

IntroToLinearDynamicalSystems-Lecture04

Instructor (Stephen Boyd): I guess we've started. Let me start with a couple of announcements. The first is I guess I should remind you that Homework 1 is due today at 5:00. Several people have asked what they should do with their homework, and I should have said this maybe on the first day, although I guess it's most relevant now. One thing you should not do is give it to me. I am highly unreliable. I will accept it by the way, that's why I'm warning you. You'll give it to me. It'll be in my hand. By the time I get back to Packard, who knows where it will be?

The TAs are more reliable. You can give your homework to a TA if you personally know them and trust them. But really, the correct I.O. method is that there is a box; it's actually a file cabinet; it's explained on the website. But it's near my office in Packard. That's the right place where you should submit the homework.

Okay, the next is that we have gotten more than just a handful of sort of requests or suggestions to do something like a linear algebra or matrix review session. That's for people who either never had a detailed course on this or have successfully repressed the memories of the course they may have taken. So for those people, the TA's and I have been talking about it, and we'll probably do something. It won't be a formal session. It might simply be one of the office hours. One block of office hours will simply be devoted to this topic. And we would of course announce that by email and on the website. So that would sort of be the idea.

Just out of curiosity, by a show of hands, if such a session were to be put together, how many people would go to it? We'll have a session, that simple. That's what we'll do. Okay, any questions about last time? If not, we'll continue.

So, last time we started looking at this idea of the null space of a matrix. Now the null space is a subset. It's a set of vectors, it's a subspace, in fact, as its name implies. And it is a subset of what I would call a dimension of the input dimension. That's what it is. So, it's the vectors that are mapped to zero. Now what it gives is, this gives exactly the ambiguity in solving $Y = AX$. So, it tells you exactly what you know and don't know when I give you Y , about X when I give you Y in $Y = AX$.

In particular, anything in the null space is going to map to zero. So for example, in a sensor or estimation problem, X corresponds to some pattern or input for which your sensors all read zero. So for all you know, it's either zero or it's something in the null space. If the only point of the null space is zero, then that means that's the only way that can happen. This is what it means.

Now in a design problem or in a control problem something in the null space is a very interesting thing. It's something like this; it's something that has, you might say, zero net effect. So in the case of the mass problem, it would be a sequence of forces were non-zero, so the mass goes flying off to the left, right, goes all over the place. But the point is

that N seconds later, it is back at the origin with zero velocity. So that means that basically nothing happened. Okay? So that's the idea.

Now one case of great interest is when the null space – I can't see if – It's not me but that's twisted at an angle, so maybe you guys can rotate your camera there a little bit. There we go. Too much. There we go. Actually, that's still too much. I'll let you guys figure it out back there. So you say that you have zero null space, that's the slang. That means, simply, not that the null space is zero. The null space can't be zero. The null space, the data type is it's a set of vectors. So when you say its zero, you mean it's a set whose only element is the zero vector. But the slang is the null space is zero. And some people call this 1:1, and what it means is this, it is the same as saying that X can always be uniquely determined from $Y = AX$. It basically says when does the transformation A not lose information? And 1:1, of course, means that different X 's get mapped to different Y 's. That actually, literally, what 1:1 means. Otherwise you might call it many to one, would be the other way.

In fact, one too many is not a function, because the definition of a function is that only one thing comes out. Another way to say this is that the columns of A are independent. So when you say something like zero null space, you're really just saying that columns of A are independent. So, it's two different languages for exactly the same thing. And you can check that. Why does it mean that? Because when you say $AX = 0$ implies $X = 0$, AX , one interpretation of AX is it's a linear combination of the columns of A where the X 's give you the recipe, the coefficients. So basically it says that the only way to add up a linear combination of the columns and get zero, the only way, is if actually the recipe is trivial, if all the X 's are zero. That's the same as saying the columns are independent.

Okay, now this part, everything up here is easy to check, or it's basically just a restatement of – I mean, there's nothing to it. This is not obvious, and so a few things here are not going to be obvious, I'm not going to show them now. I'll state them as unproven fact, later we'll come back and you'll be able to show all these things. This is an unproven fact. It says if A has zero null space, if and only if it has a left inverse. Now a left inverse is a matrix B , which when multiplied on the left – Sorry, when it multiplies A on the left it gives you the identity. That's what BA is.

Now you may or may not have seen this. You have probably seen the idea of a matrix inverse, but maybe not a left inverse. Actually, how many people have even heard of the concept of a left inverse for a non-square matrix? That's good because it used to be basically no one. So a left inverse means you multiply on the left and you get I . By the way, the meaning of the left inverse is actually, it's great, it's actually a decoder is what it is. Because if $BA = I$, $Y = AX$ then if you multiply by B , what you get is you get X . So XB , when you see a left inverse, it means it is a decoder or in communications, it's a perfect equalizer. In any kind of signal processing it's a perfect deconvolver or something like that. That's what a left inverse is.

Now what's interesting about this is that, sort of, of course, is a very strong statement about this first one. It says yeah, you can uniquely determine X , given $Y = AX$. This

says, actually, not only can you determine it, but you can get it by applying a linear function. That's not obvious right? It could have been that you could uniquely get X , but you had to jump through all sorts of hoops and do all sorts of signal processing or something like that to get X . It turns out no. In fact, you can decode with a linear decoder.

And the next one is this. It turns out this one is relatively easy to show. It says it's the determinant of A transposed A , is non-zero. At the moment, that means nothing to you, but it will soon. And I think this is one of the ones that's – this one, I have to be very careful here, because some of these are extremely simple. They're either three lines or they're actually very difficult. And all of the difficult ones, by the way, are equivalent to each other.

Shall I try showing this? What do you think? So I reserve the right to absolutely give up one minute into this, and to pretend that I never thought – that I knew ahead of time this was actually one of the tough ones. Okay, so we're just gonna see. We'll see if this is one of the easy ones. What we want to show is this. Let's start by saying suppose that A has a non – We're going to show that $AX = 0$ – Well, here, A is 1:1, if and only if, $\det A$ transpose A is non-zero. Okay, let's try that. All right, so let's try a few things here. Let's start by saying, assume A were not 1:1. Okay? That means, well by definition, it says that its null space is not just zero. That means that there is some X which is non-zero, but $AX = 0$. That's what it means. It means there's a non-zero element in null space.

