

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

HERBERT

GROSS:

Hi. As I was standing here wondering how to begin today's lesson, an old story came to mind of the professor who passed out an examination to his class, and one of the students said, "Professor, this is the same test you gave us last week". And the professor said, "I know, but this time I changed the answers." And I was thinking of this in terms of the fact that much of the new mathematics is essentially the old mathematics with some of the answers changed.

One of the topics that we used to belittle in the traditional curriculum because it was too easy, was the topic called linear equations. And it turns out that in the study of several variables in particular-- but it was already present in calculus of a single variable-- we very strongly used the concept of linearity.

I could've called today's lesson something old, something new. Meaning that the old topic that we were going to revisit would be that of linear functions, and the new topic would be how it manifests into the modern curriculum in the sense that one introduces a subject called linear algebra, or matrix algebra, as a standard portion of a modern calculus course whereas in the traditional calculus courses, essentially nothing was ever said about matrix algebra or linearity.

Instead I picked a more conservative title for today's lesson, I simply call it "Linearity Revisited". And as I say, it goes back to when we were in junior high school or high school, when we were taught that linear functions were very nice. For example, given the equation y equals mx plus b -- the linear equation meaning what? It graphs as a straight line, but that the two variables are related linearly, y is a constant multiple of x , plus a constant.

We were told solve for x in terms of y . And what we found was that if y equals mx

plus b , this was true-- if and only if-- x was equal to y minus b over m . What we showed was given a value of x that corresponded a value of y , and conversely given a value for y that corresponded a unique value for x .

And to put this into the language of functions, what we were saying was that if $f(x)$ equals mx plus b , then f inverse exists. In other words, what we're saying is that no two different x values can give you the same y value if the function has the form y equals mx plus b . And just about the time that we were learning to enjoy this kind of an equation our dream world was shattered, and we were told it's too bad, but most functions aren't linear.

We were given things like y equals x to the seventh plus x to the fifth, and we found that we couldn't solve for x very conveniently in terms of y . And that's what began our intermediate algebra and advanced algebra courses. In other words, the fact that most functions are non-linear. Now an interesting thing occurred though. Let me just emphasize this. And this is the key point.

In terms of calculus, we discovered-- and here's a key word coming up-- Most functions are locally linear. Now that sounds a little bit like a tongue twister, but actually back in the first part of course when we talked about Δy sub \tan -- a change in y to the tangent line. Notice what we were saying.

We were saying that to study $f(x)$ near f equals a , we saw that $f(a) + \Delta x$ minus $f(a)$ was equal to f' of a times Δx plus $k \Delta x$ where the limit of k -- if Δx went to 0-- was 0 itself. Provided of course that f was differentiable at x equals a otherwise you couldn't write down f' of a here.

The interesting point is this. But if you look just at this term over here, this expresses Δf as a linear function of Δx . The part that makes this thing non-linear is the term called $k \Delta x$. But that's the term that's going to 0 as a second order infinitesimal. So what we're really saying is this-- That provided that f is differentiable at x equals a -- In other words locally we mean this-- near x equals a we can say that Δf is approximately f' of a times Δx .

That's what we call Δf sub tan, recall. And what we mean by approximately here is that error $k \Delta x$ goes to 0 very, very rapidly as Δx goes to 0. And what we mean by locally is this-- suppose f' exists also when x equals b . We can again compute Δf near x equals b Now Δf is equal to what? Approximately f' of b times Δx plus that error term which goes to 0 very rapidly.

We again call this thing here Δf tan, but the thing to keep in mind is since f' of a need not equal f' of b , Δf tan is different at a and at b . In other words, even though it's always true where f is differentiable, that we can say that Δf is approximately Δf tan the value of Δf tan depends on the value of x that we're near. And that's what we mean by saying that approximating Δf by Δf tan is a local property.

Now I think that sometimes by putting these things into words it sounds harder than it really is. So I think what might be nice is if we just look at a specific illustration, a problem which I deliberately picked to be as simple a nonlinear example as I can think of.

