

IntroToLinearDynamicalSystems-Lecture17

Instructor (Stephen Boyd): Welcome back. I've got to say, the schedule is a bit bizarre where you're off for a week and then come back for a week and a half, basically.

Let me go down to the pad here. I'll start by making just a couple of announcements. You might want to zoom in a little bit here. The first one is, just to remind you, the final is going to be next Friday to Saturday or Saturday to Sunday. And as always, if for some reason these don't work for you, just let us know and we'll figure out an alternative time for you to take it. Other administrative things are homework, so I guess homework eight, which you have now, right, and that's due this Thursday, I think. Is that right? Male Speaker:

Yeah.

Instructor (Stephen Boyd): So that's due Thursday. But we're going to pipeline a little bit and we're going to launch homework nine later today. That's after we figure out what it is and all that sort of stuff. So we'll launch homework nine later today and that's due next Thursday, that's obviously, the last one. Except, of course, with your final you'll be given a homework ten, which you can turn in during the first week of January.

Okay. Let's, if anyone remembers what we were doing, it seems, this is the topic we were doing before our quarter was interrupted by the schedule. We were talking about the idea of the gain of a matrix in a direction. So this is actually very, very important. This is sort of at the heart of the overloading $Y = AX$. If you have $Y = AX$ where A , Y , and X are scalars, it's, you don't, I mean, it's very easy, that's simply multiplication. When you have a matrix here, it's much more interesting. Because in fact, the amount by which the size of Y divided by the size of X , that's the gain of the matrix A in the direction X , this varies with direction. That's exactly what makes it interesting.

Now there are cases when it doesn't vary with direction. You've seen one. One is when A is orthogonal. So if A is square and its columns are orthonormal, then $Y = AX$, the norm of Y is the norm of X , the gain is one in all directions. So watch, you're going to get the complete picture today, on that.

So last time we asked the question, what is the maximum gain of a matrix. And we answered it. And the answer was well, it's, first of all it's got a name we gave it a name. The name is the norm of a matrix. And it is given by the square root of the largest, whoops, there we go, there it is the square root of the largest eigen value of $A^T A$. Now, you have to be very, very careful here, because A is not necessarily symmetric. In fact, it's not even necessarily square. $A^T A$, of course, is always square, however, and positive semi definite. It's nothing norm of a matrix and that tells you exactly what the maximum amplification factor of a matrix is when you consider it the mapping of $Y = AX$. But we know more. We know things like this, the maximum gain input direction is the eigen vector associated of $A^T A$, associated with [inaudible]. So that would be the input, which is amplified by the largest factor. The input

that is amplified by the smallest factor, input direction is QN . That's the eigen vector of $A^T A$ associated with [inaudible]. Okay. And of course, this can be zero, which is just a longwinded way of saying A has a null space. Because if you're in the null space, it means you've come in non zero, you come out zero; it means the gain is zero. So this is a quantitative idea that generalizes the qualitative of null space. Okay. This brings us to the whole story. And the whole story is the singular value decomposition, otherwise known as the SVD singular value decomposition. Now, it's been a while. Historically it's been around for quite a while, maybe 100 years, that's certainly 100 years. It was certainly well known in the '20s and '30s. It's been used in statistics. You'll hear other names for it. The one you'll hear most is PCA, which is principle component analysis. And there's probably a couple other words for it in other strange fields. That came out the wrong way. That was not to imply that statistics is a strange field, although, it is a bit odd, actually. So but anyway, so you'll hear other names for it. Certainly, in some areas of physics it's got some name, and I guess, in other contexts it's called things like the carbon and low ebb expansion, that's something else you'll hear about the KLX. You'll hear all sorts of names. But the main ones are singular value decomposition and principle component analysis. Okay. And the singular value decomposition is a triple matrix decomposition. By the way, we've seen a bunch of them by now. Well, not that many, actually, but we've seen a bunch, and it is very important to keep your matrix decompositions clear in your mind. We've seen diagonalization, that's something like $A = T \Lambda T^{-1}$. We've seen QR factorization. We have seen an orthogonal eigen decomposition. That's like $A = Q \Lambda Q^T$, and this is yet, another. By the way, there's lots of others as well, maybe not a lot but five or so other common ones. So here it is. It says that any matrix can be written out this way. It's $A = U \Sigma V^T$ where A is end by end with rank R . This intermediate matrix Σ is diagonal and its size is R by R . So it's exactly equal to the rank its size. And the singular values, that's the Σ one through ΣR , these are ordered, always, from largest to smallest. So Σ_1 is the largest and Σ_R is the smallest here and these are positive. U and V are both matrices with orthonormal columns. So $U^T U = I$ and $V^T V = I$. So that's what these are. So it looks like this, U and then Σ , which is diagonal, and then V^T can look any size, like that, for example look like, for example. Okay. That would be a typical way this might look. Okay. So this intermediate dimension is the rank of the matrix. Now, you can write this many other ways. If you write the columns of U as U_1 through U_R and the columns of V as V_1 through V_R , then you can light this new Σ transpose out as what people call as a dyadic expansion. So that's a dyadic expansion, which of course, means you're writing the matrix as a linear combination of dyads and the dyads are $U_i V_i^T$. Dyad is just a word for rank one matrix, also called an outer product, though, in other fields. Actually, there are no other fields. This is math. So – So Σ_i here, these are called the singular – these are the non zero singular values of A . So there's actually a bit of confusion and we'll get to that in a bit, and I'll warn you about that, as to whether singular values can be zero. In the definition I've just given you, the singular values are all positive period. So it doesn't make any sense to say $\Sigma_5 = 0$. But we'll see that, in fact, there's a lot of confusion on the streets about this and people do refer to zero singular values. But we'll get to that. Okay. Now, these V_i s are called the right or input singular vectors of A , we're going to see Y in a minute, and the U_i s are the left for output singular vectors. And

