Updated May 20, 2024

1 Chapter I: Real Vector Spaces

Definition 1 (Function)**.** A function or map $f : X \to Y$ between the sets X and Y is a subset $f \subseteq X \times Y$ such that

1. $x \in X \implies \exists y \in Y$ such that $(x, y) \in f$
2. $(x, y); (x, y') \in f \implies y = y'$

If $(x, y) \in f$, we label y as $f(x)$.

Definition 2 (Inverse Image)**.** Given sets X and Y, and the function $f : X \to Y$, the inverse image of $T \subseteq Y$ by f is $f^{\leftarrow}(T) = \{x | f(x) \in T\}$.

Definition 3 (Real Vector Space)**.** A real vector space (or simply a vector space) is a non-empty set X and two functions, $X \times X \to X : (\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x}+\mathbf{y}$ and $X \times \mathbb{R} \to X : (\mathbf{x}, a) \mapsto \mathbf{x}a$ such that

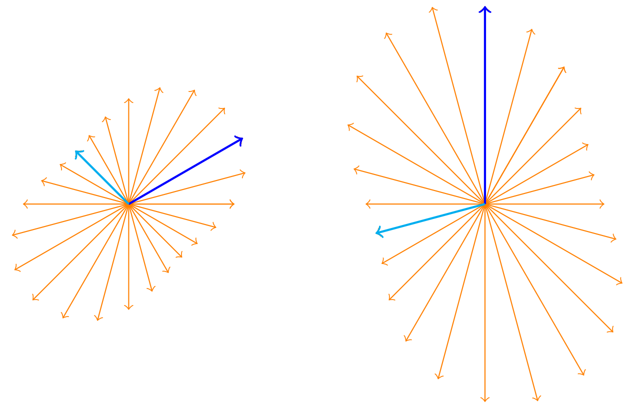
1. $+$ is commutative, associative, has an identity, admits inverse
2. $\mathbf{x}1 = \mathbf{x}$
3. $(\mathbf{x} + \mathbf{y})a = \mathbf{x}a + \mathbf{y}a, \mathbf{x}(a + b) = \mathbf{x}a + \mathbf{x}b$
4. $(\mathbf{x}a)b = \mathbf{x}(ab)$

Elements of X are called vectors.

Example 1. *Arrows in 2-d space with parallelogram addition form a real vector space.*

Definition 4 (Linear Function/Map on Vector Space)**.** A function $\mathbf{A}$ from vector space X to vector space Y is linear if for all $\mathbf{x}, \mathbf{y} \in X$ and $a \in \mathbb{R}$, $\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{A}(\mathbf{x}) + \mathbf{A}(\mathbf{y})$ and $\mathbf{A}(\mathbf{x}a) = \mathbf{A}(\mathbf{x}a)$.

Example 2. *Both X and Y consist of vectors in two-dimensional plane. Map $\mathbf{A}$ scales a vector in X by 1.5 and rotates it counterclockwise by $60°$ to get its image in Y. $\mathbf{A}$ is linear.*



$X \xrightarrow{\mathbf{A}} Y$

2 Chapter II: Affine Spaces

2.1 Spaces

Definition 5 (Affine Space)**.** An affine space with vector space T is a non-empty set X and a difference function $\mathbf{d} : X \times X \to T$ such that for any $x, y, z \in X$

1. $\mathbf{d}(x, y) + \mathbf{d}(y, z) = \mathbf{d}(x, z)$
2. The restricted map $\mathbf{d}_x : \{x\} \times X \to T : (x, y) \mapsto \mathbf{d}(x, y)$ is bijective

The affine space is represented as (X, T).

Comment. *That is, points are separated by vectors. Vectors go from one point to another in the affine space. Any two points are separated by vectors but it is not necessary from definition that given a starting point and a vector, there is an ending point such that the vector goes from the starting point to the ending point.*

Comment. *The bijective condition allows us to define $x + \mathbf{t}$ for $x \in X$, $\mathbf{t} \in T$ as $y \in X$ such that $\mathbf{d}_x(x, y) = \mathbf{t}$.*

Comment. *The first argument in the definition of $\mathbf{d}_x(x, y)$ is redundant. It is kept to indicate that $\mathbf{d}_x(x, y)$ operates on vectors starting at x and ending at points in the affine space rather than directly on points in the affine space. An alternative notation could have been $\mathbf{d}_x((x, y))$ with the understanding that (x, y) is a vector from x to y.*

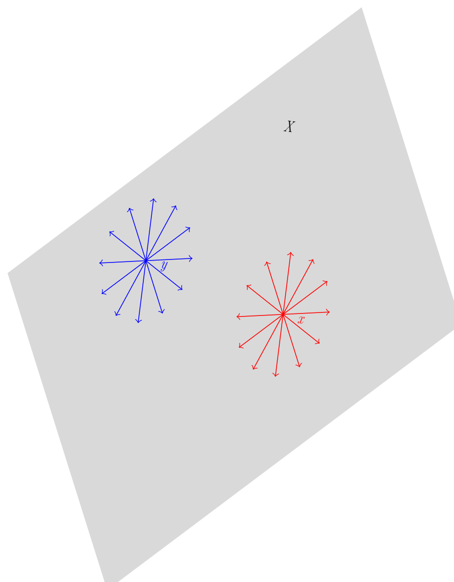
Definition 6 (Tangent Space)**.** Given an affine space X with a difference function $\mathbf{d}$ and $x \in X$, the tangent space to X at x, denoted T_xX is the vector space $\{x\} \times X$ with the operations

1. $(x, y) + (x, z) = \mathbf{d}_x^{\leftarrow}(\mathbf{d}_x(x, y) + \mathbf{d}_x(x, z))$
2. $(x, y)a = \mathbf{d}_x^{\leftarrow}((\mathbf{d}_x(x, y))a)$

Comment. *A tangent space to X at x consists of differences of all points of X from x. These differences, called tangents, are vectors from the definition of an affine space. Thus, tangents in a tangent space are vectors, except that they all start from the same point. That is, tangents are **bound vectors** but elements of the vector space underlying the affine space are **free vectors**.*

Comment. *The restricted difference function $\mathbf{d}_x$ is called a freeing map. Its inverse $\mathbf{d}_x^{\leftarrow}$ that gives corresponding bound vector starting at x is called a binding map. A distance function returns free vectors.*

Example 3. *The red vectors are part of T_xX and the blue vectors are part of T_yX.*



Definition 7 (Dimension of Affine Space)**.** The dimension of an affine space is the dimension of the space of its free vectors.

Example 4. *All points on this page form an affine space of dimension two.*

Definition 8 (Weighted Average of Points in Affine Space)**.** Given affine space (X, T), for $x, y \in X$, $(1-\lambda)x + \lambda y$ is defined as $x + \lambda\mathbf{d}(x, y)$.

Comment. *The above definition is generalized in obvious way to a weighted average of multiple points in an affine space.*

Example 5. *For the affine space of all points on this page, one-fourth of top-left corner and three-fourth of bottom right-corner is a point one-fourth distance along the diagonal from bottom right corner to top-left corner.*

2.2 Maps

Definition 9 (Affine Map)**.** If X and Y are affine spaces with corresponding vector spaces T and S, respectively, a map $A : (X,T) \to (Y,S)$ is affine if for any $x, x' \in X, \lambda \in \mathbb{R}$,

$$A((1-\lambda)x + \lambda x') = (1-\lambda)A(x) + \lambda A x'.$$

Result 1. *An affine map $A : (X,T) \to (Y,S)$ induces a linear map, called the linear part of A, $\mathbf{A} : T \to S$ such that points separated by $\mathbf{t} \in T$ in X are mapped to points separated by $\mathbf{At} \in S$ in Y.*

Proof. $\mathbf{A}$ is well-defined: Suppose $x, y, x', y' \in X$ such that $\mathbf{d}(x,y) = \mathbf{d}(x',y')$. Then,

$$\begin{aligned}
&\frac{1}{2}x + \frac{1}{2}y' = x + \frac{1}{2}\mathbf{d}(x,y') = x + \frac{1}{2}(\mathbf{d}(x,y) + \mathbf{d}(y,x') + \mathbf{d}(x',y')) \\
&= x + \frac{1}{2}(\mathbf{d}(x,y) + 2\mathbf{d}(y,x') + \mathbf{d}(x',y) + \mathbf{d}(x,y)) \\
&= x + \mathbf{d}(x,y) + \mathbf{d}(y,x') + \frac{1}{2}\mathbf{d}(x',y) = x' + \frac{1}{2}\mathbf{d}(x',y) = \frac{1}{2}x' + \frac{1}{2}y.
\end{aligned}$$

Then, by Definition 9, $\frac{1}{2}A(x) + \frac{1}{2}A(y') = \frac{1}{2}A(x') + \frac{1}{2}A(y)$. That is, $A(x) + \frac{1}{2}\mathbf{d}(A(x), A(y')) = A(x') + \frac{1}{2}\mathbf{d}(A(x'), A(y))$. Substituting

$\mathbf{d}(A(x), A(y')) =$
$\mathbf{d}(A(x), A(y)) + \mathbf{d}(A(y), A(x')) + \mathbf{d}(A(x'), A(y')) =$
$2\mathbf{d}(A(x), A(y)) + 2\mathbf{d}(A(y), A(x')) + \mathbf{d}(A(x'), A(y)) + \mathbf{d}(A(x'), A(y')) - \mathbf{d}(A(x), A(y)),$

we get

$$\begin{aligned}
&\frac{1}{2}A(x') + \frac{1}{2}\mathbf{d}(A(x'), A(y)) + \frac{1}{2}(\mathbf{d}(A(x'), A(y')) - \mathbf{d}(A(x), A(y))) \\
&= \frac{1}{2}A(x') + \frac{1}{2}\mathbf{d}(A(x'), A(y)) \\
&\implies \mathbf{d}(A(x'), A(y')) = \mathbf{d}(A(x), A(y)).
\end{aligned}$$

$\mathbf{A}$ is a linear map: Let $x, y, z \in X$ such that $\mathbf{d}(x,y) = \mathbf{s}, \mathbf{d}(y,z) = \mathbf{t}$ and $\lambda \in \mathbb{R}$. Then, $\mathbf{A}(\mathbf{s}) + \mathbf{A}(\mathbf{t}) = \mathbf{d}(A(x), A(y)) + \mathbf{d}(A(y), A(z)) = \mathbf{d}(A(x), A(z))$ (by Definition 5) $= \mathbf{A}(\mathbf{s} + \mathbf{t})$.
And $\mathbf{A}(\lambda\mathbf{s}) = \mathbf{d}(A(x), A(x + \lambda\mathbf{d}(x,y))) = \mathbf{d}(A(x), A((1-\lambda)x + \lambda y))$ (by Definition 8) $= \mathbf{d}(A(x), (1-\lambda)A(x) + \lambda A(y))$ (by Definition 9) $= \mathbf{d}(A(x), A(x) + \lambda\mathbf{d}(A(x), A(y)))$ (by Definition 8) $= \lambda\mathbf{d}(A(x), A(y)) = \lambda\mathbf{A}(\mathbf{s})$. □

Example 6. *Take a picture of this page with a phone, possible slanted at an angle. The mapping from a point in the page to the corresponding point on the picture is an affine map.*

Definition 10 (Affine Isomorphism)**.** Given affine spaces (X, T) and (Y, S), an affine map $A : X \to Y$ is an affine isomorphism if there is an affine map $B : Y \to X$ such that AB and BA are identity maps.

Example 7. *Shadows from a distant light source such as sun of objects in a two-dimensional plane on another two-dimensional plane form an affine isomorphism. Shadows of objects in a three-dimensional space on a two-dimensional plane do not.*

Definition 11 (Affine Automorphism)**.** Given affine space (X, T), an affine isomorphism $X \to X$ is an affine automorphism.

3 Chapter III: Dual Spaces

Definition 12 (Covariant and Contravariant Vectors)**.** Given a vector space X, vectors in X are called contravariant vectors. A linear map from X to $\mathbb{R}$ is called a covariant vector or a dual vector or a linear functional.

Definition 13 (Dual Space)**.** The space of linear functionals (covariant vectors) on a vector space X is denoted $L(X, \mathbb{R})$ or X^*, and is called the dual space of X.

