Differential Geoemtry

Hasan Poonawala

Contents

1 Manifolds

The essence of a manifold, a type of set, is that you are unable to use a single system of numbers to consistently describe the whole manifold. You start by choosing one system of numbers (coordinates), but are then forced to switch to a different system of numbers for describing the manifold.

Example 1. We use latitude $\phi \in [-90^\circ, 90^\circ]$ and longitude $\theta \in (-180^\circ, 180^\circ]$ to describe locations of points on the Earth, like in GPS systems. One issue is that the North Pole, a single point, has latitude 90° but an infinite set of possible longitude values. Another way to think about this is that points with the same latitude correspond to a circle, except for $\phi = \pm 90^\circ$, where the circle at other latitudes – infinite points – reduces to point as $\phi \to \pm 90^\circ$.

One consequence of this issue is that using the change in latitude and longitude as a proxy for measuring distance traveled on the Earth's surface works much better at the Equator than at the poles. If you had to choose, you would rather 'circle the world by traveling East-West' at the poles than at the Equator.

Even if you stay within a limited region of a manifold, in many applications, you may need to deal with multiple coordinate systems and enforce consistency between coordinate descriptions. For example, you might want to combine multiple LIDAR sensor readings taken by a robot moving in a room. The room doesn't change with the robot's pose, but the spherical coordinates assigned to things in the room by using the LIDAR sensor depend on the robot's pose, since the LIDAR readings are relative to the robot's frame, not the room's. To combine these LIDAR readings into a map of the room requires enforcing consistency between the multiple descriptions of things in the room. When we describe motion on manifolds, we also need to develop notions of velocities and accelerations that are consistent with these coordinates, and transformations between coordinates.

Unfortunately, the price of all this consistency is a steep learning curve.

Definition 1 (Topological Manifold)**.** A manifold M (manifold) is a second-countable Hausdorff topological space that is locally homeomorphic to Euclidean space $\mathbb{R}^m$. The dimension of the manifold becomes m.

A space is second-countable if every cover has a finite sub-cover. A space is Hausdorff if for any pair of points we can find two mutually disjoint sets that contain only one of the pair.

Definition 2 (Differentiable Manifold)**.** A manifold M (manifold) is a second-countable Hausdorff topological space that is locally diffeomorphic to Euclidean space $\mathbb{R}^m$.

A manifold Q is a second-countable Hausdorff space with a set of compatible charts that cover Q. The charts are local diffeomorphisms to $\mathbb{R}^n$, the compatibility makes the charts an atlas. One either uses chart ϕ to map an open set of Q to its coordinates in $\mathbb{R}^n$, or to assign an open set of points on the manifold to an open set of $\mathbb{R}^n$, which effectively defines coordinates.

Example 2 (Euclidean Space $\mathbb{R}^n$)**.** Every finite-dimensional Euclidean space $\mathbb{R}^n$ is a topological manifold. The identity map is a trivial homeomorphism (and diffeomorphism) mapping this manifold to $\mathbb{R}^n$, that is, $m = n$.

1.1 Coordinates

The notion of the differentiable manifold being locally diffeomorphic to $\mathbb{R}^m$ means that there is a differentiable bijective map $\varphi\colon U \mapsto \mathbb{R}^m$ where $U \subset M$ is an open subset of M. Since the range space of φ is $\mathbb{R}^m$, the diffeomorphism φ is assigning m-dimensional coordinates to points on the manifold. These m-dimensional coordinates are sometimes referred to as **intrinsic** coordinates. We can perform operations on points in U through operations on their intrinsic coordinates. This indirect operation makes sense precisely because φ is smooth and bijective.

Example 3 (Coordinates for Euclidean Space)**.** Recall that we can describe points in Euclidean space without using coordinates, but to perform computations we assigned coordinates by choosing a frame. That frame can be chosen in many ways, with consistency achieved using homogenous transformations. Note that since we assign coordinates to points in a Euclidean space, a single frame assigns unique coordinates to all points in Euclidean space. By contrast, we can't assign a unique latitude and longitude to all points on the Earth.

Manifolds locally look like Euclidean space $\mathbb{R}^m$, but they do not behave like $\mathbb{R}^m$ in a global sense. One consequence of this behavior is that one cannot always assign global coordinates to a manifold, that is, one cannot assign a unique vector $\mathbb{R}^m$ to every point of M and still perform meaningful calculations. The lack of global coordinates is often why we consider embedded manifolds, where we can use the coordinates $\mathbb{R}^n$ together with the constraints h_i to perform calculations.

1.2 Embedded Manifolds

For our purposes, we consider manifolds that are embedded submanifolds. We use l constraints on $\mathbb{R}^n$ to specify M, where $l < n$. Each constraint is represented by a smooth function $h_i\colon \mathbb{R}^n \to \mathbb{R}$, so that

$$M = \{x \in \mathbb{R}^n \colon h_i(x) = 0, \forall i \in \{1, \ldots, l\}\} \tag{1}$$

If a point $x \in \mathbb{R}^n$ belongs to the manifold M, then x is said to be the **extrinsic** coordinates of that point on the manifold. The manifold M is said to be embedded in $\mathbb{R}^n$.

Whitney Embedding Theorem An important question is whether there are manifolds for which it is impossible to define extrinsic coordinates, meaning that these manifolds have a geometry so complex that they cannot be embedded into any $\mathbb{R}^n$. The answer is that it is always possible to embed a smooth manifold. Therefore, we can always talk in terms of both extrinsic and intrinsic coordinates for a smooth manifold.

Example 4 (Sphere)**.** We can define the surface $\mathcal{S}^2$ of a sphere of radius 1 as an embedded submanifold of $\mathbb{R}^3$ using the constraint $x^2 + y^2 + z^2 = 1$. That is,

$$\mathcal{S}^2 = \{(x, y, z) \in \mathbb{R}^3 \colon x^2 + y^2 + z^2 - 1 = 0\}$$

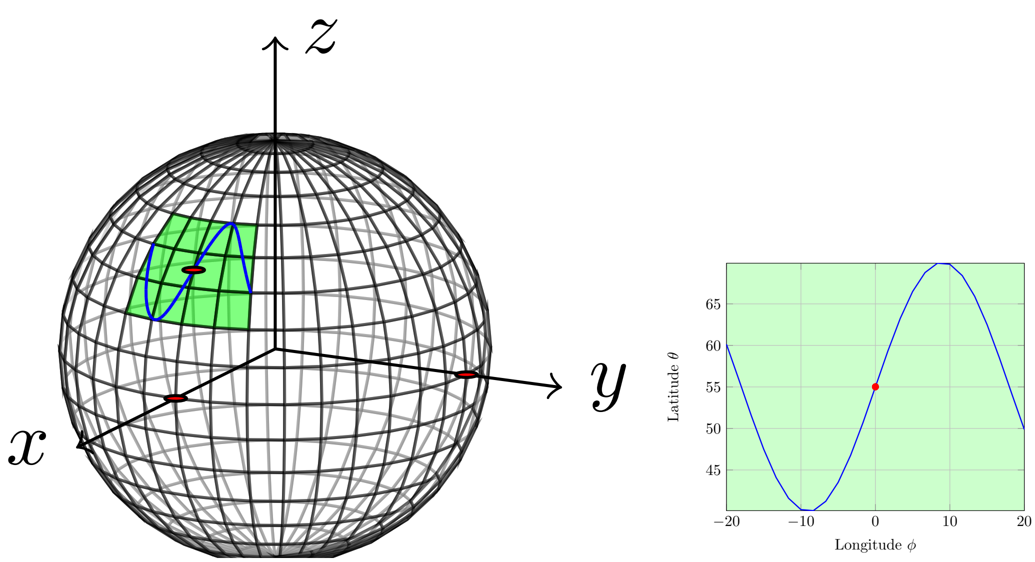

Figure 1: The surface of the sphere (left) can be mapped to $\mathbb{R}^2$ (right) where the x and y axes correspond to longitude and latitude respectively. Since this mapping is a diffeomorphism, the inverse map exists.

Example 5 (Coordinates for $\mathcal{S}^2$)**.** Just as points in Euclidean space can be assigned different coordinates depending on how we define a reference frame, we may be able to assign multiple coordinates to points on a Manifold. The main difference is that these coordinates may only work for a subset of the manifold, unlike Euclidean space.

One example of coordinates for the surface $\mathcal{S}^2$ of a sphere is the longitude ϕ and latitude θ with respect to a point $p^* \in M$ corresponding to $(\phi, \theta) = (0, 0)$ and a great circle ($\theta = 0$) passing through p^*. We can define the map $\varphi: \mathcal{S}^2 \mapsto \mathbb{R}^3$ as

$$x = \cos\theta\cos\phi \tag{2}$$
$$y = \cos\theta\sin\phi \tag{3}$$
$$z = \sin\theta \tag{4}$$

As required, φ is a smooth map. In fact, it is smooth for all values of $\mathbb{R}^2$, but is not injective (one-to-one) for all $\mathbb{R}^2$. To make sure that φ is a diffeomorphism, we define its domain as $-\pi/2 < \theta < \pi/2$ and $-\pi < \phi \leq \pi$. One can check that indeed $x^2 + y^2 + z^2 = 1$ for all extrinsic coordinates (x, y, z) corresponding to intrinsic coordinates (θ, ϕ).

1.3 Fibre Bundles

A fibre bundle is a space M that is locally a product space, but not globally so. That is, we can decompose M locally into the Carteisan product $B \times F$, where B is the base space and F is the fibre space.

> Specifically, the similarity between a space E and a product space $B \times F$ is defined using a continuous surjective map $\pi: E \to B$ The space E is known as the total space of the fiber bundle, B as the base space, and F the fiber.
>
> Mappings between total spaces of fiber bundles that "commute" with the projection maps are known as bundle maps. A bundle map from the base space itself (with the identity mapping as projection) to E is called a section of E. Fiber bundles can be specialized in a number of ways, the most common of which is requiring that the transition maps between the local trivial patches lie in a certain topological group, known as the structure group, acting on the fiber F.

