Introduction to Differential Geometry Notes: Lecture 1

Mason Price

February 7, 2023

1 Euclidean Space

Definition: Euclidean space is $\mathbb{R}^n$, together with the **inner product**. For $x, y \in \mathbb{R}^n$,

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i \tag{1}$$

Properties: For all $x, y, z \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$:

1. $\langle x, y\rangle = \langle y, x\rangle$ (symmetric)
2. $\langle \alpha x + y, z\rangle = \alpha\langle x, z\rangle + \langle y, z\rangle$ (linearity, or biliearity, because it's true for all variables).
3. $\langle x, x\rangle \geq 0$ (positive definite)

Definition: The length/**norm** of $x \in \mathbb{R}^n$ is

$$\|x\| = \langle x, x\rangle^{1/2} \tag{2}$$

Remark: there are other "norms" available.
Notation: $(e_i), 1 \leq i \leq n$ is the standard basis.

1.1 Cauchy-Schwartz

$$|\langle x, y\rangle| \le \|x\| \cdot \|y\| \ \ \forall x, y \in \mathbb{R}^n \tag{3}$$

(intuitively, this comes from the fact that $\langle x, y\rangle = \|x\|\|y\|\cos\theta$).

Proof: Let, for $\lambda \in \mathbb{R}$,

$$\begin{aligned} f(\lambda) &:= \|\lambda x + y\|^2 \\ &= \langle \lambda x + y, \lambda x + y\rangle \\ &= |\lambda|^2\|x\|^2 + 2\lambda\langle x, y\rangle + \|y\|^2 \end{aligned}$$

(by linearity).

This is a quadratic polynomial in λ. For all λ, $f(\lambda) \ge 0$. Hence the discriminant of the quadratic is $\Delta \le 0$, (because there should either be zero or one solution).

$$\Delta = (2\langle x, y\rangle)^2 - 4\|x\|^2\|y\|^2 \le 0$$

$$\iff |\langle x, y\rangle| \le \|x\|\|y\|$$

and we have equality in Caucy-Schwartz iff

$$\Delta = 0$$

$$\iff \|\lambda x + y\|^2 = 0$$

$$\iff \lambda x + y = 0$$

$$\iff x, y \text{ are collinear. } \Box$$

i.e. “the inner product is maximal when x, y are collinear.”

1.2 Triangle Inequality

$$\|x + y\| \le \|x\| + \|y\| \tag{4}$$

Proof:

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle \\ &\le \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \text{ (by Cauchy-Schwartz)} \\ &= (\|x\| + \|y\|)^2 \\ \therefore \ \ \|x + y\| &\le \|x\| + \|y\| \end{aligned}$$

Remark: Points and vectors are not always equivalent.
Geometric Interpretation: $\vec{x} \in \mathbb{R}^n$, $\vec{x}\,' = \overrightarrow{PQ}$.

$$P = 0 = (0, 0, \ldots, 0) \in \mathbb{R}^n$$

Let R be a point such that $\overrightarrow{PR} = \vec{x} + \vec{y} \iff \overrightarrow{QR} = \vec{y}$.

[figure] Triangle inequality in PQR:

$$|PR| \leq |PQ| + |QR|$$

equality when the three vectors are collinear, because then Q is on the line PR. (i.e. the route $P \to R$ pass through Q).

2 Basic notions of metric spaces.

Definition: A **metric space** is a set X, equipped with a map $d : X \times X \to \mathbb{R}$ called the "distance" or "metric" (function), such that $\forall x, y \in X$:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(y, x) = d(x, y)$
4. Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$

A metric space is sometimes written (X, d).

Examples: $\mathbb{R}^n$ is a metric space with $d(x, y) := \|x - y\|$. Triangle inequality:

$$\begin{aligned} d(x, y) &= \|x - y\| \\ &= \|x - z + z - y\| \;\; \forall z \in \mathbb{R}^n \\ &\leq \|x - z\| + \|z - y\| \\ &= d(x, z) + d(z, y) \end{aligned}$$

Example: Extended real line. "Compactification of $\mathbb{R}$".

[figure]

We can choose to *define* "∞" = "$-\infty$", and then add ∞ to $\mathbb{R}$ with $\mathbb{R} \cup \{\infty\}$. Or, we can do them separately as in $\mathbb{R} \cup \{\infty\} \cup \{-\infty\}$.

To extend the real line, first we define

$$\begin{aligned} f : \mathbb{R} &\to I \\ x &\mapsto \frac{x}{1+|x|} \end{aligned}$$

where $I = (-1, 1)$. The function f is bijective, since

$$\begin{aligned} f^{-1} : (-1, 1) &\to \mathbb{R} \\ x &\mapsto \frac{x}{1-|x|} \end{aligned}$$

Then, let $J = [-1, 1]$, and $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$. Intuitively

$$\begin{aligned} \mathbb{R} &= (-\infty, \infty) \\ \overline{\mathbb{R}} &= [-\infty, \infty] \end{aligned}$$

We extend f to $\overline{f} : \overline{\mathbb{R}} \to J$. Of course, $\overline{f}|_{\mathbb{R}} = f$. Then,

$$\begin{aligned} \overline{f}(+\infty) &= 1 \\ \overline{f}(-\infty) &= -1 \end{aligned}$$

We define the metric on this space to be

$$d_{\overline{\mathbb{R}}}(x, y) = \left| \, \overline{f}(x) - \overline{f}(y) \, \right| \quad \forall x, y \in \overline{\mathbb{R}}$$

As sets: $\mathbb{R} \subset \overline{\mathbb{R}}$, but as metrics $(\mathbb{R}, d_{\mathbb{R}}) \not\subset (\overline{\mathbb{R}}, d_{\overline{\mathbb{R}}})$. This is in contrast to $\mathbb{R}$ and $\mathbb{R}^2$, where the distances are the same. For example, you can see that the metrics are different because

$$d_{\overline{\mathbb{R}}}(0, x) = \left| \, 0 - \frac{x}{1+|x|} \, \right| = \frac{x}{1+|x|} \text{ if } x > 0$$

and

$$\frac{x}{1+|x|} < x = d_{\mathbb{R}}(0, x)$$

So the two metric spaces are different. But the point is that $d_{\overline{\mathbb{R}}}(+\infty, x)$ is defined.

$$\begin{aligned} d_{\overline{\mathbb{R}}}(+\infty, x) &= \left| \, 1 - \frac{1}{1+|x|} \, \right| \\ &= \frac{x}{1+x} \text{ if } x > 0 \end{aligned}$$

What's happening is that the extended line is getting shrunk via f into an interval from -1 to 1.