Result 2. *If X is a vector space, then X^* is a vector space.*

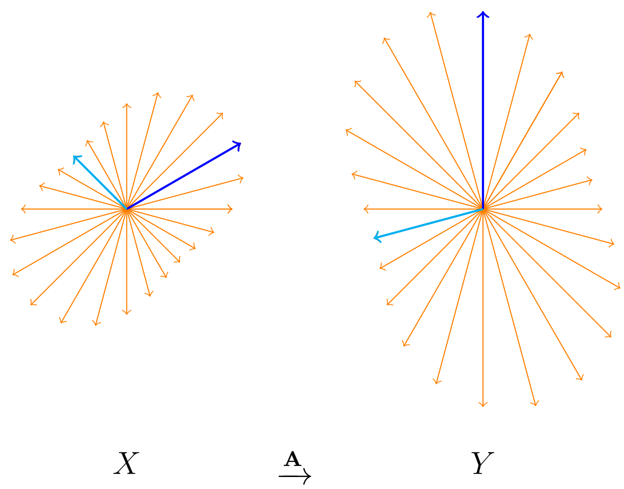
Proof. Define addition and scalar multiplication in X^* as pointwise addition of linear functionals and pointwise scalar multiplication. The result is also a linear functional. X^* satisfies Definition 3 of a vector space. $\square$

Example 8. *If X is the space of vectors in a 2-dimensional x-y plane, the x-component of the vector, the y-component of the vector, the length of the vector, the cosine of the angle between the vector and the x-axis are all elements of X^*.*

Definition 14 (Dual Map)**.** Given vector spaces X and Y, and a linear map $\mathbf{A} : X \to Y$, $\mathbf{A}^*$, the dual map to $\mathbf{A}$ is $\mathbf{A}^* : Y^* \to X^* : \mathbf{f} \mapsto \mathbf{f} \circ \mathbf{A}$.

Comment. *If $\mathbf{A} : X \to Y$ is a linear map and $\mathbf{f} : Y \to \mathbb{R}$ is a linear functional on Y, then $\mathbf{g} = \mathbf{A}^*(\mathbf{f}) : X \to \mathbb{R}$ is a linear functional on X such that for all $\mathbf{x} \in X$, $\mathbf{g}(\mathbf{x}) = \mathbf{f}(\mathbf{A}(\mathbf{x}))$.*

Example 9. *This is a continuation of Example 2.Both X and Y consist of vectors in two-dimensional plane. Map $\mathbf{A}$ scales a vector in X by 1.5 and rotates it counterclockwise by 60° to get its image in Y. Then, $\mathbf{A}^*$ maps any rule (same as linear functional) that assigns a number for each vector in Y to a rule that assigns to each vector in X, the number for its corresponding vector in Y. For example, if the rule in Y returns the polar angle (azimuth) of a vector, $\mathbf{A}^*$ maps this to a rule which returns the polar angle of a vector in X plus 60°. If the rule in Y returns the length of a vector, $\mathbf{A}^*$ maps this to a rule which returns 1,5 times the length of a vector in X.*



Result 3. *dim(X^*)=dim(X).*

Proof. Choose a basis $\beta = \mathbf{b}_1, \cdots, \mathbf{b}_n$ for X. Define n linear functionals

$$\mathbf{b}^i : X \to \mathbb{R} : (a^1, \cdots, a^n) \mapsto a^i \textbf{ for } i = 1, \cdots, n.$$

Consider a linear functional $\mathbf{f} \in X^*$. For any $\mathbf{x} = (x^1, \cdots, x^n) \in X$,

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\sum_{i=1}^{n} x^i \mathbf{b}_i) = \sum_{i=1}^{n} x^i \mathbf{f}(\mathbf{b}_i) = \sum_{i=1}^{n} \mathbf{f}(\mathbf{b}_i)\mathbf{b}^i(\mathbf{x}).$$

Thus, $\mathbf{f} = \sum_{i=1}^{n} \mathbf{f}(\mathbf{b}_i)\mathbf{b}^i$. That is, linear hull of $\mathbf{b}^1, \cdots, \mathbf{b}^n$ is X^* (or $\mathbf{b}^1, \cdots, \mathbf{b}^n$ span X^*). Further, $\mathbf{b}^1, \cdots, \mathbf{b}^n$ are linearly independent because $\sum_{i=1}^{n} \alpha_i \mathbf{b}^i(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i x^i = 0$ for all $\mathbf{x}$ only if $\alpha_1 = \cdots = \alpha_n = 0$. Thus, $\mathbf{b}^1, \cdots, \mathbf{b}^n$ is a basis of X^*. □

Definition 15 (Dual Basis)**.** Given basis $\mathbf{b}_1, \cdots, \mathbf{b}_n$ of a vector space X, the basis of X^*, $\mathbf{b}^1, \cdots, \mathbf{b}^n$, defined in the proof above, is the dual basis.

Example 10. *For the two-dimensional vector space in x-y plane, if the basis consists of unit vectors in x and y directions, then the dual basis consists of the x-coordinate and the y-coordinate.*

4 Chapter IV: Metric Vector Spaces

Definition 16 (Metric Tensor)**.** A metric tensor on a vector space X is a function $\mathbf{F} : X \times X \to \mathbb{R}$ which satisfies

1. $\mathbf{F}(\mathbf{x} + \mathbf{x}', \mathbf{y}) = \mathbf{F}(\mathbf{x}, \mathbf{y}) + \mathbf{F}(\mathbf{x}', \mathbf{y})$
2. $\mathbf{F}(\mathbf{x}, \mathbf{y} + \mathbf{y}') = \mathbf{F}(\mathbf{x}, \mathbf{y}) + \mathbf{F}(\mathbf{x}, \mathbf{y}')$
3. $\mathbf{F}(\mathbf{x}a, \mathbf{y}) = a\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{F}(\mathbf{x}, \mathbf{y}a)$
4. $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{F}(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in X$
5. $\mathbf{F}(\mathbf{x}, \mathbf{y}) = 0$ for all $\mathbf{y} \in X$ implies $\mathbf{x} = \mathbf{0}$

Definition 17 (Inner Product)**.** An inner product on a vector space X is a metric tensor on X which is

1. positive definite $\mathbf{F}(\mathbf{x}, \mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ or
2. negative definite $\mathbf{F}(\mathbf{x}, \mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$

Definition 18 (Metric Vector Space)**.** A metric vector space $(X, \mathbf{G})$ is a vector space X with a metric tensor $\mathbf{G} : X \times X \to \mathbb{R}$. We abbreviate $\mathbf{G}(\mathbf{x}, \mathbf{y})$ to $\mathbf{x} \cdot \mathbf{y}$.

Definition 19 (Inner Product Space)**.** An inner product space $(X, \mathbf{G})$ is a vector space X with an inner product $\mathbf{G} : X \times X \to \mathbb{R}$.

Definition 20 (Standard Inner Product on $\mathbb{R}^n$)**.** The standard inner product on $\mathbb{R}^n$ is defined by

$$(x^1, \cdots, x^n) \cdot (y^1, \cdots, y^n) = \sum_{i=1}^{n} x^i y^i.$$

Definition 21 (Lorentz Metric on $\mathbb{R}^4$)**.** The Lorentz metric on $\mathbb{R}^4$ is defined by

$$(x^0, x^1, x^2, x^3) \cdot (y^0, y^1, y^2, y^3) = x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3.$$

Comment. *The standard inner product is an inner product. Lorentz metric is a metric tensor, but not an inner product.*

Definition 22 (Length)**.** In a metric vector space $(X, \mathbf{G})$, the length of $\mathbf{x} \in X$ is $|x|_{\mathbf{G}} = \sqrt{\mathbf{x} \cdot \mathbf{x}}$.

Comment. *The length may be real or imaginary. Nonzero lengths are always real or always imaginary in an inner product space.*

Definition 23 (Norm)**.** A norm on a vector space X is a function $X \to \mathbb{R} : \mathbf{x} \mapsto ||\mathbf{x}||$ such that for all $\mathbf{x}, \mathbf{y} \in X$ and $a \in \mathbb{R}$,

1. $||\mathbf{x}|| = 0$ implies $\mathbf{x} = \mathbf{0}$,
2. $||\mathbf{x}a|| = |a| \, ||\mathbf{x}||$, and
3. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

Definition 24 (Partial Norm)**.** A partial norm on a vector space X is a function $X \to \mathbb{R} : \mathbf{x} \mapsto ||\mathbf{x}||$ such that for all $\mathbf{x}, \mathbf{y} \in X$ and $a \in \mathbb{R}$,

1. $||\mathbf{x}|| \geq 0$ for all $\mathbf{x} \in X$ and
2. $||\mathbf{x}a|| = |a| \, ||\mathbf{x}||$.

Comment. *A metric tensor assigns a number to a pair of vectors, while a norm assigns a number to each vector.*

Example 11. *In the space of vectors in x-y plane, the absolute value of x component is a partial norm, but not a norm.*

Definition 25. A metric tensor $\mathbf{G}$ can be used to get a norm: $||\mathbf{x}|| \equiv ||\mathbf{x}||_{\mathbf{G}} = \sqrt{|\mathbf{G}(\mathbf{x}, \mathbf{x})|}$, called the size of $\mathbf{x}$. The size is the same as the length $|\mathbf{x}|_{\mathbf{G}}$ if $\mathbf{G}$ is an inner product.

Definition 26. A unit vector $\mathbf{x} \in X$ in a metric vector space $(X, \mathbf{G})$ is a vector of size 1: $||\mathbf{x}||_{\mathbf{G}} = 1$.

Result 4. *In an inner product space $(X, \mathbf{G})$, for any $\mathbf{x}, \mathbf{y} \in X$, we have $\mathbf{x} \cdot \mathbf{y} \leq |\mathbf{x}| \, |\mathbf{y}|$ with equality for non-zero $\mathbf{x}, \mathbf{y}$ only if $\mathbf{y} = \mathbf{x}a$ for $a \in \mathbb{R}$.*

Proof. For any $a \in \mathbb{R}$,

$$(\mathbf{x}a - \mathbf{y}) \cdot (\mathbf{x}a - \mathbf{y}) = (\mathbf{x} \cdot \mathbf{x})a^2 - (2\mathbf{x} \cdot \mathbf{y}) + \mathbf{y} \cdot \mathbf{y} \geq 0.$$

This requires that the discriminant of the quadratic in a be non-positive:

$$\begin{aligned} &(2\mathbf{x} \cdot \mathbf{y})^2 - 4(\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y}) \leq 0 \\ &\implies \mathbf{x} \cdot \mathbf{y} \leq |\mathbf{x}| \, |\mathbf{y}|. \end{aligned}$$

With equality, there is a single value of a for which $\mathbf{x}a - \mathbf{y} = \mathbf{0}$. □

Definition 27 (Orthogonal Vectors)**.** Two vectors $\mathbf{x}$,$\mathbf{y}$ in a metric vector space are orthogonal if $\mathbf{x} \cdot \mathbf{y} = 0$.

5 Chapter V: Tensors and Multilinear Forms

Definition 28 (Multilinear Mapping)**.** A function $\mathbf{f} : X_1 \times X_2 \times \cdots X_n \to Y$ where $X_1, \cdots, X_n, Y$ are vector spaces, is a multilinear mapping if

1. $\mathbf{f}(\mathbf{x}_1, \cdots, \mathbf{x}_i + \mathbf{x}'_i, \cdots, \mathbf{x}_n) = \mathbf{f}(\mathbf{x}_1, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_n) + \mathbf{f}(\mathbf{x}_1, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_n)$
2. $\mathbf{f}(\mathbf{x}_1, \cdots, \mathbf{x}_i a, \cdots, \mathbf{x}_n) = (\mathbf{f}(\mathbf{x}_1, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_n))\, a$

for any $x_1 \in X_1, \cdots, x_n \in X_n$, and $x'_1 \in X_1, \cdots, x'_n \in X_n$, $i \in \{1, \cdots, n\}$, and $a \in \mathbb{R}$. The vector space of all such functions is denoted by $L(X_1, \cdots, X_n; Y)$.

Definition 29 (Multilinear Form)**.** A multilinear form on $\mathbf{X}$ is a multilinear mapping $\mathbf{f} \in L^n(X; Y) \equiv L(\overbrace{X, \cdots, X}^{n \text{ times}}; Y)$.

Example 12. *If* $\mathbf{X}$ *is a three-dimensional space, then* $\mathbf{f}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$*, the signed volume (following left-hand rule) of a parallelopiped formed by* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbf{X}$ *is a multilinear form..*

Definition 30 (Tensor Product of Spaces)**.** A tensor product of vector spaces $\mathbf{X}_1, \cdots, \mathbf{X}_n$ is a space $\mathbf{X}$ together with a map

$$\bigotimes : \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n \to \mathbf{X}$$

such that

1. $\bigotimes : \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n \to \mathbf{X}$ is multilinear
2. If $\mathbf{f} : \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n \to \mathbf{Y}$ is multilinear, then there is a unique linear map

$$\hat{\mathbf{f}} : \mathbf{X} \to \mathbf{Y}$$

   such that $\mathbf{f} = \hat{\mathbf{f}} \circ \bigotimes$.

