, x_{d-1}, f_0(x), \ldots, f_{n-d}(x))$. We have

$$\phi(V \cap \tilde{M}) = \{(y_1, \ldots, y_n) \in W : y_{d+1} = \ldots = y_n = 0\}$$

$$\phi(V \cap N) = \{(y_1, \ldots, y_n) \in W : y_d = \ldots = y_n = 0\},$$

$$\phi(V \cap M) = \{(y_1, \ldots, y_n) \in W : y_{d+1} = \ldots = y_n = 0, y_d \geq 0\}.$$

This checks (iii).

(iii)$\Rightarrow$ (i) Let

$$U = \{t = (t_1, \ldots, t_d) \in \mathbb{R}^d : (t_1, \ldots, t_d, 0, \ldots, 0) \in W\}.$$

It is clearly an open subset of $\mathbb{R}^d$. The composition of $\phi^{-1} : W \to V$ with the map $U \to W$ defined by sending $(t_1, \ldots, t_d)$ to $(t_1, \ldots, t_d, 0, \ldots, 0)$ is the coordinate patch from property (i).

Corollary 1. *A subset $M \subset \mathbb{R}^n$ is a d-dimensional manifold of class C^k with a boundary N if and only if $M \setminus N$ is a d-dimensional manifold of class C^k, and for any point $p \in N$ one of the equivalent properties (i), (ii), (iii) holds.*

Example 7 Let $M = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1$ be the closed disk with radius R and centered at 0. I claim that it is a smooth manifold with a boundary ∂M equal to the circle of radius 1. Clearly, being an open subset of $\mathbb{R}^2$, $M \setminus \partial M$ is a smooth 2-dimensional manifold. For any point p on the boundary we have to check one of the equivalent properties from Theorem 2. Property (ii) obviously holds since we may take $V = \mathbb{R}^2$ and $f_0 = x_1^2 + x_2^2 - 1$. Then

$$M = \{x \in V : f_0(x) \geq 0\}, \quad N = \{x \in V : f_0(x) = 0\}.$$

Let us check property (i). Assume $p = (x_1, x_2) \in \partial M$ satisfies $x_2 > 0$. Take $U = \{(t_1, t_2) \in (-1, 1) \times (-\pi/2, \pi/2)$, $V = \mathbb{R}^2$ and define $\alpha : U \to V$ by the formula $(t_1, t_2) \to ((1 - t_2)\sin t_1, (1 - t_2)\cos t_1)$. Clearly $\tilde{M} = \alpha(U) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 2, x_2 > 0\}$ is a 2-dimensional smooth manifold. The inverse map $\alpha^{-1} : \tilde{M} \to U$ is given by the formula $(x_1, x_2) \to (\tan^{-1}\frac{x_1}{x_2}, 1 - \sqrt{x_1^2 + x_2^2})$. Also

$$\det D\alpha(t) = \det \begin{pmatrix} -\sin t_1 & (1 - t_2)\cos t_1 \\ -\cos t_1 & (t_2 - 1)\sin t_1 \end{pmatrix} = (1 - t_2)$$

is not equal to 0 at each point in U. Now

$$\alpha(U \cap \mathbb{H}^2) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\} = M,$$

$$\alpha(U \cap \mathbb{H}_0^2) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\} = \partial M.$$

I leave it to you to check property (i) for other points $p \in \partial M$. To check property (iii) from Theorem 2 we take V equal to $\tilde{M}$ from above, $W = V$ and take $\phi : V \to W$ to be the map α from above.

Math. 396. Notes 5: The tangent space

Let M be a manifold of class C^k and let $f : M \to \mathbb{R}$ be a function. If M is an open subset of $\mathbb{R}^n$ we know what does it mean that f belongs to class C^k at a point $p \in M$. We want to extend this notion to arbitrary manifolds. The idea is very simple. Let $\alpha : U \to V \subset M$ be a coordinate patch of M at p with $p = \alpha(a)$. By taking the composition we obtain a function

$$f \circ \alpha : U \to \mathbb{R}.$$

We say that f belongs to class C^k at p if $f \circ \alpha$ is of class C^k at the point a. Of course this definition is legal only if we can check that it does not depend on the choice of the coordinate patch α. But it is easy. Let $\alpha_1 : U_1 \to V_1$ and $\alpha_2 : U_2 \to V_2$ be two coordinate patches of M at $p = \alpha_1(a_1) = \alpha_2(a_2)$. We know that there exists a diffeomorphism $\phi : U_1' \to U_2'$ of class C^k defined on an open subset U_1' of U_1 such that $U_2' \subset U_2$ and $\alpha_1 = \alpha_2 \circ \phi$ on U_1. This implies that for any $x \in U_1$

$$f \circ \alpha_1 = (f \circ \alpha_2) \circ \phi. \tag{1}$$

Since ϕ and ϕ^{-1} are of class C^k, the function $f \circ \alpha_1$ is of class C^k at a_1 if and only if $f \circ \alpha_2$ is of class C^k at $a_2 = \phi(a_1)$. This makes our definition legal. In particlular we can speak about continuous functions on M.

As soon as we understand what is a scalar function on M of class C^k we can define a vector function on M of class C^k. It is a map $f : M \to \mathbb{R}^m$ whose composition with each projection $\mathbb{R}^m \to \mathbb{R}$ is a scalar function of type C^k on M.

Definition. Let $M \subset \mathbb{R}^n$ and $N \subset \mathbb{R}^m$ be two manifolds of class C^k. A map $f : M \to N$ is called a map of class C^k if considered as a map $f : M \to \mathbb{R}^m$ it is a map of class C^k defined on M.

Let us try to define the derivative Df of a map $M \to N$ of class C^k at a point $p \in M$. Let us first consider the case when $N = \mathbb{R}$. So we want to define the derivative of a scalar function of class C^k on M. We can try to define it as the derivative of $f \circ \alpha$ at the point $a = \alpha^{-1}(p)$ for a coordinate patch $\alpha : U \to V$ at p. However the Chain Rule tells us that this definition will depend on α:

$$D(f \circ \alpha_1)(a) = D(f \circ \alpha)(\phi(a)) \circ D\phi(a). \tag{2}$$

To solve this problem we have to look more carefully at the definition of the derivative. Recall that the derivative of any differentiable map $f : U \to \mathbb{R}^m$ defined on an open subset of $\mathbb{R}^n$ at a point $a \in U$ is defined as the $m \times n$ matrix $Df(a)$ such that for any vector $x \in U$

$$f(x) = f(a) + Df(a) \cdot (\vec{x} - \vec{a}) + ||\vec{x} - \vec{a}||\alpha(x), \tag{3}$$

where $\lim_{x \to a} \alpha(x) = 0$. Here we put the arrow over x to indicate that we consider x as a vector in the linear space $\mathbb{R}^n$ so that the operations of addition and scalar multiplication are applied. In fact, from time to time we shall forget to put the arrow identifying a point in the space with the corresponding vector or the corresponding one-row or one-column matrix.

Recall that any $m \times n$-matrix A defines a map

$$T_A : \mathbb{R}^n \to \mathbb{R}^m, \quad \vec{x} \to A \cdot \vec{x}.$$

Here we consider $\vec{x}$ as a 1-column matrix and use the matrix multiplication. This map is *linear*, i.e. satisfies

$$T_A(\vec{x} + \vec{x}') = T_A(\vec{x}) + T_A(\vec{x}'),$$

$$T_A(\lambda \cdot \vec{x}) = \lambda \cdot T_A(\vec{x}), \quad \forall \lambda \in \mathbb{R}, \quad \forall \vec{x}, \vec{x}' \in \mathbb{R}^n.$$

Conversely any linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is equal to T_A for a unique matrix A. We write each $\vec{x} = (a_1, \ldots, a_n)$ as $\vec{x} = a_1\vec{e}_1 + \ldots + a_n\vec{e}_n$ where $\vec{e}_i = (0, \ldots, 1, 0, \ldots, 0)$ is the i-th unit vector (with 1 at the i-th spot). Then using the definition of a linear map we get

$$T(a_1\vec{e}_1 + \ldots + a_n\vec{e}_n) = a_1T(\vec{e}_1) + \ldots + a_nT(\vec{e}_n) = A \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix},$$

where

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1j} & \ldots & a_{1n} \\ \vdots & \ldots & \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{ij} & \ldots & a_{in} \\ \vdots & \ldots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mj} & \ldots & a_{mn} \end{pmatrix}$$

is the matrix with j-th column A_j equal to the vector

$$T(\vec{e}_j) = (a_{1j}, \ldots, a_{nj}).$$

Recall that we always consider vectors in $\mathbb{R}^n$ as arrows originated at 0. We can also consider vectors as arrows originated from any point $a \in \mathbb{R}^n$. Such an arrow is just an ordered pair of points (a, b), where a is its origin and b is its end-point. If we consider a and b as vectors $\vec{a}$ and $\vec{b}$ coming from the origin, then $\vec{b} = \vec{a} + (\vec{a} - \vec{b})$, so that the end-point b is determined by the vector $\vec{v} = \vec{b} - \vec{a}$. So a vector originated from a is determined by its origin point a and a vector $\vec{v}$ with origin at 0 such that the end-point is equal to the end-point of the vector sum $\vec{a} + \vec{b}$. This allows us to consider the set of vectors originated at a as the set of pairs $(a, \vec{v})$, where a is a point of $\mathbb{R}^n$, and $\vec{v}$ is a vector originated at 0, i.e. an element of the linear space $\mathbb{R}^n$.

We shall denote the set of vectors originated from a point $a \in \mathbb{R}^n$ by $T_a\mathbb{R}^n$ and call it the *tangent space of* $\mathbb{R}^n$ at the point a. If U is any open subset of $\mathbb{R}^n$ (or $\mathbb{H}^d$) and $a \in U$, we set by definition

$$T_a(U) = T_a(\mathbb{R}^n)$$

and call it the *tangent space of* U at the point a. Since U is a manifold of dimension n we would like to extend the definition of the tangent space to any manifold M.

Let U be an open subset of $\mathbb{R}^d$ (or $\mathbb{H}^n$) and $\alpha : U \to \mathbb{R}^n$ be a differentiable map defined on U. For any $(a, v = (a_1, \ldots, a_n)) \in T_a(U)$ set

$$\alpha_*(a, \vec{v}) = (\alpha(a), D\alpha(a) \cdot \vec{v}) \in T_{\alpha(a)}\mathbb{R}^n.$$

So, the derivative of α allows us to define the map

$$\alpha_{*,a} : T_a(U) \to T_{\alpha(a)}(\mathbb{R}^n), \quad (a, v = (a_1, \ldots, a_n)) \to \alpha_*(a, \vec{v}).$$

Suppose now that α is a coordinate patch of a d-dimensional manifold at a point $p = \alpha(a)$.

Definition The *tangent space* of M at a point $p \in M$ is the image $T_p(M) = \alpha_{*,a}(T_a(U))$ of the map $\alpha_{*,a}$ where $\alpha : U \to M$ is a coordinate patch of M at the point $\alpha(a)$.

In other words, $T_p(M)$ consists of all vectors in $T_p(\mathbb{R}^n)$ which can be written in the form $(p, D\alpha(a)\cdot v = (a_1,\ldots,a_n))$, where $v = (a_1,\ldots,a_n)$ runs through the set $\mathbb{R}^d$.

The set of vectors $D\alpha(a)\cdot v = (a_1,\ldots,a_n), v = (a_1,\ldots,a_n) \in \mathbb{R}^d$, is a linear subspace $D\alpha(a)(\mathbb{R}^d)$ of $\mathbb{R}^n$ equal to the image of $\mathbb{R}^d$ under the linear map $\mathbb{R}^d \to \mathbb{R}^n$ defined by the matrix $D\alpha(a)$. Recall that a linear subspace L in $\mathbb{R}^n$ is a subset L of $\mathbb{R}^n$ which is closed under operations of addition and scalar multiplication of vectors. By definition, $T_p(M)$ consists of vectors in $\mathbb{R}^n$ which originate at p and end at $\vec{p}+\vec{w}$, where $\vec{w}\in D\alpha(a)(\mathbb{R}^d)$. If we identify $T_p(M)$ with the set of the end-points $\vec{p}+\vec{w}$ of vectors $(p,\vec{w})\in T_p(M)$ then we obtain

$$T_p(M) = \vec{p} + D\alpha(a)(\mathbb{R}^d).$$

The set of points in $\mathbb{R}^n$ obtained from a linear subspace L by adding to each vector from L a fixed vector a from $\mathbb{R}^n$ is called an *affine subspace*. For example, when L is a one-dimensional subspace spanned by a nonzero vecror $\vec{v}$, we have

$$\vec{a} + L = \{\vec{a} + t\cdot\vec{v} : t\in\mathbb{R}\}.$$

This is the line through the point a parallel to the vector $\vec{v}$.

Let $T:\mathbb{R}^d\to\mathbb{R}^n$ be a linear map defined by a matrix A. To find the image $T(\mathbb{R}^d)$ we do as follows. Write any $\vec{v}\in\mathbb{R}^d$ in the form $\vec{v} = a_1\vec{e}_1+\ldots+a_n\vec{e}_n$. Then, as we have already seen before,

$$T\vec{v} = a_1A_1 + \ldots + a_nA_n,$$

where A_j is the j-th column of A considered as a vector in $\mathbb{R}^n$. Thus $T(\mathbb{R}^d)$ consists of all possible linear combinations of columns of the matrix A.

In our case

$$D\alpha(a) = \begin{pmatrix} \frac{\partial\alpha_1}{\partial t_1} & \ldots & \frac{\partial\alpha_1}{\partial t_j} & \ldots & \frac{\partial\alpha_1}{\partial t_d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial\alpha_n}{\partial t_1} & \ldots & \frac{\partial\alpha_n}{\partial t_j} & \ldots & \frac{\partial\alpha_n}{\partial t_d} \end{pmatrix}(a).$$

Thus

$$D\alpha(a)(\mathbb{R}^d) = \{\lambda_1 \begin{pmatrix} \frac{\partial\alpha_1}{\partial t_1}(a) \\ \vdots \\ \frac{\partial\alpha_1}{\partial t_d}(a)\end{pmatrix} + \ldots + \lambda_n \begin{pmatrix} \frac{\partial\alpha_n}{\partial t_1}(a) \\ \vdots \\ \frac{\partial\alpha_n}{\partial t_d}(a)\end{pmatrix} : \lambda_1,\ldots,\lambda_n\in\mathbb{R}\}.$$

Let us restrict our map α to the intersection of U with the line ℓ_j which passes through $a = (a_1,\ldots,a_n)$ parallel to the vector $\vec{e}_j$. Its parametric equation is $\vec{x} = \vec{a} + t\vec{e}_j$. The restriction of α to this line is the vector function in one variable

$$\mathbf{f}_j(t) = \alpha(a_1,\ldots,a_{j-1},a_j+t,a_{j+1},\ldots,a_n) =$$

$$(\alpha_1(a_1,\ldots,a_{j-1},a_j+t,a_{j+1},\ldots,a_n),\ldots,\alpha_n(a_1,\ldots,a_{j-1},a_j+t,a_{j+1},\ldots,a_n).$$

Its image is the curve $\alpha(\ell_j\cap U)$ on M passing through p with the velocity vector

$$\mathbf{f}_j'(0) = (\frac{\partial\alpha_1}{\partial t_j}(a),\ldots,\frac{\partial\alpha_n}{\partial t_j}(a)).$$

But this is the j-th column of the matrix $D\alpha(a)$. Comparing with the above we find that $T_p(M)$ consists of vectors $(p,\vec{w})$, where $\vec{w}$ is a linear combination of the velocity vectors of d curves $\alpha(\ell_j\cap U)$ at the point p.

