Physics 407-08
Straight lines

What is a straight line? Consider the flat two dimensional surface, with the usual rectilinear coordinates $x,\ y$. What is a straight line?

One answer is $y=\alpha x+y_0$ to express y as a function of x, a linear function of x. But why is this a straight line? And what is a straight line?

Going back to Euclid, a straight line is the shortest distance between two points. But what is distance? Ie, in order to use this definition of a straight line, we need to know what "distance" means. Distance is a structure which is imposed on space by who knows what, but is crucial to knowing what straight lines are.

Let us use the usual ideas of distance, (as developed by the Greeks 2500 years ago, who probably got it from the Babylonians up to 3000 years ago), for such rectilinear coordinates we have

$$\Delta l^2 = \Delta x^2 + \Delta y^2 \tag{1}$$

as the fundamental definition of length in terms of x and y.

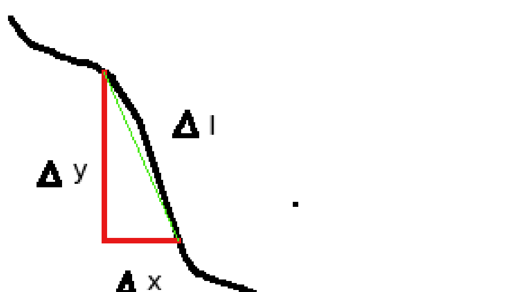

Figure 1: *Rectangular coordinates*

where Δx and Δy are little pieces of the coordinates along the curve. If we approximate the curve as a set of little "straight lines", then the total length of the line is $L=\sum \Delta l$.

Ie, the total length is $L = \sum \sqrt{\Delta x^2 + \Delta y^2}$. Let us try to make this more definite. Let us label the points along the line by λ, some completely arbitrary parameter. Then we can rewrite the above as

$$L = \sum \sqrt{\left(\frac{\Delta x}{\Delta \lambda}\right)^2 + \left(\frac{\Delta x}{\Delta \lambda}\right)^2} \Delta \lambda \tag{2}$$

or if we take the limit as $\Delta\lambda$ goes to zero, this becomes

$$L = \int \sqrt{\left(\frac{dx}{d\lambda}\right)^2 + \left(\frac{dy}{d\lambda}\right)^2} d\lambda \tag{3}$$

How do we look for the shortest line? Let us set up a whole sequence of curves all parametrized by λ and with the curves labelled by ϵ. Ie, for each ϵ, the curve is $x(\epsilon, \lambda),\ y(\epsilon, \lambda)$. All of these curves go from the fixed point $x_0,\ y_0$ at parameter λ_0 to the fixed point $x_1,\ y_1$ at the parameter λ_1. The parameters λ_1 and λ_2 are fixed– they do not depend on ϵ. (This will be so that we do not have to worry about end points in the integration by parts below). The length of each of these curves is

$$L(\epsilon) = \int_{\lambda_0}^{\lambda_1} \sqrt{(\partial_\lambda x(\epsilon, \lambda))^2 + (\partial_\lambda y(\epsilon, \lambda))^2} d\lambda \tag{4}$$

We will assume that the curves are labelled such that $x(\epsilon, \lambda_0) = x_0,\ y(\epsilon, \lambda_0) = y_0$ and $x(\epsilon, \lambda_1) = x_1,\ y(\epsilon, \lambda_1) = y_1$ for all ϵ.

Now let us take the derivative:

$$\frac{dL}{d\epsilon} = \int_{\lambda_0}^{\lambda_1} \frac{1}{\sqrt{(\partial_\lambda x(\epsilon, \lambda))^2 + (\partial_\lambda y(\epsilon, \lambda))^2}} (\partial_\lambda x \partial_\lambda \partial_\epsilon x + \partial_\lambda y \partial_\lambda \partial_\epsilon y) d\lambda \tag{5}$$

Now you can do an integration-by-parts with respect to λ to give

$$\frac{dL}{d\epsilon} = -\int_{\lambda_0}^{\lambda_1} \left(\partial_\lambda \left[\frac{1}{\sqrt{(\partial_\lambda x(\epsilon, \lambda))^2 + (\partial_\lambda y(\epsilon, \lambda))^2}} (\partial_\lambda x) \right] \partial_\epsilon x \right. \tag{6}$$

$$\left. + \partial_\lambda \left[\frac{1}{\sqrt{(\partial_\lambda x(\epsilon, \lambda))^2 + (\partial_\lambda y(\epsilon, \lambda))^2}} (\partial_\lambda y) \right] \partial_\epsilon y \right) \tag{7}$$

The end-point values in the integration-by-parts are zero because $\partial_\epsilon x(\epsilon, \lambda_0) = 0$ and similarly for all the other end point derivatives with respect to ϵ.