Comment. *The tensor product space in Definition 30 needs to have addition operation for linearity to make sense. However, it is not necessarily closed under addition so it is not necessarily a vector space. To create a vector space, linear combinations of image of* $\bigotimes$ *should be included in the space.*

Example 13. *If* $\mathbf{X}_1^*, \mathbf{X}_2^*, \cdots \mathbf{X}_n^*$ *are dual spaces for vector spaces* $\mathbf{X}_1, \mathbf{X}_2, \cdots \mathbf{X}_n$*, respectively, the following mapping is a tensor product of spaces* $\mathbf{X}_1^*, \mathbf{X}_2^*, \cdots \mathbf{X}_n^*$

$$\bigotimes : \mathbf{X}_1^* \times \mathbf{X}_2^* \times \cdots \times \mathbf{X}_n^* \to L(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbb{R}) :$$
$$(\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_n) \mapsto \mathbf{g}_1 \otimes \mathbf{g}_2 \otimes \cdots \otimes \mathbf{g}_n$$

where $\mathbf{g}_1 \otimes \mathbf{g}_2 \otimes \cdots \otimes \mathbf{g}_n$ *is the multilinear mapping*

$$\mathbf{g}_1 \otimes \mathbf{g}_2 \otimes \cdots \otimes \mathbf{g}_n : \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n \to \mathbb{R} :$$
$$(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) \mapsto \mathbf{g}_1(\mathbf{x}_1)\mathbf{g}_2(\mathbf{x}_2) \cdots \mathbf{g}_n(\mathbf{x}_n).$$

Requirement 1 in the definition of tensor product is clearly satisfied. For requirement 2, let $\mathbf{b}_{i1}, \mathbf{b}_{i2}, \cdots, \mathbf{b}_{ik_i}$ *be a basis for* $\mathbf{X}_i$ *and let* $\mathbf{b}_i^1, \mathbf{b}_i^2, \cdots, \mathbf{b}_i^{k_i}$ *be the corresponding dual basis for* $\mathbf{X}_i^*$*. Then, for* $\mathbf{h} \in L(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbb{R})$*,* $\mathbf{x}_1 \in \mathbf{X}_1, \mathbf{x}_2 \in \mathbf{X}_2, \cdots, \mathbf{x}_n \in \mathbf{X}_n$*,*

$$\begin{aligned}
\mathbf{h}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) &= \mathbf{h}\left(\sum_{j_1=1}^{k_1} a_1^{j_1}\mathbf{b}_{1j_1}, \sum_{j_2=1}^{k_2} a_2^{j_2}\mathbf{b}_{2j_2}, \cdots, \sum_{j_n=1}^{k_n} a_n^{j_n}\mathbf{b}_{nj_n}\right) \\
&= \sum_{j_1=1}^{k_1}\sum_{j_2=1}^{k_2}\cdots\sum_{j_n=1}^{k_n} a_1^{j_1}a_2^{j_2}\cdots a_n^{j_n}\mathbf{h}(\mathbf{b}_{1j_1}, \mathbf{b}_{2j_2}, \cdots, \mathbf{b}_{nj_n}) \\
&= a_1^{j_1}a_2^{j_2}\cdots a_n^{j_n}\mathbf{h}(\mathbf{b}_{1j_1}, \mathbf{b}_{2j_2}, \cdots, \mathbf{b}_{nj_n}) \\
&= \mathbf{h}(\mathbf{b}_{1j_1}, \mathbf{b}_{2j_2}, \cdots, \mathbf{b}_{nj_n})\mathbf{b}_1^{j_1} \otimes \mathbf{b}_2^{j_2} \otimes \cdots \otimes \mathbf{b}_n^{j_n}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n).
\end{aligned}$$

Thus, $\mathbf{h}$ *has a unique representation as* $h_{j_1j_2\cdots j_n}\mathbf{b}_1^{j_1} \otimes \mathbf{b}_2^{j_2} \otimes \cdots \otimes \mathbf{b}_n^{j_n}$ *where* $h_{j_1j_2\cdots j_n} = \mathbf{h}(\mathbf{b}_{1j_1}, \mathbf{b}_{2j_2}, \cdots, \mathbf{b}_{nj_n})$*. Now, if* $\mathbf{f} : \mathbf{X}_1^* \times \mathbf{X}_2^* \times \cdots \times \mathbf{X}_n^* \to \mathbf{Y}$ *is multilinear, define* $\hat{\mathbf{f}} : L(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbb{R}) \to \mathbf{Y}$ *as*

$$\hat{\mathbf{f}} : h_{j_1j_2\cdots j_n}\mathbf{b}_1^{j_1} \otimes \mathbf{b}_2^{j_2} \otimes \cdots \otimes \mathbf{b}_n^{j_n} \mapsto h_{j_1j_2\cdots j_n}\mathbf{f}\left(\mathbf{b}_1^{j_1}, \mathbf{b}_2^{j_2}, \cdots, \mathbf{b}_n^{j_n}\right).$$

The mapping is, by definition, linear. It satisfies $\mathbf{f} = \hat{\mathbf{f}} \circ \bigotimes$*. It is unique because any other linear mapping should match* $\hat{\mathbf{f}}$ *on all* $\mathbf{b}_1^{j_1} \otimes \mathbf{b}_2^{j_2} \otimes \cdots \otimes \mathbf{b}_n^{j_n}$ *and, therefore, by linearity, match on all* $L(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbb{R})$*.*

Result 5. *A tensor product of vector spaces* $\mathbf{X}_1, \cdots, \mathbf{X}_n$ *exists, and any two are isomorphic.*

Proof. Existence follows from Example 13 after replacing each $\mathbf{X}_i$ with $\mathbf{X}_i^*$. Suppose $\bigotimes : \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n \to \mathbf{X}$ and $\bigotimes' : \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n \to \mathbf{X}'$ are both tensor products. Then, from the definition of tensor product, there exist unique $\mathbf{\Psi}$ and $\mathbf{\Phi}$ such that $\bigotimes = \mathbf{\Psi} \bigotimes'$ and $\bigotimes' = \mathbf{\Phi} \bigotimes$. These imply $\mathbf{\Psi\Phi} \bigotimes = \mathbf{I}_X \bigotimes$ and $\mathbf{\Phi\Psi} \bigotimes' = \mathbf{I}_{X'} \bigotimes'$. Since $\mathbf{I}_X \bigotimes$ is multilinear, so is $\mathbf{\Psi\Phi} \bigotimes$. Then, by the uniqueness property in part 2 of the definition of tensor product, $\mathbf{I}_X = \mathbf{\Psi\Phi}$. Similarly, $\mathbf{I}_{X'} = \mathbf{\Phi\Psi}$. Thus, $\mathbf{X} \cong \mathbf{X}'$. □

Definition 31 (Tensor Product and Tensor Product of Spaces)**.** The tensor product of vector spaces $\mathbf{X}_1, \cdots, \mathbf{X}_n$ defined in Definition 30 (unique upto an isomorphism from Result 5) is represented as $\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_n$. For $\mathbf{x}_1 \in X_1, \cdots, \mathbf{x}_n \in X_n$, the image of $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ under $\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_n$ is called the tensor product of $\mathbf{x}_1, \cdots, \mathbf{x}_n$ and is denoted by $\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_n$.

Comment. *From Example 13 and Definition 31, if $\mathbf{x}_1 \in X_1, \cdots, \mathbf{x}_n \in X_n$, the tensor product $\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_n \in L(X_1^*, \cdots, X_n^*; \mathbb{R})$ is a linear functional on $X_1^* \times \cdots \times X_n^*$.*

Result 6. *For any two vector spaces $\mathbf{X}_1$, $\mathbf{X}_2$, there is an isomorphism*

$$L(X_1; X_2) \to X_1^* \otimes X_2.$$

Proof. Define

$$\begin{aligned} \mathbf{f} : X_1^* \times X_2 &\to L(X_1; X_2) \\ (\mathbf{g}, \mathbf{x}_2) &\mapsto (\mathbf{x}_1 \mapsto (\mathbf{x_2})(\mathbf{g}(\mathbf{x_1}))) \,. \end{aligned}$$

Since $\mathbf{f}$ is multilinear, by part 2 of the definition of tensor product, there is a linear map

$$\hat{\mathbf{f}} : X_1^* \otimes X_2 \to L(X_1, X_2) \text{ with } \hat{\mathbf{f}} \bigotimes = \mathbf{f}.$$

Then, $\hat{\mathbf{f}}(\mathbf{g} \otimes \mathbf{x}_2) = \mathbf{0} \implies \mathbf{f}(\mathbf{g}, \mathbf{x}_2) = \mathbf{0} \implies (\mathbf{x}_2)(\mathbf{g}(\mathbf{x}_1)) = \mathbf{0} \forall \mathbf{x}_1 \implies \mathbf{g} \otimes \mathbf{x}_2 = \mathbf{0}$. This ensures that $\hat{\mathbf{f}}$ is injective (no two elements map to same image, we have only shown it for elements of the form $\mathbf{g} \otimes \mathbf{x}_2$, technically, we should show it for linear combinations of such terms too.). Finally, $dim(X_1^* \otimes X_2) = dim(X_1^*) dim(X_2) = dim(X_1) dim(X_2) = dim(L(X_1; X_2))$. Thus, $L(X_1; X_2)$ and $X_1^* \otimes X_2$ are isomorphic. □

Comment. *Result 6 states that $X_1^* \otimes X_2$ and $L(X_1; X_2)$ are isomorphic. Previous comment states that $X_1^* \otimes X_2$ and $L(X_1, X_2^*; \mathbb{R})$ are isomorphic.*

These two are consistent because $L(X_1; X_2)$ and $L(X_1, X_2^*; \mathbb{R})$ are isomorphic. An element in $L(X_1; X_2)$ maps a vector in X_1 to an image vector in X_2. The corresponding element in $L(X_1, X_2^*; \mathbb{R})$ returns a real number by applying a functional in $\mathbf{X}_2^*$ to the image vector.

Definition 32 (Tensors)**.** Vectors in the tensor product space X_h^k of the type

$$\underbrace{X \otimes X \otimes \cdots \otimes X}_{k \text{ times}} \otimes \underbrace{X^* \otimes X^* \otimes \cdots \otimes X^*}_{h \text{ times}}$$

for some space X are called tensors on X, covariant of degree h, contravariant of degree k, or of type $\binom{k}{h}$.

Comment. *When the definition says "vectors in the tensor product space," it does not mean contravariant vectors in X (or covariant vectors in X^*). A tensor covariant of degree h and contravariant of degree k can be thought of as a rule that linearly acts on h contravariant vectors and k covariant vectors to give a real number.*

Result 7. *dim(X_h^k)=dim(X)$^{h+k}$.*

Proof. Example 13 shows that tensor products of the form $\mathbf{b}_1^{j_1} \otimes \mathbf{b}_2^{j_2} \otimes \cdots \otimes \mathbf{b}_n^{j_n}$ form a basis for $L(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbb{R})$ so $\dim(L(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbb{R})) = k_1 k_2 \cdots k_n$. The result follows from Replacing n with $h + k$, each of $\mathbf{X}_1, \cdots, \mathbf{X}_h$ with $\mathbf{X}$, each of $\mathbf{X}_{h+1}, \cdots, \mathbf{X}_n$ with $\mathbf{X}^*$, and noting dim($\mathbf{X}^*$)=dim($\mathbf{X}$) (see Result 3). □

Example 14.

- *A scalar is a tensor of type $\binom{0}{0}$. It returns a number without any arguments.*

- *A contravariant vector is a tensor of type $\binom{1}{0}$. It can be applied to a covariant vector to get a number (for example, x-coordinate).*

- *A covariant vector (a linear functional) is a tensor of type $\binom{0}{1}$. It can be applied to a contravariant vector to get a number.*

- *A metric tensor is a tensor of type $\binom{0}{2}$. It is applied to two contravariant vectors to get a number.*

- *A linear mapping $X \to X$ is a tensor of type $\binom{1}{1}$, as given a vector and a functional, the functional applied to the image of the linear mapping is a scalar that is linear in the vector and also linear in the functional. This is a special case of Result 6.*

- *At a point in $\mathbb{R}^3$, for any unit vector $\mathbf{v}$, Cauchy stress tensor $\mathbf{T}(\mathbf{v})$ is a vector representing the stress (force per area) along the plane perpendicular to $\mathbf{v}$ at the point. Cauchy's postulate states that the stress is a function of the normal vector $\mathbf{v}$ only, and is not influenced by the curvature of the surface. Cauchy's stress theorem states that the stress is a linear function of $\mathbf{v}$. Since a linear combination of unit vectors is not necessarily a unit vector, we can generalize the definition to non-unit vectors as $\mathbf{T}(\mathbf{v})$ is $|v|$ times a vector representing the force per area, along the plane perpendicular to $\mathbf{v}$. Then, Cauchy stress tensor is a linear mapping from vectors in $\mathbb{R}^3$ to vectors in $\mathbb{R}^3$, and is thus, isomorphic to a tensor of type $\binom{1}{1}$ on the space of vectors in $\mathbb{R}^3$.*

Definition 33 (Contraction)**.** Given a tensor product space

$$\underbrace{X \otimes X \otimes \cdots \otimes X}_{k \text{ times}} \otimes \underbrace{X^* \otimes X^* \otimes \cdots \otimes X^*}_{h \text{ times}},$$

the linear mapping

$$\begin{aligned}
\underbrace{X \otimes X \otimes \cdots \otimes X}_{k \text{ times}} \otimes \underbrace{X^* \otimes X^* \otimes \cdots \otimes X^*}_{h \text{ times}} \to \underbrace{X \otimes X \otimes \cdots \otimes X}_{k-1 \text{ times}} \otimes \underbrace{X^* \otimes X^* \otimes \cdots \otimes X^*}_{h-1 \text{ times}} :\\
\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_{i-1} \otimes \mathbf{x}_i \otimes \mathbf{x}_{i+1} \otimes \cdots \otimes \mathbf{x}_k \otimes \mathbf{f}^1 \otimes \cdots \otimes \mathbf{f}^{j-1} \otimes \mathbf{f}^j \otimes \mathbf{f}^{j+1} \otimes \cdots \otimes \mathbf{f}^h\\
\mapsto \left(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_{i-1} \otimes \mathbf{x}_{i+1} \otimes \cdots \otimes \mathbf{x}_k \otimes \mathbf{f}^1 \otimes \cdots \otimes \mathbf{f}^{j-1} \otimes \mathbf{f}^{j+1} \otimes \cdots \otimes \mathbf{f}^h\right) \mathbf{f}^j(\mathbf{x}_i)
\end{aligned}$$

is called a contraction map and the image of a tensor $\mathbf{x}$ in $\mathbf{X}$ under this map is called a contraction of $\mathbf{x}$.