Let me come back to our old friend, the function $f(x)$ equals x squared, which as I say, is about as simple a non-linear function we can get into. Now we know that $f(x)$ equals x squared plots as the curve y equals x squared, the parabola. Let's take a couple of points on this parabola. Let's say the point $(1, 1)$, and the point $(2, 4)$, draw in the tangent lines to the curve at these two points. And we know what? That the equation of the tangent line to the curve at $(1, 1)$ is $y - 1$ over $x - 1$ equals the slope. Since y is equal to x squared the slope is $2x$, when x is 1 the slope is 2. So the equation of this tangent line is given by $y - 1$ over $x - 1$ equals 2.

At the point corresponding to x equals 2, $2x$ is 4, so the equation of the tangent line here is $y - 4$ over $x - 2$ equals 4. So now I've induced three functions that I can talk about. My original function, $f(x)$ is x squared. This straight line is the linear function just solving for y in terms of x . $g(x)$ equals $4x - 4$. And this straight line corresponds to the function $h(x)$ equals $2x - 1$.

Now the interesting point, of course, is that these two functions here are linear. They are completely different functions. Notice not only pictorially are they different, but algebraically their slopes are different, and their y-intercepts are different, and back in our course in part one, we talked about things geometrically saying look at near the point of tangency, the tangent line serves as a good approximation to the curve itself.

What were we really saying then? What we were saying was that near the point of tangency, $g(x)$, which was a linear function, could replace $f(x)$ which was a nonlinear function.

Of course when we moved too far away from a given point then when we said that $f(x)$ still had a linear approximation, we had to pick a different linear function. By the way, again because we were dealing with one independent variable and one dependent variable, it was very easy to invent the concept of a graph. As we shall show a little while, the concept of linearity extends to several variables, but you can't draw the graph as nicely.

So let me now revisit the same result here, only without reference to the graph. What we're saying is that our function is mapping the real number line into the real number line. In other words instead of putting x and y at right angles to each other, let's put x and y horizontally parallel to one another. And what we're saying is that f maps the interval from 0 to 2 onto the interval from 0 to 4.

Now what does h do? Remember h is the function $2x$ minus 1. h maps the interval from 0 to 2 onto the interval from minus 1, 2, 3. And you see this is all this diagram means f maps 0 into 0, it maps 1 into 1, it maps 2 into 4. f is the function which squares the input to yield the output. And correspondingly, h maps 0 into minus 1, it maps 1 into 1, and it maps 2 into 3.

Now the interesting point is that f and h are very different. In fact, the only time f and h have the same output is when x equals 1. Which of course we move from before because how was $h(x)$ constructed? $h(x)$ was constructed to be the line tangent to the parabola y equals x squared at the point x equals 1 y equals 1. So that should

be no great surprise.

But if we didn't know that notice it algebraically we could equate $f(x)$ to $h(x)$ conclude therefore, that means x squared must equal $2x$ minus 1 . We then transpose, and get that x minus 1 squared must be 0 whence x must equal 1 . And what we have is that near x equals 1 x squared behaves like-- and I put this in quotation marks because that's the hardest part of the course that's going to follow was what you mean by behaves like-- but x squared behaves like $2x$ minus 1 .

And what we mean by that is this-- at least in terms of a picture. If I pick a small interval surrounding x equals 1 on the x -axis, and a small interval-- like a thick dot-- surrounding y equals 1 on the y -axis here. Then as a mapping from this domain into this range, I can essentially not distinguish f from h .

The error is so small that as the size of the interval shrinks, the error goes to 0 even faster. And therefore, if I stay close enough locally to the point in question-- I cannot tell the difference between the non-linear function and the linear function.

But what I have to be careful about is this-- that what whereas x squared can be replaced by $2x$ minus 1 , near x equals 1 , near x equals 2 , x squared can be replaced again by a linear function. Namely $4x$ minus 4 . But $4x$ minus 4 is not approximately the same as $2x$ minus 1 , no matter where you look.