[figure]

3 Transformations of Euclidean Space

Reminder: Euclidean space is $\mathbb{R}^n$ with structure from the inner product.

Definition: $(X, d_X), (Y, d_Y)$ are metric spaces. A map $f : X \to Y$ is an **isometry** if it is *surjective* and *distance-preserving*:

$$d_y(f(x), f(x')) = d_x(x, x') \quad \forall x, x' \in X$$

Note: an isometry is a bijection. (Exercise)
Examples:

1. A *translation* is an isometry.
2. An *orthogonal transformation*, (e.g. a rotation), is an isometry.

What is an orthogonal transformation? $\to$ A rotation is one example, but it doesn't make sense to generalize a rotation to $\mathbb{R}^n$ because the equation for a plane in $\mathbb{R}^3 : \{\vec{x} \in \mathbb{R}^3 | \langle \vec{n}_p, \vec{x} \rangle = 0\}$ extends to

$$\langle \vec{n}_p, \vec{x} \rangle = 0 \iff \sum_{i=1}^{n} (\vec{n}_p)_i x_i = 0$$

and this only defines an $n-1$ dimensional hyper-plane, and hence it's impossible to make a 2-d angle. An appropriate generalization is an **orthogonal** transformation.

Definition: An **orthogonal transformation** is a linear map $R : \mathbb{R}^n \to \mathbb{R}^n$ such that the inner product is preserved.

$$\langle R(x), R(y) \rangle = \langle x, y \rangle \ \ \forall x, y \in \mathbb{R}^n \tag{5}$$

This implies that if M is the matrix representation of R in (e_i), then

$$^tMM = I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in M_n(\mathbb{R})$$

So the columns of M form an orthonormal basis.

Definition: An **orthogonal matrix** is a matrix $M \in M_n(\mathbb{R})$ such that

$$ {}^tMM = I_n = M{}^tM \tag{6} $$

Remark: R is orthogonal iff its matrix M is orthogonal. (This is an exercise).

Proposition: A matrix A is orthogonal iff it is **norm-preserving**, i.e.

$$ \|Ax\| = \|x\| \ \ \forall x \in \mathbb{R}^n $$

Proof: Start with the fact that A is orthogonal:

$$ \begin{aligned} \langle Ax, Ax\rangle &= \langle x, x\rangle \\ &= \|x\|^2 \\ \iff \ \|Ax\|^2 &= \|x\|^2 \end{aligned} $$

It preserves the norm because it preserves the inner product. Now, to prove the other direction of the $\iff$, we suppose $\|Ax\| = \|x\| \ \ \forall x$. Then, we start with the general statement

$$ \begin{aligned} \|x+y\|^2 &= \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle \\ \implies \ 2\langle x, y\rangle &= \|x+y\|^2 - \|x\|^2 - \|y\|^2 \end{aligned} $$

And likewise

$$ \begin{aligned} 2\langle Ax, Ay\rangle &= \|A(x+y)\|^2 - \|Ax\|^2 - \|Ay\|^2 \\ &= \|x+y\|^2 - \|x\|^2 - \|y\|^2 \\ &= 2\langle x, y\rangle \\ \therefore \ \ \langle Ax, Ay\rangle &= \langle x, y\rangle \quad \square \end{aligned} $$

Introduction to Differential Geometry Notes: Lecture 2

Mason Price

February 14, 2023

1 Isometries of $\mathbb{R}^n$

Definition: $f : \mathbb{R}^n \to \mathbb{R}^m$ is an isometry if it is distance-preserving.

Examples:

1. Translations $(n = m)$
2. Orthogonal transformations.

Review: Orthogonal transformations are linear functions $R : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\langle R(x), R(y)\rangle = \langle x, y\rangle$$

If $M = M_R$ is the matrix of R in the canonical basis of $\mathbb{R}^n$, then

$$\langle R(e_i), R(e_j)\rangle = \langle e_i, e_j\rangle = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}$$

$$\implies \quad M^t M =^t MM = I_n$$

Definition: An **orthogonal matrix** is $A \in M_n(\mathbb{R})$ such that

$${}^tAA = A^tA = I_n$$

Fact:

$$A \in M_n(\mathbb{R}) \text{ is orthogonal} \iff \|Ax\| = \|x\| \;\; \forall x \in \mathbb{R}^n$$

So it is an isometry.

Corollary: A map $f : x \mapsto Ax + b$ is an isometry $\iff$ A is orthogonal. Where $A \in M_n(\mathbb{R}), b \in \mathbb{R}^n$.

Proof:

$$\begin{aligned} \|f(x) - f(y)\| &= \|(Ax + b) - (Ay + b)\| \\ &= \|A(x - y)\| \\ &= \|x - y\| \quad \text{because } A \text{ is orthogonal.} \quad \square \end{aligned}$$

Definition: Write $O_n(\mathbb{R})$ for the set of all orthogonal $n \times n$ matrices.

Observation:

$$det(M^t M) = det(M)^2 = 1 \ \ \forall M \in O_n(\mathbb{R}) \implies det(M) \in \{+1, -1\}$$

So the determinant of an orthogonal matrix is only either $+1$ or -1.

Definition: M is **orientation-preserving** if $det(M) = 1$, (or orientation reversing if $det(M) = -1$). If

1. $det(M) = 1$, M is a **rotation.**
2. $det(M) = -1$, M is a **reflection**.

Exercise: Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be an isometry. Show that there exists $A \in O_n(\mathbb{R})$ and $b \in \mathbb{R}^n$ such that $f(x) = Ax + b \ \ \forall x \in \mathbb{R}^n$.