A non-trivial fibre bundle is the decomposition of the Mobius strip into the axial and transverse 'manifolds'. Locally, the bundle is the product $\mathbb{R} \times \mathbb{R}$, but globally not.

Two important types of fibre bundles are vector bundles and principal bundles. In vector bundles, the fibres are vector spaces, for example the tangent bundle. In principal bundles, the fibres are groups. For example, the set of reference frames for the tangent bundle. These two examples shows the concept of associated bundles, since the reference frame for each fibre of the vector bundle often belongs to a group.

A common application of fibre bundles arises when the base space is the manifold. An example is the tangent bundle TM of a manifold M, which is itself a manifold. Here, the base space is the manifold M and the fibre at $x \in M$ is the tangent space T_xM. It is precisely the tangent bundle that is most relevant to us.

1.3.1 Vector Bundles

The fibres are vector spaces, with the additional requirement that the structure group be a linear group. For our purposes, we will deal with tangent bundles, which are examples of vector bundles. To each $x \in M$ we associate a Euclidean vector space with the same dimensions as M.

1.3.2 Principal Bundles

A principal bundle P is a manifold that formalizes the product of $M \times G$, where G is a group. In our applications, the principal bundle will consist of products between a manifold and the set of bases we can assign to the tangent spaces to the manifold.

On each fibre we must have a free and transitive action given by group G, so that each fibre must be a principal homogenous space. We must have

1. An action by G on P, so that $(x, g)h = (x, gh)$
2. A projection from P onto M

The bundle is often specified along with the group by referring to it as a principal G-bundle. The group G is also the structure group of the bundle.

2 Vector Spaces, Bases, Coordinates

Before we proceed to working with tangent vectors/spaces/bundles, we must review the basics of vector spaces. See Appendix A for a complete development of a vector space from first principles. The main point here is that a basis $\mathcal{B}$ for an n-dimensional vector space V is a collection of n independent vectors $\{e_1, e_2, \dots, e_n\}$, each in V, that span the vector space. This spanning property of the basis means that any vector $v \in V$ is a linear combination of the basis vectos with coefficients $(\alpha^1, \alpha^2, \dots, \alpha^n)$.

These coefficients are typically referred to as the components or coordinates of a vector. Formally, v is a symbol, while $[v]_{\mathcal{B}} = (\alpha^1, \alpha^2, \dots, \alpha^n)$ refers to a *representation* of v with respect to $\mathcal{B}$. However, when the basis is obvious, we refer to the representation also through symbol v. Therefore, we often see the expression

$$\underbrace{v}_{\text{formally } [v]_{\mathcal{B}}} = \begin{bmatrix} \alpha^1 \\ \alpha^2 \\ \dots \\ \alpha^n \end{bmatrix}. \tag{5}$$

When the basis $\mathcal{B}$ is the standard one, this simpler notation works well enough, which is why it is taught as such in introductory classes.

2.1 Covariant and Contravariant Coordinates of A Vector

The standard representation $[v]_{\mathcal{B}}$ of a vector v with respect to a basis $\mathcal{B}$, also called the coordinates of v, are an example of contravariant coordinates. If we define a new basis $\mathcal{B}'$, whose elements create a matrix representation T with respect to $\mathcal{B}$, then the coordinates with respect to $\mathcal{B}'$ are

$$[v]_{\mathcal{B}'} = T^{-1}[v]_{\mathcal{B}}. \tag{6}$$

Let's say that the basis elements in $\mathcal{B}$ have some representation with respect to a third basis $\mathcal{B}_0$. The basis vector $e_i' \in \mathcal{B}'$ may be represented in $\mathcal{B}_0$ by multiplying the representation of the elements in $\mathcal{B}$ by T:

$$[e_i']_{\mathcal{B}_0} = T[e_i]_{\mathcal{B}_0}. \tag{7}$$

The fact that the correct transformation for the vector coordinates $[v]_{\mathcal{B}}$ are in opposition to the correct transformation for $[e_i]_{\mathcal{B}_0}$ are why the vector coordinates are called contravariant. This contravariance is also why we denote the indices of the elements of $[v]_{\mathcal{B}}$ as superscripts, instead of the more familiar subscripts.

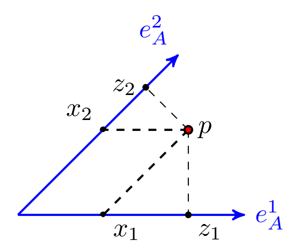

Figure 2: The basis A (blue) may be used to define two sets of coordinates for p. The coordinates (x^1, x^2) are contravariant, and the coordinates (z_1, z_2) are covariant.

A geometric way to calculate these contravariant coordinates is to draw lines parallel to the coordinate axes, as seen in Figure 2, and obtain the length of the intersection onto the coordinate axis. In Figure 2, the contravariant (and standard) coordinates are $[v]_{\mathcal{B}} = (x^1, x^2)$. These parallel lines are actually orthogonal with respect to the basis $\mathcal{B}$, and so in a sense this operation is still an orthogonal projection, but with respect to the potentially non-Cartesian basis $\mathcal{B}$.

Instead of orthogonal projections with respect to $\mathcal{B}$, that leads to contravariant coordinates, we may use orthogonal projections where orthogonality is in a Cartesian sense. In Figure 2, this Cartesian-orthogonal projection leads to coordinates (z_1, z_2). These coordinates are *covariant*, and by convention their indices are subscripts. Note that $v \neq z_1 e_A^1 + z_2 e_A^2$

2.2 Reciprocal Basis

An important idea is that we may use the basis of a vector space to define a basis for its dual space, through a reciprocal basis. The reciprocal basis is one way to bring a notion of Cartesian orthogonality into a basis that is not Cartesian.

Let $B_1 = \{e_1, e_2, \ldots, e_n$ be a basis for a vector space V, and B_2 be the collection of vectors $\{f_1, f_2, \ldots, f_n\}$ satisfying

$$e_i \cdot f_j = \underbrace{\delta_{ij}}_{\text{Kronecker delta}} = \begin{cases} 1 & \text{, if } i = j, \\ 0 & \text{, otherwise.} \end{cases} \tag{8}$$

Note that the dot product $\langle \cdot \rangle$ here is defined with respect to a Cartesian basis. Figure 3 illustrates a reciprocal basis.

If we associate the basis A with an $n \times n$ matrix M in some Cartesian basis, then the reciprocal basis has coordinates $(M^{-1})^T$ in that Cartesian basis. A unitary matrix defines a basis that is its own reciprocal basis.

The significance of the reciprocal basis is that it formalizes how to define subspaces orthogonal (relative to a cartesian basis) to other subspaces. Therefore, we may modify a vector without changing its Euclidean projection onto a subspace of interest. Equivalently, the closest point in a subspace won't change (but the distance to that closest point changes).

Formally, Consider a basis B, and a partition of the basis into B_S and $B \backslash B_S$. If B' is the reciprocal basis of B, then the partition B_S of B induces a partition of B' into B'_S and $B \backslash B'_S$/ Finally, $v \cdot w = 0$ for any $v \in \text{Span}(B_S)$ and $w \in \text{Span}(B \backslash B'_S)$

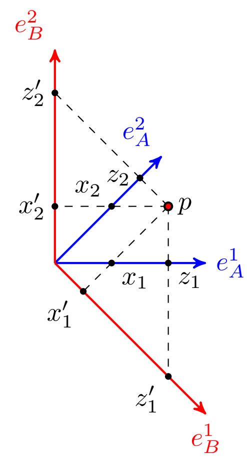

Figure 3: The reciprocal basis A (blue) may be used to define another sets of contravariant coordinates for p, and covariant coordinates for p. Additionally, moving along e_B^1 keeps the projection of p onto e_A^2 (second covariant coordinate) constant. Or, moving along e_A^1 only changes the projection along the e_B^1.

3 Tangent And Cotangent Space

Recall that we use l constraints on $\mathbb{R}^n$ to specify a manifold M, where $l < n$. Each constraint is represented by a smooth function $h_i \colon \mathbb{R}^n \to \mathbb{R}$.

We assume that the l differentials dh_i are linearly independent at each point $x \in M$. In this case, the dimension of the manifold is $m = n - l$.

3.1 Tangent Space

To an m-dimensional manifold, we can assign a tangent space $T_x M$ at each $x \in M$ which is an m-dimensional vector space specifying the set of instantaneous velocities possible at x. For an embedded manifold in $\mathbb{R}^n$, the tangent space $T_x M$ is an affine subspace of $\mathbb{R}^n$, where the origin of $T_x M$ corresponds to x.

Example 6 (Tangent Space to $\mathbb{R}^n$)**.** The tangent space $T_p\mathbb{R}^n$ to a point $p \in \mathbb{R}^n$ is itself another copy of $\mathbb{R}^n$. That is, the possible velocities for a point $p \in \mathbb{R}^n$ creates a space that similar to $\mathbb{R}^n$. The fact that Euclidean space and its tangent space is the same is the reason that one often confuses point vectors (which are coordinates) with velocity vectors. The fact that the tangent space at each point in a Euclidean space is the same ($\mathbb{R}^n$) is what allows one to treat velocities as free vectors.

We can compute the tangent space as follows: Consider all curves in passing through a point $q \in \mathbb{R}^m$. Each curve γ is a one-dimensional set parametrized by a parameter, say t, belonging to a range, where $\gamma(0) = q$. The derivative $\left.\frac{\partial \gamma}{\partial t}\right|_{t=0}$ defines a vector tangent to the curve at q, and the collection of all such vectors at q (corresponding to all possible curves through q) forms a vector space at q. When these curves are defined in intrinsic coordinates, the tangent space will turn out to be $\mathbb{R}^m$. When the curves through q are defined in extrinsic coordinates, the tangent space turns out to be a hyperplane tangent to the manifold at p, where $p = \varphi(q)$.