6 Chapter VI: Topological Vector Spaces

6.1 Continuity

Definition 34 (Metric). A metric on a set X is a function $d : X \times X \to \mathbb{R}$ satisfying

1. $d(x, y) = d(y, x)$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, z) \leq d(x, y) + d(y, z)$.

The pair (X, d) is called a metric space.

Definition 35 (Semimetric). A semimetric or pseudometric on a set X is a function $d : X \times X \to \mathbb{R}$ satisfying

1. $d(x, y) = d(y, x)$
2. $d(x, x) = 0$
3. $d(x, z) \leq d(x, y) + d(y, z)$.

Definition 36 (Continuous Function). A function $f : X \to Y$ between metric spaces (X, d) and (Y, d') is continuous at $x \in X$ if for any $0 < \epsilon \in \mathbb{R}$ there exists $0 < \delta \in \mathbb{R}$ such that if $d(x, y) < \delta$ then $d'(f(x), f(y)) < \epsilon$. A function is continuous on $S \in X$ if it is continuous at all $x \in S$. A function is continuous if it is continuous at all $x \in X$.

Definition 37 (Open Ball). Given a metric space (X, d), $x \in X$, and $0 < \delta \in \mathbb{R}$, the set $\{y | d(x, y) < \delta\}$ is called the open ball $B(x, \delta)$ of radius δ around x.

Definition 38 (Boundary Point). A boundary point or point of closure of a set S in a metric space (X, d) is a point x such that for any $0 < \delta \in \mathbb{R}$, $B(x, \delta)$ contains both points in S and points not in S.

Definition 39 (Open Set). A set in a metric space is open if it does not contain any boundary points it has.

Definition 40 (Closed Set). A set in a metric space is closed if it contains all its boundary points.

Result 8. *An open ball is an open set.*

Proof. Given a metric space (X, d), and an open ball $B(x, \delta)$ with $x \in X$ and $0 < \delta \in \mathbb{R}$, consider $x' \in B(x, \delta)$. Let $\epsilon = (\delta - d(x, x'))/2$. Clearly, $\epsilon > 0$ and for any $y \in B(x', \epsilon)$, $d(x, y) \leq d(y, x') + d(x', x) < \delta \implies y \in B(x, \delta)$. Thus, x' is not a boundary point of $B(x, \delta)$. $\square$

Result 9. *A function $f : X \to Y$ between two metric spaces is continuous at $x \in X$ if and only if for any open set V containing $f(x)$, there is an open set U containing x such that $f(U) \subseteq V$.*

Proof. **Only If**: If V is an open set containing $f(x)$, $f(x)$ is not a boundary point of V, so there exists $0 < \epsilon \in \mathbb{R}$ such that $B(f(x), \epsilon) \subseteq V$. If f is continuous at x, there exists $0 < \delta \in \mathbb{R}$ such that $d(x, y) < \delta \implies d(f(x), f(y)) < \epsilon$. Hence, $U = B(x, \delta)$ is an open set (see Result 8) containing x and $f(U) \subseteq B(f(x), \epsilon) \subseteq V$.
If: Suppose for any open set V containing $f(x)$, there is an open set U containing x such that $f(U) \subseteq V$. Given $0 < \epsilon \in \mathbb{R}$, choose $V = B(f(x), \epsilon)$. Then, there exists an open set U containing x such that $f(U) \subseteq B(f(x), \epsilon)$. By definition of open set, x is not a boundary point of U, so there exists $0 < \delta \in \mathbb{R}$ such that $B(x, \delta) \subseteq U$. Then, for any $x' \in X$ such that $d(x, x') < \delta$, $x' \in B(x, \delta) \subseteq U$ and therefore, $f(x) \in f(U) \subseteq B(f(x), \epsilon) \implies d(f(x), f(x')) < \epsilon$. $\square$

Result 10. *A map $f : X \to Y$ between two metric spaces is continuous if and only if $f^{\leftarrow}(V)$ is open for each open set V in Y.*

Proof. **Only If**: If f is continuous and V is an open set in Y, for each $x \in f^{\leftarrow}(V)$, there exists an open set U containing x such that $f(U) \in V$ (see Result 9). By definition of open set, x is not a boundary point of U, so there exists $0 < \delta \in \mathbb{R}$ such that $B(x, \delta) \subseteq U \subseteq f^{\leftarrow}(V)$. That is, x is not a boundary point of $f^{\leftarrow}(V)$. Thus, $f^{\leftarrow}(V)$ is open.
If: For any $x \in X$, consider any open set V in Y containing $f(x)$. Then, there is an open set $U = f^{\leftarrow}(V) \in X$, U contains x, and $f(U) \subseteq V$. By Result 9, f is continuous at x. $\square$

Definition 41 (Topology)**.** A topology on a set X is a specification of a family $\mathcal{T}$ of subsets of X, called the open sets of the topology, satisfying the axioms

1. empty set $\emptyset \in \mathcal{T}$,

2. For any finite family $\{U_i | i = 1, \cdots, n\}$ of open sets, $\bigcap_{i=1}^n U_i$ is open,

3. For any family $\{U_\alpha | \alpha \in A\}$ of open sets, $\cup_{\alpha \in A} U_\alpha$ is open.

The combination $(X, \mathcal{T})$ is called a topological space.

Definition 42 (Hausdorff Topology)**.** A topology $\mathcal{T}$ on a set X is a Hausdorff topology if for any two distinct points $x, y \in X$, there exist open sets $U, V \in \mathcal{T}$ such that $x \in U, y \in V$ and $U \cap V = \emptyset$.

Definition 43 (Metric Topology)**.** If X is a metric space, the metric topology on X is defined by the open sets determined by the metric.

Comment. *A topology isn't necessarily a metric topology as the open sets specified in Definition 41 need not satisfy Definition 39 of open sets based on some metric.*

Definition 44 (Metrisable)**.** A topological space $(X, \mathcal{T})$ is metrisable if $\mathcal{T}$ is the metric topology corresponding to a metric on X.

Definition 45 (Neighborhood)**.** A neighborhood of a point x in X is an open set containing x.

Definition 46 (Continuous)**.** A map $f : (X, \mathcal{T}) \to (Y, \Sigma)$ between two topological spaces is continuous if $V \in \Sigma \implies f^{\leftarrow}(V) \in \mathcal{T}$.

Comment. *The above definition is consistent with Definition 36 given Result 10.*

Definition 47 (Homeomorphism)**.** A map between two topological spaces is a homeomorphism if it is continuous, bijective, and its inverse is also continuous. Two topological spaces are homomorphic if there is a homeomorphism between them.

6.2 Limits

Definition 48 (Sequence)**.** A mapping $S : (X, \mathbb{N}) \to X : i \mapsto x_i$, where X is any set is called a sequence of points in X.

Definition 49 (Limit)**.** A sequence S of points x_i in a topological space X has the point x as a limit if every neighborhood of x contains x_i for all but finitely many $i \in \mathbb{N}$. If S has a limit x, represented as $\lim(S) = \lim_{i \to \infty} x_i = x$, then S is convergent and converges to x.

Result 11. *A function $X \to Y$ between topological spaces, where $(X, \mathcal{T})$ is metrisable, is continuous if and only if it preserves limits:*

$$\lim_{i\to\infty} x_i = x \implies \lim_{i\to\infty} f(x_i) \text{ exists and is } f(x).$$

Proof. **Only If**: Since $(X, \mathcal{T})$ is metrisable, Result 9 applies. For any neighborhood $N(f(x))$ of $f(x)$, there is a neighborhood $N'(x)$ of x such that $f(N'(x)) \subseteq N(f(x))$. Then, $f(x_i) \notin N(f(x)) \implies x_i \notin N'(x)$. If $\lim_{i\to\infty} x_i = x$, $A = \{i | x_i \notin N'(x)\}$ is finite. Then, $B = \{i | f(x_i) \notin N(f(x))\} \subseteq A$ is also finite, so $\lim_{i\to\infty} f(x_i) = f(x)$.
If: We prove the contrapositive. Consider a metric d on X such that the corresponding topology is $\mathcal{T}$. If f is not continuous at some x, then for some neighborhood $N(f(x))$ of $f(x)$, every neighborhood $N'(x)$ of x contains points y such that $f(y) \notin N(f(x))$. Consider the sequence $N_i(x) = B(x, \frac{1}{i})$ and choose $y_i \in N_i(x)$ such that $f(y_i) \notin N(f(x))$. Clearly, y_i converges to x but $f(y_i)$ doesn't converge to $f(x)$. □

Definition 50 (Weak Topology)**.** The weak topology on a vector space V is the smallest family $\mathcal{T}$ of subsets of V such that

1. $\mathcal{T}$ is a topology
2. For any linear functional $\mathbf{f} : V \to \mathbb{R}$ and open set $U \subseteq \mathbb{R}$, $f^{\leftarrow}(U) \in \mathcal{T}$.

Comment. *Open sets of a weak topology are the inverse images of open sets in* $\mathbf{R}$ *under linear functionals (and their finite unions and intersections).*

Definition 51 (Open Box Topology)**.** The open box topology on a finite-dimensional vector space V with basis $\mathbf{b}_1, \cdots, \mathbf{b}_n$ is the smallest family $\mathcal{T}$ of subsets of V such that

1. $\mathcal{T}$ is a topology
2. For every open set $U \subseteq \mathbb{R}$, each $(\mathbf{b}^i)^{\leftarrow}(U) \in \mathcal{T}$.

Comment. *Open sets of an open topology are "open boxes," sets of all points, whose ith coordinate for some i lies in an open set in* $\mathbf{R}$ *(and their finite unions and intersections).*

Result 12. *For any topological space X, the sum $\sum_{i=1}^n f_i$ of a finite set of continuous functions $f_1, \cdots, f_n : X \to \mathbb{R}$ is continuous.*

Proof. Consider the metric topology on $\mathbb{R}$ with the metric $d(x,y) = |x-y|$. Since $f_1, \cdots, f_n$ are continuous, for any $x \in X$ and $0 < \epsilon \in \mathbb{R}$, there exist open sets $U_1, \cdots, U_n \subseteq X$ containing x such that $f_i(U_i) \subseteq B(f_i(x), \epsilon/n)$ for $i = 1, \cdots, n$. Then, $U = \bigcap_{i=1}^n U_i$ is an open set (by definition of topology) containing x. For any $y \in U$,

$$\left| \sum_{i=1}^n f_i(y) - \sum_{i=1}^n f_i(x) \right| \leq \sum_{i=1}^n |f_i(y) - f_i(x)| < \sum_{i=1}^n \epsilon/n = \epsilon.$$

Thus, $\left(\sum_{i=1}^n f_i\right)(U) \subseteq B\left(\left(\sum_{i=1}^n f_i\right)x, \epsilon\right)$. □

Result 13. *For any basis $\mathbf{b}_1, \cdots, \mathbf{b}_n$ of a finite-dimensional vector space V, with some topology $\mathcal{T}$, all $\mathbf{f} \in V^*$ are continuous if and only if the vectors $\mathbf{b}^1, \cdots, \mathbf{b}^n$ of the dual basis are continuous.*

Proof. **Only If**: If all $\mathbf{f} \in V^*$ are continuous, so are $\mathbf{b}^1, \cdots, \mathbf{b}^n$.
If: Follows from Result 12 as each $\mathbf{f} \in V^*$ is a linear combination of $\mathbf{b}^1, \cdots, \mathbf{b}^n$. □

Definition 52 (Usual Topology)**.** The usual topology on a finite-dimensional vector space is the weak topology on the vector space or equivalently, the open box topology for a basis on the vector space.