This implies that any isometry is of the form

$$T_{x_0} \circ R$$

where R is an orthogonal transformation, $x_0 \in \mathbb{R}^n$ is a fixed vector, and $T_{x_0}(x) = x_0 + x$ is a translation.

Remark: In some sense we can say $\mathbb{R}^n \subset \mathbb{R}^{n+1}$.

$$\left(\begin{array}{c|c} A & b \\ \hline 0 \dots 0 & 1 \end{array}\right) \left(\begin{array}{c} x \\ \hline 1 \end{array}\right) = \begin{pmatrix} Ax + b \\ \vdots \\ 1 \end{pmatrix}$$

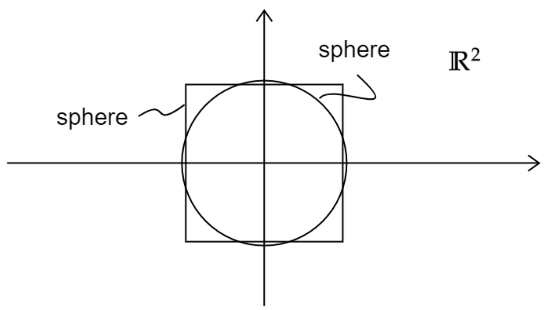
A translation is **not** linear, but we can use this to embed $\mathbb{R}^n$ in $\mathbb{R}^{n+1}$. The matrix belongs to $M_{n+1}(\mathbb{R})$.

$$\mathbb{R}^n \cong \{(x_1, \cdots, x_n, 1) \in \mathbb{R}^{n+1}\}$$

So this vector space is preserved.

2 Topology of $\mathbb{R}^n$

Let $X = \mathbb{R}^n$. (Metric space (X, d)).



"Sphere simply means the distance is the same" So it depends on the metric. Different metrics have the same "topology".

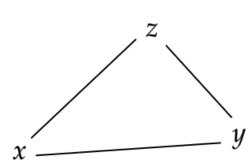
Definition: An **open ball** of center $x \in X$, with radius $\varepsilon > 0$ is

$$B(x,\ \varepsilon) := \{y \in X |\ d(x, y) < \varepsilon\}$$

(the strict '$<$' is what makes it 'open').

Exercise: Show $|d(x, z) - d(y, z)| \leq d(x, y)$. (Reverse triangle inequality).

Proof: Start with the triangle inequality:



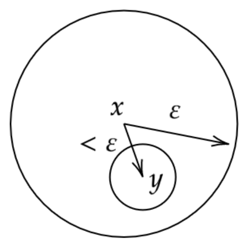
$$d(x, z) \leq d(x, y) + d(y, z)$$
$$\implies d(x, z) - d(y, z) \leq d(x, y)$$

but a metric is always positive, and $d(x,z) - d(y,z)$ may be negative, so we take its absolute value:

$$\therefore \quad |d(x,z) - d(y,z)| \leq d(x,y) \quad \square$$

Topology: A set $U \subset X$ is **open** if $\forall x \in U: \ \exists \varepsilon = \varepsilon(x) > 0$ such that $B(x,\varepsilon) \subset U$. A **closed** set is the complement of an open set.

Observation: An open ball is open because for $U = B(x,\varepsilon)$, $y \in U$, $d(x,y) < \varepsilon$, we can choose a ball with radius less than $\varepsilon - \|y - x\|$, say half of it.



Let $U = \overline{B(x,\varepsilon)} = \{y \in X \mid d(y,x) \leq \varepsilon\}$. This is a **closed ball**.

Definition: A point $x \in U$ is **interior** to U if $\exists B(x,\varepsilon) \subset U$. The interior of U, denoted $\mathring{U}$ is the set of interior points.

Note: A set U is open $\iff$ every $x \in U$ is interior $\iff$ $\forall x \in U : x \in \mathring{U} \subset U \iff U = \mathring{U}$. (i.e. an open set contains only its interior points).

Definition: $f : X \to Y$ is **continuous** at x if $\forall V \subset Y$ open, such that $f(x) \in V$, there exists an open set $U \subset X$ such that $f(U) \subset V$. ("Nearby points get mapped to nearby points".)

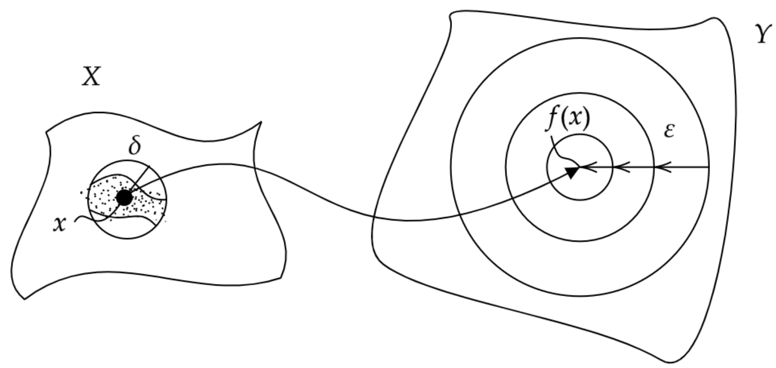
In $\mathbb{R}^n$ this becomes: $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $x \in \mathbb{R}^n$ if $\forall \varepsilon > 0 :$ $\exists \delta > 0 :$

$$\|x - y\| < \delta \implies \|f(x) - f(y)\| < \varepsilon$$

These definitions are equivalent because ε corresponds to V, (an arbitrarily small image in the range including $f(x)$), and δ corresponds to U, (the associated area in the domain including x, which gets mapped to something contained in V).

Equivalently: $\forall \varepsilon > 0 : \exists \delta > 0$ such that

$$f(U \cap B_\delta(x)) \subset B(f(x), \varepsilon)$$



Exercise: $(X, d_X), (Y, d_Y)$ are metric spaces. Show that $f : X \to Y$ is continuous if and only if $f^{-1}(U)$ is open for all $U \subset Y$ open.

Definition: $f : X \to Y$ is a **homeomorphism** if it is

1. continuous
2. bijective
3. f^{-1} is continuous

"It takes open sets to open sets".

