PROFESSOR: Hi. Recall that in our previous two lectures, we have been discussing the concepts of velocity and acceleration of particles moving along a curve. And we had pointed out originally that the most natural definition was in terms of say, i and j, or i, j, and k components-- in other words, Cartesian coordinates. And then we showed in the last lecture that tangential and normal components form a rather important system of vectors. And now, what we would like to do is to motivate the use of polar coordinate types of vectors.

And again, it's interesting point out that we could have introduced polar coordinates into our course at any one of a number of times. In fact, one of the problems with the traditional mathematics curriculum was that, frequently, topics were put in just for the sake of filling a chapter. They fit into the space. They didn't need any additional prerequisites, so we put them in there. And many a typical traditional high school algebra book could have had the chapters shuffled with no loss of continuity.

What we're trying to do in this particular course is to motivate why particular concepts would have been invented if they had not already have been invented previously. In particular, in dealing with particles moving in a plane or in space, there comes a time when one considers the very important physical application known as a central force field.

Now, in a central force field-- and by the way, this is where the word polar coordinates comes up very naturally, and why we call today's lecture "Polar Coordinates." See, in a central force field, what we mean is that the only force acting happens to be located at a fixed position. And wherever that particle is, the force is always directed along the line, either towards or away, depending upon whether it's attraction or repulsion. It's always acting this way.

If we're using Newtonian mechanics, since the force is proportional to the acceleration, it says what? That a very natural direction for acceleration would be along the direction that connects the origin-- meaning the location of the central force-- to the particle. Well, you see, once this is done, it becomes a very natural thing to say, OK, if you're going to be talking about central force fields, why not introduce new variables?

After all, it seems that the important thing now in measuring the central force is to know how far the particle is from the force and also, I suppose, to locate the position of the particle. Once you know how far away it is, you would like to know what angle the radius vector makes with the positive x-axis. And in this sense, that would be all that we would need to motivate polar coordinates.

Now, you see, once the motivation is there, we can always study the subject independently of whether the motivation had ever been given or not. In other words, now, we say, look, it's important to study polar coordinates. Why is it important? Because you want to study central force fields. But the structure of polar coordinates is going to be the same, whether we study central force fields or not.

Then, you see, once we finish our study of polar coordinates, then we say, OK, now, let's go back as a particular application to a central force field. At any rate, without further ado, let's tackle the subject of polar coordinates. Again, many of you may be very, very familiar with polar coordinates, some of you a bit rusty. To play it safe, I have divided the study guide in such a way that there will be two units devoted to polar coordinates, but that this one lecture will cover both of those two units, that this lecture, as usual, will be the overview.

And I guess this is the best you can do with mathematics beyond a point. Beyond a certain point, mathematics ceases to be a spectator sport, and you just have to dirty your hands with the computations, and that the basic theory, as simple as it may seem, is straightforward. It's just a case of picking up the computational know how together with a familiarity so that you feel at ease with the concepts. So let's just quickly go through the highlights of polar coordinates.

They start off in a very deceptively simple way. Namely, given a point P in the plane, if we elect to use Cartesian coordinates, the point P can be labeled as (x, y). On the other hand, using polar coordinates-- referring to this diagram, at least-- one would tend, as we mentioned before, to label this point (r, theta), where r is the distance of the point from the origin and theta, the angle that we're talking about over here.

Notice, by the way, that there is a very simple set of formulae by which we can switch from one pair of representations to the other. For example, to express x and y in terms of r and theta, one simply has that x equals r * cos(theta), y equals r * sin(theta). OK. And conversely, to express r and theta in terms of x and y, one has that r squared equals x squared plus y squared and the tan theta is y/x.

By the way, you might wonder why I didn't write r equals the square root of x squared plus y squared. The reason I wrote this is that we're going to get into some complications very shortly, complications which I shall skim over. And if you see how long-winded I am, skimming over is not going to be that short, but short relative to the treatment that we're going to give it in the exercises.

There are certain complications that set in, because r can be negative. By the same token, somebody says, instead of writing tan theta equals y/x, why couldn't you say that theta was the inverse tangent y/x-- use the inverse trig functions? Remember that the inverse trig functions require a principal value domain for the tangent that's what? Between minus pi/2 and pi/2.

