General Relativity

Installment II.

September 8, 2003

1 The Story So Far

From Galileo we know that physical behavior is unaffected by uniform motion. The image of a vehicle moving at constant velocity is often used to exemplify this physical law. Galileo liked ships for this; Einstein took trains for the mental voyage, which he called a gedanken (=thought) experiment.

Observer G is on the ground, not moving, and watching a train going by in heir positive x direction. (We pretend that the earth does not move.) A second observer, call heir T, is going by in a box car. T sets up a rectangular coordinate system with heir x' axis lined up with G's x axis. These being classical physicists, so we assume, with them, that they both have the same times in their two systems. That is,

$$t' = t, \tag{1}$$

which is very nearly true in daily life. This would once have been thought so obvious as to not need mention. Yet, the very fact that we feel the need to mention in a discussion of this kind is already a departure from the classical imagery.

At time t', T observes a fly in heir coordinate system at x'. At time $t = t'$, G observes the fly at coordinate x in heir system. What is the relation between the two measurements of position? (There are some minor effects that might complicate the study of this question but we may neglect them in a thought experiment. One is the tiny travel time of light between the two systems. Can you think of any others?) Back in the good old days (before March 14, 1879, say), we would have said that

$$x' = x + vt \tag{2}$$

Let $f(x)$ be a force in the x direction in G's system. Then Newton's second law says that, for an object of mass m,

$$m\frac{d^2x}{dt^2} = f(x). \tag{3}$$

We may use (2) to see what form the law takes in T's system. From (2) we see that $f(x) = f(x' - vt)$. We introduce a new function f' with the property that $f'(x') = f(x' - vt) = f(x)$. The new function is the force at the point x but expressed in the primed coordinate. A function with this property is called a (Galilean) scalar. We then see that

$$m\frac{d^2x'}{dt'^2} = f'(x'), \tag{4}$$

The form of the law is unchanged by uniform motion as just Galileo wrote This invariance of the form of the law (sometimes called covariance) under the transformation given by (1)-(2) is not true for Maxwell's equations of electromagnetism, which are covariant under a different transformation called the Lorentz transformation. Let us pause to look briefly at this transformation even though it is a bit early for this.

2 The Lorentz Transformation

Since we need to specify both position and time to identify an event such as a meeting with someone, we should not be surprised that the time of the event would be treated on the same footing as the place. If we keep on thinking for now of a space of one dimension we then have to give simply t and x to make a date. But if the person we are meeting is not on the uniformly moving train with us, we might want to tell hem t' and x', which we could find from the Galilean transformation laws, (1) and (2). As we saw, the form of Newton's laws of motion are not modified by this transformation and, as these are quite good laws, we might happily adopt the Galilean transformation in physics. But Lorentz (and Poincaré) found that invariance of form under that transformation did not apply to Maxwell's equations (which we shall write down at some point). Suffice it to say for now that Maxwell's equations do not change form when, instead of the Galilean transformation rules, we use the transformation

$$x' = \gamma(x + vt) \tag{5}$$

$$t' = \gamma(t + \frac{vx}{c^2}) \tag{6}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{7}$$

and c is the speed of light in a vacuum, that is, 3×10^{10} cm/sec.

This difference in preferences for describing how coordinates in spacetime change when we go into a moving frame seems to pose a problem of choice, but it is not so grievous as it might first appear. Light goes $300,000$ km/sec — several circumferences of the earth — in one second. In our daily experience, including our studies of planetary motion, nothing is going that fast. If $|v| << c$, then $\gamma \approx 1$ and (5)-(6) are practically indistinguishable from (1) and (2) so all is well many aspects of daily life, including most mechanical problems. However, when things are moving at high enough speeds, the two transformation laws differ markedly and there then is a real choice to be made: which of the two laws of transformation is the right one? Or is there something more fundamental in our thinking that has gone agleigh? Perhaps the physical laws do not remain the same when we go into a uniformaly moving system?

In 1905, Einstein argued for the validity of the Lorentz transformation and, to deal with the resulting dilemma, he showed how the laws of mechanics needed to be modified when the bodies in question move at high speed. We will discuss all that and look further into these transformation laws after we have studied some vector analysis by way of preparation. This will be our standard procedure — a little physics interspersed with a little mathematical discussion. This is the sort of trick used by novelists in which one jumps from place to place and protagonist to protagonist in successive chapters till, at the end, it all comes nicely together, with any luck. Such weaving together of several strands in the narrative is the plan we shall follow here. But please take a few moments now to answer the following question.

