MECHANICS II

APPM2023

Course Notes by
Warren Carlson

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

2023

Preface

This is the set of course notes for APPM2023. The notes are intended to complement the lectures and other course material, and are by no means intended to be complete. Students should consult the various references that have been given to find additional material, and different views on the subject matter.

This material is under development. Please would you report any errors or problems (no matter how big or small). Any suggestions would be gratefully appreciated.

School of Computer Science and Applied Mathematics,
University of the Witwatersrand,
Johannesburg,
South Africa

Contents

Chapter 1

Algebra and Geometry

Geometry and algebra form the basis of the quantitative study of the world. Geometry defines the spatial relation among points and curves. Algebra is the formal set of rules by which computations involving geometry are made. In this chapter we shall discuss the geometric and algebraic basis for the description of mechanical systems. We shall begin with the coordinate free description of length in a geometric context, then coordinate systems as a method for describing the relative position of a point in space. We then discuss the algebraic formulation of length measure given a coordinate system and promote the algebra to a general setting.

1.1 Coordinate Grids

Our ultimate goal in Mechanics is to understand the mechanism by which systems go. This immediately implies that we should have same good description of the system in terms of its location is space and time. The objective of this section is to get some understanding of the conditions under which we can give a good description of a system that we want to study.

René Descartes was the first to introduce the concept of coordinate systems to geometry. He assigned a pair of real numbers to each point in the plane - an x- and y-coordinate. The plane thus parameterised is known as the *Cartesian plane*. Here we shall study several examples of coordinates on the Cartesian plane and other spaces

In the simplest case, we can study the motion of an object that moves along a straight line. As a precursor to a more general discussion to follow, we study the motion of a small bead that moves along a straight wire with infinite length. In this case, we might ask where the bead is located at some point in time, and then again at another point in time. If we are to make intuitive sense of how the bead moves, we should have a way to denote the position of the bead along the wire. The most natural description of the position of the bead in its motion is the distance of the bead from some predefined point. We shall call this reference point the *origin* or *datum point*. With this choice of origin, we can define the position of the of the bead a using a single (signed) real valued number. We write $a \in \mathbb{R}$ where a positive value of a denotes a position to the right-hand side of the origin, and a negative value of a denotes a position to the left-hand side of the origin.

Since the real numbers $\mathbb{R}$ have a natural ordering, it is possible to decide if the bead has moved toward the left or toward the right of its previous position by simply comparing the current a_f value of the bead position with a previously measured value of a_i. We compute the *displacement vector*

$$\vec{a} = a_f - a_i$$

defined as the directed line segment whose initial point and final point are a_i and a_f, respectively. When this directed line segment is placed with its initial point at a_i, the final point will correspond to a_f. In this way, we can locate the final position of the bead a_f using only its initial data a_i and the displacement vector $\vec{a}$.

The set of all displacement vectors $\vec{a}$ and $\vec{b}$ along the wire defines a *vector space* $\mathbb{V}$ where

$$\alpha\vec{a} + \beta\vec{b} \in \mathbb{V}$$

when $\alpha, \beta \in \mathbb{R}$. This means that there exists a vector $\vec{c} \in \mathbb{V}$ such that

$$\vec{c} = \alpha\vec{a} + \beta\vec{b}.$$

Note that the definition of each of these displacement vectors and, indeed, the vector space $\mathbb{V}$ is defined in terms of the differences between the points in $\mathbb{R}$ corresponding to points along the wire. The space $\mathbb{R}$ is an example of an affine space.

Definition 1 (Affine Space) *An affine space is a space comprising a set of points where the difference vector between any two points is well defined, but the sum is not defined.*

The definition of an affine space uses the idea of differences between point rather than their sum. This makes intuitive sense when considering displacement vectors, but the following question now arises: Why should we need a definition of this form if we already know how to add real numbers? The answer is quite simple - for any pair of numbers a and b, we can compute their sum $c = a + b$ and claim that the sum of the two points is now well defined. However, we could equally compute the difference $a = c - b$ or $b = c - a$. This demonstrates that the definition is, at least, compatible with the summation of real numbers in $\mathbb{R}$. However, an affine space gives us something more general, which is still valid when the above argument is not, for example, we can consider a displacement vector, connecting two points in the plane. There exists no composition rule to combine two points in the plane to get a third point, however there does exist a composition rule to add two displacement vectors in the plane to get a third displacement vector. The interpretation of the third displacement vector is unchanged from that in the one dimensional case, where the vector defines a displacement between points it the plane. Therefore, once the collection of all displacement vectors is defined, the collection of points needed to define them is no longer needed. This is not the case for the space of points themselves. A byproduct of this definition is that there is also no preferred choice of origin. Can you explain why this might be the case?

Remark 1 *Physical space, the space where we live, is an affine space.*

At this point we should note that we have not specified how the position a has been assigned. To do this, we should introduce a ruler, with a regular increment, to demarcate the position of the bead. A ruler with a regular increment has the benefit that increasing the geometric distance of the bead from the origin by some fixed factor is associated to a re-scaling of the numerical value of the a by the same factor. Additionally, we might choose to transfer the distance information on the ruler, that we will call the *metric*, onto the wire where the bead moves. Transferring the metric information from the ruler to the wire corresponds to placing a unique numerical value to each point on the wire, while maintaining the ordering of the numbers on the ruler, and hence allows us to read off the position information of the bead using only the numbers associated the points on the wire, without the need for the ruler. Additionally, the distance between any two bead positions is simply the magnitude of the displacement vector connecting the bead positions. We call this process of assigning a unique numerical value to each point on the wire *coordinatisation*. At this point we have associated the space of positions along the wire with the space of real numbers, $\mathbb{R}$.

It should be clear that choosing coordinates is not a unique process. As an example, we could swap out the original ruler that has some predefined unit, perhaps *millimeters*, with another ruler with a different unit, say *inches*. Since we know that one inch corresponds to 25.4 millimeters, we can use a conversion factor 25.4 to translate between the different distance measurement scales on the millimeter ruler and the inch ruler. This conversion factor is the translation factor associated with changing the metric from millimeters to inches. Then, if the bead is at a position a measured using the inch ruler, then it will have a position

$$a' = 25.4a$$

as measured on the millimeter ruler. For each different ruler, there will be a new conversion factor. We can think of the exchange of rulers as a transformation f on the metric information that we place on the wire. Later we shall see that there are choices of coordinatisation where the conversion factor changes depending on where the coordinates are studied. This rescaling gives a coordinate transformation

$$f : \mathbb{R} \to \mathbb{R} \quad \text{where} \quad a \mapsto f(a)$$

subject to the following restriction

$$a < b \iff f(a) < f(b) \quad \text{for all } a, b \in \mathbb{R}.$$

This means that if a point a is to the left of a point b on the wire before the change of coordinate is applied, then the transformed coordinate $f(a)$ is also to the left of the point $f(b)$ under the new coordinate system. When f is also linear, that is, for any scalar c

$$f(c\,a) = c\,f(a)$$

then f is an example of an *affine transformation*.

Definition 2 (Affine Transformation) *Suppose X is an affine space with $a, b \in X$ and let f be a function*

$$f : X \to X$$

$$b-a \mapsto f(b)-f(a)$$

then f is an affine transformation.

Remark 2 *An affine transformation describes any function that preserves lines, parallelism and relative scales but not distance information or angles.*

The exchange of the millimeter ruler for an inch ruler in the description of the position of the bead on the wire is an example of an affine transformation on the coordinate grid on the wire that allows us to describe the position of the bead. We shall encounter many such transformations and it is important to understand the the connection between different coordinate descriptions of physical systems and effects of changing coordinate systems when describing certain physical quantities. Before continuing coordinate transformations, let's consider some interesting examples of coordinate systems on some well known spaces.

Example 1.1 continues the discussion of placing coordinates onto a 1-dimensional space.

Example 1.1 (Coordinate grid on a Circle) *The circle S^1 is a 1-dimensional space. This means that every point on the circle is described by a single number. Since there is no preferred point of origin S^1, we can choose an origin denoted 0 and proceed to coordinates the rest of the space, see Figure 1.1. There are may ways to coordinatize S^1, but we shall leave these technical details for later in this text.*

The simplest way to assign coordinates on S^1 is to associated points on the circumference with the distance along the circumference from the chosen origin using a mapping function f. Suppose that S^1 has a circumference of 2π. We can map a finite interval corresponding to the semi-open subset $I = [0,2\pi) \subset \mathbb{R}$ to S^1. By this mapping

$$f : \mathbb{R} \to S^1$$
$$f(x) = x \bmod 2\pi \quad \textit{and} \quad x \overset{f}{\sim} x + 2\pi$$

and the length of the interval I matches the circumference of the circle. The equivalence relation $x \overset{f}{\sim} x+2\pi$ means that the points x and $x+2\pi$ are mapped to the same point on S^1. This means that f is a many-to-one *mapping and that f describes a valid* one-to-one *mapping only on the restricted subset* $[0,2\pi) \subset \mathbb{R}$.

After, the interval $[0,2\pi)$ is mapped to the circle, f maps $[2\pi,4\pi)$ onto S^1 and so on for each interval $[2\pi n, 2\pi(n+1))$ and for each $n \in \mathbb{Z}$. The negative real numbers are similarly mapped to the circle under the generalization to $n \in \mathbb{Z}$. Clearly, f is a many-to-one mapping, and f maps the entire space $\mathbb{R}$ is mapped onto S^1, by winding the real line multiple times over the circle.

Example 1.2 extends the ideas from coordinatizing the 1-dimensional line to the 2-dimensional plane. This is done by constructing a regular coordinate grid containing multiple copies of $\mathbb{R}$.

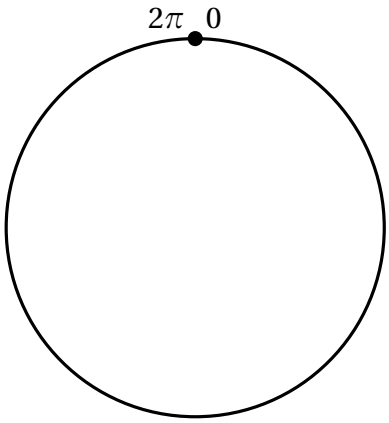


Figure 1.1: A mapping of the semi-open interval $[0, 2\pi)$ to S^1. Starting at $x = 0$ the positive real line is wrapped around the circumference of the circle, such that the point $x = 2\pi$ coincides with the image of $x = 0$.

Example 1.2 (Coordinate grid on the Cartesian Plane) *The* 2*-dimensional Cartesian plane can be covered with a by a* 2*-dimensional grid that assigns a pair of numbers to each point on the plane. The assignment of points follows from the placement of a single copy of the real line* $\mathbb{R}$ *on the plane. Then, at each point* x *along this line, a second copy of* $\mathbb{R}$ *is placed such that each each point* y *on the subsequent copies are aligned to form a grid. This grid is formed by the* Cartesian product *of the two copies of* $\mathbb{R}$ *to form a new space* $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$*. Each point on the plane corresponds to a unique point* (x, y) *in the Cartesian product space* $\mathbb{R}^2$*. Figure 1.2 shows grid formed my the multiple copies of* $\mathbb{R}$ *that covers the* 2*-dimensional plane.*

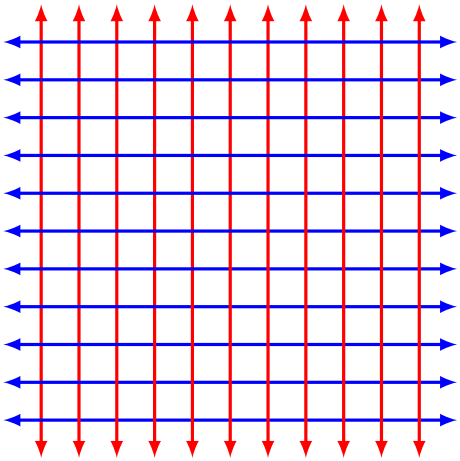


Figure 1.2: Coordinate grid on the Cartesian plane built from copies of the real line $\mathbb{R}$ to form the 2-dimensional product space $\mathbb{R}^2$ that covers the entirety of the 2-dimensional plane. Each horizontal line (blue) corresponds one copy of $\mathbb{R}$ and each vertical line (red) corresponds another copy of $\mathbb{R}$, and the arrows at the end of each line indicates that the lines representing each copy extend to an infinite distance in each direction.

An arbitrary point in the plane corresponds to a point on the grid with associated coordinate pair (x, y)*. Since each copy of the real line in the pair assigns a unique point in* $\mathbb{R}^2$*, each pair* (x, y) *uniquely defines a point on the Cartesian plane. This one-to-one mapping between point in the plane and points in* $\mathbb{R}^2$ *ensures that this coordinatization is valid everywhere on the plane.*

There is no unique process to assign coordinates to a space. An example of this fact is demonstrated in Example 1.3, where a non-uniform coordinate grid is assigned to the two dimensional plane.

Example 1.3 (Riemann Normal Coordinates on the Cartesian Plane) *An alternative coordinate system on the* 2*-dimensional Cartesian plane can be constructed by choosing an origin on the plane and then laying out a copy of the positive real line. This single copy* $\mathbb{R}$ *defines a* 1*-dimensional subspace of the plane. Then, using a protractor, we can lay out additional copies of the* $\mathbb{R}$*, each having an infinitesimal angular separation from the last, forming a radial grid, see Figure 1.3. This coordinate system is called the* Riemann Normal Coordinates *(RNC) and corresponds to the familiar plane polar coordinates on* $\mathbb{R}^2$*.*

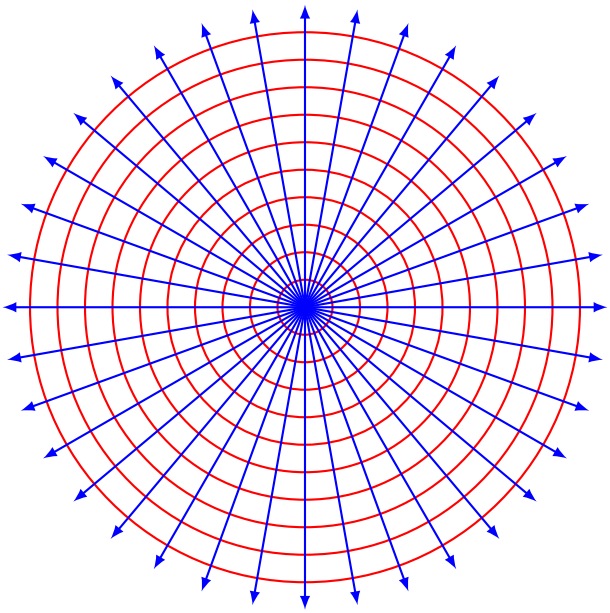


Figure 1.3: Riemann normal coordinates correspond to the standard plane polar coordinates in two dimensions. Radial lines (blue) emanating from the origin denote distance from the origin, while the angular displacement between adjacent radial lines correspond to a given angle of rotation about the origin. Circular (red) lines correspond to the loci of points at fixed radial distance from the origin.

Notice that the RNC fail to be one-to-one at the origin, since this point has a unique radial coordinate $r = 0$*, but non-unique angular position.*

The Riemann Normal Coordinates (RNC) are a good starting point to assign coordinates to a patch of a given space. In general, we do not expect that any single coordinate system to cover the whole of the space we want to study, but in the case of the 2-dimensional plane and sphere S^2, we can use these coordinates to cover the whole space, see Example 1.4. In this case, we see that RNC assign coordinates to the whole of S^2, but this coordinate system is not one-to-one at two points.

Example 1.4 (Coordinate grid on the Sphere) *A non-trivial example of Riemann normal coordinates from Example 1.3 follows by considering the case of the sphere S^2. Choose as the origin the north pole of the sphere S^2, and meet out an instance of the Riemann Normal coordinates. This corresponds, under an appropriate rescaling, to the standard spherical polar coordinates on the sphere.*

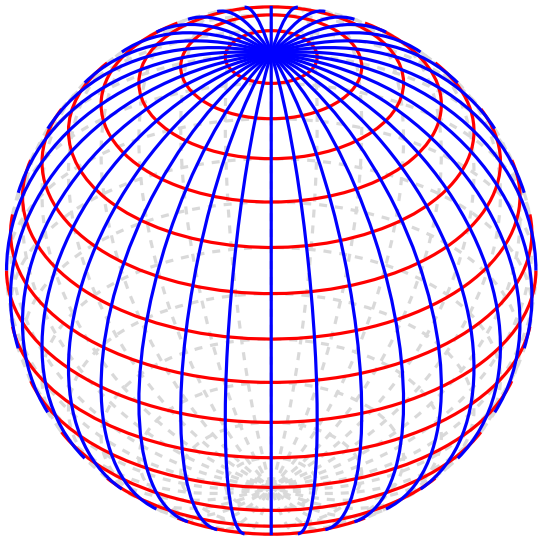


Figure 1.4: Polar coordinate grid on a sphere S^2. Radial lines (blue) emanating from the origin denote distance from the origin, while the angular displacement between adjacent radial lines correspond to a given angle of rotation about the origin. Circular (red) lines correspond to the loci of points at fixed radial distance from the origin.

