1
4D Space

"A powerful theory is simple."

A powerful theory is simple. The simpler, the more powerful it becomes. The unified field theory starts with conventions that are accepted about the simplest and most basic aspects of observation. The space within which observations are made is just about as basic as you get. This chapter scrutinizes the convention of 4D space and develops some mathematical tools.

1.1
Convention

A point in ordinary 3D space is located by three independent coordinates. The distance between any two points is determined by the Pythagorean theorem. In an ordinary 4D space the added dimension is geometric time and the Pythagorean theorem is made to work in four dimensions.

Actually, there is the possibility of adopting any number of different geometries. For example, imagine a planar geometry that lies on the surface of a sphere. In this geometry, straight lines are arcs produced from intersections between the surface of the sphere and flat planes that cut through the center of the sphere. This particular geometry is called a *2D spherical geometry*. And there are others, too, although there is no need here to delve into them. Suffice it to point out that the geometry that is adopted depends on the problem being studied. It is really your decision as to the geometry in which to imagine reality. In other words, no single geometry is correct. On the other hand, say that you want to establish a "universal" geometry, in other words, a geometry that everyone prefers. Then, you would have to go out and persuade everyone to accept it. It is not the goal in this monograph to claim that the ordinary 4D geometry being used is universal. It is simply the one that was selected. It was selected by and large because it is easy and familiar.

Let us now begin the development. Imagine that you are at the beach watching a sun set. You would probably be willing to agree that at any given

Unified Field Theory for the Engineer and the Applied Scientist. Larry Silverberg
© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 978-3-527-40788-0

instant, the sun is located somewhere. This, of course, cannot be proven. The sun could be a hallucination or some other kind of figment of the imagination. Unless you accept certain rules about the things you observe, and unless the rest of us agree to accept these rules, it would be hopeless to proceed further. It follows that a community of like-minded people are required to agree that the things seen are located in the three dimensions of space and in the dimension of time. Indeed, the practice of interpreting everything as being located in the four dimensions of space and time is arguably the most basic convention established for the backdrop of reality. For now, suffice it to accept that space has three dimensions and time is another dimension. In this chapter the three dimensions of space and the dimension of time will be brought together to produce the 4D space.

The establishment of conventions is an evolutionary process that dates back to before recorded history. Today, only the results are being practiced. But at some point, conventions were established for a unit measurement of length and for a unit measurement of time. To establish these conventions, devices were manufactured to count multiples and fractions of units. The ruler and the clock were built to measure length and time. The spatial measurement is taken by placing the ruler up against a body and comparing a pair of coincident events on the body and the ruler. Similarly, the temporal measurement is taken by comparing a pair of coincident events. The ruler and the clock provide physical standards for measurement.[1)]

The idea of dimension is more recent, dating back some 500 years when an organized method was developed to coordinate or analyze measurements. The coordination method starts with the construction of a coordinate system. The coordinate system is a reference relative to which measurements are taken. The reference defines the starting point of a set of measurements and the directions along which they are taken. The most common type of coordinate system is called the *Cartesian coordinate system*, named after Renee Descartes [2]. Using a Cartesian coordinate system, measurements are taken in perpendicular directions.

1.2
Cartesian Coordinates

The location of the sun, or of any other event, is determined by four numbers. Three of them are referred to as the spatial coordinates x_1, x_2, and x_3 and one is the temporal coordinate x_4. The four numbers can be viewed as measurements along the axes of a four-dimensional coordinate system. Each

1) It is fascinating, although beyond the scope of this work, that these conventions produce, quite arbitrarily, spatial and temporal constants. The ruler takes the unit measurement of distance to be the same everywhere in space and the clock counts periodic events that are taken to have the same period throughout time. We are so accustomed to these constants that they appear to be natural even though they are really artificially constructed. In mathematics, this is called *congruence* [1].

axis is a dimension of 4D space. The four numbers collected together form the vector $\mathbf{x} = (x_1\ x_2\ x_3\ x_4)$. Whenever a measurement is taken along a direction that is not along one of the coordinate axes, a rule for the length of a measurement is needed. Drawing from the Pythagorean theorem, the *length* of the vector **x** is (see Figure 1.1)

$$\mathbf{x} = \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2}. \tag{1.1}$$