You know how to find x' and t' given x and t from (5)-(6) or (5)-(6). Suppose you are given x' and t'. What are the transformation laws that tell you x and t for both the Galilean and Lorentzian transformations? Do these inverse transformations sound reasonable? Explain that (to yourselves).

But now, let us come to the real reason that we have opened the issue of tranformation laws at this time. For this we need to look at Figure 1, a plot of the coordinate system in spacetime that we shall often use. In daily life, we use different units for space and for time, even though we now want to consider time as a coordinate in our world, our own spacetime. This turns out to be more convenient for visualizing what is happening in relativistic matters if we use the same units for space and time. That is why some people use light years, for example. So let us fix the problem using c, which appears to be a constant of Nature and has the dimension of length/time. We can use either x/c as our space coordinate, which then is expressed in time units (as in light years), or we can take ct to mark time. Either way, the value of c becomes unity in the new units, so that takes care of that, until you need to express things in everyday units (And even that is not difficult.)

Because $c = 1$ in these natural units, Figure 1 looks quite simple. The t and x axes are those of observer G. They are at right angles to each other since nothing indicates that things need to be otherwise and G wisely chose them this way. The choice of t for the vertical axis is the general preference among relativists and is adopted here. There are two lines at 45° in the figure and these represent light particles, or photons, one moving to the right at speed c (= 1) and one to the left at speed c. (Quantum mechanics is being ignored here y there are no worries about how we can know where the photon is when we also know its precise velocity.) Had we not adjusted the units, those two photon paths through spacetime — their worldlines — would be indistinguishable from the t axis on the scale of this plot. Note also that the worldline of the peripatetic T on heir train is shown in the figure going to the right at a respectable clip.

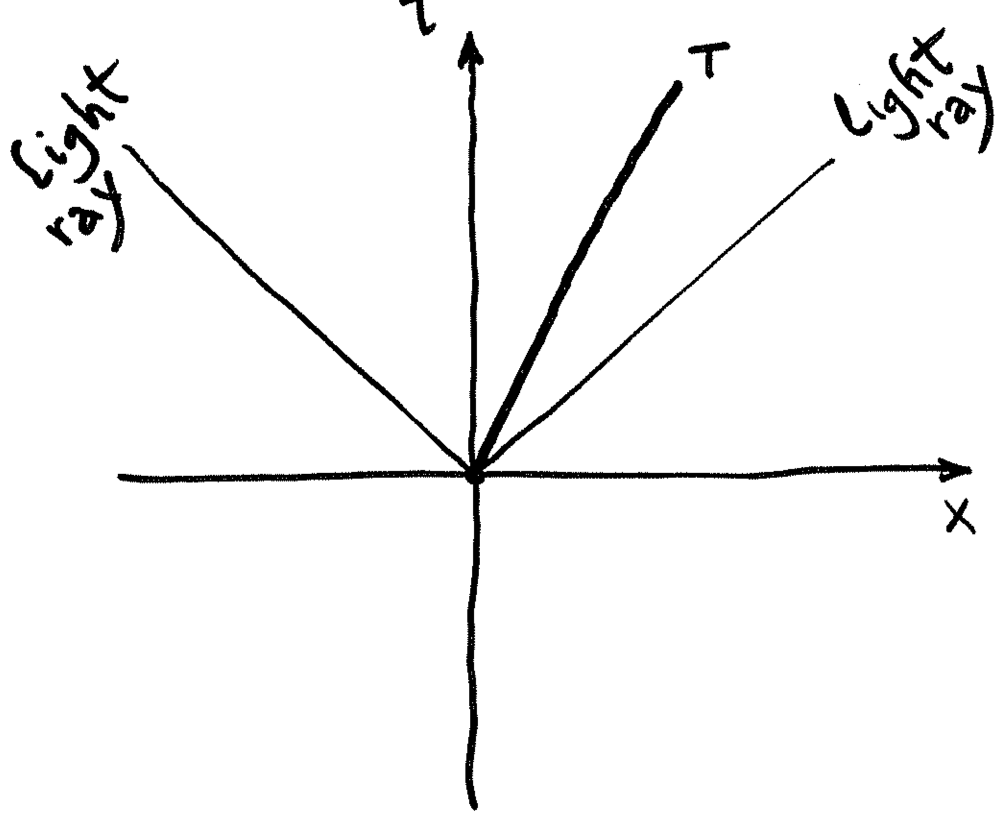

Figure 1.