And notice that the angle theta certainly is not restricted to that range. Theta could be in the second quadrant, the third quadrant. It can wind around. In other words, one of the luxuries about Cartesian coordinates was that they were terribly formally stilted. There were no two ways to represent the same point.

In the study of Cartesian coordinates, one can say, to a point, position is everything in life, that once you've changed the coordinates, you've changed the point. For example, there's no possible way for the point (x, y) to equal the point, say, (2, 3), unless x equals 2 or y equals 3. In other words, for Cartesian coordinates, the only way to points could be equal would be if they were equal coordinate by coordinate.

The complication that sets in for polar coordinates is that, unfortunately, this simple result no longer remains true. For example, if I tell you that the point (r1, theta1) is equal to the point (r2, theta2)-- that's a slash through the arrow here. This means it's false. It does not imply that I can conclude that r1 equals r2 and that theta1 equals theta2. To be sure, it's possible that r1 equals r2 and it's possible that theta 1 equals theta 2, but it doesn't have to happen.

Well, the nice thing about mathematics is that we can always give examples to back up what we're saying. Let's look at a few examples. Let's look, for example, at what happens, if we're using radian measure, if we replace the angle by some integral multiple of 2 pi added on to that angle. In other words, let's take the point (r1, theta1) and compare that point with the point whose coordinates are (r1, theta1 + 2k pi).

Look, the r values are the same here. But look at theta1 and theta1 + 2k pi. Notice that, unless k is 0, these are different angles. It's rather interesting. I know in high school this happens a lot. Youngsters will tend to say things like 30 degrees is the same as 390 degrees.

And in a way, they're right, except that they say it in such a way that it's dangerous. Certainly, one should not confuse 30 degrees with 390 degrees. What one usually means is that any trigonometric function of 30 degrees is equal to that same trigonometric function of 390 degrees. In other words, the technical way of saying it is that the trigonometric functions are periodic with period 2pi in radian measure, 360 degrees in degree measure.

In other words, for example, using radian measure, notice that the sine of theta1 is equal to the sine of theta1 + 2 pi k, where k is some integer here. But that theta1 is unequal to theta1 + 2 pi k if k is unequal to 0. If k is equal to 0, of course, this happens to be a special case.

You see, the whole point is that all it means in terms of a graph, if you were to plot the curve y equals sin(x), it is possible for what? Many different values of theta, many different values of x, to give the same value of sine x. What this means, by the way, in terms of polar coordinates, is that it certainly makes a difference whether we talk about a line making an angle of 30 degrees with the positive x-axis or making 390 degrees, because, you see, that 390 degrees seems to indicate that you've made one full circuit and then an additional 30 degrees. As far as position is concerned, you're in the same place. But for example, in terms of fuel consumption, it uses more fuel, say, to go through the 390 degrees then to go through the 30 degrees.

But at any rate, that's not the point we want to make here. The point is notice in this particular example that this is two different names for the same point. But we cannot conclude that, coordinate by coordinate, the coordinates are equal.

That's not the worst of it. The worst of it is that r and/or theta don't even have to be positive. Another way of saying that is they may be negative. For example, to capitalize on the idea of vector notation, what one frequently does is says, look, let's talk about a negative distance. And by a negative distance, we'll identify that with sense.

In other words, let's call this angle here theta. See, notice, what you're really saying here is you're assuming that the radius vector, when you're calling this angle theta, has this particular sense. Suppose you wanted to talk about this vector. Notice that this would be the right sense if the angle here happened to be what? Theta + pi-- 180 degrees, pi radians.

Now, what you say is if you're looking at this particular point, notice that with respect to this vector, you must go what? Negative r units-- in other words, you must move in the opposite sense of this particular vector to get to this point. And again, this is a touchy enough concept that we're going to do this in great detail in the learning exercises. But the point is notice, that (r, theta) is a different name for the same point as that which is named by (minus r, theta + pi).

Do you see that? Let's look at that one more time just to make sure. See, on the one hand, reading this angle and this distance, this is (r, theta). On the other hand, reading it from this angle, the value is minus r and the angle is theta + pi. And I'll get a different piece of chalk in a minute, because this doesn't seem to be writing too well. But let's not worry about that for the time being.

Let me give you an example. Look at the curve whose polar equation is r equals sine squared theta. Notice that if theta equals pi/6, the sine of pi/6 is 1/2. 1/2 squared is 1/4. So notice that r equals 1/4, theta equals pi/6 satisfies this particular equation.