In fact, this process gives an explicit way to assign coordinates to the hyperplane *using the intrinsic coordinates*. Specifically, we can consider a curve through q in $\mathbb{R}^m$ whose tangent is parallel to a coordinate axis of $\mathbb{R}^m$. We can map this curve to its extrinsic coordinates in $M \subseteq \mathbb{R}^n$, and find the derivative of the resulting curve, which gives us a vector in $\mathbb{R}^n$. The derivatives of these two curves (when seen as a curve in extrinsic and intrinsic coordinates) are related through the derivative of the diffeomorphism φ. That is, if $v \in T_x\mathbb{R}^m$ is the velocity at x in intrinsic coordinates, the corresponding derivative u of the curve in extrinsic coordinates is

$$u = \frac{\partial \varphi}{\partial q} v. \tag{9}$$

Therefore, given φ, we can use a basis for $\mathbb{R}^m$ (which serves to represent $T_q\mathbb{R}^m$) to generate a basis for T_pM. That is, if $e_1, \ldots, e_m$ is a basis for $\mathbb{R}^m$, then $\frac{\partial \varphi}{\partial q}e_1, \ldots, \frac{\partial \varphi}{\partial x}e_m$ forms a basis for T_pM.

Example 7 (Tangent Space to Sphere)**.** We have the map $\varphi\colon \mathbb{R}^m \mapsto U \subset M$. The partial derivative of φ with respect to (θ, ϕ) at $q = (\theta, \phi)$ is

$$D\varphi = \begin{bmatrix} -\cos\theta\sin\phi & -\sin\theta\cos\phi \\ \cos\theta\cos\phi & -\sin\theta\sin\phi \\ 0 & \cos\theta \end{bmatrix} \tag{10}$$

The tangent space at a point $p \in M \subseteq \mathbb{R}^3$ is simply the span of the columns of $D\varphi$ in (10) with the values $(\phi, \theta) = q$ given by $\varphi^{-1}(p)$.

Remark 1. Note that the tangent space does not change for different diffeomorphisms φ, only the basis for the tangent space will change. In fact, more formal definitions of a smooth manifold insist that the set of diffeomorphisms that map the same point $p \in M$ to different coordinate spaces $\mathbb{R}^m$ be consistent when it comes to assigning coordinates to their tangent spaces, so as to make the tangent space computations independent of the coordinates (diffeomorphism) used.

3.2 Cotangent Space

We can associate any vector space V with a dual space V^* consisting of the space of linear functionals on V. An element of V^* is a real-valued function of V, and is linear with respect to V. Since T_xM is an m-dimensional vector space we can associate a dual space T_x^*M, called the cotangent space, to it.

A natural basis for T_x^*M is the set of n linear functions whose evaluations of the n-basis vectors of T_xM form the Kronecker delta function. In effect, this natural definition is an analogue to the **reciprocal basis** of a vector space basis. Under the usual outer product, this analogy becomes equivalence.

3.3 Smooth Vector Field

A smooth vector field on a manifold M is a smooth map $f\colon M \to T_xM$. This map is typically represented as a **column** vector of m real-valued functions.

Example 8 (Vector Field on $\mathcal{S}^2$)**.** The first (or second) column of the matrix $D\varphi$ in (10) defines a vector field on $M \subseteq \mathbb{R}^3$.

3.4 Smooth Covector Field

Similarly, a smooth covector field on a manifold M is a smooth map $w\colon M \to T_x^*M$. This map is typically represented as a **row** vector of m real-valued functions.

Example 9 (Covector Field on $\mathcal{S}^2$)**.** The gradient of any smooth scalar function on $\mathcal{S}^2$ defines a covector field on $\mathcal{S}^2$. Taking the function $h(p) = x^2 + y^2 + z^2 - 1$, a covector field on $\mathcal{S}^2$ is $w(p)$ given by

$$w(p) = dh(p) = \begin{bmatrix} 2x & 2y & 2z \end{bmatrix} \tag{11}$$

3.5 Distributions and Codistributions

Let $X_1(x), \ldots, X_k(x)$ be vector fields on M that are linearly independent. A distribution $\Delta(x)$ is the point-wise linear span of these vector fields

$$\Delta = \operatorname{span}\{X_1(x), \ldots, X_k(x)\} \tag{12}$$

This definition ensures that at each $x \in M$, Δ defines a k-dimensional subspace of the tangent space T_xM at x. Similarly, a codistribution $\Omega(x)$ on M is the span of a set of linearly independent covector fields on M.

Example 10 (Tangent space as a Codistribution)**.** If we view each column of the matrix $D\varphi$ in (10) as a vector field, say $X_1(p)$ and $X_2(p)$, then at a given $p \in M$, the tangent space T_pM is precisely the distribution $\Delta = \operatorname{span}\{X_1(p), X_2(p)\}$, that is, $T_pM = \Delta$.

4 Tensors

Our use of tangent spaces will require us to define functions over them. This requirement will make us come across tensors, which we now introduce.

If we take a product space of m vector spaces, and want to map an element of this product space to $\mathbb{R}$, there are many ways to do so. For example, consider $V \times W$, and the map simply the sum of the elements. That is, given $v \in V \subseteq \mathbb{R}^{n_1}$ from the first space and $w \in W \subseteq \mathbb{R}^{n_2}$ from the second space, our map is just

$$T(v, w) = \sum_{i=1}^{n_1} v_i + \sum_{i=1}^{n_2} w_i. \quad \text{(Not a tensor.)} \tag{13}$$

This map is linear in the product space $V \times W$, but only affine in either V or W when the vector from the other space is fixed. We actually want it to be linear in each of the individual spaces, when the vector from the other spaces are fixed. These individually-linear maps from a collection of vector spaces to $\mathbb{R}$ are what tensors are. Naturally, tensors are nonlinear in the product space.

$$T(v, w) = \sum_{i=1}^{\min(n_1, n_2)} v_i w_i. \quad \text{(A tensor.)} \tag{14}$$

4.1 Covariance and Contravariance

A tensor is simply a multilinear map. However, once we decide to represent it using numbers, we have to pick bases, opening a can of worms.

We represent tensors as a collection of coefficients once we define a basis for each vector space. For example, the tensor in (14) may be given a representation

$$T_{\alpha\beta} v^\alpha w^\beta,$$

where we are using the Einstein summation convention, and $T_{\alpha,\beta}$ are the components or coefficient of the tensor T. If $n_1 = n_2 = 2$, then we have four coefficients T_{11}, T_{12}, T_{21}, T_{22} given by

$$T_{11} = 1, \quad T_{12} = T_{21} = 0, \text{ and } \quad T_{22} = 1. \tag{15}$$

Notice the conventions of superscripts and subscripts. The superscripts on v and w indicate that the computation is with respect to contravariant coordinates for V and W in some basis. If we want the tensor computation to be invariant with respect to a change of basis, we must change the coefficients $T_{\alpha\beta}$ *covariantly* with the change of bases for V and W, and therefore we use subscripts for the indices of T according to the convention. We don't have to follow these conventions that result in invariance under change-of-bases, but the tensor becomes an inconsistent way to compute properties of geometric objects without this invariance.

4.2 Tensor order

The dictates of invariance makes us categorize tensors as (m, n)-tensors, where m is the number of contravariant coordinates and n is the number of covariant coordinates. Alternatively, m is the number of arguments that are covariant, and n is the number of arguments that are contravariant. Note that Kreyszig refers to (r, s)-tensors where r is the number of *covariant* coefficients/coordinates of the tensor.

Example 11. A metric tensor is a $(0, 2)$-vector. A common example is the inertia tensor of a mechanical system, and the evaluation of this tensor on two copies of the velocity of the system gives us (twice) the mechanical energy of the system. The representation of this tensor in an appropriate basis is called the mass matrix of the system. Again, the invariance properties of tensors are useful because we don't want the mechanical energy to change when we use different bases to describe the same velocity.

4.3 Dual Vector Spaces

A dual vector or one-form is exactly a $(0, 1)$-tensor, since it is covariant, and maps vectors (contravariant objects) linearly to a real number. The dual space V^* of a vector space is the set of $(0, 1)$-tensors with arguments in V, which is itself a vector space. Dual vectors are also called covectors, and dual spaces covector spaces.

4.3.1 Gradients

The definitions of gradients are almost always introduced using a row vector containing the partial derivatives of f with respect to each coordinate.

$$\nabla f = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}, \tag{16}$$

where we have suppressed the dependence of these terms on x.

Now, if we want the derivative of f along a direction $v \in V$, we take the 'inner product':

$$df(v) = \langle \nabla f, v \rangle = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} v^i \tag{17}$$

A more formal approach is to define the gradient ∇f of a function $f\colon V \to \mathbb{R}$ as a $(0, 1)$-tensor, and so that it is an element of V^*, which makes it a linear function on V. The evaluation $\nabla f(v)$ is the (directional) derivative of f along v. We get linearity in v, but we are also allowed to take linear combinations of elements of V^*. Since we view ∇f as an element of V^*, whose basis is $\{e^1, e^2, \ldots, e^n\}$, our standard representation in (16) implies that

$$\nabla f = \frac{\partial f}{\partial x_1} e^1 + \frac{\partial f}{\partial x_2} e^2 + \cdots + \frac{\partial f}{\partial x_n} e^n \tag{18}$$

The evaluation of v by the $(0, 1)$-tensor ∇f becomes

$$\begin{aligned} \nabla f(v) &= \left(\frac{\partial f}{\partial x_1} e^1 + \frac{\partial f}{\partial x_2} e^2 + \cdots + \frac{\partial f}{\partial x_n} e^n \right) (v^1 e_1 + v^2 e_2 + \cdots + v^n e_n) & (19) \\ &= \frac{\partial f}{\partial x_i} v^j e^i(e_j) & (20) \end{aligned}$$

We see that our standard definition of the directional derivative in (17) only makes sense when the basis of the dual vector space satisfies

$$e^i(e_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

This condition is similar to that of the reciprocal basis condition, however here the bases are for different spaces, namely V and V^*. This detail about the implicit basis for V^* given that of V is left out of most discussions. This detail becomes important when dealing with coordinate changes, which is often not a concern when introducing the gradient. If we are working in a different basis for V^* that is not the standard one derived from V, then we would have to choose a different representation for ∇f. This reciprocal nature of the basis for V^* makes sense, since then moving along e_i should create an effect on $\nabla f(v)$ that is only proportional to $\frac{\partial f}{\partial x_i}$, which is true since the cartesian-orthogonal projection along other dual basis elements remains unchanged.

