




Representations

of the

Physical Universe

Todo list

<div></div> Insert 2D Plane with Coordinates	2
<div></div> Categorical diagram of the concepts POINT \leftrightarrow ASSIGNING SPECIFIC VALUE TO (x, y) , CURVE \leftrightarrow EQUATION RELATING VALUES OF (x, y) , PLANE \leftrightarrow ALL POSSIBLE VALUES OF (x, y) , GEOMETRY \leftrightarrow ALGEBRA	2
<div></div> Insert Ball Falling	4
<div></div> Two different frames: the original and one rotated 45 clockwise, showing acceleration pointing the same way but different \ddot{x}, \ddot{y}	4
<div></div> “independence of the physical universe on coordinate representation” Duality diagram physics \leftrightarrow math between something obvious and a profound symmetry principle . .	5
<div></div> Draw three axes in 3D with corresponding basis vectors . . .	6
<div></div> Show a 2-D basis in a 3-D space, and a basis of 3 vectors in a 2-D subspace	7
<div></div> Figure, showing geometric vectors and tuples of numbers as extending the abstract algebraic vectors, and an arrow from tuples to geometric labelled “representation in a coordinate system”	9
<div></div> elementary transform with $1/\sqrt{2}s$	9
<div></div> This will be a caption for a sketch of a 3D (2D?) rotation . .	13
<div></div> Graphic of orthogonality being changed under affine transform	14
<div></div> I wanna really elaborate how an inner product <i>picks</i> a frame to be orthogonal, promoting it above all the others.	14
<div></div> Insert graphic here, with an affine transformation of the plane	15

	Caption something like “If Euclid’s space allowed for not just rotations/translations but also general linear transformations like this, being perpendicular would be a coordinate-dependent idea. Note parallel lines, on the other hand, stay parallel, so affine space <i>does</i> have a notion of parallel”	15
	Graphic of polar coordinate system/spherical	16
	EXERCISE: Check invariance quickly using this method. . .	20

Copyright ©2017 by Alexander B. Atanasov No part of this book may be reproduced in any form or by any means without permission in writing from the authors. All images have been created by hand or in Wolfram Mathematica. This copyright does not extend to the graphs and figures within this work. Anyone who can reproduce these figures is free to do so and use them.

Other copyright information will be put here by the publisher.

Preface

Dedicated to \otimes ,
...

Contents

<i>Preface</i>	i
<i>Acknowledgements</i>	iv
0 Prerequisite Material	1
1 Old Grounds, Revisited	2
1.1 The Cartesian Coordinate System	2
1.2 Linear Algebra & Coordinates	6
1.3 The Notion of Length on Vector Spaces	13
1.4 Nonlinear Coordinate Systems are Locally Linear . . .	15
1.5 Einstein's Summation Convention	18
1.6 Exercises	20

Acknowledgements

The authors would like to thank:

∇ John C. Dell

\mathbb{F} Jonathan A. Osborne

$\frac{j(j+1)\hbar^2}{2m_e R^2}$ James C. Ellenbogen

$\langle \sigma \sigma \epsilon \epsilon \rangle$ David Poland

$\mathcal{M}(r, k)$ Igor B. Frenkel

$\mathcal{U}_q(\mathfrak{sl}_2)$ You Qi

$\mathcal{L}_X Y$ Vincent E. Moncrief

$\{-, -\}$ Francesco Iachello

$\int \frac{d^4 \ell}{(2\pi)^4}$ Walter D. Goldberger

$\omega_i dq^i$ Ted A. Jacobson

For the valuable conversations, lectures, and research opportunities that they offered, which led up to the development of this text.

Part 0

Prerequisite Material

Chapter 1

Old Grounds, Revisited

1.1 The Cartesian Coordinate System

One of the most revolutionary ideas in mathematics and physics draws its name from the seventeenth century philosopher René Descartes. The idea is simple: any point P in the 2D plane can be represented by a pair of real numbers (x, y) in a one-to-one correspondence. The paired numbers represent distances along two perpendicular axes that meet at a special point, O , called the **origin** and correspond to $(0, 0)$. This construction gives a **Cartesian coordinate system** for the Euclidean plane, which does something amazing: it relates the geometry of the plane to the algebra of variables and equations. Algebra could be represented geometrically, and conversely geometric problems could be solved by transforming them into algebraic ones. The two great classical branches of mathematics, the study of space and the study of quantity, have been joined.

Insert 2D Plane with Coordinates

If the above construction seems obvious to the reader, it should be the same type of obviousness that would strike one when staring at a wheel, perplexed by how such a simple concept went undiscovered for so long in so many cultures. It should speak to the nontrivial character of the Cartesian coordinate system that Descartes himself did not discover the full picture as we know it today. He introduced a variant where distances were measured from a *single* reference axis. It took many years before the idea to add a second perpendicular axis took root, and for the modern treatment to be established.