Example: (Continuity)

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2 \end{aligned}$$

Choose $x_0 \in \mathbb{R}$. Let $\varepsilon > 0$, then we want $|f(x) - f(x_0)| < \varepsilon$:

$$\begin{aligned} |f(x) - f(x_0)| &= |x^2 - x_0^2| \\ &= |x + x_0| \cdot |x - x_0| \quad \text{(because if } x < x_0\text{, you just switch the order)} \\ &= |x - x_0 + 2x_0| \cdot |x - x_0| \\ &\leq (2 \cdot |x_0| + |x - x_0|) \cdot |x - x_0| \;\; \text{by the triangle inequality} \end{aligned}$$

Choose $|x - x_0| < \delta$, then for $x_0 \neq 0$ you may assume $\delta < |x_0|$. Then, choose $\delta < \sqrt{\varepsilon/3}$ which works because then

$$\begin{aligned}
|f(x) - f(x_0)| &< (2 \cdot |x_0| + |x - x_0|) \cdot |x - x_0| \\
&< (2 \cdot \delta + \delta) \cdot \delta \\
&< \left(2\sqrt{\frac{\varepsilon}{3}} + \sqrt{\frac{\varepsilon}{3}}\right) \cdot \sqrt{\frac{\varepsilon}{3}} \\
&= \varepsilon
\end{aligned}$$

Example: **Linear Functions**

Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. (i.e. $L(\alpha x + y) = \alpha L(x) + L(y)$). Then, each entry in the vector of the result of $L(x)$ is

$$(L(x))_i = \sum_{k=1}^{n} a_{i,k} \cdot x_k$$

This is just the dot product, where in the matrix, $i =$ row, $k =$ column, and for x a column vector, $k =$ row. Then, by the definition that a norm is the sum of squares of the entries:

$$\begin{aligned}
\|L(x)\|^2 &= \sum_{i=1}^{n} \left(\sum_{k} a_{i,k} \cdot x_k\right)^2 \\
&\leq \sum_{i=1}^{n} \left(\sum_{k} a_{i,k}^2\right) \left(\sum_{k} x_k^2\right) \\
&= \sum_{i=1}^{n} \left(\sum_{k} a_{i,k}^2\right) \cdot \|x\|^2
\end{aligned}$$

Define: $\|L\|^2 = \sum_{i,k} a_{i,k}^2$. Then, the inequality above becomes

$$\|L(x)\|^2 \leq \|L\|^2 \cdot \|x\|^2$$

Note: $\|L\|$ is well-defined and a norm (metric) by identifying

$$M_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$$

(which is norm-preserving, but we still lose some information).

We have $\|Lx\| \leq \|L\| \cdot \|x\| \;\; \forall x \in \mathbb{R}^n$. Hence,

$$\|Lx - Lx_0\| = \|L(x - x_0)\| \leq \|x - x_0\| \cdot \|L\|$$

So $\forall \varepsilon > 0$, we can choose $\delta = \frac{\varepsilon}{\|L\|}$ such that

$$\|x - x_0\| < \delta \implies \|Lx - Lx_0\| < \varepsilon$$

because $\|x - x_0\| < \delta \iff \|x - x_0\| < \frac{\varepsilon}{\|L\|} \implies \|Lx - Lx_0\| < \frac{\varepsilon}{\|L\|} \cdot \|L\| = \varepsilon$.

$\square$

Corollary: Isometries are continuous.

$$Rx - Rx_0 = A(x - x_0)$$

where A is the orthogonal part of R. This follows because the translations cancel and then the orthogonal transformations are linear.

3 Differentiation

Definition: $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at $x_0 \in \mathbb{R}^n$ if there exists a linear map $L_{\varepsilon, x_0} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} = 0$$

For example, $f : \mathbb{R} \to \mathbb{R}$ is differentiable at x_0 if there exists a linear map $L : \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{x \to x_0} \frac{|f(x) - f(x_0) - L(x - x_0)|}{x - x_0} = 0$$

Equivalently:

$$|f(x) - f(x_0) - L(x - x_0)| = |x - x_0|\varphi(x)$$

with $\varphi(x) \to 0$ as $x \to x_0$. This is true $\iff \exists \alpha \in \mathbb{R}$ such that $L : x \mapsto \alpha x$

$$|f(x) - f(x_0) - \alpha(x - x_0)| = |x - x_0|\varphi(x)$$

$$\iff \frac{f(x) - f(x_0)}{x - x_0} = \alpha + \varphi(x)$$

Since $\varphi(x) \to 0$, α is just the derivative of f.

Exercise: What functions are differentiable?

4 Appendix (not from lecture)

Question: How does the definition of an orthogonal matrix imply

$$\|Ax\| = \|x\| \ ?$$

You can show it either way:

1. Start with $\|Ax\| = \|x\|$ and then show that $A^T = A^{-1}$.
2. Start with $A^T = A^{-1}$ and then show that $\|Ax\| = \|x\|$.

Proof (1): Suppose $\|Ax\| = \|x\|$. For all $x, y \in \mathbb{R}^n$,

$$\langle Ax, Ay\rangle = \frac{1}{4}\left(\|A(x+y)^2\|^2 - \|A(x-y)^2\|^2\right)$$

because $\|A(x+y)\|^2 = \|Ax+Ay\|^2 = \langle Ax+Ay, Ax+Ay\rangle$, by the definition of a norm. Then, since $\|x\|^2 = \langle x, x\rangle$,

$$\|x+y\|^2 = \langle x+y, x+y\rangle = x^2 + 2\langle x, y\rangle + y^2$$

So

$$\|A(x+y)\|^2 = \langle Ax+Ay, Ax+Ay\rangle = \|Ax\|^2 + 2\langle Ax, Ay\rangle + \|Ay\|^2$$

and

$$\|A(x-y)\|^2 = \langle Ax-Ay, Ax-Ay\rangle = \|Ax\|^2 - 2\langle Ax, Ay\rangle + \|Ay\|^2$$

Hence,

$$\|A(x+y)\|^2 - \|A(x-y)\|^2 = 4\langle Ax, Ay\rangle$$
$$\iff \langle Ax, Ay\rangle = \frac{1}{4}\left(\|A(x+y)^2\|^2 - \|A(x-y)^2\|^2\right) \tag{1}$$