Now, notice that another name for the same point using this recipe here is (minus 1/4, 7 pi/6). But without even looking, you should be able to see that it's impossible for this value of r and this value of theta to satisfy this equation. In particular, notice that sine squared theta can't be negative. Just by reading this equation, r can never be negative, therefore, how can r equals minus 1/4 possibly satisfy this equation?

And this is where the big complication comes in in terms of simultaneous equations and what have you. Here, we have what? Two different names for the same point. Yet, by one of its names, the point satisfies the equation, and by another name, it doesn't satisfy the equation.

And this is a very, very touchy thing, because you'd like to believe what? That a point belongs to a curve, not the name of a point. And that one of the difficulties with polar coordinates is that, to check whether a point belongs to a curve or not, if the point, as named by one way, doesn't satisfy the equation, you still have to check other possible names to see whether they satisfy the equation or not. And this is a very complicated topic. And this is why some of our learning exercises in this unit take so many pages. It's more or less devoted to giving you insight into this if you don't already have this insight.

But at any rate, let's continue on in general terms here. The next most important thing to talk about is to make sure that we understand what polar coordinates really mean. Namely, when someone says, I am thinking of the curve C whose polar equation is r equals sin(theta), he does not mean plot r versus theta this way and call that the curve C. You see, notice that even though you're calling this r and theta, using the axes to be at right angles this way, no matter how you slice it, you're still using Cartesian coordinates here, only what theta replacing the name of x and r replacing the name y.

See, this is not what's meant. Remember that when you're talking about polar coordinates, r specifically measures what? The distance of the particular point on C from the origin, and theta measures the angle that that radius vector makes with the x-axis. For example, what we do we mean by the curve whose pole equation is r equals sin(theta) is the circle of radius 1/2 centered on the y-axis, you see, at the point what? In Cartesian coordinates, x is 0, y is 1/2.

And I claim that that's what we mean by the curve r equals sin(theta). And how can I prove that to you? Well, the easiest way to prove that to you is by some elementary geometry. Namely, I call this point (r, theta). I now draw in these guidelines. Notice that, by definition, this length is r. Since this is an inscribed angle on a diameter, it's a 90 degree angle. Therefore, since this angle is theta and both of these angles are complements of a 90 degree angle, this angle must also be theta.

And now, if I just look at this particular diagram, I read this right triangle, and immediately, I see what? That from this right triangle, r equals sin(theta). In other words, the curve C, whose polar equation is r equals sin(theta), is this particular curve.

By the way, we often get indoctrinated into seeing things in a very natural way in terms of a native tongue. My father-in-law was brought up in Russia. He had a little grocery store. He spoke fluent English, but when he added up customers' bills, he always added in Russian. He was more at home with Russia numerals.

And this always impressed me, until I realized that I was the same way, only I use base 10 numerals instead of base 7 or something like this. And the same thing happens with the polar versus Cartesian coordinates. Even though polar coordinates are independent of Cartesian coordinates, the fact remains that we're probably more at home with Cartesian coordinates than we are with polar coordinates.

Consequently, one very common trick, when we can get away with it, is to take a polar equation and translate it into an equivalent Cartesian equation. Namely, given r equals sin(theta) and remembering that r squared is x squared plus y squared and that r sin(theta) is y, we multiply both sides of this equation by r to get r squared equals r sin(theta). This leads to x squared plus y squared equals y. If we then complete the square, et cetera, we find that, in Cartesian coordinates, this is the equation of the circle centered at the point (0, 1/2) with radius equal to 1/2,

The thing that's very, very important to note here is that, in terms of x and y, the relationship x squared plus y squared equals y is not structurally the same as the relationship r equals sin(theta). In other words, the curve C is the same whether you use its Cartesian form or whether you use its polar form. But what's very important to note is that the relationship between r and theta is not algebraically the same as the relationship between x and y.

But the important point is that, since the curve C does not measure r versus theta as being at right angles to each other, that if you were given r equals sin(theta) and you compute dr/d(theta), which, in this case, simply would be what? cos(theta)-- that does not have the slope of C.