it sort of makes sense. If A is $U \Sigma V^T$, you can see that when you sort of operate, but when you form AX , the first thing you do is multiply it by $V^T X$. But you know what that is. That's expanding V in the VI basis. That's the first thing you do. Then you scale and then you send the output along the UI basis. So these are really the basis sort of for the input that actually comes through the system and the output. So we'll get more into that in a minute. But it's clear V is associated with input and U with output. That's clear. Okay. Well, these things are, right at the moment, they're quite mysterious but I think we can clear up the mystery very quickly. If you form $A^T A$ and A is $U \Sigma V^T$ times $U \Sigma V^T$, if you transpose this you get $V^T \Sigma U^T U \Sigma V^T$, Σ is diagonal and therefore symmetric, so Σ^T is Σ , then you get $U^T U \Sigma V^T$, that goes away, and you get $V^T \Sigma^2 V$. And what that says is the following: It says that $V^T \Sigma^2 V$ is an eigen vector decomposition of $A^T A$. So $A^T A$ is a symmetric positive semi definite matrix, if you write out it's eigen vector in respect to decomposition, you get $V \Sigma^2 V^T$, that's this thing, and so those are the input singular directions. And that says that these are eigen vectors of $A^T A$. It also tells you exactly what—now we know what the singular values are. They're no longer mysteries. The singular values are the square roots, σ_i is the square root of the i eigen value of $A^T A$. Okay. And that's for i equals one to R . Those are only eigen values of $A^T A$ that are positive, then they become zero. And in particular, σ_1 is nothing but the norm of A . It's the square root of the largest eigen value of $A^T A$. So σ_1 is the maximum gain. That sort of makes sense. We'll see why that's the case in a minute. We can actually see it right now, but I'll wait a minute. Now, if you multiple A by A^T in the other direction, you get AA^T , then you just multiply this out. This time the V goes away and you get $U \Sigma^2 U^T$ and what you find are the UI are eigen vectors of AA^T . So that's what they are. These are the output directions. And the σ_i , again, the square roots of the eigen values of AA^T , and here a question comes up, and now you have two formulas of the σ_i , which is the right one? Well, they're both correct because it turns out that the eigen values of, let's see, CD are the same as the non zero eigen values of DC , whenever these are both square matrices. Was there a homework problem on that or something? I have a vague memory of this. Yes. Okay. There was. Good. So in fact, AA^T and $A^T A$, which by the way, generally have different sizes, but the non zero eigen values are the same. By the way, this is super useful. I can give an example. If I ask you what are the eigen values of PP^T , or something like that, that's a symmetric – I don't even need that, I can just do it the other way around. Let's make a general thing PQ^T , there you go. So PQ^T is a rank one matrix and I asked what are the eigen values of PQ^T . You would use this theorem to answer it. Where you'd say, well, they're the same as the non zero eigen values correspond to the eigen values of that. That's a one by one matrix. I know what the eigen values of a one by one matrix are. It's just the number. So that says that the eigen values of PQ^T are $Q^T P$ and N minus zeros if that M by M . Okay. So I'm just giving you an example but anyway, that was just a little aside. Okay. Now, we can actually see what these U s are. Now, obviously, if you write A is $U \Sigma V^T$, if you plug anything into A like AX and some vector here, and it gets multiplied by U . So anything of the form AX is a linear combination of the columns of U .

Therefore, the columns of U , actually certainly, are a subspace of the range of A but in fact, they're exactly equal to the range of A , because I can make this thing be any vector I like, including like E one through ER here. Okay. So U one through UR are an orthonormal basis for the range of A . Now, V one through VR are an orthonormal basis for the null space of V orthonormal complement, not the null space of A . In fact, it's sort of the opposite of null space of A . These are the input directions that basically are not in the null space. That's one way to say it. So if you're orthogonal to V one through VR you're in the null space of A . So that's what this means. So this connects, once again, the all sorts of things you've seen before. So let's get some interpretations now of what this means, A equals U sigma V transpose. Well, it means this, you can write it out as a dyadic expansion or you can do it as a block diagram. As a block diagram, it works like this. It says take a vector N and the first thing you do is you partially expand it. Of course, the V s are not a basis, because there's only R of them. But it says you partially expand V , you calculate the V coefficients where I equals one to R , that's V transpose X , that's an R vector, this is R vector with the coefficients of X and a VI expansion, then you scale those by positive numbers. You scale this by positive numbers. That's all you do. That's diagonal, no interaction. Then you take those positive numbers and you reconstitute the output by multiplying by this matrix's columns orthogonal. Now, by the way, when you move from here to here, that's an isometric mapping. In other words, distances going from here, this is distance in RR , that was big R , big bold R , with a superscript little r , and then this is in our, I can't remember now, M , is that right? Yes. This is NRM. So this mapping is isometric. The distances are preserved exactly, angles preserved, everything here, norms, everything. Okay. So this is kind of the idea. Now, this looks exactly the same as if A is square it looks just like an eigen value decomposition for symmetric A . Actually, it's one difference. It is symmetric A , this is diagonal and real, but they can be negative, because you can have negative eigen values. And in that case, U and V are the same. The input and output directions are the same. Part of that sort of makes sense. That's what symmetric means. Symmetric means something like you can swap the input and output and it looks the same. Because when you transpose a matrix that's what you're doing. You're switching the roles of the input and output. So this kind of makes sense. So this is the picture. By the way, you can see all sorts of things now about you get a completed picture of the gain as a function of direction now. And let's see how that works. If you want to find an input here for which the output is largest possible, that's here, that'd be the maximum gain input direction, you would do this. You'd say, well, look, norm here is the same as the norm here. So if you want to maximize the norm here, you don't even have to worry about the U , just make this as big as it can be. So it says what you really want is you want this vector, then scaled by these numbers sigma one through sigma R , to be as big as possible. Now, when you put in a vector here of unit norm, here you're going to get a V transpose X is a set of numbers whose norm is less than one. I guess that's vessels theorem, or whatever, and I believe you had a homework problem on that, too. So because the partial expansion and the VI coordinates. How would you make this come out so that – what's going to happen to this vector is this vector's going to come out, it's going to be a bunch of numbers whose squares are going to add up to less than one, less than equal to one, and then they're going to be scaled by positive numbers. If you want that to come out as big as possible, you have to put all, all of this into the first component because it is going to get