For example, if $d = 1$, we obtain the definition of the tangent line of an embedded curve.

Although the definition of $T_p(M)$ seems to be dependent on the choice of the coordinate patch, in fact it does not. The easiest way to see it is to use the formula (2). Since $D\beta(a_1)$ is a one-to-one map, we have $D\beta(a_1)(\mathbb{R}^d) = \mathbb{R}^d$, hence

$$D\alpha_1(a_1)(\mathbb{R}^d) = D\alpha_2(a_2)(D\beta(a_1)(\mathbb{R}^d)) = D\alpha_2(a_2)(\mathbb{R}^d).$$

In fact we have the following equivalent definition of the tangent space in terms of equations defining a manifold.

Theorem 1. *Let V be an open neighborhood of $p = (p_1, \ldots, p_n)$ in $\mathbb{R}^n$ such that $M \cap V = \{x \in \mathbb{R}^n : f_1(x) = \ldots = f_n(x) = 0\}$ for some functions $f_1, \ldots, f_n : V \to \mathbb{R}^n$ of class C^k with $Df(p)$ of maximal rank $n - d$. Then $T_pM = p + L$ where L is the linear subspace of $\mathbb{R}^n$ equal to the set of vectors $\vec{v} = (v_1, \ldots, v_n)$ whose coordinates satisfy the homogeneous system of linear equations:*

$$\frac{\partial f_i}{\partial x_1}(p)v_1 + \ldots + \frac{\partial f_i}{\partial x_n}(p)v_n = 0, \quad i = 1, \ldots, n - d. \tag{4}$$

In particular, the set of end-points of vectors from $T_p(M)$ is equal to the set of solutions $x = (x_1, \ldots, x_n)$ of the inhomogeneous system of linear equations

$$\frac{\partial f_i}{\partial x_1}(p)(x_1 - p_1) + \ldots + \frac{\partial f_i}{\partial x_n}(p)(x_n - p_n) = 0, \quad i = 1, \ldots, n - d. \tag{5}$$

*Proof.*Let $\alpha : U \to M$ be a coordinate patch at $p = \alpha(a)$. Let us first see that

$$D\alpha(a)(\mathbb{R}^d) \subset L. \tag{6}$$

For any $t = (t_1, \ldots, t_d) \in U$ we have

$$f_i(\alpha_1(t), \ldots, \alpha_n(t)) \equiv 0, \quad i = 1, \ldots, n - d.$$

Taking the partial derivative in the variable t_j for each $j = 1, \ldots, d$ we get

$$\sum_{k=1}^{n} \frac{\partial f_i}{\partial x_k}(p)\frac{\partial \alpha_k}{\partial t_j}(a) = 0, \quad i = 1, \ldots, n - d, j = 1, \ldots, d.$$

This shows that the vectors $A_j = (\frac{\partial \alpha_1}{\partial t_j}(a), \ldots, \frac{\partial \alpha_n}{\partial t_j}(a))$ satisfy the system of equations (4). Since each vector from $D\alpha(a)(\mathbb{R}^d)$ is a linear combinations of the vectors $A_1, \ldots, A_d$ we obtain (6). Now let us compare the dimensions of the linear subspaces $D\alpha(a)(\mathbb{R}^d)$ and L. The dimension of the first space is equal to the number of linearly independent vectors among the vectors $A_1, \ldots, A_d$. But they are exactly the columns of the matrix $D\alpha(a)$. Since the rank of this matrix is equal to d, the dimension is d. On the other hand the dimension of L is equal to $n - k$, where k is equal to the rank of the matrix of the coefficients of the system. But by the assumption, the rank of this matrix is $n - d$. Thus L is of dimension $n - (n - d) = d$. Since both $D\alpha(a)(\mathbb{R}^d)$ and L have the same dimension and one contains another, they must coincide. This proves the theorem.

Example 1. Let

$$M = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}|| = R\}.$$

be the sphere in $\mathbb{R}^n$ of radius R centered at the origin. Then the equation $f = \sum_{i=1}^n x_i^2 - R^2 = 0$ defines M at each point $p = (p_1, p_2, p_3) \in M$. Since $f'(\mathbf{x}) = 2\mathbf{x} \neq 0$ for any $\mathbf{x} \in M$ (because $0 \notin M$) we can use f to apply Theorem 1. We obtain that $T_p(M)$ is defined by the equation

$$\sum_{i=1}^n p_i(x_i - p_i) = 0 \tag{7}$$

or equivalently,

$$\sum_{i=1}^n p_i x_i = R^2.$$

The equation (7) shows that $T_p(M)$ consists of vectors originated at p and perpendicular to the radius vector $\vec{p}$. This is analogous to the case of a circle where the tangent line is perpendicular to the radius vector.

Now it is clear how to define the *derivative of a differentiable map* $f : M \to \mathbb{R}^m$ defined on a manifold $M \subset \mathbb{R}^n$. Recal that in the case when $M = U$ is an open subset of $\mathbb{R}^n$, the derivative of f at the point $p \in U$ defines a map $f_{*,a} : T_pU \to T_{f(p)}\mathbb{R}^m$. We define the derivative df_p of f at p as the map:

$$df_p : T_p(M) \to T_{f(p)}\mathbb{R}^m$$

satisfying the following property: for any coordinate patch $\alpha : U \to M$ at p, we have

$$df_p \circ \alpha_{*,a} = (f \circ \alpha)_{*,a}.$$

This property tells us that the image $df_p((p, \vec{w}))$ of a tangent vector $(p, \vec{w}) \in T_p(M)$ is obtained as follows. We write it as the image $\alpha_*(a, \vec{v})$ of the tangent vector $(a, \vec{v}) \in T_aU$ and then set

$$df_p(p, \vec{w}) = (f \circ \alpha)_*(a, \vec{v}).$$

If $\gamma : U' \to M$ is another coordinate patch of M at p, then we may assume that $\alpha = \gamma \circ \beta$, where $\beta : U \to U'$ is a diffeomorphism. Then $f \circ \alpha = f \circ (\gamma \circ \beta) = (f \circ \gamma) \circ \beta$, and

$$(f \circ \alpha)_{*,a} = (f \circ \gamma)_{*,\beta(a)} \circ \beta_{*,a}.$$

Since $(p, \vec{w}) = \alpha_*(p, \vec{v}) = \gamma_*(\beta(a), \beta_*(a, \vec{v}))$, we see that the definition of $df_p(p, \vec{w})$ is independent of the choice of the coordinate patch.

If the image of f is contained in a manifold $N \subset \mathbb{R}^m$, then the image of df_p is contained in $T_{f(p)}N$. This can be seen by using the fact (proven in homework) that tangent vectors at a point $f(p)$ are the velocity vectors of curves on N passing through $f(p)$. The image of a curve on M with the velocity vector $(p, \mathbf{v})$ is a curve on N with the velocity vector $df_p(p, \mathbf{v})$. So we have dfefined a map $df_p : T_pM \to T_{f(p)}N$.

Note that T_pM is a linear space with respect to the operations

$$(p, \mathbf{v}) + (p, \mathbf{v}') = (p, \mathbf{v} + \mathbf{v}'), \quad \lambda \cdot (p, \mathbf{v}) = (p, \lambda\mathbf{v}).$$

It is a linear subspace of $T_p\mathbb{R}^n$. The latter can be identified with $\mathbb{R}^n$. It is easy to see that the derivative $df_p : T_pM \to T_{f(p)}N$ is a linear map.

Remark. Note that we define df_p by taking an open neighborhood of p and this definition does not depend on which neighborhood one takes. For example, let us assume that the restriction of the map f to an open neighborhood V of p in M extends to a differentiable map $F : W \to \mathbb{R}^m$, where W is an open neighborhood of p in $\mathbb{R}^n$ such that $W \cap M = V$. Here "extends" means that the restriction of F to V coincides with f. Then the restriction of $dF_p : T_p(\mathbb{R}^n) \to \mathbb{R}^m$ to $T_p(M) \subset T_p(\mathbb{R}^n)$ coincides with df_p. This easily follows from the chain rule for the map $F \circ \alpha$, where $\alpha : U \to V$ is a local patch at p.

Finally let us discuss the *tangent bundle* of a manifold M. By definition this is a subset of $M \to \mathbb{R}^n$ which consists of pairs $(p, \mathbf{v})$, where $(p, \mathbf{v}) \in T_pM$. For example, $T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$. Since $M \subset \mathbb{R}^n$ we have

$$TM \subset \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}.$$

I claim that TM is a manifold. In fact, let $(p, \mathbf{v}) \in TM$. Let U be an open neighborhhod of p in $\mathbb{R}^n$ such that $M \cap U = f^{-1}(0)$, where $f : U \to \mathbb{R}^{n-d}$ is a map of class C^k with $\text{rank}(Df(p)) = n - d$. Let $U' = U \times \mathbb{R}^n$. This is an open neighborhhod of $(p, \mathbf{v})$ in $\mathbb{R}^{2n}$. It is easy to see that $TM \cap U'$ can be given by the equations

$$f(\mathbf{x}) = 0, \quad Df(p) \cdot \mathbf{y} = 0,$$

where we use the coordinates x_i in the first factor $\mathbb{R}^n$ and the coordinates y_i in the second factor $\mathbb{R}^n$ of $\mathbb{R}^{2n}$. Let $F : \mathbb{R}^{2n} \to \mathbb{R}^{n-d} \times \mathbb{R}^{n-d} = \mathbb{R}^{2n-2d}$ be the map defined by $F(\mathbf{x}, \mathbf{y}) = (f(\mathbf{x}), Df(\mathbf{x}) \cdot \mathbf{y})$. Clearly $M \cap U' = F^{-1}(0)$. One can easily computes the derivative of F:

$$DF(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} A & Df(\mathbf{x}) \\ Df(\mathbf{x}) & 0 \end{pmatrix}.$$

Here we represent a matrix of size $2n - 2d \times 2n$ as a block matrix with A of size $(n-d)\times$ and 0 is the zero matrix of size $(n-d) \times n$. Since $\text{rank}(Df(p)) = n - d$, we easily see that $\text{rank}(DF(p, \mathbf{v})) = 2n - 2d$. This shows that TM is a manifold of dimension $2d$ near the point $(p, \mathbf{v})$. Since $(p, \mathbf{v})$ was an arbitrary point of TM, TM is a manifold.

Note that we have a natural map $\pi : TM \to M$ which sends $(p, \mathbf{v})$ to p. Its fibre over p (i.e. the pre-image of the set $\{p\}$) is equal to $T_p(M)$. The map p is differentiable since it is given by the restriction of the smooth map $T\mathbb{R}^n \to \mathbb{R}^n$ to TM. The map π is called the *canonical projection* of the tangent bundle to the base M.

Example 2. Let M be the circle of radius 1 with center at the origin. Then M is given by the following two equations in $\mathbb{R}^4$:

$$x_1^2 + x_2^2 - 1 = 0, \quad x_1y_1 + x_2y_2 - 1 = 0.$$

For any $t \in \mathbb{R}$ let $A_t = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$. It represents the rotation transformation of $\mathbb{R}^2$ about the angle t. Let $p = (\cos t, \sin t) \in M$. Then T_pM consists of pairs $(p, \mathbf{v})$, where $\mathbf{v} = \lambda(-\sin t, \cos t), \lambda \in \mathbb{R}$. Clearly $A_{-t} \cdot \mathbf{v} = (0, \lambda)$. Consider the map $\phi : TM \to M \times \mathbb{R}$ defined by $\phi(p, \mathbf{v}) = (p, \lambda)$. It is easy to see that this is a diffeomorphism from TM to $M \times \mathbb{R}$. So, we can view the tangent bundle of a circle as an infinite cylinder.

Math. 396. Notes 6: Vector fields.

Let G be an open subset of $\mathbb{R}^n$, and I an interval in $\mathbb{R}$. Consider a continuous function and $\mathbf{F} : (a, b) \times G \to \mathbb{R}^n$. We are looking for differentiable functions $\mathbf{f} : I \to G$ such that for any $t \in I$:

$$\mathbf{f}'(t) = \mathbf{F}(t, \mathbf{f}(t)). \tag{1}$$

If $\mathbf{F}(t,\mathbf{x}) = (F_1(t,\mathbf{x}),\ldots,F_n(t,\mathbf{x}))$ and $\mathbf{f}(t) = (f_1(t),\ldots,f_n(t))$, then (1) is equivalent to the system of equalities of scalar functions defined on I:

$$\frac{df_i(t)}{dt} = F_i(t, f_1(t),\ldots,f_n(t)), \quad i = 1,\ldots n. \tag{2}$$

We say that $\mathbf{f}$ is a *solution of the system of differential equations of the first order* defined by the function F.

Since F is continuous, it is clear that each solution belongs to class C^1.

An additionla condition on a solution defined by

$$\mathbf{f}(t_0) = \mathbf{x}_0 \tag{3}$$

for some $t_0 \in I, \mathbf{x} \in G$ is called an *initial condition*. We shall show now that under certain condition on $\mathbf{F}$, a solution satisfying an initial condition exists and is unique.

Beforfe we begin the proof, we make the following definition:

Definition. We say that $\mathbf{F} : I \times G \to \mathbb{R}^n$ satisfies the Lipschitz condition if for any $t_0 \in I$ and any $\mathbf{x}_0 \in G$ there exists an interval $(-a + x_0, x_0 + a)$ contained in I, an open neighborhood U of $\mathbf{x}_0$ in G and a number $r \geq 0$ such that

$$||\mathbf{F}(t,\mathbf{x}) - \mathbf{F}(t,\mathbf{y})|| \leq r||\mathbf{x} - \mathbf{y}||.$$

for any $t \in (-a + x_0, x_0 + a)$ and any $\mathbf{x}, \mathbf{y} \in G$.