Notice that in the case of S^2, the anti-podal point of the origin, corresponding to the south pole, is at a fixed "radial" distance from the origin. Beyond this antipodal point, the radial lines wrap back along S^2. This means that the Riemann normal coordinates provide a one-to-one mapping between only a subset of $\mathbb{R}^2$ onto S^2. There are two points on S^2 where the coordinate mapping breaks down, the north pole (with many possible angular displacement assignments) the south pole (with many possible radial distance assignments). More generally, RNC can be extended to more than two dimensions.

The torus is another example of a 2-dimensional space with interesting geometric and topological properties that is easily coordinatized using a rectangular coordinate grid. This is demonstrated in Example 1.5.

Example 1.5 (Coordinate grid on the Torus) *The 2-dimensional surface of the torus T^2 corresponds to a product of copies of the circle, attached in a specific way so that the joined copies form a surface with a* handle. *If we zoom into any single point on the T^2, we find a local description that suggests that the space is formed by a product of two copies of the circle, that is $T^2 \simeq S^1 \times S^1$. However, this is not a good description of the whole torus since there is a special arrangement of circles to form this surface. Figure 1.5 gives on choice of coordinates on the surface of the torus.*

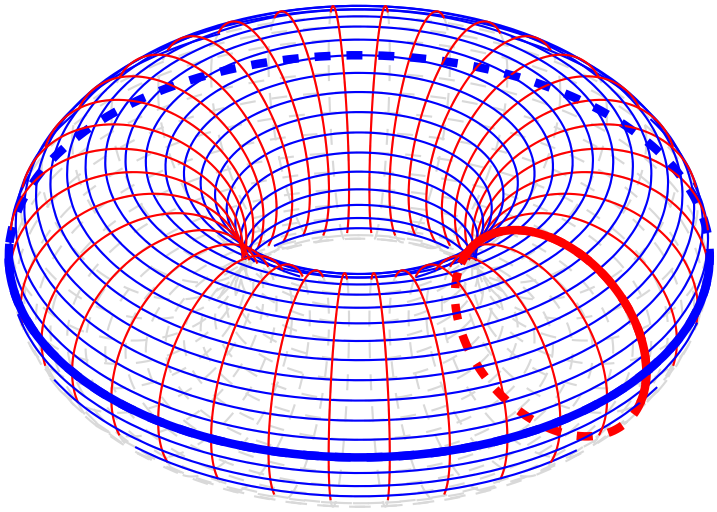


Figure 1.5: Coordinate grid on a torus. The red circle with circumference c_{red} and blue circle with circumference c_{blue} define the size of the torus. "Cutting" the torus along these two lines transforms the 2-dimensional surface of the torus in three dimensions into a flat, 2-dimensional rectangle corresponding to a subset of the 2-dimensional Cartesian plane.

We can add parameters to the surface of the torus by "cutting" the surface along the red line and straightening the resulting shape to form a cylinder, and then cutting along the blue line and then flattening the torus into a rectangle. We can then assign edge lengths c_{red} and c_{blue}, corresponding to the circumference of the red and blue circles drawn on T^2. We can identify the red edges of this rectangle with a pair of adjacent vertical lines in Figure 1.2 and similarly, identify the blue edges of this rectangle with a pair of adjacent horizontal lines in Figure 1.2. The surface T^2 now corresponds to a rectangular piece of the Cartesian plane when the red edges of length are identified (this means that the red line on the left-hand side of the rectangle is the red line on the right-hand side of the rectangle) and similarly, and the blue edges are identified. Clearly, there is more than one choice of rectangle in Figure 1.2 that can be matched, so there is a many to one matching between the $\mathbb{R}^2$ and T^2, and each $c_{red} \times c_{blue}$ rectangular block on the Cartesian plane is a good coordinate patch for T^2.

The length along a path that coincides with one of the coordinate curves on the space is easily determined by simply computing the difference between the final and initial position along that path. This is identical to the process by which we measure the length of an object using a ruler,

since the construction of the coordinate grid transfers the information from the ruler onto the space we wish to study. However, if the coordinate curve does not have a constant sized increment, or if the path deviates from the coordinate curve, then the path length is not determined by this simple procedure an a new method of computing lengths is needed. We shall discuss this method in the next section.

1.2 Euclidean Geometry

The solution to the problem of measuring the distance between marked points on a triangle drawn on a flat plane was known in the ancient world to the Greeks, Egyptians, Mesopotamians and Babylonians. The problem is described as follows. Given a right-angled triangle with given length and height, what is the extent of the diagonal edge? There exist many constructions for this length. An intuitive construction of this length relies on only the relation between similar triangles in $\mathbb{R}^2$.

Consider the triangle in Figure 1.6. Label the vertices of a triangle; an edge joining vertices in a triangle is a directed line element that is labelled by the vertices it intersects. In general, we may refer to a given edge $\vec{AB}$ and with length AB. We can prove by construction the similarity of these triangles as follows. Consider triangles $\triangle ABC$ and $\triangle ABD$ then

$$\begin{aligned} C\hat{A}B &= D\hat{A}B \quad \text{[common angle]} \\ A\hat{B}C &= A\hat{D}B \quad [90^\circ] \\ A\hat{C}B &= A\hat{B}D \quad \text{[sum of internal angles in triangle].} \end{aligned}$$

By equality of internal angles, we conclude that triangles $\triangle ABC$ and $\triangle ADB$ are similar and write $\triangle ABC \sim \triangle ADB$. Similar arguments supply $\triangle ABC \sim \triangle DCB$ and $\triangle ABD \sim \triangle DCB$. The size of an angle is a measure of displacement of a point in space relative to some given reference point from a third point. This displacement defines the ratios of the edge length between these points. The similarity of triangles in the collection $\{\triangle ABC, \triangle ABD, \triangle DCB\}$ implies a corresponding set of relations among ratios of edge lengths from one triangle with each of the other triangles in the collection. So,

$$\frac{AB}{AC} = \frac{AD}{AB} \quad \text{and} \quad \frac{DC}{BC} = \frac{BC}{AC}$$

and so

$$AB^2 = AD \cdot AC \quad \text{and} \quad BC^2 = DC \cdot AC.$$

and the sum of these becomes

$$AB^2 + BC^2 = AD \cdot AC + DC \cdot AC = (AD + DC) \cdot AC,$$

where $AD + DC = AC$, from which we determine

$$AB^2 + BC^2 = AC^2. \tag{1.1}$$

Equation (1.1) is the celebrated *Pythagorean Theorem* in Euclidean geometry.

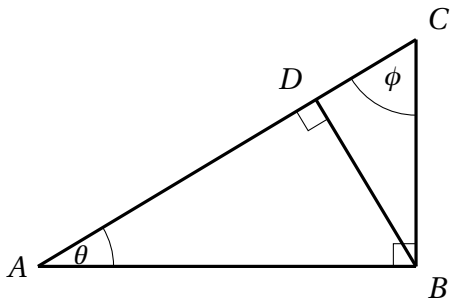


Figure 1.6: The triangle $\triangle ABC$ with right-angle $A\hat{B}C$ has internal angles equal to those in triangles $\triangle ADB$ and $\triangle DCB$. Equality of the internal angles among these triangles ensures that the ratios of the appropriate edge lengths of among triangles is maintained.

Remark 3 (Pythagorean Tripple) *An integral solution to* (1.1) *is called a* Pythagorean Triple. *Pythagorean triples are generated by the relation*

$$\left(n^2-m^2\right)^2+\left(2n^2m^2\right)^2=\left(n^2+m^2\right)^2 \tag{1.2}$$

for $n,m\in\mathbb{Z}$. *Note that this formulation follows directly from the factorisation of quartic polynomials.*

We shall use (1.1) as the basis of measurement in Euclidean space. We shall use this idea of length to define an operation that maps vectors to scalars that is compatible with our usual ideas of length. To construct the an expression for the lengths of vectors in the plane, we may use the *Law of Cosines* which follows from the geometry of the plane. Given the labelled triangle construction in Figure 1.7, we find the following relations among the edges,

$$\begin{aligned}(\vec{b}-\vec{a})\cdot(\vec{b}-\vec{a})&=\left\|\vec{b}-\vec{a}\right\|^2\\ \vec{a}\cdot\vec{a}+\vec{b}\cdot\vec{b}-2\vec{a}\cdot\vec{b}&=\|\vec{a}\|^2+\left\|\vec{b}\right\|^2-2\|\vec{a}\|\left\|\vec{b}\right\|\cos(\theta)\\ \vec{a}\cdot\vec{b}&=\|\vec{a}\|\left\|\vec{b}\right\|\cos(\theta)\end{aligned}$$

By the laws of cosines, we find a natural between the magnitudes of two vectors and the angle between them, and component wise multiplication and sum of two vectors. We call this the *dot* or *inner product*. Clearly, the inner product of a vector with itself returns the Pythagorean measure of its length. We have used only algebra and the Euclidean law of cosines to reach this outcome. We can extend this definition to more that two dimensions by adding more components to each vector, without changing any part of the formal relation or reference to coordinate directions.

The definition of relative distance defined by (1.1) and the *Law of Cosines* shall form the basis of all measurements of length to follow in these notes. In particular the dot product shall define the mathematical machinery that we shall use to measure lengths of objects.

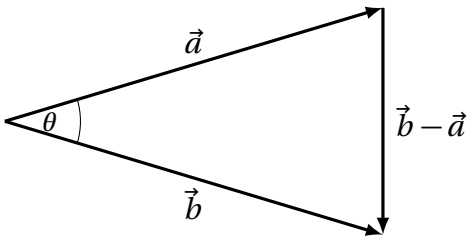


Figure 1.7: The *Law of Cosines* implies that the lengths adjacent edges and the angle between them encodes the length of the third edge of any triangle in the plane.

1.3 Coordinate Systems and Their Properties

The state of a particle depends on its position and momentum. For us to reason about these properties, we need a coordinate system that maps spatial positions onto a mathematical framework. Recall that the classical universe is an Affine Space. In whichever coordinate system we work, we must first select a reference point in our system and call it the origin.

René Descartes was the first to introduce the concept of coordinate systems to geometry. He assigned a pair of real numbers (x, y) to each point in the plane. The plane thus parameterised is known as the *Cartesian plane*.

1.3.1 The 2-Dimensional Cartesian Coordinate System

We begin by choosing two orthogonal directions in the space $\mathbb{R}^2$ and label them with the corresponding unit vectors $\hat{x}$ and $\hat{y}$ directions. The unit vectors in these two directions, being orthogonal, are linearly independent. Thus, since $\mathbb{R}^2$ is a 2-dimensional space, these unit vectors form an orthonormal basis for the plane. Every point in the plane can then be expressed as some linear combination of these unit vectors.

Figure 1.9 shows how the coordinate position of a marked point in $\mathbb{R}^2$ is described by a position vector $\vec{p}$ with a given extent in each of the coordinate unit vector directions. The general position $\vec{p}$ can be written as

$$\vec{p} = a\,\hat{x} + b\,\hat{y},$$

where $a = \vec{p}\cdot\hat{x}$ and $b = \vec{p}\cdot\hat{y}$. The pair (a, b) are called the *coordinates* of $\vec{p}$. In a purely symbolic sense, we may think of the above as

$$\vec{p} = \begin{pmatrix}\hat{x} & \hat{y}\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}$$

where standard row and column matrix multiplication is used to evaluate this product. The unit vectors $\hat{x}$ and $\hat{y}$ above are usually omitted on account of an implicit understanding of which basis is being used. Indeed, since every vector in the plane can now be uniquely identified with its coordinates, we can simply identify $\vec{p}$ with the tuple of coordinates (a, b),

$$\vec{p} := \begin{pmatrix}a\\ b\end{pmatrix}$$

where the position in the column of numbers is sufficient to define the different coordinate direction components.

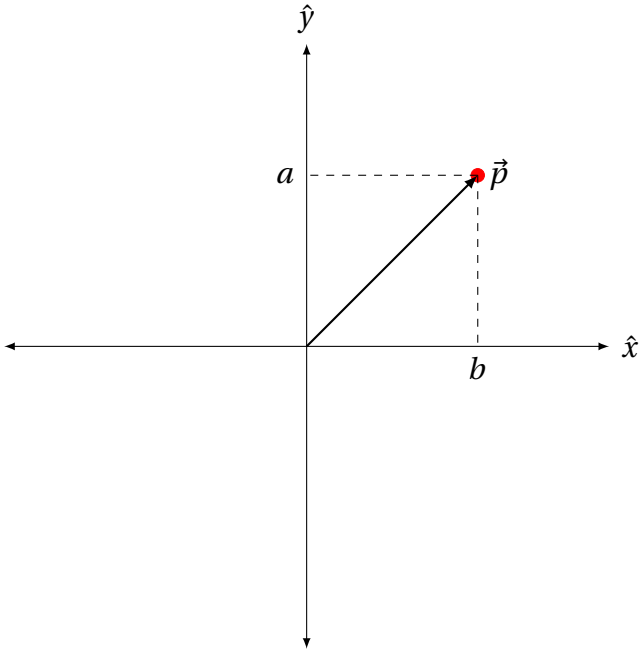


Figure 1.8: The coordinate direction axes in a 2-dimensional coordinate system that define the coordinate position of a marked point with position vector $\vec{p} = (a, b)^\top$.

When constructing a 2-dimensional Cartesian coordinate system to describe a given problem, we must choose three parameters, namely, the origin, the $\hat{x}$ direction and the $\hat{y}$ direction. Notice that the requirement that the $\hat{x}$ and $\hat{y}$ directions are orthogonal implies that any choice of one leaves one of two choices for the other (see Figure 1.9).

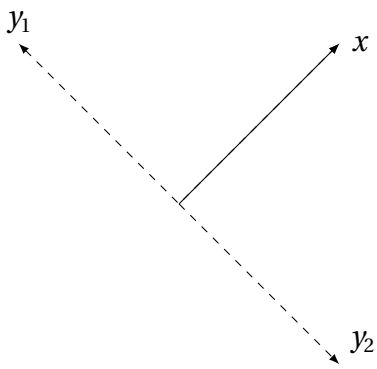


Figure 1.9: The choice of orthogonal coordinate direction axes in a 2-dimensional coordinate system.

Assume we have chosen a single point in the Affine Space of the Classical Universe to represent the origin of our coordinate system. There are infinitely many 2-dimensional Cartesian Coordinate Systems rooted at this single point. Consider first coordinate systems that are related by a *rotation* operation. Consider three 2-dimensional Cartesian coordinate systems that are rotated with respect to each other, as in Figure 1.10. Without loss of generality, assume that one of these is the standard (x, y) Cartesian coordinate system, which is represented with a horizontal x-axis and a vertical y-axis. Consider a point p coordinates (x, y) relative to a set of coordinate axes $\hat{x}$ and $\hat{y}$, (x', y') relative to a set of coordinate axes $\hat{x}'$ and $\hat{y}'$, and (x'', y'') relative to coordinate axes $\hat{x}''$, $\hat{y}''$, where $\hat{x}'$ and $\hat{y}'$ are obtained by rotating $\hat{x}$ and $\hat{y}$ through an angle θ; and $\hat{x}''$ and $\hat{x}''$ are obtained by rotating $\hat{x}'$ and $\hat{y}'$ through an angle ϕ. Let $\vec{p}$ be the directed line segment starting at the origin of the coordinate system and ending at p. Then $\vec{p} = p\hat{p}$ where $p = \left\|\vec{p}\right\|$ is the length of the vector $\vec{p}$. This is presented in Figure 1.10.