Now if AX is zero, then surely A transpose AX is zero because I could write it his way, that's the zero vector, and A transposed times any zero vector is zero. But I could also write it this way and write A transpose AX . I can re-associate. Ah, but now we have a problem. We have a square matrix here multiplying a non-zero vector and we get zero. And that implies that the determinant of A transpose A is zero. Okay? So we just showed that if this fails to hold, this fails to hold.

And now we do the opposite. Now let's assume – that's one way. Let's assume now $\det A$ transpose A is zero. Assume that's zero. That means the matrix A transpose A is singular. That means there's some non-zero vector, which gets mapped to zero like that, where X is non-zero. Now let me ask you a question here. If I had scalars, and I wrote down – Actually, can I just cancel the A transpose off of here? That's an excellent question. Can I just write this as A transpose $A = 0$, as A transposed times $AX = 0$, and now say A is non-zero, so I can just A transposed non-zero and just remove that. Can I do that? Can you cancel matrices? No, in general, no you absolutely cannot. So it's entirely possible you can have the product of two matrices, both non-zero, and you get zero. So you can't cancel matrices. Okay?

That's not going to help, but watch this. This is actually – Here I'll show you a trick, you may have to pan up here. It's going to set the pace for the class. I can move it up. All right, what happens is this. We know this, and here's the trick. If this is true, I can certainly multiple on the left by A transpose and get that because X transpose times a vector that's zero is surely zero. That's the case. Now I'm going to rewrite this in a different way. I'm going to write this as AX transposed times that. And that's just one of

the rules for transposition. You transpose a product by transposing each thing and reversing the order. Okay?

Now you may know that the – Well, I guess its part of the prerequisite. I think actually we're going to cover it in a few slides anyway. But in fact, the transpose of a vector times itself is actually the square of its norm. That's actually this. This is actually nothing more than the sums of the squares of the entries. This is $\sum x_i^2$. Now when a sum of squares of numbers is zero, every single number is zero, period. And in fact, what this says is, therefore AX is zero. That's just what we wanted, because now I have an X in the null space of A , and its non-zero, and therefore A is not 1:1. Okay?

By the way, I won't do this very often. I would actually encourage you to see how much of – Most of the statements in the notes are things you should check. These things get boring after a while, but you should actually try to check as much as you can. Most of it is quite straightforward. Some of this stuff for now is not.

Okay, so let me say something about the interpretations of the null space. Well, suppose $Y = AX$ describes a measurement set up or something like that. Then what it says then is something in the null space is undetectable. It is a pattern that is undetectable from the sensors. Because when you have that pattern, all the sensors just give you flat zero reading. And you don't know if maybe what you're looking at is zero. In fact, it gives you the exact same signature, zero. It also says this, it says that X and $X + v$ – is you take any vector and you add anything in the null space, they are completely indistinguishable according to your measurements, totally indistinguishable because they give you the same readings, the same Y , the same AX .

Now in fact it characterizes the ambiguity in the measurement. That's what it does. Now if $Y = AX$ represents something like an output resulting from an input or action, X , then basically it says that something in the null space is actually an action. It is a non-zero action that has no net result. That's what it means, and in fact in this case, the null space characterizes the freedom of input choice for a given result. So actually having a big null space in a measurement problem is, generally speaking, bad because it means there's lots of ambiguity, if what you're trying to do is recover X .

If on the other hand, X represents an action that you can take, and Y is an outcome, having a big null space is great because it says there are lots of X 's. And you have an organized way, in fact, of finding X 's that will satisfy the requirements. So that's good, and it means that you can drop back, and you can optimize something else. You can drop back and find a small X or whatever you like.

Now the range of a matrix is another one of these fundamental subspaces, or whatever. And it's denoted with a calligraphic R or a script R or something like that. It's range of A . This is also a subspace. It is not a – it is a subspace of the output dimension. So it is actually a subset of R^n and it is simply the set of all vectors you can hit. It is the set of all vectors that have the form AX where X ranges over all of R^n . So this is a subset of R^m .

Okay, unless your matrix is square, the null space and range are sets of vectors of completely different dimensions.

Now you can interpret it this way, it's a set of vectors that can be hit by the linear mapping $Y = AX$. That's what it means. It's the set of all things you can achieve, or something like that. But it's also, in terms of the columns of A , it's very simple. It's the span of the columns, so null space zero means that the columns are independent. If the range – Here the range is simply the span of the columns. Okay? And you can also think of it this way. It's the set of right hand sides, if you write it as $AX = Y$ for which $AX = Y$ has a solution, so that's the idea.

Now here it has all sorts of interesting interpretations and applications in different areas. So for example, $Y = AX$ represents a design problem where X is an action, Y is an outcome. Range of A is the set of all things you can actually achieve. So if someone says, "I would like this outcome." But if that outcome is not in the range, it means, "Sorry, I can't do it. I can't get that outcome. Maybe I can get close." We'll talk about that. "But I can't get that outcome." In the case of a sensor reading, it is also very interesting. $Y = AX$, the range means it is a possible sensor measurement. What would be a practical use for the range in a case like that?

Student:[Inaudible]

Instructor (Stephen Boyd): Got it, that's it precisely. So what would happen is this, if in fact the measure – If you believe the system is modeled by $Y = AX$, Y is a vector of measurements. If Y comes and the first thing you do with Y is check, is it in the range? If the answer is yes, no problem, you can then proceed to whatever you're doing. If the answer is no, you want to throw an exception. You want to throw an exception to whoever passed you the Y , and basically says sorry, that Y could not have possibly been something of the form AX . By the way, that might mean one of your sensors is off. So this could be a sensor detection routine or something like that. And you could do all sorts of interesting stuff there. How would you check? Suppose you knew that one sensor was bad, what do you do? What's the algorithm? How do you check it? That can be a homework problem, by the way. Want to make a note of that? That's simple.

So let's talk about it. How do you do that? If I give you $Y = AX$, X is in R^{10} , Y is in R^{20} . Roughly speaking, you've got about a 2:1 redundancy of sensors over parameters you're estimating. But there's a trick. One of the sensors has failed. How would you find it? So, I want you to tell me how to find out which sensor. I want you to come back and say, "Sensor 13 is no good." How do you do it?