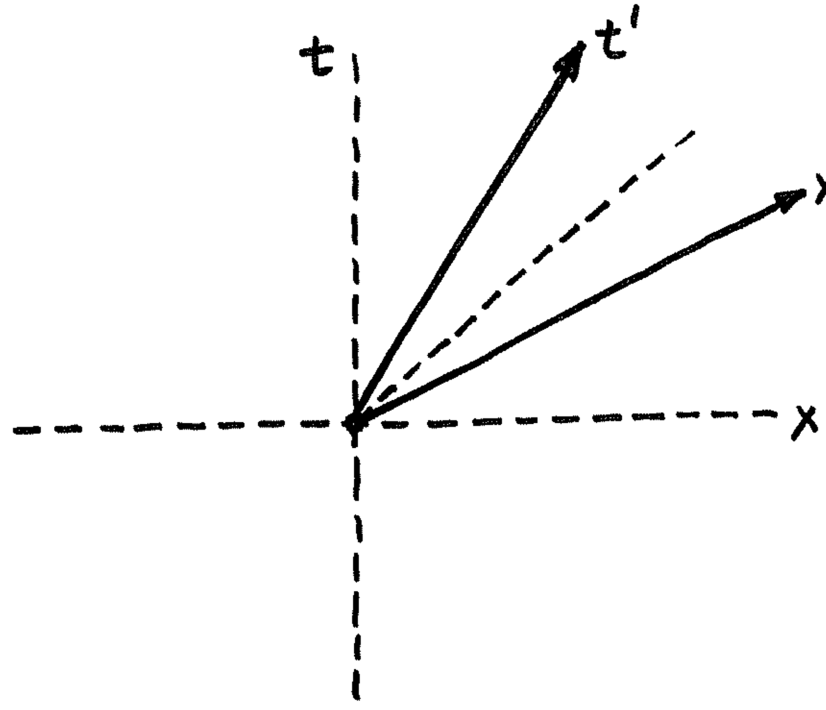

Figure 2.

In the coordinate system shown, G is not moving in space but is moving in time, that is, vertically in this coordinate system. How shall we draw the coordinate system of T? Naturally, we shall assume that like G, T in heir reference frame is moving in the direction of heir time axis. That means that the world line of T coincides with the t' axis. But where shall we place the x' axis?

We could look at the Lorentz transformation for guidance in drawing the x' axis, but let us try working from Figure 1 first. There are two features that seem significant there. The t and x axes are orthogonal to each other and the worldlines of light particles bisect the angles between the axes because light is traveling with respect to G at the speed c ($= 1$, in natural units). The orthogonality is just a matter of convenience, but the speed of light is apparently a constant of nature and it represents some important physics. So let us assume that light moves at speed $c = 1$ for observer T as well as for G. If you ask why, you will not be alone. One answer is that this is what follows from the Lorentz transformation. It also follows from some closely related assumptions made by Einstein in 1905, but in fact he was just trying to rationalize some features of electromagnetic theory. So the real answer to the question is that the world is made that way and the miracle is that Maxwell figured out the right equations for electromagnetism. It is no wonder that, on studying Maxwell's equations, Ludwig Boltzmann wrote "Was it a god who wrote these lines?' But as we shall see, these features of space are really an expression of its geometric structure.

If the speed of the photons is to be $c = 1$ in system T, the light ray must bisect the angle between the t' and x' axes (as you should verify). This is shown in Figure 2 where the coordinate axes of T are drawn in the system of G. What is striking is that the new axes are not orthogonal. This shows at least one case where oblique coordinates arise in a somewhat natural way and, as you can imagine, they can be forced on us for other reasons when we get into complicated geometrical situations. Let us then consider vector analysis in oblique coordinate systems before going on with the physics problems. In doing that we shall begin to learn the vectorial notation that will be needed in studying relativity, so buckle down for a longish haul.