You might say well look it, don't these two straight lines intersect at the particular point? The answer is yes they do. But even at the point that they intersect, there was no neighborhood in which these lines can serve as approximations for one another.

Those are two straight lines that intersect at a constant angle, and as soon as you leave the point of intersection there is a significant arrow. Meaning an arrow which does not go to 0 more rapidly than the change in x . You don't have that higher or the infinitesimal over here. At any rate, leaving this to the exercises and the supplementary notes for you to get more out of, in summary, let's just say this-- If f is continuously differentiable at x equals a , then locally-- meaning near x equals a -- f

behaviors linearly.

$f(x)$ is approximately $f(a)$ plus f' of a times the quantity x minus a , and you see once x is chosen to be a this is a number, this is a number, Δx here is the only variable on the right hand side.

So what we're saying is that $f(x)$ is a what? Linear function of Δx . And the more interesting point is since this is all review so I say what I mean by interesting point is what? That we don't have to just review this way, we did this simply to refresh your memories as to how linearity was playing a big role in calculus of a single variable.

Now what we're going to do is extend the result to several variables. Let me just say that at the outset. That this concept does extend to n variables, but n equals 2 yields a particularly good geometric insight.

For example, let's suppose I look at two equations and two unknowns. Well actually, I'll use u and v instead. Let those be variables. Also we can think of this as a function. I have $u(x, y)$ is x squared minus y squared, Whereas $v(x, y)$ is $2xy$. Notice that these are not linear, because here we have things appearing to second part, your squares, and here we have what? The variables multiplying one another.

These are not linear equations, but the beautiful point is-- if you look at this way-- is even without a picture, I can think of this as a mapping which maps two dimensional space into two dimensional space. And how does this mapping take place? It maps the point or the pair, or the 2-tuple-- whichever way you want to say it-- (x, y) into the 2-tuple (u, v) , where u is x squared minus y squared, and v is $2xy$.

In other words, \bar{f} -- and notice I put the bar underneath simply to indicate that E^2 is a vector space, and we have a function that's mapping what? A vector into a vector, so I indicate that f is a vector function here. It maps a vector into a vector. And how does the mapping take place? It maps the 2-tuple (x, y) into the 2-tuple x squared minus y squared comma $2xy$ u, v .

Now the thing is that as long as we only have n equals 2, we can still draw a picture, but not a picture as nice as what existed when n was equal to 1. See pictorially, f -

bar maps the xy plane into what we can call the uv plane but notice that since the domain of f -bar has two degrees of freedom-- a two dimensional vector space-- Notice that the domain of f -bar is the entire xy plane. Whereas the range of f -bar is the entire uv plane--

In other words, I can now view f -bar as a mapping which carries points in the xy plane into points in the uv plane, and this will be exploited more later in the course, but the idea is this. Let's take a look for the time being. Let's see what f -bar does to the point $(2, 1)$.

Remember u is x squared minus y squared, so at the point $(2, 1)$, u becomes what? 2 squared minus 1 squared, which is 3 . On the other hand, $2xy$ is 2 times 2 times 1 , which is 4 . So f -bar can be viewed as mapping the point $(2, 1)$ into the point $(3, 4)$.

Now you recall that calculus isn't interested in what's happening at a particular point, it's interested in what's happening in the neighborhood of a particular point. So the major question is how does f -bar behave near the point $(2, 1)$. In other words, what is f -bar $(2 + \delta x, 1 + \delta y)$, when δx and δy are quite small. That's the question that we're raising over here.

What we're saying is we know that $(2, 1)$ maps into $(3, 4)$. We also know or we'd like to believe that a point near $(2, 1)$ maps into a point near $(3, 4)$. Well if we call this point $(2 + \delta x, 1 + \delta y)$, then the corresponding image over here should be $(3 + \delta u, 4 + \delta v)$.