Student: You put in two inputs.

Instructor (Stephen Boyd): You know what, actually it's passive. You don't have – You can't do that. You don't have the – it's passive. You can't control X , you're just given readings, that's all. You had a suggestion.

Student: Take the [inaudible] of Y [inaudible].

Instructor (Stephen Boyd): I don't think you can take the inner product of Y with the rows of the sensor matrix because they have different dimensions. The rows of the sensor matrix are of size R_n , the input. And the output is of size n – if it was a ten in our example, the other's R_{20} , but you're close. What would you do? [Crosstalk]

Instructor (Stephen Boyd): Find the null space of what? Student:

[Inaudible]

Instructor (Stephen Boyd): Let me ask you this. Suppose I have a sensor suite, and I say, "Please take out sensor 13, what does that do to the matrix? How do you express that in terms of what you do with $Y = AX$, Y used to be 20 high, now it's now 19. How do you get that y and what do you do with A ? What do you do?"

Student: You delete the 13th row.

Instructor (Stephen Boyd): You delete the 13th row of A . Okay, so that's what you do. So basically, if everything is fine, if it's $Y = AX$ that's what it's been for a while, and I come along and say, "Nuh, uh, sensor 13 is going out for maintenance." Then you have a reduced A and you remove the 13th row. That's what it means. Does everybody have this? So now you have a 19×10 matrix. Okay.

You can calculate the range of that 19×10 matrix. Let's suppose – Let's say you can do that. You will within one lecture. You can calculate the range of that matrix, and you can check if the reduced measurement is in the range. If it's in the range, what does it mean after I remove the 13th row? You have to be careful. What does it mean? It's kind of complicated.

Student: [Inaudible]

Instructor (Stephen Boyd): I'm going to turn it around and say that it means that the 13th – It means, if now you remove a row, and you look at the reduced Y , and it's in the range of the reduced matrix, what it means is this; the 13th sensor might not have been bad. Sorry, it's a strange statement, but that's exactly what it means. What if, I remove the 13th row; I look at the range of that matrix. I look at the range of that matrix or something like that, and actually now the reduced Y is not in the range. It means you still have a bad sensor. It means you're still wrong. It means the 13th was not the probably. I'm assuming there's only one sensor failure.

I have a feeling that I'm just making things more confused. This is what we call a homework problem. You will be seeing this shortly. Trust me. Okay, so. Or you can go back and see if any of this makes sense later. All right?

Now, there's a special case, which is when you can hit all vectors in the output space. That's when the range is R_m . That means you can hit everything. That says, you can solve $AX = Y$ for any Y . That has all sorts of obvious interpretations. We'll get to that. Another way is to say that the columns of A span R_m . That's what it means. Now in this case, this is not obvious. It says there is a right inverse of A . So, this says, for any Y you give me, I can find an X for which $AX = Y$. This third bullet here is much more specific. It actually says, not only that, but in fact I can construct the X as a linear function of Y . That's what it means.

So here, if in fact there's AB with $AB = I$, that's a right inverse. What it is says is this, it says that ABX is X , well of course because IX is X . On the other hand, it just says – You know what, I shouldn't have used X . Sorry. There we go. It was perfectly okay, just horrible notation.

What this says is – this is quite specific. It says, “Oh, you want an Ax for which $AX = Y$, no problem. Let's take your Y and multiply by B .” So B actually is an automated design procedure. That's what it is. If this is a design problem, it's a design procedure. It says, “You want this velocity and position? Multiply by B that gives you a force program that does it.” Okay, so that's what it is.

This is the same as saying that the rows of A are independent. Again, you can check. And it's basically, now it gets tricky. This is the null space of A^T is zero. Now, some of these look complicated. Some actually are, most are not, but I am going to defer a lecture until we show all of the details. So this is the null space of A^T is empty. And you can say this, these two are actually basically the same thing because to say that the null space of a matrix is equal to just zero is basically saying that the columns are independent. But the columns A^T are the rows of A . And so, this one and this one are simply the English equivalent of that statement. Okay? And this one would probably follow up from one on the previous page or something like that.

I want to warn you, some of these are not obvious. But by the way, I would not – You should sit down and sit down and see how many of these you can actually show. You won't be able to show some of them, like that one. Maybe that one, I'm not sure, right away. But within a lecture you will be able to.

Okay, so let's look at various interpretations of range. Suppose a vector V is in the range of a matrix, and a vector W is not in the range. We'll see what does this mean. If $Y = AX$ represents a measurement. Then if you observe V that means it's a possible, or maybe a better name is a consistent sensor signal. We just talked about that a few minutes ago when I confused all of you, most of you any way. So if however, you get a sensor measurement which is not in the range, that basically says it's impossible, if you really believe $Y = AX$ or it means it's inconsistent. That's what it means.

Now if $Y = AX$ represents the result of an output given an action, or an input or something like that, then it says, if you're in the range, it's something that's possible. It's achievable. However, something not in the range is something that simply cannot be

achieved. That's what it means. So, in this case R_A characterizes possible results or achievable outputs. That's what it does.

Okay, now so far we have talked about the size of A has been arbitrary, it doesn't have to be square. And in fact, once you find my greatest complaint with, aside from the fact that they are profoundly boring, linear algebra classes. The first linear algebra classes, my first complaint is that they focus on square matrices too much, which actually in engineering has essentially no use. We'll see why, but in fact, most times in engineering when you have, in any kind of engineering application, you have a square matrix, except we'll see some examples later, but in general it means something is deeply wrong. We'll see why. You'll know why later, why in fact matrices should either be fat or skinny or something like that.

It means kind of that you set things up so there are just enough knobs to turn to do what you need to do. That's silly. The whole point is you either want to exploit redundancy or something like that. There are cases where you want to do that. If you want to measure something, if you want to measure six things, and sensors are super duper expensive, and for some reason you can only afford six, okay fine. But in general, this makes absolutely no sense. It's not robust either. In fact, if you have seven sensors to measure six things, at least then you can actually add a nice layer of software that will actually detect and remove a sensor that fails, automatically. Using the method which I deeply confused all of you with, and which you will see on the homework shortly.