3 Vector Analysis

3.1 Arrows

For now, let us stay in Euclidean space. The number of dimensions can be anything, if it is not too large, but the illustrations we use will be mainly in the (Euclidean) plane, to which we are confined by inadequate skill at drawing. For our present purposes, a vector may be represented as an arrow in the space. This convenience is going to be denied us once we get the hang of the notation and go to more general spaces, so it might be well to prepare for this eventuality by beginning to think about vectors more abstractly. A reading of the book on vector spaces by Halmos, for example, would make for nice bedtime reading for the serious student

To denote vectors, let us use boldface Latin letters. As to pictures, the most vivid illustration of a vector is an arrow representing a displacement; the length of the arrow is the displacement distance and the arrow points in the direction of the displacement. Suppose we are taken from point P to point Q by a displacement $\mathbf{a}$ and then to a third point (that shall remain nameless) by another displacement $\mathbf{b}$. The total displacement is by a vector $\mathbf{c}$ which is the sum of $\mathbf{a}$ and $\mathbf{b}$ as illustrated in Figure 3 by the parallelogram law. What could be simpler? Certainly not the abstract statement of all this, which we do not attempt here. What you see though is what a luxury it is to draw the vector in the space in which we are using it.

Subtraction of a vector, as in $\mathbf{a}-\mathbf{b}$, is effected by adding the negative of the subtracted vector. The negative of a vector is obtained by reversing the direction of the arrow representing it. This is done algebraically by introducing the notion of multiplication of a vector by a real number. So suppose that $|\mathbf{a}|$ is the length of $\mathbf{a}$. Then, if q is a positive real number, $q\mathbf{a}$ is a vector in the same direction as $\mathbf{a}$ with length $q|\mathbf{a}|$. If q is a negative real number, $q\mathbf{a}$ is a vector in the direction opposite to that of $\mathbf{a}$ and with length $-q|\mathbf{a}|$.

This summary would be inadequate if you did not know all this already. So if you do not know it all quite well, immediate remediation is needed. This can be obtained by reading an elementary book on vectors or asking a helpful friend or instructor.

3.2 Notation

If we are in a space of dimension n, we need n linearly independent vectors to make a basis for representing any vector in the space. For our purposes, by saying two vectors are linearly independent we mean that they are not colinear. Let us denote these non-colinear basis vectors as $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. A concise way to write this is by means of what is called the *range convention*. In that case we simply denote the collection of basis vectors as $\mathbf{e}_i$ where it is understood that i ranges over the values $1, 2, \ldots, n$. If then $\mathbf{e}_i$ is a basis, we may represent any vector $\mathbf{a}$ as a linear combination of the $\mathbf{e}_i$. That is, we may write that

$$\mathbf{a} = a^1\mathbf{e}_1 + a^2\mathbf{e}_2 + \ldots + a^n\mathbf{e}_n \tag{8}$$

The numbers $a^1, a^2, \ldots, a^n$ are called the components of $\mathbf{a}$ and we may write them as a^i in view of the range convention. (It is understood that $i = 1, 2, \ldots, n$.)

People new to this way of writing things often ask why they are written this way. Many are especially bothered by the fact that the index on the components is written as a superscript where it may be confused with an exponent. If you happen to know two people with the same given names, do you not manage to keep things straight? I myself know of a few different people called Engelbert, but I know which is which. Anyway, it is all a matter of

convention. (You may recall this exchange from the 'Hunting of the Snark,' where there is much wisdom to be found,
What's the good of Mercator's
North Poles and equators,
Tropical zones and meridian lines?
So the Bellman would cry.
And the crew would reply,
They are merely conventional signs.)

Another way to write (8) is

$$\mathbf{a} = \sum_{i=1}^{n} a^i \mathbf{e}_i \ , \tag{9}$$

which has a nice feel to it. But an even more convenient version of this is provided by the *summation convention* that states: whenever a suffix is repeated with one index up and one down it is understood that this suffix is to be summed over its full range. Thus, on adopting this convention, we may write (9) as

$$\mathbf{a} = a^i \mathbf{e}_i \ . \tag{10}$$

Once you have used this notation for a while, it will become second nature to you.

3.3 Inner Product

As we are in Euclidean space, we may use such old standards as the inner, or dot, product. If θ is the angle between two vectors $\mathbf{a}$ and $\mathbf{b}$, then we may write

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta. \tag{11}$$

The dot product of two orthogonal vectors is then zero and the dot product of a vector with itself is the square of its length. All this is the situation in the wondrous space of Euclid. Of course, this plot will thicken, but let us see how things look when we use the notation that we have just introduced.

