PROFESSOR: OK. Now where am I with this problem? Well, last time I spoke about what the situation's like as alpha goes to infinity. And I want to say also a word about-- more than a word-- about alpha going to 0. And then the real problems come when alpha is in between the real problem the situations these ill-posed problems that come from inverse problems trying to find out what's inside your brain by taking measurements at the skull. All sorts of applications involve a finite alpha. And I'm not quite ready to discuss those topics. Roughly speaking.

I'll write down a reminder now. What happened when alpha went to infinity? When alpha went to infinity, it enforced this part became the important part. So as alpha went to infinity you the limit was, u infinity, so I'll call it u infinity. Well so u infinity was a minimizer of this one, u infinity minimized bu minus d squared in fact in my last lecture I was taking bu equal d as an equation that had exact solutions, and saying how did we actually solve bu equal d. So u infinity minimizes bu minus d. But that might leave some freedom, if b doesn't have that many rows, if it's rank is not that big, then this doesn't finish the job. So among these if there are many and that's what were interested in.

u at infinity. That limit will minimize the other bit. au minus b square. Does that makes sense somehow? This is in the problem here, for any finite alpha, as alpha gets bigger and bigger we push harder, and harder on this one, so we get a 1 that's a winner for this one, but the trace of this first part is still around and if there are many winners then having this first part in there will give us the, among the winners, the one that does the best on that turn.

And small alpha going to 0 will just be the opposite right? This finally struck me over the weekend, you know like I could divide this quantity, this whole expression by

alpha, so then I have a 1 there, and a 1 over alpha here and now as alpha goes 0 this is the big term. So now u shall I call this u 0. There OK, for brilliant notation right, so this is this produces a u alpha. In that limit it convert is to a u infinity that focuses first on this problem, but in the other limit, when alpha's going to 0 is faster and then this term is biggest. u 0 minimizes au minus b square and if there are many minimizers among these well you know what I'm going to write u 0, right see I put a little hat there. Did I? I don't know I haven't stayed with these hats very much but maybe I'll add them. u hat minimizes the term that's not so important, bu minus d square.

OK, so today's lecture is still about these limiting cases. As I said the scientific problems, ill-posed problems. Especially these inverse problems, give situations in which these limiting problems are really bad, and you don't get to the limit, you don't want to. The whole point is to have a finite alpha. And but choosing that alpha correctly is the art, let me just say why, why so I almost, I'm started anticipating what I'm not ready to do properly. So, I'll say y a finite alpha, on Wednesday, y because noisy data. Even with, because of the noise at best u is only determined because of the noise up to some order, say order of, some small quantity delta that measures the noise. This is like a measure of the noise.

Then there's no reason to do what we did last time like forcing bu equal d. Then we don't, there's no point in forcing bu equal d if that equation, if the d in that equation has noise in it, then pushing it all the way to the limit is unreasonable, and may produce a very you know a catastrophic illness. So that's when, so it's really the presence of noise, the presence of uncertainty in the first place that says OK, a finite alpha is fine, you're not looking for perfection, what you're looking for is some stability, some control of the on the stability. OK, right. But now, so that's Wednesday.

Today let me, go I didn't give an example, so today, two topics, one is an example with bu equal d. that was the last lectures and that's the case when alpha goes to infinity, then secondly is something called a pseudoinverse. You may have seen that expression, the pseudoinverse of a, and sometimes it's written a with a dagger or a

with a plus sign. And that is worth knowing about. So this is a topic in linear algebra. It would be in my linear algebra book, but it's a topic that never gets into the 1806 course cause it's sort of a little late and that will appear as alpha goes to 0. Right so that's what today is about, it's linear algebra because I'm not ready for the noise yet.

But it's the noisy data that we have in reality and that's why, in reality, we alpha will be chosen finite. OK so part one then is to do a very simple example with bu equal d Here is the example OK. So this is my sum of squares in which I plan to let alpha to go to infinity. So a is the identity matrix and b is 0. So that quantity is simple. Here I have just one equation so p is 1 p by n is 1 by 2 I've just one equation u1 minus 2 equals 6, and in the limit, as alpha go to infinity, I expect to see that that equation is enforced. So there's two ways to do it, we can let alpha go to infinity and look at u alpha going toward u infinity, maybe with our little hats.

Or the second method which is the nul space method. Which is what I spoke about last time, the nul space method solves the constraint bu equal d which is just u1 minus u2 equal six. OK that's, maybe I'll start with that one, which looks so simple of course just to solve u1 minus 2 equal 6. I mean there everybody would say OK solve it for u2 equals u1 minus 6 so here is that there's a method any sensible person would use. But this course doesn't, OK the sensible method would be u2 is u1 minus 6 plug that into the squares and minimize. So when I plug this in, of course this is exact, and this becomes u, so I'm minimizing, minimizing u1 squared plus, what was it? u1 minus 6 square.