If gradients are actually $(0, 1)$-tensors, and (16) is only a representation of this tensor in some basis of V^*, then why do gradient descent algorithms update a vector of parameters θ by adding a scaled form of the transpose of this representation to the parameters? The answer is that the vector of a certain magnitude as measured by the standard Euclidean metric that will have the smallest value of $df(v)$ ends up having the same components (with a negative sign) as the gradient, given the related bases for V and V^*. This notion of magnitude comes from a metric tensor. If we chose a different metric, we would get a different update direction.

5 Connections

Suppose we have a function $f\colon M \to Y$, where M is a manifold, and Y is some topological space. Given distinct points $x_1 \in M$ and $x_2 \in M$, how should we compare the values $f(x_1)$ and $f(x_2)$? That is, how do we connect the images from two different points? Mathematically, what topology or algebra should we use for the images at different points on the manifold?

The trivial connection is to just use the topology/algebra given by Y, so that we don't care which points of M produced the elements of Y that we are comparing. For example, if a function assigns a vector in some space V to each point in a manifold, then the difference in the function would be simply $f(x_2) - f(x_1) \in V$. The reason we need non-trivial connections is that the trivial connection is sometimes naive.

One example of needing a non-trivial connection is when we define directional derivatives of vector fields (section of a tangent bundle) along curves on a manifold by directly taking differences between the vectors at two different points, and then proceeding with a limit. The issue is that this directional derivative will then depend on the basis we choose for the tangent spaces, which may depend on the coordinates for the manifold. In effect, we lose geometric behavior (invariance to coordinate transformations) because our method for comparing objects in the image space ignores the geometry of the underlying domain (the manifold).

5.1 Differential Geometry and Connections

There are two uses for connections in differential geometry. One is to move a geometric object at one point to another point locally. For example, parallel transport of a vector along curves. It suffices to define an isomorphism between the image spaces at two different points. w The other is to differentiate geometric objects along curves in the manifold, giving rise to covariant derivatives.

Naturally, these two uses can be related by insisting that the local shifting corresponds to a zero rate of change of a covariant derivative. In general, we see that most definitions enforce this relationship, even though it's not clear that it is required:

> ...the usual notion of connection is the infinitesimal analog of parallel transport. Or, vice versa, parallel transport is the local realization of a connection.
>
> Thus the connection ∇ defines a way of moving elements of the fibers along a curve, and this provides linear isomorphisms between the fibers at points along the curve:
>
> $$\Gamma(\gamma)_s^t = E_{\gamma(s)} \to E_{\gamma(t)} \tag{22}$$
>
> from the vector space lying over $\gamma(s)$ to that over $\gamma(t)$. This **isomorphism** is **known as the parallel transport map** associated to the curve.

Parallel transport can be approximated using Schild's ladder. By the Ambrose-Singer holonomy theorem, the result of parallel transport along closed curves is related to the curvature of connection ∇.

In differential geometry, we always exploit smoothness and differentiability, so that all connections are Ehresmann connections.

5.2 Types of Connections

A function on a manifold defines a fibre bundle. The connections we define depend on the topology of the fibres, and what we need from the connection. In general, we distinguish between connections on vector bundles and those in principal bundles.

There are

1. Ehresmann connections are connections on fibre bundles, whether vector or principal
2. Linear or Kozsul connections, which effectively define covariant derivatives on vector bundles, are linear Ehresmann connections on vector bundles.
3. Principal connections are Ehresmann connections on principal G-bundles
4. Affine connections are linear connections on tangent bundles, but may also be defined as principal connections on the frame bundles.
5. Cartan connections are a specialization of Principal connections.

5.3 Ehresmann connections

This section is almost verbatim from Wikipedia. First, we must define the vertical and horizontal bundles. Given $\pi\colon E \to M$, which induces $d\pi\colon TE \to TM$ then we define the vertical bundle V as

$$V = \ker(d\pi) \tag{23}$$

The horizontal bundle H is the complement of V in TE, so that TE is the direct sum of H and V:

$$TE = H \oplus V \tag{24}$$

The horizontal bundle has the following properties:

- For each $e \in E$, H_e is a vector subspace of the tangent space T_eE to E at e, called the horizontal space H_e of the connection at e
- H_e depends smoothly on e
- For each e, $H_e \cap V_e = \{0\}$
- For any $e \in E$, and $v \in T_eE$, $T_eE = H_e + V_e$

The Ehresmann connection on E is this smooth subbundle H of TE.

Connection form. Another way to define the connection is by defining v the projection onto V along H, such that $H = \ker\, v$. Moreover, $v^2 = v$, and the projection of an element in V is the element itself.

Horizontal lift. The Ehresmann connection defines a general form of parallel transport, called the horizontal lift. A lift of a curve γ in M is a curve $\tilde{\gamma}$ through E such that $\pi(\tilde{\gamma}) = \gamma$. A lift is horizontal if, in addition, every tangent of the curve lies in the horizontal subbundle of TE:

5.3.1 Connections on Vector Bundles

Let E be a smooth vector bundle on a differentiable manifold M. The set of sections on E is $\Gamma(E)$. A connection on E is a smooth $\mathbb{R}$-linear map

$$\nabla\colon \Gamma(E) \to \Gamma(T^*M \otimes E)$$

such that the Leibniz rule holds for all smooth f on M and sections s on E:

$$\nabla(fs) = df \otimes s + f\nabla s \tag{25}$$

This connection on a smooth manifold immediately leads to a covariant derivative

$$\nabla_X \colon \Gamma(E) \to \Gamma(E)$$

along a section $X \in \Gamma(TM)$ by contracting X with the resulting covariant index in the connection:

$$\nabla_X(s) = (\nabla(s))(X). \tag{26}$$

The covariant derivative satisfies the linearity property in X and the Leibniz property in s:

$$\nabla_{fX} s = f\nabla_X s; \quad \nabla_X fs = \partial_X(fs) + f\nabla_X s, \tag{27}$$

where the derivative is the directional one. In practice, we define linear transformations using Christofel symbols.

5.4 Kozsul Connection

An Ehresmann connection H on E is said to be a linear (Ehresmann) connection or **Kozsul** connection if H_e depends linearly on $e \in E_x$ for each $x \in M$.

5.4.1 Affine Connections

An affine connection is a linear connection on a tangent bundle. The linear connection immediately defines a covariant derivative and vice versa. Given a basis for the tangent space, we may provide Christoffel symbols that dictate the projection operation central to linear connections on vector bundles.

Note that the term ‘affine’ comes from the fact (source) that the tangent spaces being connected are affine, so that the modern definition of a linear connection implies an affine transformation between two affine spaces.

5.4.2 Covariant Derivatives From Christoffel Symbols

At q and $q' = q + dq$, we have tangent spaces T_qQ and $T_{q'}Q$.

Suppose we have a basis vector $e_i \in T_{q'}Q$. We can translate this basis vector in the ambient space, which is Euclidean so that translation is in the usual sense, and move its point of attachment to q which is the origin of T_qQ. However, most likely this translated vector $\bar{e}_i$ does not lie in T_qQ. So, we project the translated version of e_i to $e'_i \in T_qQ$. A simple way to project is linearly.

Under parallel translation in the ambient space and linear translation of the result onto T_qQ, we now get an element $e'_i \in T_qQ$. Now, there's an existing i^{th} basis vector in T_qQ, and we just defined a new version from e'_i. We can say that the difference between them is another vector $de_i \in T_qQ$, which has coordinates

$$de_i = (de_i^j)e_j. \tag{28}$$

At the end of this process, we can write the valid vector equation in T_qQ as $e'_i = e_i + de_i$ for each i.

The parallel translation might be unique, but the projection need not be. Whatever projection rule we choose, we will define some correction terms de_i that help relate the basis vectors from two different tangent spaces. We can make the coordinates of de_i depend linearly on the difference dq between q and q'. This linearity makes sense, since it implies that as $dq \to 0$, $e'_i \to e_i$. We can make this linear map depend on q, so that up to a first order approximation, we will define

$$de_i^j = \Gamma^j_{ik}(q)dq^k, \tag{29}$$

so that

$$e'_i = e_i + \Gamma^j_{ik}(q)dq^k e_j \tag{30}$$

The function Γ^j_{ik} gives the components of the affine connection, or equivalently, the covariant derivative.

We have a vector field $X(q) = X^i(q)e_i(q)$. How do we define the derivative? It is clear that $X(q+dq) - X(q)$ makes no sense.