Any function $\mathbf{F}$ as above of class C^1 such that all its partial derivatives are bounded in G satisfies the Lipschitz condition. In fact, fix t and $\mathbf{x}, \mathbf{y}$ and consider the function $\phi_i(\tau) = F_i(t, \mathbf{x} + \tau(\mathbf{y} - \mathbf{x}))$. Clearly $\phi_i(0) - \phi_i(1) = F_i(t,\mathbf{x}) - F_i(t,\mathbf{y})$. By the Mean Theorem, there exists $\tau_0 \in (0,1)$ such that $\phi_i(0) - \phi_i(1) = \phi_i(\tau_0)'$. This implies

$$F_i(t,\mathbf{x}) - F_i(t,\mathbf{y}) = \frac{dF_i(t, \mathbf{x} + \tau(\mathbf{y} - \mathbf{x})))}{d\tau}(\tau_0) = \sum_{k=1}^{n}(y_k - x_k)\frac{\partial F_i}{dx_k}(\mathbf{x} + \tau_0\mathbf{y}).$$

Since all partial derivatives are bounded by some constant C, we get

$$|F_i(t,\mathbf{x}) - F_i(t,\mathbf{y})| \leq nCM,$$

where $M = \max\{|y_k - x_k|, k = 1,\ldots,n\}$. By choosing a larger constant C, if needed, we may assume that this is true for all $i = 1,\ldots,n$. Now,

$$|||\mathbf{F}(t,\mathbf{x}) - \mathbf{F}(t,\mathbf{y})||^2 = \sum_{i=1}^{n}|F_i(t,\mathbf{x}) - F_i(t,\mathbf{y})|^2 \leq$$

$$\sum_{i=1}^{n} n^2C^2M^2 = n^3C^2M^2 \leq n^3C^2\sum_{k=1}^{n}|y_k - x_k|^2 = n^3C^2M^2||\mathbf{y} - \mathbf{x}||^2.$$

Taking the square root from the both sides we get what we need.

The condition of boundeness can be easily achieved if $\mathbf{F}$ admits a continuous extension to the closure of G.

Now we are ready to prove the following:

Theorem (Cauchy). *Keep the previous notation. Assume that* $\mathbf{F}$ *satisfies the Lipschitz condition in* $I \times G$. *Let* B *be a closed ball of radius* R *with center at* $\mathbf{x}_0 \in G$ *and let* $(-a+t_0, t_0+a)$ *be chosen small enough such that* $||\mathbf{F}(t,\mathbf{x})|| \le R/a$ *for all* $t \in (-a+t_0, t+0+a)$ *and* $\mathbf{x} \in B$. *Then there exists a unique solution* $\mathbf{f} : (-a+x_0, x_0+a) \to B \subset \mathbb{R}^n$ *of (1) satisfying the initial condition (3).*

Proof. Let $t_0 \in (a,b)$. A function $\mathbf{f} : J \to G$ is a solution of (1) defined on some interval $J \subset I$ containing t_0 if and only if

$$\mathbf{f}(t) = \mathbf{f}(t_0) + \int_{t_0}^{t} \mathbf{F}(\tau, \mathbf{f}(\tau))d\tau. \tag{4}$$

To verify this it is enough to differentiate both sides and use the Fundamental Theorem of Calculus.

Let E be the set of continuous functions $\mathbf{g} : (-a+t_0, t_0+a) \to B$ satisfying $\mathbf{g}(t_0) = \mathbf{x}_0$. Here a and B is chosen as in the assertion of the theorem. For any $\mathbf{g} \in E$ the function

$$\Phi(\mathbf{g})(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{F}(\tau, \mathbf{g}(\tau))d\tau \tag{5}$$

belongs to E. In fact,

$$||\Phi(\mathbf{g})(t) - \mathbf{x}_0|| = \Big| \int_{t_0}^{t} \mathbf{F}(\tau, \mathbf{g}(\tau))d\tau \Big| \le \int_{t_0}^{t} |\mathbf{F}(\tau, \mathbf{g}(\tau))|d\tau \le (R/a)a = R,$$

so that $\Phi(\mathbf{g})((-a+x_0, x_0+a)) \subset B$. Obviously, $\Phi(\mathbf{g})(t_0) = \mathbf{x}_0$.

Since F satisfies the Lipschitz condition we have $||\mathbf{F}(t,\mathbf{x}) - \mathbf{F}(t,\mathbf{y})|| \le K||\mathbf{x} - \mathbf{y}||$, where $\mathbf{x}, \mathbf{y} \in B$. Thus, for any $\mathbf{g}, \mathbf{h} \in E$ we have

$$||\Phi(\mathbf{g})(t) - \Phi(\mathbf{h})(t)|| = || \int_{t_0}^{t} (\mathbf{F}(\tau, \mathbf{g}(\tau)) - \mathbf{F}(\tau, \mathbf{h}(\tau))d\tau|| \le$$

$$\int_{t_0}^{t} ||(\mathbf{F}(\tau, \mathbf{g}(\tau)) - \mathbf{F}(\tau, \mathbf{h}(\tau))||d\tau \le K \int_{t_0}^{t} ||\mathbf{g}(\tau) - \mathbf{h}(\tau)||d\tau \le 2KR|t - t_0|.$$

Here we use that $\mathbf{g}(t), \mathbf{h}(t) \in B$ so, by the triangle inequality, $||\mathbf{g}(\tau) - \mathbf{h}(\tau)|| \le 2R$.

Next

$$||\Phi(\Phi(\mathbf{g})) - \Phi(\Phi(\mathbf{h}))|| = || \int_{t_0}^{t} (\mathbf{F}(\tau, \Phi(\mathbf{g})(\tau)) - \mathbf{F}(\tau, \Phi(\mathbf{h})))(\tau))d\tau|| \le$$

$$\int_{t_0}^{t} ||(\mathbf{F}(\tau, \Phi(\mathbf{g})(\tau)) - \mathbf{F}(\tau, \Phi(\mathbf{h})(\tau))(\tau))||d\tau \le K \int_{t_0}^{t} ||\Phi(\mathbf{g})(\tau) - \Phi(\mathbf{h})(\tau)||d\tau \le$$

$$2KR \int_{t_0}^{t} |\tau - t_0|d\tau = 2K^2 R \frac{|t - t_0|^2}{2}.$$

Starting from $\mathbf{g}_0 = \mathbf{g}$ define by induction $\mathbf{g}_n = \Phi(\mathbf{g}_{n-1})$. Similarly define $\mathbf{h}_n$. Continue as above we easily get

$$||\mathbf{g}_n(t) - \mathbf{h}_n(t)|| \le 2RK^n \frac{|t - t_0|^n}{n!}. \tag{6}$$

Since $n!$ goes to infinity faster than the exponential function, we get that $||\mathbf{g}_n(t) - \mathbf{h}_n(t)|| < 1$ for n large enough.

Now let us show that the sequence $\{\mathbf{g}_n(t)\}$ uniformly converges to a function $\mathbf{f}$ satisfying (4). Consider the infinite sum of functions

$$S = \mathbf{g}_0 + \sum_{n=1}^{\infty} (\mathbf{g}_n - \mathbf{g}_{n-1}).$$

Its partial sum

$$S_n = \mathbf{g}_0 + (\mathbf{g}_1 - \mathbf{g}_0) + (\mathbf{g}_2 - \mathbf{g}_1) + \ldots + (\mathbf{g}_n - \mathbf{g}_{n-1}) = \mathbf{g}_n.$$

Set $\mathbf{h} = \mathbf{g}_1$. Then $\mathbf{h}_n = \mathbf{g}_{n+1}$ and we have shown above that, for n large,

$$||\mathbf{h}_n(t) - \mathbf{g}_n(t)|| = ||\mathbf{g}_{n+1}(t) - \mathbf{g}_n(t))|| < 1.$$

Thus we obtain that the terms of the infinite series S (after we drop a finite number of first terms) are less or equal than the terms of the infinite series $c(1+q+q^2+\ldots+q^n+\ldots)$, where $q < 1$. Since this series absolutely converges, our series uniformly converges. But this means that the sequence of its partial sums $\mathbf{g}_n(t)$ is uniformly convergent. Now let us see that the limit $\mathbf{f}(t) = \lim_{n\to\infty} \mathbf{g}_n(t)$ satisfies the assertion of the theorem. We have

$$\mathbf{g}_n(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{F}(\tau, \mathbf{g}_{n-1}(\tau))d\tau$$

Applying the Lipschitz condition, we get $||\mathbf{F}(t,\mathbf{g}) - \mathbf{F}(t,\mathbf{g}_n)| \leq K||\mathbf{g} - \mathbf{g}_n||$. This implies that the sequence $\phi_n = \mathbf{F}(t, \mathbf{g}_n(t))$ uniformly converges. It remains to take the limit of both sides and to use the standard fact about passing to the limit under integral.

To finish the proof it remains to show the uniqueness of the solution. Assume $\mathbf{f}$ and $\mathbf{g}$ satisfy (4) and $\mathbf{f}(t_0) = \mathbf{g}(t_0)$. Then $\mathbf{g}, \mathbf{h} \in E$ and $\Phi(\mathbf{f}) = \mathbf{f}, \Phi(\mathbf{g}) = \mathbf{g}$. Applying (6) we get (observing that $\mathbf{f}_n = \mathbf{f}, \mathbf{h}_n = \mathbf{h}$ for all n) that

$$\lim_{n\to\infty} ||\mathbf{f}_n(t) - \mathbf{h}_n(t)|| = ||\mathbf{f}(t) - \mathbf{h}(t)|| = 0.$$

This of course implies that $\mathbf{f}(t) = \mathbf{g}(t)$ for all $t \in (t_0 - a, t_0 + a)$. The proof now is finished.

Definition A vector field of class C^k on a manifold M is a map $\xi : M \to TM$ of class C^k such that $\pi(\xi(p)) = p$ for any $p \in M$. Here $\pi" TM \to M$ is the canonical projection of TM to M.

In plane words a vector field chooses for a point $p \in M$ a tangent vector $\xi(p) = (p, \xi_p)$. The dependence on the point must be of class C^k.

When M is an open subset of $\mathbb{R}^n$, a vector field is nothing more than a function $\mathbb{F}$ of class C^k defined on M and taking values in $\mathbb{R}^n$.

Let $\mathcal{D}$ be an open subset of $\mathbb{R}^n$ and

$$\mathbf{F} = (f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})) : \mathcal{D} \to \mathbb{R}^n$$

be a vector field defined on $\mathcal{D}$. We assume that the derivative $DF(\mathbf{x})$ exists at any point $\mathbf{x} \in \mathcal{D}$ and its entries are continuous functions on $\mathcal{D}$. Let $\mathcal{D}'$ be the subset of points $(t, \mathbf{x})$ in $\mathbb{R}^{n+1}$ such that $t \in (a, b), \mathbf{x} \in \mathcal{D}$, where $a < 0 < b$. A function $\phi(t, \mathbf{x})$ defined in $\mathcal{D}'$ is called a flow line of the vector field if, for all $(t, \mathbf{x}) \in \mathcal{D}'$,

$$\frac{\partial \phi(t, \mathbf{x})}{\partial t} = \mathbf{F}(\phi(t, \mathbf{x})), \quad \phi(0, \mathbf{x}) = \mathbf{x}. \tag{1}$$

When we fix $\mathbf{x}$, the function $\phi(t, \mathbf{x})$ describes a path in $\mathcal{D}$ such that its value at $t = 0$ is $\mathbf{x}$, and at each point of the path the velocity vector is equal to the value of the vector field at this point. When we fix $t = t_0$ the function $\phi(t_0, \mathbf{x})$ describes a map from $\mathcal{D}$ to $\mathbb{R}^n$. This map takes a point $\mathbf{x}$ "flow" along the path $t \to \phi(t, \mathbf{x})$ until time t_0.

Consider the derivative $D_{\mathbf{x}}\phi(t, \mathbf{x})$ of $\phi(t, \mathbf{x})$ as a function in $\mathbf{x}$. This is a square matrix of size n which depends on t and $\mathbf{x}$. Differentiating the first equationin (1), using the chain rule, we obtain

$$D_{\mathbf{x}} \frac{\partial}{\partial t} \phi(t, \mathbf{x}) = \frac{\partial}{\partial t} D_{\mathbf{x}} \phi(t, \mathbf{x}) = D_{\mathbf{x}} \mathbf{F}(\phi(t, \mathbf{x})) = D\mathbf{F}(\phi(t, \mathbf{x})) \circ D_{\mathbf{x}} \phi(t, \mathbf{x}).$$

The equation

$$\frac{\partial}{\partial t} D_{\mathbf{x}} \phi(t, \mathbf{x}) = D\mathbf{F}(\phi(t, \mathbf{x})) \circ D_{\mathbf{x}} \phi(t, \mathbf{x}) \tag{2}$$

is called the *equation of first variation.*

Recall that for any square matrix $A = (a_{ij})$ the sum of its diagonal elements is called the *trace* of A:

$$Tr(A) = a_{11} + a_{22} + \ldots + a_{nn}.$$

If we take $A = D\mathbf{F}(\mathbf{x})$, then the trace of A is called the *divergence of a vector field* $\mathbf{F}$ and is denoted by $div(\mathbf{F})$:

$$div(\mathbf{F}) = Tr(D\mathbf{F}(\mathbf{x})) = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \ldots + \frac{\partial f_n}{\partial x_n}.$$

We want to prove the following

Theorem.

$$div(\mathbf{F}(\mathbf{x})) = \frac{d}{dt} \det D_{\mathbf{x}} \phi(t, \mathbf{x})|_{t=0}.$$

Let us fix $\mathbf{x}$ and consider the matrix $D_{\mathbf{x}}\phi(t, \mathbf{x})$ as a matrix $A(t)$ whose entries are functions in t. Let $B(t)$ be the matrix $D\mathbf{F}(\phi(t, \mathbf{x}))$, also depending on t. The equation (2) gives us

$$A(t)' = B(t)A(t). \tag{3}$$

Here the derivative of $A(t) = (a_{ij}(t))$ is the matrix of derivatives $(a_{ij}(t)')$. The determinant $|A(t)|$ of matrix $A(t)$ is a function of t which we can differentiate.