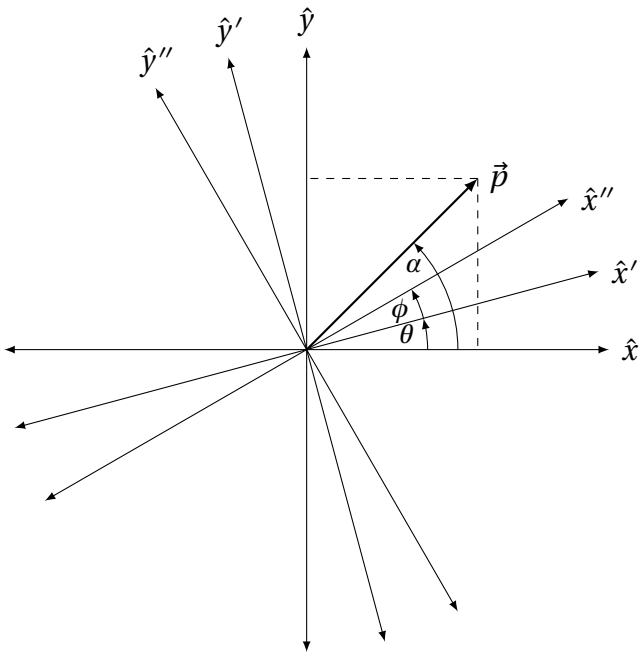


Figure 1.10: Multiple rotation of a coordinate system with a designated point P marked by the displacement vector $\vec{p}$. The $(\hat{x}'', \hat{y}'')$-coordinate system is rotated by ϕ relative to the $(\hat{x}', \hat{y}')$-coordinate system, which is itself rotated by θ relative to the $(\hat{x}, \hat{y})$-coordinate system. The vector $\vec{p}$ now has a different component representation in each coordinate system.

Next, consider the relative angular displacements of the each set of coordinate axes. The following identities are useful in the discussion to follow,

$$\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$$
$$\sin(\alpha+\beta) = \cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta).$$

From Figure 1.10 we find

$$x = p\cos(\alpha) \quad \text{and} \quad y = p\sin(\alpha).$$

and

$$x' = p\,\cos(\alpha-\theta) \quad \text{and} \quad y' = p\,\sin(\alpha-\theta).$$

We can decompose the compound angle expressions using the angular composition identities

$$\begin{aligned} x' &= p\,(\cos(\alpha)\cos(-\theta)+\sin(\alpha)\sin(-\theta)) \\ &= p\,\cos(\alpha)\cos(\theta)-p\,\sin(\alpha)\sin(\theta) \\ &= x\,\cos(\theta)+y\,\sin(\theta) \end{aligned}$$

and

$$\begin{aligned} y' &= p\,(\cos(\alpha)\sin(-\theta)+\sin(\alpha)\cos(-\theta)) \\ &= -p\,\cos(\alpha)\sin(\theta)+p\,\sin(\alpha)\cos(\theta) \\ &= -x\,\sin(\theta)+y\,\cos(\theta). \end{aligned}$$

We can express these relations in matrix form as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

The matrix,

$$R(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

specifies the rotation. Note that $R(\theta)$ carries coordinate pair (x, y) through an angle θ to (x', y') in the new coordinate system. The matrix R is called a *rotation matrix*. Note also, $R(-\theta)$ carries coordinate pair (x', y') through an angle θ to (x, y) in the original coordinate system, such that a rotation through an angle θ followed by a rotation through an angle $-\theta$ leaves the point (x, y) unchanged. Additionally,

$$R(-\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}^{\top}.$$

Therefore,

$$R(-\theta)R(\theta) = R^{\top}(\theta)R(\theta) \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

or,

$$R^{\top}(\theta)R(\theta) = \mathbb{1},$$

which implies that

$$R^{\top}(\theta) = R^{-1}(\theta).$$

Successive rotations can be represented by successive multiplications of the rotation matrix. If P has coordinates (x'', y'') with respect to $\hat{x}''$ and $\hat{y}''$, where these axes make an angle ϕ with $\hat{x}'$ and $\hat{y}'$, then,

$$x'' = x'\,\cos\big(\phi\big) + y'\,\sin\big(\phi\big) \quad \text{and} \quad y'' = -x'\,\sin\big(\phi\big) + y'\,\cos\big(\phi\big),$$

which is now a rotation of (x', y') through an angle ϕ. In matrix form

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

which is equal to a rotation through an angle $\phi + \theta$. Then,

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \cos(\phi + \theta) & \sin(\phi + \theta) \\ -\sin(\phi + \theta) & \cos(\phi + \theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

and $\phi + \theta$ is the angle between $\hat{x}$ and $\hat{y}$, and $\hat{x}''$ and $\hat{y}''$.

Remark 4 *Rotating a vector about the origin of a fixed coordinate system through an angle θ is equivalent to fixing that point and rotating the coordinate axes about the origin through an angle $-\theta$.*

Not all 2-dimensional Cartesian coordinate systems rooted at the same origin can be related by a rotation. For instance, the coordinate system resulting from swapping the x and y axes is *not* a rotation of the original coordinate system. The transformation matrix in this case is given by

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x'_p \\ y'_p \end{pmatrix},$$

It is clear that any 2-dimensional Cartesian coordinate system rooted at the same origin as the original (x, y)-system is either a rotation of the (x, y)-system, a swap of axes, or *both* a swap of axes *and* some rotation of the result. This latter transformation is simply the composition of the other two

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} x'_p \\ y'_p \end{pmatrix}.$$

Schematically, such transformations can be written as

$$\vec{r}_p = M \vec{r}'_p,$$

where M is either the rotation matrix presented earlier or the composition matrix above. Notice that $M^\top M = \mathbb{1}$, so M is an orthogonal matrix. We write $M \in O(2)$ to mean M is an orthogonal matrix of size 2. $O(2)$ is the orthogonal group of order 2. Observe that if M is a pure rotation then $\det(M) = 1$, and if M involves a switching of axes, then $\det(M) = -1$. The class of all pure rotations of two dimensional Cartesian coordinate systems is called $SO(2)$, the special orthogonal group of order 2.

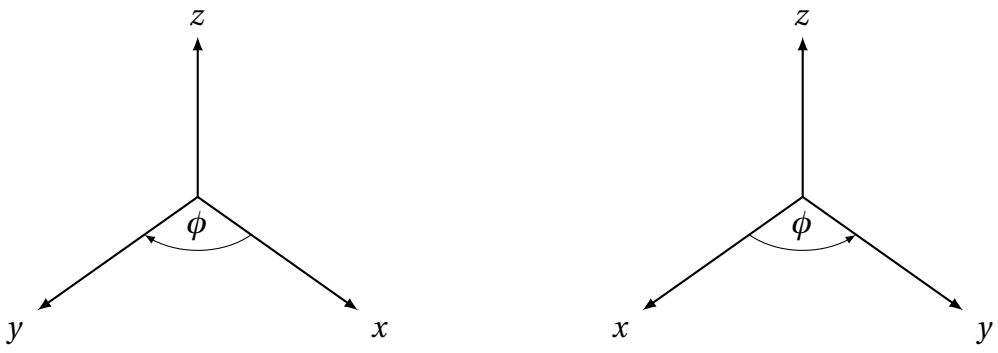
1.3.2 The 3-Dimensional Cartesian Coordinate System

A 3-dimensional Cartesian coordinate system is fully specified by four parameters. It is necessary assign to some point in the system the special property of being the *origin* of the system. Thinking of all other points in the universe by their difference vector from the origin we now have a vector

space $\mathbb{R}^3$. The other three parameters are the orthogonal unit vectors $\hat{x}, \hat{y}$ and $\hat{z}$. As in two dimensions the coordinates of any point can be written in terms of these unit vectors

$$\vec{p} = p_x\hat{x} + p_y\hat{y} + p_z\hat{z} = \begin{pmatrix}\hat{x} & \hat{y} & \hat{z}\end{pmatrix}\begin{pmatrix}p_x\\p_y\\p_z\end{pmatrix} := \begin{pmatrix}p_x\\p_y\\p_z\end{pmatrix}.$$

As before, once we have chosen two of these (say $\hat{x}$ and $\hat{y}$) we have two possible choices for the other (say $\hat{z}$). In this instance there is a physical significance attached to this choice - it determines the *handedness* of the resulting coordinate system.



Left-Handed Right-Handed

Figure 1.11: The relative orientation of the left-handed and right-handed coordinate systems.

Remark 5 (Handedness of a Coordinate System) *In a* Right-Handed *coordinate system, if you place your right hand at the origin, and point your fingers down the x axis, and then curl your hand towards the y -axis, your thumb will point up the z axis. In a* Left-Handed*y coordinate system we apply the same rule replacing the right hand with the left hand.*

The right-hand rule is used to determine the direction of the third axis given two other coordinate axis directions in three dimensions. We use this to give an orientation among triples of orthogonal coordinate unit vectors and define the vector cross-product of two vectors to give a formal mathematical operation that generates this orientation. We usually prefer to work in a Right-Handed coordinate system. In a right-handed system $\hat{x} \times \hat{y} = \hat{z}$ holds, while in a left-handed system $\hat{x} \times \hat{y} = -\hat{z}$ holds. The usual cross-product formulas are phrased for right-handed coordinate systems.

If we fix the origin, two sets of 3-dimensional Cartesian Coordinate systems are related to each other by a 3-dimensional Orthogonal Transformation

$$\vec{r}_p = M\vec{r}'_p,$$

where $M \in O(3)$. This fact is easy to take on face-value as a generalization of the result for 2-dimensional Cartesian Coordinates, and indeed, this result generalizes correctly to n-dimensional

space. This generalisation deserves some discussion; to this end, suppose we connect two sets of Cartesian Coordinates in n dimensions by an invertible linear transformation, Q. Then distances between points in space should be preserved by Q. No matter what n-dimensional Cartesian space we consider, the distance $\mathrm{d}(\vec{a},\vec{b})$ between two prescribed points, with position vectors $\vec{a}$ and $\vec{b}$, should be the same. Thus, we require

$$\mathrm{d}(\vec{a},\vec{b}) = \mathrm{d}(Q\vec{a}, Q\vec{b}).$$

Stated differently,

$$(\vec{a}-\vec{b})^\top(\vec{a}-\vec{b}) = (Q\vec{a}-Q\vec{b})^\top(Q\vec{a}-Q\vec{b}).$$

Simplifying this yields

$$(\vec{a}-\vec{b})^\top(\vec{a}-\vec{b}) = (\vec{a}-\vec{b})^\top Q^\top Q(\vec{a}-\vec{b}).$$

And since this must be true for all $\vec{a},\vec{b}\in\mathbb{R}^n$ it follows that

$$\vec{x}^\top\vec{x} = \vec{x}^\top Q^\top Q\vec{x} \quad \forall\vec{x}\in\mathbb{R}^n.$$

Let us name this central matrix $L = Q^\top Q$. Choose $\vec{x} = (0,0,\ldots,0,1,0,0,\ldots 0)^\top$ with the 1 in the i-th position. Then,

$$1 = \vec{x}^\top\vec{x} = \vec{x}^\top L\vec{x} = L_{ii}.$$

Since this holds for all i, the diagonal elements of L must all be 1's.

We can learn more about the structure of the transformation matrix L by considering the following neat construction. Choose $\vec{x} = (0,0,\ldots,0,1,0,\ldots,0,1,0,\ldots,0)^\top$ with 1's in the i-th and j-th positions and 0 everywhere else. This yields

$$2 = \vec{x}^\top\vec{x} = \vec{x}^\top L\vec{x} = L_{ii} + L_{jj} + L_{ij} + L_{ji}.$$

But we know from the previous result that $L_{ii} + L_{jj} = 1+1 = 2$. Combining these statements results in

$$L_{ij} + L_{ji} = 0.$$

But $L = Q^\top Q$ and so $L^\top = L$, meaning $L_{ij} = L_{ji}$. Putting these last two results together, we get $L_{ij} = 0, i \neq j$. Thus L is a diagonal matrix with 1s on the diagonal - the identity matrix. We have shown that $Q^\top Q = I$, and hence Q must be an orthogonal matrix to preserve distances. This confirms that in general transformations between n-dimensional Cartesian (orthogonal) coordinate Systems must be via orthogonal matrices. On the other hand, for every orthogonal transformation, the *angle* between two vectors is preserved by the transformation. This is because the *inner product* is preserved,

$$x^\top y = x^\top \mathbb{1} y = x^\top Q^\top Q y = (Qx)^\top(Qy).$$

So, if we started with an orthonormal basis, an orthogonal transformation will keep our basis orthonormal, and hence every orthogonal transformation takes a Cartesian coordinate system to another Cartesian Coordinate System.

In summary, every transformation between Cartesian coordinate systems is an orthogonal transformation, and every orthogonal transformation maps Cartesian coordinate systems to Cartesian coordinate systems. In other words the transformations mapping Cartesian coordinate systems to Cartesian coordinate systems are *precisely* the orthogonal transformations.

We remark further that the orthogonal matrices always have a determinant of either $+1$ or -1,

$$\det\left(Q^2\right) = \det(Q)\det(Q) = \det\left(Q^\top\right)\det(Q) = \det\left(Q^\top Q\right) = \det(\mathbb{1}) = 1.$$

Thus $\det(Q) = \pm 1$. Those with determinant $+1$ can be thought of as rotations. We call the collection of all of these matrices the *special orthogonal group* $SO(n)$. Those with determinant -1 can be thought of as first interchanging axes and then rotating. In 3-dimensions, transformations of determinant $+1$ preserve the *handedness* of the coordinate system, whilst transformations of determinant -1 reverse the *handedness* of the coordinate system.

1.3.3 Other Linear Coordinate Systems

In linear algebra, any set of linearly independent vectors that spans the space of interest can be thought of as a *basis* for that space. These vectors need not be of unit length, and they need not be mutually orthogonal. Any point in space can be expressed as a linear combination of these basis vectors. The coefficients in this expansion are the *coordinates* of the point in the linear basis.

We now consider the transformation equations between standard Cartesian coordinates and some linear coordinate system with basis $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$. These vectors need not be of unit length and they need not be mutually orthogonal. The only requirement is that they are *linearly independent*.

Let us consider an arbitrary point in space, $\vec{p}$. This point is written in both the original Cartesian coordinates and in the linear coordinate system as follows,

$$\vec{p} = p_1\hat{x}_1 + p_2\hat{x}_2 + p_3\hat{x}_3,$$
$$\vec{p} = p_1'\hat{x}_1 + p_2'\hat{x}_2 + p_3'\hat{x}_3.$$

Operating with $\hat{x}_1^\top$ in both equations yields

$$p_1 = p_1'\hat{x}_1^\top\vec{v}_1 + p_2'\hat{x}_1^\top\vec{v}_2 + p_3'\hat{x}_1^\top\vec{v}_3.$$

Similarly,

$$p_2 = p_1'\hat{x}_2^\top\vec{v}_1 + p_2'\hat{x}_2^\top\vec{v}_2 + p_3'\hat{x}_2^\top\vec{v}_3,$$
$$p_3 = p_1'\hat{x}_3^\top\vec{v}_1 + p_2'\hat{x}_3^\top\vec{v}_2 + p_3'\hat{x}_3^\top\vec{v}_3.$$

Thus,

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} \hat{x}_1^\top\vec{v}_1 & \hat{x}_1^\top\vec{v}_2 & \hat{x}_1^\top\vec{v}_3 \\ \hat{x}_2^\top\vec{v}_1 & \hat{x}_2^\top\vec{v}_2 & \hat{x}_2^\top\vec{v}_3 \\ \hat{x}_3^\top\vec{v}_1 & \hat{x}_3^\top\vec{v}_2 & \hat{x}_3^\top\vec{v}_3 \end{pmatrix}\begin{pmatrix} p_1' \\ p_2' \\ p_3' \end{pmatrix} = \begin{pmatrix} \hat{x}_1^\top \\ \hat{x}_2^\top \\ \hat{x}_3^\top \end{pmatrix}\begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix}\begin{pmatrix} p_1' \\ p_2' \\ p_3' \end{pmatrix} = XV\begin{pmatrix} p_1' \\ p_2' \\ p_3' \end{pmatrix}$$

In the above, X is the matrix whose column vectors are $\hat{x}_1$, $\hat{x}_2$ and $\hat{x}_3$ while V is the matrix whose column vectors are $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$. We are free to write the matrices X and V in any coordinate

system, as long as we use the same one for both. If we use the original Cartesian Coordinates, then $X = \mathbb{1}$, and the columns of V are the coordinates of the vectors $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$ in our original Cartesian Coordinate System.

With X and V defined as above, the same result can be derived directly

$$\vec{p} = X \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = V \begin{pmatrix} p_1' \\ p_2' \\ p_3' \end{pmatrix}.$$

And so,

$$\begin{pmatrix} p_1' \\ p_2' \\ p_3' \end{pmatrix} = X^{-1} V \begin{pmatrix} p_1' \\ p_2' \\ p_3' \end{pmatrix}. \tag{1.3}$$

Since X is orthogonal, $X^{-1} = X^\top$. In the general case of transforming between any two linear coordinate systems, the last step no longer applies, but the remainder of the argument is valid, and (1.3) gives the transformation rule between *any* two sets of linear coordinate systems. We can continue in this line to show in a different way that the matrix of transformation between two Orthonormal coordinate systems is orthogonal.