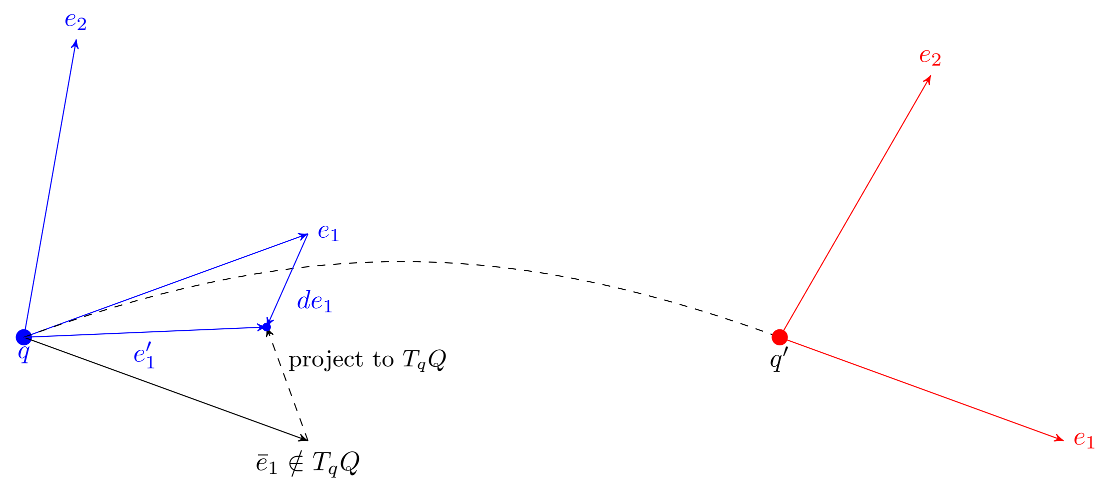

Figure 4: The basis element e_1 of $T_{q'}Q$ is mapped to $e_1' \in T_qQ$ through a translation in ambient space followed by a linear projection onto T_qQ. The difference between the original basis element $e_1 \in T_qQ$ and this mapped one from $T_{q'Q}$ is the vector $de_1 \in T_qQ$.

We have to first pull $X(q + dq)$ back to T_qQ. Now, $X(q + dq)$ is some linear combination of the basis vectors $e_i \in T_{q+dq}Q$. To map $X(q + dq)$, we simply use the *same* linear combination of the mapped basis vectors $\tilde{e}_i$ (denoted as e_i' in the previous section). We do this pulling back through the coordinates.

$$\tilde{e}_i = e_i(q) + \Gamma^j_{ik}(q)dq^k e_j \tag{31}$$

$$\begin{aligned}\tilde{X} &= X^i(q + dq)\tilde{e}_i \\ &= (X^i(q) + \partial_j X^i dq^j)\Big(e_i(q) + \Gamma^j_{ik}(q)dq^k e_j\Big) \\ &= X^i(q)e_i + \partial_j X^i dq^j e_i + \Gamma^j_{ik}(q)X^i dq^k e_j\end{aligned} \tag{32}$$

where it appears that the product $\partial_j X^k dq^j dq^k$ is ignored. Now, $\tilde{X} - X$ at any q is

$$dX = \tilde{X} - X = \partial_j X^i dq^j e_i + \Gamma^m_{ik}(q)X^i dq^k e_m. \tag{33}$$

We may use the above expression to compute the covector $\partial X/\partial q_j$. It's i^{th} component is

$$(\nabla_j X)^i = \partial_j X^i + \Gamma^i_{jk} X^k \tag{34}$$

5.5 Principal Connection

The frame bundle of a manifold M is a special type of principal bundle in the sense that its geometry is fundamentally tied to the geometry of M.

6 Riemannian Metrics

A metric on a vector space in $\mathbb{R}^m$ is a $(0, 2)$-tensor. It's representation is therefore an $m \times m$ matrix. For robotic systems, the metric tensor is known as the inertia tensor, which is also a Riemannian metric. The Riemannian property comes from the fact that we can define a continuous tensor field.

Once we define a metric $\mathbb{G}_q\colon T_qQ \times T_qQ \mapsto \mathbb{R}$, we get an affine connection, covariant derivative, and hence geodesics. The coordinate-independent equations of a mechanical system are then

$$\nabla_{\dot{q}}\dot{q} = \mathbb{G}_q^\sharp(-dV + f + f_d)$$

where V is a conservative potential, f_d is the dissipation, and f are the external generalized forces.

Our quest is that for *any* (suitable) coordinate system, we can achieve motions once we estimate $\mathbb{G}_q$, f_d, and dV in that coordinate system. This approach opens the door to sensorimotor control. Note that these quantities have been difficult to extract from data thus far.

In general, the bases for a vector and covector space are unrelated, until we can define a metric, or more accurately, a $(0, 2)$-tensor.

6.1 Preservation of Exterior Products

Let's say we equip $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$ with bases B_1 and B_2 respectively. Let the metrics on $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$, and the exterior product, be the usual ones. So, if we have representations v, for $v \in \mathbb{R}^n$ and ω for $\omega \in (\mathbb{R}^n)^*$, we get

$$\|v\|_I^2 = v^T v \tag{35}$$
$$\|\omega\|_I^2 = \omega^T \omega \tag{36}$$
$$\langle \omega, v \rangle = \omega^T v \tag{37}$$

Suppose we transform $\mathbb{R}^n$ by a linear transformation represented as T_1 in B_1 and $(\mathbb{R}^n)^*$ by a linear transformation represented as T_2 in B_2. Then, we would see that the $v \mapsto T_1^{-1} v$ and $\omega \mapsto T_2^{-1} \omega$.

$$\|v'\|_I^2 = v^T T_1^{-T} T_1^{-1} v \tag{38}$$
$$\|\omega'\|_I^2 = \omega^T T_2^{-T} T_2^{-1} \omega \tag{39}$$
$$\langle \omega', v' \rangle = \omega^T T_2^{-T} T_1^{-1} v \tag{40}$$

So, if we want the linear functional evaluations of vectors to be invariant after the change of bases, we need $T_2 = T_1^{-T}$. Let $M^{-1} = T_1^T T_1$. Then,

$$\|v'\|_I^2 = v^T T_1^{-T} T_1^{-1} v = \|v\|_M^2 \tag{41}$$
$$\|\omega'\|_I^2 = \omega^T T_1^T T_1 \omega = \|\omega\|_{M^{-1}}^2 \tag{42}$$
$$\langle \omega', v' \rangle = \omega^T v \tag{43}$$

While changing the basis of $\mathbb{R}^n$ to vectors with cartesian coordinates T_1, and the basis of $(\mathbb{R}^n)^*$ correspondingly to covectors with coordinates T_1^{-T}, we preserve linear functionals, but not the metric. To preserve the metric, we must choose $T_1^T T_1 = I$, meaning cartesian transformations.

6.2 In Coordinates

Once we define coordinates, we get the Euclidean metric for free on $T_q Q$ through the natural basis. However, we may prefer a different size measure for vectors defined using the natural basis. This metric is the Riemannian metric $M(q)$. As the derivation in Section **??** shows, to choose metric $M(q)$ on $T_q Q$ requires us to choose metric $M^{-1}(q)$ on $T_q^* Q$ to keep the results of linear functionals invariant.

The metric also leads to the sharp map $\mathbb{G}_q^\sharp : T_q^* Q \to T_q Q$ which maps torques (cotangents) to velocities (tangents). In effect, $F \mapsto M^{-1}(q) F \in T_q Q$. Similarly, we get the flat map $\mathbb{G}_q^\flat : T_q Q \to T_q^* Q$ which maps velocities to torques. In effect, $v \mapsto M(q) v \in T_q^* Q$. The point of these maps is to ensure that taking the outer product between cotangent and tangent is the same as evaluating the metric after transformation:

$$F^T v = \langle F, v \rangle = \langle G^\sharp(F), v \rangle_M = \langle M^{-1}(q) F, v \rangle_M = F^T M^{-T}(q) M(q) v = F^T v$$

Similarly $\langle F, v \rangle = \langle F, G^\flat(v) \rangle_{M^{-1}}$.

Alternatively, we may say that the metric on $T_q Q$ induces a metric for $T_q^* Q$ under the natural basis, or a basis for $T_q^* Q$ under the natural metric, and the natural exterior product makes sense either way.

6.3 Errors On Manifolds

Consider two points $q \in Q$ and $r \in Q$, with the understanding of the current and reference configurations. We may define an smooth function $\phi: Q \times Q \to \mathbb{R}$. It is an error function if it is positive definite. In other words $\phi(q, r) \geq 0$ with equality if and only if $q = r$. It is symmetric if $\phi(q, r) = \phi(r, q)$ for all $q, r \in Q$.

6.4 Transport Map and Velocity Error

Let $\mathrm{d}_1\phi$ and $\mathrm{d}_2\phi$ denote the differentials of ϕ with respect to the first and second arguments. A map $\mathcal{T}_{(q,r}: T_rQ \to T_qQ$ is a **transport map** if it is compatible with the configuration error, meaning that

$$\mathrm{d}_2\phi = -T^*_{(q,r)}\mathrm{d}_1\phi, \tag{44}$$

where $\mathcal{T}^*_{(q,r}: T^*_qQ \to T^*_rQ$ is the dual map of $\mathcal{T}_{(q,r}$. Intuitively, this map correctly relates the steepest direction of decreasing errors at the two points q and r. For Euclidean-distance based errors, $T^*_{(q,r)} = I$ for all q, r.

Given $\dot{q} \in Q$ and a velocity $\dot{r} \in T_rQ$, the velocity error is

$$\dot{e} = \dot{q} - T_{(q,r)}\dot{r}, \tag{45}$$

where $\dot{r}$ has been transported into T_qQ.

It is then possible to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(q(t), r(t)) = \mathrm{d}_1\phi(q(t), r(t)) \cdot \dot{e}(t) \tag{46}$$

7 Lie Groups

Definition 3 (Lie Group)**.** A Lie group is a finite dimensional smooth manifold G together with a group structure on G, such that the multiplication $G \times G \to G$ and the attaching of an inverse $g \mapsto g^{-1}: G \to G$ are smooth maps.

Example 12 ($SO(3)$)**.** The space of rotation matrices forms a Lie group under matrix multiplication. The dimension of the manifold of rotation matrices is 3. $SO(3)$ is also called a Matrix Lie Group.

7.1 Lie Algebra and the Tangent Space of Lie Groups

To every Lie group G we can associate a Lie algebra whose underlying vector space $\mathfrak{g}$ is the tangent space of the Lie group at the identity element.

