

Instructor (Stephen Boyd): So, let's start with a couple of announcements. I think we're a little bit behind on updating the website of the readings. So it should be kind of obvious. You should read the chapter on geometrical problems. That's kind of obvious, I guess because we're going to start that today. The other is you should read this appendix on numerical linear algebra. It's not very long; it's like 15 pages, just because we're gonna hit this material next week, and what's that?

Student:

[Inaudible].

Instructor (Stephen Boyd): Oh, it's more like 15, yeah, not 50. You wanna read this because the fact is you don't want to hear all this material for the absolute first time when I cover it like next Tuesday or something. So that's just a hint that we're not gonna be moving slowly on this material.

Okay, I want to finish up something about an experiment design which actually relates to our next topic, which is geometrical optimization problems, so experiment design, you recall from last time is this. You're gonna make a number of measurements and you have a pallet of possible measurements you're allowed to take.

That's V1 through VP, parameterize these things, and you're given a budget, for example, just a total number. You're just told, you know make 40 measurements out of this pallet of 50 possible measurements. Which 40 would you take?

And we talked about that last time. When you commit to a set of 40, which is actually nothing, but a set of integers on you know, M1 thorough MP that add up to 40. When you decide on your allocation, you'll be scored on a covariance matrix. That's your error covariance matrix.

Now, that's obviously a vector optimization problem because one person's error covariance matrix could be very small in one direction, big in another and then you could have the opposite, so that's kind of obvious.

The first thing is you have to scalarize. Then the second is there is this issue, it's a smaller issue, it's the issue of a relaxation from an integer problem to a continuous problem, and we talked about that last time.

So the most common scalarization is de-optimal experiment design. So in de-optimal experiment design, you minimize the log determinant of the covariance matrix. So this is a covariance matrix. This is, of course, convex clearly in the [inaudible] here. And this corresponds geometrically to minimizing the volume of the corresponding confidence ellipsoid.

This is sort of like if you're doing a problem involving positive definite matrices where you want it small, log determinant or determinant is something like the least squares there. It's kind of like if you can't come up with and defend another measure, then you might as well just go with log volume or something like that. So it's something like least squares.

Okay. So this is the de-optimal experiment design. Now, one of the things is when we start doing applications, it gets extremely interesting to do things like taking out a problem that comes up as an application. This is an experiment design. This is the optimal experiment design and you actually worked out a dual for this problem.

What's often the case is that a dual of a practical problem will actually turn out to have an interpretation, which is another practical problem, and it can be one from a totally different field or something like that. After the fact, everything will make sense, but it will not be remotely obvious at first that the two things are connected.

So in this case, a dual, which I'll work out shortly is this, so these two are duals, and of course, what I mean by that in the loose sense, what I mean is that I transform this slightly, formed a Lagrange dual, and then made a few more small transformations of that one.

Maybe I solved a [inaudible] variable. Maybe we'll get to that. This is called – so these are duals in the loose sense. However, they are duals in the following sense. Any feasible point here provides a lower bound for this problem period.

Okay, there's an easy transformation – if you solve this problem, you can solve that one and vice versa. There's zero duality gap here, absolutely zero in this case, because the weak form of Slater holds. Slater condition says you have to have a point that's strictly feasible.

Well, sorry, the strong form of Slater's condition hold here because I'll just take all the [inaudible] to be one over P, okay? And then the strong form holds, but anyway, so zero duality gap between these.

Now, this problem actually has a very simple interpretation, [inaudible] as a constant, so you're maximizing $\log \det W$ or if you like, you're minimizing say $\log \det, \text{ say, } W^{-1}$. That's the log of the volume with a constant multiplier of the half or something like that.

And then also an additive constant, but then an additive constant and a factor of a half or something like that, that's the log of a volume of an ellipsoid, and this would be $K^T W^{-1} D K$ less than one, and this is the problem of computing the minimum volume of ellipsoid that covers the points VK .

Okay, so I mean this sort of, by the way, this problem comes up a lot as, well. In fact, I'll mention statistical application of this right off the bat. Actually we'll get to some of that

in the next lecture, so maybe I'll wait. So what this says, and you can work out what complimentary slackness is. Complimentary slackness says that if [inaudible] is positive, that means you actually carry out an experiment of type is [inaudible] three is positive, that means you'll actually use B3.