So let's reduce the problem to one unknown, that this is the nul space method, nul space method is to solve the equation and remove unknowns. Remove p unknowns coming from the p constraints and here p is 1, OK, and by the way, can we just guess, or not, guess but pretty well be sure what's the what's the minimizer here? Anybody just tell me what u1 would minimize that? Make just a guess maybe ? I'm looking for a number sort of halfway between 0 and 6 somehow. You won't be surprised that the u1 is 3.

And then, from this equation, I should learn that u2 is mine is 3 u2 not u2 infinity.

Now I've got too many u2 infinity is minus 3, anyway simple calculus if you just set the derivative, this 0 you'll get 3 and then you get minus 3 for u2, so that's the nul space method, except that I didn't follow my complicated qr orthogonalization. And I just want to do that quickly to reach the same answer.

And to say, why don't I just do this anyway? This is what? This would be the row, this would be the standard method in the first month of linear algebra would be to use the row reduced echelon form, which of course is going to be really, really simple for this matrix in fact that's already in row reduce echelon form. Elimination as row reduction is nothing to do to improve that and then solve and then plug in and then go with it.

OK, well the thing is, that row reduced echelon form of stuff you teach is not for large systems guaranteed stable. It's not numerically stable and the option of using-- of orthogonalizing is the right one to know for a large system.

So you you'll have to, allow me on this really small example to use a method that I described last time. And I just want to recap with an example on the small system. OK, so what was that method? So this is the nul space method using qr now. Now MATLAB command qr, so what did we qr of b prime do you remember that we took that the that's the MATLAB command that eventually will, or is actually already in the notes for this section, and those notes will get updated but that's step one in the nul space method.

qrb prime and this gives me a chance to say what's up with this qr out rhythm. I mean after lu, qr is the most important algorithm in MATLAB and the so what does it do? b prime the transpose of b is just 1 minus Right? OK. Now what does Gram Schmidt do to that matrix? Well the idea of Gram Schmidt is to produce or so normal columns so the most basic Gram Schmidt idea would say, so what would Gram and Schmidt say? They'd say well we only have one column, and all we would have to do was normalize it

So Gram Schmidt would lead you, would produce the normalize saying times square root of 2.

That would be the q, and this would be the r, 1 by 1 in Gram Schmidt. OK, but here's the point, that the q r algorithm in MATLAB which no longer uses the Gram Schmidt idea, instead uses a householder idea and one nice thing about this is that it produces not just this column, but another one, it produces a column for the, of the, it completes the basis to a full or normal basis.

So it finds a second vector so ordinary Gram Schmidt was just had one column times one number what qr actually does is ends up with two columns. And well everybody can see what the other column in that's has length 1 of course and is orthogonal to the first column and now that is multiplied by 0. So this is what qr does after, if we have this 2 by 1 matrix, it produces a 2 by 2 times a 2 by 1. And you might say, it was wasting it's time, because to find this part because it's multiplied by 0 but what are we learning from the vector? From this 1, 1 vector or 1 over square root of 2? 1 over square root of 2 vector?

What good can that do us?

It's the nul space of b. So b was 1 minus 1, So let me just so that's the connection with nul space of b. if I look to if I look at vectors there's my matrix b, and if I'm solving bu equal d, if I'm solving bu equal d, then u is u particular . And u nul space, and if I want u nul space, then that's where this and these whatever extra columns this might be p columns and then this would be n minus p columns that's what that's good for, and of course that column is tells me about the nul space which for this matrix is one dimensional and easy to find, OK.

So that may be just to, so you know the difference between Gram Schmidt's qr r which stops with if you had one column you end with one column, and the MATLAB house holder qr which finds a full square matrix. OK, which is good to know and here we've we found a use for it. Ok, So then the algorithm that I gave last time and I'll give the code in the notes goes through the steps of finding a u particular, and actually the u particular that it would find if I is happens to be 3 minus 3 happens to be the actual winner, and the, therefore, the u nul space that that algorithm would find if I went through all the steps you would see that because I'm in this special

case of b being 0 and so on.

That the vector that it would choose this is the basis for the nul space, but it would choose 0 of that basis vector and would come up with that answer. OK so that's what the algorithm from last time would have done to this problem. I also over the weekend, thought OK, if it's all true, I should be able to use my first method. The large alpha method and just find the answer to the original problem and let alpha go to infinity. Are you willing to do that? That might take a little more calculation but let me try that.