the greatest amplification factor. And that means you want $V^T X$ to be E one. And there's only one way to do that. And that's to choose X equals V_1 . That's the first column of V . You can check this, you can verify this many other ways. I'm just explaining how this works. Okay. So that's the idea. U_1 one is the highest gain output direction. And of course, you have AB^{-1} and σ_1 and U_1 one. In fact, of course, you have AVI equals $\sigma_1 I$ times U_1 . So it basically says that the I 's singular input direction is amplified by gain factor σ_1 . The σ_1 's are gain factors. So let's see, actually, if we can figure out how this works. Am I missing something here, Page 30? No, I'm not. Here we go. All right, here we go. So let's take this four by four matrix. Obviously, that's so small you can figure out what happens yourself. Actually, that's on the boundary, to tell you the truth. Even a four by four matrix, you can't look at it, no person can look at a four by four matrix, except in totally obvious cases like when it's diagonal, and say something intelligent about the gain varies. I mean, you can say a few things, you can mumble a few things and wave your arms, but you really can't say much more. So you can imagine what happens when you have a 500 by 500 matrix. But anyway, let's take A is in our four by four, and the singular values are 10, 7, .1, and .05. So the first thing, we can all ask sorts of questions about what does that mean about the matrix and its gain properties. Well, the first is this, the rank of A is four, because you have four singular values. So A is $U \sigma V^T$, U , V , and σ are all four by four in this case. And what this says is that the gain varies from as small as .05 to it as big as 10. So that is a 200 to 1 ratio of gain. So if you want to know how isotropic is this mapping, in the sense of how much can the gain vary with direction, the answer's 200 to 1. We'll talk, actually later this lecture, about what that isotropy means and isotropy means, lack of isotropy. So the gain varies from .05 so it has no null space. However, there is an input direction, which is scrunched by factor of 20 to 1. And there's another input direction, which is amplified by a factor of 20 to 1. Okay. So we can be much more specific about it. We could actually say that if you have an input, if X , mostly, is in the plane given span by V_1 and V_2 , these are orthogonal, that's the two most sensitive input directions, if you're in that plane, then you get amplified by somewhere between seven and ten. That's what it means. Okay. So there's one plane of input directions for which this system roughly has a gain on the order of ten, you know, between seven and ten. By the way, the outputs come out along the directions span by U_1 and U_2 . Those are the two most sensitive output directions, or the highest gain output directions. Now, at the other side, there is another plane, by the way, orthogonal to the original plane that requires, by the way, some serious visualization skills, advanced visualization skills, since this is an R^4 . But you have another plane orthogonal to the original high gain plane, now you have a low gain plane inputs that along in this low gain plane they get scrunched by factor between zero, 5, and .1. Okay. And they're going to come out, well, they'll hardly come out, but when they come out, I mean, they'll be greatly attenuated, but they will come out along U_3 and U_4 span by that low gain [inaudible] plane. That's the picture. Okay. Now, depending on the application, you might say that A is rank two. That's going to depend entirely on the applicant. Obviously, it's not rank two. Mathematically it's rank four, right there. However, in some cases you might say that A is rank two. As someone said, it's sort of the gain, if this is some kind of communication channel or wireless system, that's a gain of plus 20 DB, that's like minus 26 or something like that. You might say, you know what, that's below the noise floor. That's worth nothing. Okay.

Let's look at another sort of baby example here. Let's take a two by two matrix. Now, for a two by two matrix you do not need singular value composition to understand what it does, that's for sure. That much is for sure. Okay. With the singular value of 1 and .5, basically what it says is you'd take something like X here, there'd be an input coordinate system, that's V_1 and V_2 , and you would resolve X into its components. That would be this and this. This V_1 gets multiplied by σ_1 and then comes out along U_1 . So whatever that is, if that's V_1 , let's say that about .55. Then you've got .55 times U_1 , where's U_1 , that's over here. This should be bigger, shouldn't it? A little bit. I didn't draw it right, I think. Well, anyway, it's about right. That should come out about here and then for U_2 you get, maybe, I don't know, .7 V_2 , .7 V_2 should get crunched by .5 and you should get .35 V_2 , and indeed, that's, at least visually, a third. So that's about right and that's the output. That's the picture. Okay. That's the singular value decomposition. It has spread, in fact, to almost all fields. Then any field that uses any math, any field that does any computation. So you will hear about it in different contexts, you'll hear about it in the statistics, you hear about it in the finance, you'll hear about it in basically anything. Signal processing control, wireless stuff, networking it goes on, and on, and on. So you'll hear all about it. So it's actually a very good thing to know. It's probably it's the last real topic for this class and it's very important and you'll be seeing it unless you – well, anyway, you'll be seeing it in many, many contexts. There's actually still some weird backward fields where it hasn't hit yet. It's coming to those fields. So there's a bunch more. I think, actually, let's see, the last couple of years it's sort of hitting fluid mechanics. So that's the one I know about most recently. Which is amazing, because it's this incredibly complicated very sophisticated field and they just didn't know about this and it's a big deal now because it explains all sorts of stuff there. Anyway, it's also used in tons and tons of other things. It's used in, let's see, it's used in the current standard for DSL, is basically straight out of the SVD, so all multi-input, multi-output wireless systems straight from SVD. It's a lot of methods for compression and things like that, straight out of the SVD. Another one, and this is – it's only a small stretch, but it turns out, in fact, things like Google's page rank. You may have heard this, straight from the SVD, the zero of order of the page rank, SVD. It's nothing else. So anyway, I just wanted to point out that this is actually an incredibly important topic. I want to say that because others, obviously, are much less important, for example, our friend, the Jordan canonical form. But we teach that just so you'd know about those things and can defend yourself, if needed. If you're caught in a dark alley late at night with a bunch of mathematicians. So, okay, who are pissed off. Okay. So let's do a couple of applications, but like I say, you go to Google and type this or some of its synonyms like PCA and there'll be literally millions and millions of hits. The whole field's based on this so you'll be seeing a lot of this. The first thing is, actually, I can now tell you the whole story on the pseudo-inverse. This is for any A not equal to zero. I've already fixed the notes, I just haven't posted it, but I just noticed this mistake this morning. So for any A non zero has a singular value decomposition so people don't really say that the zero matrix doesn't have a SVD or something, or it gives you a headache, I don't know, something like that. I mean, V and U would be zero by, they'd have to be zero by something. So any non zero matrix has a SVD. So if you have a non zero matrix and its SVD is $U \sigma V^T$, then the Moore-Penrose inverse is $V \sigma^{-1} U^T$. That's it. And you can check this agrees exactly with the two formulas for