You'll actually do a positive number of those experiments. Then you must have one minus VK transposed W VK 0. That means VK is on the boundary of the smallest enclosing, the minimum volume enclosing ellipsoid. So the picture is this – a very simple one. We looked at this last time, but very briefly. You have something like 20 choices of experiments to carry out here.

Okay, now these experiments are pretty much the same as these, I mean there's a negative sign; it doesn't matter. But the vector V is smaller and that corresponds to, you know, twice these measurements are the same as these and have twice the signal to noise ratio.

So it would be foolish to choose any of these and indeed, none are chosen. Now, over here, all the weight is put on these two, and this sort of makes sense because the only difference among these is the angle and given the choice of measurements, you would take measurements that are to the extent possible orthogonal. Okay.

So it's actually short of chosen, this point and this point and this is actually put a weight of .5 on each of – meaning if you were to do some, if you were to give an budget of experiments, half should be here and half should be here.

That's what it means and it kind of makes sense. There would be no reason to skew them because they're approximately, they're 30 degrees apart or 35 degrees. I mean they're as close as you're going to get to orthogonal. They're the most mutually – they have the largest angle and so they give the most mutual information or whatever, not any information [inaudible], but they're maximally mutually informative if you take ones that have a high angle, so that's what these are.

Now the interesting thing is the dual has a perfect interpretation. You take these 20 points and you compute the minimum volume ellipsoid that covers them centered at zero and that is, in fact, the one drawn here, and sure enough it touches two points. These two here, and these two actually are gonna be the ones you use in the experiment design, so that's another way to understand in the experiment design you're choosing points that are far.

Now, it's not as simple or as – it's not as simple as choosing the points that are farthest. Now, that would be a greedy scheme to choose the experiments that have the highest signals to noise ratio. That's not the case because in that case, if one sensor had a very high signal to noise ratio, you'd basically only take that one. You actually compute the minimum volume ellipsoid that covers these points and then these give you the other ones that fill in other directions and so on.

Okay. So that's that. And I think I'll skip the derivation of the dual, other than just to mention a few things about it. It should be straightforward for you now. There's no way you could look at this and just say, "Oh, okay." But, you know, it's something that in a few minutes you should be able to trace through. It's quite straightforward.

The only interesting thing here, I don't know if we've done enough problems on this or we've gone over this enough, is actually when you have a matrix problem like this, you actually have to calculate the gradient of the log determinant of inverse of a matrix.

That's complicated. The gradient is a linear function on matrices, and you actually have to, I mean don't just take, the formula is basically it's minus X inverse. But you have to be very, very careful to understand what that means.

That's actually covered in an appendix in the book, so I'd recommend reading that. In fact, let's add that to our list, and I don't remember what appendix this is. Okay, so it's the one that covers, let's say, that, what does it mean and so on and so forth.

Okay. All right, so how do you form, when you reformulate it this way and have a matrix constraint here, it's quite straightforward to introduce a Lagrange multiplier. Normally, it would be new transposed times you know, X minus B . Here it's the matrix equality. The transpose is an inner product. In inner product for matrices is traced in general [inaudible] Y . These are some metrics of trace XY . So you write this as Trace Z times the matrix residual here.

Okay, so that's the origin of that. Otherwise everything else kind of works out the same and I'll skip the details. Just to say the bigger picture is, basically for any problem you look at, and have time to actually think about carefully, it's actually worth your while to work out the dual because the dual often has a very interesting interpretation, give you some better idea about how it works and all that sort of stuff.

Okay. We'll move on to our last sort of generic family of applications and these are in geometry, so geometric problems. These came up actually all the time, and we'll start with some of the ones that are least obvious, the extreme [inaudible] volume of ellipsoid. So we'll look at that. These are some of the most interesting and least obvious applications. In fact, it's the least obvious what you can do in these cases.

So the L' owner John ellipsoid of a set C , which need not be convex, I mean, at all, is the minimum volume ellipsoid here that covers the set. Okay, so that's the definition of the L' owner John ellipsoid. Actually it's unique. We'll actually by formulating it as a convex problem, we'll prove that right away.