Now, in defining the functions of x and y with respect to ϵ they can be chosen completely arbitrarily. The only way that $\frac{dL}{d\epsilon}$ can be zero for an arbitrary choice of those functions is if each term is zero for all λ. Let me assume that ϵ has been chosen so that the desired curve, the straight line, occurs for $\epsilon = 0$. Then the equations become

$$\partial_\lambda \left[\frac{1}{\sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2}} (\partial_\lambda x(0,\lambda)) \right] = 0 \tag{8}$$

$$\partial_\lambda \left[\frac{1}{\sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2}} (\partial_\lambda y(0, lambda)) \right] = 0. \tag{9}$$

These equations look like a mess. However, let us redefine the parameter λ to a parameter I will call s

$$s = \int_{\lambda_0}^{\lambda} \sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2} d\lambda \tag{10}$$

so

$$\frac{1}{\sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2}} \partial_\lambda = \partial_s \tag{11}$$

Furthermore writing λ as a funtion of s,

$$\sqrt{(\partial_s x(0,\lambda(s)))^2 + (\partial_s y(0,\lambda(s)))^2} = 1 \tag{12}$$

[To do this in more detail, since

$$s = s(\lambda(s)) \tag{13}$$

so taking the derivative with respect to s on both sides,

$$1 = \partial_\lambda s(\lambda) \partial_s \lambda(s) \tag{14}$$

by the chain rule, and thus

$$\partial_\lambda s = \frac{1}{\partial_s \lambda} \tag{15}$$

Now,

$$\partial_\lambda = \partial_\lambda s \partial_s = \sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2}\partial_s \tag{16}$$

Ie, we can write all of the ∂_λ by ∂_s. . If we do this in $\sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2}$ we find that

$$\sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2} = \sqrt{(\partial_\lambda x(0,\lambda))^2 + (\partial_\lambda y(0,\lambda))^2}\sqrt{(\partial_s x(0,\lambda(s)))^2 + (\partial_s y(0,\lambda(s)))^2} \tag{17}$$

or

$$\sqrt{(\partial_s x(0,\lambda(s)))^2 + (\partial_s y(0,\lambda(s)))^2} = 1 \tag{18}$$

] Ie, s is the length along the curve to the point λ. Then the equations for the "shortest line" become

$$\partial_s^2 x = 0 \tag{19}$$
$$\partial_s^2 y = 0 \tag{20}$$

which have the trivial solution

$$x = x_0 + \alpha s \tag{21}$$
$$y = y_0 + \beta s \tag{22}$$

The above condition on s becomes $\alpha^2 + \beta^2 = 1$.

Clearly these solutions are what we usually consider as straight lines.

What have we used?

1) parametrize the curve.

2) finding the length of the curve in terms of the derivatives of x and y along the curve.

Note that a critical part of this derivation was the use of Pythagoras theorem. The length of a little piece of the curve is equal to the sum of squares of the changes in x and y along that little piece of the curve.

This length function, this thing that determines the length as a function of the changes in the coordinates, is a structure called the metric. It will turn out to be the most important thing in General relativity.

Polar Coordinates

There is another popular set of coordinates that is often used to describe the plane, namely polar coordinates. These are defined by the transformation

$$r\cos(\theta) = x \tag{23}$$

$$r\sin(\theta) = y \tag{24}$$

Ie, we again have two numbers, r and θ to define any point on the plane. We note that that there is something funny about these. At $r = 0$ (i.e., $x = y = 0$) any value of θ gives exactly the same point. Secondly, for x and y, any value of x and y designates a different point in the plane. But for r and θ this is no longer true. In addition to the problem at $r = 0$ there is another problem in that $r,\ \theta + 2\pi$ is exactly the same point as $r,\ \theta$. Ie, the coordinate label of a point is multi-valued as far as θ is concerned. We can either accept this or we can demand that θ only take values between, say, π and π.