Then, since $\|Ax\| = \|x\| \ \ \forall x \in \mathbb{R}^n$,

$$\|A(x+y)\|^2 = \|x+y\|^2$$

and

$$\|A(x-y)\|^2 = \|x-y\|^2$$

So, by equation (1)

$$
\begin{aligned}
\langle Ax, Ay \rangle &= \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right) \\
&= \langle x, y \rangle
\end{aligned}
$$

$$\therefore \langle Ax, Ay \rangle = \langle x, y \rangle \tag{2}$$

Then, since $\langle Ax, Ay \rangle = \langle x, A^T Ay \rangle$, (which I will prove next), equation (2) becomes

$$\langle x, A^T Ay \rangle = \langle x, y \rangle$$

Which is true for all x, y if and only if

$$A^T A = I_n$$

$$\iff A^T = A^{-1}$$

Lemma: Proof that $\langle Ax, Ay \rangle = \langle x, A^T Ay \rangle$:
In general, $\langle x, y \rangle = x^T y$, by matrix multiplication, so

$$
\begin{aligned}
\langle Ax, Ay \rangle &= (Ax)^T Ay \\
&= x^T A^T Ay \\
&= (x)^T (A^T Ay) \\
\therefore \langle Ax, Ay \rangle &= \langle x, A^T Ay \rangle
\end{aligned}
$$

where the second step is justified by the properties of a matrix transpose, (namely, $(AB)^T = B^T A^T$).

Proof (2): Suppose that $A^T = A^{-1}$, i.e. A is orthogonal. Then,

$$
\begin{aligned}
\|Ax\| &= \sqrt{\langle Ax, Ax \rangle} \\
&= \sqrt{\langle A^T Ax, x \rangle} \quad \text{by the lemma.} \\
&= \sqrt{\langle x, x \rangle} \quad \text{because } A \text{ is orthogonal.} \\
\therefore \|Ax\| &= \|x\|
\end{aligned}
$$

$$\square$$

It's also worth noting that this immediately implies the transformation A is an isometry.

$$\|A(x-y)\| = \|x-y\|$$

Introduction to Differential Geometry Notes: Lecture 3

Mason Price

February 25, 2023

1 Differentiation in $\mathbb{R}^n$

Recall: Let $U \subset \mathbb{R}^n$ be an open set, with $f : U \to \mathbb{R}^m$ continuous. Then f is **differentiable** at $x_0 \in U$ if there exists an $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} = 0$$

where $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is the set of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. Or alternatively,

$$\|f(x) - f(x_0) - L(x - x_0)\| = o(x - x_0)$$

where o is a function such that $o(x - x_0) \to 0$ as $x \to x_0$. Recall:

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$$

if $f(x) = o(g(x))$ for a function o such that $o(g(x)) \to 0$ as $x \to x_0$. ("f goes to 0 faster than g").

Proposition 1 *The linear map $L = L_{f,x_0}$ if it exists, is unique.*

This is similar to the fact that the derivative of a function of a single variable is unique, i.e.

$$\frac{f(x) - f(x_0)}{x - x_0} \mapsto f'(x_0)$$

but instead of just a number, f' is a linear map and it is determined by a matrix.

Proof: Let $L, L' : \mathbb{R}^n \to \mathbb{R}^m$ be two such maps. Then, the difference between where they map to is

$$\begin{aligned}
\|(L - L')(x - x_0)\| &= \|(L - L')(x - x_0) + f(x) - f(x) + f(x_0) - f(x_0)\| \\
&\leq \|f(x) - f(x_0) - L(x - x_0)\| + \|f(x) - f(x_0) - L'(x - x_0)\| \\
&\qquad \text{by the triangle inequality} \\
&= o(x - x_0) + o(x - x_0) \\
\therefore \|(L - L')(x - x_0)\| &= o(x - x_0)
\end{aligned}$$

This is true in general, so it must also be true in particular for $x - x_0 = re_i$, where $|r| = \|x - x_0\|$. Then,

$$o(r) = |r| \cdot \|L - L'\| \cdot \|e_i\|$$

where $\|L - L'\|$ is the Euclidean norm when $M_{m,n}$ is identified with $\mathbb{R}^{mn}$, (i.e. $M_{m,n} \cong \mathbb{R}^{mn}$). Then,

$$o(s) = \frac{o(r)}{|r|} = \|L - L'\| \quad \text{as } x \to x_0$$

$$\implies \|L - L'\| = 0$$

$$L = L'$$

Alternatively, we can also show this using an argument based on the entries of the matrix:

$$\begin{aligned} o(r) &= |r| \cdot \|L - L'\| \cdot \|e_i\| \\ &= \left(\sum_j (a_{i,j} - a'_{i,j})^2 \right)^{1/2} |r| \quad \text{for all } i \end{aligned}$$

by the definition of the norm. Then, since $\frac{o(r)}{|r|} \to 0$ as $r \to 0$, we have that $a_{i,j} = a'_{i,j}$, for a matrix representation $(a_{i,j})$ of L. Hence we have shown that if there are two linear maps that satisfy the definition, then they must be the same map.

Definition 1 *The unique linear map L is the **differential** (or derivative) of f at x_0, denoted:*

$$df_{x_0} \quad \textit{or} \quad f'(x_0)$$

Examples:

1. Linear maps are differentiable. In particular, $dL_{x_0} = L$.

2. Isometries are differentiable, because the translations in f cancel out, leaving the orthogonal part which is a linear map, so the same argument applies. i.e. if $B : \mathbb{R}^n \to \mathbb{R}^m$ is an isometry, then $dB_{x_0} = R$, where R is the linear/orthogonal part of B.