1.3.4 Curvilinear Coordinates

When dealing with orthonormal coordinate systems it is quite clear what the coordinate axes are. This concept generalizes to all coordinate systems. In particular, it generalizes to non-orthogonal coordinate systems as well. If we set all the coordinates to zero except for one, and allow that one coordinate to vary, we trace out the *coordinate axis* of that coordinate. Check that this concept yields the standard $x-$, $y-$ and z-axes in the Cartesian Coordinate systems. In arbitrary linear coordinate systems, the coordinate axes are lines emanating from the origin along the directions of the basis vectors.

A related concept is that of *coordinate curves*. Instead of setting the other coordinates to zero if we simply fix them as some prescribed constants, while allowing our chosen coordinate to vary, we obtain a *coordinate curve* for that coordinate. Each coordinate is then associated with a *family* of coordinate curves. In *linear* coordinate systems, these *curves* are lines parallel to the basis vectors.

Finally, we consider the idea of coordinate surfaces. In a 3-dimensional coordinate system, we can set one of the coordinates to be a constant value and allow the *other two* to vary. The figure that they trace out is called a coordinate surface. It should be clear that the coordinate surfaces of linear coordinate systems are planes.

So far we have covered only *linear* coordinate systems. That is, we've considered coordinate systems that made use of the vector-space nature of 2-dimensional and 3-dimensional space to assign coordinates to each point. In all of these systems the coordinate curves are lines and the coordinate surfaces are planes. However, it is not necessary to confine ourselves to linear representations of space. Indeed, any parameters that unambiguously label every point in space

can be thought of as coordinates. It is possible, and often useful, to use nonlinear coordinate systems for solving problems.

We consider here a family of coordinate systems collectively known as *Curvilinear Coordinates*. The name *Curvilinear Coordinates* is derived from the fact that the coordinate curves and coordinate surfaces are not necessarily straight lines and planes in these coordinate systems, but curves and curved surfaces. For every set of curvilinear coordinates we construct, we will consider the following aspects.

1.3.5 Transformation Equations

Consider the curvilinear coordinates $s^1, s^2, \ldots, s^n$. In order for a unique set of curvilinear coordinates to correspond to every point in space, we need to stipulate a correspondence between the Cartesian Coordinates of some point (in a well-specified Cartesian System), and the Curvilinear Coordinates of the same point. We achieve this through *transformation equations*

$$\begin{aligned} x^1 &= x^1\left(s^1, s^2, \ldots, s^n\right), \\ x^2 &= x^2\left(s^1, s^2, \ldots, s^n\right), \\ &\vdots \\ x^n &= x^n\left(s^1, s^2, \ldots, s^n\right). \end{aligned}$$

It is customary to refer to the coordinates with superscripts (s^i) rather than subscripts (s_i). There is a reason for this, but for now we simply accept this as convention.

We require that these transformation equations are *well-behaved*. We mean by this

1. The equations must be *locally invertible*. In some neighbourhood of every point there is an expression for the curvilinear coordinates in terms of the Cartesian Coordinates. If this was not the case the coordinates would not 'uniquely' label the points in some region of space.

2. The Jacobian Matrix of the transformation must be non-singular. This is necessary to ensure that the number of curvilinear coordinates at a point correspond to the number of coordinates that described that point and ensures that the transformation is well defined.

3. The transformation equations must be *differentiable*, and the derivatives must be continuous. Often they are *smooth* (infinitely differentiable). This requirement ensures that the coordinate curves and surfaces are indeed *curved*, as opposed to jagged.

The inverse function theorem tells us that point 3 above *follows* from points 1 and 2. From point 3, we can always invert the equations of transformation at some given point in space, so that we can write

$$\begin{aligned} x^1 &= x^1(s^1, s^2, \ldots, s^n) \\ &\vdots \end{aligned} \qquad \text{and} \qquad \begin{aligned} s^1 &= s^1(x^1, x^2, \ldots, x^n) \\ &\vdots \end{aligned}$$

$$x^n = x^n(s^1, s^2, \ldots, s^n) \qquad\qquad s^n = s^n(x^1, x^2, \ldots, x^n).$$

An informal motivation for these facts will follow after we've defined tangent vectors.

1.4 Coordinate Curves and Coordinate Surfaces

We will normally distinguish a particular curvilinear coordinate system according to the shapes of the coordinate curves (in the 2-dimensional case) or coordinate surfaces (in the 3-dimensional case). We will derive these objects for each coordinate system we discuss.

1.4.1 Parametric Curves

Parametric curves give a simple way to assign coordinates from one space to another. We can assign coordinates on the real line $\mathbb{R}$ to the real line $\mathbb{R}$ by linear transformation

$$f : \mathbb{R} \to \mathbb{R}$$
$$f(x) = a\,x + b = x'$$

where $a, b \in \mathbb{R}$. The mapping f assigns to each $x \in \mathbb{R}$ a unique image $x' \in \mathbb{R}$ for every $x \in \mathbb{R}$. For each choice of $a, b \in \mathbb{R}$ there corresponds a different map f. Clearly, the choice of f is not unique. When the spaces involved in the coordinate mapping are more complicated than $\mathbb{R}$, we should expect that the mapping function f will be more complicated, too. As an example of a more complicated mapping, suppose we consider a coordinatisation map f that assigns to each point in $\mathbb{R}$ a point on the unit circle S^1. Next, we consider one example of such a mapping known as a *projective map.*

Example 1.6 (The Projective Mapping $\mathbb{R} \to S^1$) *An efficient way to assign numbers to each point on the circle using a continuous mapping from the real line is by considering displacement vectors that connect points on the line to points on the circle. Suppose we mark the origin O as the zero point on the real line $\mathbb{R}$ and as the centre of the unit circle S^1, choose the point P as the point on the circle directly above O, and let Z be any point in $r \in \mathbb{R}$. Clearly*

$$O = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad P = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \textit{and} \quad Z = \begin{pmatrix} r \\ 0 \end{pmatrix}$$

We can think of P as the 'north pole' of the circle. Now construct the following position vectors

$$\vec{p} = P - O = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \textit{and} \quad \vec{v} = Z - P = \begin{pmatrix} r \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} r \\ -1 \end{pmatrix}$$

Notice that $\vec{p}$ connects the origin to the North pole and $\vec{v}$ connects the north pole to the point in $\mathbb{R}$. The line line element OP will intersect the circle S^1 at some point on its circumference. This line element with be directed along $\vec{v}$ but will have a length that depends on the position of Z. We can

construct a new vector that joins the north pole to the point of intersection between the line element OP and the circle using a linear combination of vectors $\vec{p}$ and $\vec{v}$,

$$\vec{w} = \vec{p} + \lambda\vec{v} = \begin{pmatrix} \lambda r \\ 1-\lambda \end{pmatrix} \quad \text{and} \quad \|\vec{w}\| = 1$$

where λ is a scaling parameter which adjusts the contribution of $\vec{v}$ in the definition of $\vec{w}$ and changes the length of $\vec{w}$ along the line connecting the north pole an z. An appropriate choice of λ will define the displacement vector $\vec{w}$ connecting the north pole to the point a on the circle. The parameter λ is called a Lagrange Multiplier.

If we enforce the requirement that the radius of the circle is 1*, then we can determine the value of the multiplier λ. We can break down the calculation as follows.*

$$\|\vec{w}\| = \begin{pmatrix} \lambda r \\ 1-\lambda \end{pmatrix} \cdot \begin{pmatrix} \lambda r \\ 1-\lambda \end{pmatrix} = \lambda^2 r^2 + (1-\lambda)^2 = 1$$

We compute the squared magnitude of $\vec{w}$ as a function of λ and set the value equal to 1

$$\lambda = \frac{2}{r^2+1}$$

and then solve λ as a function of the parameters of the problem.

$$\vec{w} = \begin{pmatrix} \frac{2r}{r^2+1} \\ \frac{r^2-1}{r^2+1} \end{pmatrix}$$

Finally, we recover a component wise expression for the value of $\vec{w}$. Now for every value of the parameter r, we assign a unique point to the circle. This argumentation can be generalized to higher dimensions.

Lagrange Multipliers are often used in optimisation problems. A single Lagrange multiplier is used in Example 1.6 to assign a specific value to a component in a vector sum, subject to a geometrical restriction. Lagrange Multipliers are especially useful in problems where we compare the lengths of parallel vectors. We shall revisit this concept later on when considering the constrained motion of objects.

Suppose we were to consider some general functions f that take input parameters r, s and t and output values in $\mathbb{R}, \mathbb{R}^2$ or $\mathbb{R}^3$. The generic functions f map elements from one space with some number of dimensions (number of inputs to f) into another space with a different number of dimensions (number of outputs from f). The functions a, b and c describe the components of the outputs of f. In each case, the number of input parameters determines the dimension of the output of f, regardless of the number of dimensions of the space where this output is sent. For simplicity, only the input and output sequences are listed.

$t \mapsto f(t)$**:** f map the real number line, $\mathbb{R}$, back to itself, where f can only stretch or compress parts of it.

$(r,s) \mapsto (a(r,s), b(r,s))$**:** f maps the 2-dimensional plane, back to the 2-dimensional plane with, where f can rotate, stretch or compress parts of it.

$(r,s) \mapsto (a(r,s), b(r,s), c(r,s))$**:** f maps the 2-dimensional plane into a subspace of 3-dimensional euclidean space. The outcome is some continuous surface in 3 dimensions.

$t \mapsto (a(t), b(t), c(t))$**:** f maps the real line to a 1-dimensional curve in 3-dimensional Euclidean space. The outcome of this mapping is some smooth, continuous line in 3 dimensions.

In general, we shall use the word *curve* to mean a continuously connected, smooth, subspace of another space. In particular, we shall consider curves in $\mathbb{R}^3$. It is important to note that there is a class of coordinate transformation for each of the parameters r, s and t that leave the output unchanged and we consider these next.

Definition 3 (Affine Transformation) *An* affine transformation *is any transformation that preserves collinearity and ratios of distances.*

An *affine transformation* is any transformation that preserves collinearity (i.e., all points lying on a line initially still lie on a line after transformation) and ratios of distances (e.g., the midpoint of a line segment remains the midpoint after transformation). In particular, we may transform the parameter t along 1-dimensional curves such that it lies in the range from 0 to 1 and think of this number as marking the percentage of the total length of the curve ($t = 0$ at the beginning of the curve and $t = 1$ at the end).

1.4.2 Tangent Vectors

One of the reasons we require the transformation equations to be differentiable is that the derivatives give us useful information about the structure of the transformation. In particular, consider what happens when we fix all the coordinates at some point and allow one of them to vary infinitesimally. The motion thus produced is in the 'characteristic direction' of that coordinate. It is tangent to the coordinate curve at that point. Formally, we write the *tangent vector* associated with the coordinate s^i as

$$\vec{e}_i = \frac{\partial}{\partial s^i} \begin{pmatrix} x^1(s^1, s^2, \ldots, s^n) \\ x^2(s^1, s^2, \ldots, s^n) \\ \vdots \\ x^n(s^1, s^2, \ldots, s^n) \end{pmatrix} = \left(\frac{\partial \vec{x}}{\partial s^i} \right).$$

This idea is best understood through examples. The most basic example is Cartesian coordinates. Example 1.7 demonstrates the most simple *transformation equations* corresponding to the identity transformation of the coordinates curves. In this case the transformation equations are trivial.

Example 1.7 (Trivial Cartesian Coordinate Transformation) *Consider the identity coordinate transformation of the Cartesian coordinate system,*

$$x = x(x,y,z) = x, \quad y = y(x,y,z) = y, \quad \textit{and} \quad z = z(x,y,z) = z.$$

So, the tangent vectors are

$$\vec{e}_x = \frac{\partial}{\partial x}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \hat{x}, \quad \vec{e}_y = \frac{\partial}{\partial y}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \hat{y}, \quad \textit{and} \quad \vec{e}_z = \frac{\partial}{\partial z}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \hat{z},$$

These are the unit vectors *in the x, y and z directions respectively.*

For a general linear coordinate system with basis vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, the tangent vectors at any point can be easily shown to be exactly the basis vectors. This makes sense - the coordinate curves run parallel to the basis vectors.

The tangent vectors are not always unit vectors. When they are not unit vectors, we can *normalize* them to obtain unit vectors in the coordinate directions. The sizes of the tangent vectors are called the *metric coefficients*, h_i, of the coordinate system. We write,

$$\vec{e}_i = h_i \hat{s}_i,$$

where $\hat{s}_i$ is a unit vector in the same direction as $\vec{e}_i$ that we can think of as the unit vector in the i-th coordinate direction.

Consider the Jacobian Matrix, J, of the coordinate transformation.

$$J = \left(\frac{\mathrm{d}\,\vec{x}}{\mathrm{d}\,\vec{s}}\right) = \begin{pmatrix} \left(\frac{\partial x^1}{\partial s^1}\right) & \cdots & \left(\frac{\partial x^1}{\partial s^n}\right) \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial x^n}{\partial s^1}\right) & \cdots & \left(\frac{\partial x^n}{\partial s^n}\right) \end{pmatrix}.$$

Clearly, the columns of J are the tangent vectors

$$J = \begin{pmatrix} \vec{e}_1 & \vec{e}_2 & \ldots & \vec{e}_{n-1} & \vec{e}_n \end{pmatrix}.$$

The matrix will be singular if and only if the tangent vectors are linearly *dependent*; but this means that the tangent space would collapse (its dimension would decrease), and the coordinate curves would coincide. We do not want this to occur because it implies that locally (very zoomed in), the coordinate space itself will collapse and not be invertible. This explains why a sensible transformation must have a non-singular Jacobian Matrix at all points, and keeps our intuition sharp about a requirement that might otherwise seem arbitrary and unnatural. The interested reader should read up on the *Inverse Function Theorem* (or the *Constant Rank Theorem*) to formally understand this point.

For the *well behaved* coordinate systems that we will consider, it follows that the tangent vectors are always linearly independent and will hence always form a basis for the tangent space, which is always n-dimensional. When we zoom in very close to a point in coordinate space,

the space begins to look like the tangent space. If we consider the region of space enclosed by varying each coordinate over an increasingly small interval, this region will begin to resemble the fundamental parallelogram or parallelepiped of the tangent space. The (signed) area/volume of this entity is given by the determinant of the matrix formed by placing the tangent vectors as column vectors i.e. the Jacobian Matrix. This reasoning leads us to a result well known in Multivariable Calculus

$$\mathrm{d}\,x^1\,\mathrm{d}\,x^2\ldots\mathrm{d}\,x^n = |\det(J)|\,\mathrm{d}\,s^1\,\mathrm{d}\,s^2\ldots\mathrm{d}\,s^n\,.$$

In many of the examples we study, the tangent vectors will be *orthogonal*. In this case the *unit vectors* will be orthonormal, and we will have

$$\det(J) = \det\begin{pmatrix}\vec{e}_1 & \ldots & \vec{e}_n\end{pmatrix} = \det\begin{pmatrix}h_1\hat{s}_1 & \ldots & h_n\hat{s}_n\end{pmatrix} = h_1\ldots h_n\,\det\begin{pmatrix}\hat{s}_1 & \ldots & \hat{s}_n\end{pmatrix}.$$

This last matrix of unit vectors is an orthogonal matrix - we can see this by pre-multiplying it by its transpose

$$\begin{pmatrix}\hat{s}_1^\top\\ \hat{s}_2^\top\\ \vdots\\ \hat{s}_n^\top\end{pmatrix}\begin{pmatrix}\hat{s}_1 & \hat{s}_2 & \ldots & \hat{s}_n\end{pmatrix} = \begin{pmatrix}\hat{s}_1^\top\hat{s}_1 & \hat{s}_1^\top\hat{s}_2 & \ldots & \hat{s}_1^\top\hat{s}_n\\ \hat{s}_2^\top\hat{s}_1 & \hat{s}_2^\top\hat{s}_2 & \ldots & \hat{s}_2^\top\hat{s}_n\\ \hat{s}_3^\top\hat{s}_1 & \hat{s}_3^\top\hat{s}_2 & \ldots & \hat{s}_3^\top\hat{s}_n\\ \vdots & \vdots & \ddots & \vdots\\ \hat{s}_n^\top\hat{s}_1 & \hat{s}_n^\top\hat{s}_2 & \ldots & \hat{s}_n^\top\hat{s}_n\end{pmatrix} = \begin{pmatrix}1 & 0 & \ldots & 0\\ 0 & 1 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & 1\end{pmatrix} = \mathbb{1}$$

Now the determinant of an orthogonal matrix is either $+1$ or -1, with the sign depending on the order in which we supply the rows. The absolute value of this is always $+1$, and thus we get

$$|\det(J)| = h_1 h_2\ldots h_n,$$

with the convention that we have chosen all the h_i to be positive. Thus when the tangent vectors are orthogonal, we can compute the area/volume elements very easily. We simply take the product of the metric coefficients.