Lemma. *Suppose (3) holds. Let $|A(t)|$ be the determinant of $A(t)$ and let $B(t) = (b_{ij})$. Then*

$$\frac{d}{dt}|A(t)| = Tr(B)|A(t)|.$$

Proof. Let $\bar{A}_1(t), \ldots, \bar{A}_n(t)$ be the rows of $A(t)$. We need the following *formula for the derivative of determinant*:

$$\frac{d}{dt}|A(t)| = \det \begin{pmatrix} \bar{A}_1(t)' \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} + \det \begin{pmatrix} \bar{A}_1(t) \\ \bar{A}_2(t)' \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} + \ldots + \det \begin{pmatrix} \bar{A}_1(t) \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t)' \end{pmatrix}. \tag{4}$$

First one checks this formula for 2×2-matrices.

$$\frac{d}{dt} \det \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix} = [a(t)d(t) - b(t)c(t)]' = a'd + ad' - b'c - bc' =$$

$$(a'd - b'c) + (ad' - bc') = \det \begin{pmatrix} a(t)' & b(t)' \\ c(t) & d(t) \end{pmatrix} + \det \begin{pmatrix} a(t) & b(t) \\ c(t)' & d(t)' \end{pmatrix}.$$

In the general case the proof is by induction. First we expand the determinant of along the first row to get

$$|A(t)| = a_{11}|A_{11}| - a_{12}|A_{12}| + \ldots + (-1)^{1+n} a_{1n}|A_{1n}|.$$

Then differentiate

$$|A(t)|' = a'_{11}|A_{11}| - a'_{12}|A_{12}| + \ldots + (-1)^{1+n} a'_{1n}|A_{1n}| +$$

$$a_{11}|A_{11}|' - a_{12}|A_{12}|' + \ldots + (-1)^{1+n} a_{1n}|A_{1n}|'.$$

Then apply the inductive formula for the derivatives of the determinants $|A_{1j}|$ of $(n-1) \times (n-1)$-matrices .

Now let $\bar{B}_i(t)$ be the rows of $B(t)$. Then

$$A(t)' = \begin{pmatrix} \bar{A}_1(t)' \\ \bar{A}_2(t)' \\ \vdots \\ \bar{A}_n(t)' \end{pmatrix} = B(t)A(t) = \begin{pmatrix} \bar{B}_1(t) \\ \bar{B}_2(t) \\ \vdots \\ \bar{B}_n(t) \end{pmatrix} A(t) = \begin{pmatrix} \bar{B}_1(t)A(t) \\ \bar{B}_2(t)A(t) \\ \vdots \\ \bar{B}_n(t)A(t) \end{pmatrix}.$$

Here each row is the product of the rows $\bar{B}_i(t)$ and the matrix $A(t)$ (which is a row matrix). Comparing the rows, we get

$$A_i(t)' = B_i(t)A(t) = b_{i1}\bar{A}_1(t) + b_{i2}\bar{A}_1(t) + \ldots + b_{in}\bar{A}_n(t).$$

Now we have

$$\det \begin{pmatrix} \bar{A}_1(t)' \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} = \det \begin{pmatrix} \sum_{j=1}^n b_{1j}\bar{A}_j(t) \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} =$$

$$\det \begin{pmatrix} b_{11}\bar{A}_1(t) \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} + \det \begin{pmatrix} b_{12}\bar{A}_2(t) \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} + \ldots + \det \begin{pmatrix} b_{1n}\bar{A}_n(t) \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} = b_{11} \det \begin{pmatrix} \bar{A}_1(t) \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} = b_{11}|A(t)|.$$

Here we use the properties of the determinants (note that all determinants except the first one are equal to zero because they have two proportional rows). Similarly, we get

$$\det \begin{pmatrix} \bar{A}_1(t) \\ \bar{A}_2(t) \\ \vdots \\ \bar{A}_i(t)' \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} = \det \begin{pmatrix} \bar{A}_1(t) \\ \bar{A}_2(t) \\ \vdots \\ \sum_{j=1}^n b_{ij}\bar{A}_j(t) \\ \vdots \\ \bar{A}_n(t) \end{pmatrix} = b_{ii}|A(t)|.$$

Thus equation (4) gives us

$$|A(t)|' = (b_{11} + b_{22} + \ldots + b_{nn})|A(t)| = Tr(B)|A(t)|.$$

The lemma is proven.

Proof of the Theorem:

Recall that in our case, the entries of $B(t) = D_{\mathbf{x}}F(\phi(t,\mathbf{x}))$ are equal to

$$b_{ij} = \frac{\partial f_i}{\partial x_j}(\phi(t,\mathbf{x}))$$

Thus we obtain

$$\frac{d}{dt}|D_{\mathbf{x}}\phi(t,\mathbf{x})| = (\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(\phi(t,\mathbf{x}))|D_{\mathbf{x}}(\mathbf{F}(\phi(t,\mathbf{x}))|.$$

Taking $t = 0$, we get

$$\frac{d}{dt}|D_{\mathbf{x}}\phi(t,\mathbf{x})|_{t=0} = (\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(\phi(0,\mathbf{x}))|D_{\mathbf{x}}(\phi(0,\mathbf{x}))| = div\mathbf{F}(\mathbf{x}). \tag{5}$$

Here we use that, by definition $\phi(0,\mathbf{x}) = \mathbf{x}$ and the derivative of the idenity map $\mathbf{x} \to \phi(0,\mathbf{x})$ is the idenity matrix.

Let us give an interpretation of the assertion of the theorem. Recall that if $v_1,\ldots,v_n \in \mathbb{R}^n$ and A is a square matrix, then the (oriented) volume of the parallelopiped formed by the vectors $Av_1,\ldots,Av_n$ is equal to the (oriented) volume of the parallelopiped formed by the vectors $v_1,\ldots,v_n$ times the determinant of A. Let us take $A = D_{\mathbf{x}}\phi(t,\mathbf{x})$ and $v_1,\ldots,v_n$ be vectors emanating from the point $\mathbf{x}$ which form the parallelopiped $P(0)$ with nonzero volume. Then the linear approximation of $\phi(t,\mathbf{x})$ transforms these vectors to $Av_1,\ldots,Av_n$ emanating from the point $\phi(t,\mathbf{x})$. Let $P(t)$ be the corresponding parallelopiped. Thus the derivative $D_{\mathbf{x}}(\phi(t,\mathbf{x})$ transforms $P(0)$ to $P(t)$. If $P(0)$ is small enough, $P(t)$ is close to the image of $P(0)$ under the map $\phi(t,\mathbf{x})$. Let $\mathcal{V}(t)$ be the oriented volume of $P(t)$. Then, by above,

$$|D_{\mathbf{x}}\phi(t,\mathbf{x})| = \mathcal{V}(t)/\mathcal{V}(0).$$

Thus, the theorem says that

$$div\mathbf{F}(\mathbf{x}) = \frac{d}{dt}\mathcal{V}(t)/\mathcal{V}(0)|_{t=0} = \frac{1}{\mathcal{V}(0)}\frac{d}{dt}\mathcal{V}(t)|_{t=0}.$$

Math. 396. Notes 6: Alternating forms

We start with the determinant of a matrix. Let

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1j} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mj} & \ldots & a_{mn} \end{pmatrix}$$

Let us view A as a collection of its rows

$$\bar{A}_i = (a_{i1},\ldots,a_{in}), \quad i = 1,\ldots,m.$$

Thus any function on the set M_{mn} of $m \times n$ matrices can be viewed as a function on the Cartesian product $(\mathbb{R}^n)^m$ of m copies of $\mathbb{R}^n$.

Definition A *determinant* is a function

$$\det : (\mathbb{R}^n)^m \to \mathbb{R}$$

satisfying the following properties:

(i) for any $j = 1, \ldots, n$ and any fixed vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n$ the function $\mathbb{R}^n \to \mathbb{R}$ defined by $\mathbf{v} \to \det(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n)$ is linear;

(ii) $\det(\mathbf{v}_1, \ldots, \mathbf{v}_n) = 0$ if $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$;

(iii) $\det(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$, where $\mathbf{e}_i = (0, \ldots, 0, 1, \ldots, 0)$ with 1 at the i-th spot.

Notice that properties (i) and (ii) imply

(ii)' If the i-th argument $\mathbf{v}_i$ is switched with the j-th argument $\mathbf{v}_j$ we get

$$\det(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n) = -\det(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n).$$

In fact, applying (i) and (ii), we get

$$0 = \det(\mathbf{v}_1, \ldots, \mathbf{v}_i + \mathbf{v}_j, \ldots, \mathbf{v}_i + \mathbf{v}_j, \ldots, \mathbf{v}_n) =$$
$$\det(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n) + \det(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) +$$
$$\det(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n) + \det(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) =$$
$$= \det(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n) + \det(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n).$$

Let us see that properties (i)-(iii) determine the function det uniquely. Write any vector $\mathbf{v}_i = (a_{i1}, \ldots, a_{in})$ in the form $\mathbf{v}_i = a_{i1}\mathbf{e}_1 + \ldots + a_{in}\mathbf{e}_n$. Using the property (i) we have

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \det(\sum_{j=1}^{n} a_{1j}\mathbf{e}_j, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{j=1}^{n} a_{1j} \det(\mathbf{e}_j, \mathbf{v}_2, \ldots, \mathbf{v}_n) =$$

$$\sum_{j_1=1}^{n} \sum_{j_2=1}^{n} a_{1j_1} a_{2j_2} \det(\mathbf{e}_{j_1}, \mathbf{e}_{j_2} \mathbf{v}_3, \ldots, \mathbf{v}_n) = \sum_{j_1, j_2, \ldots, j_n = 1}^{n} a_{1j_1} \ldots a_{j_n} \det(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \ldots, \mathbf{e}_{j_n}).$$

Now by property (ii), all the terms in the sum corresponding to the collection of not all distinct indices $(j_1, \ldots, j_n)$ must be equal to zero. So we may assume that the indices $(j_1, \ldots, j_n)$ are all distinct. In other words $(j_1, \ldots, j_n)$ is a *permutation* of numbers $\{1, 2, \ldots, n\}$. Using property (ii)' we get

$$\det(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \ldots, \mathbf{e}_{j_n}) = \epsilon(j_1, \ldots, j_n) \det(\mathbf{e}_1, \ldots, \mathbf{e}_n).$$

Here $\epsilon(j_1, \ldots, j_n)$ is the sign of the permutation $(j_1, \ldots, j_n)$. It is equal to 1 if one needs to make even number of switches to put the permutation in the standard order , and -1 otherwise.

Finally we obtain

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{(j_1, j_2, \ldots, j_n) \in S_n} \epsilon(j_1, \ldots, j_n) a_{1j_1} \ldots a_{nj_n} \det(\mathbf{e}_1, \ldots, \mathbf{e}_n), \qquad (1)$$

where the sum its taken over the set S_n of all permutations of the numbers $1, 2, \ldots, n$. Using property (iii), we get the formula for the determinant

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{(j_1, j_2, \ldots, j_n) \in S_n} \epsilon(j_1, \ldots, j_n) a_{1j_1} \ldots a_{nj_n}. \qquad (2)$$

This proves the uniqueness of the determinant function. To prove the existence, one defines the function det by induction. First if $n = 1$, i.e. $\mathbb{R}^n = \mathbb{R}$, we set

$$\det(a) = a.$$

Now assuming that we know what is the determinant of square matrices of size $(n-1)$ we set

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}. \tag{3}$$

Here M_{ij} is the determinant of the matrix obtained from the matrix A by deleting the i-th row and the j-th column. One checks that this function satisfies all the properties (i) -(iii). We skip this verification referring to any (good) textbook in linear algebra. Notice that as a corollary of the uniqueness theorem, one can use formula (3) for any $i = 1, \ldots, n$. Also one may use similar formula

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}. \tag{3'}$$

This is called the expansion of the determinant along the $j-th$ column.

Now we shall extend the notion of the determinant first by refusing to demand property (iii), and second to define the function for any number of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in any linear space V.

Definition. An *alternating k-form*(or an *skew-symmetric k-form*) on a linear space V is a function

$$\omega : V^k = V \times \ldots \times V \to \mathbb{R}$$

satisfying the following properties:

(i) for any $j = 1, \ldots, k$ and any fixed vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k$ the function $\mathbb{R}^n \to \mathbb{R}$ defined by $\mathbf{v} \to \omega(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k)$ is linear;

(ii) $\omega(\mathbf{v}_1, \ldots, \mathbf{v}_k) = 0$ if $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$.
Similarly to property (ii)' of the determinant we have

(ii)' $\omega(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_k) = -\omega(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_k)$.

We denote the set of alternating k-forms on V by $A^k(V)$. Obviously the sum of alternating k-forms and the scalar multiple of an alternating k-form is an alternating k-form.

If you are unfamiliar with the notion of an abstract linear space, think about V as the space $\mathbb{R}^n$ or its linear subspace $V \subset \mathbb{R}^n$.

Examples 1. If $k = 1$, any alternating 1-form is just a linear function on V.

2. Let $V = \mathbb{R}^3$, for any vector $\mathbf{a} \in \mathbb{R}^3$ the function

$$\omega_{\mathbf{a}}(\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{a}.$$

is an alternating 2-form.

3. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis in V. This means that any vector $\mathbf{v} \in V$ can be written uniquely in the form

$$\mathbf{v} = a_1\mathbf{e}_1 + \ldots + a_n\mathbf{e}_n.$$

Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$. Write

$$\mathbf{v}_i = \sum_{j=1}^{n} a_{ij}\mathbf{e}_j, \quad i = 1, \ldots, k. \tag{4}$$

Consider the matrix

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{k1} & \ldots & a_{kn} \end{pmatrix}.$$

For any subset $J = \{j_1, \ldots, j_k\}$ of the set $[1, n] = \{1, \ldots, n\}$ of k elements consider the $k \times k$-submtarix of A defined by choosing the columns $A_{j_1}, \ldots, A_{j_k}$

$$A_J = \begin{pmatrix} a_{1j_1} & \ldots & a_{1j_k} \\ \vdots & \vdots & \vdots \\ a_{kj_1} & \ldots & a_{kj_k} \end{pmatrix}.$$

Define

$$\psi_J(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \det A_J \tag{5}$$

Using the properties of the determinant function, we see that the function ψ_J is an alternating k-form. It is called a *basic alternating k-form*. Since the number of subsets of $[1, n]$ consisting of k elements is equal to

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

we have $\binom{n}{k}$ basic functions ψ_J.