What we can say is that whatever the image of $(2 + \delta x, 1 + \delta y)$ is it has the form $(3 + \delta u, 4 + \delta v)$, and all we have to do is find δu and δv . This is the pictorial idea of what's happening. Now the point is that δu and δv are very difficult to find. After all, u and v are non-linear functions. To invert them is either difficult, or downright impossible-- one of the other-- in many cases.

The thing that's easy to find is δu tan, and δv tan. Remember δu tan was the postulate of u with respect to x times δx , plus the postulate of u with respect to y times δy . Since u is equal to x squared minus y squared, that

means $\Delta u \approx 2x \Delta x - 2y \Delta y$. We're interested in this at the point $(2, 1)$.

Letting x be 2, and y be 1, we see that $\Delta u \approx 4 \Delta x - 2 \Delta y$. Since v is equal to $2xy$, the postulate of v with respect to x is $2y$ the postulate with v with respect to y is $2x$. Therefore, $\Delta v \approx 2y \Delta x + 2x \Delta y$. Since we're evaluating this at $x = 2, y = 1$, we see that $\Delta v \approx 2 \Delta x + 4 \Delta y$.

Now here's the key point. This is always $\Delta u \approx$. This is always $\Delta v \approx$. Well, the local thing comes in is that we know that because u and v are continuously differentiable functions of x and y , that near the point $(2, 1)$, we can replace Δu by $\Delta u \approx$ Δv by $\Delta v \approx$, and we wind up with what? Δu is approximately $4 \Delta x - 2 \Delta y$. Δv is approximately $2 \Delta x + 4 \Delta y$.

But the key point now is that this is a system of linear equations. You see Δu is a linear combination of Δx and Δy , and Δv is also a linear combination of Δx and Δy . In other words, as long as u and v are continuously differentiable functions of x and y , we can approximate locally Δu and Δv by linear approximations.

Notice how linear systems come into play. Now I've been emphasizing the case $n = 2$ just so we could draw a picture. Notice that no matter how many variables we have. Well, in fact, let me just summarize this in terms of x and y first. And then we'll generalize it to n -variables in a minute.

The key point for two variables, and what happens for two variables happens for any number. But as we've often done in this course, we emphasize the two variable case because we can still visualize the picture. Even though the graph idea is hard to see, because we're mapping two dimensions into two dimensions.

But at least the domain and the range are easy to see separately, but if u is a continuously differentiable function of x and y near the point (x_0, y_0) , then Δu is

exactly the postulate of u with respect to x times Δx , plus the postulate of u with respect to y times Δy , plus an error term, $k_1 \Delta x + k_2 \Delta y$, where k_1 and k_2 go to 0, as Δx and Δy go to 0.

If we just look at this spot alone, Δu is linear up to this as a correction term. In other words, the non-linearity part of Δu is going to 0, as a second order infinitesimal, and the reason I keep harping on this point is that no matter how complex the theory gets in the rest of this particular block, the key step is always going to be that when you have a continuously differentiable function you can essentially-- as long you stay locally-- you can essentially throw away the nasty part.

You can essentially throw away this error term, because it goes to 0 so rapidly that if you stay close enough to the point (x_0, y_0) , no harm comes from neglecting this term. What you must be careful about is that as soon as you pick a large enough neighborhood so that this term is no longer negligible, then even though this part here is still Δu sub tan, Δu sub tan is no longer a good approximation for Δu . At any rate, n -variables what we're saying is suppose w is a function of x_1 up to x_n .

Then if w happens to be continuously differentiable at the point corresponding to \bar{x} equals \bar{a} , meaning in terms of n -tuples x_1 up to x_n is the point a_1 comma up to a_n , then what we're saying is that Δw can be replaced by-- now this has been mentioned the text, I don't remember whether we've mentioned this in previous lectures or not. It's rather interesting that when you deal with more than three independent variables we somehow don't like to use the word Δw sub tan. Because tangent indicates a tangent line or a tangent plane which is a geometric concept.