Categorical diagram of the concepts **POINT** \leftrightarrow **ASSIGNING SPECIFIC VALUE TO (x, y)** , **CURVE** \leftrightarrow **EQUATION RELATING VALUES OF (x, y)** , **PLANE** \leftrightarrow **ALL POSSIBLE VALUES OF (x, y)** , **GEOMETRY** \leftrightarrow **ALGEBRA**

Coordinates soon extended beyond just being pairs of numbers (x, y) . Their use applied just as well to describing points in 3D space in reference to a fixed origin by a triple of coordinates (x, y, z) . Soon enough after, they were generalized to arbitrarily high dimensions: any point in n dimensional space could be represented by an n -tuple (x_1, x_2, \dots, x_n) . Despite some controversy at the time of their invention, these tuples of numbers would subsequently be used to lay the foundations for modern physics and mathematics. Modern linear algebra, multivariable calculus, and the myriad of connections between algebra and geometry begin with the concept of a coordinate.

Since then, the use of coordinate systems has proven indispensable to scientists and mathematicians throughout history. Newton used Descartes' ideas to formulate his infinitesimal calculus. Maxwell used it to analyze electromagnetic fields, discovering mathematically that light is a wave in the electromagnetic field. Einstein, going further, made use of coordinate systems on a four dimensional space of three dimensions for physical space and one for time to formulate his theory of relativity and gravitation. Today, physicists and engineers base their calculations in the frames of coordinate systems. In mathematics, Descartes' idea planted the roots for what would turn into the modern field of algebraic geometry.

When studying a geometry in some n -dimensional space, we pick an origin and set of axes to form a coordinate system. Once a Cartesian coordinate system has been set, can represent any point as a tuple of n real numbers. Because of this, n -dimensional space is frequently just denoted by \mathbb{R}^n , the set of n -tuples of real numbers. It is the aim of this chapter to illustrate, in as much clarity as possible, why these two concepts are different.

We'll begin this illustration with an example. Let's visualize the simple picture of a ball falling. we could set the origin at some point on the ground, and pick one axis parallel to the ground, "distance", and one perpendicular, "height". We can decide to measure the axes in meters, or we could decide to measure in feet (nothing stops us

from making bad choices of coordinates). The physical point P where the ball lies is represented by $(x, y) = (0 \text{ m}, 10 \text{ m})$.

Insert Ball Falling

We can now study motion of the ball, free of geometry, as just two functions $x(t)$ and $y(t)$ on which we can do arithmetic and calculus. If we are given **equations of motion** corresponding to gravitational acceleration downwards:

$$\frac{d^2y}{dt^2} = -g \text{ m/s}^2, \quad \frac{d^2x}{dt^2} = 0 \text{ m/s}^2$$

with initial conditions,

$$x(0) = 0 \text{ m}, \quad y(0) = 10 \text{ m} \quad \frac{dx}{dt}(0) = 10 \text{ m/s} \quad \frac{dy}{dt}(0) = 0 \text{ m/s}$$

then we can perform our well-known kinetic calculations for the system, and see how the system evolves in the time direction¹. In this case we get explicitly:

$$x(t) = 10t \text{ m}, \quad y(t) = -\frac{g}{2}t^2 \text{ m}$$

This solution to the equations of motion describes exactly how the coordinates representing the position of the ball change in time. Thus, this can be used to predict the motion of the ball itself. By use of coordinates, this problem of geometric evolution is recast as a calculus problem on real-valued functions of time, and this can be easily solved.

But the coordinate system that we chose—apart from being easy to do calculations in—was arbitrary. We could have just as well worked in a 45° rotated frame, and done the same calculations, with accelerations now being present along the directions of both axes. If we solved this, we would have gotten different-looking equations for the different coordinates $x'(t), y'(t)$, but they would correspond to the exact same *dynamics of motion*.

Two different frames: the original and one rotated 45 clockwise, showing acceleration pointing the same way but different \ddot{x}, \ddot{y}

¹One might ask “what is the difference between dynamics and geometry?”, and indeed, when time is introduced as an additional dimension later in the book, we can view dynamics of an n dimensional system as simply a special case of geometry in $n + 1$ dimensions: n space and 1 time.

The ball will fall down towards the ground according to the force of gravity. That is the way the world works. It doesn't matter what coordinate system we set up to do that calculation, we should get the exact same result. Plainly, Nature doesn't *care* what coordinate system we use: the events that happen in the natural world are independent of the languages that humans use to describe them. This fact, obvious as it may be, plays tremendous importance in modern physics and will be central to the rest of this book.

As obvious as this fact may be from the physical viewpoint, it gives rise to an extremely strong constraint on the mathematical language that we use to describe the world. **FINISH HERE**

"independence of the physical universe on coordinate representation" Duality diagram physics \leftrightarrow math between something obvious and a profound symmetry principle

It is funny that coordinate systems, which appear in almost every serious calculation in physics and could be considered the most powerful tool yet developed, must necessarily be so arbitrary. There is no canonical "God-given" frame of reference to choose, but to do any calculations, we must nonetheless choose *one*. If this strikes the reader as alien and strange, an esoteric quality reserved to only mathematics, they should turn their mind to the nature of language itself. It is no less arbitrary in its many forms, and no less crucial for developing a deep understanding of the workings of the world. In an even more similar way, the different scales and tunings that form the basis for various forms music around the world are just as arbitrary. No one scale can be demonstrated to be *the* god given one, there are only those that are easier to work with than others.