Hey, what do you think? Do you think we're allowed to, even though we've posted Homework two? I think we can do it, we'll see. All right. We just have to make – it may have to wait until Homework three, we'll see.

Now we're going to look at a square matrix. You see, it's invertible or non-singular if its determinant is non-zero. And this is equivalent to many, many things. The things like the columns of A are a basis for R^n . That means they span R^n ; that means its range is everything. And it means that they're independent which means that, in this case, the other half of being a basis, so leave that. It means a null space is zero. That's the slang. The null space is the set consisting of the zero vector. I don't think – I quit. I'm going to start saying null space is zero.

It also says the rows of A are basis for R^n . It says $Y = AX$ is a unique solution for every Y in R^n . So, it means it has now – there's a matrix that is both its left and its right inverse, and it's the only one. So, if you're a left inverse of a square matrix, you're also a right inverse. Okay? It's the same as saying the null space of A is just zero, and the range is everything. And it's the same as saying the determinate of $(A^T)A$ and $\det(A)A^T$ is non-zero. By the way, for a square matrix, this is easy to show because here, I can write this, $\det(A^T)A$ is $\det(A^T) \times \det(A)$, right? That I can do. Can I do that here? If I wrote here, that's $\det(A) \times \det(A^T)$, what would you say to me?

Student: Syntax error.

Instructor (Stephen Boyd): You'd say syntax error, exactly. Why? Because \det takes as argument only a square matrix, and in this problem there is M and there is N and they need not be the same. So this is a syntax error, which is a terrible, terrible thing, right? Because it means the most casual parsing would reveal the error. We'll see later that there are some deeper semantic errors. That's where the syntax works out but the meaning is wrong. Strangely, a semantic error is actually less of a crime. Okay, but in this case, they're square, so we can write these out, and these are very straightforward.

So what's the interpretation of an inverse? Well, I mean, this is kind of obvious. It basically undoes the mapping associated with A . Actually, the cool part is its both – it can work as equally well as a post equalizer or a pre-equalizer or a prefilter, and I'm using some of the language of communications here. In other words, you can in this case, you can send something through a channel, and then apply B to reconstruct what happened. That's a traditional equalizer. Or, before you send the signal, you can process it through B , and that has lots of names like, one is like a pulse shape or pre-equalizer, or something like that, in communication. So that's the idea.

Okay, so that's that. That you've seen. I'll mention one other thing involving that. One interpretation of inverse, in fact you can also use it for things like left inverses and right inverses, but let's take B is A^{-1} , and let's look at the rows of the inverse of a matrix. Well, let's write this, this is $Y = AX$. If you write it out column by column is $Y = X^{-1}A^{-1} = X_n A_n$, of course $X = BY$. This first one is $Y = AX$ and this is just $X = BY$, because that's the inverse here. Now, if I write this way, I want to write out Y column by column, this one, and I'm going to write X out row by row. This is XI is BI until they transpose Y , because when you multiply a matrix B by a vector Y , if you do a row interpretation it says you take the inner product of each row of B , with Y . Did I get that right? I did.

Hey, that says look. These XI 's I can plug them in here and I can rewrite it this way. And now you see the most beautiful thing. You see what the rows of the inverse of a matrix are. They are things that extract the coefficients of – so this basically says something like Y is a linear combination of the A 's. Now if you want to know what linear combination, or how do you get the mixer, what's the recipe? It says you take the inner product with the rows of the inverse matrix and these give you exactly the coefficients in the expansion of Y in the basis A^{-1} through A_n . Okay?

Sometimes people call A_1 through A_n and B_1 through B_n , dual bases. That's actually kind of common now in some areas of signal processing. They call it – this goes back a long way, and it's interesting, actually to work because we're going to look at some other stuff later. But let me ask you something. What can you say about this? What is that equal to? What is it?

Student: [Inaudible]

Instructor (Stephen Boyd): Yep, that's one. If $I = J$, and its zero if $I \neq J$. Okay? By the way, actually very soon we're going to talk about things like orthogonality and things like that. So some people refer to these as biorthogonal sets or something like that. In

other words, it's kind of weird, it says that if the B's are orthogonal to the other A's but it doesn't – what do you know about this? What can you say about (AIT) AJ? Not much. There is only one thing I can say. I can say something intelligent about that. What can I say intelligent about that? It's what?

Student:It's non-zero. Instructor:

How do you know its non-zero?

Student:[Inaudible]

Instructor (Stephen Boyd):So I would have a column of A0, what's wrong with that? I could remove it. So, remove it.

Student:[Inaudible]

Instructor (Stephen Boyd):Thank you. It's all the same thing. Yeah, A could not be invertible if it had a column that was zero. Right. Okay. All right. All of these things are very abstract now, but we'll see them in applications and it will – Yeah?

Student:[Inaudible]

Instructor (Stephen Boyd):No, they're not. Ah, but a basis doesn't mean that they're orthogonal. We'll get to that. Yeah. We'll get to that. By the way, if your background is physics or its something like applied physics or something like that, most, a lot of the matrices you encounter will be not only square but symmetric. And so, a lot of operands will be self-adjointed. And that means a lot of things because that's what you've always seen. If that's the case, how's my diagnosis? That's a diagnosis there. Okay, so in that case this class is great for you. You're going to see all sorts of stuff, and you're going to learn all the things that are almost always true, like eigenvalues are real for you, right? Yeah. Got it. Good. You'll be fine, everything will be great.

Actually, then it comes up. You'll see it's going to come up in physics, too. So, we'll see. There will be cool cases where it's not symmetric. And then weird things will happen and all your fellow physicists will be deeply and profoundly confused because for them everything is symmetric, all bases are orthogonal and so on.

Rank of a matrix. The rank of a matrix is simply the dimension of its range. And here are some facts, these are not obvious. Things like the rank of A is the rank of AT. Another way to say this is the rank of A is the maximum number of independent columns or rows of A you can have. And that means immediately the rank of a matrix cannot be any more than the minimum of its input or output dimension, number of rows or columns. I actually said those backwards, but that's fine. That's what it means. If you have a 3 x 5 matrix, its rank could not possibly be more than three. It is at most three.