Comment. *The equivalence in the definition follows from Result 13.*

Result 14. *The metric given by any norm on a finite-dimensional vector space defines the usual topology on the vector space.*

Comment. *Proof is outlined in Exercise VI.3.8 in Dodson and Poston (1991). The result allows us to assume the usual topology on a vector space, even if the norm is not specified.*

Result 15. *If V, W are finite-dimensional vector spaces, then all linear maps $\mathbf{A} : V \to W$ are continuous in the usual topology on V.*

Proof. Choose a basis $\mathbf{b}_1, \cdots, \mathbf{b}_1$ for W and arbitrary open intervals $I_i \in \mathbf{R}$ that define an open box

$$U = \left\{(w^1, \cdots, w^n) \in W | w^1 \in I_1, \cdots, w^n \in I_n\right\} = \bigcap_{i=1}^n \left(\mathbf{b}^i\right)^{\leftarrow}(I_i).$$

The maps $\mathbf{b}^i \circ \mathbf{A} : V \to \mathbb{R}$ for $i = 1, \cdots, n$ are compositions of linear maps and hence linear. By definition of weak topology, $(\mathbf{b}^i \circ \mathbf{A})^{\leftarrow}(I_i)$ are open in V. Then, their intersection is also open. Now,

$$\mathbf{v} \in \bigcap_{i=1}^{n} \left(\mathbf{b}^i \circ \mathbf{A}\right)^{\leftarrow}(I_i) \iff \mathbf{b}^i \circ \mathbf{A}(\mathbf{v}) \in I_i \text{ for each } i$$
$$\iff \mathbf{A}(\mathbf{v}) \in \left(\mathbf{b}^i\right)^{\leftarrow}(I_i) \text{ for each } i \iff \mathbf{A}(\mathbf{v}) \in U.$$

So $\mathbf{A}^{\leftarrow}(U) = \bigcap_{i=1}^{n} (\mathbf{b}^i \circ \mathbf{A})^{\leftarrow}(I_i)$, which we showed to be open. We have shown that $\mathbf{A}^{\leftarrow}(U)$ is open for an arbitrary open box in W. Since an arbitrary open set $U' \in W$ is a union of open boxes, $\mathbf{A}^{\leftarrow}(U')$ is a union of open sets in V, and hence, an open set in V. Thus, $\mathbf{A}$ is continuous. □

Definition 53 (Product Topology)**.** Given topological spaces $X_1, \cdots, X_n$, the product topology on the set $X_1 \times X_2 \times \cdots \times X_n$ is the collection of all unions of sets of the form

$$U_1 \times \cdots \times U_n \subseteq X_1 \times \cdots \times X_n$$

where each U_i is open in X_i.

6.3 Compactness

Hypothesis 1 (Intermediate Value Hypothesis)**.** If a function $f : [0,1] \to \mathbb{R}$ is continuous and for some $v \in \mathbb{R}$, we have $f(1) < v < f(0)$ or $f(0) < v < f(1)$, then there exists $x \in [0,1]$ such that $f(x) = v$.

Axiom 1 (Completeness Axiom)**.** Intermediate Value Hypothesis is true.

Result 16. *A sequence $J_1, J_2, \cdots$ of closed subintervals J_i of the closed interval $[a,b]$ such that $J_{i+1} \subseteq J_i$ satisfies $\bigcap_{i \in \mathbb{N}} J_i \neq \emptyset$. That is, there exists $x \in [a,b]$ which is in every J_i.*

Comment. *The omitted proof relies on Intermediate Value Hypothesis.*

Result 17. *If $S : \mathbb{N} \to [a,b] \subset \mathbb{R} : i \mapsto x_i$ is a sequence of points in the interval $[a,b]$, then S has at least one convergent subsequence.*

Proof. Start with the interval $[a,b]$ with infinitely many x_i and find a sequence of closed intervals, left half or right half of the previous interval containing infinitely many x_i. The result then follows from Result 16. □

Definition 54 (Bounded)**.** A set $S \subseteq \mathbb{R}$ is bounded if there exists a bound $b \in \mathbb{R}$ such that $x \in S \implies |x| \leq b$.

Definition 55 (Compact)**.** A set C in a finite-dimensional vector space X is compact if

1. It is closed in the usual topology.
2. For any linear functional $\mathbf{f} \in X^*$, $\mathbf{f}(C) \subseteq \mathbb{R}$ is bounded.

Result 18. *A set $S \subseteq X$ is bounded if and only if the coordinates of points in S are bounded by some $b \in \mathbb{R}$.*

Comment. *The omitted proof is similar to the proof of Result 13.*

Definition 56 (Induced Topology)**.** If S is a subset of a topological space X, the induced topology on S is the collection of sets $\{S \cap U | U \text{ open in } X\}$.

Result 19. *A sequence in a subspace S of X converges to $x \in S$ in the induced topology on S if and only if it converges to x as a sequence in X.*

Proof. **Only If**: Suppose a sequence in S converges to $x \in S$ as a sequence in X. Then, every neighborhood of x in X contains x_i for all but finitely many $i \in \mathbb{N}$. Then, this is also true for every neighborhood of x in S.
If: Suppose a sequence in S converges to $x \in S$ in the induced topology on S. Then, every neighborhood $N(x)$ of x in X contains x_i for all but finitely many $i \in \mathbb{N}$ because so does neighborhood $N(x) \cap S$ of x in S. $\square$

Result 20. *If C is a compact set in any finite-dimensional vector or affine space X, and $f : C \to \mathbb{R}$ is a continuous function with respect to the induced topology on C, then $f(C)$ is bounded and closed.*

Proof. The proof assumes X is a vector space. Choose a basis. From Result 18, C's coordinates are bounded by some $b \in \mathbb{R}$. We first show that C has convergent subsequence property. Let S be a sequence of points $c_i = (c^1(i), \cdots, c^n(i))$. Since $c^1(i) \in [-b, b]$, by Result 17, there is a subsequence S^1 of S such that the sequence of its first coordinates converges to some x^1. Applying the same logic to S^1, we can find a subsequence S^2 such that the sequence of its second coordinates converges to some x^2. Repeating this process for all coordinates, we get a subsequence $\bar{S}$ of S whose jth coordinate converges to x^j. Then, $\bar{S}$ converges to $(x^1, \cdots, x^n)$, which must lie in C since C is compact. By Result 19, $\bar{S}$ converges in the induced topology on C.

Now, if $f(C)$ is not bounded, there is a sequence x_i in $f(C)$ that goes to infinity with all its subsequences. For each x_i choose $c_i \in C$ such that $f(c_i) = x_i$. From convergent subsequence property, the sequence $c_1, c_2, \cdots$ has a subsequence converging to a point $c \in C$. Then, in this subsequence, $\lim_{i\to\infty} x_i = \lim_{i\to\infty} f(c_i) = f(\lim_{i\to\infty} c_i) = f(c)$, where the second equality follows from continuity of f (see Result 11). Since we have found a convergent subsequence of x_i, $f(C)$ must be bounded.

If x is a boundary point of $f(C)$, choose a sequence x_i in $f(C)$ converging to x, a sequence $c_i \in C$ such that $f(c_i) = x_i$, and a convergent subsequence $c'_1, c'_2, \cdots$ of the sequence $c_1, c_2, \cdot$ with limit $c \in C$. Then, the sequence $x'_1, x'_2, \cdots$, where $x'_i = f(c'_i)$ converges to x, and $x = \lim_{i\to\infty} x'_i = \lim_{i\to\infty} f(c'_i) = f(\lim_{i\to\infty} c'_i) = f(c)$. Thus, $x \in C$ so $f(C)$ is closed. $\square$

7 Chapter VII: Differentiation and Manifolds

7.1 Differentiation

Definition 57 (Derivative at a Point on Affine Space)**.** If $f : X \to X'$ is a map between affine spaces. a derivative of f at $x \in X$ is a linear map

$$(\mathbf{D}_x f) : T_x X \to T_{f(x)} X'$$

such that for any neighborhood N in $L(T_x X; T_{f(x)} X')$ of the zero linear map, there is a neighborhood N' of $\mathbf{0} \in T_x X$ such that if $\mathbf{t} \in N'$, then

$$\mathbf{d}'\left(f(x), f(x+\mathbf{t})\right) = \mathbf{d}'_{f(x)}\left((\mathbf{D}_x f)(\mathbf{t}) + \mathbf{A}(\mathbf{t})\right)$$

for some $\mathbf{A} \in N$, where $\mathbf{d}'$ is the distance function in X'.

Comment.

1. *We used parentheses to indicate that $(\mathbf{D}_x f)$ is a mapping, not a composition of two mappings $\mathbf{D}_x$ and f. $\mathbf{D}_x$ is undefined.*

2. *This mapping takes bound vectors in tangent space to X at x (orange arrows in the following figure) to bound vectors in tangent space to X' at $f(x)$ (red arrows in the following figure). That is, movements in X around x are mapped to movements around $f(x)$ in X'.*

3. *The mapping is linear: if you move twice as far in the same direction from x, the mapping takes you twice as far from $f(x)$ in the same direction.*

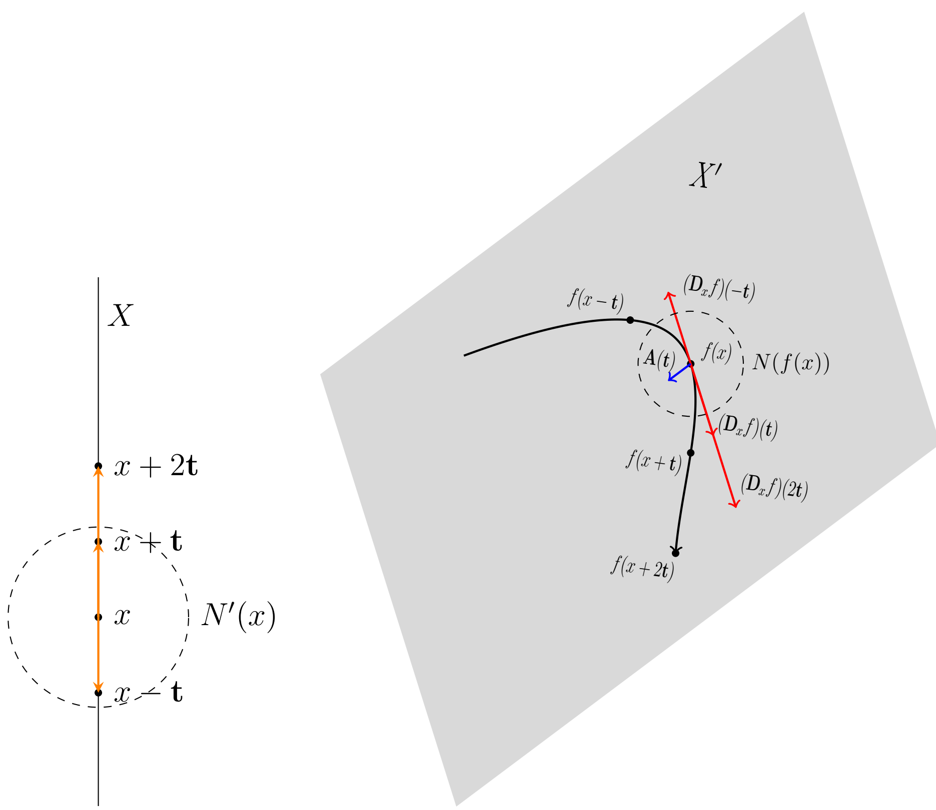
4. *Since the movement is a linear approximation to f, in general, if we move $\mathbf{t}$ from x, the mapping will not yield the vector from $f(x)$ to $f(x+\mathbf{t})$. However, as $\mathbf{t}$ becomes small, it gets arbitrarily close.*

5. *The left hand side of the equation in Dodson and Poston (1991) is $\mathbf{d}'\left(f(x+\mathbf{t}), f(x)\right)$. That seems to be an error.*

6. *The left side in the above equation is the distance (in X') between $f(x)$ and $f(x+\mathbf{t})$. The right side is the (free vector corresponding to) a bound vector, the value of the derivative applied to $\mathbf{t}$, plus a "correction term" $\mathbf{A}(\mathbf{t})$ (blue arrow in the following figure).*

7. *The definition specifies that the correction term is smaller than a small linear mapping of* $\mathbf{t}$ *(that is, in a neighborhood of a zero map from* T_xX *to* $T_{f(x)}X'$*, shown with the dotted circle in* X' *in the following figure) if* $\mathbf{t}$ *is sufficiently small (in an appropriately chosen neighborhood of* $\mathbf{0}$ *in* T_xX*, shown with the dotted circle in* X *in the following figure).*

8. *Even though no topologies are assumed for affine spaces* X *and* X'*, neighborhoods can be defined by the usual topologies, which follow from any norm (Result 14). It is not clear how neighborhoods in* $L(T_xX; T_{f(x)}X')$ *are defined. One possibility is that a neighborhood of the zero linear map in* $L(T_xX; T_{f(x)}X')$ *is the set of all maps from some neighborhood of* $\mathbf{0}_X$ *to some neighborhood of* $\mathbf{0}_{X'}$*. That is,* N *is defined as a neighborhood of the zero linear map in* $L(T_xX; T_{f(x)}X')$ *if* $N = \{\mathbf{A}|\mathbf{A}(\mathbf{t}) \in M\ \forall \mathbf{t} \in N'\}$ *for some neighborhood* N' *of* $\mathbf{0}_X$ *in* X *and some neighborhood* M *of* $\mathbf{0}_{X'}$ *in* X'*.*