Arguing the same as in the proof of the formula (2) for the determinant we find that, for any $\omega \in A^k(V)$,

$$\omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \sum_{(j_1, j_2, \ldots, j_k)} a_{1j_1} \ldots a_{kj_k} \omega(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}).$$

Here $(j_1, \ldots, j_k)$ is an ordered sequence of k distinct numbers from the set $[1, n] = \{1, \ldots, n\}$. Note that in particular

$$A^k(V) = \{0\} \quad \text{if } k > n.$$

Let $I \subset [1, n]$ consist of k elements (we write $\#I = k$). Let S_I be the set of all permutations of the elements from I. That is the set of ordered k-tuples $(j_1, \ldots, j_k)$ where all $j_i \in I$. We can write the previous equality as

$$\omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \sum_{J \subset [1,n], \#J = k} \big(\sum_{(j_1, j_2, \ldots, j_k) \in S_J} a_{1j_1} \ldots a_{kj_k} \omega(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}) \big) =$$

$$\sum_{J \subset [1,n], \#J = k} \big(\sum_{(j_1, j_2, \ldots, j_k) \in S_J} \epsilon(j_1, \ldots, j_k) a_{1j_1} \ldots a_{kj_k} \big) \omega(\mathbf{e}_{j'_1}, \ldots, \mathbf{e}_{j'_k}).$$

Here $\{j'_1, \ldots, j'_k\} = J$ and $j'_1 < \ldots < j'_k$. The expression in the brackets can be easily recognized as the determinant of the matrix A_J. Thus

$$\omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \sum_{J \subset [1,n], \#J = k} \omega(\mathbf{e}_{j'_1}, \ldots, \mathbf{e}_{j'_k}) \psi_J(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k).$$

This shows that

$$\omega = \sum_{J\subset[1,n],\#J=k} \lambda_J \psi_J, \tag{6}$$

where

$$\lambda_J = \omega(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}), \quad J = \{j_1, \ldots, j_k\}, \quad j_1 < \ldots < j_k.$$

Remark. The set $A^k(V)$ is a linear space with respect to addition and scalar multiplication of functions. The equality (6) shows that the basic functions ψ_J span $A^k(V)$. One can also prove that the basic functions are linearly independent. In fact, if

$$\sum_{I\subset[1,n],\#I=k} \lambda_I \psi_I = 0,$$

we evaluate this sum on $(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k})$, $\quad j_1 < \ldots < j_k$. For each subset $I \neq J = \{j_1, \ldots, j_k\}$ of $[1, n]$ the matrix A with rows $\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}$ contains the zero column corresponding to the index $i \in I, i \notin J$. Thus the matrix A_I has a zero column, and hence zero determinant. Thus $\psi_I(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}) = 0$. On the other hand, if $I = J$, the matrix A_I is the identity matrix, and hence $\psi_J(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}) = \det A_J = 1$. So

$$0 = \sum_{I\subset[1,n],\#I=k} \lambda_I \psi_I(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}) = \lambda_J \psi_J(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}) = \lambda_J.$$

This shows that ψ_J's are linearly independent functions. Therefore

$$\dim A^k(V) = \binom{n}{k}.$$

When $k = 1$, $A^(V)$ is the space of linear functions on V. It is usually denoted by V^* and is called the *dual space* of V. If $(e_1, \ldots, e_n)$ is a basis in V, then the linear functions $(e_1^*, \ldots, e_n^*)$ defined by

$$e_i^*(e_j) = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

form a basis in V^*. In fact, for each $f \in V^*$ we have

$$f = \sum_{j=1}^{n} f(e_j) e_j^*,$$

which can be easily checked by comparing the values of both sides at each basis vector e_i. Since V and $(V^*)^*$ have the same dimension (if V is finite-dimensional), they are isomorphic linear spaces. However, one can do better and define an isomorphism which does not depend on the choice of a basis (such an isomorphism is called a *natural* or *canonical* isomorphism). For this we have to consider any element $v \in V$ as a linear function on V^*. It is defined by

$$v(f) = f(v).$$

This defines a linear map from V to $(V^*)^*$. Its kernel is zero, since $f(v) = 0$ for all $f \in V^*$ implies $v = 0$ (if $v \neq 0$ we consider the linear function f defined by $f(v) = 1$ and $f(w) = 0$ if w is not proportional to v). Now the dimensions of V and $(V^*)^*$ are the same so the map $V \to (V^*)^*$ is an isomorphism.

Example 4. Consider Example 2. We have

$$\omega_{\mathbf{a}}(\mathbf{v}_1,\mathbf{v}_2) = \alpha_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & x_3 \end{vmatrix} - \alpha_2 \begin{vmatrix} x_1 & x_3 \\ y_1 & x_3 \end{vmatrix} + \alpha_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} =$$

$$\alpha_1\psi_{23}(\mathbf{v}_1,\mathbf{v}_2) - \alpha_2\psi_{13}(\mathbf{v}_1,\mathbf{v}_2) + \alpha_3\psi_{12}(\mathbf{v}_1,\mathbf{v}_2).$$

We also check that

$$\omega_{\mathbf{a}}(\mathbf{e}_2,\mathbf{e}_3) = (\mathbf{e}_2 \times \mathbf{e}_3)\cdot\mathbf{a} = \mathbf{e}_1\cdot\mathbf{a} = \alpha_1,$$

$$\omega_{\mathbf{a}}(\mathbf{e}_1,\mathbf{e}_3) = (\mathbf{e}_1 \times \mathbf{e}_3)\cdot\mathbf{a} = -\mathbf{e}_2\cdot\mathbf{a} = -\alpha_2,$$

$$\omega_{\mathbf{a}}(\mathbf{e}_1,\mathbf{e}_2) = (\mathbf{e}_1 \times \mathbf{e}_2)\cdot\mathbf{a} = \mathbf{e}_3\cdot\mathbf{a} = \alpha_3.$$

Example 5. Let $f_1,\ldots,f_k \in A^1(V)$. Define $f_1\wedge f_2\wedge\ldots\wedge f_k \in A^k(V)$ by the formula

$$f_1\wedge f_2\wedge\ldots\wedge f_k(\mathbf{v}_1,\ldots,\mathbf{v}_k) = \det\begin{pmatrix} f_1(\mathbf{v}_1) & f_1(\mathbf{v}_2) & \ldots & f_1(\mathbf{v}_k) \\ f_2(\mathbf{v}_1) & f_2(\mathbf{v}_2) & \ldots & f_2(\mathbf{v}_k) \\ \vdots & \vdots & \vdots & \vdots \\ f_k(\mathbf{v}_1) & f_k(\mathbf{v}_2) & \ldots & f_k(\mathbf{v}_k) \end{pmatrix}.$$

Using the properties of the determinant we check that this is an alternating k-form. Let $(e_1,\ldots,e_n)$ be a basis in V and $(e_1^*,\ldots,e_n^*)$ be the dual basis. For each subset $I = \{i_1 < \ldots\ldots < i_k\} \subset [1,n]$ consider the alternating k-form

$$e_I^* = e_{i_1}^* \wedge \ldots \wedge e_{i_k}^*. \tag{8}$$

Note that for any $\mathbf{v} = a_1e_1 + \ldots + a_ne_n$ we have $e_i^*(\mathbf{v}) = a_i$. By definition $e_I^*(\mathbf{v}_1,\ldots,\mathbf{v}_k) = \det A_I$, where A_I is the $(k\times k)$-submatrix of the matrix A whose columns are the coordinates of $\mathbf{v}_j$'s with respect to the basis $(e_1,\ldots,e_n)$ defined by rows with indices $(i_1,\ldots,i_k)$. Thus we see that e_I^* coincides with the basic function ψ_I.

A generalization of the previous example leads to an important operation over alternating forms.

Definition. Let $\omega \in A^k(V), \eta \in A^m(V)$. Define

$$\omega\wedge\eta \in A^{m+k}(V)$$

by the following formula

$$\omega\wedge\eta(\mathbf{v}_1,\ldots,\mathbf{v}_{k+m}) = \frac{1}{k!m!}\sum_{(j_1,\ldots,j_{k+m})\in S_{k+m}} \epsilon(j_1,\ldots,j_{k+m})\omega(\mathbf{v}_{j_1},\ldots,\mathbf{v}_{j_k})\eta(\mathbf{v}_{j_{k+1}},\ldots,\mathbf{v}_{j_{k+m}}). \tag{9}$$

We shall see the reason for inserting factorials.

It is easy to see that this function is linear in each variable. Let us check that it satisfies property (ii) of alternating forms. Let $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$ with $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$. Obvioulsy, all terms in the sum (7) such that $i,j \in \{j_1,\ldots,j_m\}$ or $i,j \in \{j_{k+1},\ldots,j_{k+m}\}$ are equal to zero (because ω and η are alternating forms). Consider the term in the sum with $i \in \{j_1, \dots, j_m\}, j \in \{j_{k+1}, \dots, j_{k+m}\}$. There is another term in the sum corresponding to the switching i and j in the permutation $(j_1, \dots, j_{k+m})$. Since they enter into the sum with different sign $\epsilon(j_1, \dots, j_{k+m})$ they cancel each other. So the resulting sum is zero. Thus our operation is well defined and gives us an alternating $(k+m)$-form.

Example 6. If $f_1, f_2 \in A^1(V)$, then

$$f_1 \wedge f_2(\mathbf{v}_1, \mathbf{v}_2) = f_1(\mathbf{v}_1)f_2(\mathbf{v}_2) - f_1(\mathbf{v}_2)f_2(\mathbf{v}_1) = \det \begin{pmatrix} f_1(\mathbf{v}_1) & f_1(\mathbf{v}_2) \\ f_2(\mathbf{v}_1) & f_2(\mathbf{v}_2) \end{pmatrix}.$$

This agrees with the definition from Example 5.

Lemma. Let $f_1, \dots, f_k, f_{k+1}, \dots, f_m \in V^*$. Then

$$(f_1 \wedge \dots \wedge f_k) \wedge (f_{k+1} \wedge \dots \wedge f_m) = f_1 \wedge \dots \wedge f_k \wedge f_{k+1} \wedge \dots \wedge f_m.$$

Proof. This immediately follows from the following *Laplace Formula*. Let A be a square matrix of size n. For every subset I of $[1, n]$ with $1 \leq k < n$ elements denote by A_I the submatrix of A formed by the first k rows and columns with indices from I. Similarly, for any subset J of $n-k$ elements denote by B_J the submatrix formed by the last $n-m$ rows and columns with indices from J. Then

$$\det A = \sum_{I \subset [1,n], \#I = k} \epsilon(I, \bar{I}) \det A_I \det B_{\bar{I}}.$$

Here $\bar{I}$ denotes the set $[1, n] \setminus I$, and $\epsilon(I, \bar{I})$ is the sign of the permutation $(i_1, \dots, i_k, j_1, \dots, j_{n-k})$, where $I = \{i_1 < \dots < i_k\}, \bar{I} = \{j_1 < \dots < j_{n-k}\}$. To prove the Laplace formula, one notices that the right-hand side satisfies all the properties of the determinant function with respect to rows. By the uniqueness of the determinant we obtain that we want.

Theorem (Properties of the wedge product).
(i) (linearity)

$$(\lambda\omega + \lambda'\omega') \wedge \eta = \lambda\omega \wedge \eta + \lambda'\omega' \wedge \eta, \quad \forall \omega, \omega' \in A^k(V), \forall \eta \in A^m(V), \forall \lambda, \lambda' \in \mathbb{R};$$

(ii) (supercommutativity)

$$\omega \wedge \eta = (-1)^{km} \eta \wedge \omega, \quad \forall \omega \in A^k(V), \forall \eta \in A^m(V);$$

(iii) (associativity)

$$(\omega \wedge \eta) \wedge \gamma = \omega \wedge (\eta \wedge \gamma), \quad \forall \omega \in A^k(V), \forall \eta \in A^m(V), \forall \gamma \in A^l(V).$$

Proof. (i) Use that

$$(\lambda\omega + \lambda'\omega')(\mathbf{v}_{j_1}, \dots, \mathbf{v}_{j_k}) = \lambda\omega(\mathbf{v}_{j_1}, \dots, \mathbf{v}_{j_k}) + \lambda'\omega'(\mathbf{v}_{j_1}, \dots, \mathbf{v}_{j_k}).$$

Plugging this into the sum (8) we easily get

$$(\lambda\omega + \lambda'\omega') \wedge \eta(\mathbf{v}_1, \dots, \mathbf{v}_{k+m}) = \lambda\omega \wedge \eta(\mathbf{v}_1, \dots, \mathbf{v}_{k+m}) + \lambda'\omega' \wedge \eta(\mathbf{v}_1, \dots, \mathbf{v}_{k+m}).$$

This checks (i). The sum $k!m!\eta \wedge \omega(\mathbf{v}_1, \ldots, \mathbf{v}_{k+m})$ is the sum of the terms

$$\epsilon(j_1, \ldots, j_{k+m})\eta(\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_m})\omega(\mathbf{v}_{j_{m+1}}, \ldots, \mathbf{v}_{j_{k+m}}).$$

The sum $k!m!\omega \wedge \eta(\mathbf{v}_1, \ldots, \mathbf{v}_{k+m})$ consists of the terms

$$\epsilon(j_1, \ldots, j_{k+m})\omega(\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_k})\eta(\mathbf{v}_{j_{k+1}}, \ldots, \mathbf{v}_{j_{k+m}}).$$

We can write each last term as

$$(-1)^{km}\epsilon(j_{k+1}, \ldots, j_{k+m}, j_1, \ldots, j_k)\omega(\mathbf{v}_{j_1}, \ldots, \mathbf{v}_{j_k})\eta(\mathbf{v}_{j_{k+1}}, \ldots, \mathbf{v}_{j_{k+m}}).$$

Here we use that we need km switches in order to go from the permutation $(j_1, \ldots, j_{k+m})$ to the permutation $(j_{k+1}, \ldots, j_{k+m}, j_1, \ldots, j_k)$ (each $j_1, \ldots, j_k$ should jump over each $j_{k+1}, \ldots, j_{k+m}$). Obviously this term is equal to $(-1)^{km}$ times a term in the first sum. Thus each term in the sum for $\omega \wedge \eta$ is equal to $(-1)^{km}$ times a term in the sum for $\eta \wedge \omega$. This proves the property (ii).

Finally let us prove property (iii). By Example 5, one can choose a basis in $A^k(V)$ which consists of forms $\psi_I = e^*_{i_1} \wedge \ldots \wedge e^*_{i_k}$. Writing η and ω as a linear combination of such forms, and applying property (i), we may assume that $\omega = f_1 \wedge \ldots \wedge f_k$ and $\eta = g_1 \wedge \ldots \wedge g_m, \gamma = h_1 \wedge \ldots \wedge h_s$ for some linear functions $f_1, \ldots, f_k, g_1, \ldots, g_m, h_1, \ldots, h_s$. By the Lemma

$$(\omega \wedge \eta) \wedge \gamma = ((f_1 \wedge \ldots, \wedge f_k) \wedge (g_1 \wedge \ldots \wedge g_m)) \wedge (h_1 \wedge \ldots \wedge h_s) =$$

$$(f_1 \wedge \ldots, \wedge f_k \wedge g_1 \wedge \ldots \wedge g_m) \wedge (h_1 \wedge \ldots \wedge h_s) = f_1 \wedge \ldots, \wedge f_k \wedge g_1 \wedge \ldots \wedge g_m \wedge h_1 \wedge \ldots \wedge h_s.$$

Similarly, we check that

$$\omega \wedge (\eta \wedge \gamma) = f_1 \wedge \ldots, \wedge f_k \wedge g_1 \wedge \ldots \wedge g_m \wedge h_1 \wedge \ldots \wedge h_s.$$

This proves the theorem.