I'm hoping you know then approaches this answer, this is the answer, I'm looking for. OK so get your mind just suppose I had to do that minimization again, now I'm not using the nul space method, so I'm not reducing, I'm not getting u2 out of the problem. I'm not, I'm doing the minimum as it stands, and so what do I get? Well I've got two variables u1 and u2 . So I take the derivatives is expect u1. I'm minimizing, everybody, when I point, I'm pointing at that top line. So it's 2 u 1 and what do I have here? 2 alpha u 1 minus u 2 minus 6 equaling 0. Is that-- did I take the u1 derivative correctly?

Now if I take the u2 derivative I get to u2s? And now the chain rule is going to give me a minus sign, so it would be a minus 2 alpha, u1 minus u2 minus 6 equals 0. So those two equations will determine u1 and u2 for a finite alpha. And then I'll let alpha head to infinity and see what happens. OK, first I'll multiply by a half and get rid of those useless 2s' and then solve this equation.

OK, so what do I have here? I've got a matrix u1 is what multiplying 1 plus alpha, u2 has a minus alpha. And this line u1 has a minus alpha, u2 has a 1 minus, minus, plus alpha, am I right? Times u1 u2 equals-- what's my right hand side? I guess the right hand side has alphas in it. 6 alpha and minus 6 alpha I think. OK. two equations, two unknowns, that's what the these are the normal equation for this problem, written out explicitly and probably I can find the solution and let alpha go to infinity. You could say, what are you doing Professor Strang to this elementary calculation, but there is something sort of satisfying about seeing a small example

actually work. At least to me.

OK so how do I solve those equations? Well good question. Should I, with a 2 by 2 matrix. Can I do the unforgivable and actually find it's inverse? I mean, it's like not allowed in true linear algebra to find the inverse, but maybe we could do it here. So u1, u2, is going to be the inverse matrix, which is, so my little recipe for finding inverses is take the entries of, this entry goes up is up here, that entry goes down there, well you couldn't see the difference this entry stays the same but she stays in place those change sign and then I have to divide by the determinant. So what was the determinant of this 1 plus 2 alpha plus also squared minus alpha squared? When I get 1 plus 2 alpha and that's the inverse matrix now that's multiplying 6 alpha and minus 6 alpha, OK.

And if I can do that multiplication, I have well there's this factor 1 over 1 plus 2 alpha and what do I have? 6 alpha, 6 alpha squared, minus 6 alpha square it I think 6 alpha? 6 alpha squared minus is 6 alpha squared plus minus 6 alpha, I think it's that is 6 alpha and ready for the great moment, let alpha go to infinity, and what do I get? As alpha goes to infinity the 1 becomes insignificant, the alpha cancels the alpha so that's approaches 3 minus 3 So there you see the large alpha method in practice. OK.

And you see what, well it's something quite important here. Something quite important and it's connected with the pseudoinverse. The pseudoinverse so now, I want to, we got this answer. And what I want to say is that the alpha, the limiting alpha system has produced this pseudoinverse. So now I have to tell you about the pseudoinverse, and what it means, and basically. The essential thing that it means is, the pseudoinverse is it gives the solution u which has no nul space component. That's what the pseudoinverse is about. I'll draw a picture to say what I'm saying, but it's this fact that means that this part which was this number. Is the output. This is the pseudoinverse of b applied to 6, 6.

You see the point? v hasn't got an inverse. v is 1 minus 1 for rectangular matrix. And it's, it's not invertible, in the normal sense. I can't find a two sided inverse, a b

inverse, doesn't exist. But a pseudoinverse counts.

So just give a MATLAB as long as I've written a MATLAB command here. Why don't I write the other MATLAB command? u is the pseudoinverse. You remember that pseudo starts with a letter p so p i n v of b multiplying d. That's what we got automatically. And it's what we get, and the reason we got the pseudoinverse. So let me just say what was special here. What was special that produced this pseudoinverse-- that I'm going to speak about more-- was this choice. a equal the identity and b was 0 the fact that we just put the norm of u squared there-- well, the idea is this produces the pseudo inverse.

And if you like-- so can I say a little more about this pseudoinverse before drawing the picture that shows what it's about? So I took this thing and let alpha go to infinity. OK, so I could equally well have divided it by alpha the whole thing, if I divide the whole thing by alpha that won't change the minimizer are certainly the same u's will win and now I see while were out we're going to 0 and that's where the pseudoinverse is usually seen we take the given problem which does not completely determine u1 and u2 and we throw in a small amount of norm u squared and find the minimum for that, right.