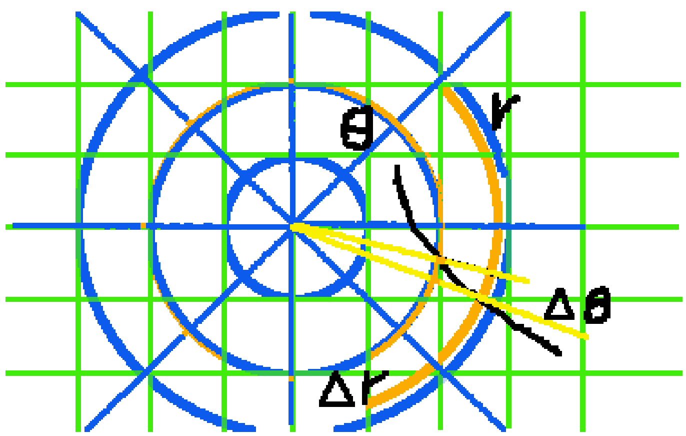

Figure 2: *Polar coordinates on rectangular*

Lets ask about straight lines. We can again look at a line defined by a curve $r(\lambda)$, $\theta(\lambda)$. Again we take a tiny part of the curve, with Δr and $\Delta\theta$. But now it is clear that the length of the curve is not given by

$$\Delta L^2 = \Delta r^2 + \Delta\theta^2 \tag{25}$$

Clearly the length of the side designated by $\Delta\theta$ has a different length depending on the value of r.For a small enough value of $Delta\theta$ and Δr, the length of the θ side is $r\Delta\theta$. Ie, at large r the length of a small change in $\Delta\theta$ increases.

Thus the correct formula is

$$\Delta L^2 = \Delta r^2 + r^2 \Delta\theta^2 \tag{26}$$

The coefficients multiplying the squares of the small changes in the coordinates are called the metric coefficients, designated by

$$g_{rr} = 1 \tag{27}$$

$$g_{\theta\theta} = r^2 \tag{28}$$

where

$$\Delta L^2 = g_{rr}\Delta r^2 + g_{\theta\theta}\Delta\theta^2 \tag{29}$$

or for the continuous curve

$$L = \int_{\lambda_0}^{\lambda_1} \sqrt{g_{rr}\left(\partial_\lambda r\right)^2 + g_{\theta\theta}\left(\partial_\lambda \theta\right)^2} d\lambda \tag{30}$$

At this point you might wonder why I use two, apparently redundant, subscripts. Why not just g_r and g_θ? This will be clear (I hope) below, when we find that sometimes the coefficients also contain cross terms, and the metric looks a lot like a matrix. For now just accept this a peculiarity (perversion?) of general relativists.

Note that the metric coefficients are in general functions of the coordinates. (In this case only $g_{\theta\theta}$ is.)

Again, we can try to derive equations for a straight line by defining a whole array of curves labelled by ϵ

$$L(\epsilon) = \int_{\lambda_0}^{\lambda_1} \sqrt{\left(\partial_\lambda r(\epsilon,\lambda)\right)^2 + r(\epsilon,\lambda)^2\left(\partial_\lambda \theta(\epsilon,\lambda)\right)^2} \tag{31}$$

Again we can take the derivative with respect to ϵ and do an integration-by-parts so that we only have terms like $\partial_\epsilon r$ and $\partial_\epsilon \theta$ without λ derivatives of those terms.

We get

$$\frac{dL}{d\epsilon} = \int_{\lambda_0}^{\lambda_1} \partial_\epsilon r \left[-\partial_\lambda \left(\frac{1}{\sqrt{(\partial_\lambda r)^2 + r^2 (\partial_\lambda \theta)^2}} \partial_\lambda r \right) + \frac{r}{\sqrt{(\partial_\lambda r)^2 + r^2(\partial_\lambda \theta)^2}} (\partial_\lambda \theta)^2 \right] d\lambda$$
$$+ \int_{\lambda_0}^{\lambda_1} \partial_\epsilon \theta \left[\partial_\lambda \frac{1}{\sqrt{(\partial_\lambda r)^2 + r^2(\partial_\lambda \theta)^2}} (-r^2 \partial_\lambda \theta) \right] d\lambda \quad (32)$$