As soon as we have the associativity we can always define the wedge product of any number of alternating forms.

$$\omega_1 \wedge \omega_2 \wedge \omega_3 = \omega_1 \wedge (\omega_2 \wedge \omega_3) = (\omega_1 \wedge \omega_2) \wedge \omega_3,$$

$$\omega_1 \wedge \omega_2 \wedge \omega_3 \wedge \omega_4 = \omega_1 \wedge (\omega_2 \wedge \omega_3 \wedge \omega_4) = (\omega_1 \wedge \omega_2 \wedge \omega_3) \wedge \omega_4,$$

and so on.

Let $T : V \to W$ be a linear map between linear spaces. Recall that this means that, for any $\mathbf{v}_1, \mathbf{v}_2 \in V$, and any scalars λ_1, λ_2,

$$T(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2) = \lambda_1 T(\mathbf{v}_1) + \lambda_2 T(\mathbf{v}_2).$$

For any linear function $f : W \to \mathbb{R}$ its composition with T is a linear function on V. We denote it by $T^*(f)$. That is,

$$T^*(f) = f \circ T.$$

Let us generalize this to any alternating k-form. For any $\omega \in A^k(W)$ define $T^*(\omega) \in A^k(V)$ by the formula

$$T^*(\omega)(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \omega(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_k)). \tag{10}$$

It is obvious that this definition is legal, i.e. the function $T^*(\omega) : V^k \to \mathbb{R}$ is indeed an alternating k-form. The form $T^*(\omega)$ is called the *inverse image* of ω under the linear map T.

It easily follows from the definition that for any scalars, λ, λ' and any $\omega, \omega' \in A^k(W)$, we have

$$T^*(\lambda\omega + \lambda'\omega') = \lambda T^*(\omega) + \lambda' T^*(\omega').$$

In other words,

$$T^* : A^k(W) \to A^k(V)$$

is a linear map.

The operation of the inverse image also behaves well with respect to the wedge product.

Theorem 2. *For any* $\omega \in A^k(W), \eta \in A^m(W)$,

$$T^*(\omega \wedge \eta) = T^*(\omega) \wedge T^*(\eta).$$

Proof. We can write any alternating k form as a linear combination of the wedge products of linear forms. Since T^* is a linear map, and $\wedge$-product satisfies property (1) of Theorem 1, it suffices to check the assertion in the case when ω and η are the wedge products of linear functions, i.e. to prove that

$$T^*((f_1 \wedge \ldots \wedge f_k) \wedge (g_1 \wedge \ldots \wedge g_m)) = (T^*(f_1) \wedge \ldots \wedge T^*(f_k)) \wedge (T^*(g_1 \wedge \ldots \wedge g_m)).$$

Applying the Lemma, we have to check that

$$T^*(f_1 \wedge \ldots \wedge f_k \wedge g_1 \wedge \ldots \wedge g_m) = (T^*(f_1) \wedge \ldots \wedge T^*(f_k)) \wedge (T^*(g_1 \wedge \ldots \wedge g_m)).$$

By associativity, and induction on $m + k$ it suffices to verify that

$$T^*(f \wedge g) = T^*(f) \wedge T^*(g).$$

This now follows easily from the definition of the wedge products of linear functions.

To see explicitly how T^* acts on forms, it suffices to find an explicit formula for $T^*(\psi_I)$ for basic functions ψ_I. Indeed we can write any $\omega \in A^k(W)$ in the form $\omega = \sum_I \lambda_I \psi_I$ and obtain

$$T^*(\omega) = \sum \lambda_I T^*(\psi_I).$$

So, if we know $T^*(\psi_I)$, we know everything. Recall that the functions ψ_I depend on the choice of a basis in W. Let $h_1, \ldots, h_m$ be a basis in W and $e_1, \ldots, e_m$ be a basis in V. The linear map T is completely determined by the images $T(e_j)$ of each e_j. Let

$$T(e_j) = \sum_{i=1}^{m} a_{ij} h_i \quad j = 1, \ldots, n.$$

We can put this data in the matrix

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}.$$

It is called the *matrix of* T with respect to the basis $e_1, \ldots, e_n$ in V and the basis $h_1, \ldots, h_m$ in W. For any $\mathbf{v} = \alpha_1 e_1 + \ldots + \alpha_n e_n \in V$, we have

$$T(\mathbf{v}) = T(\alpha_1 e_1 + \ldots + \alpha_n e_n) = \alpha_1 T(e_1) + \ldots + \alpha_n T(e_n) =$$

$$\alpha_1(a_{11}h_1 + \ldots + a_{m1}h_m) + \ldots + \alpha_n(a_{1n}h_1 + \ldots + a_{mn}h_m) =$$

$$(\alpha_1 a_{11} + \ldots + \alpha_n a_{1n})h_1 + \ldots + (\alpha_1 a_{m1} + \ldots + \alpha_n a_{mn})h_m.$$

This shows that the coordinates of $T(\mathbf{v})$ with respect to the basis $h_1, \ldots, h_m$ are equal to the vector $(\beta_1, \ldots, \beta_m)$, where

$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} = A \cdot \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

This shows that the knowledge of A completely determines T. Now let us see how T transforms linear functions ψ_i on W to linear functions on V. Let $h_1^*, \ldots, h_k^*$ be the dual basis of $(h_1, \ldots, h_m)$ and $(e_1^*, \ldots, e_n$ be the dual basis of $(e_1, \ldots, e_n)$. By definition,

$$T^*(h_i^*)(e_j) = h_i^*(T(e_j)) = h_i^*(a_{1j}h_1 + \ldots + a_{mj}h_m) = a_{1j}h_i^*(h_1) + \ldots + a_{mj}h_i^*(h_m) = a_{ij}.$$

Comparing this with

$$(a_{i1}e_1^* + \ldots + a_{in}e_n^*)(e_j) = a_{i1}e_1^*(e_j) + \ldots + a_{in}e_n^*(e_j) = a_{ij},$$

we find that

$$T^*(h_i) = a_{i1}e_1^* + \ldots + a_{in}e_n^*, \quad i = 1, \ldots, m.$$

This shows that the matrix of $T^* : A^1(W) \to A^1(V)$ with respect to the dual bases $h_1^*, \ldots, h_m^*$ and $e_1^*, \ldots, e_n^*$ is equal to the transpose A^t of the matrix A of T.

Now everything is ready to get the formula for $T^*(\psi_I)$. By Theorem 2,

$$T^*(\psi_I) = T^*(h_{i_1}^* \wedge \ldots \wedge h_{i_m}^*) = (\sum_{j=1}^n a_{i_1 j} e_j^*) \wedge \ldots \wedge (\sum_{j=1}^n a_{i_m j} e_j^*) =$$

$$\sum_{(j_1,\ldots,j_m) \in S_m} a_{i_1 j_1} \ldots a_{i_m j_m} e_{j_1}^* \wedge \ldots \wedge e_{j_m}^* =$$

$$\sum_{1 \le j_1' < \ldots < j_m' \le n} \Bigg(\sum_{(j_1,\ldots,j_m) \in S_{\{j_1',\ldots,j_m'\}}} \epsilon(j_1, \ldots, j_m) a_{i_1 j_1} \ldots a_{i_m j_m} \Bigg) e_{j_1'}^* \wedge \ldots \wedge e_{j_m'}^* =$$

$$\sum_{J \subset [1,n], \#J = k} \det(A_{I,J}) e_{j_1'}^* \wedge \ldots \wedge e_{j_m'}^* = \sum_{J \subset [1,n], \#J = k} \det(A_{I,J}) \psi_J, \tag{11}$$

where

$$A_{I,J} = \begin{pmatrix} a_{i_1 j_1} & \ldots & a_{i_1 j_1} \\ \vdots & \vdots & \vdots \\ a_{i_m j_1} & \ldots & a_{i_m j_m} \end{pmatrix}$$

obtained from the matrix A by choosing the columns with indices $i_1 < \ldots < i_m$ from I and rows with indices $j_1 < \ldots < j_m$ from J.

Example 7. Assume $m = k$, so that the matrix of T is square. Then $A^k(W)$ is spanned by one basic function $\psi_{1,\dots,n} = h_1^* \wedge \dots \wedge h_m^*$ and $A^n(V)$ is spanned by one basic function $\psi'_{1,\dots,n} = e_1^* \wedge \dots \wedge e_n^*$. Then

$$T^*(\psi_{1,\dots,n}) = \det(A)\psi'_{1,\dots,n}, \tag{12}$$

where A is the matrix of T with respect to the chosen bases in V and W.

Example 8. Let $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ with standard bases formed by the unit vectors. Any linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by the formula

$$T(\mathbf{v}) = A \cdot \mathbf{v},$$

where A is the matrix of size $(m \times n)$ whose columns are the images of the basic vectors $\mathbf{e}_1, \dots, \mathbf{e}_m$. The dual basis to the standard basis in $\mathbb{R}^n$ consists of the coordinate functions x_i since

$$x_i(\mathbf{e}_j) = \delta_{ij}.$$

To distinguish the dual bases in V and W let us denote the dual basis in W by $y_1, \dots, y_m$. Thus

$$T^*(y_i) = a_{i1}x_1 + \dots + a_{in}x_n, \quad i = 1, \dots, m.$$

This agrees with our usual presentation of a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Now

$$T^*(y_{i_1} \wedge \dots \wedge y_{i_k}) = \sum_J \det(A_{I,J}) x_{j_1} \wedge \dots \wedge x_{j_k}. \tag{13}$$

Math. 396. Notes 7: Orientation

Let V be a vector space of finite dimension n. For any two bases (or *frames*) $\mathcal{E} = (e_1, \dots, e_n)$ and $\mathcal{E}' = (e'_1, \dots, e'_n)$ in V we can define the transition matrix $C_{\mathcal{E},\mathcal{E}'} = (c_{ij})$ by writing

$$e_i = \sum_{j=1}^{n} c_{ij} e'_j, \quad i = 1, \dots, n.$$

We say that $\mathcal{E}, \mathcal{E}'$ are *equivalent* if

$$\det C_{\mathcal{E},\mathcal{E}'} > 0.$$

The following well-known lemma from Linear Algebra shows that this defines an equivalence relation on the set of frames in V:

Lemma.
(i) $C_{\mathcal{E},\mathcal{E}} = I_n$, *the identity matrix.*
(ii) $C_{\mathcal{E}',\mathcal{E}} = C^{-1}_{\mathcal{E},\mathcal{E}'}$.
(iii) $C_{\mathcal{E},\mathcal{E}''} = C_{\mathcal{E},\mathcal{E}'} C_{\mathcal{E}',\mathcal{E}''}$

We have the following

Proposition 1. *There are two equivalence classes of frames in V.*

Proof. Take any frame $\mathcal{E} = (e_1, \dots, e_n)$ and consider its equivalence class. Suppose $\mathcal{E}'$ does not belong to this equivalence class. Then, by definition, $\det C_{\mathcal{E},\mathcal{E}'} < 0$. For example we may take

$\mathcal{E}' = (e_2, e_1, \ldots, e_n)$ if $n > 1$ or $\mathcal{E}' = -e'_1$ if $n = 1$. In both cases we have $\det C_{\mathcal{E},\mathcal{E}'} = -1$. Now we have two equivalence classes. Take any frame $\mathcal{E}''$. By property (iii) from the Lemma we have

$$\det C_{\mathcal{E},\mathcal{E}''} = -\det C_{\mathcal{E}',\mathcal{E}''}.$$

So one of the determinants is positive, and hence $\mathcal{E}''$ is equivalent to either $\mathcal{E}$ or to $\mathcal{E}'$.

Definition. An *orientation on a vector space* V is a choice of one of the two equivalence classes of frames. A vector space together with a choice of its orientation is called an *oriented vector space.*

Example 1. Let $V = \mathbb{R}^n$. We can choose the orientation which contains the canonical basis $(\vec{e}_1, \ldots, \vec{e}_n)$. This orientation is called the *positive* (or *right*) orientation of $\mathbb{R}^n$. The other orientation is called the *negative* (or *left*) orientation. It contains the basis $(\vec{e}_2, \vec{e}_1, \ldots, \vec{e}_n)$, if $n > 1$, or $-\vec{e}_1$ if $n = 1$.

Example 2. Let V be a one-dimensional subspace of $\mathbb{R}^n$ spanned by a non-zero vector $\vec{v}$. Then any basis in V is equal to $\{\lambda\vec{v}\}$ for some non-zero number λ. There are two equivalence classes of frames: one consists of all bases for which $\lambda > 0$, another one consists of all bases for which $\lambda < 0$. An orientation picks up one of these sets. Of course, geometrically this corresponds to the choice of a direction of the line, or, equivalenctly, the choice of one of its two rays (half-lines).

Definition. Let $T : V \to W$ be a linear invertible map between two oriented vector spaces (necessary of the same dimension). We say that T *preserves the orientations* if for any basis $\mathcal{E} = (e_1, \ldots, e_n)$ from the orientation of V, the basis $T(\mathcal{E}) = (T(e_1), \ldots, T(e_n))$ belongs to the orientation of W.

Remark 1. Note that in this definition it is enough to check this property only for one basis $\mathcal{E}$ from the orientation of V. In fact, if $\mathcal{E} = (e'_1, \ldots, e'_n)$ is a basis equivalent to $\mathcal{E}$ then

$$e_i = \sum_{j=1}^{n} c_{ij} e'_j, \quad (c_{ij}) = C_{\mathcal{E},\mathcal{E}'},$$

implies

$$T(e_i) = \sum_{j=1}^{n} c_{ij} T(e'_j).$$

Thus

$$C_{\mathcal{E},\mathcal{E}'} = C_{T(\mathcal{E}),T(\mathcal{E}')}.$$

So, if $\det C_{\mathcal{E},\mathcal{E}'} > 0$, we get $\det C_{T(\mathcal{E}),T(\mathcal{E}')} > 0$ and $T(\mathcal{E})$ is equivalent to $T(\mathcal{E}')$.

Also notice that if $T : V \to V'$ and $T' : V' \to V''$ preserve the orientations, then $T' \circ T : V \to V''$ and $T^{-1} : V' \to V$ preserve the orientations.

Now let us define an orientation on a manifold. We would like to equip each tangent space $T_p(M)$ with an orientation which behaves "continuously" when we move from one point to another. The precise meaning of this will be explained shortly.

First let us consider the simplest case when $M = U$ is an open subset of $\mathbb{R}^d$. Then each $T_p(U)$ can be identified with $T_p(\mathbb{R}^d)$ and the latter with the space $\mathbb{R}^d$. Choose the positive orientation on $\mathbb{R}^d$ from Example 1 and use it to define an orientation on each $T_p(U)$. We shall always assume that the orientation on $T_p(U)$ is chosen in this way. We call it the *standard orientation* on $T_p(U)$.