the Moore-Penrose inverse or pseudo-inverse you've seen so far. So you've seen when A is skinny and full rank there's this formula, and you've seen when A is fat and full rank there's this formula. But this formula, here, works for any matrix that is non zero, any at all, and it basically does what U think U do. If you want to sort of invert this, you'd sort of make a V transpose over there. Sigma square invertible diagonal can be, of course, inverted, and then you do something like inverting U , it would be inverting U , if U were square, you'd make it U transpose and you get this. You switch the role of the input and output directions, invert the gains along these directions and that's what you get. Okay. So that's the general Moore-Penrose inverse. It coincides with these, in these two cases, but it's now generalized to other situations. And in fact, in the general case here's the meaning of it. So let's look at the general case when A is not full rank. Okay. So A is not full rank. If A is not full rank and you want to do least squares, least norm things get tricky. So let's see how that works. If A is not full rank the first thing you do is you say, well, how close can you get to Y with something of the form AW . This a least squares problem except now A is not full rank. And so up until now, you actually did not know how to solve this. You inform your trusted A transpose, A inverse and right at that point you have a problem because A transpose A is not invertible if A is not full rank. Okay. However, if you want to minimize this, there are still – you can form this least squares problem and ask how close can you get? Now, because A is not full rank the answer's not going to be a single point, it's going to be actually the whole set of points an affine set. So it's going to be the set of points for which AZ minus Y , the deviation is as small as it can be. That's an affine set. It's a set of least squares solutions. Now, in that set, you can then ask for the smaller, the minimum norm least squares approximately solution. And that's exactly what A dagger Y gives you. So this is quite complicated but it's worth thinking about and you should be able to sort of see what it does now. So this is the picture. So X zero inverse is the minimum norm least squares, and I really should put approximate here because I don't like writing, I don't even like, I'm going to fix that. I didn't even like saying least squares solution. Because the whole point is it's not a solution. I know that that's what everyone says, but it irritates me. So I'm going to change it. It's an approximate solution. Okay. Now, we can also do the pseudo-inverse of the regularization. So let's see how that works. So let's see how that's going to work. So let's let U be positive and then this is the regularized, this is ticking of regularization, or ridge regression you call it in statistics, or something like that, and here you have two things. You have sort of your mismatch and then you have something that penalizes the size of X . By the way, it's clear, this is going to be related to pseudo-inverse, because pseudo-inverse finds among the X s that minimizes this, it finds the one of smallest norm. So it's clearly related, these are clearly very close. Now, this is a perfectly good formula. It's simply A transpose A plus μI inverse, A transpose Y . This inverse is valid, period. A can be fat, skinny, full rank, it don't matter, it can be zero here, it's fine, no problem, okay, because the μI takes care of everything and this is always positive definite. Okay. Now, as μ goes to zero this limit gives you this pseudo-inverse. And in fact, it turns out if you want to get the pseudo-inverse this is another formula for it. It's the limit of A transpose A plus μI inverse times A transpose. So it's the limit. Another way to understand what the pseudo-inverse is, if you don't like the SVD, you can think of it this way, it's the limit of regularized least squares in the limit as μ goes to zero. Now, in some cases this is super simple. If A is skinny and full rank, then this makes sense when

Mu equals zero and it just converges to our old friend, this thing. But this formula's quite different in other cases. If A is fat and full rank, $A^T A$ is definitely not invertible. Okay. But $A^T A + \mu Y$, as long as μ is positive, is invertible. And this inverse makes sense and this will actually converge to this A value. In fact, in that case it would have converged, of course, to $A^T A$ quality inverse, which is the other formula for it. But the other cool thing is that even if A is not full rank, this converges to that, to the pseudo-inverse. That's what it is. So the pseudo-inverse is something like the limit of infinitely small regularization or something like that. As usual, the important part to know what it means not to know simply that it exists and so on. Okay. So now you know the full story on the pseudo-inverse. Now, like a lot of other factorizations, like the QR and I think we've seen a couple others, maybe just QR, you can extend, in factorization, by zero-padding matrices on the left and right. You do that to make certain a zero-padding some and then padding out other things in convenient ways. One way is to like, for example, fill out an orthonormal basis. So you can do that with SVD, too. So if you take – that's the original SVD, like this, so R is the rank, these are all positive numbers. What you do is you find a matrix U_2 , which is complimentary to U_1 , and you find a complimentary matrix to V_1 . So that's V_2 . Complimentary means that the columns of the V_2 are together with the columns of V_1 , form an orthonormal basis. And you know lots of ways to do that. One way is to run QR by appending, for example, the identity matrix after U_1 through UR that would do. So what you're going to get now is $U_1 U_2$ is actually now a square matrix and it's the output size M by M . And you'll get a matrix $V_1 V_2$. By the way, these look, when written like this, they're fat, they're actually square because U is tall here. Right, so this is actually a square matrix, although visually when printed it doesn't look like it. That's square as well. That's the input direction and both are orthogonal. So now what you do is to make up for the fact that you've just added a bunch of extra columns on U and V , no problem, you zero pad σ and you get something like this. Now σ is no longer square. It's diagonal in the weird sense that off the diagonal N equals zero. But generally people don't talk about diagonal matrices that are not square. I mean, like, I don't know why they don't, but they generally don't. And then these are just the zero padding the square. By the way, some of these could be zero, depending on A . So for example, if A is skinny and full rank, like this, it means it's rank R and that says that when you write it as $U \sigma$ and V^T , one of these is going to be square. I guess this one. That's your V^T . So in this case, if A is skinny and full rank, you don't have to actually add anything to V . For example, if A were fat and full rank, you wouldn't have to add anything to U . Okay. I hope I got those right. I might not of, but too bad. Okay. So you zero pad like this and now you can write, once again, A is $U \sigma V^T$. So you have A as this thing and it works out this way. And people call this the full SVD of the matrix. In this form what happens is these are actually orthogonal matrices U and V . But σ now has exactly the same size, as the original matrix A is diagonal, I mean, in the sense that's the only place where it can have – but it can also have zero entries. And the one we looked up before, if you want to distinguish it from this, is called the compact or the economy SVD. Oh, I should mention this, it's actually quite important in matlab, for example, if you just do this I think you get the full thing. Although I don't remember. Is that correct? Okay. And if you want the compact one, I think you write something like this. That's econ, the string econ, I think. Anyone have a laptop with them? I know Jacob

was, he's not here, so anyway. Okay. By the way, this is quite important. Let me mention a couple of reasons this is important. You don't want this – if A is, for example, I don't know, 100 by $10,000$ then just forget the computation, let's talk storage, here. So you can be in big trouble if you type this. If you type this it's going to return a V , which is $10,000$ by $10,000$. Okay. And if that's not big enough of a dense matrix to make trouble, you know, make this $100,000$. And now, for sure, you're in trouble. So unless you want the big one, you could actually be in big trouble. Here, this one would do the right thing. I would return something that was, if this was full rank, it would return 100 by 100 matrix, no problem, 100 by 100 diagonal matrix, definitely no problem, and in this case $100,000$ by 100 matrix, again, that would be no problem. Okay. But if you try this one it won't work. Okay. So I just mention this. But it is important. Okay. Now, here, by the way, is where the confusion on zero singular values comes in. So it's up there and let me explain it in this context. Now, let's just take a matrix A that's in \mathbb{R} like seven by seven. And let's say it's rank six. Okay. So the SVD, or the compact SVD, or economy SVD looks like this. It's U , that's six Y , and you get a six by six diagonal, and then you get a V transpose, which is going to be six by seven. Okay. V is seven by six. Okay. Now, this is what happens in that case. And the singular values are σ_1 through σ_6 . However, when a matrix is square a lot of people will talk about σ_7 , in this case, and they'll say it's zero. And there are actually applications where you want to talk about that. Because, in fact, what you might be interested in is what the minimum gain of this matrix. Well, it's rank six. It's got a null space. The minimum gain is zero. And then people will talk about V_7 . V_7 is, of course, nothing but, in this case, describes null space of A . So it turns out while σ_1 , V_1 , U_1 , these things are sort of unambiguous. But if someone writes something like σ_{\min} , you have to be kind of careful. You mean the minimum of the positive ones or you mean the actual minimum, in which case it would be zero or something like this. So and things like this you just have to ask because it just depends on the context as to whether a person's referring to the smallest positive one, or the smallest one period, one zero. Of course, if things are full rank or whatever there's no issue here. There's no ambiguity. Okay. So let's look at the geometric interpretation of how this works. So this is a full SVD. That's to make it simple to do the interpretation. You have $A = U \Sigma V^T$. We know exactly what orthogonal matrices do geometrically. They either rotate, or they reflect, or something more complicated, but equivalent. They preserve angles, they preserve distances, they preserve norms. Right. So it's an isometry. Anything related to distances, or angles, or whatever, is preserved by V and U . Now, like rotations. Okay, so this says to operate by A on a vector you carry out three operations. First, you multiply by V^T , then Σ , then U . So the first one is something like – and I mean rotate loosely in the sense of – of course, a rotation is given by an orthogonal matrix, but it could also be a reflection or some combination of the two. Okay. So the first thing you do is you rotate. Then you stretch along the axis by Σ . Now, you put in the differing gains. So far there's been no difference in gains anywhere because if you multiply by an orthogonal matrix it's actually complete isotropic in the sense that the gain is one all directions. Now you put in the gain and isotropy. So you put in the unevenness in the gains now. By the way, some people call this, I've heard dilate and, I don't even know how to pronounce this, but it might be dilatate. Does anyone actually know how to pronounce that? For a long time I thought it wasn't a word. But then I found it in enough