Assume that we have a basis $\mathbf{e}_i$ and two vectors $\mathbf{a} = a^i \mathbf{e}_i$ and $\mathbf{b} = b^j \mathbf{e}_j$. (Here i and j are dummy indices since they are summed over.) Then

$$\mathbf{a} \cdot \mathbf{b} = a^i b^j \mathbf{e}_i \cdot \mathbf{e}_j \ . \tag{12}$$

If the coordinate axes *were* chosen to be mutually orthogonal and the basis vectors had unit length, we would have

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \tag{13}$$

where δ_{ij} is called the Kronecker delta; $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$. The Kronecker delta is our first instance of an object with two indices; both of them are subject to the range convention and the summation convention. Also, this delta is symmetric in its indices: $\delta_{ij} = \delta ji$. We may then write

$$\mathbf{a} \cdot \mathbf{b} = a^i b^j \delta_{ij} \tag{14}$$

where a double sum is implied.

Let

$$a_i = a^j \delta_{ji} \tag{15}$$

On the left of this definition we have one lower free index, i. On the right, we have the *same* free index and the dummy index j (repeated). The grammar is that the same free indices must always appear on each side of an equality and this rule is very helpful in locating errors. So is the taboo on having and index appear more than twice in an expression. The process indicated in (15) is called lowering the index and it gives us a new set of numbers a_i with the index downstairs. Of course, because of the simple nature of the delta, the two sets of components (for that is what they are) are equal for the same index but this is a feature of this special case — Euclidean space with orthonormal basis vectors.

If you find this notation new, it may be helpful to relate it to one you may have seen before. (Even if you have not seen it before, it may be helpful.) Let us represent a^i as a vertical column of numbers, $a^1, a^2, \ldots$. And let us represent a_i as a horizontal row of numbers $a_1, a_2, \ldots$. In other words, our vector $\mathbf{a}$ has two kinds of components, though the distinction is not really needed (except perhaps conceptually) when we deal with rectangular Cartesian coordinates and it is usually not even mentioned in that case. Indeed, the two sets of components are in fact equal at this stage and we may write, for $n = 2$, that

$$a^i = \begin{pmatrix} a^1 \\ a^2 \end{pmatrix} \; ; \qquad a_i = \begin{pmatrix} a_1 & a_2 \end{pmatrix} . \tag{16}$$

Then, for example,

$$\mathbf{a} \cdot \mathbf{b} = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} a^1 \\ a^2 \end{pmatrix} = a_1 b^1 + a_2 b^2 . \tag{17}$$

This form of expressing the dot product is not quite the same as the one we shall be using since here the order of the vectors matters in this notation. (Those who know quantum mechanics will see an analogy also to the bra and ket notation of that subject.)

3.4 Oblique Coordinates

When we go through the same exercise that we just performed, but with oblique coordinate axes, some new aspects of vector analysis emerge. In Figure 4, we have a sketch of a two-dimensional situation with two coordinates whose axes are at an oblique angle to one another. Also shown is an arrow representing a vector $\mathbf{a}$ and two short arrows along the coordinate axes indicating the basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. We wish to write $\mathbf{a} = a^i \mathbf{e}_i$ but what is the geometric significance of the components? That is, how do we project the vector $\mathbf{a}$ onto the coordinate axes? In the figure, the standard projection is shown in which lines from the head of the arrow are drawn parallel to the axes and these cut the axes to give us the two vectors $a^1\mathbf{e}_1$ and $a^2\mathbf{e}_2$.

We may also compute the inner product of two vectors, $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{a}\cdot\mathbf{b} = a^i b^j \mathbf{e}_i \cdot \mathbf{e}_j \ . \tag{18}$$

Now let

$$g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j \ . \tag{19}$$

Then the inner product is written as

$$\mathbf{a}\cdot\mathbf{b} = a^i b^j g_{ij} \tag{20}$$

We see that, in the case of oblique coordinates, the situation resembles that with rectangular coordinates but that the Kronecker delta is replaced by g_{ij} (which is the same as g_{ji}). In this relatively simple situation, we can write δ_{ij} and g_{ij} as two-dimensional arrays of numbers, that is, as matrices. The latter case is a touch more complicated, but it is still quite computable and you should compute the individual values of g_{ij} if you haven't already visualized them. More than this, we may also lower the index on a^i:

$$a_j = a^i g_{ij} \ . \tag{21}$$

Here, the indices were written a little differently than before to give you an idea how things work, but the rules for their manipulation are the same. What is really different now is that the components a_i are not the same as the a^i and this points to an alternative mode of projection onto coordinate axes in the case of oblique coordinates.