You see, this is what I want you to understand. If I compute dr/d(theta) and one would interpret that as a slope, it's not a slope of the curve C. It's a slope of this particular curve which wasn't C. In other words, this is the curve that measures r versus theta in terms of Cartesian coordinates. That dr/d(theta) just shows you how r is changing with respect to theta, it does not tell you how the curve is rising at that point.

And again, more information is given on this in terms of exercises. But that's what I do want you to understand. It does not mean that you can't compute the slope of this curve at any point. It does mean that if you want the slope and you compute it by letting it equal dr/d(theta), you will get an answer this way, but it won't be the slope that you're looking for.

In fact, what I hope that you can see rather simply here is the following. Say I want the slope of the tangent line to the curve at this particular point, it doesn't really make much difference whether I use the angle phi-- which the tangent line makes with the positive x-axis-- or whether I use phi or "phi," or, I don't know-- This angle is called either "phi" or "phi". And this angle is called "psi" or "psi." I get all mixed up with the Greek letters.

You know, in fact, coming into work today, I was listening to a disc jockey, and he was making fun of how in drugstores now, you can get a full course meal, but you can't get any medicines anymore. And he was talking about his friend who got all A's in pharmacy school, and they flunked him out because he didn't know how to make a sandwich. I almost flunked out of math because I have trouble with the Greek alphabet here.

But look, forget about that. The thing I want to see is I want to know the direction of this line. If it's convenient to use phi, I'll use phi. If it's convenient to use psi, I'll use psi. What is phi? Phi is the angle that the tangent line makes with the positive x-axis. What is psi? Psi is the angle that the tangent line makes with the radius vector.

Now, as I'll show you in some of our exercises, to use the definition that the slope is tan(phi) and translating that from Cartesian coordinates into polar coordinates is a very, very messy job. What turns out to be very, very interesting, at least from my point of view, is that not only is the angle psi-- "psi"-- more natural to use than phi when you're dealing with polar coordinates, but it almost turns out that because it was more natural, there's a simpler formula for it.

In other words, whereas the formula for tan(phi) is very, very complicated in polar coordinates, the formula for tan(psi) is very simply given by r divided by dr/d(theta). In other words, if I take r, divide that by dr/d(theta), what I get is the tangent of this angle. Now, look. Once I have the tangent of this angle, it's very simple to construct a tangent line. That's what I want you to see.

Look. Suppose this is the point P, and this is my pole O, for polar coordinates. And given that r was some function of theta, suppose now I've computed dr/d(theta), I've divided that into r, and I've now found what tan(psi) is. What I do is I take the line of action that joins the origin to P, I construct the angle psi, draw this line, and whatever that line is, that is the line which is tangent to my particular curve at the point P.

But the thing I want to see is that we do not need Cartesian coordinates at all in order to be able to tackle calculus properties when an equation is written in polar coordinates. But what there is a tendency to do is what? That when we're dealing with polar coordinates, since we're more at home with Cartesian coordinates, we will often be tempted to switch everything into Cartesian coordinates, which is perfectly fair game. But the important thing is to remember that all of our calculus results can be derived independently of whether Cartesian coordinates were ever invented.

For example, the final concept I want to talk about in this lecture is how one would have studied area if one had only polar coordinates and had never had Cartesian coordinates. By the way, notice also that, in terms of polar coordinates, one is more interested in sectors than in rectangular grids. In other words, if you look at something ranging from theta1 to theta2, you think of something caught between two rays here.

And let's suppose I wanted the area of this particular region, where this particular curve happened to have the form, say, r equals g(theta)-- I don't care what it is. Let's just call it r equals g(theta). Now, the interesting point, again-- and I keep saying "interesting point" because they are interesting points-- that if you have really taken Part 1 of this course seriously, you're going to be amazed to see how much free mileage you get out of Part 2 just by translating things back in to our general theorems of Part 1.

Let me see how I could find the area of this particular region. The idea is I'll take a little increment of area here between these two black lines. Remember, my basic building blocks are now r values. What I'll do now is I'll pick the smallest value of r that gives me an arc that lies inside this segment. then I'll pick the biggest value of r that encloses this segment.

Notice, what I do is I'll let [capital R sub capital M] denote radius. See, that's this distance from here to here? I'll let [little r sub little m] represent this distance. What I'm saying is, if I now swing two arcs, my area theorems from the first semester-- my area axioms from the first part of course-- are still valid. Namely, the smaller r-- the smaller sector-- is contained inside the delta A that I'm looking for. And the larger sector contains the region I'm looking for. In other words, the region that I'm looking for, delta A, is caught between the areas of these two sectors.