Definition. An *orientation on a d-dimensional manifold M* is a choice of an orientation on each $T_p(M)$ satisfying the following property: For any $p \in M$ there exists a coordinate patch $\alpha : U \to M$ with $p \in \alpha(U)$ such that the map $\alpha_{*,\alpha^{-1}(p)} : T_{\alpha^{-1}(p)}(U) \to T_p(M)$ preserves the orientations.

As we shall see later not every manifold admits an orientation. The following theorem explains the difficulty.

Theorem 1. *A manifold M admits an orientation if and only if there exists a set of coordinate patches $\alpha_i : U_i \to M, i \in I$, such that*
(i) $M = \bigcup_{i\in I} \alpha_i(U_i)$;
(ii) for each $i, j \in I$ with $\alpha_i(U_i) \cap \alpha_j(U_j) \neq \emptyset$,

$$\det D\psi_{\alpha_i\alpha_j} > 0.$$

We express this condition by saying that α_i and α_j overlap positively. Here for any two coordinates patches $\alpha : U \to M, \beta : U' \to M$ with $V = \alpha(U) \cap \beta(U') \neq \emptyset$ (overlapping patches) we define $\psi_{\alpha,\beta}$ to be the diffeomorphism of open subsets of $\mathbb{R}^d$

$$\psi_{\alpha,\beta} = \beta \circ \alpha^{-1} : \alpha^{-1}(V) \to \beta^{-1}(V).$$

Proof. Suppose we have an orientation on M. By definition, we can find for each $p \in M$ a coordinate patch $\alpha : U \to M$ such that $\alpha_{*,q} : T_q(U) \to T_{\alpha(q)}(M)$ preserves the orientations. Letting p run through M we find a set of coordinate patches α_i satisfying property (i) of the Theorem. In fact, if M is compact, we can find finitely many such coordinate patches. Now for any two overlapping coordinate patches α_i and α_j and $p \in \alpha_i(U_i) \cap \alpha_j(U_j)$ we have

$$\alpha_{*,\alpha_i^{-1}(p)} : T_{\alpha_i^{-1}(p)}(U_i) \to T_p(M)$$

and

$$\alpha_{*,\alpha_j^{-1}(p)} : T_{\alpha_j^{-1}(p)}(U_j) \to T_p(M)$$

both preserve the orientations. By Remark 1,

$$(\psi_{\alpha_i,\alpha_j})_{*,\alpha_i^{-1}(p)} = \alpha^{-1}_{*,\alpha_j^{-1}(p)} \circ \alpha_{*,\alpha_i^{-1}(p)} : T_{\alpha_i^{-1}(p)}(U_i) \to T_p(M) \to T_{\alpha_j^{-1}(p)}(U_j)$$

preserves the orientations. But then the determinant of the derivative matrix $D(\psi_{\alpha_i,\alpha_j})$ must be positive at $\alpha_i^{-1}(p)$. This checks property (ii).

Conversely, assume we have found a set of coordinate patches satisfying properties (i) and (ii) from Theorem 1. Then we put an orientation on each $T_p(M)$ in the following way. Find a coordinate patch $\alpha_i : U_i \to M$ with p in the image. Put the orientation on $T_p(M)$ by choosing the equivalence class of the basis $\frac{\partial}{\partial u} = (\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d})$ of $T_p(M)$ defined by the columns of the derivative matrix $D(\alpha_i)$ at the point $\alpha_i^{-1}(p)$. This definition of the orientation on $T_p(M)$ does not depend on the choice of α_i. In fact, if $\alpha_j : U_j \to M$ is another coordinate patch with p in the image, it defines the orientation by picking the basis $\frac{\partial}{\partial v} = (\frac{\partial}{\partial v_1}, \ldots, \frac{\partial}{\partial v_d})$ formed by the columns of the matrix $D(\alpha_j)(\alpha_j^{-1}(p))$. By condition (ii), the determinant of the matrix $D(\psi_{\alpha_i,\alpha_j}) = (\frac{\partial v_i}{\partial u_j})$ is positive at $\alpha_i^{-1}(p)$. But clearly, this matrix is the transition matrix $C_{\frac{\partial}{\partial u},\frac{\partial}{\partial v}}$ from the basis $\frac{\partial}{\partial u}$ to the basis $\frac{\partial}{\partial v}$. Thus the two orientations coincide.

Definition. A collection of coordinate patches satisfying properties of Theorem 1 is called a *collection of orientation patches.*

Remark 2. As we see from Theorem 1, an orientation on M is equivalent to the existence of a collection of orientation patches. Suppose we have such a collection $\{\alpha_i\}_{i\in I}$. Take some coordinate patch $\alpha : U \to M$ not from this collection. Then $\alpha_{*,q} : T_q(U) \to T_{\alpha(q)}(M)$ is either preserves the orientation or not. In the former case for any α_i overlapping with α at $p = \alpha(q)$ the determinant of the matrix $D(\psi_{\alpha,\alpha_i})$ is positive at q. Assume that U is connected. Then $\det D(\psi_{\alpha,\alpha_i})$ stays positive in the whole open set $\alpha^{-1}(\alpha(U)\cap\alpha_i(U_i))$. This shows that we can add α to our collection to obtain a larger collection of orientation patches. In the latter case, we change the α to $\alpha' = \alpha \circ g$, where $g : \mathbb{R}^d \to \mathbb{R}^d$ either interchanges the first two coordinates (if $d > 1$), or multiplies the first coordinate by -1 (if $d = 1$). Then the new patch α' defines the map $\alpha'_{*,g^{-1}(q)} : T_{g^{-1}(q)}(g^{-1}(U)) \to T_p(M)$ which preserves the orientations. Now the coordinate patch α' can be added to our collection.

Remark 3. Suppose we choose an orientation on M. Now replace the orientation in each $T_p(M)$ to another one. Then this defines another orientation on M. In fact if $\alpha_i : U_i \to M$ is a collection of orientation patches for the original orientation of M, we change all of then to the patches α'_i following the recipe from the previous Reamrk. The collection $\{\alpha'_i\}$ will be a collection of orientation patches. The new orientation on M is called the *reversed* or *opposite* orientation for the given one.

Now it is time to give examples.

Example 3. Any 1-dimensional manifold M admits an orientation. In fact we can choose an orientation on M in the sence of parametrized curves. It is defined by the velocity vector $\alpha'(t) \in T_p(M)$ at each point of $p = \alpha(t) \in M$ where $\alpha : (a, b) \to M$ is a coordinate patch at p. We noticed before that the orientation on the curve defined by coordinate patches which overlap positively, i.e. with $\psi'_{\alpha,\beta}(t) > 0$ in $\alpha^{-1}(M) \cap \beta^{-1}(M)$.

Example 4. Any n-dimensional manifold M in $\mathbb{R}^n$ is orientable. In fact, it can be covered by open subsets in $\mathbb{R}^n$, hence it is open in $\mathbb{R}^n$. Recall that our manifolds are manifolds without boundary. We shall discuss the orientation on manifolds with boundary later.

Example 5. Obviously, if M is covered by one coordinate patch it is orientable. Suppose M can be covered by 2 coordinate patches α and β. Then M admits an orientation. In fact if these two patches do not overlap positively, we change α to α' as in Remark 2 and obtain a collection of orientation patches $\{\alpha', \beta\}$.

For example, let us see that the unit sphere

$$S^n = \{x = (x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : ||x||^2 = 1\}.$$

is orientable. For this we shall exhibit two coordinate patches which cover S^n. Consider the point $N = (0, \ldots, 0, 1) \in S^n$ (the *north pole*). Given a point $(u_1, \ldots, u_n) \in \mathbb{R}^n$ let us join the point $(u_1, \ldots, u_n, 0) \in \mathbb{R}^{n+1}$ and N with a line. Its parametric equation is

$$(x_1, \ldots, x_{n+1}) = (tu_1, \ldots, tu_n, 1 - t).$$

So when $t = 0$ we get the point N and when $t = 1$ we get the point $(u_1, \ldots, u_n, 0)$. This line intersects the sphere at a unique point besides N. To get this point we find the value of the parameter t for which $||(tu_1, \ldots, tu_n, 1 - t)|| = 1$. It is easy, we have

$$||(tu_1, \ldots, tu_n, 1 - t)||^2 = t^2(u_1^2 + \ldots + u_n^2) + 1 - 2t + t^2 = 1 \Leftrightarrow t = \frac{2}{u_1^2 + \ldots + u_n^2 + 1}.$$

Notice that we discarded the solution $t = 0$ because it corresesponds to the point N. Now this value of t gives us the intersection point

$$(x_1, \ldots, x_{n+1}) = (\frac{2u_1}{u_1^2 + \ldots + u_n^2 + 1}, \ldots, \frac{2u_n}{u_1^2 + \ldots + u_n^2 + 1}, 1 - \frac{2}{u_1^2 + \ldots + u_n^2 + 1}).$$

This formula defines a smooth coordinate patch

$$\alpha_1 : \mathbb{R}^n \to S^n$$

Its image is the set $S^n \setminus \{N\}$. In fact, for any $a = (a_1, \ldots, a_{n+1}) \in S^n \setminus \{N\}$ we join N and a with the line

$$(x_1, \ldots, x_{n+1}) = (ta_1, \ldots, ta_n, t(a_{n+1} - 1) + 1).$$

This line intersects the subspace $\{x_{n+1} = 0\}$ at the unique point

$$(u_1, \ldots, u_n, 0) = (\frac{a_1}{1 - a_{n+1}}, \ldots, \frac{a_n}{1 - a_{n+1}}, 0).$$

It is immediately checked that this defines the inverse of the map $\alpha_1 : \mathbb{R}^d \to S^n \setminus \{N\}$. Note that in the denominator $a_{n+1} \neq 1$ because we take a different from N. So we can view S^n as a compactification of $\mathbb{R}^n$ by adding one point at infinity.

Now to cover the whole S^n we consider another coordinate patch α_2 doing similar construction replacing the north pole N with the south pole $S = (0, \ldots, 0, -1)$. We give the corresponding formula for $\alpha_2 : \mathbb{R}^n \to S^n \setminus \{S\}$:

$$(x_1, \ldots, x_{n+1}) = (\frac{2v_1}{v_1^2 + \ldots + v_n^2 + 1}, \ldots, \frac{2v_n}{v_1^2 + \ldots + v_n^2 + 1}, -1 + \frac{2}{v_1^2 + \ldots + v_n^2 + 1}).$$

Notice that under this map the north pole is equal to the image of the origin in $\mathbb{R}^n$.

For the future use let us give the formula for the map

$$\psi_{\alpha_1,\alpha_2} : \mathbb{R}^n \setminus \{0\} = \alpha_1^{-1}(S^n) \to \mathbb{R}^n \setminus \{0\} = \alpha_2^{-1}(S^n).$$

Comparing the last coordinates we find

$$||u||^2||v||^2 = (u_1^2 + \ldots + u_n^2)(v_1^2 + \ldots + v_n^2) = 1.$$

Comparing the first n coordinates we find

$$\frac{v_i}{||v||^2 + 1} = \frac{u_i}{||u||^2 + 1} = \frac{u_i}{1 + 1/||v||^2} = \frac{u_i||v||^2}{||v||^2 + 1}.$$

This gives

$$v_i = u_i||v||^2 = \frac{u_i}{||u||^2} \quad i = 1, \ldots, n$$

One can compute the derivative matrix for this diffeomorphism:

$$\begin{pmatrix} \frac{\partial v_1}{\partial u_1} & \ldots & \frac{\partial v_1}{\partial u_d} \\ \frac{\partial v_2}{\partial u_1} & \ldots & \frac{\partial v_2}{\partial u_d} \\ \vdots & \vdots & \vdots \\ \frac{\partial v_n}{\partial u_1} & \ldots & \frac{\partial v_n}{\partial u_d} \end{pmatrix} = \frac{1}{||u||^4} \begin{pmatrix} ||u||^2 - 2u_1^2 & -2u_1u_2 & -2u_1u_3 & \ldots & -2u_1u_n \\ -2u_2u_1 & ||u||^2 - 2u_2^2 & -2u_1u_3 & \ldots & -2u_2u_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -2u_nu_1 & -2u_nu_2 & \ldots & -u_nu_{n-1} & ||u||^2 - 2u_n^2 \end{pmatrix}.$$

We can write this matrix in the form

$$-2||u||^{-4}(A - \lambda I_n),$$

where

$$A = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \cdot \begin{pmatrix} u_1 & \dots & u_n \end{pmatrix} = \begin{pmatrix} u_1^2 & u_1u_2 & \dots & u_1u_n \\ u_2u_1 & u_2^2 & \dots & u_2u_n \\ \vdots & \vdots & \vdots\vdots & \\ u_nu_1 & u_nu_2 & \dots & u_n^2 \end{pmatrix}, \quad \lambda = ||u||^2/2.$$

Its determinant is equal to $(-2)^n$ times the value of the characteristic polynomial of A at λ. The matrix A has rank 1 and its characteristic polynomial is equal to

$$\det(A - tI_n) = (-1)^n(t^n - \text{Trace}(A)t^{n-1}) = (-1)^n(t^n - ||u||^2t^{n-1}).$$

When we plug in $t = \lambda$ we get

$$\det((\frac{\partial v_i}{\partial u_j})) = (-2)^n||u||^{-4n}(-1)^n(\frac{||u||^{2n}}{2^n} - \frac{||u||^{2n}}{2^{n-1}}) = -\frac{1}{||u||^{2n}}.$$

So the two coordinate patches overlap negatively.

Example 6. Suppose $M \subset \mathbb{R}^n$ is a manifold of dimension $n-1$ (a *hypersurface*). In this case we can define M (locally) by one equation. Given a coordinate patch $\alpha : U \to M$ about a point p it defines a basis $\frac{\partial}{\partial u_1}, \dots, \frac{\partial}{\partial u_{n-1}}$ in $T_p(M)$. We can form the vector product

$$\frac{\partial}{\partial u_1} \times \dots \times \frac{\partial}{\partial u_{n-1}}$$

Recall that for any $n-1$ vectors $\vec{v}_1, \dots, \vec{v}_{n-1}$ in $\mathbb{R}^n$ the vector product

$$\vec{v} = \vec{v}_1 \times \dots \times \vec{v}_{n-1}$$

is determined uniquely by the properties:
(i) $\vec{v} \cdot \vec{v}_i = 0, i = 1, \dots, n-1$;
(ii) $||\vec{v}|| = Vol(\vec{v}_1, \dots, \vec{v}_{n-1}) = \det((\vec{v}_i \cdot \vec{v}_j))^{1/2}$;
(iii) $(\vec{v}, \vec{v}_1, \dots, \vec{v}_{n-1})$ is basis of $\mathbb{R}^n$ belonging to the positive orientation.