1.4.3 Cotangent Vectors

It is worth remarking at this point that there are actually *two* kinds of direction vectors associated with any change of coordinates. The first kind is the *tangent vectors* discussed above, which we can think of as the partial derivatives $\left(\frac{\partial}{\partial s^i}\right)$. The second kind is a set of vectors known as the *cotangent vectors*. These are associated with differentials $\mathrm{d}\,s^i$. From the chain rule we can write out the differential as

$$\mathrm{d}\,s^i = \left(\frac{\partial\,s^i}{\partial\,x^1}\right)\mathrm{d}\,x^1 + \left(\frac{\partial\,s^i}{\partial\,x^2}\right)\mathrm{d}\,x^2 + \cdots + \left(\frac{\partial\,s^n}{\partial\,x^n}\right)\mathrm{d}\,x^n\,.$$

This last expression looks like a dot product, and can be rewritten as

$$\mathrm{d}\,s^i = \begin{pmatrix}\left(\frac{\partial\,s^i}{\partial\,x^1}\right)\\ \left(\frac{\partial\,s^i}{\partial\,x^2}\right)\\ \vdots\\ \left(\frac{\partial\,s^i}{\partial\,x^n}\right)\end{pmatrix}\cdot\begin{pmatrix}\mathrm{d}\,x^1\\ \mathrm{d}\,x^2\\ \vdots\\ \mathrm{d}\,x^n\end{pmatrix}.$$

This leads us to directly associate the differential $\mathrm{d}\,s^i$ with the vector

$$\vec{e}^{\,i} = \begin{pmatrix} \left(\frac{\partial s^i}{\partial x^1}\right) \\ \left(\frac{\partial s^i}{\partial x^2}\right) \\ \vdots \\ \left(\frac{\partial s^i}{\partial x^n}\right) \end{pmatrix} = \left(\frac{\partial s^i}{\partial \vec{x}}\right).$$

We call this vector, evaluated at some point, the i-th cotangent vector at the given point. For Cartesian coordinates, the tangent and cotangent vectors coincide.

The inner products between the cotangent vectors are known as the components of the *contravariant metric tensor*

$$g^{ij} = \vec{e}^{\,i} \cdot \vec{e}^{\,j}.$$

The cotangent vectors are always linearly independent for the *well behaved* systems that we will consider. This follows from the result of the next section.

1.4.4 Tangent and Cotangent Vector Component Relations

In general it is not necessary for the *tangent vectors* to be orthogonal. Likewise the *cotangent vectors* are not necessarily orthogonal. However, the *tangent* and *cotangent* vectors satisfy a condition of *mutual orthogonality*

$$\vec{e}_i \cdot \vec{e}^{\,j} = \vec{e}^{\,j} \cdot \vec{e}_i = \begin{pmatrix} \left(\frac{\partial s^j}{\partial x^1}\right) \\ \left(\frac{\partial s^j}{\partial x^2}\right) \\ \vdots \\ \left(\frac{\partial s^j}{\partial x^n}\right) \end{pmatrix} \cdot \begin{pmatrix} \left(\frac{\partial x^1}{\partial s^i}\right) \\ \left(\frac{\partial x^2}{\partial s^i}\right) \\ \vdots \\ \left(\frac{\partial x^n}{\partial s^i}\right) \end{pmatrix} = \sum_{k=1}^{n} \left(\frac{\partial s^j}{\partial x^k}\right)\left(\frac{\partial x^k}{\partial s^i}\right) = \left(\frac{\partial s^j}{\partial s^i}\right).$$

Where the second last step follows from the chain rule, and the last follows from the fact that the s^i are functionally independent (not constrained or related to each other by any equations). This relationship can be written more elegantly by making use of the *Kronecker δ-function*,

$$\delta_i^j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.4}$$

We can state the mutual orthogonality relation as

$$\vec{e}_i \cdot \vec{e}^{\,j} = \delta_i^j.$$

Because of this relation, the tangent and co-tangent vectors are called *reciprocal bases*. As will become apparent shortly, this is a very useful result. We now obtain some geometric insight into the mutual orthogonality identity in Figure 1.12.

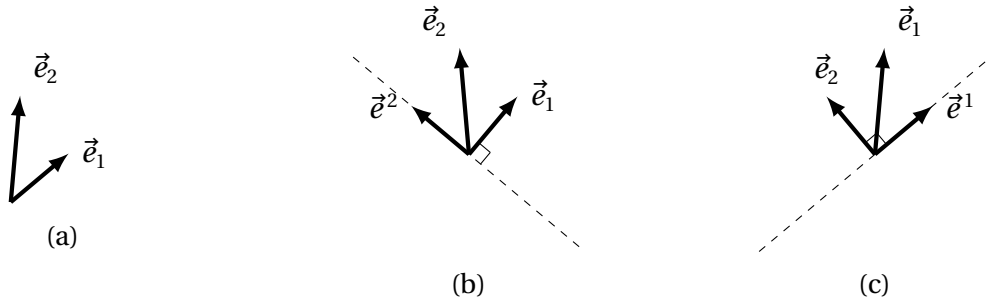


Figure 1.12: Mutual orthogonality of tangent and co-tangent vectors In figure (a) we begin with our original basis vectors (tangent vectors). Figure (b) shows how the mutual orthogonality condition completely determines the vector $\vec{e}^2$; firstly the constraint that it must be orthogonal to $\vec{e}_1$ forces it to be on the blue line illustrated above. Secondly the constraint that the dot product $\vec{e}_2 \cdot \vec{e}^2 = 1$ forces it to make an acute angle with the vector $\vec{e}_2$ and determines its length completely. Finally, figure (c) illustrates the same idea for choosing the other reciprocal vector, $\vec{e}^1$.

It is clear from the example that the mutual orthogonality condition completely specifies the basis $\left\{\vec{e}^1, \ldots, \vec{e}^n\right\}$ once the basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is known. The geometric method is purely algebraic and does not rely on the direct calculation of the co-tangent vectors. Notice how neither set of basis vectors is orthogonal in this example, but mutual orthogonality still applies.

The purely algebraic method presented above generalizes to any number of dimensions. We consider the Jacobian matrix, J, whose columns are the tangent vectors. Because the coordinate transformations are *well behaved*, J is invertible and so J^{-1} exists and is unique. Let the rows of J^{-1} be $\left(\vec{r}_{1^\top}, \vec{r}_{2^\top}, \ldots, \vec{r}_n^\top\right)$. Then,

$$\mathbb{1} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix} = J^{-1}J = \begin{pmatrix} \vec{r}_1^\top \\ \vec{r}_2^\top \\ \vdots \\ \vec{r}_n^\top \end{pmatrix} \begin{pmatrix} \vec{e}_1 & \vec{e}_2 & \ldots & \vec{e}_n \end{pmatrix} = \begin{pmatrix} \vec{r}_1^\top e_1 & \vec{r}_1^\top e_2 & \ldots & \vec{r}_1^\top e_n \\ \vec{r}_2^\top e_1 & \vec{r}_2^\top e_2 & & \vec{r}_2^\top e_n \\ \vdots & & \ddots & \\ \vec{r}_n^\top e_1 & & & \vec{r}_n^\top e_n \end{pmatrix}.$$

It clearly follows that the rows of J^{-1} are mutually orthogonal with the columns of J, and that this property defines J^{-1} so that these are the unique vectors satisfying it. Clearly then the rows of J^{-1} are the cotangent vectors $\vec{r}_i = \vec{e}^i$.

One consequence of the above argument is that, for our *well behaved* coordinate systems

$$\begin{pmatrix} \left(\frac{\partial x^1}{\partial s^1}\right) & \cdots & \left(\frac{\partial x^1}{\partial s^n}\right) \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial x^n}{\partial s^1}\right) & \cdots & \left(\frac{\partial x^n}{\partial s^n}\right) \end{pmatrix}^{-1} = \begin{pmatrix} \left(\frac{\partial s^1}{\partial x^1}\right) & \cdots & \left(\frac{\partial s^1}{\partial x^n}\right) \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial s^n}{\partial x^1}\right) & \cdots & \left(\frac{\partial s^n}{\partial x^n}\right) \end{pmatrix}$$

Stated differently, the Jacobian of the inverse of some transformation is the inverse of the Jacobian of the transformation.

As a corollary we find that if the tangent vectors are orthonormal, then $J^{-1} = J^\top$, and so the tangent vectors and the cotangent vectors coincide. What happens when the tangent vectors are orthogonal, but not orthonormal?

1.4.5 The Metric Tensor

The Cartesian tangent vectors are orthogonal, but clearly the tangent vectors are not orthogonal for all coordinate systems. In a linear coordinate system, for instance, the tangent vectors are only orthogonal if the basis vectors are orthogonal. The inner products between the tangent vectors are known as the components of the *covariant metric tensor*

$$g_{ij} = \vec{e}_i \cdot \vec{e}_j. \tag{1.5}$$

Clearly $h_i = \sqrt{g_{ii}}$. For an orthogonal system, g_{ij} is zero whenever $i \neq j$. For a general coordinate system, the tangent vectors at a given point form a *basis* for a *linear space* rooted at that point. We call this space the *tangent space* of the coordinate space at that point. The tangent vectors are sometimes referred to as the *direction vectors* at a point. The tangent space can be thought of as a very zoomed in picture of coordinate space near a point. Formally, it is the linearisation of coordinate space at the point. As we zoom closer and closer to the point, the coordinate space 'flattens out' into the tangent space.

We may construct an object **g** whose components are exactly the g_{ij} of (1.5). By indexing into **g** we may extract each component g_{ij} of **g** in a given computation. Now, the dot product of a vector $\vec{u}$ with a vector $\vec{v}$ in N-dimensions, in a given coordinate basis $\{\vec{e}_k\}$ is expressed as

$$\begin{aligned}
\vec{u} \cdot \vec{v} &= \left(u^1 \vec{e}_1 + u^2 \vec{e}_2 + \dots u^N \vec{e}_N\right) \cdot \left(v^1 \vec{e}_1 + v^2 \vec{e}_2 + \dots v^N \vec{e}_N\right) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} u^i v^j \vec{e}_i \cdot \vec{e}_j \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} g_{ij} u^i v^j.
\end{aligned}$$

It is common to omit the summation symbols and write instead

$$\vec{u} \cdot \vec{v} = g_{ij} u^i v^j,$$

where the summation over the indices i and j is implicit. This implicit summation is referred to as a *summation convention.*

A point in space has two basic sets of coordinates

1. Coordinates in terms of the Tangent Vectors (known as the *Contravariant Components of the Vector*)
$$\vec{v} = v^1 \vec{e}_1 + v^2 \vec{e}_2 + \dots + v^m \vec{e}_m.$$
Conventionally these components are written with a superscript v^i.

2. Coordinates in terms of the Co-tangent Vectors (known as the *Covariant Components of the Vector*)
$$\vec{v} = v_1 \vec{e}^1 + v_2 \vec{e}^2 + \dots + v_m \vec{e}^m$$
Conventionally these components are written with a subscript v_i.

This is nothing especially new - we know from linear algebra that we can write the same vector in terms of two different bases. What is special is the *two particular bases* we have chosen. Because of mutual orthogonality, we can determine the coordinates very easily. We simply take dot products with the tangent and cotangent vectors, respectively,

$$\vec{e}_i \cdot \vec{v} = \vec{e}_i \cdot \left(v_1 \vec{e}^{\,1} + v_2 \vec{e}^{\,2} + \cdots + v_m \vec{e}^{\,m} \right) = v_i$$
$$\vec{e}^{\,i} \cdot \vec{v} = \vec{e}^{\,i} \cdot \left(v^1 \vec{e}_1 + v^2 \vec{e}_2 + \cdots + v^m \vec{e}_m \right) = v^i$$

Thus finding the *covariant* or *contravariant* components of some point is as easy as taking the dot product with the tangent or cotangent vectors, respectively. Clearly, it follows from the relationships of the dot products that for given vectors $\vec{u}$ and $\vec{v}$,

$$\vec{u} \cdot \vec{v} = g^{ij} u_i v_j = g_i{}^j u^i v_j = g_{ij} u^i v^j = g_{ij} u^j v^i \tag{1.6}$$

We interpret covariant components u_i as the elements of row matrices and contravariant v^j as the elements column matrices. Then we may replace the explicit reference to summation indices in the summation convention for computing inner products with the multiplication rules of linear algebra and write instead

$$\vec{u} \cdot \vec{v} = U^\top \mathbf{g} V \tag{1.7}$$

where U is the row matrix with components of $\vec{u}$, V is the column matrix with components of $\vec{v}$ and $\mathbf{g}$ is the metric tensor. This result will be useful when we are transforming vector equations from one coordinate system to another. We can replace the explicit reference to summation indices by employing matrix algebra.

Example 1.8 (Simple Matrix Dot Product) *Consider the dot product between the basis unit vectors $\hat{x}$ and $\hat{y}$ in the standard 2-dimensional rectilinear coordinate system. Clearly, $\hat{x} \cdot \hat{y} = 0$. We may rewrite these vectors as column matrices*

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \textit{and} \quad \hat{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and use the 2-dimensional identity matrix

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \hat{x} \cdot \hat{x} & \hat{x} \cdot \hat{y} \\ \hat{y} \cdot \hat{x} & \hat{y} \cdot \hat{y} \end{pmatrix}$$

as the metric tensor to write

$$\hat{x} \cdot \hat{y} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

Similarly, we find

$$\hat{x} \cdot \hat{x} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1$$

and

$$\hat{y}\cdot\hat{\hat{y}}=\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}=\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}=1$$

as expected. More generally, in orthonormal coordinate basis (defined using a system of mutually orthogonal unit vectors), $\vec{e}_i\cdot\vec{e}_j=\delta_{ij}$*, then*

$$g^{ij}=g_{ij}\quad \textit{and}\quad u^i=u_i$$

and so

$$\vec{u}\cdot\vec{v}=\begin{pmatrix}u_1 & u_2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}v^1\\ v^u\end{pmatrix}=\begin{pmatrix}u_1 & u_2\end{pmatrix}\begin{pmatrix}v^1\\ v^u\end{pmatrix}=u_1v^1+u_2v^2=u^1v^1+u^2v^2.$$

as expected, and the matrix multiplication is explicit. By similarly argument,

$$\vec{u}\cdot\vec{u}=\begin{pmatrix}u_1 & u_2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}u^1\\ u^2\end{pmatrix}=\begin{pmatrix}u_1 & u_2\end{pmatrix}\begin{pmatrix}u^1\\ u^2\end{pmatrix}=u^1v^1+u^2u^2=\vec{u}^2,$$

again, as expected.

The first interesting geometric quantity that we want to compute is the length of a curve given by a function $y=f(t)$. We make use of the vector calculus of the preceding sections to write the position vector of a point on the curve and then construct a sequence of line elements at each point along that curve whose length we can sum. Suppose that the parametrisation is such that the end points of the curve correspond to $t=0$ and $t=1$. The length of a 1-dimensional parametric curve, parameterised by t, with tangent $\vec{v}(t)$ is simply the integral of the magnitude of the tangent vector. We start by constructing a vector that defines a point on the curve at a time t,

$$\vec{a}(t)=t\,\hat{x}+f(t)\hat{y} \tag{1.8}$$

and then compute the velocity of the point $\vec{a}$ along the path

$$\vec{v}(t)=\left(\frac{\mathrm{d}\,\vec{a}}{\mathrm{d}\,t}\right)$$

corresponding to the tangent vector along the curve at a time t. The length along the path $\Delta\ell(t)$ that is traversed by a point starting at $\vec{a}(t)$, at a speed $\|v\|(t)$ and for a period of time Δt is given by

$$\Delta\ell(t)=\|\Delta\vec{a}(t)\|=\left\|\left(\frac{\mathrm{d}\,\vec{a}(t)}{\mathrm{d}\,t}\right)\Delta t\right\|=\|\vec{v}(t)\|\,\Delta t.$$

We rewrite this portion of the path in the limit that $\Delta t\to 0$ and recover infinitesimal element of the path

$$\mathrm{d}\ell^2=\vec{v}(t)\cdot\vec{v}(t)\,\mathrm{d}\,t^2=g_{ij}\,\vec{v}^i(t)\vec{v}^j(t)\,\mathrm{d}\,t^2.$$

Integrating over all such infinitesimal path elements returns the length over the entire path

$$\ell=\int_{y=f(x)}\mathrm{d}\ell=\int_0^1\mathrm{d}\,t\,\sqrt{\vec{v}(t)\cdot\vec{v}(t)} \tag{1.9}$$

and metric tensor is necessary to compute the dot product of the tangent vector with itself. The formulation can be extended to any smooth parametric curve curve $\vec{a}(t)$ in any number of dimensions.