books that I realized it was a word, and then I thought it was like a weird variation. Maybe something like from the U.K. or something like that. You know just some weird thing. But it turns out no, it appears to be a word. I don't know if there're exact synonyms or not, but anyway. But you'll hear this operation when you stretch or scrunch something along the axis is a dilation. That makes sense to me. Dilatation, actually, I heard people say it. Even people who seem to know what they're talking about say it, so I suppose it is a word. Okay. You also zero pad or truncate, because when this middle action, here, actually is going to map something of different – it's going to map of different sizes. You zero pad or truncate, depending on whether the output is bigger than then input or smaller. So that's what you do. And the final thing is you rotate by U . So this means that we can calculate what happens if you have a unit ball. You can calculate exactly what happens. Take a unit ball. First, multiply a unit ball by V transpose. If you take a unit ball and you apply an orthogonal matrix to it, nothing happens. I mean, the individual vectors move. This point might move over here. But the ball is invariant. The ball is the same. So anything that's in the unit ball, if the image of it is the same unit ball. Not at the detailed level, not on a vector-by-vector basis, but on the set basis, it's the same. Okay. Now, you take this ball and you apply a dilation to it. And in this case, it's a dilation along the axis because you multiply one here, there's a singular value of two, and a singular value of .5. You take the two, that's a long the X one axis, or whatever, and you stretch it by two, and you scrunch it by a factor of two and X two. Okay. Now you take this thing and you rotate by this output, these output directions, and you get this. And you can see all sorts of cool stuff here. You can see, for example, that, indeed, you won with the largest gain output direction because the unit ball and all the things that went in here, were things of norm one. The biggest that any of them ever came out is that guy and that guy, by the way. So those are the two points that came out with the largest possible gain. And look at that, sure enough, they're aligned with U one. And what input direction, by the way, ended up here? What input direction was it?

Student:[Inaudible].

Instructor (Stephen Boyd): V one, precisely. So V one is some vector, here. Say that might be V one. V one got rotated to here or here, by the way. It wouldn't matter, it got rotated either to here or to here. And those are the two points, on this ball, that suffer, or on this sphere, that suffered the largest expansion here, and ended up here. Then it got rotated over here or here. So that's it. So it says that the image of the unit ball, image means you overload matrix vector multiply to apply to sets. You multiply a matrix by a set of vectors. You just multiple all the vectors and you get a new set. That's what this is. And it says that the image of the unit ball is an ellipsoid and the principle axis's are exactly σ U Y . Okay.

Now this allows you to understand exactly the game properties. And let's do an example. For fun let's do a couple of examples. In fact, let's start with this. Suppose the singular values are 5, 4, and 0.01, there's your singular values. What does the image of the unit ball look like? What does it look like?

Student:[Inaudible].

Instructor (Stephen Boyd): It's a pancake. Exactly! Slightly oblong pancake, but it's a pancake for sure. Everybody agree with that? And this says something like this. There is an input plane, let's say it's a mapping from \mathbb{R}^3 to \mathbb{R}^3 , it says there's an input plane, which is a plane where this thing kind of has a high-gain, like between four and five, and another gain, you might even call V^3 , you might even call that like an almost null space, because while AV^3 is not zero, that would qualify for full no space membership. It is not zero. But it comes out attenuated by a factor of 100. Okay. So it's something like almost a null space. In fact, this is exactly the kind of ideas you want to have when you do this. In fact, you'll find that all of these ideas about null space, range, all that stuff, it's wonderful, you have to understand all of it. However, in practice, these things are quite useless and quite silly, very silly, and very useless. What the SVD will do is it will give you quantitative ways of talking intelligently in the context of some application about things like null spaces, and ranges, and things like that. So here the range of this matrix is \mathbb{R}^3 . It's on two period. Up until, maybe, the third week of the class it's on two period. Now, in this context, depending on you know the details of the application, you could say that matrix actually has a range which is two dimensional. It's a lie. It has a three-dimensional range. But it's third direction is so feeble, the gain is so feeble, that, again, depending on the context, you might say for practical purposes that matrix is rank two. Okay. So this is what this is going to allow you to do. We'll see a lot of examples of this so this will become clear. Okay. So that's the pancake. What's this? What's the image? If the singular values were 5, 4, 3.9, what would the image look like? Just roughly.

Student: [Inaudible].

Instructor (Stephen Boyd): Egg, sure. Yeah. Maybe not even quite an egg. No, maybe an egg, I don't know. Something like an egg, right. So basically it says that there's one cross section that looks kind of like an egg. Maybe not even as much as a football or something like that, but an egg. That's that. Okay. How about this?

Student: [Inaudible].

Instructor (Stephen Boyd): It's what?

Student: [Inaudible].

Instructor (Stephen Boyd): A cigarette toothpick. Okay. I'll take toothpick. Maybe I miss heard that. Did someone say cigarette? Okay. Yeah. No. I'll take toothpick. Yeah. So it's a toothpick. That's what that is. And basically, actually, what you can say is, we'll see this quite precisely soon, you can actually say in this case that matrix is nearly rank one, depending on the application. I want to stress that. Matrix is nearly rank one depending on the application. So I want to stress that. So that's the idea. Okay.