Instead we replace the word tangent by lin as an abbreviation for linear. The key point being what? That this thing that we call Δw sub lin, or if you like to call it sub tan, what's in the name? Call it whatever you want. The point is that this thing that we call Δw sub lin, or Δw sub tan is the partial of f with respect to x_1 evaluated at \bar{a} times Δx_1 , plus the partial of f with respect to x of n evaluate

it at a bar times delta x_n .

And the key point is that once you have chosen a specific number \bar{a} , notice that the coefficients of δx_1 up to δx_n are numbers. They're not variables. They are numbers once \bar{a} is chosen. So that what is δw_{lin} , why we call it linear? Notice that this expression here is a linear combination of δx_1 up to δx_n .

In other words they're what? Sums of terms each involving a δx times excuse me. A δx times a constant. What we're saying is that nice functions, and what's a nice function? A nice function is one which is continuously differentiable. A nice function is locally linear. In other words, a continuously differentiable function near a particular point can be approximated by a linear function, where the error will be very small as long as you stay near the point in question.

You remember at the beginning of my lecture I said something old, something new. This finishes the old part of the course. In other words, what I've tried to motivate for you here is why if we were remodeling the pre-calculus curriculum much more emphasis should be paid to linear equations. Granted that most functions in real life are non-linear, the point remains that locally, functions are linear. OK?

That's the key point. Locally we deal with linear functions. Therefore, since all non-linear functions may be viewed as being linear locally, this motivates why we should really study systems of linear equations. In other words, this motivates the subject called linear systems. Now what is a linear system? Essentially a linear system is m equations in n unknowns.

In many cases m and n are taken to be equal, but what kind of equations are they? They are equations where all the variables appear separately to the first power multiplied only by a constant term, and by the way, let me introduce this double subscript notation rather than introducing umpteen different symbols for constants.

Notice that a very nice device here is to pick one symbol like an a , and then use two subscripts. The first subscript telling you what row the coefficient is referring to, and the second one which column are in terms of the equations. The first subscript tells

you which equation you're dealing with, and the second subscript tells you what variable it's multiplying.

For example this is what? This is the coefficient of x_1 in the first equation. This is the coefficient of x_n in the first equation. This is the coefficient of x_n in the n -th equation. Think of this as the row and the column if you will. And what we're saying then is that the solutions of this type of system of equations are really controlled by the coefficients of the x 's.

In other words, by the numbers a_{ij} , where i and j take on-- well i takes on all values from what? The number of rows. i goes from 1 to m , and j goes from 1 to n . But the a 's become very important, and this is what ultimately is going to motivate what we mean by a matrix, but before I come to that let me give you just one example of what I mean by saying that the equations are governed by the coefficients of the x 's, not by the constants on the right hand side.

By the way, notice the convention that when you have two equations with two unknowns rather than call the unknowns x_1 and x_2 , it's conventional to call the unknowns x and y . Let's take a particularly simple system here-- $x + y = b_1$, $x - y = b_2$. If we add these two equations, we get $2x = b_1 + b_2$, whereupon $x = (b_1 + b_2) / 2$. If we subtract the two equations, we get $2y = b_1 - b_2$, whereupon $y = (b_1 - b_2) / 2$.

Notice that this tells us how to solve for x and y in terms of b_1 and b_2 . Namely, to find x you take half the sum of the two b 's. To find y , you take half the difference. Now certainly the solution depends on the values of b_1 and b_2 . I'm not saying you don't change the answers by changing the constants on this side. What I am saying is that the structure by which you find the answers does not depend on b_1 and b_2 ; it's determined solely by the coefficients of x and y .

What we're saying is no matter what b_1 and b_2 are in this particular problem, to find x and y we take half the sum of the b 's, and we take half the difference. In other words, the solution depends on b_1 and b_2 numerically, but not structurally.

Well the whole idea is this-- and this is what we so often do in mathematics.