Equation (1.1) is developed in detail in Figure 1.1. Figure 1.1a shows a proof of the Pythagorean theorem. Figure 1.1b extends the definition of length to 3D space. Figure 1.1c extends the definition of length to 4D space. Notice that length is extended from 2D space to 3D space the same way it is extended from 3D space to 4D space. A 4D geometry that defines length according to Equation (1.1) is referred to as an *ordinary geometry*.[2)]

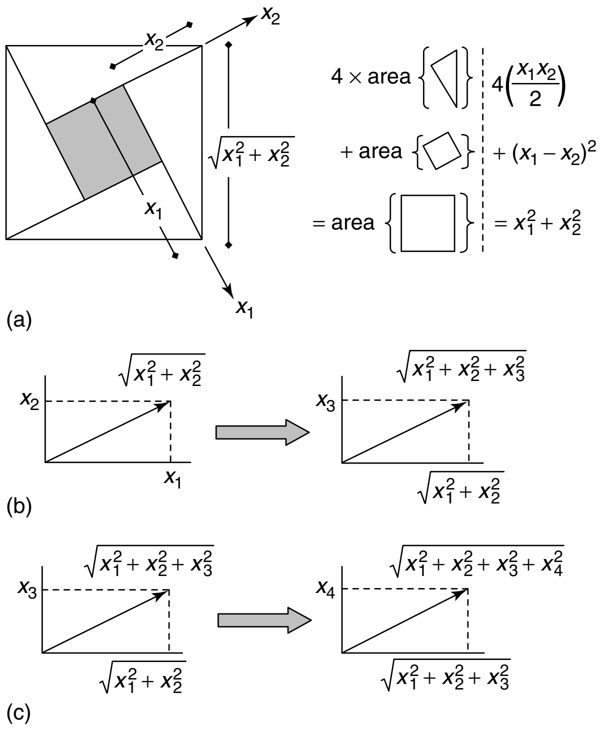

Figure 1.1 Length. (a) Proof of the Pythagorean theorem, (b) extension from 2D space to 3D space, and (c) extension from 3D space to 4D space.

2) A 4D geometry in which length is defined according to Equation (1.1) is also referred to as a 4D *Euclidean* geometry. Euclidean geometry is based on the works of Euclid [3]. Basic theorems in 4D geometry date back to the Greeks [4].

The proof of the Pythagorean theorem for 4D geometry treated the spatial coordinates and the temporal coordinate in the same manner, as though there is no real difference between them. But does that make sense? What assumptions did we make? To answer these questions, look at this proof more closely. Notice, to prove the Pythagorean theorem, that the area of a rectangle was first defined as the product of its sides. A notion of area was considered necessary to define length.

Area is basically defined as the number of unit squares in a rectangle. The Pythagorean theorem holds in a geometry in which area is defined this way. The reader should appreciate that there are other senses of length that are meaningful, too, although these senses of length would either not define area this way, or not define it at all. For example, imagine constructing a space from a grid of lines (see Figure 1.2). The distances between the lines can be infinitesimal or finite. It is fun to think of the lines as roads and the space as a city block. When traveling from **A** to **B** the length of travel is $a + b$, not $\sqrt{a^2 + b^2}$. This is another acceptable way to measure length even though it is not being adopted here.

Returning to the problem at hand, the question remains why must area be defined as the product of base and height when one of the coordinates is spatial and the other is temporal? In fact, is it even necessary to define area when one of the coordinates is temporal? And if so, how would we know whether this fourth coordinate is conventional time? The answer is that area would not necessarily be defined this way and that this definition of length does not necessarily make sense when one of the coordinates is temporal. Indeed, one should not view the temporal coordinate x_4 and conventional time t as identical. In the development below, the temporal coordinate x_4 is called *geometric time* and t is called *conventional time*. The temporal coordinate will be taken to be a fourth coordinate in an ordinary geometry, that is, a geometry that satisfies the Pythagorean theorem. However, the justification is predicated on the existence of a relationship between geometric time x_4