So let's see how the SVD comes up in estimation and inversion. Y equals X plus V , where V is the measurement noise, like this, and X is something you want to estimate. Now, probably the best way to really get more deeply than we've gotten into this,

although, to tell you the truth, the stuff you know is going to work quite well in practice and go quite far. So what you would do is you might do something like this. You might do least squares, here. You might take Y , choose to estimate X , you would take something a dagger Y , okay, which would be $A^T A^{-1} A^T Y$, and you would say that that's the X that minimizes the discrepancy between what you actually measured and what you would have measured had the perimeter been X had, or something like that. Better models, well, maybe not better, I don't know, different models are obtained by looking at probability and statistics. So here you assume V has some distribution. But there's another model, and the other model says this, that this noise, I don't know anything about this noise, but I'm willing to give you a bound on its norm. Okay. So some people call that a norm bound model or something like that, and some people call it a worse case model or something, and it's a lot more forgiving than a statistical model. We're not looking at statistical models so it doesn't matter, but a statistical model basically says V is random. So V is not out to get you in a statistical model. In this one, V can be out to get you. V can include someone intentionally trying to mess up your measurement. So V can actually include things like somebody intentionally trying to jam you. It can include something related to X . For example, V could include quantization error, in this case. So it'd be another model for quantization error. Something like that. But anyway, that's the model where you know nothing about the noise except you have a norm bound on it. Okay. Now, let's put an estimator in, a linear estimator, and let's make it unbiased, that mean $BA = I$, and you know what that means, it means that if you form $X - BY$, if there were no noise, you'd get perfect reconstruction no matter what X is. So this means it's perfect reconstruction with no noise. Now, the estimation error is this. If you form \hat{X} , that's what you guess, minus what X really is, it's BV . Now this makes sense. We've seen this before. The key now, is to do this, among left inverses of A . What you want is a small one. Why does it want to be small? It wants to be small because B is what amplifies the noise. So B does two things. It maps a measured vector into an estimate, but it also maps measurement noise or error into estimation error. So you want B small. B can't be that small because, after all, $BA = I$. So for example, you can't take B as zero, as an extreme case, obviously. Okay. Now, we can then analyze, using this model, the set of possible estimation errors and that's an ellipsoid. Actually, there's a more refined ellipsoid you can get that's smaller but we'll just do this one. So here you can get a more refined ellipsoid but for now, we'll just this. We'll just say that if you get BV and V ranges in some unit ball, then in fact, there's an ellipsoid, which is BV , as V ranges over all vectors of norm less than α . This defines an ellipsoid, and we're going to call that B uncertainty ellipsoid. Note that has nothing to do with Y . Actually, that's the key to the more refined estimate. But this is fine for now to give you a rough idea. Then it says that this is where the estimation error lies, but an ellipsoid's original model is symmetric, so you can say X is $\hat{X} - X$ utility, that's $\hat{X} - X$, it's N , $\hat{X} - X$ the uncertainty ellipsoid, but minus and plus ellipsoid are the same because an ellipsoid is invariant under negation. And that says that the true X has to align an uncertainty ellipsoid, this E uncertainty, it's centered at the estimate \hat{X} , and you can say exactly what the insert into the ellipsoid is. B here, at the singular values of B gives you the semi-axis of uncertainty ellipsoid. So you want B , if, for example, these were the semi-axis of B , then we can talk about what it means. In this case, if these were the singular values of B , therefore, these are the semi-axis lengths,

when you multiple them by alpha of the uncertainty ellipsoid, in this case, very different. This one says your ignorance is pretty much isotropic. It means that when you guess \hat{X} , where X can be compared to where \hat{X} is, it's uniform, it's kind of your ignorance in this direction is kind of, you know, it's not too much different from your ignorance in that direction. And if you want, you can imagine a GPS system or something like that where you're estimating position. And it basically says, I guess if someone forced me to guess where I am, I'd say I'm here. But if someone says how far off are you, you say, well, I can be, you know, plus/minus three millimeters this way, it's about equal in all directions. Everybody see this? If, however, this is the story, it gets really much more interesting. This says well, in fact, this is in an estimation context this is singular values of B remember. This says, in that context you would say, well, in two directions I have uncertainty on the order of plus/minus four millimeters, let's say. In another direction I've nailed it. I have way, way better. Better, I'm much, much better, estimate of the uncertainty. So in this case the uncertainty ellipsoid is a pancake. Remember for an uncertainty ellipsoid small is good. This is better still. This says, in two orthogonal directions I have a very, very high confidence in my estimate. But in one direction, it's really way, way worse. So that's what is says. Okay. So that's the picture. Okay. So that's the idea. Okay. So you can look out, for example, the worse error is the norm of B , for example, and it turns out, in fact, that the best thing, again, by this measure, is actually using the least squares estimators. So if you use least squares estimator here, in fact, you're uncertainty ellipsoid will be given by this formula will be as small as it can possibly be. So if what you care about is uncertainty ellipsoid, or if you believe this analysis and so on, then once again, our friend least squares is the best estimate you can come up with. In other words, if someone says, no, I'm not going to use least squares I'm going to use some other fancy thing, you'll get an uncertainty ellipsoid with some other left inverse of A . You'll get an uncertainty ellipsoid that's actually bigger in all directions. That's what will happen. And that comes out this way, that's B least squares, B least squares transpose less BV transpose. And that's that. So once again, we see that least squares is actually a good thing to do. So let's look at this example, our navigation using range measurements. So here I have four measurements that estimate two positions and we look at two positions. One was the just enough measurements method where you take the first two measurements, you say, look, I've got two unknowns, all I need to two measurements, I'll use the first two, you don't even need the last two range estimates, don't even need it. So you form this left inverse like this. It's got a two by two block and a block of zeros. And then we'll compare that to least squares, which is \hat{X} is A dagger least squares. That's a two by four matrix. A dagger is this matrix that blends your four ranges into your position of your X and your Y position. Okay. And from the same example, here's what we get, this is where alpha equals one. In the first case, you get this. And in the second case, you get this. Now, what is correct here, is that if you were to translate this over so they have the same center, they'd lie on top of each other, in fact. Oh, sorry. This one completely covers that one. Okay. So that's the picture. Now, there is something a little bit confusing about this and I'll say what it is and it has to do with the fact that our analysis here can be sharpened a little bit once you have Y . This idea that this set, this uncertainty ellipsoid, the idea that this set contains X minus X utility, that's completely correct. However, this does not depend on Y . Note it's independent of the actual measurement you made. It turns out, once you have the measurement you can