However, unlike in introductory classes on physics and engineering, where the goal is to become proficient at doing calculations in various coordinate systems to make predictions, this text will take the next steps and look at the "meta-theory" behind this. We will ask how the laws of physics can be formulated so that they are independent of the coordinates we use. Doing this will give us significantly greater insight into the nature of physical laws and their underlying coordinate freedom. Such an understanding will in turn lead to a significantly deeper appreciation of the role that symmetry plays in physics, and ultimately will lead us to understanding more about the physical universe than we ever could before.

1.2 Linear Algebra & Coordinates

This section assumes some base knowledge of linear algebra. As such, it should be taken as a review and expansion, rather than an introduction.

To begin this refined understanding of the nature of coordinates, let us go to the familiar setting of 3D Euclidean space (read: the physical world as we know it). A Cartesian coordinate system on this space requires a choice of origin, and a choice of axes. The first choice is relatively easy and uninteresting, and comes down to simply marking one point in the space as distinct. The choice of axes, however, is far more interesting, and gives rise to **linear algebra** and **bases** on this space.

Given an origin, a choice of axis to measure lengths along consists of picking a *direction* for the axis and a *magnitude* for the length that it measures. You should recognize that these two properties are exactly what define a vector in Euclidean space. To each axis is associated a special vector, \mathbf{e} , corresponding to moving a unit direction along this axis.

Draw three axes in 3D with corresponding basis vectors

Note that we place no requirement that the axes be perpendicular, only that they specify a valid coordinate system. Since we see that specifying an axis is in fact the same as specifying a vector corresponding to moving +1 unit along that axis, from now on we will work with vectors. What does it mean for three vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ to specify a valid coordinate system in 3D?

A valid coordinate system for 3D space one that associates to each point in 3D space a unique set of three real numbers that characterize it. This is the same in any dimension.

In the language of linear algebra, we need our coordinate system's associated vectors to both **span** the space so that we can represent any point, and be **linearly independent** so that every point will have unique representation in our coordinate system. That's all that a basis is: it specifies a set of axes for a valid coordinate system.

Definition 1.1. A set of vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is said to span a space if every point P in the space can be represented as a linear combination $a_1\mathbf{e}_1 + \dots + a_n\mathbf{e}_n$

Definition 1.2. A set of vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is said to be linearly independent if there is only one way to represent the zero vector $\mathbf{0}$ as a combination of them, namely as $\mathbf{0} = 0\mathbf{e}_1 + \dots + 0\mathbf{e}_n$.

This second definition is the same as saying every vector that we can represent in our system has a unique representation. Let's make this clear. If there were two ways to write a vector \mathbf{v} , as:

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

and

$$b_1\mathbf{e}_1 + \dots + b_n\mathbf{e}_n = \mathbf{v}$$

then subtracting these two different combinations would give a nonzero way to represent zero, contradicting linear independence:

$$(a_1 - b_1)\mathbf{e}_1 + \dots + (a_n - b_n)\mathbf{e}_n = \mathbf{0}.$$

Conversely, if there were a nonzero combination of vectors summing to zero, then we could add that combination to the coordinate representation of any point and get a *different* representation of the same point, contradicting uniqueness. So coordinate representations for all vectors are unique as long as there is only one representation for $\mathbf{0}$, namely the one where each component equals zero.

Intuitively, having a set of vector span a space means that this set has enough information to describe every point in the space. Linear independence, on the other hand, means that there is no superfluous information in the set of vectors. We cannot linearly combine vectors in some subset to get another vector in the set; each vector is adding its own unique additional piece of information, making the set able to span in a new direction.

Bases that don't span, or are not linearly independent would lead to coordinate systems like these:

Show a 2-D basis in a 3-D space, and a basis of 3 vectors in a 2-D subspace

Once an origin is chosen, each point in 3D space becomes a vector relative to that origin. Therefore, the idea that points in space are independent of the coordinate system used to describe them applies to vectors just as well. If we choose a basis for our vector space $\mathbf{e}_1, \dots, \mathbf{e}_n$, then we can express any vector \mathbf{v} by a unique combination

$\mathbf{v} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$. We then say that in this basis, we can represent \mathbf{v} by a list of numbers. Often, it is written:

$$\mathbf{v} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

but writing this as an equality is wrong.

The vector \mathbf{v} is something physical: a point relative to an fixed origin. As an example, if we chose a stationary frame as an origin, \mathbf{v} will now represent a physical velocity: the flow of water, the rush of air, the motion of the sun's core relative to Venus. It is purely geometric and has *absolutely nothing to do with numbers* until a coordinate system is chosen. On the other hand, the right hand side by itself is just a list of numbers entirely dependent on the coordinate system. If we change coordinate systems, the right hand side changes. \mathbf{v} does not.

A geometric vector, \mathbf{v} is *not* a list of numbers. Once we pick a basis, u can be *represented by* a list of numbers, but if we change into a different basis, those numbers all have to change as well. Later, this exact same idea will be what distinguishes a physical tensor from a multi-dimensional array (like the ones encountered in computer science). It can be *represented by* a multi-dimensional array once a coordinate system is chosen, but the numbers in each entry will differ depending on the coordinate system we pick.

“Hold on” you may say “the definition of vector isn’t something that has magnitude and direction. Vectors simply need to be able to be added and scaled. You can definitely add and scale lists of n numbers, so aren’t they vectors in that sense?”