So here it is. We'll parameterize the ellipsoid as the – this is the inverse image of the unit ball under an affine mapping, okay? And that's a parameterization. I should point out there's at least like four or five totally different parameterizations of an ellipsoid, right and you can write it as a quadratic form. The forward image of an affine mapping of the

unit ball, you can write it as the sub level set of a quadratic function. You can write it all sorts of different ways.

Which of these methods you use to describe an ellipsoid, to parameterize the ellipsoid will completely change the convexity properties, and so when you have different problems, you're gonna have to choose one of these. So it's not like this we could not, this will not work out, if we parameterize for example, it this way, X minus XC , a set of X , such as that.

X minus XC transposed, say P inverse, X minus XC is less than one. That's another parameterization of an ellipsoid. And the data structure to describe this ellipsoid is a center point, XC , and a positive definite matrix P . Now, it's easy to go back and forth between this and this. It's just linear algebra, okay. It's easy; however, when you formulate the problem here, you'll end up with a non-convex problem for the L'owner John problem.

Here it's gonna end up being convex, so you have to choose the right parameterization. It makes a great deal of difference which one you choose. So this is the inverse image of the unit ball under an affine mapping. It's an ellipsoid. And you know, it's not hard to go from this to this. It's not a big deal. You know, $A^T A$ is P or something.

Now, in this problem here, it turns out you can assume without loss of generality that A is positive definite. And that's not at all hard to show. I mean one is this; it's got to be non-singular. If A were singular here, then it would, of course, have a null space and in fact, this would be called a generalized ellipsoid or a cylinder because anything in the null space it would have an infinite dimension here.

That's what this, it would have a direction along which was infinite. It wouldn't be bounded in that case. Nothing wrong with that, by the way. Sometimes that comes up. That corresponds to A being positive semi-definite and not positive definite.

Now the question is why can we assume A to be positive definite? Does someone have any suggestions about why? If someone gave you an A here that was non, not symmetric, not positive definite, how can I compute, how can I determine the same ellipsoid with an A that is positive symmetric and positive and definite? How do you do it?

Student: You just flip the side of the eig indice.

Instructor (Stephen Boyd): You could flip the side of the eig indice. A might not even be symmetric, but you're on to the right track. What would you do, so let's right the SVD of A . Let's write A here as U, σ, V^T times V^T less B , this is less than 1, all right.

So this is the set of V for which this is true; right? And I need a suggestion. This is A , and I want to write this in a new form where sort of the new A and the new B are symmetric, A is symmetric positive definite.

So they're all writing the skeleton. So I need a suggestion.

Student:[Inaudible].

Instructor (Stephen Boyd):What's that?

Student:[Inaudible].

Instructor (Stephen Boyd):You're gonna choose what?

Student: U equals V .

Instructor (Stephen Boyd):Okay, choose U equals V , yeah, I mean that's kind of what we want to do, but you can't say that. I mean somebody just gave us this description with A and B . Their A is not symmetric. So we don't have the right to say choose U equals V . They said, "No, no, no, this is the A I'm giving you."

So one thing you can do is this. This is a vector and if I multiply that on the left by any orthogonal matrix, it doesn't change the norm. That's the same. So I could multiply, let's just do this in our heads; let's multiply this vector on the left by first U transpose.

Student:Mm-hmm.

Instructor (Stephen Boyd):And then by V , and because I was multiplying by matrices, the first was whatever, so we're gonna write it this way. V, U transposed times U sigma V transposed V plus V . Everybody cool with that. I mean that's identical. This is equal to that because that's an orthogonal matrix. Everybody cool with that? Okay, and what do you have now?

Now we're done.

Student:You have V .

Instructor (Stephen Boyd):Yep, now you have this and everything's cool. This V , sigma V transposed plus then something here which is V, U transposed B –

Student:[Inaudible].

Instructor (Stephen Boyd):Yeah, I did, yeah, somewhere; where here?

Student:Yes.