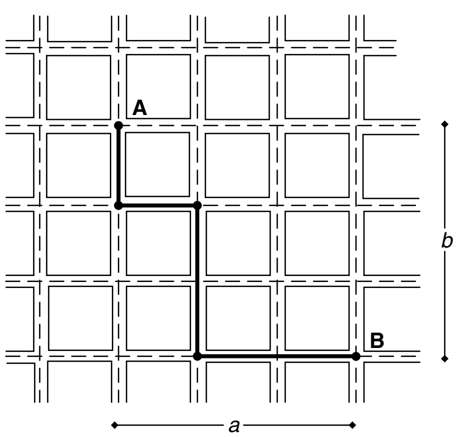

Figure 1.2 Road length.

and conventional time t, which will be developed shortly. From the outset, it should be understood that the immediate goal is to develop, as a convention, an ordinary 4D geometry with x_4 as the temporal coordinate. The basic question is whether there exists a relationship between x_4 and t and what it is.[3)]

1.3 Time as a Fourth Dimension

To develop the relationship between geometric time x_4 and conventional time t, we need to think more about how time is measured. Do we really know when an event actually occurs? Do we measure it directly and, if not, what do we actually measure? The development below is divided into three steps. In the first step, the relationship between conventional time and the images that we see is discussed. This step clarifies the difference between the time of a measurement and the time of an event. In the second step, the concept of the complex number is reviewed. The complex number is reviewed because of its central role and because there is a lot of confusion surrounding it. The aim of the review is also, in part, to trace the complex number back to its origin, so that this part of the development can be as intuitive as the rest. Finally, in the third step, the relationship between x_4 and t is exposed.

1.3.1 Images

Imagine that it is a clear night and you gaze up and see thousands of stars. The stars appear to be on the surface of a sphere, equidistant from you. In fact, this is precisely the sensation replicated in a planetarium, where stars are projected onto a spherical surface. But do the stars in the sky really lie on a spherical surface? The spatial image that you see is really of events that occurred at very different times, in some cases thousands of years apart, at very different distances.

Now, look down at your hand. The spatial image that you construct of it is formed from events that occurred almost simultaneously. It follows that depth perception and the images of the bodies that are formed in our minds are largely a result of relative spatial information. Spatial images of bodies are perceived by following changes in patterns. Without these patterns, our minds place the spatial images at equal distances to us to form images on a sphere.

When an event occurs, it takes time for the signal of the event to be communicated to your eye or to a clock. Time is counted in terms of the signal

3) The relationship between x_4 and t was the principle question answered in the theory of relativity. It was first studied soon after light was recognized to be a wave, upon which it was realized that the time during which an event is measured corresponds to the instant the signal reaches the clock and not when the event actually occurs [5]. The measured time was called *retarded* time. The theory of relativity provides a means for accounting for retarded time.

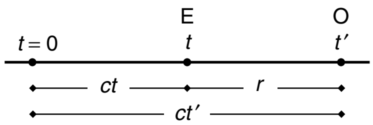

Figure 1.3 The communication line.

that reaches you. In fact, you do not know precisely where the signal comes from or the precise instant it occurs because events at different distances and at different instances can reach you at the same time. This is illustrated by the *communication line* shown in Figure 1.3.

The origin of the communication line is at time $t = 0$. An event E occurs at time t and observer O records the measurement at time t'. The spatial distance between event E at time t and the measurement by O at time t' is r. Time t' and time t are related by

$$ct' = r + ct, \tag{1.2}$$

where c is the speed of the signal. The observer only records the event at t'; the distance r, the time t of the event, and the speed c of the signal are not measured directly.

1.3.2
Complex Numbers

The communication line shown in Figure 1.3 is a primitive system showing time t and distance r on the same axis. This representation will be modified with the help of complex numbers. Using complex numbers, a geometric relationship between the time t of the event E and the distance r between the event and the measurement will be developed. The geometric relationship will produce a fourth dimension. Before proceeding with that, though, it will be instructive to review complex numbers.