Here's a very basic fact, some people actually make a big deal about this and they circle it, and it's in some fancy box and everything and it's called sometimes, the fundamental theorem of linear algebra or something like that. They make a big deal. They have a theme song, and anyway. This is sometimes called the fundamental theorem of linear algebra. It's beautiful, but it's not that big a deal. Actually, you'll find out it's actually much cooler to be able to do stuff with linear algebra than to celebrate pretty formulas. You know, it's like $E_i + 1 = 0$, you get over that and everything's fine.

What this says is very interesting. It says that the rank of the matrix, that's the dimension of the set of things you can hit. That's this, plus the dimension of the null space, that's the dimension of sort of the ambiguity set. The size of the set of things that get crunched to zero is equal to N , which is the input dimension. My interpretation of this, well I'm sure lots of people interpret it this way, is this is something like conservation of dimension or if you like, something like degrees of freedom. I mean, maybe that's a better way to say it. Don't take any of this too seriously, but it's just to give you the idea.

So the rank of a matrix is actually the dimension of the set hit, but let's make this very specific. Let's take A is $\mathbb{R}^{N \times R}$, well we just had something that was 20×10 . That's it; so by the way, the name for that technically is that's a skinny matrix because it's got 20 rows and ten columns. I'll use that terminology. It's going to come up a lot. So I have a skinny matrix, $\mathbb{R}^{20 \times 10}$, and I'm going to tell you that the rank of A , by the way could it be 11? No, but I'm going to make it eight. Okay? So there it is, it's a 20×10 matrix whose rank is eight.

Now what does this mean? It means that if you ask, what can you hit, what's the set of things that have the form AX ? Well, the answer to that is this. First of all, that's a subspace of \mathbb{R}^{20} , so you have twenty long vectors and there's a subspace. It could have a dimension up to ten, but in fact it only has a dimension of eight. That's what it says. It's got a dimension of eight. So basically it says, if X is an action and Y is a result, it says that the set of things I can effect or do is actually eight dimensional. That's what I can do. I can do eight dimensions worth of stuff.

But wait a minute, in this case I had ten input knobs, so roughly speaking, you would look at that and say, if what you can do is only eight dimensional, you have ten input knobs, you have two redundant knobs. Guess what. That's the dimension of the null space, so the dimension of the null space of A must be two because they have to add up to ten. That's the idea here. It basically says something about the total number of degrees of freedom going in minus the degree of ambiguity is equal to the total dimension of what you can affect. That's kind of what it says. It's a good idea to have a rough idea of what this is, and it will get much more specific in specific cases.

This one is very loose. Each dimension of input is either crushed to zero or ends up an output. That's boy, I wouldn't put that in anything but notes. Because it's actually, in some sense, I didn't even know what the dimension of an input is. What I'm saying is that if any of you said this, boy, we'd come down hard on you. I mean, unless it was

casual, in casual conversation, but in anything other than that you would not want – On the other hand, as a vague idea it kind of explains what's going on.

Now there's another interpretation of rank, which is actually very, very nice. It's the idea of coding, and it comes up in ideas like transmission and actually it comes up in compression, all sorts of things. So here it is, if you have a product of matrices, then the rank is less than the minimum of the ranks of the matrices, of the things that you multiply. This sort of makes sense because if you look over here, and you think about what you can do, it basically says – this one is actually very, very easy to show. This one up here. But now what this says is if I write a matrix A in factored form, if I write B as M by R matrix multiplied by an R by N , you can think of R as an intermediary dimension, the intermediate dimension. It basically says, I have something that looks like this, that's A , but I'm going to write it this way, and I'm going to factor it, and that means basically I'm going to write it as a product of two things. Actually, if you like instead of one piece of code, it's gonna call two, one after the other. Okay?

And what this says is, in this case R is the intermediary dimension, it's the intermediate dimension here. That's what it is. Now if you factor something like this the rank is less than R , but in fact, if the rank is R you can factor it this way. So I can write it this way. Now this doesn't look like it has any – It's interesting, it's actually extremely interesting, and tons of super practical stuff is based on just this idea. But we'll just look at one now. Actually later in the class we're going to do something even cooler which is this. Very soon you'll know how to factor a matrix into one to make a factorization like this.

This has lots of applications immediately. Like, suppose I want to transmit Y , given X , but I pay for the number of real numbers I have to transfer between them, right? If N is 100, and M is 200, but R is 3, this makes a lot of sense because what happens is first I do this processing on this transmit side. I transmit three numbers, or whatever I said the rank was, and then they're reconstructed over here. Already these ideas should have ideas of efficient transmission, compression; all that kind of stuff is close to here. It is very close to this.

Let's just look at one application. Let's suppose I need to calculate a matrix vector product $Y = AX$. Now you would think that can't be that big a deal. Why would that be a big deal? And it would not be a big deal if A were, for example, $5,000 \times 5,000$. Right? Or it depends on what we're talking here, A could be $20,000 \times 20,000$, and you could do this doing NPI and a whole bunch of processes, or something like that. Or A could be 100×100 and you could do this in a – actually, 40×40 and you could do this in a shockingly fast way. Okay? I'm just multiply $Y = AX$. This could be – What are you doing when you're multiplying by A ? This could be an equalizer. It could be precoder. It could be some estimation you're doing, it could be anything. So we're not worried about that.

Now, suppose you knew that A could be factored, and let's suppose – The interesting case is when the rank is much smaller than either M or N . But let's look at this. Now, if I compute

$Y = AX$ directly, that's MN operations because I calculate the inner product of each row of A with Y , and each of those requires N multiplies $N-1$ adds. You know, roughly. Okay? That would be MN operations. With a typical sort of gigaflop machine – well my laptop comes out faster than that, so it's pretty fast. That's okay because as fast as you can do this, there'll be some application where you want to do it faster, or you want to handle bigger things. That would be MN operations.

Now suppose you do it this way, you write A as BC , and you associate it this way. You first operate by C , and then by this. Well, if you multiply by C first, it's going to cost you RN operations, then it costs you MR , and that's $(M + N) \times R$. Okay? Just to plug in some numbers here let's suppose – Let's make both M and N equal to 1,000, and let's take $R = 10$. So it's a 1,000 x 1,000 matrix and its rank is 10. The saving now is pretty obvious. In the first case it's 106 and in the second one it's 102,000, so it's a million versus 20K. Okay? That's a 50:1 savings in multiplying AX . Has everybody got this? This is the idea.