Instructor (Stephen Boyd):There we go. Thank you. Okay. And I'm gonna call this A tilde and I'm gonna call that B tilde, and I'm done because I just took the original ellipsoid. I mean this would be the code you'd write at the top of something where you

decided in a group of people doing whatever development that the data structure of an ellipsoid is the inverse image of an affine mapping, okay?

However, you don't put in the specifications that the matrix should be positive definite. Okay. Somebody passes in this, but for your method, you need this. This is the stuff you write at the top that translates it to an equivalent one with positive definite A . Okay. Everybody okay on this?

Student: You could have [inaudible] right?

Instructor (Stephen Boyd): No, it's a singular value decomposition. It's not eig in value decomposition, singular value decomposition and it's non-singular with the original matrix A , so they're all positive, so we're cool.

Student: [Inaudible] A is not singular, then $\tilde{A}$ is. If A is singular –

Instructor (Stephen Boyd): Oh, pardon me, right if A is square and non-singular and therefore U and V are square orthogonal – they're orthogonal matrices.

Student: The whole point of this was that A was not squared.

Instructor (Stephen Boyd): No, no, no, no. The whole point – A is squared. Good question. Actually you can do this, by the way, if A is not squared, too, but you'd have to – you can do this in that case, too.

What A has to be in that case is, well, let's see, it has to be fat and full rank or skinny, mini fat and full rank, I think is what it – it has to be; sorry, skinny and full rank is the most general. If you want to allow people to describe an ellipsoid in this way in the most general way, A , I believe can be skinny and full rank. That'll do the trick.

Student: You're not losing information [inaudible].

Instructor (Stephen Boyd): Yeah, exactly, so in that case we're cool because and in everything I did here was kosher, so let's just, you know, in that case you have to check, go back and check that U has the right – well, I guess $U^T U$ was always cool. So, okay. Question?

Student: Why did we embark on this journey in the first place?

Instructor (Stephen Boyd): Well, I haven't said yet.

Student: Not the journey of the ellipsoid, the journey of positive symmetrizing the matrix?

Instructor (Stephen Boyd): Now, would we embark on a journey – well, we shouldn't answer that. I was gonna say, "Would we embark on a journey that was not needed?" The

answer is yes, we do that every day, but they're kind of weird things. I don't usually have it in writing, those weird little journeys and side trips. So obviously, this is gonna come up, right? Indeed it will in two lines, so, but thank you for putting the posting the sign there which is, you know, why are we doing this? Okay, so let's move on.

Student: Why does it have to be – so you're assuming A is not singular?

Instructor (Stephen Boyd): That's right – oh, why? If A is singular, this doesn't describe an ellipsoid; it describes something called the generalized ellipsoid. It's called a cylinder, in fact, because it will be infinite in some directions, in fact, specifically in the null space of A . Along the null space, any point in the null space of A can get infinitely big and satisfy this, and everything's cool; right.

So in that case, so that's not an ellipsoid. By the way, such a thing has infinite volume. And if we're minimizing volume, it's not gonna be something. It's sort of infinitely bad; its not gonna be of interest. So that's why.

Okay. Now, I'm gonna use this parameterization and yes, I wanted a symmetric positive definite for a reason and now you know why. Here is it. By the way, you see this problem here; this problem is actually correctly the L^1 John ellipsoid problem even when A is not symmetric; okay. It is because I assure you, $\log \det A$ is gonna be the volume, it's gonna be the volume while multiplied times the volume of the unit ball on that space. If that's gonna be the log of the volume, actually it's already added to that of this ellipsoid period, or one over it is, all right.

So, however, the log of the determinant of some non-symmetric matrix is some insane function that is not in general convex; in fact, it's never convex – just not. So however, if I assume A has symmetric positive definite, then this problem here is convex and that means we can solve it.

So that's it. So here I minimize $\log \det A$ inverse, and you have to check about the volume and things like that, but I mean that does transform – the volume transforms by the factor $\det A$ inverse here.

Now, to say that – I mean I can rewrite this way. I don't have to write it this way. I can write this, right, for all B and C . This just basically says that for every point in C , you're in the ellipsoid. You're in the ellipsoid if and only if $\|AB + B\|$, when you take the norm, is less than 1.