You’re completely right, and this is the main source of confusion in understanding this topic. In most courses, we can freely call any list of numbers a ‘vector’. After all, you can add lists and scale them so they do form a ‘vector space’, as you point out. This boils down to a really unfortunate linguistic degeneracy in mathematics terminology. The type of vectors that we see in physics (acceleration, force, electric field, etc.) are **geometric vectors** that have nothing *a priori* to do with lists of numbers until we represent them as such by using coordinate systems. On the other hand, abstract structures that we can add and multiply by scalars are **algebraic vectors**, and lists of numbers

are an example of that. Geometric vectors are also an example of algebraic vectors, since they, too, can be added and scaled.

Figure, showing geometric vectors and tuples of numbers as extending the abstract algebraic vectors, and an arrow from tuples to geometric labelled “representation in a coordinate system”

To avoid confusing lists of numbers with the geometric vectors in the physical world, we will refer to the former as **tuples** rather than vectors.

So, returning to the geometric vector \mathbf{v} , a correct way to write it would be (using standard matrix multiplication convention):

$$\mathbf{v} = (\mathbf{e}_1 \dots \mathbf{e}_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a_1 \mathbf{e}_1 + \dots + a_n \mathbf{e}_n. \quad (1.1)$$

Once we pick a basis, that column of coordinates means something. If we denote our basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ by B , then we will use the notation of square brackets to mean not just a column of numbers, but a full combination of the vectors in B :

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_B := (\mathbf{e}_1 \dots \mathbf{e}_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a_1 \mathbf{e}_1 + \dots + a_n \mathbf{e}_n. \quad (1.2)$$

Now we can begin working towards the goal of understanding coordinate independence of physics by first understanding how coordinates transform as we go from one basis to another. Changes of basis in linear algebra are infamously frustrating to students, because one never knows when to apply a transformation or its inverse. In the following discussion, we will use notation as a tool to make this as painless as possible.

Let us do a very simple example to start. In the 2-D plane, say we have our original basis $\mathbf{e}_1, \mathbf{e}_2$ and we rotate by $\pi/4$ radians to get a new basis. Say we have the vector $\mathbf{v} = \mathbf{e}_1 + \mathbf{e}_2$, i.e. with coordinate representation $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in the original basis.

Now our new basis is the old one rotated by $\pi/4$ so

elementary transform with $1/\sqrt{2}$ s

$$\begin{aligned}\mathbf{e}'_1 &= \frac{\sqrt{2}}{2}\mathbf{e}_1 + \frac{\sqrt{2}}{2}\mathbf{e}_2 \\ \mathbf{e}'_2 &= -\frac{\sqrt{2}}{2}\mathbf{e}_1 + \frac{\sqrt{2}}{2}\mathbf{e}_2.\end{aligned}$$

Using matrices, this can be written compactly as²:

$$\begin{pmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

This transformation relates the actual basis vectors themselves. On the other hand, if we wanted to know the *coordinate representations* of the vectors \mathbf{e}'_1 and \mathbf{e}'_2 , then in the new basis they would obviously be represented in coordinates as:

$$\mathbf{e}'_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\text{new}} \quad \mathbf{e}'_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\text{new}}$$

while in the old basis their representations are:

$$\mathbf{e}'_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}_{\text{old}} \quad \mathbf{e}'_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}_{\text{old}}.$$

If we know that we can describe a point as $\begin{bmatrix} x' \\ y' \end{bmatrix}_{\text{new}}$ in the new basis, then we can easily get its description in the old basis as:

$$\begin{aligned}\begin{bmatrix} x' \\ y' \end{bmatrix}_{\text{new}} &= x'\mathbf{e}'_1 + y'\mathbf{e}'_2 \\ &= x' \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}_{\text{old}} + y' \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}_{\text{old}} \\ &= \left[\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}_{\text{old} \leftarrow \text{new}} \begin{pmatrix} x' \\ y' \end{pmatrix} \right]_{\text{old}}\end{aligned}$$

²The tuple of basis vectors \mathbf{e}_i is written as a row rather than a column to be consistent with Equation (1.1). Then the coordinates are represented in a column. Because of the way we do matrix multiplication, then the matrix must act on the right. It's an issue of styling and indexing, and not physically meaningful.

This is the same matrix that related the basis vectors. It's not too hard to see why, and in fact it's just associativity:

$$\begin{aligned} \begin{bmatrix} x' \\ y' \end{bmatrix}_{\text{new}} &= (\mathbf{e}'_1 \quad \mathbf{e}'_2) \begin{pmatrix} x' \\ y' \end{pmatrix} \\ &= (\mathbf{e}_1 \quad \mathbf{e}_2) \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \\ &= \left[\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}_{\text{old} \leftarrow \text{new}} \begin{pmatrix} x' \\ y' \end{pmatrix} \right]_{\text{old}}. \end{aligned}$$

This also means that given a coordinate tuple $\begin{pmatrix} x \\ y \end{pmatrix}$ in the old basis, we can get the coordinates in the new basis by going the other way and inverting:

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix}_{\text{old}} &= \left[\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}_{\text{old} \leftarrow \text{new}} \begin{pmatrix} x' \\ y' \end{pmatrix} \right]_{\text{old}} \\ \Rightarrow \begin{pmatrix} x' \\ y' \end{pmatrix} &= \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}_{\text{new} \leftarrow \text{old}}^{-1} \begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

For our vector \mathbf{v} , represented as $(1, 1)$ in our original basis, in the new basis, we would have:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\text{old}} = \left[\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}_{\text{new} \leftarrow \text{old}}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]_{\text{new}} = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}_{\text{new}}.$$

Indeed, $\mathbf{e}_1 + \mathbf{e}_2 = \sqrt{2}\mathbf{e}'_1 + 0\mathbf{e}'_2$.