By the way, later in the class we're going to look at very good methods for not just – very soon you'll know how to factor a matrix this way. If it is low ranked you'll factor it. Later in the class, we'll do some things that are very cool. We'll actually look at how to do approximate factorizations. That's even cooler. You will be given a matrix A ; you will know how to find, for example, an approximation of A that's ranked ten. Okay? This has tons of uses all over the place. Okay, that's what it means. If you see a matrix that's low rank, it says actually, you could operate on it quickly or it means it's got something to do with compression or something like that.

Student:How do you know if a matrix is low rank in the first place?

Instructor (Stephen Boyd):How do you know if a matrix is low rank in the first place? You will know, the sort of exact answer, if you want exact rank, you'll know in one lecture. If you want to know – much more interesting question, much more practical is this, given a matrix, how do you know if it's approximately rank 10. You'll know that in about five weeks. By the way, did you notice that I didn't answer your question at all? Okay. But I did it forcefully. So, that's okay. We put your question on a stack, it will be popped. Okay?

One last topic is change of coordinates. Now, the standard basis vectors are these E_i 's like this, and you obviously have this expansion. I mean, that's kind of silly, but it's a pedantic way of writing this out. But you would say in this case that the X_i are the coordinates of X in the standard basis. Normally, you just say X_i are the coordinates. Okay? Or the entries of the coordinates or something like that. That's how you write it.

Now, if you have another basis in R^n , if T one through ten is another basis, you can expand a vector X in another basis, and we'll call the coordinates \tilde{X}_1 up to \tilde{X}_N . Okay? And of course, the fact that that's a basis tell us that – That tells us two things. It says that the T 's are actually independent, which means there's only one way to write it this way. And it means, also, that any X on the left hand side here can be written in this form. Okay? It turns out we can get these \tilde{X} very easily, by a couple of matrix

operations. So here \tilde{X} and the coordinates of these, and what we're going to do is we're going to write this out as a matrix, so we'll write that as $X = T[\tilde{X}]$. Now, this is basically a character representation of this, so you write it this way. But now, immediately we get the answer. T is invertible, and therefore \tilde{X} is simply $T^{-1}x$, and indeed it turns out that the inner product of the i th row of the inverse of a matrix extracts the i th coordinates. Okay?

For example, if someone says – I mean actually, where would this come up all the time? It comes up in things like vision and robotics. And actually, it comes up in navigation as well. So, for the rest of we just use a fixed coordinate system, and seem to go through life just fine, but for seem reason in things like dynamics and navigation and things like that, everyone has to have their own personal coordinate system, and so you have to transform things from one to the other. For example, in robotics, every link of the robot arm has its own personal coordinate system, and you have to transform them. I don't know why people do this, but they do. This is kind of the thing that would allow you to do that. Now you know what it is, it's sort of the inverse of a matrix and it the rows extract the coordinates, and things like that.

Now, you can also look at what happens to a matrix. If you have $Y = AX$, this is all in the standard coordinate system, and now suppose you write X in the T coordinates. And then we're going to write Y also in the T coordinate, and the T coordinates will be \tilde{X} and \tilde{Y} , then just by working it out you get $\tilde{Y} = (T^{-1}AT)\tilde{X}$, that's what it looks like. So, this thing is something that's going to come up a bunch of times, not a bunch of times, but it will come up several times, and it should light up a light in your brain. First of all, it's got a name, its called similarity transformation, that's the first. For the moment this is all very abstract, but you should think of it as basically representing $Y = AX$ but you represent both X and Y in a different basis, and you'll get something that looks like that. So we'll get back to that later.

Now for some reason at the end of this lecture, there's some stuff that's more elementary than all of this and should have been at the beginning. But that's okay, we'll do it now. It has to do with just a quick review of norms and inner products and things like that. Now, the Euclidian norm of a vector is actually the square root of the sum of the squares. Some of the squares you can write as $X^T X$. That's the inner product of a vector with itself. And it's the square root of the sum of the squares. And it's supposed to measure the length of a vector. And indeed it measures the length of a vector with $N = 1$. It's the absolute value. When $N = 2$ it's actually the Euclidian distance in the plain, and for three it's Euclidian distance in R^3 . But it makes sense to talk about then norm in R^500 . It makes perfect sense, and people talk about it all the time.

Now there are various things related to this, but I'll mention some of the basic ideas. So, it's supposed to measure length. It's supposed to be a generalization of absolute value. By the way, some people have actually – the double bars are to distinguish it from the absolute value. But presumably if you were super, super cool, and kind of just grew up from a child onward using vectors and matrices, you would probably write this this way, and there are people who have reached that level of development, of linear algebraic development that they write it this way.

The same way we've got to the level where zero, we don't mark it; we don't put a doodad on it. We don't put an arrow on it, like our friends in physics. We don't make it bold or something like that. So for zero we're fully developed. Zero is totally overloaded for us. It's the same thing for the number, the complex number, matrices, vectors, everything. This is a hold over. Actually, it's 150 years old, maybe more; something like that. There's only one minor catch here. Unfortunately, there's a bunch of fields where people – it's useful to interpret this as the element-wise absolute value. That's the hitch there, but it's supposed to be a generalization of this.

So it satisfies a bunch of properties, these are all easy to show. This one, I believe you did show, or will be showing. Is that right? No, it's close to something you will show or will be showing, which is because showing, which is the Cauchy–Schwarz inequality. Is that right? On this homework? I'm always a couple of homeworks ahead, so I can't remember. Okay, this is a triangle inequality. It's kind of obvious. It says basically, if that's X and that's Y , that's $X + Y$, and it says that the length of this is no bigger than the length of X plus the length of Y .

The picture makes it totally obvious when they would be equal, and that's when X and Y are aligned. This is called definiteness. It says if the norm is zero, the vector is zero. That's easy to show here because the square root of a sum of squares is zero; the sum of the squares is zero. But if a sum of numbers that are non-negative is zero, each number is zero. That means each X_i^2 is zero. Therefore, each X_i is zero.