First let us observe that we can write g_{ij} as an array as in

$$g_{ij} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \ . \tag{22}$$

In the present case, we have $g_{11} = g_{22} = 1$ and $g_{12} = g_{21} = \cos\vartheta$ where ϑ is the angle between the (oblique) coordinate axes. A little trigonometry then reveals that if we drop perpendiculars from the head of the arrow of $\mathbf{a}$ onto the coordinate axes we obtain segments on the axes whose lengths are given by $a_i = g_{ij}a^j$. This alternate way of generating components is not so meaningful geometrically. A more fruitful approach is obtained by introducing the so-called reciprocal basis.

We first introduce the reciprocal matrix g^{ij} defined so that

$$g_{ij}g^{jk} = \delta_i^{\ k} \tag{23}$$

where

$$\delta_i^{\ k} = \begin{cases} 1 & \text{for} \quad i = j \\ 2 & \text{for} \quad i \neq j \end{cases} \ . \tag{24}$$

Now we note that

$$g^{ij}a_j = g^{ij}g_{jk}a^k = \delta_k^{\,i}a^k = a^i \, . \tag{25}$$

Just as we used g_{ij} to lower an index, we can use its reciprocal, g^{ij}, to raise an index.

The main aim just now is to get the hang of these manipulations by poring over them. But do notice that

$$\mathbf{a} = a^i\mathbf{e}_i = \delta_i^{\,j}a^i\mathbf{e}_j = g_{i\ell}g^{\ell j}a^i\mathbf{e}_j = \left(g_{\ell i}a^i\right)\left(g^{\ell j}\mathbf{e}_j\right) . \tag{26}$$

This suggests that we may also write

$$\mathbf{a} = a_i\mathbf{e}^i \tag{27}$$

where

$$\mathbf{e}^i = g^{ij}\mathbf{e}_j \, . \tag{28}$$

There are as many of the $\mathbf{e}^i$ as of the $\mathbf{e}_j$ and they form what is known as the reciprocal basis since

$$\mathbf{e}^i \cdot \mathbf{e}^j = \left(g^{ik}\mathbf{e}_k\right) \cdot \left(\, g^{j\ell}\mathbf{e}_\ell\right) = g^{ik}g^{j\ell}g_{k\ell} = g^{ik}\delta_k^{\,j} = g^{ij} \, . \tag{29}$$

The last equality in this string is a definition really, but it demands to be made. We also note that

$$\mathbf{e}_i \cdot \mathbf{e}^j = g_{ik}\mathbf{e}^k \cdot \mathbf{e}^j = g_{ik}g^{kj} = \delta_i^{\,j} \, . \tag{30}$$

From this we conclude that $\mathbf{e}_i \perp \mathbf{e}^j$ for $i \neq j$ and that $\mathbf{e}_i \cdot \mathbf{e}^j = 1$ for $i = j$. The projection of $\mathbf{a}$ onto the reciprocal axes is illustrated in Figure 5. Two forms of projection onto oblique axes thus gives us two sets of bases (dual to each other, one sometimes says) with two sets of coordinates. All this prefigures some important geometry that we must try to appreciate on our way to the theory of gravity.

Before we go on though, there is one last point to be made about this notation. In the Euclidean space, the position vector of a point (the fly of the example) is

$$\mathbf{x} = x\mathbf{e}_1 + y\mathbf{e}_2 + \ldots \, . \tag{31}$$

This expression does not take advantage of the range and summation conventions. So we let $x^1 = x, x^2 = y, \ldots$ and now we can write

$$\mathbf{x} = x^i\mathbf{e}_i \, . \tag{32}$$

The square of the distance of the point from the origin is then

$$\mathbf{x}^2 = \mathbf{x} \cdot \mathbf{x} = g_{ij}x^ix^j = (x^1)^2 + (x^2)^2 + 2x^1x^2\cos\varphi \, , \tag{33}$$

as it ought to be.