We now consider several example calculations for the length, area and volume in different coordinates systems.

1.5 Coordinate System Examples

In this section we consider several examples of length, area and volume in different coordinate systems.

There is always more than one way to assign coordinates to a space. It will be useful to be able to translate between these different coordinate systems. As was stated that RNC correspond to the classic example of 2-dimensional plane polar coordinates. We call these curvilinear coordinates in 2-dimensions since there exists functions that translate the coordinate curves in one choice of coordinates to coordinate curves in another. Here we consider two parameters, the distance of a point from the origin r and its angle with the x-axis θ. The parameters r and θ completely characterise any point in the plane. We can write transformation equations for this coordinate system

$$x = r\cos(\theta) \quad \text{and} \quad y = r\sin(\theta). \tag{1.10}$$

These equations can be inverted - thus every point corresponds to an unique pair (r,θ). It is understood that we restrict θ to be in some interval of length 2π so that no two values of θ correspond to the same *physical angle*. Figure 1.3 demonstrates the standard (x,y) and (r,θ) coordinate description of a point in the 2-dimensional plane.

Are these transformation equations *well-behaved*? Certainly they are invertible - without even using equations we can see that every point in the plane has a unique polar representation and a unique Cartesian representation, and thus there is a one-to-one correspondence between them. The equations are also differentiable. As for the Jacobian Matrix, we shall leave the answer to this question to the discussion of tangent vectors. It turns out that it is always non-singular (which comes as no surprise, since we convinced ourselves that they were invertible). Recall that these are the family of curves formed when we let one coordinate vary while keeping the other constant.

If we vary r while keeping θ constant, we get a ray emanating from the origin at the angle θ out to infinity. These are the radial coordinate curves depicted in Figure 1.3. These curves are *lines*, and thus we see that there is no curvature associated with the r coordinate. If we vary θ while keeping r constant we get a *circle* of radius r and centre the origin. These are the circular coordinate curves depicted in Figure 1.3. Clearly these curves are not straight lines, so there is curvature associated with the θ coordinate. Fundamental geometry then tells us that the angle between the two families of curves is always 90degree. We will see this more formally when we consider the tangent vectors for polar coordinates.

Notice that the coordinate curves for each coordinate are a family of curves, not just one curve. When we fix the other parameter, then we get a specific curve in the family. For instance the

θ-curves are the circles centre the origin, but if we specifically look at the curve when $r=5$, then we single out the circle of radius 5. Similarly the r-curves are the rays from the origin, but if we specifically consider the curve where $\theta = 90$degree then we single out the positive y-axis.

We often make use of polar coordinates when solving problems involving circular motion or circular symmetry. Circles are coordinate curves in polar coordinates, so motion on a circle is described by just one of coordinate. Thus, it is convenient to reason in polar coordinates about physical systems where there is circular motion or circular symmetry.

It is a very simple matter to obtain the tangent vectors for any coordinate system once we have written down the transformation equations

$$\vec{e}_r = \frac{\partial}{\partial r}\begin{pmatrix} x(r,\theta) \\ y(r,\theta) \end{pmatrix} = \frac{\partial}{\partial r}\begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix},$$
$$\vec{e}_\theta = \frac{\partial}{\partial \theta}\begin{pmatrix} x(r,\theta) \\ y(r,\theta) \end{pmatrix} = \frac{\partial}{\partial \theta}\begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix} = r\begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}.$$

We notice that $\vec{e}_r$ is a unit vector, and $\vec{e}_\theta$ has length r. So, we can write

$$\hat{r} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \quad \text{and} \quad \hat{\theta} = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix},$$

and

$$\vec{e}_r = \hat{r} \quad \text{and} \quad \vec{e}_\theta = r\hat{\theta}.$$

The unit vectors $\hat{r}$ and $\hat{\theta}$ are often used when solving problems about circular motion. The former points radially outwards, and the latter points tangentially and counter-clockwise. In Figure 1.13, notice how the unit vectors are indeed tangent to the corresponding coordinate curves.

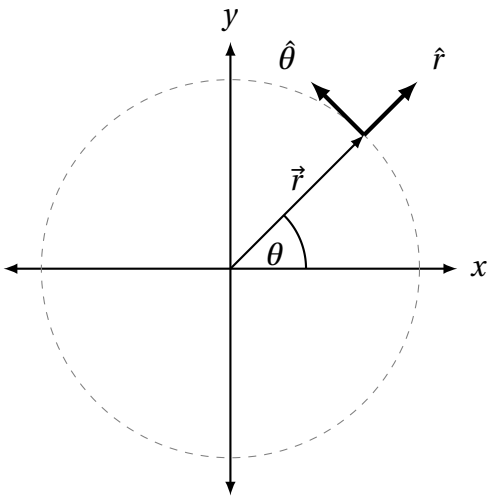


Figure 1.13: Unit vectors tangent to coordinate curves in 2-dimensions.

Earlier we remarked that the two classes of coordinate curves in polar coordinates are orthogonal. This is easy to show by using the tangent vectors

$$\begin{aligned}\vec{e}_r \cdot \vec{e}_\theta &= \begin{pmatrix}\cos(\theta) & \sin(\theta)\end{pmatrix} \cdot \begin{pmatrix}-r\sin(\theta) \\ r\cos(\theta)\end{pmatrix} \\ &= -r\cos(\theta)\sin(\theta) + r\cos(\theta)\sin(\theta) \\ &= 0.\end{aligned}$$

Thus the unit vectors are orthogonal for all values of r and θ, and hence everywhere in space. Clearly the two families of coordinate curves are always orthogonal.

We noticed earlier that $\hat{r}$ is a unit vector - and hence the r coordinate does not curve (or stretch) space. Indeed the coordinate curves are straight rays. We also saw that $hat\theta$ has a coefficient of curvature of r. Thus θ is a curved coordinate. This makes sense physically as θ is an angle, and its corresponding coordinate curves are circles.

Example 1.9 (Circumference of a Circle) *We can determine the circumference of a circle in the (x,y)-plane by starting with a polar coordinate representation of a position of a point on the circle using (1.10). In this coordinate system, a point $\vec{p}(t) = r\cos(\theta)\hat{x} + r\sin(\theta)\hat{y}$ traces a circular path of radius r, centered at the origin of the coordinate system, that is parameterised t. Tangents to this path are given by the parametric curve $\dot{\vec{p}}(t)$. Then, it follows*

$$\vec{p}(t) = r\begin{pmatrix}\cos(\theta(t)) \\ \sin(\theta(t))\end{pmatrix} r\hat{r} \quad \text{and} \quad \dot{\vec{p}}(t) = \dot{r}\begin{pmatrix}\cos(\theta(t)) \\ \sin(\theta(t))\end{pmatrix} r\dot{\theta}\begin{pmatrix}-\sin(\theta(t)) \\ \cos(\theta(t))\end{pmatrix} = \dot{r}\hat{r} + r\dot{\theta}\hat{\theta}$$

The radial coordinate is constant for motion along the circumference of a circle, so $\dot{r} = 0$ and we have immediately

$$\dot{\vec{p}}(t) = r\dot{\theta}\hat{\theta}.$$

Next we use (1.9) to integrate the lengths of each tangent vector along the path defined by the circumference of the circle,

$$\ell = \int_0^1 \mathrm{d}t\,\sqrt{\dot{\vec{p}}(t)\cdot\dot{\vec{p}}(t)} = \int_0^1 \mathrm{d}t\,\sqrt{r^2\dot{\theta}^2\hat{\theta}\cdot\hat{\theta}} = r\int_0^1 \mathrm{d}t\,\dot{\theta} = r\int_0^1 \mathrm{d}t\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right) = r\int_0^{2\pi}\mathrm{d}\theta$$

where the change of coordinates $\theta = 2\pi t$ ensures that a single circuit of the circle occurs in the time interval $t \in [0,1]$. Then,

$$\ell = r\int_0^{2\pi}\mathrm{d}\theta = 2\pi r$$

which is exactly what we should expect from the standard definition of the circumference of a circle.

Example 1.10 (Area of a disk) *We now consider the area element in 2-dimensional Polar Coordinates. We may obtain this in several ways. Our first method is a purely geometric argument. Consider the area element enclosed between coordinate curves of infinitesimal distance apart as depicted in Figure 1.14.*

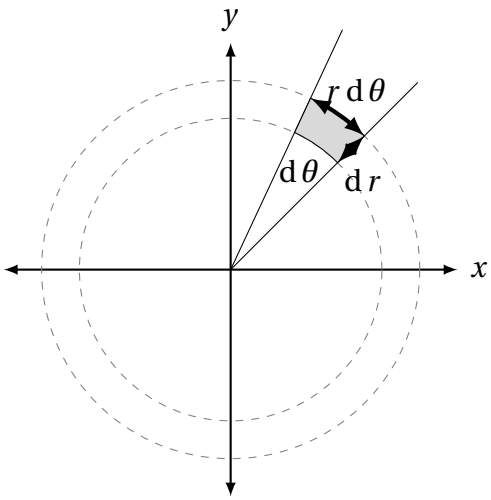


Figure 1.14: 2-dimensional-Polar-Coordinate-Area-Element

The area element is highlighted in Figure 1.14. We notice that because the curves are at 90 *degrees to each other, the area element starts to resemble a rectangle as we make its sides very small (the curvature on the 'circle' side flattens out in the limit). Thus we can write the area of the infinitesimal element as the product of the lengths of the sides. The area element is clearly given by*

$$\mathrm{d}A = \mathrm{d}r(r\,\mathrm{d}\theta) = r\,\mathrm{d}r\,\mathrm{d}\theta\,.$$

Notice how the orthogonality of the coordinate curves was required for this argument to work. Had the curves not been orthogonal, simply multiplying the side lengths in the limit would not give the correct area element. Notice also that the 'side lengths' of our infinitesimal area elements are our coefficients of curvature. This is no coincidence. With these two points in mind, we can now see that our geometric argument is in fact analogous to the argument presented in general in the preceding section, that the Jacobian Determinant can be given as the product of the coefficients of curvature provided that the tangent vectors are orthogonal. Thus in this case we could immediately write out

$$\mathrm{d}A = (1)(r)\mathrm{d}r\,\mathrm{d}\theta = r\,\mathrm{d}r\,\mathrm{d}\theta\,.$$

The third way of deriving this result is to evaluate the Jacobian determinant directly. This yields the same answer. As an example of the use of the area element, we will now compute the area of the disk of radius R,

$$\mathit{Area} = \int_S \mathrm{d}A = \int_0^R \int_0^{2\pi} \mathrm{d}r\,\mathrm{d}\theta\,r = 2\pi \int_0^R \mathrm{d}r\,r = \pi R^2.$$

We will see more of this type of integral when we study rigid bodies later in the course.

1.5.1 2-Dimensional Elliptical Coordinates

An example of a coordinate system in which the tangent vectors are not in general orthogonal is the elliptic coordinate system (u,ϕ), in which for some prescribed a and b, the transformation

equations are given by

$$x = a\,u\,\cos(\phi) \quad \text{and} \quad y = b\,u\,\sin(\phi).$$

Here we allow ϕ to vary between 0 and 2π and $u > 0$. It should be clear that polar coordinates are a special case of elliptic coordinates in which $a = b = 1$ and $r = u$. We get a stretched polar coordinate system when $a = b \neq 1$. The name should leave little surprise that the coordinate curves in elliptic coordinates are rays and ellipses. This is depicted in Figure 1.15.

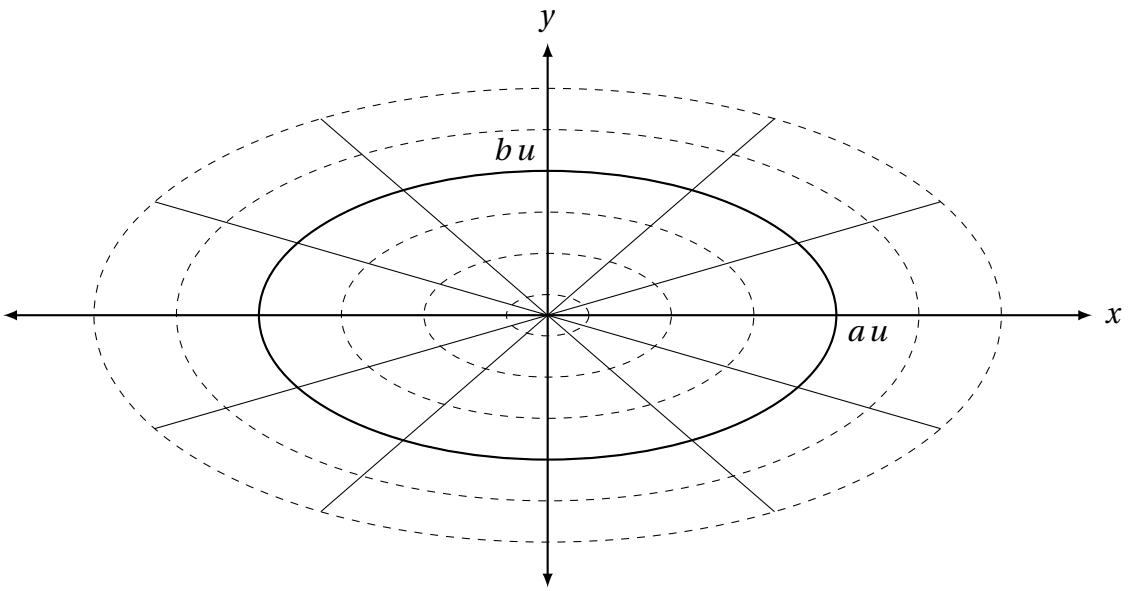


Figure 1.15: 2-dimensional elliptic coordinate curves. Notice that this coordinate system has a distorted area element. It is not in general the same as the product of the coefficients of curvature. This is because the coordinate curves intersect at an angle, and the infinitesimal unit of area is now a parallelogram.

It is evident in Figure 1.15 that the curves are *not* in general orthogonal. This is best seen by computing the tangent vectors

$$\vec{e}_u = \begin{pmatrix} a\,\cos(\phi) \\ b\,\sin(\phi) \end{pmatrix}, \quad \vec{e}_\theta = \begin{pmatrix} -u a\,\sin(\phi) \\ u b\,\cos(\phi) \end{pmatrix}.$$

The dot product is

$$\vec{e}_u \cdot \vec{e}_\theta = \begin{pmatrix} a\,\cos(\phi) & b\,\sin(\phi) \end{pmatrix} \begin{pmatrix} -a\,u\,\sin(\phi) \\ b\,u\,\cos(\phi) \end{pmatrix} = \frac{1}{2} u \left(b^2 - a^2\right) \sin(2\phi).$$

Clearly, this inner product is zero for all ϕ when $a = b$. When $a \neq b$ then the coordinate unit vectors are orthogonal only when $\phi = \pi n$ or $\phi = \frac{\pi}{2} + \pi n$ and $n \in \mathbb{Z}$.

Remark 6 *There is no known closed form solution for the circumference of an ellipse in terms of its semi-major axis lengths a and b. This quantity must be determined numerically, usually using computer-based numerical methods that implement the arc-length formula of* (1.9).