Now, there are some things related to norm. For example, it's extremely common in a lot of applications where N varies. By the way, if N doesn't vary in an application, like if N , if you're doing control or something like that, you don't – when N is fixed, it's like six and it's the three positions and the three momentums. You don't divide by N or something like that. But in other areas, like signal processing, where you get vectors of different lengths. For example, an audio clip is a vector, but N depends on what the length of the audio segment is. And you want to talk about different things like that. Or it could be any other signal.

It's extremely common to normalize the norm by square root N . That's the same as taking the sum of the squares of the elements, dividing by N . This is called the mean square value of the vector X . And that is obviously the root mean square value, that's the RMS value of a vector. And, it just gives you a rough idea of the typical size of that the entries would be. There's actually some things you could say about it. But so if the RMS value of a vector is one or one point three, it means that basically if you look at a random entry it should be on the ord. of one point three.

By the way, if a vector has RMS value of one point three, how big could the entries be? Could they be 1,000? Could you have an entry that's 1,000?

Student: It depends on [inaudible].

Instructor (Stephen Boyd): I'll simplify your answer. Ready? Yes. That's his answer. That's correct. You can have a vector whose RMS value is one point three and you could have an entry in it that's 1,000. That only works if you have a really big vector. You would have a lot of zeros and then a really huge one. You couldn't have a four long vector that had an RMS value of one point three and have an entry of 1,000. There's things you can say about that, but that's the rough idea.

Now, norm is length from zero. If you take, since you have a difference, it says the norm of a difference gives you a distance. This is very important. This allows you to talk about the distance between two things. This is what allows you to do things, talk about the distance between, for example, two audio signals, or two images. You can say, and in fact here you might even give an RMS value, and in both cases it would be very interesting. What would you do? You might have the original and some encoded and subsequently decoded signal, and you would say, "I was within in 1 percent." The RMS error is 1 percent that means the norm of the difference divided by the norm of, say, the original signal is about point-zero-one. That's what that would mean.

Now, the inner product, of course we've been using it all the time. There's lots of ways to write it, but one way is this. It's actually just $X^T Y$, nothing else. In fact, we won't write this. Actually, it's kind of cool. There's other ways people write this. I'll show you some of them. Here's one that you see. Again, we're going toward physics now. This would be the notations there. There are others. You could write $X \cdot Y$ that's one way. And I've seen this. You could also do dot, and depending on what subfield they're in, your dot is either little or big thick thing. These are different notations for inner product. Does anyone know any others because there's probably some others too?

All right, so the inner product, you can work out these properties is quite straightforward. In fact, I don't think I'll say much more about that, I'll zoom on. Because these are kind of obvious.

The Cauchy-Schwarz inequality is very important and it says this, that the inner product of two vectors in absolute value is less than or equal to the norm of the product. From this, you can immediately derive, for example, the triangle inequality right away. But that's what this says. And this is something, I guess you show on the homework, or have shown, or will show in the next half day or something like that.

The angle between two vectors is simply the arc cosine of the inner product divided by the product of the norms. Now, this number here is between plus and minus one. And the arc cosine, let's draw the cosine. And we usually take it between zero and 180, or something like that. So we don't distinguish between minus 10 degrees and plus 10 degrees. It's the unsined angle, is the way to say it. So it's the angle between things. And this allows you to do all sorts of cool stuff. To talk about two images, and say that they have an angle of 3.7 degrees with respect to each other. It makes perfect sense. Or an image can come in and you can have a database of 10,000 images and you could ask which of the 10,000 images is the one that just came in. Have the largest inner product

width, which would mean the smallest angle. And this would make sense, and you can already start guessing these things have applications.

It's kind of weird to talk about the angle between two vectors, in for example R one million because that's what you're doing if you're looking at 1K by 1K images. After a while – don't try to visualize, by the way, R one million. Or at least work your way up there because you could be in big – yeah. Start with R_4 , when you're comfortable at R_4 , maybe R_5 or R_8 , something like that. There also might be some drugs you might have – some pharmacological things are needed to really, really understand R .

Actually, Tom Kover gives a wonderful talk, basically showing beyond a shadow – makes it completely clear. You know, we pretend, so people ask, “What are you doing here?” When we're talking about the angle between two images, “There's no such thing as the angle between two photos.” And I go, “Yeah, sure, its 3 degrees.” Then I'd say, “You don't understand, we do this kind of applied math stuff, and the way it works is we talk about distances and angles, but in huge dimensional spaces.” And they say, “Can you really visualize that?” At that point you have the choice of either lying and saying, “Oh yeah.”

Or what you would really say is something like this, you'd say, “Well, no not really.” The truth is not even close, right? But you'd say, “Well not really, but we use our pictures that we draw on two dimensional paper or in our offices, you get a grad Student: to say hold your hand up here, and then you get in there and you poke around this way.” You say, “We use that intuition from R_3 to guide us in making image compression algorithms in our one million, right?” So you might ask how useful is your intuition from R_3 ?

Tom Kover has a series of examples showing that you know nothing. He'll show you all sorts of things. They'll be true for $N = 3, 4, 5$, for six on they're so false it's amazing. Thing are like wow – I should collect some of these from him because they're really good, and there are about four of them. One is like if you pick a point at random in a unit ball in arc 20, it's basically; I think they call it a sphere hardening. There's something like a 99.99999 percent chance that you are within 1 percent of the boundary of the sphere. Just totally insane things, but okay, just to let you know, so that at least we were honest about it.

It's very important to understand what it means to have a positive inner product or a negative inner product or something like that. Is should say this, if X and Y are aligned, that means they point in the same direction, that says that the inner product is the product of the norms. If they're opposed, that means that the inner product is as small as it can be. If they're orthogonal, it means that there's a 90 degree angle between them. It means $X^T Y$ is zero. And it's very important to understand what it means to have a positive inner product or negative inner product.

So positive inner product basically means that the angle between them is between zero and 90. If you have two things, as you twist them out, the angle goes from 90, and the inner product is quite positive. It's as positive as it can be when they're aligned. As they

twist along like this, and when they get to 90 degrees, the inner product is zero, then the inner product starts getting negative, like that. So a very, very crude thing to say would be something like this. If the inner product between two vectors is positive, they point roughly in the same direction. And I mean really roughly. I mean they're not fighting each other. If you add them, that's actually one of the characterizations, that the norm of $X + Y$ is bigger than the minimum of the two. You make progress.