Because the solution to our equations depends on the coefficients of the x 's, we somehow want to focus our attention on the coefficients. And we don't need the x 's in there, because we can sort of think of the x 's as being a place value type of situation. In other words, x_1 can be thought of as being the first column. x_2 to the second column. The first equation can be thought of as the first row. The second equation, the second row.

And what this motivates is a concept called an m by n matrix. Now this sounds like a very ominous term, an m by n matrix. But the point is it's not a very ominous term. It's in fact, I think that it's too-- in fact the word matrix essentially indicates an array, and that's all this thing is. By an m by n matrix, we simply mean a rectangular array of numbers, arranged to form m rows-

In other words, the first number tells you the number of rows, and the second number tells you the number of columns. Now there's certainly nothing logical about that in terms of our game idea. Just memorize this, it's a rule of the game or a definition. Somebody could've said, why didn't you give the columns first and then the rows? Well we could've, but one of them had to come first.

And the convention is that one refers to the rows first, and then the columns. An m by n matrix then is what? It's a rectangular array of numbers consisting of m rows and n columns. By way of an example-- by the way, to indicate that's you're talking about a matrix, one usually encloses the array in brackets, or in parentheses. It doesn't make any difference. I will use whichever one strikes my fancy at the moment.

And it happens to be brackets right now. But if I write down this array-- What is it now? $[1, 1, 1; 1, -1, 2]$; this is a rectangular array of numbers consisting of what? Two rows and three columns. And so this is an example of a 2 by 3 matrix. A 2 by 3 matrix.

Now again, we don't want to invent this thing vacuously. Let's keep track of what this matrix is coding for us in terms of a system of equations. Well. For example,

suppose we have the system of equations z_1 is equal to $y_1 + y_2 + y_3$. z_2 is equal to $y_1 - y_2 + 2y_3$, and we want to think of the y_1 , y_2 , and y_3 as being the variables. z_1 and z_2 as being the constants here.

What is the matrix of coefficients here? Well the matrix would be what? The coefficient of the first variable in the first column is 1, second variable first column is 1, third variable first row is 1. You see? Second equation, first variable coefficient is 1. Second equation second variable coefficient is -1.

Second equation, third variable. Coefficient is 2. So using our matrix coding system, the matrix of coefficients would be what? $[1, 1, 1; 1, -1, 2]$. Which is exactly the matrix that we wrote down over here. And to put this into a different perspective so to see what we're driving at, let's take a second example where we first start out with three equations and four unknowns. Three linear equations and four unknowns. And then we'll write the matrix for this afterwards.

But let the equations be $y_1 = x_1 + 2 * x_2 + x_3 + x_4$. $y_2 = 2 * x_1 - x_2 - x_3 + 3 * x_4$. $y_3 = 3 * x_1 + x_2 + 2 * x_3 - x_4$. if I want to write the matrix of coefficients, what do I do? I simply leave the variables out, and write down what? My first row would be what? $[1, 2, 1, 1]$. My second row would be $[2, -1, -1, 3]$. My third row would be $[3, 1, 2, -1]$.

In other words, my matrix of coefficients now would be what kind of a matrix? It would be a rectangular array of numbers consisting of three rows and four . Columns. All right? And that would be called a 3 by 4 matrix.

Again know this-- in this coding system, the number of rows corresponds to the number of equations. And the number of columns corresponds to the number of variables that are formed in linear combinations. To summarize this again-- the matrix of coefficients in our second example is the 3 by 4 matrix $[1, 2, 1, 1; 2, -1, -1, 3; 3, 1, 2, -1]$.

Well again, let's recall that when we do mathematics, we don't like to introduce notation for the sake of notation. And simply to be able to have a way of conveniently writing the coefficients, but not being able to use it efficiently would be

a rather stupid thing to do.

Why invent new notation if it's not going to help us effectively solve new problems? This is why in mathematics we've been emphasizing the game idea whereby what we really care about is structure. We care about structure, not about the terms themselves. And to motivate when I'm driving at, let me return to examples one and two. And bring up a question that has great impact-- and even if we don't appreciate it right now in terms of a practical application, let's at least see what's happening.