Now, in the general n -dimensional case, we start with a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Now let's make a general transformation into a new basis so that:

$$\begin{aligned} \mathbf{e}'_1 &= A_{11}\mathbf{e}_1 + A_{21}\mathbf{e}_2 + \dots + A_{n1}\mathbf{e}_n \\ \mathbf{e}'_2 &= A_{12}\mathbf{e}_1 + A_{22}\mathbf{e}_2 + \dots + A_{n2}\mathbf{e}_n \\ &\vdots \\ \mathbf{e}'_n &= A_{1n}\mathbf{e}_1 + A_{2n}\mathbf{e}_2 + \dots + A_{nn}\mathbf{e}_n, \end{aligned}$$

or in matrix form:

$$(\mathbf{e}'_1 \quad \dots \quad \mathbf{e}'_n) = (\mathbf{e}_1 \quad \dots \quad \mathbf{e}_n) \mathbf{A}_{(\text{old} \rightarrow \text{new})}, \quad (1.3)$$

$$\mathbf{A} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}.$$

As before, we see that \mathbf{A} is the transformation taking the old basis \mathbf{e}_i to the new one \mathbf{e}'_i . \mathbf{A} also takes the new coordinate representations and tells us how they'd look like in the old basis (note the reversal).

$$\begin{aligned} \begin{bmatrix} a'_1 \\ \vdots \\ a'_n \end{bmatrix}_{\text{new}} &= (\mathbf{e}'_1 \quad \cdots \quad \mathbf{e}'_n) \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} \\ &= (\mathbf{e}_1 \quad \cdots \quad \mathbf{e}_n) \mathbf{A} \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} \\ &= \left[\mathbf{A}_{(\text{old} \leftarrow \text{new})} \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} \right]_{\text{old}} \\ &= \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_{\text{old}}. \end{aligned}$$

Note how when \mathbf{A} acts from the right on the \mathbf{e}_i , it takes the old basis to the new basis. On the other hand when it acts on the coordinate tuple, from the left, it took it from new coordinate old coordinate. For this reason we can label \mathbf{A} as $\mathbf{A}_{(\text{old} \leftrightarrow \text{new})}$ and the side on which it acts (basis vectors vs. coordinate tuples) automatically tells us whether it transforms those quantities from old to new or vice-versa.

So then it is the *inverse* \mathbf{A}^{-1} that tells us how our old coordinates for a vector \mathbf{v} will look like in our new basis.

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \mathbf{A}_{(\text{old} \leftarrow \text{new})} \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} \Rightarrow \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} = \mathbf{A}_{(\text{new} \leftarrow \text{old})}^{-1} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad (1.4)$$

From looking at equations (1.3) and (1.4), we see the central idea. For any vector \mathbf{v} , if we *vary* the basis \mathbf{e}_i to a different basis, \mathbf{e}'_i , then the coordinates a_i will vary the *other* way to a'_i , so that the geometric vector

$$\mathbf{v} = a_1 \mathbf{e}_1 + \cdots + a_n \mathbf{e}_n = a'_1 \mathbf{e}'_1 + \cdots + a'_n \mathbf{e}'_n$$

is *invariant* regardless of coordinate choice:

$$(\mathbf{e}'_1, \dots, \mathbf{e}'_n) \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} = (\mathbf{e}_1, \dots, \mathbf{e}_n) \mathbf{A} \mathbf{A}^{-1} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = (\mathbf{e}_1, \dots, \mathbf{e}_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \mathbf{v}$$

We say that the basis vectors \mathbf{e}_i **co-vary** and the coordinates a_i **contra-vary** with the change of basis. The idea, although it sounds simple, is rather hard to get the feel of. It's worth thinking a good bit about how coordinates and bases need to vary in opposite ways so that the physical object represented by the coordinates stays the same regardless of how we look at it.

This will be a caption for a sketch of a 3D (2D?) rotation

When you rotate your character in a video game (and in real life too, by the way), the world rotates *contrary* to the direction that you've rotated in. That's because the coordinates of what you see have *contra-varied* while your basis vectors, given by the direction you face have *co-varied*. The end result is that despite changing your coordinate system, physics stays the same: invariant. The universe did not rotate itself just because you did. This extends beyond just rotations to *all* linear transformations. As an example: if you applied a scaling transformation ($\times 100$) to change your units of length: $\text{cm} \rightarrow \text{m}$, the coordinates corresponding to those lengths would change in the opposite way, $100 \rightarrow 1$, ($\div 100$) to represent the same distance.