To say that the ellipsoid covers C is to say that for any point in C , $\|AB + B\|$ is less than or equal to 1. So that's this statement. Now, by the way, we know immediate that it's a convex problem.

Some people, by the way would glorify it and make it very complicated and call this a semi-infinite problem. Why? Because that is a convex constraint – in fact it's a two-norm constraint on the variables. By the way, the variable here are A and B . So actually

reading this kind of stuff requires serious concentration because it's a very standard convention that symbols like A and B and C and α and β are parameters.

And variables are things at the end of the alphabet like U , V , W , X , Y . So this is totally turned around and the variables are A and B and the parameters – well B is a dummy parameter, but nevertheless.

Okay, so you have to read this correctly. That is a two-norm constraint. It's nothing more, but it's 1 for every point in C . Oh, if C is infinite, this is a so-called semi-infinite problem. Okay. Now, you may not be able to have a tractable representation of this semi-infinite constraint. In some cases you can. I mean one is a finite set, so here I give you a bunch of points and you have to calculate the minimum volume ellipsoid that covers them.

Now, in this case, it's easy to say this because you just write that. This is a convex problem. So now you know something. And of course, it's not just the Lowner John ellipsoid for the set consisting of these points. But in fact, the convex [inaudible] those points and that's a polyhedron described by vertices.

So you just learned something which is by no means obvious, and that is that if someone gives you a polyhedron defined by its vertices, then computing the minimum volume ellipsoid that covers those points is a completely tractable problem. Okay. Everybody got that?

And in case you think, you know, sort of this is obvious and everything works and all that, watch out, because if I give you a bunch of points like this and I give you a polyhedron described by the vertices and I ask you for example, to solve a variation on this which is please compute the maximum volume ellipsoid that fits inside it, that's actually MP hard, okay?

So I'll show you, of course, I'll focus on the ones that actually we can do, but watch out. One step off this path and things get very complicated and these are not obvious facts. They're just not obvious period. Okay. Let me mention a couple of things about this one. This has huge applications, sort of choosing maximum volume ellipsoid around some points.

And let me sort of just give you one or two, they're beautiful applications. Here's a really very nice one. Suppose you just simply have V_1 up to V_n , doesn't really matter, you know, 10,000. It doesn't matter. These are measurements – they're vector measurements. They're in our tenth.

That's it; they're vector measurements. And what you want to do now is check are there any measurements that like stand way out. In other words, are there outliers. That's the question here. Are there gross – are there some measurements here that are big and somehow don't fit with the set; okay.

So how would you solve a problem like that? It's vague, comes up all the time. I want to identify outliers in a bunch of samples. What would I do? Make some suggestions, actually. Let's start with unsophisticated ones. So what would you – what would be the first thing you might do?

Student:[Inaudible].

Instructor (Stephen Boyd):Sure, yeah, absolutely, so the simplest thing is you might – let's subtract the mean of these from here. So no they're sort of centered roughly around zero. That would be reasonable. Then actually at that point, you might just look at the norms of all these things, and if all the norms were on the order of you know, half, one, two and a handful of them have the norm of 50, then there's your outlier problem solved, right?

So basically you have a cloud of points like this and a couple of just way out there, okay. Everybody agree? So that would be the first thing you might do. Then you might say well, no, no, that's not right. What I'll do is I'll take these points; I'll subtract the mean. That will center them and now I'll calculate the empirical covariance matrix of them. That gives you the ellipsoid that kind of covers the cloud the best. Everybody cool on that?

Then I might even change coordinates so that ellipsoid was uniform, was a ball. That's a very hard change of coordinates – the one I mentioned before. I don't recommend it without a lot of practice.

So you change coordinates and the ellipsoid becomes a ball, came out right. And now it means roughly speaking the data kinds of varies equally and now if you start seeing points that are way out these could be your outliers. Everybody okay on that?

Now, the problem with that is that that one – well, that would actually, that would, sorry, that would actually work – that would probably work okay. Turns out a much more sophisticated method than that is actually to compute the minimum volume ellipsoid that covers these, okay.