(where the integrals are all fromλ_0 to λ_1). Again the only way that this could be zero for arbitrary sets of paths labelled by ϵ is if the terms factors multiplying the derivatives with respect to ϵ are zero at each point. I.e.,

$$-\partial_\lambda \left(\frac{1}{\sqrt{(\partial_\lambda r)^2 + r^2(\partial_\lambda \theta)^2}} \partial_\lambda r \right) + \frac{r}{\sqrt{(\partial_\lambda r)^2 + r^2(\partial_\lambda \theta)^2}} (\partial_\lambda \theta)^2 = 0 \quad (33)$$
$$\partial_\lambda \left(\frac{1}{\sqrt{(\partial_\lambda r)^2 + r^2(\partial_\lambda \theta)^2}} (-r^2 \partial_\lambda \theta) \right) = 0 \quad (34)$$

Again, we define the new parameter

$$s = \int_{\lambda_0}^{\lambda} \sqrt{(\partial_\lambda r)^2 + r^2(\partial_\lambda \theta)^2} d\lambda \quad (35)$$

along the solution curve (again chosen to have $\epsilon = 0$) s again is the path length along the solution curve. Writing the equations in terms of s instead of λ we again get

$$\partial_s^2 r - r\,(\partial_s \theta)^2 = 0 \quad (36)$$
$$\partial_s (r^2 \partial_s \theta) = 0 \quad (37)$$

This set of equations is messier than is the one in x,y coordinates but still solvable. There are some tricks which will be useful later on which I might as well introduce here.

Integrating the second equation, we get

$$r^2 \partial_s \theta = C \quad (38)$$

(remembering that both θ and r are functions of s.) C is just an integration constant. It will turn out to be plus or minus the radius of closest approach of the curve to $r = 0$.

At this point we can use a trick, Instead of trying to integrate the second order equation for r, we use the fact that since s is the length, we have

$$ds = \sqrt{(\partial_s r)^2 + r^2(\partial_s\theta)^2}ds \tag{39}$$

or

$$(\partial_s r)^2 + r^2(\partial_s\theta)^2 = 1 \tag{40}$$

Ie, this is another non-linear differential equation but in terms only of the first derivatives. Substituting for the equation for $\partial_s\theta$ we get

$$(\partial_s r)^2 + \frac{C^2}{r^2} = 1 \tag{41}$$

or

$$\int \frac{dr}{\sqrt{1-\frac{C^2}{r^2}}} = \int ds \tag{42}$$

or defining $z = r/C$

$$C\int_{r_0/C}^{r/C} \frac{zdz}{\sqrt{z^2-1}} = s \tag{43}$$

$$\pm\sqrt{r^2 - C^2} = s - s_C \tag{44}$$

where s_C is an integration constant.

If $s - s_C$ goes through 0, the sign selected for the left hand side must change. Note that r cannot be less than $|C|$ or the square root would be imaginary. Furthermore, at $s = s_C$, r attains that smallest value of $|C|$. It is however not necessary that $s = s_C$ lies on the path from the intial to final point.

To determine the various constants of integration, we must also solve the equation for θ. We have

$$\frac{d\theta}{ds} = \frac{C}{r^2} = \frac{C}{((s-s_C)^2 + C^2)} \tag{45}$$

from which we have

$$(\theta - \theta_C) = \arctan(\frac{s-s_C}{C}) \tag{46}$$

As s increases, θ increases if $C > 0$ and decreases for $C < 0$.

Setting $\delta\theta = (\theta_2 - \theta_1)|_{mod 2\pi}$ where we take $\delta\theta$ as lying between $-\pi$ and π, we see that C must have the same sign as $\delta\theta$ (because $s_1 < s_2$). ($|_{mod2\pi}$ means that we add or subtract integer multiples of 2π until the answer lies in the specified range). θ_C is the angle of closest approach to $r = 0$ of the straight line (ie, it is the value of θ when $r = |C|$ along the straight line.)

We thus have the following equations to determine the constants s_c, s_1, and θ_C.

$$r_0^2 - C^2 = s_C^2 \tag{47}$$

$$r_1^2 - C^2 = (s_1 - s_C)^2 \tag{48}$$

$$\theta_0 - \theta_C = \arctan(-\frac{s_C}{C}) \tag{49}$$

$$\theta_1 - \theta_C = \arctan(\frac{s1 - s_C}{C}) \tag{50}$$

These can be simplified if we realise that the equations for r and θ in terms of s can be simplied to

$$r\cos(\theta - \theta_C) = C \tag{51}$$

to give

$$\tan(\theta_C) = \frac{r_1 - r_0}{r_0 \tan(\theta_0) - r_1 \tan(\theta_1)} \tag{52}$$