If you have a negative inner product, it means the angles between 90 and 180. This allows us to combine things like a half space. If you have a set of X 's who have a negative inner product where they're give vector Y , that's this. You draw the vector Y and the vectors that have a zero inner product with Y , basically those that are orthogonal to Y , that's the hyper plain with Y as a normal vector. That's here. This shaded region, these are actually vectors whose inner product is negative with Y . So they have an angle more than 90 degrees, that's sort of the picture. This is called the half space, and these things come up and all sorts of things. A half space is the solution of a linear inequality.

I'll start in on the material of the next lecture, which you've probably seen in one or two other contexts, maybe in the context of Fourier or somewhere. So we'll talk about orthonormal sets of vectors. The Graham-Schmidt procedure or when you right it out as a matrix, in matrix terminology it's called the QR factorization. And then we'll talk about various things we can show with this.

We start with the idea of an orthonormal set of vectors. So a set of vectors is called normalized if norm U_i is one. So normalized just means the length is one. And by the way, some people call those unit vectors. I don't know why. Other people call them direction vectors, vectors whose norm is one. So that's very common notation for it. Unit vectors is a bit weird because for most people unit vectors means e_1, e_3, e_5 , but still.

A set of vectors is orthogonal if different vectors in the set – another way to say this is mutually orthogonal. It says that any two different vectors in the set have a 90 degree angle between them. Okay? And it's orthonormal if both. Now, watch out because people will simply say, you will say on the streets, U_1 through U_K are orthonormal vectors. That is basically completely wrong. By the way, there's nothing wrong with saying U_1 through U_K are normalized vectors. That's cool because to be normalized is an attribute of a vector alone. To be orthogonal, it makes no sense to say – if it made sense then you could have your set of orthogonal vectors and mine, and we could throw them together. If you have five and I have seven, now we have 12 orthogonal vectors. It doesn't work that way.

Normalized vectors works, no problem, we can combine your normalized vectors with mine, and now we have twelve normalized vectors. Orthogonality refers to a set of vectors. It is not an attribute of a vector. Okay. How do you write that in vector notation? Very simple, if you have some vectors U_1 through U_K , you put them in a matrix like this. You just simply make them the columns of a matrix, call that U . And to say this, is to simply say that $U^T U$ is I , period. That's it. Why? Because $U^T U$ is U^{-1} transpose up to U_K transpose like this. And then you multiply by U^{-1} through U_K . And

you can see now, the II entry of this matrix is basically $UI^T UI$. That's one and the off diagonals are zero.

What would be the matrix way of saying that a set of vectors is orthogonal but not necessarily normalized? How would you say it?

Student:[Inaudible]

Instructor (Stephen Boyd): $U^T U$ is diagonal. That means they are – that's a mutually orthogonal set. How do you say – what's the matrix way of saying a set of vectors is normalized? How do you say it?

Student:[Inaudible]

Instructor (Stephen Boyd): In terms of the $U^T U$, it would be the diagonal entries of $U^T U$ would be all ones. How about the off diagonal entries? Who knows? That depends on the angles between them. That's what it would say.

Now, you can quickly show that orthonormal vectors are independent. How? Well, if you write something this way, and you multiply it by UI , you will get that I is zero. And that says, if you have an orthonormal set of vectors, U_1 through U_K , it turns out it spans our view. You have to be very, very careful here, because if I say that I have an orthonormal set of vectors, that's exactly the same as saying $U^T U = I$, exactly the same. This is the matrix way to say that.

Watch out because it is not the same as that. Although, that will hold in some special cases. Okay? In general, this is not true. An actually, there's a quick way to remember it. If you have a bunch of vectors and you stack them up. So here's three vectors in \mathbb{R}^{10} , so I write them this way, and I write this and so you get a 3×3 matrix here. Right? Like that. That's $U^T U = I$. Okay?

Now, suppose I got confused and switched them around and I wrote, $(U) U^T = I$. By the way, there's no syntax crime here. None because this would be a 10×10 matrix. This would be a 10×3 matrix. And that would be a 3×10 matrix, and there is no syntax crime. So a casual parsing doesn't reveal any problem with this. The parser goes right through it. Okay? But as a quick way to know that you're in deep, deep trouble, and that this, when you multiply two matrices together, the rank of this matrix is three. The rank of that matrix is three, and the rank of the product, therefore, is at most three. And what's the rank of the identity matrix in $\mathbb{R}^{10} \times \mathbb{R}^{10}$? It's ten, so we have a problem here. So this is a semantic error if you do this. It's a semantic error.

By the way, we're going to find out that that 10×10 matrix is not the identity. What it is is really interesting, and we're going to see it later. But for the moment, just remember it is not the identity. Okay, I'll mention the geometric properties, which are actually very interesting. Suppose you have an orthonormal set of vectors, which is to say a matrix whose columns are orthonormal. Ooo, that was slang, did you hear that? But I'm trusting

you that you are cool on that and aren't going to make semantic and other errors. That was slang. That's how, by the way, people would say it. Otherwise, you'd sound like a pedant right? If you're like, "Excuse me don't you mean the set of its columns is orthonormal." You shouldn't talk to people like that. You should avoid them.

So suppose the columns of U are orthonormal. It says the following, if you take $W = UZ$. It says the norm of W is the norm of Z . Now note that the norm of W is different – W is possibly a different size from the size of Z , so what this says is, when you multiply by a matrix whose columns are orthonormal, you don't change the norm. That's what it means. Another way to say it is, people say it this way, $W = UZ$ is isometric. "Iso" means the same, and "metric" means that it has to do with the metric or distance properties. And it basically says if you have two points like this, and the distance is three between them. If they get mapped by you, it says that the distance between if that's like X and Y , the distance between UX and UY in possibly a higher dimension will also be three. So it preserves these things.

It's very easy to show this. That's just a quick calculation. I think maybe this is a good time to quit for this lecture.

[End of Audio]

Duration: 76 minutes