1.3 The Notion of Length on Vector Spaces

Let us consider the property of orthogonality. It's well known that for geometric vectors, there's more that we can do than just add, scale, and transform them: we can take dot products³ between them. When we have two geometric vectors in space, their dot product is a well defined number. If it is zero, then the vectors are orthogonal to one another. From the dot product and the magnitudes, it is possible to calculate the angle between two given vectors.

When vectors are represented in terms of tuples of numbers, the dot product was taught to us as "multiply component by component, and then sum that up". This is not, in general, what the dot product really is.

³And in 3-D we can also take a cross product. This operation, and generalizations, will be discussed in the following chapters.

Consider a basis transformation as below:

$$\begin{aligned}\mathbf{e}'_1 &= 2\mathbf{e}_1 + \mathbf{e}_2 \\ \mathbf{e}'_2 &= \mathbf{e}_1 + 2\mathbf{e}_2.\end{aligned}$$

It's easy to compute the inverse of this matrix and see how the new coordinates should work, but it's worthwhile looking at this geometrically. It is not a rotation, but more of a “stretching”. Notice that while in our original perspective, if we viewed \mathbf{e}_1 and \mathbf{e}_2 as orthogonal vectors, in the *new* perspective, they are *no longer* orthogonal. This is very important: linear transformations in general do not preserve orthogonality.

Graphic of orthogonality being changed under affine transform

This linear transformation has stretched our vector space and changed the notion of distance. While rotations keep distances preserved, more general linear transformations don't care about a notion of distance.

If we were to take the “dot product” $\mathbf{e}'_1 \cdot \mathbf{e}'_2$ just by multiplying corresponding coordinates and summing them up, then in the new basis we'd get zero, but in the original basis we would *not*. This dot product actually changes depending on the coordinate system that we use! In some sense, this is expected: all we're doing is multiplying contravariant coordinates together and summing them up. The result should be contravariant as well (in fact doubly contravariant).

The failure of the dot product to be invariant is intimately related to the fact that transformations can change lengths. This should not be too surprising. After all, the length of a vector is defined by the square root of its dot product with itself. If the dot product we learned is not invariant under general coordinate transformations, what is the right way to measure length?

It is here that there is a big subtlety. A vector space on its own does not have a notion of length. We've just seen choosing different bases would give rise to different length scales and notions of “perpendicular” as well. Endowing a vector space with a way to universally tell what the length of a vector is, or whether two vectors are perpendicular is actually adding *extra structure* to the space. It picks a whole class of specific coordinate systems and says “these are the orthogonal reference frames; the others are skewed and stretched perspectives”. This allows us to measure length using an invariant *inner product*.

I wanna really elaborate how an inner product *picks* a frame to be orthogonal, promoting it above all the others.

Euclidean space, as well as the world in which we live in, both have an natural way to measure length between two points that is invariant of

the coordinate system used. A vector space on its own does not, and so it is called an *affine space*. In affine space, although there are notions like “parallel”, there is not a notion of distance. Adding an inner product to affine space gives rise to Euclidean space.

Insert graphic here, with an affine transformation of the plane

Caption something like “If Euclid’s space allowed for not just rotations/translations but also general linear transformations like this, being perpendicular would be a coordinate-dependent idea. Note parallel lines, on the other hand, stay parallel, so affine space *does* have a notion of parallel”

This will be discussed in much greater detail in the following two chapters.

The takeaway from this discussion is that general linear transformations don’t preserve distances. Because of this, the usual definition of the dot product gives different results depending on the choice of coordinate system. To make length an *invariant* regardless of coordinate choice, the definition of the inner product between two vectors needs to be appropriately modified. Without an *inner product* to measure distance, \mathbb{R}^n is *not* Euclidean space \mathbf{E}^n , and is called affine space. Affine space looks the same regardless of the linear transformations we apply, while Euclidean space only looks the same when we rotate or translate our frame (because those are the two transformations that don’t change lengths).

The reason we haven’t encountered this problem so far in physics is simply because we always have worked in orthonormal frames given by an orthonormal basis like $\hat{\mathbf{i}}, \hat{\mathbf{j}}$, where the inner product in fact *is* just the sum of the products of corresponding components. \mathbb{R}^n without an inner product is affine, not Euclidean. Often however, when talking about Euclidean space, authors refer to it as \mathbb{R}^n instead of \mathbf{E}^n simply because \mathbf{E}^n is modeled as \mathbb{R}^n endowed with an inner product.

1.4 Nonlinear Coordinate Systems are Locally Linear

Perhaps you may be wondering why we’ve spent so much time on changing between coordinate systems represented by basis vectors centered at a fixed origin. Consider the change between Cartesian and polar coordinates. What does this have to do with the linear changes of coordinates that we’ve been discussing?

We could use something like a polar system of (r, θ) or a spherical system (r, ϕ, θ) . These coordinate systems are not representable in terms of axes, but instead look like this:

Graphic of polar coordinate system/spherical

This is an example of a non-linear coordinate transformation. They are more commonly referred to as **curvilinear**. Whereas linear ones map lines to lines, curvilinear ones more generally map lines to curves. The idea for making sure that the equations of physics still stay true for non-linear coordinate transformations is to note that just like a curve locally looks like a line, a *non-linear* transformation locally looks like a *linear* one. The linear transformation that it locally looks like is called the **Jacobian** J . If the laws of physics are invariant under linear transformations locally at each point, then *globally*, they will be invariant under non-linear ones as well. That is why we cared about studying covariance and contravariance for linear transformations: more complicated cases can be reduced to their local linear behavior.