The points that are on the surface of that ellipsoid actually are ones which you would declare as candidates for outliers, okay? And in fact, the way this often works is you take those points and you remove them and then you redo whatever you were doing, some least squares problem, some singular [inaudible]. It really doesn't matter; you do the processing.

What you're looking for is something like this. You want to remove a small number of points, do whatever signal processing or statistics you were doing before and have the answer all of the sudden get way better; right?

So you want to remove eight points from the set of 10,000, do some signal processing and all the sudden have like an excellent fit. This is a hint that at least that those eight

included the ones that were messing you up before. Everybody see what I'm saying? That would be outlier [inaudible].

Okay, so this is called – the method. This is called in statistics – it's called ellipsoidal peeling. It's the greatest phrase, right, so you have a giant pile of data. You stick an ellipsoid around it and you remove those points. Once you remove those points, by the way, if you're quality of estimation of whatever signal processing or statistics you're doing didn't get like really good, you do it again.

You've removed those points and the ellipsoid shrinks down to fit the next one and this is called ellipsoidal peeling and you do this in waves or whatever until you like what you get or until you have no data points anymore, in which case you went too far. Yeah?

Student: I was gonna ask a question. Let's say you do it, you peel like two or three layers and you don't see any substantial change. Does that tell you that maybe, you know, you didn't have any variable outliers before?

Instructor (Stephen Boyd): Yes, it might tell you that, yeah. Generally, when that happens, when you do a couple of cycles of ellipsoidal peeling and it doesn't get better, generally you don't admit to people that you ever did that.

So yeah, you just go back and say, "It's not working." That's what you do, although you're prepared if someone says, "How do you know that there's not – maybe there's just like five or six outliers, just like bad data points that are completely messing up your, you know PCA or whatever you're doing, or your imagery construction or whatever it is?"

That's how, and then you'd say, "Well, you know, I did do some, I did some ellipsoid peeling. And nothing got better." By the way, we've already talked about it last time, but what's another method for dealing with outliers – I mean just grossly? What?

Student: [Inaudible] different norm?

Instructor (Stephen Boyd): Yeah, you'd go to an L1-type norm or a Huber norm, or something like that – Huber is not a norm, but you'd go to a robust norm. By the way, you could do the same thing there, the peeling thing.

Student: [Inaudible].

Instructor (Stephen Boyd): Take all of them out. You take all the ones that are on the boundary. If you don't want to take as many out – if this is 10,00 and these are an R10, I forget how many are gonna be on the boundary, but it's generically something like 50, so 50 out of 10,000.

Student: [Inaudible] do it again.

Instructor (Stephen Boyd): Yeah. I do minimum volume ellipsoid. I think generically about 50-55 or something like that would be on the boundary. By the way, if you take out the 55 and all of a sudden, like your image appears and that's like somebody's head. You're doing medical imaging, then you know the following, you know that A, there were some outliers, No. 1 and No. 2, they were within the 55 year removed.

Now, if you care, you can go back and start putting points in one at a time until the head goes away, okay, and then you say, all right, that, you know, so if you care, but on the other hand if you have 10,000 measurements and you throw away 55, you know, it would seem unusual if suppose instead you threw away 22 outliers and whatever, 33 non-valid data points.

Loss of 33 points, I mean I'm just talking – this is very hand-waving, but loss of 33 points out of 10,000 or I should say 9,950 is not gonna really make a big difference, so that's what you do, so you can add them back in or whatever you like.

These are actually fun methods. They can be shockingly effective, and I should add actually on the whole issue of outlier detection – I should make one comment about it. When an outlier is like totally off scale, anybody can do it. So if like you're getting a bunch of measurements, you know, and a bunch of them are like you know, 1, 2, 3, minus 5, minus 6 and all the sudden, one is like 1E plus 37, okay. You don't need advanced methods to detect that as an outlier.

That's because a bit flipped in a floating-point representation or something, you know, in the exponent. You don't need this for that, okay. On the other hand, if the outliers, you know, basically don't – if an outlier is just like it's just off a little tiny bit. It's just like some like 1 over [inaudible], and you know, it's not Gaussian, but it's not huge, not big enough to mess up your algorithm or you have enough data that it gets averaged out nicely, then you don't have to remove it.