Now, how do you find the area of a sector? What you do is you take the area of the entire circle and divide it by the fractional part of the circle that you're taking. For example, the area of the circle whose radius is [capital R sub M] is pi [R sub M] squared.

Now, what portion of the circle am I taking? The angle is delta theta. I'm using radian measure, so the entire swinging angle would have been 2 pi, so I'm taking delta theta over 2 pi of the entire circle. So that's the area of the larger sector.

What is the area of the smaller sector? The area of the smaller sector is the area of the entire circle pi [r sub little m] squared times, again, what? Delta theta over 2 pi. And because the region is embedded between these two, its area is caught between these two areas. And if I now solve for the delta A, I simplify, I cancel the pi's, I get that delta A is caught between what? 1/2 [capital R sub M] squared delta theta and 1/2 [little r sub m] squared delta theta.

I now divide through by delta theta. And I make a non-crucial assumption here that delta theta is greater than 0, remembering that delta theta is negative. All I have to do is reversed the signs of the inequalities.

The only reason I assumed that delta theta was positive is so that I wouldn't change the direction of the inequalities. A similar demonstration will hold when delta theta is negative. The important point is I then divide through by delta theta. I get delta A over delta theta. It's caught between 1/2 [little r sub m] squared and 1/2 [capitalized

R sub M] squared.

Now, as I let delta theta approach 0, what does that mean? I'm going to let delta theta close in over here-- approach 0. What's happening here? This is a fixed value of r. Notice that [little r sub m] and [capital R sub M], as delta theta approaches 0, they're both being pushed closer and closer to r. In other words, as delta theta approaches 0, [r sub little m] and [capital R sub M] both approach r, provided that r is a continuous function of theta-- in other words, that there are no breaks in the curve here. OK?

The important point, then, is what? We can then take the limits as delta theta approaches 0. And we find that dA/d(theta) is the limit as delta theta approaches 0-- delta A divided by delta theta, which simply is what now? We come back here. As delta theta approaches 0, [little r sub m] and [capital R sub M] both approach r, and therefore, this common limit becomes 1/2 r^2, therefore integrating this between what limits?

Between theta1 and theta2, I find that the area of my region is the integral from theta1 to theta2, 1/2 r^2 d(theta), which, of course, means what? This integral 1/2-- r is g(theta), so I square that times d(theta). And notice, that this is a function of theta alone. And so I can find this particular area.

What's crucial to understand is that that area does not change just because you're dealing with polar coordinates rather than Cartesian coordinates. But rather the form of the equation changes. What I mean by that is something like this.

Let's suppose I have a region like this. This region is inanimate. It doesn't know whether you're looking at in polar coordinates or in Cartesian coordinates. In Cartesian coordinates, it's bounded above by the curve y equals f1(x), and it's bounded below by the curve y equals f2(x). From what we studied in the first semester, the area of the region r in Cartesian coordinates is the integral from 0 to A, f1(x) minus f2(x) dx.

On the other hand, if I call this curve r equals g(theta), in terms of polar coordinates, and the initial angle is theta1, and the terminal angle here is theta2, then the area of that same region r is given to be 1/2 integral from theta1 to theta2, g(theta)^2 d(theta).

Mathematically, this integral in terms of x and this integral in terms of theta look completely different. But the crucial point is they are simply different expressions for the same answer. Which of the two is the better one to use? It depends on the particular problem. If, for example, a problem begs for a polar coordinates interpretation, use polar coordinates. In fact, polar coordinates and Cartesian coordinates can both be bad sets of equations. Who knows? It's not the point.

The point is that we now have two coordinate systems. There are others that we will define as the course goes on. But the important point is that we are now ready to tackle motion in the plane for central force fields if we so desire.

At this particular moment, I do so desire. So the chances are that next time, we will be talking about velocity and acceleration vectors in polar coordinates. At any rate, until next time, good bye.

Funding for the publication of this video was provided by the Gabriella and Paul Rosenbaum Foundation. Help OCW continue to provide free and open access to MIT courses by making a donation at ocw.mit.edu/donate.