As an example, consider going from a Cartesian to a polar coordinate system. We have $x = r \cos \theta$ and $y = r \sin \theta$. Certainly, this is not a linear transformation of coordinates. There is sinusoidal dependence on θ in this transformation. Physics and geometry, however, do not have laws in terms of absolute coordinates (it doesn't make sense to say "That object is located at 50 meters") but only in terms of relative distances (you'd instead say "That object is located 50 meters *relative to me*"). It is the changes over relative distances between points that we care about, and these are obtained by integrating the *infinitesimal* changes at each point.

So although x, y do not depend linearly on theta, through the use of the chain rule, we have a local linear relationship in their infinitesimal changes:

$$\begin{aligned} dx &= \cos \theta \, dr - r \sin \theta \, d\theta \\ dy &= \sin \theta \, dr + r \cos \theta \, d\theta \end{aligned}$$

At any given point, this relationship can be written as a linear change of basis.

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix}$$

So every nonlinear transformation from some coordinate system $x_1 \dots x_n$

to $x' \dots x'_n$ has the local linear transformation law:

$$\begin{aligned} \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}_{\text{old}} &= \begin{pmatrix} \frac{\partial x_1}{\partial x'_1} & \cdots & \frac{\partial x_1}{\partial x'_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial x'_1} & \cdots & \frac{\partial x_n}{\partial x'_n} \end{pmatrix}_{\text{old} \leftarrow \text{new}} \begin{pmatrix} dx'_1 \\ \vdots \\ dx'_n \end{pmatrix}_{\text{new}} \\ \Rightarrow \begin{pmatrix} dx'_1 \\ \vdots \\ dx'_n \end{pmatrix}_{\text{new}} &= \begin{pmatrix} \frac{\partial x_1}{\partial x'_1} & \cdots & \frac{\partial x_1}{\partial x'_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial x'_1} & \cdots & \frac{\partial x_n}{\partial x'_n} \end{pmatrix}_{\text{new} \leftarrow \text{old}}^{-1} \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}_{\text{old}} \end{aligned}$$

This is exactly the analogue of Equation (??), so indeed the changes in coordinates dx_i can be called *contravariant*, just like the coordinates were for the linear transformation case. All of this is just an extension of the principle of local linearity from calculus.

Similarly, we can express the new derivative operators in terms of the old ones by using the chain rule. For polar coordinates we have

$$\begin{aligned} \frac{\partial f}{\partial r} &= \frac{\partial x}{\partial r} \frac{\partial f}{\partial x} + \frac{\partial y}{\partial r} \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial \theta} &= \frac{\partial x}{\partial \theta} \frac{\partial f}{\partial x} + \frac{\partial y}{\partial \theta} \frac{\partial f}{\partial y} \end{aligned}$$

or more compactly we can relate just the differential operators themselves:

$$\begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix}$$

so that more generally:

$$\begin{pmatrix} \frac{\partial}{\partial x'_1} \\ \vdots \\ \frac{\partial}{\partial x'_n} \end{pmatrix}_{\text{new}} = \begin{pmatrix} \frac{\partial x_1}{\partial x'_1} & \cdots & \frac{\partial x_1}{\partial x'_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial x'_1} & \cdots & \frac{\partial x_n}{\partial x'_n} \end{pmatrix}_{\text{new} \leftarrow \text{old}}^T \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix}_{\text{old}} \quad (1.5)$$

If you look carefully, you will see that this accordingly mirrors the covariant change of vectors in Equation (??). So while the infinitesimal changes in the coordinates themselves are contravariant, just like the linear coordinates themselves, the *differential operators* corresponding to changes in these coordinates become *covariant*, just like the vectors \mathbf{e}_i in the linear case. This is the first correspondence that will hint that our basis vectors \mathbf{e}_i actually *correspond* to the differential operators $\frac{\partial}{\partial x_i}$.

Indeed, as we begin to move away from 3-D and n -D Euclidean space, we will see why the old notions of unit vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ are better viewed as the operators $\frac{\partial}{\partial x}, \frac{\partial}{\partial y},$ and $\frac{\partial}{\partial z},$ and in general $\mathbf{e}_i \rightarrow \frac{\partial}{\partial x_i}$. This change of notation will allow us to easily pass onto far more general spaces than the Euclidean ones we've gotten used to.

1.5 Einstein's Summation Convention

Row tuples, column tuples, matrices representing basis transformations (old to new and new to old), co-variance and contra-variance. These ideas have constituted the entirety of this first chapter, and although hopefully they have not been too difficult conceptually, the matrix manipulation even at this early level is already a pain. We have to write everything in terms of tuples, and we need to arbitrarily decide which ones are rows and which ones are columns, and which matrices are transposed so that all the matrix multiplications make sense, as defined in linear algebra. A young physicist named Albert Einstein used a convention of writing all these equations so that we did not have to explicitly write out tuples of abstract basis vectors and coordinates.

The first step is to avoid explicitly writing out row tuples and column tuples. To do this, instead of writing out a whole tuple to represent a vector like (e_1, \dots, e_n) we will simply write e_i . e_i should be viewed as the whole vector, rather than a specific i th component of \mathbf{v} . The index i is *free* and can be anything.