Result 21. *If f has a derivative at x, it is unique.*

Proof. Let $\mathbf{d}$ and $\mathbf{d}'$ be distance functions in X and X', and d and d' be metrics in X and X'. Suppose there are two derivatives $\mathbf{D}_x^1 f$ and $\mathbf{D}_x^2 f$, whose values differ for some $\mathbf{u}$ in $T_x X$. Suppose $d'((\mathbf{D}_x^1 f)(\mathbf{u}), (\mathbf{D}_x^2 f)(\mathbf{u})) = \delta > 0$. Consider a neighborhood N in $L(T_x X; T_{f(x)} X')$ of the zero linear map such that for all $\mathbf{A} \in N$, $d'(\mathbf{A}(\mathbf{u}), \mathbf{0}_{X'}) \leq \delta/3$. By definition of derivative, there is a neighborhood N' of x such that if $k\mathbf{u} \in N'$, then,

$$\begin{aligned} \mathbf{d}'\left(f(x), f(x + k\mathbf{u})\right) &= \mathbf{d}'_{f(x)}\left((\mathbf{D}_x^1 f)(k\mathbf{u}) + \mathbf{A}^1(k\mathbf{u})\right), \\ \mathbf{d}'\left(f(x), f(x + k\mathbf{u})\right) &= \mathbf{d}'_{f(x)}\left((\mathbf{D}_x^2 f)(k\mathbf{u}) + \mathbf{A}^2(k\mathbf{u})\right) \end{aligned}$$

for some $\mathbf{A}^1, \mathbf{A}^2 \in N$. If follows

$$\mathbf{d}'\left((\mathbf{D}_x^1 f)(k\mathbf{u}), (\mathbf{D}_x^2 f)(k\mathbf{u})\right) = \mathbf{d}'\left(\mathbf{A}^2(k\mathbf{u}), \mathbf{A}^1(k\mathbf{u})\right).$$

Then, by triangle inequality,

$$\begin{aligned} k\delta &= d'\left((\mathbf{D}_x^1 f)(k\mathbf{u}), (\mathbf{D}_x^2 f)(k\mathbf{u})\right) \\ &= d'\left(\mathbf{A}^2(k\mathbf{u}), \mathbf{A}^1(k\mathbf{u})\right) \\ &\leq d'\left(\mathbf{A}^2(k\mathbf{u}), \mathbf{0}_{X'}\right) + d'\left(\mathbf{0}_{X'}, \mathbf{A}^1(k\mathbf{u})\right) \leq 2k\delta/3, \end{aligned}$$

a contradiction. □

Result 22. *If X and X' are affine spaces, the derivative of $f : X \to X'$ at $x \in X$, if it exists, is given by*

$$(\mathbf{D}_x f)(\mathbf{t}) = \lim_{h \to 0} \mathbf{d}'^{\leftarrow}_{f(x)} \left(\frac{\mathbf{d}'\left(f(x), f(x + h\mathbf{t})\right)}{h} \right)$$

where $\mathbf{d}'$ is the distance function in X'.

Comment. *Proof omitted. The expression gives the value of the derivative for vector $\mathbf{t}$ in the tangent space $T_x X$. The derivative specifies a value for each $\mathbf{t}$. Note that scaling t scales the derivative by the same factor. The job of $\mathbf{d}'^{\leftarrow}_{f(x)}$ is to convert a free vector to a bound vector.*

Definition 58 (Differentiability on Affine Spaces)**.** If X and X' are affine spaces, a map $f : X \to X'$ is differentiable at $x \in X$ if f has a derivative at x. It is differentiable if it is differentiable at x for all $x \in X$.

Definition 59 (Derivative on Affine Space)**.** If (X, T) and (X', T') are affine spaces with distance functions $\mathbf{d}$ and $\mathbf{d}'$, respectively, and $f : X \to X'$ is differentiable, the derivative of f is the map

$$\hat{\mathbf{D}}f : X \to L(T; T') : x \mapsto \mathbf{d}'_{f(x)}(\mathbf{D}_x f)\mathbf{d}_x^{\leftarrow}.$$

Comment. *The derivative at a point is defined as an operator that takes tangent vectors in one tangent space to tangent vectors in another tangent space. The derivative of the map (not just at a point) is defined to take a point (for example, x) in the affine space to the derivative at that point ($(\hat{\mathbf{D}}f)(x) = \mathbf{d}'_{f(x)}(\mathbf{D}_x f)\mathbf{d}_x^{\leftarrow}$), modified such that it takes free vectors (not bound) (for example, $t \in T_x X$) to free vectors (not bound) ($(\mathbf{d}'_{f(x)}(\mathbf{D}_x f)\mathbf{d}_x^{\leftarrow})(\mathbf{t}) \in T_{f(x)}Y$). This enables us to consider the value of the derivative at different points applied to the same free vector.*

Definition 60 (Higher Order Derivative)**.** If (X, T) and (X', T') are affine spaces, $(\hat{\mathbf{D}}^k f) : X \to L^k(T; T') = \hat{\mathbf{D}}(\hat{\mathbf{D}}^{k-1} f)$ for $k > 1$ with $\hat{\mathbf{D}}^1 f = \hat{\mathbf{D}} f$.

Comment. *From Result 6,* $(\hat{\mathbf{D}}^k f) : X \to \underbrace{T^* \otimes T^* \otimes \cdots \otimes T^*}_{k \text{ times}} \otimes T'$.

Definition 61 (Partial Derivative)**.** Given affine spaces (X, T) and (X', T') with basis $(\mathbf{b}_1, \cdots, \mathbf{b}_n)$ for T and basis $(\mathbf{e}_1, \cdots, \mathbf{e}_m)$ for T', a partial derivative of $f : X \to X'$ is

$$(\partial_j f^i) : X \to \mathbb{R} : x \mapsto \lim_{h \to 0} \frac{\mathbf{e}^i(f(x + h\mathbf{b}_j)) - \mathbf{e}^i(f(x))}{h}$$

where $(\mathbf{e}^1, \cdots, \mathbf{e}^m)$ is the dual basis for dual space T'^*.

Comment. *A partial derivative is defined with respect to "directions" in both domain affine space and range affine space, once we define these directions. That requires specifying bases in each space. The value of a partial derivative is real as the components $\mathbf{e}^i$ are real.*

7.2 Manifolds

Definition 62 (Manifold)**.** A C^k manifold modeled on an affine space X is a Hausdorff topological space M together with a collection of open sets $\{U_a | a \in A\}$ in M and corresponding maps $\phi_a : U_a \to X$, such that

1. $\cup_{a \in A} U_a = M$,
2. For each $a \in A$, ϕ_a defines a homeomorphism $U_a \to \phi_a(U_a)$, and
3. if $a, b \in A$ and $U_a \cap U_b \neq \emptyset$, then the composite $\phi_b \circ \phi_a^{\leftarrow}$ on the set $\phi_a(U_b)$ on which it is defined, is C^k.

The pairs (U_a, ϕ_a) are called charts on M, and the set $\{(U_a, \phi_a) | a \in A\}$ is called an atlas.

Comment. *In the third condition in the definition above, why shouldn't the composite map $\phi_b \circ \phi_a^{\leftarrow}$ be defined on the set $\phi_a(U_a \cap U_b)$ rather than $\phi_a(U_b)$?*

Definition 63 (Smooth Manifold)**.** A smooth manifold is a C^∞ manifold.

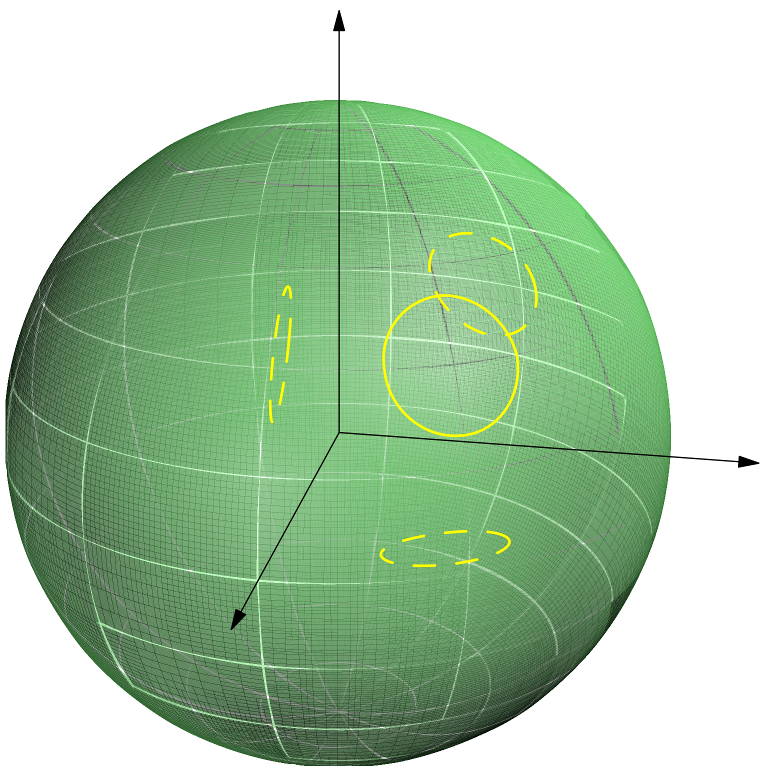
Definition 64 (Dimension of Manifold)**.** The dimension $\dim(M)$ of a manifold M is the dimension of the affine space it is modeled on. An n-manifold is a manifold with dimension n.

Example 15. *Consider unit 2-sphere, as in Exercise VII.2.1 of Dodson and Poston (1991), defined as $S^2 = \{(x^1, x^2, x^3) \in \mathbb{R}^3 | (x^1)^2 + (x^2)^2 + (x^3)^2 = 1\}$. It is a 2-dimensional manifold embedded in $\mathbb{R}^3$ with charts (U_{1+}, ϕ_{1+}), (U_{1-}, ϕ_{1-}), (U_{1+}, ϕ_{1+}), (U_{1-}, ϕ_{1-}), (U_{1+}, ϕ_{1+}), and (U_{1-}, ϕ_{1-}) defined with "flattening maps" as*

$$\begin{aligned}
U_{i+} &= \{(x^1, x^2, x^3) \in S^2 | x^i > 0\},\ i = 1, 2, 3, \\
U_{i-} &= \{(x^1, x^2, x^3) \in S^2 | x^i < 0\},\ i = 1, 2, 3 \\
\phi_{1s} &: U_{1s} \to \mathbb{R}^2 : (x^1, x^2, x^3) \mapsto (x^2, x^3),\ s \in \{+, -\}, \\
\phi_{2s} &: U_{2s} \to \mathbb{R}^2 : (x^1, x^2, x^3) \mapsto (x^1, x^3),\ s \in \{+, -\},\ \textit{and} \\
\phi_{3s} &: U_{3s} \to \mathbb{R}^2 : (x^1, x^2, x^3) \mapsto (x^1, x^2),\ s \in \{+, -\}.
\end{aligned}$$

The following figure shows the images of a circle in S^2 mapped to $\mathbb{R}^2$ using charts (U_{1+}, ϕ_{1+}), (U_{2+}, ϕ_{2+}), and (U_{3+}, ϕ_{3+}). Even though all charts map to the same $\mathbb{R}^2$, the figure shows them on three different planes for clarity. One of these images can be mapped to another through composite maps mentioned in Definition 62 of a manifold. For example,

$$\phi_{2+} \circ \phi_{1+}^{\leftarrow} : \mathbb{R}^2 \to \mathbb{R}^2 : (x^2, x^3) \mapsto (\sqrt{1 - (x^2)^2 - (x^3)^2}, x^3).$$



Definition 65 (Differentiability on Manifolds)**.** A map $f : M \to N$ between smooth manifolds is differentiable at $x \in M$ if for some charts (U, ϕ) on M, (V, ψ) on N, with $x \in U$, $f(x) \in V$, the map $\psi \circ f \circ \phi^{\leftarrow}$ is differentiable at $\phi(x)$. The map f is C^k at x if the map $\psi \circ f \circ \phi^{\leftarrow}$ is C^k at $\phi(x)$. The map is differentiable if it is differentiable at all $x \in M$. The map is C^k if it is C^k at all $x \in M$.

Comment. *Basically, the differentiability of a map on manifolds is defined in terms of differentiability of the corresponding map between their underlying affine spaces.*

Definition 66 (Diffeomorphism on Manifolds)**.** A homoemorphism $f : M \to N$ between C^k manifolds is a C^k diffeomorphism if both f and $f^{\leftarrow}$ are C^k. Two manifolds are diffeomorphic if there is a diffeomorphism between them.