So yeah. Let me say it somehow. I take, I take the b transpose b plus the 1 over alpha i, now 1 over alpha is still going to infinity in my, in this lecture. So 1 over alpha this is, the whole thing is headed for 0 times the norm of u square. This is the u1 squared plus u2 squared, OK and that inverse, that quantity inverse approaches the well, once I, I'm not giving the complete formula, but that's is what entering here and it produces leads to, may I see the vague word leads toward the pseudoinverse b plus.

Yeah and I'll do better with that. OK, I want to go on to the picture. OK, so right. Do you know the most important picture of linear algebra? The whole picture what a matrix is actually doing? Here we have a great example to draw that picture. So here's the picture that 1806 is, it's at the center of 1806. For our 2 by 1 matrix. So this is our matrix is b equals 1 minus 1 this is the picture for that matrix. OK, so that

matrix has a row space. The row space is the set of all vectors there are a combinations of the rows, but there's only one row, so the row space is only a line I guess it's probably that line.

So the row space of b of the, my matrix, is all multiples of 1 minus 1, so it's a line. Let's put the 0 point in. OK, then the matrix also has a nul space. The nul space as the side of solutions to bx equals 0, it's a line, in fact it's a perpendicular line. So this is the nul space of b, and it contains, oh what does it contain? All the solutions to bu equals 0 which, in this case, are all multiples of 1 1. And just to come back to my early comment, that's what the qr, that's the extra half of the qr algorithm is telling us, it's giving us and it's a beautiful basis for the nul space.

And so the key point is that the nul space is always perpendicular to the row space, which of course we see here. This z is what we had to compute when there were, p components and not just 1 and now where is, let's see, what else goes into this picture? Where are the solutions to my equation bu equal d? So my equation was my equation was u1 minus u2 equal a particular number 6 and where are the solutions to u1 minus u2 equals 6.

OK, so now I want to draw that. Where are all the solutions to the u1, u2 plane? OK, so one solution is take c equal to 3. 3 minus 3-- the combination 3 minus 3-- which is right there is my particular solution, so u particular, or u row space, is 3 minus 3. That solves the equation and it lies in the row space. And now if you understand the whole point of linear equations, where are the rest of the solutions? How do I draw the rest of the solutions?

Well to a particular solution I add on any nul space solution. The nul space solutions go this way. So I add on, well this is my whole line of all solutions, so this is the line of all solutions. And now they key question is which solution has, is the smaller? When, so this is the idea this pseudoinverse, when there are many solutions pick the smallest one. Pick the shortest one, it's the most stable somehow. It's the natural one and which one is it? OK which, so here is the origin. What point on that line is closest to the origin? What point minimizes u1 square plus u2 square?

Everybody can see this guy, that minimi-- so that's the pseudoinverse says wait a minute when you've got a whole line of solutions, just tell me a good one. Tell me the special one. And special one is the one in row space.

And that's the one, that the a pseudoinverse picks. So the pseudoinverse of a matrix, the general rule is, and so part of the lecture was the fact that, as alpha goes to infinity in this problem, the pseudoinverse will do it, or I could say just directly what is the pseudoinverse do? The pseudoinverse so b plus the pseudoinverse chooses, it chooses up if you like, up that's the b plus, that the solution I can't say b inverse d for everybody knows my equation is bu equal d. So this is my equation bu equal d. And my particular solution, my pseudo solution, my best solution, is going to be b plus d and it's going to be in the row space because it's the smallest solution.

So if you meet the idea of pseudoinverses now you know what it's talking about. We've got because we don't have a true inverse, we have a whole line of a solutions, would we want to pick one, and this pseudoinverse picks this one. It's the one in the row space, and it's the shortest because these are orthogonal because these are orthogonal u is up plus un and because those are orthogonal the length of u squared by Pythagoras is the length of up squared plus the length of un squared. And which one is shortest? The one that has no un.

That orthogonal component might as well be 0 if you want the shortest. So all solutions have this, and this is the length of a the shortest one. OK.

So that tells you what the pseudoinverse is. At least it tells you what it is for 1 by 2 matrix. As long as I'm trying to speak about the pseudoinverse let me complete this thought. But you saw the idea, the thought was there are there two ways to get it again. The nul space method that goes for it directly, or the big alpha method that we checked actually works. So that was a point of this board here. That's a big alpha method, it also produces in the limit as alpha goes to infinity up. And there's a little-- it doesn't have-- if alpha was a 1000, I wouldn't get exactly the right answer cause this would be 2001 in the denominator.