Definition 4 (Lie Algebra)**.** A Lie algebra is a vector space $\mathfrak{g}$, together with a non-associative operation called the Lie bracket, an alternating bilinear map

$$\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}, \;\; (x, y) \mapsto [x, y],$$

satisfying the Jacobi identity:

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] \;=\; 0, \forall\, x, y, z \in \mathfrak{g}.$$

Remark 2. Alternating: $[x, x] = 0$; Non-associative: $[[x, y], z] \neq [x, [y, z]]$

Example 13 (Lie Algebra of $SO(3)$)**.** The Lie algebra of $SO(3)$ consists of a vector space $\mathfrak{so}(3)$ and a Lie bracket given by the usual matrix commutator. $\mathfrak{so}(3)$ is the set of 3×3 real skew-symmetric matrices, and the Lie bracket is

$$[R_1, R_2] = R_1R_2 - R_2R_1.$$

Properties of $\mathfrak{so}(3)$:

- Linear
- Interpretation as cross product
- $x^TSx = 0$ for any x.
- $S(Ra) = RS(a)R^T$

7.2 Exponential Map

The exponential map turns out to be natural. This means the following diagram (taken from Lerman's notes) commutes for any Lie group morphism $\phi\colon H \to G$:

$$\begin{tikzcd} \mathfrak{h} \arrow[r, "\delta\phi"] \arrow[d, "\exp"'] & \mathfrak{g} \arrow[d, "\exp"] \\ H \arrow[r, "\phi"] & G \end{tikzcd}$$

Definition 5. The exponential of $X \in \mathfrak{g}$ is given by $\exp(X) = \gamma(1)$ where $\gamma\colon \mathbb{R} \to G$ is the unique one-parameter subgroup of G whose tangent vector at the identity is equal to X.

Moreover, we have that $\gamma_X(s+t) = \gamma_X(s)\gamma_X(t)$, $\gamma_X(0) = I$

Example 14. For a Matrix Lie Group, the exponential of X is

$$\exp(X) = \sum_{k=0}^{\infty} \frac{X^k}{k!} = I + X + \frac{1}{2}X^2 + \frac{1}{6}X^3 + \cdots.$$

Example 15. The unit circle centered at 0 in the complex plane is a Lie group (called the circle group) whose tangent space at 1 can be identified with the imaginary line in the complex plane, $\{it : t \in \mathbb{R}\}$. The exponential map for this Lie group is given by

$$it \mapsto \exp(it) = e^{it} = \cos(t) + i\sin(t),$$

that is, the same formula as the ordinary complex exponential.

7.3 Adjoints

The adjoint of a linear map $f\colon V \to W$ is the map $f^*\colon W^* \to V^*$, which satisfies

$$f^*(\phi)(v) = \phi(f(v)), \text{ equivalently, } \langle f^*(\phi), v\rangle = \langle \phi, f(v)\rangle$$

The Representation Theorem states that given a finite dimensional inner product (vector) space V, every linear functional on V may be represented as an inner product with a unique element in V. The exterior product of $T(v)$ with $w \in W^*$ is effectively a linear functional on V, and so the adjoint map is providing the representation of that linear functional as an element of V^*, or V through the canonical transformation. Effectively, the adjoint of f gives us a way to perform an exterior product in W as either an exterior or inner product in V.

If the map f has a matrix representation A in the bases of V and W, then the map f^* has a representation A^T when using the canonical bases for V^* and W^*.

Example 16 (Jacobian)**.** In velocity kinematics, the Jacobian $J(q)$ defines a map from joint velocities at q to task-space velocities $\dot{q} \to J(q)\dot{q} = \xi$. Therefore, $J(q)^T$ defines a map from task-space forces (co-vectors in task space) to joint-torques (co-vectors in joint-space), which then is mapped to an acceleration through the sharp map ($J(q)^T F_{ext} \mapsto M^{-1}J(q)^T F_{ext}$).

7.4 Adjoint Representations in Groups

A representation of a group G on a vector space V over a field K is a group homomorphism from G to $GL(V)$, the general linear group on V. That is, a representation is a map

$$\rho\colon G \to \mathrm{GL}(V)$$

such that

$$\rho(g_1 g_2) = \rho(g_1)\rho(g_2), \qquad \text{for all } g_1, g_2 \in G.$$

In algebra, a *homomorphism* is a structure-preserving map between two algebraic structures of the same type (such as two groups, two rings, or two vector spaces). A homomorphism commutes with the algebraic structure operation.

We can look at representations of a given group on any vector space[1]. But there is exactly one distinguished vector space that comes automatically with each group: its own Lie algebra[2]. This representation is the **adjoint representation**.

The adjoint action $I_g(h) = ghg^{-1}$ is a homomorphism[3]. We may see that

$$I_g(e) = geg^{-1} = gg^{-1} = e, \qquad \forall g \in G.$$

The property $I_g(e) = e$ means that any curve through e on the manifold G is mapped by this homomorphism to another (not necessarily the same) curve through e. Therefore the adjoint representation maps any tangent vector (of a curve on G) in T_eG to another tangent vector in T_eG. In contrast left- (and right-) translations L_g map tangent vectors in T_eG to tangent vectors in $T_{\mathbf{g}}G$.

The induced map (by I_g) of any tangent vector in T_eG (an element of the Lie algebra) to another tangent vector in T_eG is called the adjoint transformation of T_eG induced by g. This induced map defines a representation of the group G on T_eG, because T_eG is a vector space.

- adjoint map: $I_g(h) = ghg^{-1}$
- The derivative of I_g at the origin is $\mathrm{Ad}_g\colon T_eG \to T_eG$; $\mathrm{Ad}_g(dI_g)_e = gXg^{-1}$
- the map
$$\mathrm{Ad}\colon G \to \mathrm{Aut}(\mathfrak{g}),\ g \mapsto \mathrm{Ad}_g$$
is a group representation called the adjoint representation of G. Note that Ad converts g into a linear map no $\mathfrak{g}$, which is what we wanted
- Differentiating the adjoint map at origin produces the adjoint action of Lie algebra: $\mathrm{ad}_x(y) = [x, y]$:
$$\mathrm{ad}(X) = (d\ \mathrm{Ad})_e(X)$$
- The adjoint representation of the Lie algebra is then $\mathrm{ad}\colon \mathfrak{g} \to Der(\mathfrak{g})$, where $Der(\mathfrak{g})$ is the Lie algebra of $Aut(\mathfrak{g})$
- Ad and ad are related through the exponential map: Specifically, $\mathrm{Ad}_{exp(x)} = exp(\mathrm{ad})$ for all x in the Lie algebra. It is a consequence of the general result relating Lie group and Lie algebra homomorphisms via the exponential map.

The naturality of the exponential functions leads to the following diagrams:

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "\mathrm{Ad}_g"] \arrow[d, "\exp"'] & \mathfrak{g} \arrow[d, "\exp"] \\ G \arrow[r, "I_g"] & G \end{tikzcd} \qquad \begin{tikzcd} \mathfrak{g} \arrow[r, "\mathrm{ad}"] \arrow[d, "\exp"'] & \mathfrak{gl}(\mathfrak{g}) \arrow[d, "\exp"] \\ G \arrow[r, "\mathrm{Ad}"] & GL(\mathfrak{g}) \end{tikzcd}$$

In other words, for all t

1. $g\ \exp(tX)g^{-1} = \exp(t\ \mathrm{Ad}_g(X))$
2. $\mathrm{Ad}(\exp(tX)) = \exp(t\ \mathrm{ad}(X))$

7.5 Orientation Group SO(3)

The orientation group SO(3) is given by

$$\mathrm{SO}(3) = \{R \in \mathbb{R}^{9\times 9}\colon R^T R = I, \det(R) = 1\} \tag{47}$$

The identity element is I, and the group operation is matrix multiplication.

[1]The representation maps to the group of general linear transformations on this vector space
[2]Taken from this blog.
[3]It is also a realization, not a representation, since it's domain is a group G and not a vector space

7.5.1 Lie Algebra

Moreover, this group is a Lie group. Consider a curve $R(t)$, where $R(t) \in \mathrm{SO}(3)$ for each $t \in [0,1]$. For every r, we know that $R(t)R(t)^T = I$. Taking the derivative with respect to time, we get

$$\dot{R}(t)R(t)^T + R(t)\dot{R}(t)^T = 0 \tag{48}$$

$$S(t) = \dot{R}(t)R(t)^T \implies S + S^T = 0 \tag{49}$$

We may then obtain that

$$\dot{R}(t) = S(t)R(t) \tag{50}$$

The derivative of $R(t)$ when $R(t) = I$, is simply $S(t)$, which is a skew-symmetric matrix. Therefore, the Lie algebra $\mathfrak{so}(3)$ is the set of skew-symmetric matrices.

There's a natural homeomorphism from $\mathbb{R}^3$ to $\mathfrak{so}(3)$ given by

$$S(\omega) = \hat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \tag{51}$$

An alternative definition for the Lie Algebra is the set of all the left invariant vector fields on the Lie group. A vector field X on a Lie group G is left invariant if

$$(dL_g)(X(x)) = X(L_g(x)) = X(gx) \tag{52}$$

It can be shown that this definition is equivalent to the tangent space at identity, which is a practical way of determining the Lie algebra.

Due to the homeomorphism between $\mathfrak{so}(3)$ and $\mathbb{R}^3$, we may then write

$$\dot{R}(t) = S(\omega(t))R(t) \tag{53}$$

for some curve $\omega(t)$ in $\mathbb{R}^3$.