You'll notice that if I look at these systems of equations over here, notice that the first two equations tell me how to express z_1 and z_2 in terms of y_1 , y_2 and y_3 . On the other hand, the second system of equations tells me how to express y_1 , y_2 , and y_3 in terms of x_1 , x_2 , x_3 and x_4 . Now without belaboring the point because the arithmetic is quite trivial here, a very natural question that might come up next this is let's look at our old friend the chain rule again.

Since the z 's are expressed in terms of the y 's, and the y 's are expressed in terms of the x 's, it seems that by direct substitution, I should be able to express the z 's in terms of the x 's namely. I replace y_1 by this linear combination of the x 's. I replace y_2 by this linear combination of the x 's. I replace y_3 by this linear combination of the x 's.

| then combine the y 's in terms of the x 's as indicated here. And that should give me the z 's in terms of the x 's. Leaving that hopefully as a trivial exercise, we come to the next example that I'd like to mention here, and that is suppose you were told to express z_1 and z_2 in terms of x_1 , x_2 , x_3 and x_4 . The point is that with the amount of arithmetic mentioned before we could easily show that $z_1 = 6x_1 + 2x_2 + 2x_3 + 3x_4$.

While $z_2 = 5x_1 + 5x_2 + 6x_3 - 4x_4$, by a straightforward substitution. The point is that somehow or other, we would like to be able to handle this substitution more efficiently. Is there a neater way of being able to transform the z 's into the x 's by way of the y 's? In other words is there a way of replacing the y 's by the x 's, and then finding z 's in terms of x 's in a convenient, mechanical way that will save us much

steps?

Not so much in these easy examples where you have 2 by 3, and 3 by 4 systems, but cases where you might have 10 equations, and 10 unknowns. Or 10 equations and 12 unknowns. And the answer is, there is a way. Of course you knew there was going to be a way. Otherwise we wouldn't be leading up to it in this particular way, and as so often happens, there usually happens to be a real life situation that motivates why we invent something called matrix algebra.

In terms of our present illustration, the chain rule that we're just talking about expressing the z's in terms of the y's, and then the y's in terms of the x's motivates what we mean by matrix "multiplication". And you may notice that I put "multiplication" here in quotation marks. The reason I put in quotation marks is unfortunately the word "multiplication" has a connotation of multiplying numbers together.

Don't think of it that way think of multiplication meaning what? A way of combining two matrices to form another matrix. There's going to be no logic behind this other than one very famous piece of logic. That is knowing what the answer was supposed to be we make up our rules to guarantee us that we will get the appropriate answer.

I remember when I was an undergraduate in college. The big type of humor that was going around at that time was the idea of, somebody would give you the answer, and you have to make up the question. Oh they were silly little things like, if the answer to the question was 9w what was the question? And the question would be, do you spell your last name with a v Herr-Wagner, and the answer would be 9w.

And these were funny jokes at that time. I don't know whether their funny now or not. But the funny point is this. That this joke, which might not be that funny is exactly how we motivate definitions and rules in mathematics. We start with the answer, and then go back, and answer the question. Knowing in advance that somehow or other, the matrix that expresses the z's in terms of the y's is given by this. And the matrix that expresses the y's in terms of the x's, is given by this matrix.

Somehow or other what we would like to do is invent a way of combining these two matrices to give me the matrix that expresses this answer. In other words, if I start knowing what the answer is supposed to be-- what is the matrix that expresses the z's in terms of the x's? Is the matrix whose first row is $[6, 2, 2, 3]$. And whose second row is $[5, 5, 6, -4]$. In other words the matrix would be what? $\begin{bmatrix} 6 & 2 & 2 & 3 \\ 5 & 5 & 6 & -4 \end{bmatrix}$.

And without even looking at any mechanical rule, the question that comes up is, how can I invent a rule that will tell me how to multiply this 2 by 3 matrix by this 3 by 4 matrix to obtain this 2 by 4 matrix? 2 by 4 matrix.