Recall also that $\vec{v}_1 \times \dots \times \vec{v}_{n-1}$ is defined by considering the matrix A with columns equal to the vectors $\vec{v}_j$ and setting

$$\vec{v}_1 \times \dots \times \vec{v}_{n-1} = (\det A_1, -\det A_2, \dots, (-1)^{1+n}\det A_n),$$

where A_i is the submatrix of A obtained by deleting the i-th row. Now if

$$\vec{w}_i = \sum_{j=1}^{n-1} c_{ij}\vec{v}_j, \quad i = 1, \dots, n-1,$$

then the matrix A is replaced with the matrix $A \cdot C$, and we obtain

$$\vec{w}_1 \times \dots \times \vec{w}_{n-1} = \det(C)\vec{v}_1 \times \dots \times \vec{v}_{n-1}.$$

Since we also know that

$$Vol(\vec{w}_1, \ldots, \vec{w}_{n-1}) = |\det C| Vol(\vec{v}_1, \ldots, \vec{v}_{n-1})$$

we obtain

$$\frac{\vec{w}_1 \times \ldots \times \vec{w}_{n-1}}{||\vec{w}_1 \times \ldots \times \vec{w}_{n-1}||} = (\det C/|\det C|)\frac{\vec{v}_1 \times \ldots \times \vec{v}_{n-1}}{||\vec{v}_1 \times \ldots \times \vec{v}_{n-1}||}.$$

In other words, the normalized vector products differ by ± 1. Suppose M is orientable. Then at each point $p \in M$ we can choose a basis $(\vec{v}_1, \ldots, \vec{v}_{n-1})$ in $T_p(M)$ from the orientation of $T_p(M)$. The vector

$$\mathbf{n}(p) = \frac{\vec{v}_1 \times \ldots \times \vec{v}_{n-1}}{||\vec{v}_1 \times \ldots \times \vec{v}_{n-1}||}$$

depends only on the orientation of $T_p(M)$ and defines a function

$$\mathbf{n} : M \to S^{n-1} \subset \mathbb{R}^n$$

which satisfies

$$\mathbf{n}(p) \cdot \vec{v} = 0, \quad \forall (p, \vec{v}) \in T_p(M).$$

For any coordinate patch $\alpha : U \to M$ from a collection of orientation patches, the composition $\mathbf{n} \circ \alpha$ is given by taking the normalized vector product of the columns of the derivative matrix $D\alpha$ and hence belongs to the class $C^k(U)$ if M is of class C^k. In particular, $\mathbf{n}$ is a continuous function on M. Conversely, if there exists a function $f : M \to S^{n-1}$ of the same class as one of the manifold M such that for any $(p, \vec{v}) \in T_p(M)$

$$f(p) \cdot \vec{v} = 0$$

then M is orientable and $f = \mathbf{n}$. We just choose an orientation on $T_p(M)$ by requiring that its basis $\vec{v}_1, \ldots, \vec{v}_{n-1}$ makes the basis

$$(f(p), \vec{v}_1, \ldots, \vec{v}_{n-1})$$

of $\mathbb{R}^n$ a positively oriented.

Example 7. Now let us give an example of a non-orientable manifold. Consider the map $f : S^2 \to \mathbb{R}^4$ defined by the formula

$$f(x_1, x_2, x_3) = (x_1^2 - x_2^2, x_1x_2, x_1x_3, x_2x_3).$$

We take M to be the image of this map. Notice that

$$f(\vec{x}) = f(\vec{y}) \Leftrightarrow \vec{y} = \pm\vec{x}.$$

In fact, comparing the last three coordinates and multiplying them we find

$$(x_1x_2x_3)^2 = (y_1y_2y_3)^2.$$

If $x_1x_2x_3 = y_1y_2y_3 \neq 0$, we get $x_i = y_i$ (by comparing the last three coordinates) so $\vec{x} = \vec{y}$. If $x_1x_2x_3 = -y_1y_2y_3 \neq 0$ we get $x_i = -y_i$ so that $\vec{y} = -\vec{x}$. Assume finally that $x_1x_2x_3 = y_1y_2y_3 = 0$. If $x_3 = 0$, we get $f(x_1, x_2, x_3) = (x_1^2 - x_2^2, x_1x_2, 0, 0) = (y_1^2 - y_2^2, y_1y_2, 0, 0)$. If $x_1x_2 \neq 0$, this gives $x_1x_2 = y_1y_2, x_1^2 - x_2^2 = y_1^2 - y_2^2$. From this it is easy to deduce that $(x_1, x_2) = (y_1, y_2)$ and hence $\vec{x} = \vec{y}$. If $x_1x_2 = y_1y_2 = 0$, we get $x_1^2 - x_2^2 = y_1^2 - y_2^2$. If $x_1 = 0$, then $-x_2^2 = y_1^2 - y_2^2$ implies $y_1 = 0$ and again $\vec{x} = (0, x_2, 0) = \pm(0, y_2, 0)$. Similarly we consider the case $x_2 = 0$. Now, if $x_3 \neq 0$, and

$x_1 = 0$, we get $(-x_2^2, 0, 0, x_2x_3) = (y_1^2 - y_2^2, 0, 0, y_2y_3)$. Again it is easy to see that y_1 must be equal to 0 and $\vec{x} = \pm\vec{y}$. Similarly we consider the case $x_3 \neq 0, x_2 = 0$.

Thus we see that the image M of the map f is bijective to the set of pairs of *antipodal points* on the sphere (i.e. points obtained one from another by multiplication by -1). It is also can be thought as the set of lines in $\mathbb{R}^3$ passing through the origin (each line cuts out a unique pair of antipodal points on the sphere). Clearly the map f is injectve on each half-sphere

$$S_i^2 = \{(x_1, x_2, x_3) \in S^2 : x_i > 0\}, \quad i = 1, 2, 3.$$

Also

$$M = f(S_1) \cup f(S_2) \cup f(S_3)$$

since each pair of antipodal points has a representave in one of the half-spheres S_i. Consider the stereographic projection from the south pole $(0, 0, -1)$

$$\alpha_3 : \mathbb{R}^2 \to S^2, \quad (v_1, v_2) \to (\frac{2v_1}{v_1^2 + v_2^2 + 1}, \frac{2v_2}{v_1^2 + v_2^2 + 1}, -1 + \frac{2}{v_1^2 + v_2^2 + 1}).$$

The pre-image of the hemi-sphere S_3 is equal to the open subset

$$U_3 = \{(v_1, v_2) \in \mathbb{R}^2 : -1 + \frac{2}{v_1^2 + v_2^2 + 1} > 0\} = \{(v_1, v_2) \in \mathbb{R}^2 : v_1^2 + v_2^2 < 1\}.$$

Thus α_3 defines a bijective map from U_3 to S_3. Combining it with the map f we get a coordinate patch

$$\beta_3 = f \circ \alpha_3 : U_3 \to f(S_3) \subset M.$$

Similarly we find the coordinate patches

$$\beta_i : U_i \to f(S_i) \subset M, \quad i = 1, 2.$$

For this we use the stereographic projection from the points $(0, -1, 0)$ (for $i = 2$) or from $(-1, 0, 0)$ (for $i = 1$). We get

$$\alpha_2 : (v_1, v_2) \to (\frac{2v_1}{v_1^2 + v_2^2 + 1}, -1 + \frac{2}{v_1^2 + v_2^2 + 1}, \frac{2v_2}{v_1^2 + v_2^2 + 1}),$$

$$\alpha_1 : (v_1, v_2) \to (-1 + \frac{2}{v_1^2 + v_2^2 + 1}, \frac{2v_1}{v_1^2 + v_2^2 + 1}, \frac{2v_2}{v_1^2 + v_2^2 + 1}).$$

Now it is easy to see that the maps α_i differ by composition with the coordinates switch. So the three coordinate patches β_i of M overlap negatively. Suppose M is orientable with some collection of orientation patches $\mathcal{P}$. As is explained in Remark 2, I can add each β_i to $\mathcal{P}$ one by one, replacing each β_i with β_i' if needed, to include them in a larger collection of orientation patches. Then the modified patches β_i must overlap positively. But this is impossible since changing one β_i to β_1', say β_1, will leave β_2 and β_3 overlap negatively. Changing two β_i's to β_i', say β_1, β_2, will leave β_1, β_2 overlap negatively. Thus we have shown that M is not orientable. It is a compact smooth 2-dimensional manifold (without boundary). It is called the *real projective plane* and is denoted by $\mathbb{RP}^2$ or $\mathbb{P}^2(\mathbb{R})$.

Now assume that M is a manifold with non-empty boundary N. We define the orientation on M in the similar way. First we do it in the case when M is an open subset U of $\mathbb{H}^d = \{(u_1, \ldots, u_d) \in$

$\mathbb{R}^d : u_d \geq 0\}$. We again identify $T_p(U)$ with $\mathbb{R}^d$ and take the positive orientation. Now we define an orientation on any d-dimensional M as a choice of an orientation on each $T_p(M)$ such that for some coordinate patch $\alpha : U \to M$ with $p \in \alpha(U)$ the map $\alpha_{*,\alpha^{-1}(p)} : T_{\alpha^{-1}(p)}(U) \to T_p(M)$ preserves the orientation. I leave it to you to state the analog of Theorem 1 in the case of manifolds with boundary. Note that an orientation on M induces an orientation on the boundary N. In fact the standard orientation of $\mathbb{H}^d$ defines the canonical basis $(\vec{e}_1, \ldots, \vec{e}_{d-1})$ in $\{(u_1, \ldots, u_d) \in \mathbb{H}^d : u_d = 0\} = \mathbb{R}^{d-1}$. The image of this basis under $\alpha_{*,\alpha^{-1}(p)}$ will define an orientation in $T_p(N) \subset T_p(M)$. Here we assume that $p \in N$. If we restrict the collection of orientation patches $\alpha_i : U_i \to M$ to the subsets $U_i \cap \{(u_1, \ldots, u_d) \in \mathbb{H}^d : u_d = 0\}$ and identify the latter with an open subset of $\mathbb{R}^{d-1}$ (by forgetting about the last coordinate u_d), we get a collection of coordinate patches of N.

In particular, we obtain that the boundary of an orientable manifold with a boundary is orientable. For example, this gives another proof that spheres S^n are orientable (being the boundary of the ball). This also proves that the real projective plane cannot be diffeomorphic to the boundary of any orientable manifold of dimension 3.

Finally let us define the volume form on M.

Theorem 2. *Let M be an orientable manifold of dimension d. There exists a unique differential d-form ω_{vol} on M such that for any point $p \in M$ and a basis $\vec{v}_1, \ldots, \vec{v}_d$ of $T_p(M)$ belonging to the orientation of M one has*

$$\omega_{\text{vol}}(p, \vec{v}_1, \ldots, \vec{v}_d) = Vol(\vec{v}_1, \ldots, \vec{v}_d).$$

Proof. The uniqueness is obvious, since two such forms have the same value on each set of linearly independent vectors and equal to zero on any set of d linearly dependent vectors. To construct ω_{vol} we choose a collection $\alpha_i : U_i \to M$ of orientation patches on M and define ω_{vol} on $\alpha_i(U_i)$ by

$$\omega_{\text{vol}}(p) = Vol(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d}) du_1 \wedge \ldots \wedge du_d$$

where $du_1, \ldots, du_d$ is the dual basis to the basis $\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d}$ in $T_p(M)$ defined by the patch α_i. This definition does not depend on the choice of a patch about the point p. In fact, a different patch defines a different basis $\frac{\partial}{\partial v_1}, \ldots, \frac{\partial}{\partial v_d}$ with the transition matrix $C = (\frac{\partial u_i}{\partial v_j})$ whose determinant is positive. Then

$$Vol(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d}) du_1 \wedge \ldots \wedge du_d = Vol(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d}) \det(C) dv_1 \wedge \ldots \wedge dv_d =$$

$$Vol(\frac{\partial}{\partial v_1}, \ldots, \frac{\partial}{\partial v_d}) dv_1 \wedge \ldots \wedge dv_d.$$

Now if we have any basis $\vec{v}_1, \ldots, \vec{v}_d$ from the orientation of $T_p(M)$, we write

$$\vec{v}_i = \sum_{j=1}^{n} c_{ij} \frac{\partial}{\partial u_j}, i = 1, \ldots, d$$

and obtain

$$\omega_{\text{vol}}(p, \vec{v}_1, \ldots, \vec{v}_d) = Vol(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d}) du_1 \wedge \ldots \wedge du_d(\vec{v}_1, \ldots, \vec{v}_d) =$$

$$Vol(\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_d}) \det(c_{ij}) = Vol(\vec{v}_1, \ldots, \vec{v}_d).$$

Example 8. If M is an open subset of $\mathbb{R}^n$ with its standard orientation, then

$$\omega_{\text{vol}} = dx_1 \wedge \ldots \wedge dx_n.$$

Indeed the value of $dx_1 \wedge \ldots \wedge dx_n$ on any positively oriented set of n vectors is equal to the determinant of the matrix of its coordinates. Since it is positive it is equal to the volume formed by these vectors.

Example 9. Let $d = 1$. Let $\alpha : U \to M \subset \mathbb{R}^n$ be a coordinate patch. Then the image of $(t, 1) \in T_t(U)$ under $\alpha_{*,t}$ is the velocity vector $\alpha'(t)$ so in $\alpha(U)$ the volume form is given by

$$\omega_{\text{vol}} = ||\alpha'(t)||dt.$$

Example 10. Let M be an orientable hypersurface in $\mathbb{R}^n$ and $\mathbf{n} = (\mathbf{n}_1, \ldots, \mathbf{n}_n) : M \to S^{n-1}$ be the unit normal vector function from Example 6. Then ω_{vol} is equal to the restriction of the following form on M:

$$\Omega = \mathbf{n}_1 dx_2 \wedge \ldots \wedge dx_n - \mathbf{n}_2 dx_1 \wedge dx_3 \ldots \wedge dx_n + \ldots + (-1)^{n-1}\mathbf{n}_n dx_1 \wedge dx_2 \ldots \wedge dx_{n-1}.$$

In fact, it is easy to see using Example 6 that

$$\Omega(p, \vec{v}_1, \ldots, \vec{v}_n) = \mathbf{n} \cdot (\vec{v}_1 \times \ldots \times \vec{v}_{n-1}) = \big(\vec{v}_1 \times \ldots \times \vec{v}_{n-1}/||(\vec{v}_1 \times \ldots \times \vec{v}_{n-1})||\big) \cdot (\vec{v}_1 \times \ldots \times \vec{v}_{n-1}) =$$

$$||(\vec{v}_1 \times \ldots \times \vec{v}_{n-1})|| = Vol(\vec{v}_1, \ldots, \vec{v}_{n-1}).$$

For example, when M is given by an equation $F(x_1, \ldots, x_n) = 0$ we get

$$\mathbf{n} = \pm \text{grad } F/||\text{grad } F||$$

and

$$\Omega \wedge dF = \pm||\text{grad } F||dx_1 \wedge \ldots \wedge dx_n.$$