3. All elementary functions $\mathbb{R} \to \mathbb{R}$.

4. $(x, y) \mapsto \langle x, y \rangle$ is differentiable. (Exercise)

5. $x \mapsto \|x\|^2$ is differentiable. (Exercise)

2 Higher Differentiation

Recall: $f : \mathbb{R} \to \mathbb{R}$ is differentiable if and only if for some subset $U \subset \mathbb{R}$, for all $x_0 \in U$, $f'(x_0)$ exists, i.e.

$$f' : x \mapsto f'(x)$$

Definition 2 *Let $f : U \to \mathbb{R}^m$, with $U \subset \mathbb{R}^n$ open. Then, f is* ***continuously differentiable*** *(C^1) if the map*

$$df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$
$$x \mapsto df_x$$

is ***continuous****, where $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is again the set of all linear maps.*

NOTE: df_x is a linear map, but df is **not**. It is a map into the set of all linear maps, because it does not specify the particular point x at which we are taking the derivative. It is like the collection of all linear maps that df_x can be.

Remember that we can identify $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with $\mathbb{R}^{mn}$, as in

$$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{nm}$$

by making a vector whose entries are the same as the entries of the matrix representation of any such linear map. Then we can also identify

$$\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)) \cong \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{nm}) \cong \mathbb{R}^{n^2m}$$

EXERCISE: Are the previous maps C^1?

Definition 3 *A map f is C^k, (k times* ***continuously differentiable****) if f is k times differentiable, and*

1. *f is differentiable*
2. *df is differentiable*
3. *$d(df)$ is differentiable*
4. $\vdots$
5. *$d^k f$ is continuous.*

Definition 4 *A* ***diffeomorphism*** *is a bijective, differentiable map $f : U \to V$ such that f^{-1} is differentiable.*

Theorem 1 *If f is differentiable at x_0, then f is continuous at x_0, (i.e. differentiability implies continuity). More precisely, if $\|df_{x_0}\| < c$, then $\exists \delta > 0$:*

$$\|f(x_0 + h) - f(x_0)\| \leq c\|h\| \quad \forall \; \|h\| < \delta$$

PROOF: By the definition of differentiability,

$$\|f(x_0 + h) - f(x_0) - L(x - x_0)\| = o(h)$$

Then by the triangle inequality, letting $L = df_{x_0}$,

$$\|f(x_0 + h) - f(x_0)\| \leq \|df_{x_0}(h)\| + o(h)$$

as $h \to 0$.

$$\implies \|f(x_0 + h) - f(x_0)\| \leq \|df_{x_0}\| \cdot \|h\| + \varepsilon\|h\|$$

by Cauchy-Schwartz. Then, if $\|df_{x_0}\| < c$, we can find some ε such that

$$\|f(x_0 + h) - f(x_0)\| \leq c \cdot \|h\|$$

where $o(h) = \|h\|\varphi(h)$ and $\varphi \to 0$ as $h \to 0$. So we can construct a continuity argument, where $\forall\varepsilon > 0, \exists\delta > 0$ such that

$$c - \|df_{x_0}\| \geq \varepsilon$$

3 Directional Derivative / Partial Derivative

3.1 Textbook:

" If $F : U \to \mathbb{R}^m$ is differentiable, with $U \subset \mathbb{R}^n$, then the m coordinate functions $F^j(x^1, \ldots, x^n)$, (i.e. where $1 \leq j \leq m$), have partial derivatives $\partial F^j/\partial x^i = F^j_{x^i}$, with respect to each of the n coordinates x^i. From our definition of $dF_{x_0} : \mathbb{R}^m \to \mathbb{R}^n$, it follows that the matrix of this linear map is given by the matrix of first derivatives of F at x_0,

$$(F^j_{x^i})_{x_0}$$

the familiar Jacobian matrix. "

3.2 Lecture:

For a map $f : \mathbb{R}^n \to \mathbb{R}^m$, say $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$, then

$$f^i_j : (\ \ldots,\ x,\ \ldots) \to \mathbb{R}$$

where each $(\ldots)$ is fixed, and x occurs in the i^{th} place. This is the j^{th} partial derivative of f at x_i.

Definition 5 *Let $f : U \to \mathbb{R}^m$, and $\alpha = (p_1, \ldots, p_n) \in U$, then there exists a $\delta > 0$ such that for $j = 1, \ldots, n$:*

$$\begin{aligned} f_j : (p_j - \delta, p_j + \delta) &\to \mathbb{R}^m \\ t &\mapsto (p_1, \ldots, p_{j-1},\ t,\ p_{j+1}, \ldots, p_n) \end{aligned}$$

The function $f_j(p_j) = f(\alpha)$ is well-defined. We can ask: " Is $f_j : \mathbb{R} \to \mathbb{R}^m$ differentiable at p_0? i.e. does

$$\lim_{h \to 0} \frac{f(\alpha + h \cdot e_j) - f(\alpha)}{h}$$

exist? If this exists, we say that f has a **partial** j^{th} derivative at α. The value is

$$D_j f(\alpha) = f_j'(\alpha) = d(f_j)_\alpha$$

Proposition 2 *Let $U \subset \mathbb{R}^n$ be open, and $f : U \to \mathbb{R}^m$ be differentiable at $\alpha \in U$. Then, $D_j f(\alpha)$ exists for all j, and*

$$df_\alpha(h) = \sum_{j=1}^{n} h_j D_j f(\alpha)$$

where $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ is not necessarily small, (it could be any vector). In particular, if $e_1, \ldots, e_n$ and $u_1, \ldots, u_m$ are canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, then

$$df_\alpha(e_j) = \sum_{i=1}^{m} D_j f_i(\alpha) u_i, \quad 1 \le j \le n$$

"The directional derivative of f at α (i.e. the linear map approximating f near α) applied to the basis element e_j is equal to the sum of the partial derivatives, with respect to the same j^{th} variable, over all the component functions of f at α applied to the basis elements u_i".