Example 1.11 (Area of an Ellipse in 2-Dimensional Elliptic Coordinates) *The coordinate grid in Figure 1.15 is distorted in a way that distinguishes it from that of the the polar coordinate grid in Figure 1.3. The coordinate curves will intersect such that the dot product between tangent vectors along the coordinate curves will vary along the path of each coordinate curve. Clearly $\vec{e}_u \cdot \vec{e}_\theta$ is only zero across all values of $\{u, \phi\}$ if we set $a = b$. For all other elliptical systems, the coordinate curves are not generally orthogonal. There will be places where the coordinate curves are orthogonal. Can you see this geometrically? Can you obtain these positions algebraically?. This means that the* coefficients of curvature *cannot be directly used in this case to obtain the area element. Instead we must compute the determinant of the Jacobian Matrix directly*

$$\det(\vec{e}_u \quad \vec{e}_\theta) = \det\begin{pmatrix} a\cos(\phi) & b\sin(\phi) \\ -ua\sin(\phi) & ub\cos(\phi) \end{pmatrix} = ab\,u\cos^2(\phi) + ab\,u\sin^2(\phi) = ab\,u.$$

So the area element in Elliptic Coordinates is $\mathrm{d}A = ab\,u\,\mathrm{d}u\,\mathrm{d}\phi$ *. We compute the area of an ellipse with a x-radius of a and a y-radius of b:*

$$Area = \int_S \mathrm{d}A = \int_0^1 \int_0^{2\pi} \mathrm{d}u\,\mathrm{d}\phi\,ab\,u = 2\pi ab \int_0^1 \mathrm{d}u\,u = \pi ab.$$

Notice that while this is easy to compute by other methods, with Elliptical Coordinates we could just as easily compute the area of an Elliptical Segment between two specified angles:

$$Area = \int_S \mathrm{d}A = \int_0^1 \int_\alpha^\beta \mathrm{d}u\,\mathrm{d}\phi\,ab\,u = ab \int_0^1 \mathrm{d}u\,u \int_\alpha^\beta \mathrm{d}\phi = \frac{\beta-\alpha}{2}ab.$$

We next consider some examples in 3-dimensions.

1.5.2 3-Dimensional Polar Cylindrical Coordinate

The 3-dimensional Polar Cylindrical Coordinate system uses three coordinates (ϕ, ρ, z) as shown in Figure 1.16. The coordinate ϕ defines an angle between the x-axis and some plane of interest (which passes through the z-axis) and is commonly referred to as the *azimuthal angle* or *azimuthal coordinate*. The coordinate ρ then gives the distance to travel in that plane without lifting. Finally, the coordinate z gives the height of the point of interest.

Using these definitions and Figure 1.16, we can write down the transformation equations

$$x = \rho\cos(\phi), \quad y = \rho\sin(\phi), \quad z = z,$$

where it is understood that $\rho \geq 0, 0 \leq \phi \leq 2\pi$. We can think of polar cylindrical coordinates as a *projection* of 2-dimensional *polar coordinates* into 3-dimensions by the addition of the z

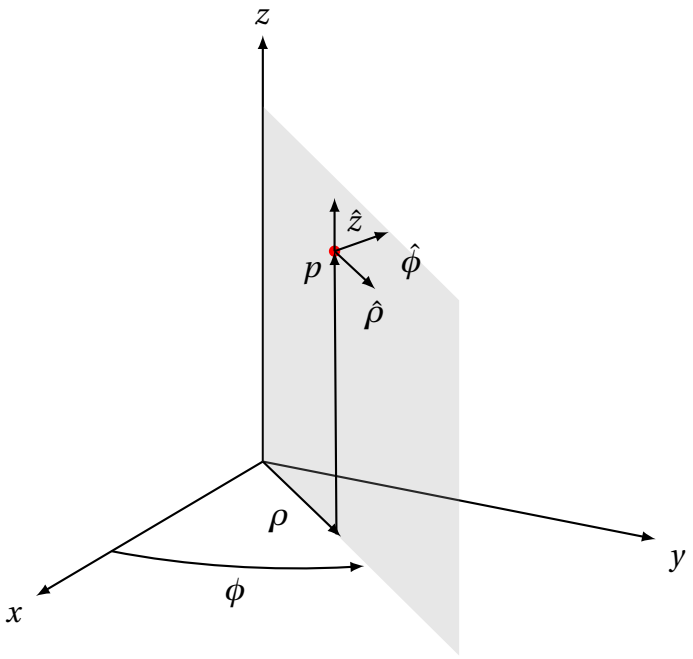


Figure 1.16: 3-dimensional polar cylindrical coordinate unit vectors.

coordinate. These transformation equations are clearly well-behaved. Now consider the *coordinate curves* for cylindrical coordinates. If we keep both ρ and z fixed and vary ϕ, we get a circle of an arbitrary radius centred at an arbitrary point on the z-axis. This is the ϕ-coordinate curve. If we keep both ϕ and z fixed and allow ρ to vary, we get a ray emanating from an arbitrary point on the z-axis in some arbitrary direction in a plane parallel to the $x-y$ plane. This is the ρ - coordinate curve. If we keep both ρ and ϕ constant and allow z to vary we get a vertical line passing through an arbitrary point in the $x-y$ plane specified by our choice of ρ and ϕ. This is the z-coordinate curve.

Remark 7 *For each of the 'coordinate curves' above, we've described a member of a family of curves. The coordinate curves for a coordinate are always a family of curves.*

Now we consider the *coordinate surfaces* for cylindrical coordinates. If we keep ρ fixed and allow ϕ and z to vary, then we get a cylinder of radius ρ and infinite height centred on the z-axis. For different values of ρ we will get cylinders of different radius, and this family of surfaces is the family of coordinate surfaces associated with varying ϕ and z. We can call it the ϕ, z family of coordinate surfaces. This coordinate surface is what gives cylindrical coordinates their name. Keeping ϕ fixed and varying the other two parameters produces a half-plane with one side along the z-axis. Similarly, keeping z fixed and varying the other two parameters produces a plane parallel to the $x-y$ axis.

When motion or symmetry exists along one of these coordinate surfaces or coordinate curves, then cylindrical coordinates will be a good choice of coordinates for the problem. Notice that the coordinate surfaces that were planes aren't interesting to us as we can consider planar motion using one of our 2-dimensional coordinate systems. Thus the cylinder is the truly interesting

object here. When our system is constrained to move in a spiral or some other motion on the surface of a cylinder, then Polar Cylindrical Coordinates may be a good choice.

As in any other coordinate system, we can obtain the tangent vectors directly from the transformation equations

$$\begin{aligned}\vec{e}_\rho &= \begin{pmatrix}\cos(\phi) & \sin(\phi) & 0\end{pmatrix}^\top = \hat{\rho},\\ \vec{e}_\phi &= \begin{pmatrix}-\rho\sin(\phi) & \cos(\phi) & 0\end{pmatrix}^\top = \rho\hat{\phi},\\ \vec{e}_z &= \begin{pmatrix}0 & 0 & 1\end{pmatrix}^\top = \hat{z},\end{aligned}$$

It is also relatively immediate what the sizes of these vectors are, and hence their relationships to the corresponding unit vectors. We notice in particular that the only coordinate with curvature is ϕ. This makes sense as it is the only coordinate for which the coordinate curve was not a line or ray. The unit vectors in cylindrical coordinates are illustrated in Figure 1.16. As per usual these vectors are tangential to the coordinate curves and in the direction of increase of their corresponding coordinates. ($\hat{z}$ points upwards, $\hat{\rho}$ points outwards, $\hat{\phi}$ points counter-clockwise).

Figure 1.16 suggests that the tangent vectors are orthogonal. Indeed, this should be quite clear from thinking of cylindrical coordinates as polar coordinates with an additional coordinate z. Formally, we can compute the pair-wise inner products between the tangent vectors

$$\vec{e}_\rho\cdot\vec{e}_\phi = \begin{pmatrix}\cos(\phi) & \sin(\phi) & 0\end{pmatrix}\begin{pmatrix}-\rho\sin(\phi)\\ \rho\cos(\phi)\\ 0\end{pmatrix} = 0,$$

$$\vec{e}_\rho\cdot\vec{e}_z = \begin{pmatrix}\cos(\phi) & \sin(\phi) & 0\end{pmatrix}\begin{pmatrix}0\\ 0\\ 1\end{pmatrix} = 0,$$

$$\vec{e}_z\cdot\vec{e}_\phi = \begin{pmatrix}0 & 0 & 1\end{pmatrix}\begin{pmatrix}-\rho\sin(\phi)\\ \rho\cos(\phi)\\ 0\end{pmatrix} = 0.$$

Thus indeed we see that the tangent vectors are orthogonal at *all* points in space. The families of coordinate curves are then orthogonal at all points in space.

Example 1.12 (Volume of a Cylinder in 3-Dimensianal Polar Cylindrical Coordinates) *In 3-dimensional polar cylindrical coordinates, we consider three coordinates (ϕ,ρ,z), see Figure 1.16. This orthogonality of the tangent vectors enables us to write down the volume element directly in terms of a product of the coefficients of curvature*

$$\mathrm{d}V = (\rho\,\mathrm{d}\phi)(1\,\mathrm{d}\rho)(1\,\mathrm{d}z) = \rho\,\mathrm{d}\phi\,\mathrm{d}\rho\,\mathrm{d}z.$$

As a simple application we compute the volume of a cylinder of radius R and height H

$$\textit{Volume} = \int_{\textit{cylinder}}\mathrm{d}V = \int_0^H\int_0^R\int_0^{2\pi}\mathrm{d}z\,\mathrm{d}\rho\,\mathrm{d}\phi\,\rho = \int_0^H\mathrm{d}z\int_0^R\mathrm{d}\rho\,\rho\int_0^{2\pi}\mathrm{d}\phi = \pi R^2 H.$$

1.5.3 3-Dimensional Polar Spherical Coordinates

In $3-d$ polar spherical coordinates we consider coordinates (ρ,θ,ϕ), where ϕ is once again the angle between the x-axis and the plane the point of interest makes with the z-axis and is called, again, *azimuthal angle* or *azimuthal coordinate*, θ is the angle *within* this plane between the z-axis and the point of interest and is called the *polar angle* or *polar coordinate*. The coordinate and ρ in this case is the *distance from the origin to the point of interest* and is commonly called the *radial position* or *radial coordinate*. Note in particular that ρ for spherical coordinates is different from ρ for cylindrical coordinates. See Figure 1.17.

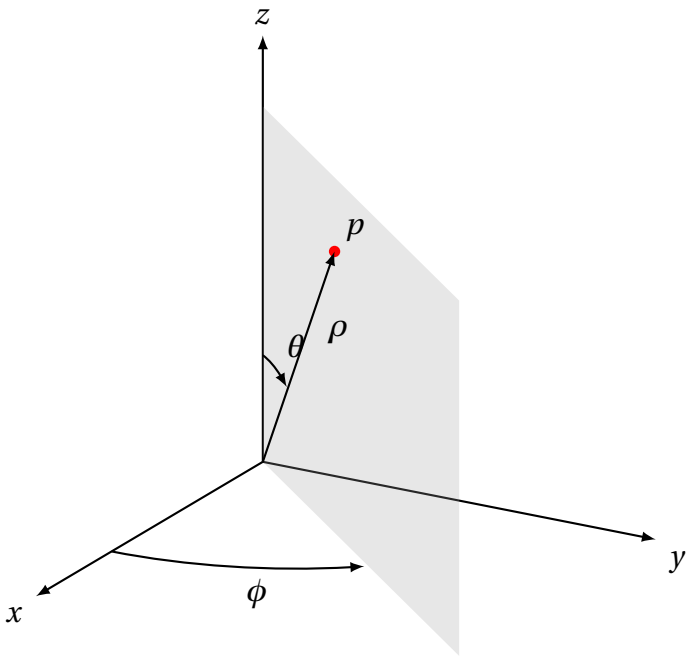


Figure 1.17: 3-Dimensional Polar Spherical Coordinate System.

As before, we shall relate these spherical coordinate system to other coordinate systems. From the Figure 1.17 and some elementary trigonometry, we arrive at the following transformation equations for spherical coordinates

$$x=\rho\,\sin(\theta)\cos\big(\phi\big),\quad y=\rho\,\sin(\theta)\sin\big(\phi\big),\quad \text{and}\quad z=\rho\,\cos(\theta),$$

where it is understood that $\rho\geq 0$, $0\leq\phi\leq 2\pi$ and $0\leq\theta\leq\pi$. (Challenge question, why do we not allow θ to vary all the way to 2π?) Notice that as we now have *two angles* we expect both of them to have a *curvature*.

If we keep θ and ρ constant and allow ρ to vary, we obtain a ray from the origin extending to infinity. This is a ρ-coordinate curve for spherical coordinates. If we keep ρ and θ constant and allow ϕ to vary, we get a circle centred somewhere on the z-axis and in a plane parallel to the $z=0$ plane. This is a ϕ-coordinate curve. If we keep ϕ and ρ constant and allow θ to vary we

obtain a semi-circle centred at the origin and with diameter on the z-axis (radius ρ, in a plane at angle ϕ with the x-axis). This is a θ-coordinate curve.

If we keep ρ constant and allow the other two parameters to vary we obtain a half-plane with one edge the z-axis. If we keep ρ constant and allow the other two parameters to vary, we obtain a sphere centred at the origin. If we keep θ constant and allow the other two parameters to vary, we obtain a cone with its apex at the origin and its axis of symmetry along the z-axis. These last two coordinate surfaces are of interest to us in problem solving. The sphere gives the coordinates the name *Spherical Coordinates*. But unlike with Cylindrical Coordinates there is a second non-trivial coordinate surface - the Cone. These coordinates are thus *also* suitable for motion that is constrained to a cone, or for conical spirals etc, as well as problems with conical symmetry.

As always we obtain the tangent vectors by taking partial derivatives of the transformation equations

$$\vec{e}_\rho = \frac{\partial}{\partial \rho}\begin{pmatrix} \rho\,\sin(\theta)\cos(\phi) \\ \rho\,\sin(\theta)\sin(\phi) \\ \rho\,\cos(\theta) \end{pmatrix} = \begin{pmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{pmatrix} = \hat{\rho},$$

$$\vec{e}_\theta = \frac{\partial}{\partial \theta}\begin{pmatrix} \rho\,\sin(\theta)\cos(\phi) \\ \rho\,\sin(\theta)\sin(\phi) \\ \rho\,\cos(\theta) \end{pmatrix} = \begin{pmatrix} \rho\,\cos(\theta)\cos(\phi) \\ \rho\,\cos(\theta)\sin(\phi) \\ -\rho\,\sin(\theta) \end{pmatrix} = \rho\begin{pmatrix} \cos(\theta)\cos(\phi) \\ \cos(\theta)\sin(\phi) \\ -\sin(\theta) \end{pmatrix} = \rho\,\hat{\theta},$$

$$\vec{e}_\phi = \frac{\partial}{\partial \phi}\begin{pmatrix} \rho\,\sin(\theta)\cos(\phi) \\ \rho\,\sin(\theta)\sin(\phi) \\ \rho\,\cos(\theta) \end{pmatrix} = \begin{pmatrix} -\rho\,\sin(\theta)\sin(\phi) \\ \rho\,\sin(\theta)\cos(\phi) \\ 0 \end{pmatrix} = \rho\,\sin(\theta)\begin{pmatrix} -\sin(\phi) \\ \cos(\phi) \\ 0 \end{pmatrix} = \rho\,\sin(\theta)\,\hat{\phi},$$

Here we see the coefficients of curvature are non-trivial for both θ and ϕ, as expected.

We wish to compute the angles between the tangent vectors. Before we begin we observe that the corresponding unit vectors have the same direction as the tangent vectors and are easier to work with. So we compute the inner products between the unit vectors

$$\begin{aligned}
\hat{\rho}^\top\hat{\theta} &= \begin{pmatrix} \sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) & \cos(\theta) \end{pmatrix}\begin{pmatrix} \cos(\theta)\cos(\phi) \\ \cos(\theta)\sin(\phi) \\ -\sin(\theta) \end{pmatrix} \\
&= \sin(\theta)\cos(\theta)\big(\cos^2(\phi)+\sin^2(\phi)\big)-\cos(\theta)\sin(\theta) \\
&= \sin(\theta)\cos(\theta)-\cos(\theta)\sin(\theta) \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
\hat{\rho}^\top\hat{\phi} &= \begin{pmatrix} \sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) & \cos(\theta) \end{pmatrix}\begin{pmatrix} -\sin(\phi) \\ \cos(\phi) \\ 0 \end{pmatrix} \\
&= \sin(\theta)\big(-\sin(\phi)\cos(\phi)+\sin(\phi)\cos(\phi)\big) \\
&= 0.
\end{aligned}$$

$$\begin{aligned}\hat{\theta}^{\top}\hat{\phi} &= \begin{pmatrix}\cos(\theta)\cos(\phi) & \cos(\theta)\sin(\phi) & -\sin(\theta)\end{pmatrix}\begin{pmatrix}-\sin(\phi)\\ \cos(\phi)\\ 0\end{pmatrix}\\ &= \cos(\theta)\big(-\cos(\phi)\sin(\phi)+\sin(\phi)\cos(\phi)\big)\\ &= 0.\end{aligned}$$

The tangent vectors are *orthogonal*.