Result 23. *Let M be a smooth manifold modeled on an affine space (X, T) with $u \in M$. Let (U, ϕ), (U', ϕ') be charts on M with $u \in U \cap U'$, and*

$\mathbf{t}, \mathbf{t}' \in T$. *Define relation* $\sim$ *by*

$$(U, \phi, \mathbf{t}) \sim (U', \phi', \mathbf{t}') \iff \left(\left(\hat{\mathbf{D}}\left(\phi' \circ \phi^{\leftarrow}\right)\right)(\phi(u))\right)(\mathbf{t}) = \mathbf{t}'.$$

Then,

1. $\sim$ *is an equivalence relation.*
2. *If* $(U, \phi, \mathbf{t}) \sim (U', \phi', \mathbf{t}')$ *and* $(U, \phi, \mathbf{s}) \sim (U', \phi', \mathbf{s}')$, *then*

$$(U, \phi, \mathbf{t} + \mathbf{s}) \sim (U', \phi', \mathbf{t}' + \mathbf{s}')$$

and

$$(U, \phi, \mathbf{t}a) \sim (U', \phi', \mathbf{t}'a) \text{ for all } a \in \mathbb{R}.$$

Hence, the set of $\sim$ *equivalence classes is a vector space.*

Comment. *Proof omitted. Two pairs of charts (around* u*) and tangent vectors in* T *belong to the same equivalence class if the derivative of the composite map* $\phi' \circ \phi^{\leftarrow}$ *from the image under the first chart to the image under the second chart maps the first tangent vector to the second tangent vector. The composite map here is between affine spaces and hence its derivative is defined already.*

Comment. *We used slightly different notation from Dodson and Poston (1991). In the definition of the equivalence relation* $\sim$*, they specify the right side as* $\hat{\mathbf{D}}_{\phi(u)}(\phi' \circ \phi^{\leftarrow})\mathbf{t} = \mathbf{t}'$ *but the notation* $\hat{\mathbf{D}}_{\phi(u)}$ *has not been defined. We instead use* $((\hat{\mathbf{D}}(\phi' \circ \phi^{\leftarrow}))(\phi(u)))(\mathbf{t})$*. Here,* $\hat{\mathbf{D}}(\phi' \circ \phi^{\leftarrow})$ *is the derivative of the composite map* $\phi' \circ \phi^{\leftarrow}$*. It is a map from points in* X *to maps from free vectors in* T *to free vectors in* T*. Its value at the point* $\phi(u)$ *is* $(\hat{\mathbf{D}}(\phi' \circ \phi^{\leftarrow}))(\phi(u))$*, a map from free vectors in* T *to free vectors in* T*. Finally, its value at free vector* $\mathbf{t}$ *is* $((\hat{\mathbf{D}}(\phi' \circ \phi^{\leftarrow}))(\phi(u)))(\mathbf{t})$*, a free vector in* T*. The notation could have been simplified to* $(\mathbf{D}_{\phi(u)}(\phi' \circ \phi^{\leftarrow}))(\mathbf{t}) = \mathbf{t}'$ *if* $\mathbf{t}$ *and* $\mathbf{t}'$ *were bound vectors, but they need to be free vectors in this definition.*

Definition 67 (Tangent Space to a Manifold)**.** The tangent space T_uM to M at u is the vector space consisting of the set of $\sim$ equivalence classes defined in Result 23. Each equivalence class is a tangent vector to M at u.

Comment. *A tangent vector to a manifold is a vector in the complicated vector space of equivalence classes defined in Result 23. It can be represented as a vector in the affine space underlying the manifold once we pin down a chart* (U, ϕ)*. The vector will be different if we choose a different chart. However, all those vector representations will be related by an equivalence class defined in Result 23.*

Result 24. *Let* M *be a smooth manifold modeled on an affine space* (X, T) *and embedded in* $\mathbb{R}^n$ *with inclusion* $\iota : M \to \mathbb{R}^n$*. Let* (U, ϕ)*,* (U', ϕ') *be charts on* M *with* $u \in U \cap U'$*, and* $\mathbf{t}, \mathbf{t}' \in T$*, and* $\sim$ *be as defined in Result 23. Then*

$$(U, \phi, \mathbf{t}) \sim (U', \phi', \mathbf{t}') \iff \left(\left(\hat{\mathbf{D}}\left(\iota \circ \phi^{\leftarrow}\right)\right)(\phi(u))\right)(\mathbf{t}) = \left(\left(\hat{\mathbf{D}}\left(\iota \circ \phi^{\leftarrow}\right)\right)(\phi'(u))\right)(\mathbf{t}').$$

Thus, each tangent vector to M *at* u *is uniquely represented by a single vector* $\mathbf{s} \in T_{\iota(u)}\mathbb{R}^n$ *such that for any* $(U, \phi, \mathbf{t})$ *in the equivalence class of that tangent vector, the derivative of the composite map* $\iota \circ \phi^{\leftarrow}$ *from the image under the chart* ϕ *to* $\mathbb{R}^n$ *maps* $\mathbf{t}$ *to* $\mathbf{s}$*.*

Comment. *Proof omitted but probably follows from Result 23 and Inverse Function Theorem as hinted in Exercise VII.2.5. The notation to the right of the equivalence operator has been changed and* $T_u\mathbb{R}^n$ *has been changed to* $T_{\iota(u)}\mathbb{R}^n$ *to be consistent with the rest of the notation.*

Result 25. *An* m*-dimensional manifold* M *can be mapped smoothly and injectively into* $\mathbb{R}^n$ *for* $n \geq N$ *where* $N \leq 2m + 1$*. The map* $\iota : M \to \mathbb{R}^n$ *is called inclusion.*

Comment. *Proof is supposed to be nontrivial. If* $M \subset \mathbb{R}^n$*, we indicate inclusion as* $\iota : M \hookrightarrow \mathbb{R}^n$ *to emphasize that the domain of the inclusion is a subset of its range.*

Definition 68 (Derivative on Manifold)**.** Let M be a manifold modeled on an affine space (X, T) and N be a manifold modeled on an affine space (Y, S). If $f : M \to N$ is differentiable at u with chart (U, ϕ) on M, chart (V, ψ) on N, $u \in U$, and $f(u) \in V$, then the derivative of f at u maps equivalence classes in T_uM to $T_{f(u)}N$:

$$\begin{aligned} \mathbf{D}_u f : &T_uM \to T_{f(u)}N : \\ &[(U, \phi, \mathbf{t})] \mapsto \left[\left(V, \psi, \mathbf{D}_{\phi(u)}\left(\phi \circ f \circ \phi^{\leftarrow}\right) \mathbf{t}\right)\right] \end{aligned}$$

where $\mathbf{t} \in T$. Here, $(U', \phi', \mathbf{t}') \in [(U, \phi, \mathbf{t})]$ iff (U', ϕ') is a chart on M with $u \in U'$, $\mathbf{t}' \in T$, and

$$\left(\left(\hat{\mathbf{D}}\left(\phi' \circ \phi^{\leftarrow}\right)\right)(\phi(u))\right)(\mathbf{t}) = \mathbf{t}'.$$

Similarly, $(V', \psi', \mathbf{s}') \in [(V, \psi, \mathbf{s})]$ iff (V', ψ') is a chart on M with $f(u) \in V'$, $\mathbf{s}, \mathbf{s}' \in S$, and

$$\left(\left(\hat{\mathbf{D}}\left(\psi' \circ \psi^{\leftarrow}\right)\right)(\psi(f(u)))\right)(\mathbf{s}) = \mathbf{s}'.$$

Comment. *Exercise VII.I.6 in (Dodson and Poston, 1991) shows that $\mathbf{D}_u f$ is well-defined. That is, the definition does not depend on the specific choice of charts (U, ϕ) and (V, ψ). Essentially, this requires one to show that if (U, ϕ) and (U', ϕ') are charts on M with $u \in U \cap U'$, (V, ψ) and (V', ψ') are charts on N with $f(u) \in V \cap V'$, $\mathbf{t}, \mathbf{t}' \in T$, and $(U, \phi, \mathbf{t}) \sim (U', \phi', \mathbf{t}')$, then,*

$$\left(V, \psi, \mathbf{D}_{\phi(u)}\left(\phi \circ f \circ \phi^{\leftarrow}\right)\mathbf{t}\right) \sim \left(V', \psi', \mathbf{D}_{\phi'(u)}\left(\phi' \circ f \circ \phi'^{\leftarrow}\right)\mathbf{t}'\right).$$

Comment. *If we fix one chart for M and another for N, then the derivative $\mathbf{D}_u f$ maps a tangent vector in the affine space underlying M (corresponding to a tangent vector in M, an equivalence class) to a tangent vector in the affine space underlying N (corresponding to a tangent vector in N, an equivalence class). Picking specific charts (like picking basis or coordinate system) is innocuous as other chart choices will change vectors in affine spaces but they will still correspond to the same tangent vector equivalence classes in manifolds.*

7.3 Bundles and Fields

Result 26. *Suppose M is an n-manifold modeled on an affine space X with atlas $\{(U_a, \phi_a) | a \in A\}$. Let $TM = \cup_{u \in M} T_u M$ and $TU_a = \cup_{u \in U_a} T_u M \subseteq TM$. Then, $\{(TU_a, \mathbf{D}\phi_a) | a \in A\}$ is an atlas making TM a $2n$-manifold modeled on $X \times X$, where $\mathbf{D}\phi_a : TU_a \to TX$ is defined by $\mathbf{D}\phi_a|_{T_u M} = \mathbf{D}_u \phi_a$.*

Comment. *For each $a \in A$, the atlas $\{(TU_a, \mathbf{D}\phi_a)\}$ takes a bound tangent vector on manifold from the set TU_a to a bound tangent vector in $X \times X$ (specifying the starting and the ending points of the vector). Since X is an affine space of dimension n, $X \times X$ is an affine space of dimension $2n$, so TM is a $2n$-manifold. The map $\mathbf{D}\phi_a$ is defined to be between manifolds and is constructed by using maps $\mathbf{D}_x \phi_a$ between vector spaces at each point x.*

The derivative of a map from one affine space to another was defined in Definition 59 and the derivative of a map from one manifold to another was defined in Definition 68. $\mathbf{D}_u\phi_a$ is the derivative of a map ϕ_a from a manifold to an affine space, which is a trivial manifold.

Proof. TM is a manifold if it satisfies the three conditions in Definition 62. First, the domains TU_a of all charts cover TM. This follows from $TM = \cup_{u\in M} T_uM$.
Second, each chart $\mathbf{D}\phi_a$ defines a homeomorphism. Assign a topology to TM with open sets consisting of finite intersections and arbitrary unions of sets of form $(\mathbf{D}\phi_a)^{\leftarrow}(W)$ where $a \in A$ and W is an open set in $X \times X$. This ensures all $\mathbf{D}\phi_a$'s are continuous (see Result 10). $\mathbf{D}\phi_a$ defines a homeomorphism (Definition 47) because

1. $\mathbf{D}\phi_a$ is injective: Consider $\mathbf{t}, \mathbf{t}' \in TU_a$ and $\mathbf{t} \neq \mathbf{t}'$.

   (a) If $\mathbf{t} \in T_uM, u \in U_a, \mathbf{t}' \notin T_uM$, then $(\mathbf{D}\phi_a)(\mathbf{t}) \neq (\mathbf{D}\phi_a)(\mathbf{t}')$ because $(\mathbf{D}\phi_a)(\mathbf{t}) \in T_{\phi_a(u)}X$ but $(\mathbf{D}\phi_a)(\mathbf{t}') \notin T_{\phi_a(u)}X$.

   (b) If $\mathbf{t}, \mathbf{t}' \in T_uM, u \in U_a$, then $(\mathbf{D}\phi_a)(\mathbf{t}) \neq (\mathbf{D}\phi_a)(\mathbf{t}')$ because $\mathbf{D}\phi_a|_{T_uM}$, defined as $\mathbf{D}_u\phi_a$ is an isomorphism and hence injective. ϕ_a is a homeomorphism (see Definition 62) but why is $\mathbf{D}_u\phi_a$ an isomorphism? From Definition 68, once a chart is fixed, $\mathbf{D}_u\phi_a$ is a linear map from tangent vectors in X (corresponding to equivalence classes in T_uM) to tangent vectors in X. Fixing that chart to be (U_a, ϕ_a), $\mathbf{D}_u\phi_a$ is an identity map, and hence, an isomorphism.

2. Similarly, $\mathbf{D}\phi_a$ is surjective because $\mathbf{D}_u\phi_a$ is an isomorphism.

3. $\mathbf{D}\phi_a$ is continuous (by the choice made above of the topology on TM).

4. $(\mathbf{D}\phi_a)^{\leftarrow}$ exists (because $\mathbf{D}\phi_a$ is injective) and is continuous (see Result 10) if for any open set U in TM, $((\mathbf{D}\phi_a)^{\leftarrow})^{\leftarrow}(U)$ is open in $X \times X$. This holds if it holds for all U of the form $(\mathbf{D}\phi_b)^{\leftarrow}(W)$ because open sets in TM are finite intersections and arbitrary unions of sets of the form $(\mathbf{D}\phi_b)^{\leftarrow}(W)$. Thus, $(\mathbf{D}\phi_a)^{\leftarrow}$ is continuous if $((\mathbf{D}\phi_a)^{\leftarrow})^{\leftarrow}((\mathbf{D}\phi_b)^{\leftarrow}(W))$ is open in $X \times X$ for any open set W in $X \times X$. But

$$((\mathbf{D}\phi_a)^{\leftarrow})^{\leftarrow}((\mathbf{D}\phi_b)^{\leftarrow}(W)) = (\mathbf{D}\phi_b \circ (\mathbf{D}\phi_a)^{\leftarrow})^{\leftarrow}(W)$$

   because if $\mathbf{y} \in ((\mathbf{D}\phi_a)^{\leftarrow})^{\leftarrow}((\mathbf{D}\phi_b)^{\leftarrow}(W))$, then $(\mathbf{D}\phi_b \circ (\mathbf{D}\phi_a)^{\leftarrow})(\mathbf{y}) \in W$ and if $\mathbf{y} \in (\mathbf{D}\phi_b \circ (\mathbf{D}\phi_a)^{\leftarrow})^{\leftarrow}(W)$, then $(\mathbf{D}\phi_a)^{\leftarrow}(\mathbf{y}) \in ((\mathbf{D}\phi_b)^{\leftarrow}(W))$.