PROOF: Let $L = df_\alpha$, and $h \in \mathbb{R}$. Then, by the definition of df_α,

$$\begin{aligned} f(\alpha + h \cdot e_j) - f(\alpha) &= L(h \cdot e_j) + o(h) \\ &= h \cdot L(e_j) + o(h) \end{aligned}$$

Therefore,

$$\boxed{\lim_{h \to 0} \frac{f(\alpha + h \cdot e_j) - f(\alpha)}{h} = L(e_j)}$$

On the other hand, this is $D_j f(\alpha)$. Therefore, we conclude that **the partial derivative is a linear map given by the differential of a basis vector.** That is,

$$D_j f(\alpha) = df_\alpha(e_j) \quad \forall j = 1, \ldots, n$$

This is the j^{th} column of the matrix of df_α. If $h \in \mathbb{R}^n, h = (h_1, \ldots, h_n)$, then

$$h = \sum_{i=1}^{n} h_i \cdot e_i \quad \text{where } h_i \in \mathbb{R}$$

So

$$\begin{aligned} L(h) &= L\left(\sum_{i=1}^{n} h_i \cdot e_i\right) \\ &= \sum_{i=1}^{n} h_i \cdot L(e_i) \end{aligned}$$

$$\therefore \boxed{L(h) = \sum_{i=1}^{n} h_i D_i f(\alpha)}$$

□

NOTE: The matrix $M(df_\alpha) \in M_{m,n}(\mathbb{R})$. The j^{th} column is

$$\mathrm{col}_j(M(df_\alpha)) = df_\alpha(e_j) = D_j f(\alpha)$$

So, therefore the matrix representation of the differential is

$$M(df_\alpha) = \left(\frac{\partial f_i}{\partial x_j}\right)_{i,j}$$

Definition 6 *The **directional derivative** of $f : U \to \mathbb{R}^m$ at $\alpha \in U$, with respect to ξ is*

$$df_\alpha(\xi)$$

and is denoted as

$$D_\xi f(\alpha)$$

In previous classes, the directional derivative was probably defined as

$$\lim_{h\to 0} \frac{f(\alpha + h\xi) - f(\alpha)}{h}$$

and this is an equivalent definition.

NOTE: The partial derivative is the directional derivative in the direction of a basis vector e_j. Recall that

$$D_j f(\alpha) = D_{e_j} f(\alpha)$$

Then, if $f : \mathbb{R}^n \to \mathbb{R}^m$, we can see the relationship between the partial derivative and directional derivative as follows.

$$\begin{aligned} D_\xi f(\alpha) = df_\alpha(\xi) &= \sum_{j=1}^{n} \xi_j D_j f(\alpha) \\ &= \sum_{j=1}^{n} \xi_j \frac{\partial f}{\partial x_j}\bigg|_{x=\alpha} \\ &= \langle \xi, df_\alpha \rangle \end{aligned}$$

Proposition 3 CHAIN RULE: *Given $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$, with maps*

$$f : U \to V$$

$$g : V \to \mathbb{R}^k$$

Assume that f is differentiable at x, and g is differentiable at $f(x) = y$. Then $g \circ f$ is differentiable at x and

$$d(g \circ f)_x = dg_{y_0} \circ df_{x_0}$$

NOTE: This is analogous and consistent with the case where f and g are real valued functions, i.e.

$$(g \circ f)'(x) = g'(f(x)) f'(x)$$

because in the general case, we're just leaving the place where the functions are evaluated implicit.

Introduction to Differential Geometry
Notes: Lecture 4

Mason Price

1 Chain Rule

Proposition 1 *For open sets $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$, with maps $f : U \to V$, $g : V \to \mathbb{R}^k$. If f and g are differentiable at x_0 and $y_0 = f(x_0)$, respectively, then $g \circ f$ is differentiable and*

$$d(g \circ f)_{x_0} = dg_{f(x_0)} \circ df_{x_0} \tag{1}$$

PROOF: Let $T := df_{x_0}$, $S := dg_{f(x_0)}$, and define L to be the linear composition

$$L = S \circ T$$

Now, define

$$\varphi(h) := (g \circ f)(x_0 + h) - (g \circ f)(x_0) - L(h) \tag{2}$$

and our proof amounts to showing that

$$\lim_{h \to 0} \frac{\|\varphi(h)\|}{\|h\|} = 0$$

Let

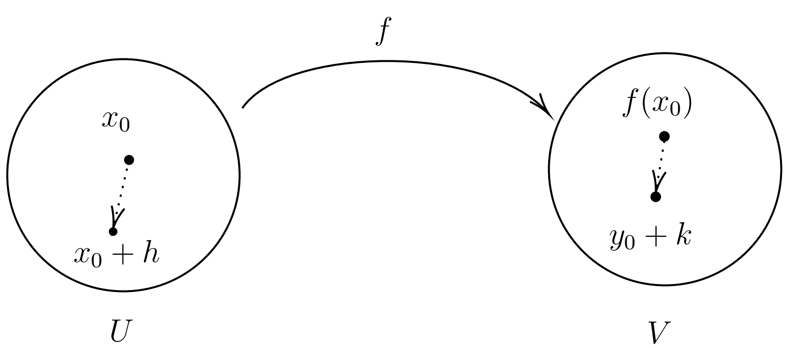
$$\varphi_f(h) := f(x_0 + h) - f(x_0) - Th \tag{3}$$

for some $h \in \mathbb{R}^n$ and $x_0 + h \in U$. Also let

$$\varphi_g(k) := g(y_0 + k) - g(y_0) - Sk \tag{4}$$

for some $k \in \mathbb{R}^n$ and $y_0 + k \in V$, where again $y_0 := f(x_0)$. Then, we know by the assumption that f and g are differentiable that

$$\lim_{h \to 0} \frac{\|\varphi_f(h)\|}{\|h\|} = 0$$
$$\lim_{k \to 0} \frac{\|\varphi_g(k)\|}{\|k\|} = 0$$



As you approach x_0 with $h \to 0$, we want also that you approach y_0 in V, so that you only need to control h. Let $k(h) := f(x_0 + h) - f(x_0)$. Then,

$$\lim_{h\to 0} k(h) = 0$$

because f is continuous. Hence, you should be able to show by an ϵ, δ argument that this implies

$$\|k(h)\| \le c\|h\|$$

if $\|h\|$ is small enough. Then,

$$\lim_{h\to 0} \frac{\|\varphi_g(k(h))\|}{\|h\|} = 0$$

Now consider the difference between where the map $(g \circ f)$ sends x_0 and $x_0 + h$.