A tremendous amount of confusion surrounds the complex number. After all, what is $\mathrm{i} = \sqrt{-1}$? The confusion that surrounds the complex number is traced to the way it became popular; it is a matter of history. However, there is nothing abstract about it as the following explains. Let us first recall that, before the coordinate system had become popular, geometry's principle role was to deal with shapes constructed from intersecting line segments, surfaces, and volumes. The line segments, surfaces, and volumes were represented by positive numbers and zero. Negative numbers are not necessary for the construction of shapes. In fact, not surprisingly, the earliest treatments with negative numbers and with complex numbers were met with skepticism.

As geometry evolved, directed line segments, called *rays*, were introduced to help with the analysis performed in geometry. The ray was introduced during the same period as the coordinate system because both kept track of direction.

In fact, the coordinate system and the ray facilitated negative numbers because they too keep track of direction. Moreover, the ray, or the *vector* as it is called today, turned out to act a lot like a number.

For the purpose of this discussion, let us extend the number system to include all real numbers, positive, zero, and negative, and let us restrict our attention to rays $\mathbf{R} = (x, y)$ that can be placed anywhere in a plane (see Figure 1.4).

Any two rays can be placed end-to-end and combined by the addition operation

$$\mathbf{R}_1 + \mathbf{R}_2 = (x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2),$$

and a ray can be lengthened by the multiplication operation

$$a\mathbf{R} = a(x, y) = (ax, ay).$$

These two operations can be performed in any order and produce the same results (see Figure 1.5). They obey the five rules of ordinary arithmetic, namely, the associative rules of addition and multiplication, the commutative rules of addition and multiplication, and the distributive rule. Therefore, rays can be manipulated just like numbers. This is briefly the foundation of real vector algebra.

The concept of complex numbers properly arises by introducing one more operation, that is, the rotation of a ray. Let us rotate a ray 90° (see Figure 1.6). This is represented by the operation

$$\mathrm{i}\mathbf{R} = \mathrm{i}(x, y) = (-y, x),$$

where i is simply a prefix that means rotate the ray 90° counterclockwise [6].

It is easy to check for yourself that the rotation of a ray and the combining and lengthening operations of rays can be performed in any order to produce the same results. They also satisfy the rules of ordinary arithmetic. This produces

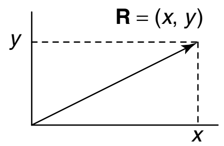

Figure 1.4 Ray.

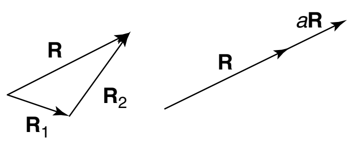

Figure 1.5 Real vector algebra.

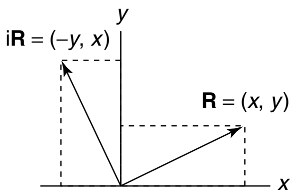

Figure 1.6 The 90° rotation.

the general operation

$$(a + \mathrm{i}b)(x, y),$$

where a, b, x, and y are real numbers and i is still the 90° rotation. To obtain $(a + \mathrm{i}b)(x, y)$, a ray (x, y) that is lengthened by an amount a is added to the ray (x, y) that has been lengthened by an amount b and rotated 90°. Today, it is standard to write $(x, 0)$ as simply x and so $(0, y) = \mathrm{i}(y, 0)$ is written as iy. The ray (x, y) can be viewed as the sum of $(x, 0)$ and $(y, 0)$ rotated 90°, that is,

$$(x, y) = (x, 0) + \mathrm{i}(y, 0) = x + \mathrm{i}y.$$

This means that we can write the general operation $(a + \mathrm{i}b)(x, y)$ as

$$(a + \mathrm{i}b)(x + \mathrm{i}y).$$

Rays, together with rotations, can be manipulated like numbers. Since the operation i acts like a number, it can be regarded as a number. This is briefly the foundation of complex algebra.[4)]

1.3.3
The Temporal Coordinate x_4

Having discussed the relationship between the measurement time of an event and the time of an actual event, and having reviewed the complex number, we are now ready to create an ordinary fourth dimension. Simply write Equation (1.2) as

$$ct' = r - \mathrm{i}x_4,$$

where we let

$$x_4 = \mathrm{i}ct. \tag{1.3}$$

4) The confusion surrounding the complex number would likely be overcome if, when introducing the concept for the very first time, the student is taught that i is a 90° rotation of a ray, that $\mathrm{i}^2 = -1$ is a 180° rotation of a ray, and that the complex number is a general operation used when combining, lengthening, and rotating rays. Our young people would then see why there is nothing imaginary or complex about the complex number.