Example 1.13 (Volume in of a Sphere in 3-Dimensional Polar Spherical Coordinates) *This enables us to think of the infinitesimal volume element as a rectangular box (as opposed to a parallelepiped), and most importantly to write the volume element in terms of the product of the coefficients of curvature.*

$$\mathrm{d}V = \rho^2 \sin(\theta)\mathrm{d}\rho\ \mathrm{d}\theta\ \mathrm{d}\phi\,.$$

As a simple application we compute the Volume of a Sphere of radius R

$$\textit{Volume} = \int_{\textit{Sphere}} \mathrm{d}V = \int_0^{2\pi}\int_0^{\pi}\int_0^{R} \mathrm{d}\rho\ \mathrm{d}\theta\ \mathrm{d}\phi\ \rho^2\sin(\theta) = \frac{1}{3}2\pi R^3\int_0^{\pi}\mathrm{d}\theta\ \sin(\theta) = \frac{4}{3}\pi R^3.$$

1.5.4 Other Coordinate Systems

It is impossible to provide an exhaustive list of all possible coordinate systems and their associated transformation equations. There are *literally* an infinite number of *well behaved* transformation equations and consequent coordinate systems. Below are some examples of coordinate transformation equations slightly off the beaten track. For each of these coordinate systems, it is left to the reader to perform the following steps,

1. Identify the correct ranges for caviling coordinates.
2. Identify what part of the $x-y$ plane or $x-y-z$ space is being mapped (if not the whole thing).
3. Identify coordinate curves/surfaces.
4. Find tangent vectors and cotangent vectors.
5. Investigate Orthogonality of tangent vectors and other properties.
6. Find area/volume elements.

Try using the `Mathematica` instruction

```
ParametricPlot[{formula for x, formula for y}, {x, xmin, xmax}, {y, ymin, ymax}]
```

to view the region covered by the coordinates

1. Canonical Hyperbolic Coordinates in 2-dimensions,

$$x = r\,(\cosh(t)-1) \text{ and } y = \sinh(t),$$

where $r, t \in \mathbb{R}$ are varying.

2. Modified Hyperbolic Coordinates in 2-dimensions,

$$x = a\,r\,(\cosh(t)-1) \text{ and } y = b\,\sinh(t),$$

where $a, b > 0$ are fixed.

3. Ellipsoid Coordinates in 3-dimensions,

$$x = a\rho\,\sin(\theta)\cos\big(\phi\big), \quad y = b\rho\,\sin(\theta)\sin\big(\phi\big) \quad \text{and} \quad z = c\rho\,\cos(\theta)$$

where $a, b, c > 0$ are fixed.

1.6 Coordinate Transformations

Recall that a *scalar field* assigns a value to each point in some domain. A *vector field* is a function that assigns a vector to every point in some domain. Formally,

$$V : S \to T,$$

where S and T are open subsets of vector spaces. To each $\vec{x} \in S$, we assign $\vec{y} \in T$ according to the rule $\vec{y} = V(\vec{x})$. When we associate some coordinate system with S and T, then we can think of V as associated with The function that takes the *coordinates* of $\vec{x}$ and maps them onto the *coordinates* of $\vec{y}$

$$\vec{V}\left(x^1, x^2, \ldots, x^n\right) = \begin{pmatrix} y^1\left(x^1, x^2, \ldots, x^n\right) \\ y^2\left(x^1, x^2, \ldots, x^n\right) \\ \vdots \\ y^m\left(x^1, x^2, \ldots, x^n\right) \end{pmatrix}.$$

One way of *visualizing* a vector field is by attaching a small vector to every point in space. So at the $\vec{x}$ point we would 'draw' the vector $\vec{V}(\vec{x})$. Consider the following vector fields,

$$\vec{V}(x,y) = \begin{pmatrix} y^1(x,y) \\ y^2(x,y) \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \tag{1.11}$$

Which is simply the constant vector field with assigns a single fixed vector value to each point in $\mathbb{R}^2$. A more complicated example is

$$\vec{V}(x,y) = \begin{pmatrix} y^1(x,y) \\ y^2(x,y) \end{pmatrix} = \begin{pmatrix} \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{x}{\sqrt{x^2+y^2}} \end{pmatrix}, \tag{1.12}$$

which is an example of a unit (normalised) vector field in 2-dimensions. Similarly,

$$\vec{V}(x,y,z) = \begin{pmatrix} y^1(x,y,z) \\ y^2(x,y,z) \\ y^3(x,y,z) \end{pmatrix} = \begin{pmatrix} \frac{-2y}{\sqrt{x^4+4y^2+z^2}} \\ \frac{x^2}{\sqrt{x^4+4y^2+z^2}} \\ \frac{-z}{\sqrt{x^4+4y^2+z^2}}. \end{pmatrix}, \tag{1.13}$$

is an example of a unit (normalised) vector field in 3-dimensions. Graphical representation of the vector fields in (1.11), (1.12) and (1.13) are depicted in Figure 1.18, Figure 1.19 and Figure 1.20, respectively.

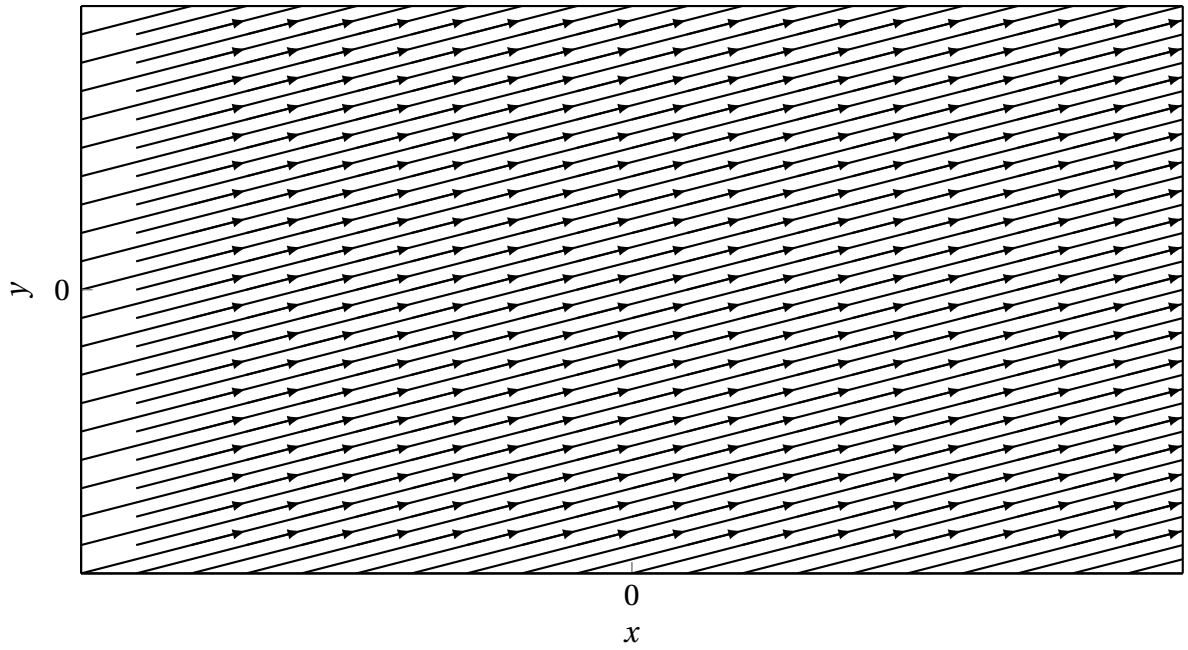


Figure 1.18: The constant 2-Dimensional normalised vector field defined in (1.11).

In all three examples above, we have implicitly made the assumption that both the original space and the target space are parameterized with *Cartesian coordinates*. This enabled us to write out equations for our vector fields taking coordinates as arguments and returning vectors in coordinate representation. Other examples of vector fields include the flow of a fluid in some container (at each point the fluid has a vector direction/velocity of flow), the electric field around some charge, and the 'wind velocity' field often shown on weather reports.

Recall when transforming scalar fields we transformed what we *put in as an argument to the function*. For vector fields we must also transforms what comes out of the function. Depending on

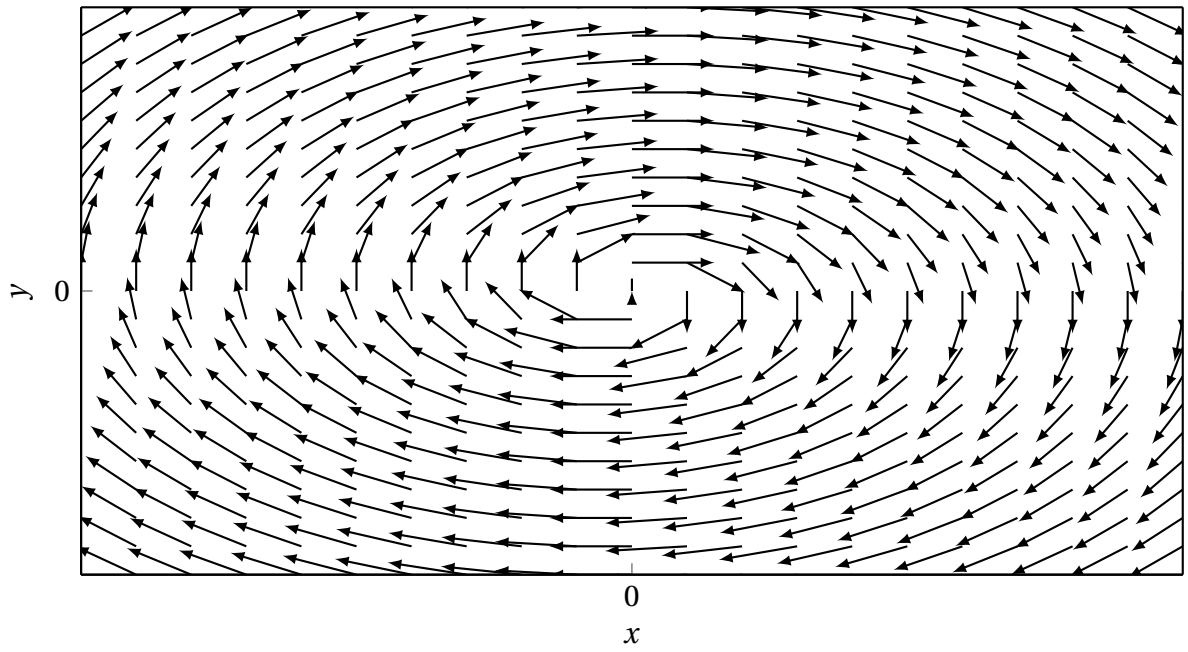


Figure 1.19: The 2-Dimensional normalised vector field defined in (1.12).

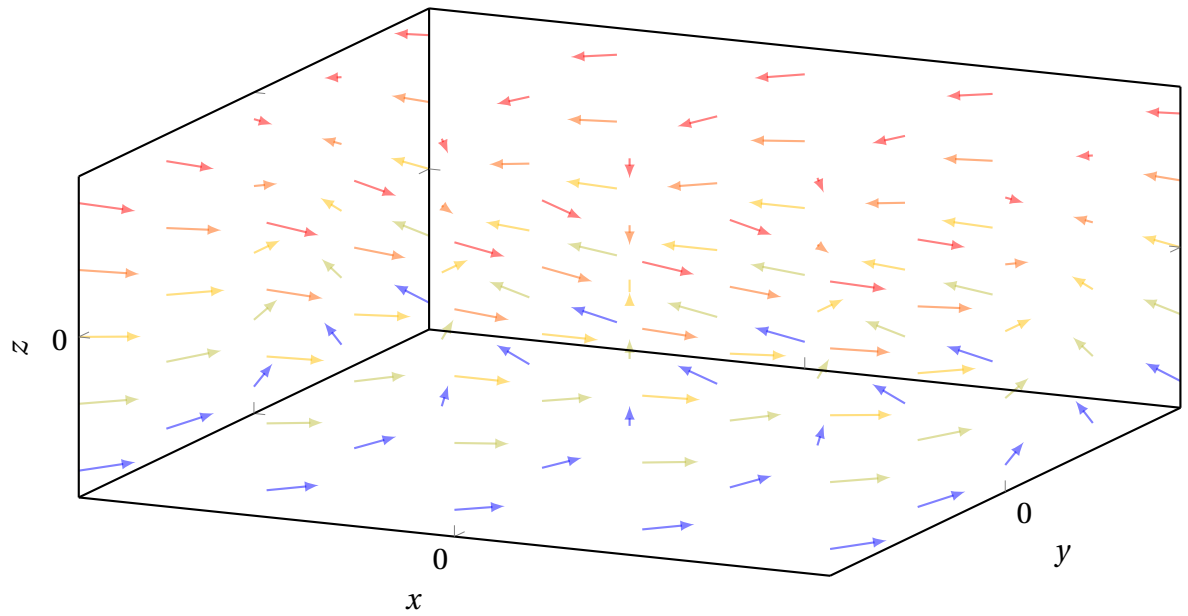


Figure 1.20: The 3-Dimensional normalised vector field defined in (1.13). The different arrow colours correspond to different z-values in 3-dimensional space.

whether we are considering covariant or contravariant coordinates, this will transform differently. Let us begin by considering the first example above,

$$\vec{f}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

We can think of this as two scalar components

$$f_x(x,y) = x \quad \text{and} \quad f_y(x,y).$$

We know that these are both covariant and contravariant components for f in Cartesian coordinates. Let us now convert this to polar coordinates. First we transform the scalar components. This part is familiar from transforming scalar functions

$$f_x(r,\theta) = f_x(x(r,\theta), y(r,\theta)) = f_x(r\,\cos(\theta), r\,\sin(\theta)) = r\,\cos(\theta)$$
$$f_y(r,\theta) = f_y(x(r,\theta), y(r,\theta)) = f_y(r\,\cos(\theta), r\,\sin(\theta)) = r\,\sin(\theta).$$

But the components are still Cartesian. To convert these to polar coordinates we have to decide whether we wish to consider the covariant components or the contravariant components. We will generally be interested in the covariant components

$$\begin{aligned} f_r &= \vec{e}_r \cdot \vec{f} \\ &= (\cos(\theta), \sin(\theta)) \cdot (r\,\cos(\theta), r\,\sin(\theta)) \\ &= r\,\cos^2(\theta) + r\,\sin([]2]\theta \\ &= r, \end{aligned}$$

and

$$\begin{aligned} f_\theta &= \vec{e}_\theta \cdot \vec{f} \\ &= (-r\,\sin(\theta), r\,\cos(\theta)) \cdot (r\,\cos(\theta), r\,\sin(\theta)) \\ &= -r^2\,\sin(\theta)\cos(\theta) + r^2\,\sin(\theta)\cos(\theta) \\ &= 0. \end{aligned}$$

As we expect from the shape of the vector field, the tangential component is zero.

Exercises

Exercise 1.1 *Show that the exchange of a millimeter ruler with an inch ruler corresponds to applying an affine transformation to the space whose original metric is given in millimeters.*

Exercise 1.2 *Embed T^2 into $\mathbb{R}^2$ and then give an explicit realization realization of a coordinate mapping from a rectangular subset of $U \subset \mathbb{R}^2$ onto T^2. Show that this mapping is many-to-one outside of U.*

Exercise 1.3 *Show by direct construction that Riemann normal coordinates provide a many to one mapping between $\mathbb{R}^2$ and S^2 and then define a subset of $\mathbb{R}^2$ where Riemann normal coordinates provide a one-to-one mapping from $\mathbb{R}^2$ onto S^2. Embed S^2 into $\mathbb{R}^2$ and then give an explicit coordinatization of S^2 for an appropriate transformation of Riemann normal coordinates.*

Exercise 1.4 *Use the standard definition of the vector dot product in N-dimensional Euclidean space in rectilinear coordinates to prove that*

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{N} a_i\, b_i$$

Exercise 1.5 *Consider a marked point in $p \in \mathbb{R}^3$. Let α, β and γ denote the angles subtended at the origin by the vector $\vec{p}$ and each of the coordinate axes $\hat{x}$, $\hat{y}$ and $\hat{z}$. Show that*

$$\cos^2(\alpha) + \cos^2(\beta) + \cos^2(\gamma) = 1.$$

Exercise 1.6 *Compute the cotangent vectors in elliptic coordinates directly and then by inverting the Jacobian matrix. Make sure you get the same results. Verify that they are mutually orthogonal with the tangent vectors.*

Exercise 1.7 *As an exercise, show that the cotangent vectors in an arbitrary linear system are the rows of the inverse transformation matrix associated with that system (hint, $V^{-1}V = I$). Clearly these do not coincide with the tangent vectors in general.*

Exercise 1.8 *Show that the matrix of transformation between two orthonormal coordinate systems is orthogonal.*