Thus, $(\mathbf{D}\phi_a)^{\leftarrow}$ is continuous if $\left(\mathbf{D}\phi_b \circ (\mathbf{D}\phi_a)^{\leftarrow}\right)^{\leftarrow}(W)$ is open in $X \times X$ for any open set W in $X \times X$, or equivalently by Result 10, if $\mathbf{D}\phi_b \circ (\mathbf{D}\phi_a)^{\leftarrow}$ is continuous. But by chain rule, $\mathbf{D}\phi_b \circ (\mathbf{D}\phi_a)^{\leftarrow} = \mathbf{D}\left(\phi_b \circ \phi_a^{\leftarrow}\right)$, which is continuous because $\phi_b \circ \phi_a^{\leftarrow}$ is continuously differentiable by the definition of charts in a manifold.

Third, for $a, b \in A$, the composite map $(\mathbf{D}\phi_b)(\mathbf{D}\phi_b)^{\leftarrow}$ is continuously differentiable (and similarly for higher order derivatives). The composite map takes bound tangent vectors in X to bound tangent vectors in X. That is, points in $X \times X$ to $X \times X$. We now convert these bound vectors to a pairs of starting points and free vectors, thus converting the composite map to a map from points in $X \times T$ to points in $X \times T$. For this, we define a map converting a bound vector to a pair of starting point and a free vector:

$$A : X \times X \to X \times T : (x, y) \mapsto (x, \mathbf{d}(x, y))$$

where $\mathbf{d}$ is the distance function in X. Now consider this modified composite map:

$$\begin{aligned} A \circ \left(\mathbf{D}\left(\phi_b \circ \phi_a^{\leftarrow}\right)\right) \circ A^{\leftarrow} &: \phi_a(U_a \cap U_b) \times T \to \phi_b(U_a \cap U_b) \times T \\ &(q, \mathbf{t}) \mapsto \left(\left(\phi_b \circ \phi_a^{\leftarrow}\right) q, \hat{\mathbf{D}}\left(\phi_b \circ \phi_a^{\leftarrow}\right) \mathbf{t}\right) = (p, \mathbf{s}). \end{aligned}$$

The domain and range of this modified composite map are direct sum (see Dodson and Poston, 1991, p. 180) $X \oplus T$ of affine space X and tangent space T. The derivative of this map at $(q, \mathbf{t})$ is

$$\mathbf{D}_q\left(\phi_b \circ \phi_a^{\leftarrow}\right) \oplus \mathbf{D}_{\mathbf{t}}\left(\mathbf{D}_q\left(\phi_b \circ \phi_a^{\leftarrow}\right)\right) : T_qX \oplus T_{\mathbf{t}}\left(T_qX\right) \to T_pX \oplus T_{\mathbf{s}}\left(T_pX\right).$$

However, $\mathbf{D}_q\left(\phi_b \circ \phi_a^{\leftarrow}\right)$ is linear (all derivatives are) so its derivative is a constant and doesn't depend on $\mathbf{t}$. Thus, the derivative of the modified composite map can be "identified with" (is isomorphic to?) with

$$\mathbf{D}_q\left(\phi_b \circ \phi_a^{\leftarrow}\right) \oplus \mathbf{D}_q\left(\phi_b \circ \phi_a^{\leftarrow}\right) : T_qX \oplus T_qX \to T_pX \oplus T_pX.$$

Now going back to the original composite map, by Chain Rule,

$$\begin{aligned} \mathbf{D}_{(q,\mathbf{t})}\left((\mathbf{D}\phi_b) \circ (\mathbf{D}\phi_a)^{\leftarrow}\right) &= \mathbf{D}_{(q,\mathbf{t})}\left(\mathbf{D}(\phi_b \circ \phi_a^{\leftarrow})\right) \\ &= \mathbf{D}_{(q,\mathbf{t})}\left(A^{\leftarrow} \circ \left(A \circ \mathbf{D}(\phi_b \circ \phi_a^{\leftarrow}) \circ A^{\leftarrow}\right) \circ A\right) \\ &= \mathbf{A}^{\leftarrow} \circ \left(\mathbf{D}_q\left(\phi_b \circ \phi_a^{\leftarrow}\right) \oplus \mathbf{D}_q\left(\phi_b \circ \phi_a^{\leftarrow}\right)\right) \circ \mathbf{A} \end{aligned}$$

which exists and is continuous because $\phi_b \circ \phi_a^{\leftarrow}$ is continuously differentiable by the definition of charts in a manifold. □

Example 16. *Consider the manifold S^2 with its associated charts in Example 15. It is embedded in $\mathbb{R}^3$ with inclusion $\iota : S^2 \hookrightarrow \mathbb{R}^n : (u^1, u^2, u^3) \mapsto (u^1, u^2, u^3)$. Consider $u = (u^1, u^2, u^3) \in S^2$ with $u^1, u^2, u^3 > 0$. Let us use the notation $\overrightarrow{(v^1, v^2, v^3)}$ to indicate the vector from $(0, 0, 0)$ to (v^1, v^2, v^3). A vector $\mathbf{v} = \overrightarrow{(v^1, v^2, v^3)}$ with $v^1, v^2, v^3 \in \mathbb{R}$ is tangent to S^2 at u in $\mathbb{R}^3$ if $u^1v^1 + u^2v^2 + u^3v^3 = 0$. By Result 24, each such $\mathbf{v}$ represents (an equivalence class corresponding to) a tangent vector in T_uS^2, a tangent to S^2 at u. Denote this by $[\mathbf{v}]$. Consider $(U_{1+}, \phi_{1+}, \mathbf{t}_{1+}) \in [\mathbf{v}]$. This requires $((\hat{\mathbf{D}}(\iota \circ \phi_{1+}^{\leftarrow}))(\phi_{1+}(u)))(\mathbf{t}_{1+}) = \mathbf{v}$. Since $\iota \circ \phi_{1+}^{\leftarrow} : \overrightarrow{(x^2, x^3)} \mapsto (\sqrt{1 - (x^2)^2 - (x^3)^2}, x^2, x^3)$, we have*

$$
\begin{aligned}
&(\hat{\mathbf{D}}\iota \circ \phi_{1+}^{\leftarrow}((u^2, u^3)))(\overrightarrow{(w^2, w^3)}) \\
&= \lim_{h \to 0} \frac{1}{h}[(\sqrt{1 - (u^2 + hw^2)^2 - (u^3 + hw^3)^2}, u^2 + hw^2, u^3 + hw^3) \\
&- (\sqrt{1 - (u^2)^2 - (u^3)^2}, u^2, u^3)] \\
&= (\frac{-u^2w^2 - u^3w^3}{\sqrt{1 - (u^2)^2 - (u^3)^2}}, w^2, w^3) = (\frac{-u^2w^2 - u^3w^3}{u^1}, w^2, w^3).
\end{aligned}
$$

Then, $(U_{1+}, \phi_{1+}, \mathbf{t}_{1+}) \in [\mathbf{v}]$ with $\mathbf{t}_{1+} = \overrightarrow{(w^2, w^3)}$ if $(\frac{-u^2w^2 - u^3w^3}{u^1}, w^2, w^3) = (v^1, v^2, v^3)$. This is true if $(w^2, w^3) = (v^2, v^3)$. Thus, $(U_{1+}, \phi_{1+}, \overrightarrow{(v^2, v^3)}) \in [\mathbf{v}]$. Similarly, $(U_{2+}, \phi_{2+}, \overrightarrow{(v^1, v^3)}) \in [\mathbf{v}]$ and $(U_{3+}, \phi_{3+}, \overrightarrow{(v^1, v^2)}) \in [\mathbf{v}]$. The following figure shows a red tangent vector to S^2 in $\mathbb{R}^3$ and its projections (in red) under ϕ_{1+}, ϕ_{2+}, and ϕ_{3+} on three planes as the members of the corresponding tangent vector equivalence class. The tangent space to S^2 at u is the set of all such $[\mathbf{v}]$:

$$T_uS^2 = \left\{ [\mathbf{v}] \middle| \mathbf{v} = \overrightarrow{(v^1, v^2, v^3)}, v^1, v^2, v^3 \in \mathbb{R}, u^1v^1 + u^2v^2 + u^3v^3 = 0 \right\}.$$

The red and blue vectors in the following figure are part of the tangent space to S^2 at the same point. Finally, aggregating T_uS^2 over all $u \in S^2$, we get

$$
\begin{aligned}
&TS^2 = \cup_{u \in S^2} T_uS^2 = \\
&\left\{ \left(\left(u^1, u^2, u^3\right), \left[\overrightarrow{(v^1, v^2, v^3)}\right] \right) \right| \\
&\left. u^1, u^2, u^3, v^1, v^2, v^3 \in \mathbb{R}, (u^1)^2 + (u^2)^2 + (u^3)^2 = 1, u^1v^1 + u^2v^2 + u^3v^3 = 0 \right\}.
\end{aligned}
$$

The red, blue, pink, and orange vectors in the following figure are part of TS^2. TS^2 *is a manifold with atlas* $\{(TU_a, \mathbf{D}\phi_a) | a \in \{1+, 1-, 2+, 2-, 3+, 3-\}\}$. *One chart in this atlas is*

$$\mathbf{D}\phi_{1+} : TU_{1+} \to T\mathbb{R}^2 : \left(\left(u^1, u^2, u^3\right), \overrightarrow{\left(v^1, v^2, v^3\right)} \right) \mapsto \left(\left(u^2, u^3\right), \overrightarrow{\left(v^2, v^3\right)} \right)$$

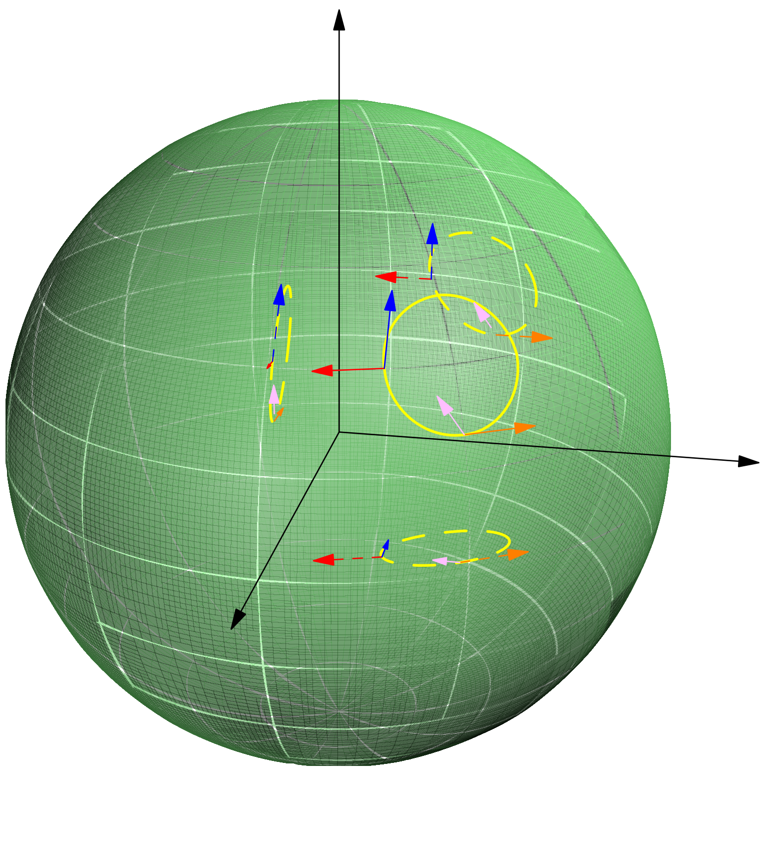
where $(u^1)^2 + (u^2)^2 + (u^3)^2 = 1$ *and* $u^1v^1 + u^2v^2 + u^3v^3 = 0$. *To see that* TS^2 *is 4-dimensional, note that*

$$\{((u^1, u^2, u^3) | u^1, u^2, u^3 \in \mathbb{R}, (u^1)^2 + (u^2)^2 + (u^3)^2 = 1\}$$

is 2-dimensional, and for a given (u^1, u^2, u^3),

$$\{(v^1, v^2, v^3)) \,| v^1, v^2, v^3 \in \mathbb{R}, u^1v^1 + u^2v^2 + u^3v^3 = 0\}$$

is also 2-dimensional.



Index

References

C.T.J. Dodson and T. Poston. *Tensor Geometry*. Springer-Verlag, 2nd edition, 1991.