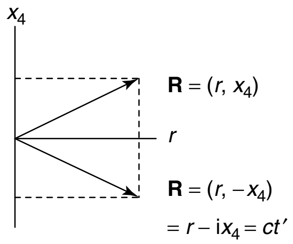

Figure 1.7 Geometric time on a perpendicular axis.

Notice in Equation (1.3) that geometric time x_4 is the distance that a signal travels in one unit of conventional time (see Figure 1.7).

Figure 1.7 shows that the measurement ct' can be viewed as a ray in a complex plane; the coordinates of the ray are r and $-x_4$. We never actually measure r or x_4 directly, only ct'. Since r and x_4 are never actually measured, the length of the ray in Figure 1.7 is freely defined by the Pythagorean theorem.[5] Notice below that the length $\sqrt{r^2 + x_4^2}$ cancels out the measurement of ct':

$$
\begin{aligned}
ct' &= r + ct = r - ix_4 \\
&= \sqrt{r^2 + x_4^2}\left[\frac{r}{\sqrt{r^2 + x_4^2}} - i\frac{x_4}{\sqrt{r^2 + x_4^2}}\right].
\end{aligned}
$$

By placing x_4 on an axis, we are interpreting it to be real. But, according to Equation (1.3), x_4 is pure imaginary if conventional time t is regarded as real. This contradiction is resolved by simply regarding conventional time as pure imaginary. The conventional view is to think of t as real and x_4 as pure imaginary. But, it is equally correct to view x_4 as real and t as pure imaginary. When x_4 is viewed as real, as we did in Figure 1.7, the coordinates x_1, x_2, x_3, and x_4 become real coordinates of an ordinary four-dimensional space.

The development above shows that physical reality can be represented in an ordinary 4D space. In particular, the question of the use of the Pythagorean theorem when a coordinate is temporal was answered. The relationship between geometric time x_4 and conventional time t was found and their relationship with the actual measurement t' of the event was clarified.

1.4 The Hypercube

Four-dimensional geometry is presently omitted from one's undergraduate studies. The reason appears to be the widely held belief that 4D geometry

5) Another consequence of only measuring $r + ix_4$ and not r and x_4 individually is that functions of measurements are always of the form $f(z)$ in which $z = r + ix_4$ is complex. This is a more restrictive form than $f(r, t)$. In complex analysis functions of the form $f(z)$ are called *analytic functions*.

is complicated and only loosely connected to reality. After all, what is a 4D shape and what connection does it have to reality? You will find out in this monograph that physical behavior has everything to do with ordinary 4D geometry. As to visualizing 4D geometry, you will see that it is unnecessary, so this is not an obstacle. However, a few features of four dimensions will need to be understood. For example, what do the faces of a 4D cube look like? How many faces does a 4D cube have?

The important features of four dimensions can be revealed by constructing a 4D cube. So, we now construct a 1D cube (line), next a 2D cube (square), then a 3D cube (ordinary cube), and, finally, the 4D cube (hypercube). Figures 1.8 through 1.10 show the construction process. The n-dimensional cubes are constructed in four steps, labeled (a), (b), (c), and (d). In step (a) of the construction of the n-dimensional cube, its vertices are listed in a table. The vertices consist of the 2^n combinations of 0 and 1. In steps (b) and (c), the faces of the n-dimensional cube are constructed from the constraints in the table. For example, in Figure 1.8, the faces of a 2D cube consist of the lines 1–2, 2–4, 4–3, and 3–1. Line 1–2 is the constraint $x_2 = 0$, line 2–4 is the constraint $x_1 = 1$, line 4–3 is the constraint $x_2 = 1$, and line 3–1 is the constraint $x_1 = 0$. The unwrapped faces of the n-dimensional cube are put together, drawn, and the coordinates are labeled. In step (d), the unwrapped faces of the n-dimensional cube are wrapped. In the case of the 4D cube, step (d) was not performed because of the difficulty of representing a fully formed 4D cube in a 2D drawing. The last construction step performed for the 4D cube was step (c) showing its unwrapped faces. Before continuing further, the reader would benefit by taking a moment to study these figures. The development of the 4D cube can become rather intuitive.