Now look in the notes, I'm going to do this in great detail. There will be many exercises on this for you to sharpen your teeth on. But for now I just want to hit this main point because the lecture is quite long. Your attention span probably is starting to be taxed. And so I just want to show you what the recipe is because my feeling is that this is something you have to hear before you can really read it without becoming panicked by the notation.

The idea is this-- first of all to multiply two matrices, all we ever require is that the number of columns in the first matrix equals the number of rows in the second matrix. And if that sounds complicated to you, simply think in terms of the chain rule again. The number of columns in the first matrix tells you how many unknowns there are in the first system of equations.

And that number of unknowns gives you the number of equations in the second system. In other words, the number of columns in the first matrix must match the number of rows in the second matrix. Notice we don't care about the number of rows in the first one matching the number of columns in the second, all we care is that the number of columns in the first matrix-- namely three here-- match the number of rows of the second, which is three.

Then the rule works in a very interesting mechanical way that makes use of the dot product. Namely what you do is, suppose I want to find the term in the product of

these two matrices that occupies the second row, third column. What I do is I take the second row. In other words I take the row that comes from the first matrix. I take the column value from the second matrix. In other words, I have what? Second row, third column. And I form the usual dot product that we've talked about. I dot the second row with the third column.

And what would I get if I did that? $1 * 1 = 1$; $-1 * -1 = 1$; and $2 * 2 = 4$. So $1 + 1 + 4 = 6$. So in this product matrix, the term in the second row, third column will be 6. The term in the second row, third column will be 6. Second row, third column will be 6.

Now leaving it as an exercise for the time being, and reading it in the supplementary notes, I'm sure you'll be able to put this all together. It's not nearly as difficult as it sounds hearing it the first time.

I think the most difficult part is rationalizing why one would invent such a definition in the first place. The answer is very simple: we invent the definition to solve a particular problem. Coming back here again, all I'm saying is that if I invent-- for example let me just give you one more checking out point here.

Let me see what the term would be in the first row, second column. To find the term in the first row, second column, I take the first row of the first matrix. Dot it with the second column of the second matrix. See first row dotted with second column, the answer will give me what? The term in the product that's in the first row, second column. Let's check that.

$1 * 2 = 2$; $1 * -1 = -1$; $1 * 1 = 1$. $2 - 1 + 1 = 2$. And therefore, the term in the first row, second column should be 2. It is. You see there's no more motivation to how we multiply these two matrices than the fact that it solves the problem that we want solved.

To find the term that's in the i -th row, j -th column of the product, dot the i -th row of the first matrix with the j -th column of the second matrix. More generally, you can always multiply an m by n matrix by an n by p matrix. What's the key factor? You don't care about the number of rows in the first, you don't care about the number of

columns in the second. What you do care about is what?

That the number of columns in the first matrix be equal to the number of rows in the second, and if you do that when you multiply an m by n matrix by an n by p matrix, notice that the result will be what? An m by p matrix. In other words, the number of rows is governed by the number of rows in the first matrix and a number of columns is governed by the number of columns in the second matrix.

Notice by the way, that this tells us right away that when we want to multiply two matrices it makes a difference in which order that they're written. If we were to take that 2 by 3 matrix, and the 3 by 4 matrix, and interchange them, we don't have the appropriate match up of rows and columns. You can't dot a 2-tuple with a 4-tuple.

The very fact that we say dot the row with the column, the dot product is only defined for two n -tuples. We insist that the n -tuples be the same. The n has to be the same to dot two n -tuples.

Let me summarize today's lecture by saying that in overview, know this-- hopefully we have reestablished the need for linear systems of equations, and secondly, once we have understood what the need for linear systems is, we are now introducing a mechanism whereby we can solve linear systems more efficiently than what we were taught in the past as to how to solve them.

You see what I'm going to do for the next few lectures now is concentrate on a new game called the game of matrix algebra. But that will unfold gradually as we develop the next two lectures. And so until our next lecture, so long.

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.