But as 1,000 becomes a 100,000,000 and alpha goes to infinity, I guess the exact one. OK, so here I was going to draw the picture, so if I draw row space, can I imagine this is a row space was dimension is the rank of the matrix, perpendicular to it is a nul space who's dimension is the rest the rank, and that was the rank that I always call r and this will have the dimension n minus r the number of this is exactly these are the two things that MATLAB found here. These where the r vectors in the row space turned into columns, and these were the n minus r, but that was only one vectors in the nul space. so normally were up in n dimensions not just 2 with two dimensions I just had lines in n dimensions I have a r dimensional sub-space perpendicular to an n minus r dimensional sub-space, and now b. What does b do? OK.

So suppose I take a vector un in in the nul space of b. Then b takes it to 0. So can I just draw that with an narrow this'll be 0. b u n is o, that's the whole idea. Ok but a vector in a row space, is not taking to 0, b will take that those are the into the I better draw here column space of b. OK which I'm drawing as a sub-space whose dimension is also, is the same rank r, that's the great fact about matrices, that the number of independent rows equals a number of independent column. So this guy heads off to some bu row space.

OK and if I've complete the picture, as I really should, there's another sub-space over here, which happen to be the 0 sub-space in this example, but usually it's here it's the nul space of b transpose. In that example b transpose was 1 minus 1 and that it's column was independent, so there was nul space. So I had a simple picture and that's why I wanted to draw you a bigger picture with it, it's dimension will be, well not n minus r, but if b is n by n, let's say, then it turns out that this nul space will be have [? dimension ?] n-r no problem.

OK now in the last three minutes I want to draw the pseudoinverse. So what I'm saying is that every matrix b, every rectangular square matrix b, has these four spaces. Four fundamental sub-spaces they've come to be called. OK and the nul space is a bunch of is of vectors would be takes to 0 B takes any vector into it's column space, so now let me just draw what happens to u equal u nul space plus u

row space. So this was a guy in the row space. if I backed up, b, what will b do it multiplies this vector This vector has a part that's in a nul space, and a part that's in the row space, but when I multiply by b what happens to this part? Gone when I multiply that by b where does it go? There, so this, all these guys feed into that same point. bu is also going there. That's why it's not invertible Of course. That's why it's not invertible.

Here I guess, yeah, you're up here I, so sorry a yeah. This was a nul space b. I didn't write in what was the nul space. OK, so the matrix it couldn't be invertible, and actually cause it has a nul space, and they all send those, so what is a pseudoinverse finally? last moment this pseudoinverse is the matrix it's like an inverse matrix that comes backwards. Right? It reverses what b does. What it cannot do, is reverse stuff that's appeared at 0.

No matrix could send 0 back to u n right? by multiply by the 0 vector I'm only going to get the 0 vector. So the pseudoinverse has to what it can do, it can send this stuff back to this. This is what the pseudoinverse does. if I had a different color chalk I would use it now but let use two arrows or even three. This is what the pseudoinverse does. It takes the column space and sends it back to the row space. And because these have the same dimension r, the point is inside, b is this is this r by r matrix that's quick, that's cool. It's totally invertible. And b plus inverts it. So from row space tp columns space goes b from columns back to row space, comes the pseudoinverse but I can't call it a genuine inverse because all this stuff, including 0. The best I can do is send those all back to 0. There now I've really wiped out that figure but I'll put the three arrows there that makes it crystal clear.

So this, those three hours are indicating what the pseudoinverse does. It takes the column space. It's column space is the row space. The column space of b plus is the row space of b. You know they're sort of and these two spaces that's where the pseudoinverses ally and b kills the nul space and b plus kills the nul space-- the other nul space. The nul space of b transpose. Anyway that pseudoinverse is at the center of the whole theory here. If I you know when I take out books from the library about regular rising these squares they begin by explaining the pseudo inverse.

Which as we've seen is arises as alpha goes to infinity or 0 whichever n were at and what I have still to do next time is, what happens if I'm not prepared to go all the way to the pseudoinverse because it's it blows up on me. And I after and I want to find finite alpha what should that alpha be? And that alpha will be determined by as I said by the somehow by the noise level in the system, right? And just to emphasize another example that I'll probably mention you know ct scans, mri all those things that are trying to reconstruct the results from limited number of measurements, measurements that are not really enough to perfect reconstruction, so this is the theory of imperfect reconstruction. Like if I can invent an expression, having met perfect reconstruction in the world of wavelets and signal processing this is the subject of imperfect reconstruction and I'll hope to do justice to it on Wednesday. OK thank you.