	1	2	3	4
x_1	0	1	0	1
x_2	0	0	1	1

(a)

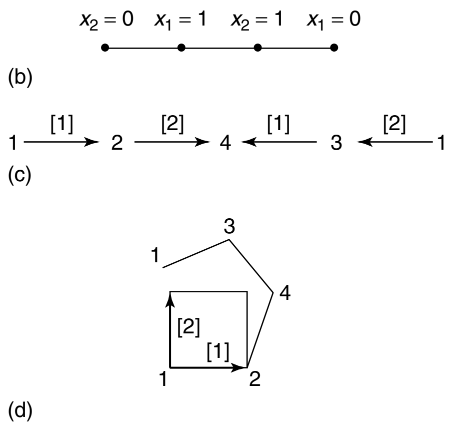

Figure 1.8 Constructing a 2D cube. (a) Vertices, (b) Constraints, (c) Unwrapped faces, and (d) Wrapping the faces.

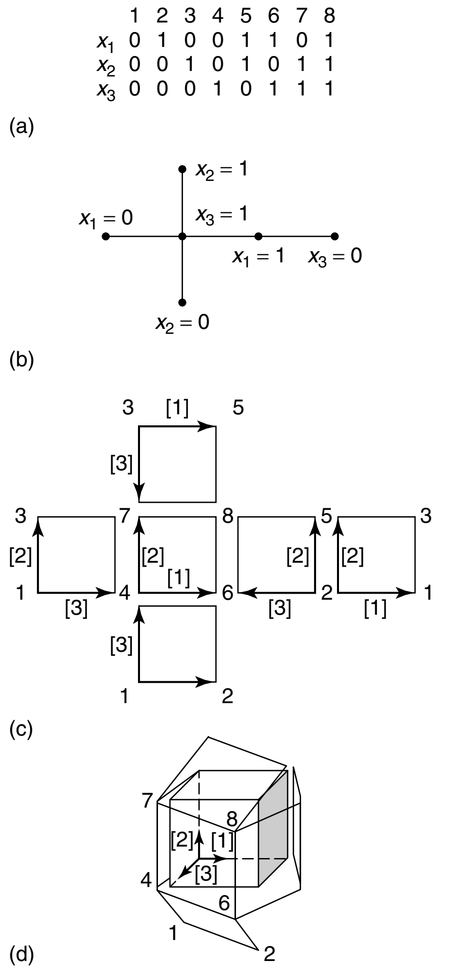

Figure 1.9 Constructing a 3D cube. (a) vertices, (b) constraints, (c) unwrapped faces, and (d) wrapping the faces.

The important features that have been uncovered are that the faces of a 4D cube are 3D cubes and that there are eight of them. For future reference, the four 3D cubes located at $x_1 = 1$, $x_2 = 1$, $x_3 = 1$, and $x_4 = 1$ will be referred to as *positive* cubes and the four 3D cubes located at $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, and $x_4 = 0$ as *negative* cubes.