Having determined θ_C we can determine C, s_C and s_1

Infinite straight lines

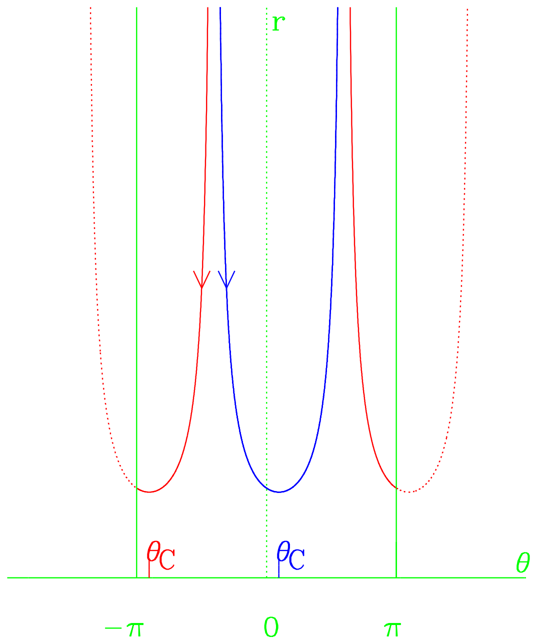

Figure 3: *The graph of an infinite straight line in polar coordinates. The blue and red ones are parallel lines going on opposite sides of r=0 with the same value of $|C|$. The arrows indicate the direction of travel (increasing s) along the lines.*

From the equation for $\theta_0 - \theta_C$ we see that for an infinite straight line, s_C goes to infinity (since s_0, the value at the intial point, is assumed to be 0). There is an infinite path length from the initial point at infinity to the point of closest approach to r=0. Similarly $s_1 - s_C$ goes to infinity. Thus $\theta_0 - \theta_C$ goes to $-sign(C)pi/2$ and $\theta_1 - \theta_C$ goes to $sign(C)\pi/2$. Since r is always positive, $\cos(\theta - \theta_C)$ is the same sign as C Ie, for positive C the curve goes from $\theta - \theta_C = -\pi/2$ through $\theta = 0$ to $\theta - \theta_C = \pi/2$. For negative C, the curve goes from $\theta - \theta_C = \pi/2$ up through $\theta - \theta_C = \pi$ to $\theta - \theta_C = 3\pi/2 \equiv -\pi/2$ The figure shows two paths coming in from infinity at $\theta \approx -\pi + .3$ going the $r = C = \pm 1.5$ at $\theta_C = .3$ and $\theta_C = -\pi + .3$ and then going to infinity again at $\theta = \pi/2 + .3$. In one case, the blue curve, the lines goes above the point $r = 0$ for positive θ, while in the other it goes below.

Conical deficit

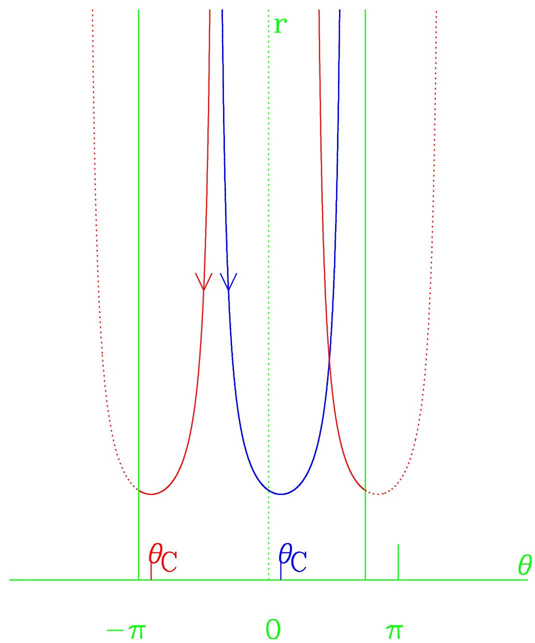

Figure 4: *The two lines of the previous figure but now in a space where the identification is $\theta + 2\pi - \epsilon \equiv \theta$. Again the two straight lines start out as parallel near $\theta = 80^o$ but on passing past r=0, they then cross. two straight lines going on opposite sides of r=0 always cross in space.*

Since the equations for r and θ do not care about the fact that $\theta + 2\pi$ is equivalent to θ, we can ask if we can make different identifications. Eg, what would happen if we set $\theta + 2\pi - \epsilon$ to be the same as θ instead? Straight lines expressed as functions of r and θ would be the same. However, now any two lines which started off as parallel but with opposite values of C would eventually intersect. Ie, consider two lines, the first with $\theta_C = \pi/2$ and $C = 1$ and the second with $\theta_C = -\pi/2$ and $C = -1$ Near $\theta = 0$ these two lines are parallel to each other. However, while one goes to $\theta = \pi$, the other goes to $\theta = -\pi$. But $\theta = -\pi$ is equivalent to the angle $\theta = \pi - \epsilon$.