actually get a smaller uncertainty ellipsoid. And it's not very hard to show. I won't do it now. But now we're going to explain why, for example, if one estimator says you have to be in the shaded set, another one says you have to be in this shaded set, then for sure you have to be in the intersection. And the intersection of those two is some weird thing that looks like this. It's this little guy here. Everybody see that? And then you'd say, wow, so although this was not a good method, it actually did give us information because it ruled out, for example, all these points over here that this model allowed you to keep. It turns out that if you do the more subtle analysis of uncertainty ellipsoid, you get something that's the same shape as this but it's smaller, like it's scaled down. Something like that. So, okay. I just mention that because it's, well, I confused myself this morning when I looked at this. That's what it is. This doesn't matter. The main point is you can understand uncertainty and how a measurement system works now by looking at the singular values of B . That's what you can do. Okay. I think what I'll do is I might actually, I'll skip this thing, which is just a completion of squares argument, that is a method to show, and if you have any left inverse it's going to end up with BV transposed bigger than the B least squares B least squares transpose. So I'm just going to skip that because it's just a completion of squares argument, because I want to get on to one other topic. Now, what we're going to do is – I guess, I can make some other comments about, I'll make just a couple more points about estimation and inversion. So in estimation and inversion, let's go back to that and look how that works. You know Y equals AX say plus V , like that. Now in some cases, in fact, I'll just tell you in a case where, I don't know, that I was involved in, and it was geophysical inversion. So X had a dimension around 10,000, something like that, and Y was a bunch of magnetic anomaly data, and it was something like 300,000 or something, if these numbers are not right, the story's going to make the right point. So 10,000 variables, 300,000 measurements like, no problem. In fact, that sounded like a 30 to 1 ratio of measurements to parameters you want to estimate. That's a pretty good ratio, right, 1.1 to 1 is not so great, 30 to 1, that's sounding pretty good. You've got about 30 times more measurements than you need. Everybody cool on this? Okay. So they'd say, no problem. A is full rank, by the way. And so you would imagine you would form this. You would form that. Something like that. Okay. And in fact, you can do this and you would get things. Things would come out and it would be shown in beautiful color that would show you the substructure, the presumed substructure and things like that. So that would be it. Okay. No problem. Now, this matrix A , I can tell you what its' singular values were. Okay, its' singular values were this. Actually, how many singular values does it have if it's 300k by 10k and it's full rank? So how many singular values does it have?

Student:

[Inaudible].

Instructor (Stephen Boyd): It's 10k, 10,000. So I'll just jump into the story. It had about 20 significant singular values. After that they dropped off precipitously and were way small. They were like below the noise floor. Okay. Everybody see what I'm saying? So the sad thing, at that point, people had to accept the following: They paid a lot of money to fly an airplane with this squid, you know, magnetic anomaly detector around. They got

300,000 measurements, 30 times more measurements than they had on those, 30 times. Which you would think would be way, that's a lot of measurement redundancy, 30 to 1. And it was very, very sad, because they thought they took 300,000 measurements and in fact, they didn't know it, but they actually took only 20. So they thought they were in the 30 to 1 measurement to data position, but in fact, they were in something like a 1 to 500. So what it meant was this \hat{X} , though very pretty, very nice when rendered in color, basically, was completely meaningless. Now, you can estimate about 20 perimeters in X if you wanted, but certainly not 10,000. Okay. So this is an example of the kind of reasoning or understanding you would have when you fully absorb the idea of SVD. And it's sort of this step beyond what happens when you first just sit around and just talk about rank, and range, and null space, and things like this. This matrix has no null space. There was no arrangement of subsurface, there's no X for which AX is zero. None. None. So in fact, without the noise you can get X exactly, in fact, by many methods. Because there's lots of left inverses of A . This is your understanding as of week 2.5. Okay. Now, it's much more subtle and it basically says for all practical purposes you found A was full rank, it was actually like rank 20. You have a question?

Student:

[Inaudible].

Instructor (Stephen Boyd): I'd like to say yes, but the answer is no, I don't know why. You know, I can make up a story though, about it. It'll be pretty good, too. Yeah, I could make up a story. I could probably even get the number 20, I mean, with a long enough story. Right. But no, I don't know why that is. It just is. So that's a very good question. Actually, it's because of that that you'd want to then – once you've come to the realization that although you have 300,000 pieces of data, you actually only have 20 measurements, independent measurements.

The next question is what are you going to augment it with? Do you want to drill a couple of exploratory wells, you want to do magnetic anomaly, gravitation anomaly, or do you want to set up some explosives. These are all getting other different measurements. And it would be interesting there to find out which one of those would augment these to give you. And all you do, completely trivial, is you add some more rows to A . Any measurement is a row to A . You add more rows to A and find out what happens to your singular values. And if you go from 20 non singular, you know, significant ones to 50, you just added 30 more measurements.

Student: [Inaudible].

Instructor (Stephen Boyd): Oh, absolutely! Yeah. The additional measurements, beyond the number of the perimeter you're estimating, that goes into smoothing, blending, averaging, this kind of thing. I mean, that is, the matrix we just looked at, I don't know if you remember it, I remember it, the two by four matrix, I mean, that kind of says it all. You should think of a left inverse in that case and actually, I love the names, sensor fusion is one, it's a great name, sensor blending. And basically, it's a fat matrix kind of

with all relatively small numbers. It takes all of those measurements in, scrunches them together, and gets very good measurements. So beyond any measurement beyond 10k, would have gone into averaging out the noise and stuff like that. I don't know if I answered your question. As a polite way, he said, no, that's what he said, but oh well, that happens. Okay. Our next topic is, also, a topic on which there are entire classes, like whole ten week classes, and it's a huge area. It's the idea of a sensitivity of linear equations to data error. So I'll say a little bit about that. It's not something we've looked at in this class. In this class, when you use matlab, for example, which in turn uses LAPACK, matlab does nothing, right. When you type $A \backslash V$, it doesn't actually matter what's happening, but something like the equivalent of forming $A^T A^{-1} A$ transpose times B is happening. And you know that that's not done in perfect precision. It's done with floating point numbers or whatever, and that means that the numbers that come out might not be exactly right. And you're used to that because, I guess, when you solve some problem and you look at the X is supposed to be equal to Y , you'd take the norm. You'd look at $X - Y$ and then entries would be on the order of 10^{-15} , and that'd be your hint that what you're seeing is just the artifacts of round off errors and stuff like that. That's a huge field. It's called numerical analysis. And we'll see a little bit of it now, about how this works. But singular value becomes positional central to the whole thing. So let's just do a square matrix, just for fun. $Y = AX$. Well, of course, that means $X = A^{-1}Y$, end of story. Not much more to say. Actually, there's going to be more to say, and it's actually quite interesting. Now, suppose you have a noise or an error in Y . For example, $Y = Y + \delta Y$. Now, in numerical analysis, δY is like a floating point round off error, for example. So there it's really small, right, it could be on the order of, you know, 10^{-10} , or, I mean, it depends on the context, but we're talking errors out there in the 10th, 15th, 20th digit. Right. Actually, in an engineering context, you want to imagine δY , it's generally much bigger. If Y is a measurement, for example, and you use a 12 bit ADC, that's only 4,096 numbers. So what that says is, the fourth digit, did I get that right, 4,096, yes, the fourth digit is absolute nonsense, if I have a 12 bit ADC. If I measure something with 12 bits, anything, it doesn't matter what it is, and then it says basically that whatever the true Y is, I don't know, but a δY was added to it that was on the order of one in 4,096. Okay. By the way, that's pretty good. A lot of high-speed stuff operates with much smaller precisions like 8-bits and things like that. Okay. So when you're doing engineering, if you're doing economics δY is probably on the order of Y , basically. I mean, it's like friends in economics say, but never on record, you know, when we talk about accuracy of things, they'll say things like, well, I'll tell you the truth, if we get this sign right, we're happy. But they'll never say that on the record. I don't know why. It has to do with the sociology of the field or something. Anyway, so they're happy if they get the sign right. That means that this δY is on the order of Y . But anyway, all right. Well, if Y becomes $Y + \delta Y$, X , obviously, becomes $X + \delta X$, where δX is $A^{-1} \delta Y$. That's simple enough. And that says that how much X has changed, because Y changed, is simply less – the worse case change is norm A^{-1} times norm δY . That's it. And what that says is the following, if the inverse of a matrix is large, and I mean large in norm, then it says small errors in Y can lead to large errors in X . By the way, it does not say does lead. I guess, it would be English, actually, but anyway. It does not say it must lead. That's not the case. For you to actually suffer