1.5 The 4D Right-hand Rule

The wrapping of the eight cubes to form the 4D cube reveals a unique right-hand orientation for the cubes [7]. Look again at the unit 4D cube shown in

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_1	0	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
x_2	0	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
x_3	0	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
x_4	0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1

(a)

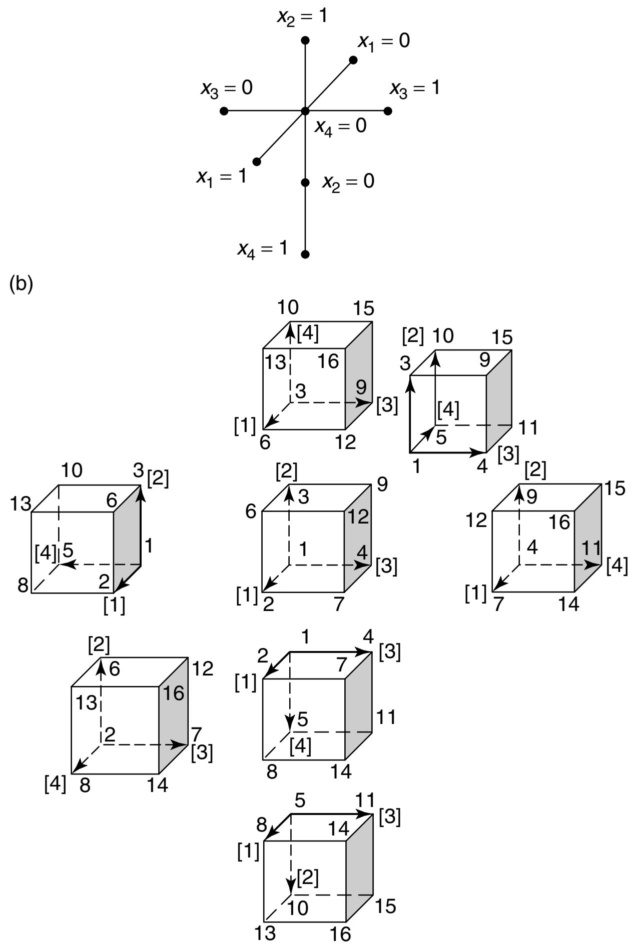

Figure 1.10 Constructing a 4D cube. (a) vertices, (b) constraints, (c) unwrapped faces, and (d) wrapping the faces.

Figure 1.10. Notice as a convention we had let the right-hand orientation of the $x_4 = 1$ positive cube be (1 2 3). As shown, once the right-hand orientation of the $x_4 = 1$ cube had been chosen this way, the right-hand orientations of the other positive cubes became (1 3 4) for the $x_2 = 1$ cube, (1 4 2) for the $x_3 = 1$

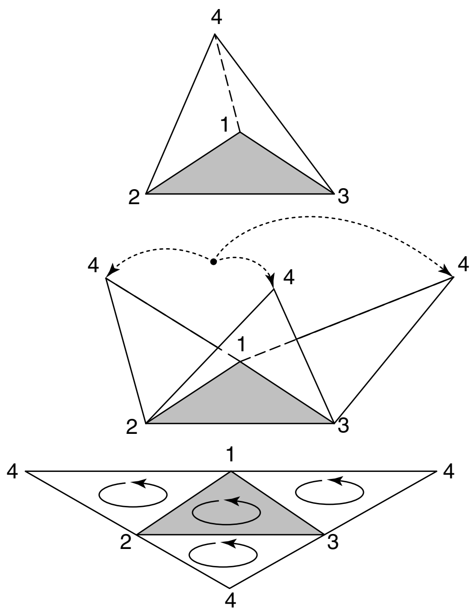

Figure 1.11 4D Right-hand rule.

cube, and (2 4 3) for the $x_1 = 1$ cube. The negative cubes follow left-hand orientations. If the orientation of any one cube had been selected differently, for example, if we had tried to choose (1 4 3) as the right-hand orientation of the $x_2 = 1$ cube, we would have found, during construction, that the vertices of the cubes in Figure 1.10 do not match up.

Figure 1.11 shows a diagram that provides a convenient way of remembering the correct orientations of the 3D cubes [8]. The diagram is of a tetrahedron. The vertices of its base are the right-handed triplet (1 2 3) and the top vertex is 4. The right-hand orientations of the other three triplets are uncovered by opening up the tetrahedron like the pedals of a flower. Each triangle reveals a right-handed triplet by following its vertices in the counterclockwise direction. As shown, (1 2 3), (1 4 2), (1 3 4), and (2 4 3) are each right-handed.