7.5.2 Adjoint Representation

We have that for $R \in \mathrm{SO}(3)$, $R(a \times b) = (Ra) \times (Rb)$. We may view $S\colon \mathbb{R}^3 \to \mathfrak{so}(3)$ as a map. Therefore,

$$(Rw) \times b = (Rw) \times (RR^Tb) = R(w \times R^Tb) = RS(w)R^Tb \tag{54}$$

$$\implies S(Rw) = RS(w)R^T = (\mathrm{Ad}(R))\,(S(w)) \tag{55}$$

$\mathfrak{so}(3)$ output of Ad_R $\quad$ $\mathrm{Ad}_R(X)$ $\quad$ Ad creates Ad_R $\quad$ $\mathfrak{so}(3)$ Input to Ad_R

gXg^{-1}

8 Dynamical Systems, Tangents, Vector Fields

Vector fields are used to define differential equations, since they pick elements from the tangent space at each point of a space.

Consider an autonomous nonlinear differential equation on a state space $X \subseteq \mathbb{R}^n$ given by

$$\dot{x} = f(x). \tag{56}$$

The function $f(x)$ is precisely a vector field on X, and $f(x) \in T_xX$.

Consider a nonlinear system with inputs of the form

$$\begin{aligned} \dot{x} &= f(x) + g_1(x)u_1 + \cdots g_m(x)u_m \\ &= f(x) + G(x)u \end{aligned} \tag{57}$$

where $f(x)$, $g_1(x)$, $\cdots, g_m(x)$ are smooth vector fields on M, $u \in \mathbb{R}^m$, and the i^{th} column of $G(x)$ is $g_i(x)$. We assume $M = \mathbb{R}^n$ for simplicity. System (57) is known as an affine input system, since the dynamics (vector field) are affine in the input u. Note that the vector field can still be nonlinear in the state x.

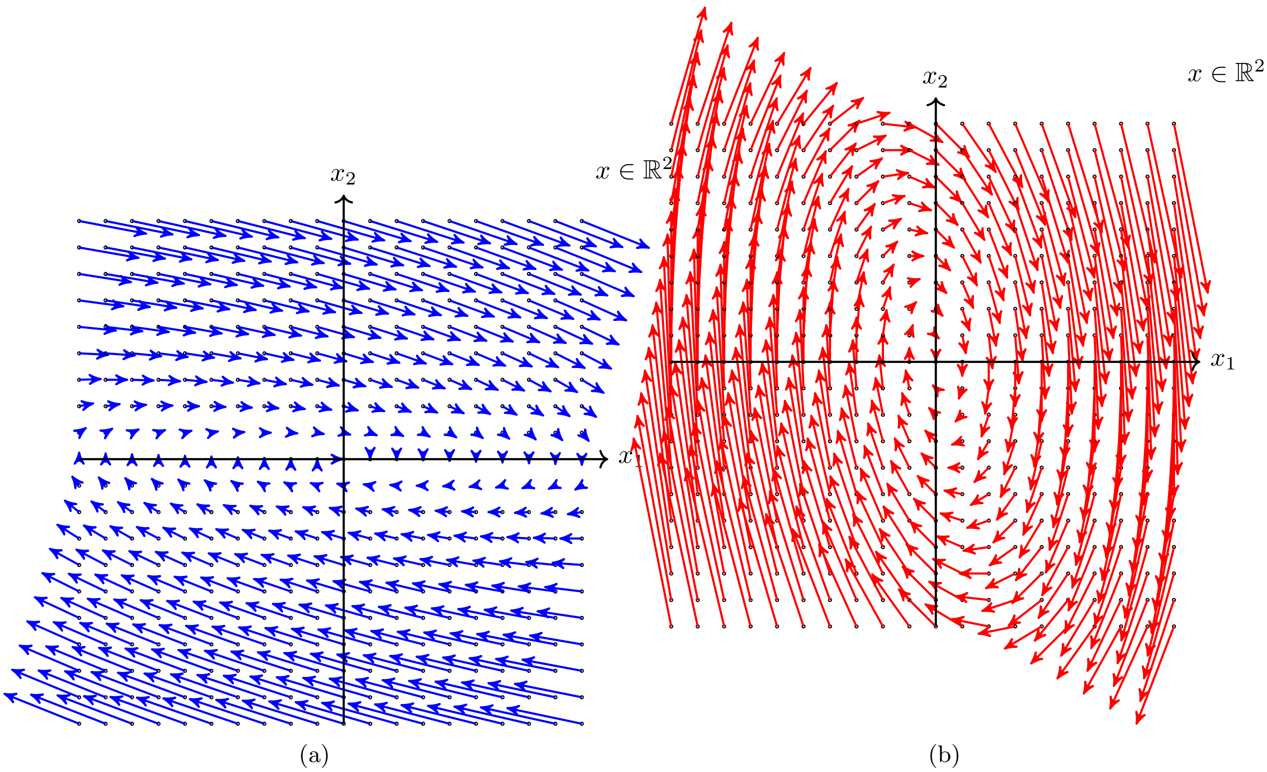

Figure 5: Two vector fields on $\mathbb{R}^2$

Example 17 (Linear Systems)**.** Compare (57) to the linear system $\dot{x} = Ax + Bu$.

Example 18 (Angular Velocity)**.** Suppose we have a rigid body at orientation given by angle $\theta(t)$ about some axis $\overrightarrow{k}$ at time t. We typically refer to its angular velocity as $\dot{\theta}$, assuming a frame conveniently aligned with $\overrightarrow{k}$. Its angular velocity ω is $\omega = \dot{\theta}k$ in the same frame that defines k.

Now, suppose we have a point $p(t)$ on this body. We have learned that its velocity is $\omega \times p(t)$. What happens when the orientation is given by a rotation matrix?

Or, how do we compute the velocity of a point $q(t)$ when

$$q(t) = R(t)q + d(t)?$$

Derive $\dot{R}(t) = SR(t)$ when $RR^T = I$. Explain that this version works in world frame.

Example 19 (Rigid Body Dynamics)**.** Let $x(t) \in \mathbb{R}^3$ be the location of the origin of a frame attached to a rigid body at time t, relative to an inertial frame. Let $R(t)$ be its orientation relative to that same inertial frame. Let I_0 anad $\omega_0(t)$ be the rotational inertia and angular velocity of the rigid body in the inertial frame. The dynamics of the rigid body pose in the inertial frame are given by

$$\dot{x}(t) = v(t) \tag{58}$$

$$m\dot{v}(t) = f(t) \tag{59}$$

$$\dot{R}(t) = S(\omega_0)R(t) \tag{60}$$

$$\frac{d}{dt}\left(I_0\omega_0\right) = \tau_0 \tag{61}$$

The orientation dynamics is easier to express in a body-fixed frame

$$I\dot{\omega} + \omega \times I\omega = \tau \tag{62}$$

Remark 3 (Quadrotor Dynamics). A quadrotor is often treated as a rigid body on which acts three torques and a thrust at its center of mass.

8.1 Lie Algebra

Choosing a feedback control $u = k(x)$ for the system (57) is like choosing a vector field out of the distribution implied by $f(x)$ and $G(x)$. Our ability to dictate the evolution of x with time therefore depends on this distribution. We now introduce some algebraic operations that help analyze the possible behaviors allowed by a distribution consisting of a finite set of linearly independent vector fields.

8.2 Lie Bracket

Let f and g be differentiable vector fields on $\mathbb{R}^n$. The Lie bracket of f and g, denoted $[f, g]$, is a vector field on $\mathbb{R}^n$ given by

$$[f, g](x) = \frac{\partial g}{\partial x} f(x) - \frac{\partial f}{\partial x} g(x) \tag{63}$$

where $\frac{\partial g}{\partial x}$ is the Jacobian of $g(x)$.

Note that the Lie Bracket maps two vector fields on $\mathbb{R}^n$ into another vector field on $\mathbb{R}^n$.

Example 20 (Lie bracket as Commutation)**.**

We also denote $[f, g]$ as $\mathrm{ad}_f(g)$, so that we can define repeated Lie brackets with respect to f through the recursion $\mathrm{ad}_f^k(g) = [f, \mathrm{ad}_f^{k-1}(g)]$, where $\mathrm{ad}_f^0(g) = g$.

8.3 Lie Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field on $\mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}$ be a scalar function. The Lie derivative of h with respect to f, denoted $L_f h$, is given by

$$L_f h = \frac{\partial h}{\partial x} f(x) \tag{64}$$

The Lie derivative yields another scalar function, implying that we can define repeated Lie brackets as $L_f^k h = L_f(L_f^{k-1} h)$, where $L_f^0 h = h$.

The Lie Brackets and Derivatives satisfy the following identity:

$$L_{[f,g]} h = L_f L_g h - L_g L_f h. \tag{65}$$

Combined with the definition of repeated Lie brackets/derivatives, this identity gets used in showing many results.

8.4 Involutivity

A distribution $\Delta = \mathrm{span}\{X_1, \ldots, X_k\}$ is involutive if and only if there exist scalar functions $\alpha_{ijk} : \mathbb{R}^n \to \mathbb{R}$ such that

$$[X_i, X_j] = \sum_{k=1}^{m} \alpha_{ijk} X_k, \ \ \forall \ i, j, k \tag{66}$$

In other words, a distribution is involutive if it is closed with respect to the Lie bracket operation.

9 Feedback Linearization

The motivation for feedback linearization is to create a systematic procedure for designing controllers for nonlinear systems of the form (57) using well-known linear system design principles, when possible. We use ideas from differential geometry to characterize when a system can be feedback linearized.

Example 21. Feedback Linearization Consider the system

$$\dot{x}_1 = a \sin x_2 \tag{67a}$$
$$\dot{x}_2 = -x_1^2 + u \tag{67b}$$

It isn't clear how to choose u so as to influence x_1. If we change the variables, locally, through the transformation

$$y_1 = x_1 \tag{68a}$$
$$y_2 = a \sin x_2 = \dot{x}_1, \tag{68b}$$

The dynamics become

$$\dot{y}_1 = y_2 \tag{69a}$$
$$\dot{y}_2 = a\cos(x_2)(-x_1^2 + u) \tag{69b}$$

Choosing

$$u = \frac{1}{a\cos x_2} v + x_1^2 \tag{70}$$

yields

$$\dot{y}_1 = y_2 \tag{71a}$$
$$\dot{y}_2 = v \tag{71b}$$

which we know how to design for and analyze.