the largest increases here, ΔY has to be in the largest gain input direction of A inverse. That's what has to happen. Okay. But unless you can rule that out, it's possible, and it says you can get larger. And now, this says, basically, you really can't solve for X given Y with small errors. And that means A can be considered singular in practice. Okay. So I mean, we can do a quick example here. And by the way, this clears up something a bit odd, which I will explain now. Take a matrix that's singular. So it's N by N matrix and it's singular. Then you ask someone here's a measurement, Y equals AX . I've measured it. You know, this Y equals AX , and X and Y plus V , so there's a small noise. V is ΔY here. That's it. So that's your measurement system. And then you say, please estimate X . And someone who's taking the first two weeks of [inaudible] will say, you can't, your A is singular. Of course, I can't estimate X , it's ridiculous. In fact, you know if it's rank N minus one, there's a whole line of things with no noise, give you the exact same answer. It's impossible, so forget it. Get yourself a better measurement setup. Okay. So you go back into the lab, and it's a fact, actually, that if you'd take a matrix A and you add, if you pick sort of entry at random or if you add a small number to any random number to the entries, with probability of one, the matrix will be non singular. So you go to the lab. All you have to do is find some physical measurement thing, tap on it with a hammer. I guarantee you A is now non singular. So a very gentle tap, just tap it once, I'd say, it's taken care of, no problem. And they look at A , A changed in the fifth digit. And I say, check it, it's non singular, and it will be. By the way, it will absolutely be non singular. Okay. Now, and I'd say, go ahead, and then I'd walk away. But actually I should walk away very quickly because it's obviously not going to work. A inverse, A had a singular value just exactly zero. After I tapped the measurement system with a hammer, A had a singular value that was one E minus eight. Okay. A inverse now has a singular value that's ten to the eight. The norm of A inverse ten to the eight and that says that, if you go down the pad here, it says that small errors in Y can be amplified by as much as a factor of ten to the eight. Okay. So if you wanted say, a couple of digits of that, you know, if you wanted X to be accurate to like, you know, to .01, no problem, Y has to be accurate to ten to the minus ten. Okay. Which is not with a 16-bit A to D is going to give you or anything like that. Okay. So I mean, this all completely obvious. You know, it would be idiotic if taking a little rubber hammer and tapping on a measurement setup, you know, or if I told the pilot in the magnetic anomaly thing to just do another thing and I'd pushed on the control stick or something like that, and that made it non singular. Anyway, it was non singular there, anyway. But that doesn't affect anything. In fact, it's singular value analysis that tells you this. Okay. So what this says is that a matrix can be invertible, matlab, which is to say LA pack, will happily invert the matrix for you. By the way, if it gets too close to singular, then LA pack will issue a warning and say something like R con warning matrix is singular to working precision. At that point, you're so far out of it, it means that basically, with double precisions it's finding out. But the bad part for engineering, in fact, for all applications, is – I guess it's the good part, is that the really evil ones are when you invert it, you see entries like ten to the nine, ten to the twenty, you know something's fishy. But the worse part is when it's kind of plausible and it's giving answers that's plausible but completely wrong. That can happen, actually. Okay. So – all right. Now, I think I'll mention one more thing and then we'll quit for today. What you really want to do is it doesn't really make an sense to say that the norm of ΔY is – I mean, if I ask you, if norm ΔX is .1, is big or small, well, it depends

on what X is. If the norm of X is 100 then the norm of ΔX is .1, it's pretty good. That's like a .1 percent error. If the norm of X is .01 and norm ΔX is .1, that's terrible. That means that your ΔX is ten times bigger than the X you're interested in. So you should really be looking at relative things that develop errors. Now to do that you do this, you say, if Y is AX and norm Y is less than norm A , norm X , and therefore, if you work out ΔX over X this thing is the relative change in X . That's less than equal to the norm of A times the norm of A inverse times the relative change in Y . And this number comes up all the time. The product of the norm of A and the norm of A inverse, that's called the condition number and it's traditionally called κ of A . And it's the maximum singular value of A divided by the minimum, always bigger than one. And this is read informally something like this. It says that the relative error in solution X is less than the relative error in the input data times the condition number. So the condition number is what amplifies or can amplify relative error. So that's the picture. Or, again, there's a misspelling here. But in terms of, if you take the log two of this, and again, very roughly, you'd say something like this. The number of bits of accuracy in the solution is about equal to the number of bits of accuracy in the data, that's the log two of this thing, minus log two κ . So when you solve $Y = AX$ for X given Y , you lose accuracy. By the way, if the condition number is one you lose no accuracy. Matrices that have condition number one, I'll give you an example, would be an orthogonal matrix. They have condition number one. All singular values are the same, gain factor all directions equal. That's why people who do numerical work love them. And by the way, it's also why a lot of transforms you'll see, like DCT, FFT, or DFT, a lot of transforms you'll see, actually, are orthogonal. I mean, there's many reasons for it, but one nice part about it is when you use transform like that, and then the inverse, and stuff like that, you actually are not losing bits of accuracy. Okay. So this is the picture. So we'll continue from here next time.

[End of Audio]

Duration: 78 minutes