This is smaller than π, and the two lines must cross. There exist no parallel lines if one runs on one side of r=0 and the other on the other side. $r = 0$ is a singular point.

The point $r = 0$ exists, but clearly behaves very very strangely.

And there is no way to tell from the coordinates that there is anything strange at $r = 0$.

Skew Coordinates

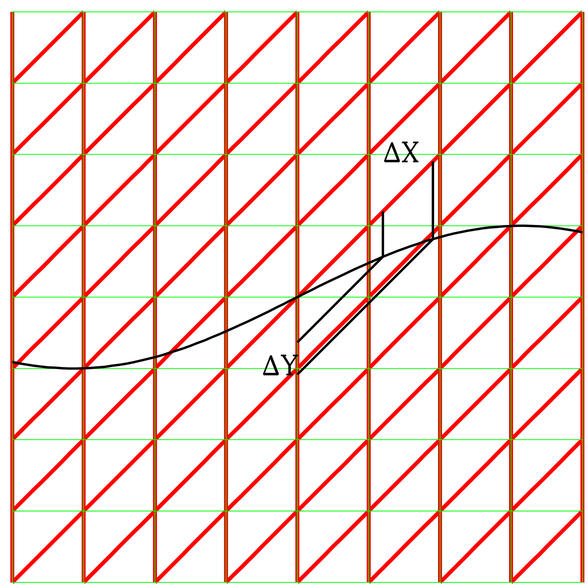

Figure 5: *Skew coordinates– green are original xy coordinates, while red are XY coordinates. Note the changes in coordinates on a little piece of the curve are not orthogonal*

Let us now, instead of using polar coordinates, use a set of skew coordinates. In this case we take the coordinates $X,\ Y$ such that

$$X = x \tag{53}$$

$$Y = x - y \tag{54}$$

This set of coordinates is not perpendicular to each other. The Y axis runs at 45 degrees to the X axis. This would clearly alter the Pythagorean theorem if we try to express ΔL in terms of ΔX and ΔY. One way we can do it, is to express Δx and δy in terms of ΔX and ΔY. Ie, if the line segment has coordinate changes between the ends of $\Delta X,\ \Delta Y$, then $\Delta x = DeltaX$ and $\Delta y = \Delta X - \Delta Y$. Ie,

$$\Delta L^2 = \Delta x^2 + \Delta y^2 = 2\Delta X^2 + \Delta Y^2 - 2\Delta X \Delta Y \tag{55}$$

Ie, we note that for these coordinates, the Pythagorean theorem requires not just the squares of the changes in the coordinates, but also cross terms between the coordinates. We call

$$g_{XX} = 2 \tag{56}$$

$$g_{YY} = 1 \tag{57}$$

$$g_{XY} = g_{YX} = -1 \tag{58}$$

Instead of defining just one cross term in the metric, we have defined two, g_{XY} and g_{YX} which are moreover equal to each other. This is so that we can write the metric as

$$\Delta L^2 = \sum_{i=X,Y} \sum_{j=X,Y} g_{ij} \Delta x^i \Delta x^j \tag{59}$$

where $x^X \equiv X$ and $x^Y \equiv Y$.

If we calculate the straight lines in this coordinate system, we get exactly the same equations as in the first case

$$\frac{d^2X}{ds^2} = 0 \tag{60}$$

$$\frac{d^2Y}{ds^2} = 0 \tag{61}$$

with the same solutions.

$$X = \alpha s + X_1 \tag{62}$$

$$Y = \beta s + Y_1 \tag{63}$$

$$\tag{64}$$

However in this case, we have

$$2(\frac{dX}{ds})^2 + (\frac{dY}{ds})^2 - 2\frac{dX}{ds}\frac{dY}{ds} = 1 \tag{65}$$

or

$$2\alpha^2 + \beta^2 - 2\alpha\beta = 1 \tag{66}$$