$$\begin{aligned}(g \circ f)(x_0 + h) - (g \circ f)(x_0) &= g(f(x_0 + h)) - g(f(x_0)) \\ &= g(y_0 + k(h)) - g(y_0) \\ &= S(k(h)) + \varphi_g(k(h))\end{aligned}$$

by Equation (4). Then, substituting this into Equation (2), we get

$$\begin{aligned}\varphi(h) :&= (g \circ f)(x_0 + h) - (g \circ f)(x_0) - L(h) \\ &= S(k(h)) + \varphi_g(k(h)) - L(h) \\ &= S(\varphi_f(h)) + \varphi_g(k(h))\end{aligned}$$

because

$$\begin{aligned}S(\varphi_f) &= S(f(x_0 + h) - f(x_0) - Th) \\ &= S(f(x_0 + h) - f(x_0)) - (S \circ T)(h) \\ &= S(k(h)) - L(h)\end{aligned}$$

So, using

$$\varphi(h) = S(\varphi_f(h)) + \varphi_g(k(h)),$$

we can say that

$$\frac{\|\varphi(h)\|}{\|h\|} \le \|S\| \cdot \frac{\|\varphi_f(h)\|}{\|h\|} + \frac{\|\varphi_g(k(h))\|}{\|h\|}$$

The linear map S is bounded, and we already know

$$\lim_{h\to 0} \frac{\|\varphi_f(h)\|}{\|h\|} = 0$$

$$\lim_{k\to 0} \frac{\|\varphi_g(k)\|}{\|k\|} = 0$$

by the differentiability of f and g. Hence,

$$\lim_{h\to 0} \frac{\|\varphi(h)\|}{\|h\|} = 0$$

$\square$

EXERCISE: Given $f : U \to \mathbb{R}^n$ and $g : U \to \mathbb{R}^n$ with $x \in U$. If f, g are differentiable then the differential of

$$\begin{aligned} h \colon U \times U &\to \mathbb{R} \\ (x, y) &\mapsto \langle f(x), g(x) \rangle \end{aligned}$$

is

$$dh_{x_0}(x) = \langle f(x_0), dg_{x_0}(x) \rangle + \langle df_{x_0}(x), g(x_0) \rangle.$$

This generalizes the product rule, i.e., for

$$f, g : \mathbb{R} \to \mathbb{R}$$

$$(f \cdot g)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$$

2 Tangent Spaces

For a function $f : \mathbb{R}^n \to \mathbb{R}$, recall that the differential of f at a point x_0 is the linear map which best approximates f near x_0, that is,

$$df_{x_0} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}).$$

Then, since it is a linear map into $\mathbb{R}$, we can say that the differential applied to a vector $v \in \mathbb{R}^n$ will be a real number, i.e.,

$$df_\alpha(v) \in \mathbb{R},$$

where the differential is evaluated at a point α. Now, any linear map can be identified with a matrix, and in particular the differential df_α can be identified with

$$df_\alpha = \frac{\partial f}{\partial x_i}(\alpha).$$

Hence, applying df_α to $v = (v_1, \dots, v_n)$ gives

$$\begin{aligned} df_\alpha(v) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\alpha) \cdot v_i \\ &= \langle df_\alpha, v \rangle. \end{aligned} \tag{5}$$

The assignment, for $v \in \mathbb{R}^n$,

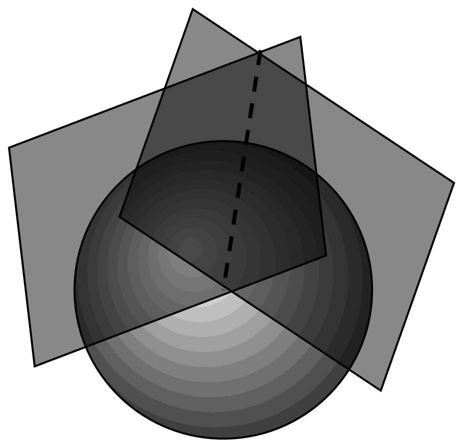
$$\begin{aligned} \partial_v :\ C^1(\mathbb{R}^n, \mathbb{R}) &\to \mathbb{R} \\ f &\mapsto \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\alpha) v_i \\ f &\mapsto D_v f(\alpha), \end{aligned}$$

where $D_v f(\alpha)$ is the directional derivative and $C^1(\mathbb{R}^n, \mathbb{R})$ is the space of continuous functions from $\mathbb{R}^n$ to $\mathbb{R}$, (i.e., the place where f lives). This represents the tangent if it satisfies the product rule.

Definition 1 *Let $M = \mathbb{R}^n$ and $x_0 \in M$. Then, the **tangent space** of M at x_0 is denoted $T_{x_0}M$ and is defined as*

$$T_{x_0}M := \{x_0\} \times \mathbb{R}^n \tag{6}$$

NOTE: We adjoin x_0 to the front just to make sure it is clear we are talking about x_0. In particular, $T_{x_0}M$ and $T_{y_0}M$ are disjoint if $x_0 \neq y_0$ even though they both include $\mathbb{R}^n$. This is the case when we're considering the *whole* tangent space, but it will also become relevant once we define, for example, tangent planes. As in the picture below, it means nothing if two tangent planes intersect in the space where they are embedded, those two tangent spaces are still disjoint.



REMARK: $T_{x_0}M$ is defined to be a vector space, so we need to define addition and scalar multiplication. We can do these simultaneously for a scalar $\lambda \in \mathbb{R}$ and two points in the tangent space $(x_0, x), (x_0, y) \in T_{x_0}M$ by defining

$$\lambda(x_0, x) + (x_0, y) := (x_0, \lambda x + y). \tag{7}$$

Definition 2 *The **tangent bundle** of $U \subset M$ is the disjoint union of all the tangent spaces,*

$$TU := \bigsqcup_{x_0 \in U} T_{x_0}M, \tag{8}$$

together with a projection

$$\begin{aligned} \pi : TU &\to U \\ (x_0, x) &\mapsto x_0 \end{aligned} \tag{9}$$

REMARK: TU can be identified with $U \times M$ by the map

$$\begin{aligned} TU &\to U \times M \\ (\{x_0\}, x) &\mapsto (x_0, x), \end{aligned}$$

because x_0 can