The reason it is preferable to view e_i as the whole vector rather than a specific i th component of it is because of the same reason given at the end of Section 1.1, that we never care about just a specific component, but rather the vector as a whole.

The second step is to be able to differentiate between *covariant* quantities and *contravariant* quantities. The convention is this: if the quantity is covariant, like a basis vector, then write its index *downstairs*: \mathbf{e}_i . On the other hand, if a quantity is contravariant (like a component of a vector) then write its index *upstairs* as v^i instead⁴. Then we can write Equation (1.1) as

$$\mathbf{v} = \sum_i a^i \mathbf{e}_i \quad (1.6)$$

⁴If you are worried that this will be confused with exponentiation, don't be. In practice, such confusion almost never arises.

For another example, the gradient operator was written $\nabla = (\partial/\partial x_1, \dots, \partial/\partial x_n)$ and is now written simply as $\partial/\partial x^i$ (adopting upper indices for x^i since co-ordinates contra-vary). This means that Equation (1.5) can be written as

$$\frac{\partial}{\partial x'^i} = \sum_j \frac{\partial x^j}{\partial x'^i} \frac{\partial}{\partial x^j}. \quad (1.7)$$

Since i is free to be anything (while j is bound since it is being summed over), this equation holds for every i in $1, \dots, n$ and so is indeed an equation relating two *vectors* on both sides. Note also that this means the transformation matrix can be written as $\frac{\partial x^j}{\partial x'^i}$. By writing just a summation and not any explicit matrices, we avoid having to worry about unnecessary troubles like transposes, etc. Note also that this matrix has both upper and lower indices so that the upper index in $\frac{\partial x^j}{\partial x'^i}$ gets multiplied with the lower index $\frac{\partial}{\partial x^j}$ to cancel out, and all that remains is the lower index i . All the trouble of seeing what's covariant, what's contravariant, and what's invariant is washed away into just seeing how the upper indices and lower indices cancel out in the end.

Sometimes in the mathematics literature in places where co- and contra-variance are of lesser importance, upper and lower indices are ignored and the notation describing something like a matrix-vector multiplication looks like this:

$$(\mathbf{A}\mathbf{v})_i = \sum_j \mathbf{A}_{ij} \mathbf{e}_j.$$

That is, the i th component of the product $\mathbf{A}\mathbf{v}$ is the sum over j of the ij th component of \mathbf{A} with the j th component of \mathbf{v} . For a matrix-matrix multiplication this would look like

$$(\mathbf{AB})_{ik} = \sum_j \mathbf{A}_{ij} \mathbf{B}_{jk}.$$

Notice the pattern: when we wish to multiply these objects, we sum over a common index (j in the above equations) that we make both objects share. The point is *an index is always repeated and summed over* when we do a multiplication.

Einstein took the bold, but ultimately brilliant step of making the convention that if we ever write an index twice, that *automatically means* that we are summing over it (unless we explicitly say we aren't). The above equations, in Einstein's scheme, now become

$$\begin{aligned} (\mathbf{A}\vec{v})_i &= \mathbf{A}_{ij} \vec{e}_j \\ (\mathbf{AB})_{ik} &= \mathbf{A}_{ij} \mathbf{B}_{jk}. \end{aligned}$$

A dot product between \mathbf{v} and itself would simply be $\mathbf{v} \cdot \mathbf{v} = e_i e_i$ in Einstein's convention. Note since i is an index that is summed over (and doesn't appear in the end result), we could just as well replace it by any symbol $e_i e_i = e_j e_j = e_{\smile} e_{\smile} = \mathbf{v} \cdot \mathbf{v}$.

Now returning back to co-variance and contra-variance, this scheme makes every equation in the chapter shockingly short. Equation (1.1) becomes $\mathbf{v} = a^i \mathbf{e}_i$, Equation (??) becomes $\mathbf{e}'_i = A_i^j \mathbf{e}_j$, Equation (??) becomes $a'_i = (A^{-1})_i^j a_j$. Equation (1.7) is further reduced to just

$$\frac{\partial}{\partial x'^i} = \frac{\partial x^j}{\partial x'^i} \frac{\partial}{\partial x^j}. \quad (1.8)$$

Similarly the transformation of infinitesimal coordinate changes is now just

$$dx'^i = \frac{\partial x'^i}{\partial x^j} dx^j. \quad (1.9)$$

and both of these are really just the chain rule plain and simple.

EXERCISE: Check invariance quickly using this method.

By doing exercises, this method is very quick and steady to get the hang of. We will adopt it for the rest of the book.

1.6 Exercises

Index

- affine space, 15
- basis, 6
- contravariance, 12
- coordinate system
 - Cartesian, 2
- covariance, 12
- equation of motion (EOM), 4
- euclidean space, 15
- gradient operator, 19
- indices
 - upper & lower, 19
- inner product, 14
- Jacobian, 16
- linear algebra, 6–13
 - linear independence, 6
 - span, 6
- origin, 2
- summation convention, 18
- vector
 - algebraic, 8
 - as a tuple, 3, 9
 - contravariant, 13
 - covariant, 13
 - geometric, 8
 - orthogonality, 13