Suppose we get a closed-loop response $y(t)$ by using some control $v = -Ky$. The response in the original coordinates is

$$x_1(t) = y_1(t) \tag{72a}$$
$$x_2(t) = \sin^{-1}\frac{y_2(t)}{a} \tag{72b}$$

9.1 Single Input Systems

A system $\dot{x} = f(x) + g(x)u$ is feedback linearizable if there exists a diffeomorphism $T: U \to \mathbb{R}^n$ defined on an open region $U \subseteq \mathbb{R}^n$ containing the origin, and nonlinear feedback $u = \alpha(x) + \beta(x)v$, with $\beta(x) \neq 0$ on U, such that the transformed state $y = T(x)$ satisfies the system of linear equations $\dot{y} = Ay + bu$ where A and b represent as a chain of integrators.

Since $y = T(x)$, and T is a diffeomorphism, we can derive

$$\dot{y} = \frac{\partial T}{\partial x}\dot{x} \tag{73}$$
$$\implies Ay + bv = \frac{\partial T}{\partial x}(f(x) + g(x)u) \tag{74}$$
$$\implies AT(x) + bv = \frac{\partial T}{\partial x}(f(x) + g(x)u) \tag{75}$$

Going by each component of the n equations, we get

$$T_2 = L_f T_1 + L_g T_1 u \tag{76}$$
$$T_3 = L_f T_2 + L_g T_2 u \tag{77}$$
$$\vdots$$
$$v = L_f T_n + L_g T_n u \tag{78}$$

Since $T(x)$ is independent of u, but v depends on u, we get

$$L_g T_1 = L_g T_2 = \cdots = L_g T_{n-1} = 0 \tag{79}$$
$$L_g T_n \neq 0 \tag{80}$$

thereby reducing the n components to

$$T_{i+1} = L_f T_i, \quad i \in \{1, \ldots, n-1\} \tag{81}$$
$$v = L_f T_n + L_g T_n u \tag{82}$$

We now work to eliminate T_i for $i \geq 2$. We do this by using the relationship between Lie brackets and Lie derivatives in (65). This relationship implies that (81) and (82) become

$$L_{\mathrm{ad}_f^k(g)} T_1 = 0, \quad k \in \{0, 1, \dots, n-2\} \tag{83}$$

$$L_{\mathrm{ad}_f^{n-1}(g)} T_1 \neq 0 \tag{84}$$

If we can find T_1 satisfying the conditions above, we can find $T_2, \dots T_n$ inductively, and then find u.

First of all, we need $\mathrm{ad}_f^k(g)$ for $k \in \{0, \dots, n-1\}$ to be independent so that (84) is satisfiable. For (83) to have a solution, we know that $\mathrm{ad}_f^k(g)$ for $k \in \{0, \dots, n-2\}$ must lead to an involutive distribution, by the Frobenius Theorem (see below).

Theorem 1. *A system $\dot{x} = f(x) + g(x)u$ is feedback linearizable if and only if there exists an open region $U \subseteq \mathbb{R}^n$ containing the origin in which*

1. $\mathrm{ad}_f^k(g)$ *for* $k \in \{0, \dots, n-1\}$ *are linearly independent in* U*.*
2. $\Delta = \mathrm{span}\{g, \mathrm{ad}_f(g), \dots, \mathrm{ad}_f^{n-2}(g)\}$ *is involutive in* U*.*

9.2 Frobenius Theorem

This theorem is concerned with the existence of a solution to a system of partial differential equations in terms of a distribution corresponding to those PDEs.

Definition 6. A distribution $\Delta = \mathrm{span}\,\{X_1, \dots, X_m\}$ on $\mathbb{R}^n$ is said to be completely integrable if and only if there are $n-m$ linearly independent functions $h_1, \dots, h_{n-m}$ satisfying the system of partial differential equations

$$L_{X_i} h_j = 0, \text{ for } 1 \leq i \leq m, 1 \leq j \leq n-m \tag{85}$$

Theorem 2 (Frobenius Theorem)**.** *A distribution Δ is completely integrable if and only if it is involutive.*

A Vector Spaces

Definition 7 (Group)**.** A group G is a set together with a binary operation $\cdot$ that satisfies the following properties for all $a, b, c \in G$:

(i) Closure: $a \cdot b \in G$;

(ii) Associativity: $a \cdot (b \cdot c) = (a \cdot) b \cdot c$;

(iii) Existence of identity element $e \in G$ such that $a \cdot e = e \cdot a = a$;

(iv) Existence of inverse element $d \in G$ such that $d \cdot a = a \cdot d = e$.

Example 22. Real numbers form a group under addition.

Example 23. Real numbers without 0 form a group under multiplication.

Definition 8 (Field)**.** A field $\mathbb{F}$ is a set together with two operations – addition $+\colon \mathbb{F} \times \mathbb{F} \mapsto \mathbb{F}$ and multiplication $\cdot\colon \mathbb{F} \times \mathbb{F} \mapsto \mathbb{F}$ – that satisfy the eight axioms listed below.

(i) Addition and multiplication are associative

(ii) Addition and multiplication are commutative

(iii) Existence of additive and multiplicative identity elements

(iv) Existence of inverse element for addition for each $v \in V$

(v) Existence of inverse element for multiplication for each $v \in V$ except for the additive identity

(vi) Distributivity of multiplication with respect to addition

Example 24. Real numbers are a field under usual addition and multiplication.

Definition 9 (Vector space)**.** A vector space over a field $\mathbb{F}$ is a set V together with two operations – vector addition $+\colon V \times V \mapsto V$ and scalar multiplication $\cdot\colon \mathbb{F} \times V \mapsto V$ – that satisfy the eight axioms listed below, for all $u, v, w \in V$ and $a, b \in \mathbb{F}$.

(i) Addition is associative: $u + (v + w) = (u + v) + w$;

(ii) Addition is commutative: $u + v = v + u$;

(iii) Existence of identity element $0 \in V$ such that $v + 0 = v$;

(iv) Existence of inverse element $x \in V$ such that $v + x = 0$;

(v) Compatibility of scalar multiplication with respect to field multiplication: $a \cdot (bv) = (a \cdot b)b$;

(vi) Existence of identity element $e \in \mathbb{F}$ under scalar multiplication such that $ev = v$;

(vii) Distributivity of scalar multiplication with respect to vector addition: $a \cdot (u + v) = a \cdot u + b \cdot v$;

(viii) Distributivity of scalar multiplication with respect to field addition: $(a + b) \cdot u = a \cdot u + b \cdot u$.

Example 25. The set of n-tuples of real numbers, denoted $\mathbb{R}^n$, over the field of real numbers form a vector space when addition and scalar multiplication of these n-tuples are taken to be element-wise addition and scalar multiplication. The 0 vector is the vector with all elements 0, and the inverse of $v \in \mathbb{R}^n$ is $-v = (-1)\cdot v$.

Definition 10 (Vector Space Basis)**.** A basis B of a vector space V is a set of vectors in V such that all other vectors can be written as a finite linear combination of the elements of B.

Remark 4 (Basis for $\mathbb{R}^n$). A basis for vector space $\mathbb{R}^n$ contains exactly n linearly independent vectors.

Remark 5 (Coordinates for $\mathbb{R}^n$). A basis for $\mathbb{R}^n$ equips each point $x \in \mathbb{R}^n$ with a coordinate given by the n coefficients of the basis vectors in the linear combination that yields x.

Definition 11 (Inner Product Space)**.** An inner product on a vector space V defined over a field $\mathbb{F}$ is a function $\langle \cdot, \cdot \rangle\colon V \times V \mapsto \mathbb{F}$ with the following properties

(i) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all $x, y \in V$;

(ii) $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$, for all $x, y, z \in V$ and $a, b \in \mathbb{F}$;

(iii) $\langle x, x \rangle \geq 0$, for all $x \in V$, and $\langle x, x \rangle = 0 \iff x = 0$.

An inner product space is a vector space equipped with a suitable inner product.

An inner product defines the notion of angle between two vectors, specifically defining when two vectors are orthogonal (perpendicular) to each other.

Example 26. Vector space $\mathbb{R}^n$ equipped with the usual dot product forms an inner product space. Two vectors in $\mathbb{R}^n$ are orthogonal when the angle between them is 90°.

Definition 12 (Norm)**.** A norm on a vector space V defined over field $\mathbb{F}$ (which is a subfield of the complex numbers $\mathbb{C}$) is a function $p\colon V \mapsto \mathbb{R}$ with the following properties:
For all $a \in \mathbb{F}$ and $x, y \in V$,

(i) $p(x + y) \leq p(x) + p(y)$;

(ii) $p(ax) = |a|p(x)$;

(iii) If $p(x) = 0$ then $x = 0$.

A **norm** defines a notion of **size** of vectors.

Example 27. An inner product space V with field $\mathbb{R}$ may be equipped with a norm p as follows:

$$p(u) = \sqrt{\langle u, u \rangle}.$$

Remark 6. For real vector spaces defined over $\mathbb{R}$, the symbol $\|\cdot\|$ is often used to denote the norm, instead of $p(\cdot)$.

Definition 13 (Metric)**.** A metric on a space X is a function $d\colon X \times X \mapsto \mathbb{R}$ with the following properties

(i) $d(x, y) \geq 0$ for all $x, y \in X$, and $d(x, y) = 0 \iff x = y$;

(ii) $d(x, y) = d(y, x)$, for all $x, y \in X$;

(iii) $d(x, y) \leq d(x, z) + d(z, y)$, for all $x, y, z \in X$.

A **metric** defines a notion of **distance** on a space.

Example 28. An inner product space V may be equipped with a norm $\|\cdot\|$, which then defines a metric $d\colon V \times V \to \mathbb{R}$ as

$$d(u, v) = \|u - v\|.$$