1.5.1 The Right-hand Indices

The two most common vector operations in 3D vector analysis are the dot product and the cross product. The following extends them to four dimensions. The extension of the dot product to four dimensions is immediate. In 3D, the dot product of the two vectors **B** and **C** is $\sum B_i C_i$ in which the sum is from 1 to 3. In 4D, the sum is merely from 1 to 4. The dot product is written in vector form as $\mathbf{B} \cdot \mathbf{C}$.

The extension of the cross product to 4D space is also straightforward when right-hand indices are employed. In 2D, the 2D right-hand indices ε_{ij} ($i = 1, 2$; $j = 1, 2$) are defined to be as zero except $\varepsilon_{12} = 1$ and $\varepsilon_{21} = -1$. Using the 2D right-hand indices, one can construct a 2D vector **A** that is perpendicular to the vector **B**. The components of the vector **A** are[6)]

$$A_i = \sum_{j=1}^{2} \varepsilon_{ij} B_j. \tag{1.4a}$$

In 3D, the 3D right-hand indices ε_{ijk} ($i = 1, 2, 3$; $j = 1, 2, 3$; $k = 1, 2, 3$) are defined to be zero except $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1$ and $\varepsilon_{321} = \varepsilon_{213} = \varepsilon_{132} = -1$. Using the 3D right-hand indices, the components of the 3D vector **A** that is perpendicular to the vectors **B** and **C** are

$$A_i = \sum_{j=1}^{3} \sum_{k=1}^{3} \varepsilon_{ijk} B_j C_k. \tag{1.4b}$$

Notice in 3D that the nonzero right-hand indices are defined in accordance with the 3D right-hand rule. Similarly, in 4D, the 4D right-hand indices ε_{ijkl} ($i = 1, 2, 3, 4$; $j = 1, 2, 3, 4$; $k = 1, 2, 3, 4$; $l = 1, 2, 3, 4$) are defined to be zero when any two indices are repeated and the nonzero indices are defined in accordance with the 4D right-hand rule. The components of the 4D vector **A** that is perpendicular to the vectors **B**, **C**, and **D** are

$$A_i = \sum_{j=1}^{4} \sum_{k=1}^{4} \sum_{l=1}^{4} \varepsilon_{ijkl} B_j C_k D_l. \tag{1.4c}$$

The right-hand order is conveniently remembered using Figure 1.11. To show that **A** in Equation (1.4a) is perpendicular to **B**, that **A** in Equation (1.4b) is perpendicular to **B** and **C**, and that **A** in Equation (1.4c) is perpendicular to **B**, **C**, and **D**, one simply verifies by calculation that $0 = \sum A_i B_i$, that $0 = \sum A_i C_i$, and that $0 = \sum A_i D_i$. In 2D, the sums are from 1 to 2, in 3D from 1 to 3, and in 4D from 1 to 4 (see Exercise 1.2). The cross product is written in 3D vector form as $\mathbf{A} = \mathbf{B} \times \mathbf{C}$. No similar vector forms exist for the cross products of 2D vectors and of 4D vectors.

1.6 Exercises

1.1 Let **A** and **B** be 4D vectors. Show that $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta$, where $|\mathbf{A}|$ is the length of **A**, $|\mathbf{B}|$ is the length of **B**, and θ is the angle between them. *Hint*: Any two vectors **A** and **B** lie in a plane.

6) Equation (1.4a) is the same as the equation $\mathbf{A} = \mathrm{i}\mathbf{B}$ where i is a 90° counterclockwise rotation and **A** and **B** are 2D vectors. So, the 2D right-hand indices are equivalent to the imaginary number i. The *n*D right-hand indices extend i to *n* dimensions.

1.2 Show in Equation (1.4c) that **A** is perpendicular to **B**, **C**, and **D**.

1.3 The sides of a right triangle on the surface of a sphere are s_a, s_b and the diagonal is s_c. Show that $\cos(s_c/R) = \cos(s_a/R)\cos(s_b/R)$, where R is the radius of the sphere radius. This is the Pythagorean theorem for a 2D spherical geometry. Also, show that when R approaches infinity this equation reduces to the ordinary Pythagorean theorem.