But there's a band of situations right in the middle where being smart actually is gonna pay off. It's that band in the middle, it's when the outliers – the outliers are small enough well they don't really hurt anybody. If they're big enough, anybody can spot outliers, but there's a pretty big band right in the middle where being smart is gonna make the difference between successful signal processing statistics or whatever this is, whatever application it is and not. So that's just my comment on this.

I guess this is silly – that's the case for all problems, this band. Yes.

Student: [Inaudible] continue all the measurements. It's just to [inaudible]?

Instructor (Stephen Boyd): Yeah, so you're removing outliers and so the argument, so first I'll tell you the truth. The truth is say you're allegoristic, so it cannot be defended; however, you can say in its defense, the following: if I simply moved, it removed the values of V with large norm, that would be something that is not scale invariant.

If you were to rescale the data or transform the data, I would remove a different set because I can make a difference out of points look big. Okay, the interesting thing is if you do minimum volume ellipsoid, that's actually affine invariant.

You'll select the same point because think about it, if you have a bunch of points, and you transform them, you know, like by any linear mapping, you wriggle them all around and you re-calculate the minimum volume ellipsoid, you get a commutative diagram. So the minimum volume ellipsoid that covers a set of points is affinely invariant. It commutes with an affine change of coordinates.

Therefore, if someone says why are you going to the trouble, why not just take the norm and remove the ones with the largest norm, right, you would say, I'm doing something a little more specific, I'm minimizing the one that is large, but my measure of large doesn't depend on the coordinates used or the scaling used; it's more sophisticated.

Funny, if you draw that story out long enough, people won't notice that it actually doesn't answer the question, which is why. It's merely a justification. Okay.

All right, so that's an application of this. There's sort of a dual problem, that's maximum problem inscribed the ellipsoid. So here I have an ellipsoid – we'll get to this. Here we're gonna describe the ellipsoid, by the way, as the forward image, not the backward image of – it's the forward image of a unit ball under an affine mapping.

And again, I can assume, I won't go through the details that B is positive definite here. Now the volume is proportional to $\det B$ and the problem is to maximize $\log \det B$ subject to this.

What this says is another way to say this. It basically says that the [inaudible] over the norm ball of BU plus D is in sync, so that's another way to say it. If you want the semi-infinite representation, you would write BU plus D is in C for all norm U less than 1.

Now we know, by the way, that is a convex constraint on B and D . Why – because for each U in the unit ball, this is a convex constraint. And remember what the variables are here. It's B and D . This is not obvious. U is fixed. That's an affine expression. It's affine in B and D . Actually, it's linear in B and D , in fact.

And so you have the constraint that a linear function should be in a set. Sorry – this is the case of C is convex; sorry. If C is convex, this is a convex constraint. Okay.

Now, you still have to be able to evaluate this to make this interesting, and there's actually a very simple case when you can and that is when this simplest case is when this thing – when the set C is a polyhedron because I can tell you whether an ellipsoid described this way lies inside a half space.

That's easy. And the way you see that, that's an easy calculation. We want to know is it true that BU plus D – let's write a half space, so the half space is the set of the X such

that, let's say, $F^T X$ is less than G , okay. That's our half space, and I want to know is it true that $BU + D$ is in H or all $\|U\| \leq 1$. That's the question, okay.

And the answer is easy. Yeah. We can calculate that. The question is whether $F^T(BU + D) \leq G$ or with a question mark like that for all $\|U\| \leq 1$. Everyone agree?

Now, by the way, we're focusing on this as a function of B and D . We're getting there. Now, the second term is this, $F^T D$. It has nothing to do with U like that, and here I'm gonna write this in a different way. I'm gonna write this as $B^T F$, U^T . And this should be true for all $\|U\| \leq 1$.

Now, if someone walks up to you and says, and you form an inner product with a vector and says, "You know, I need the inner product of a vector with U , U ranges over the unit ball. We know exactly what these numbers range over. They range between plus/minus the norm of $B^T F$."

Okay, in fact to be quite precise, that's [inaudible], but in the general case, general norm, it's plus/minus the dual norm, the dual of that norm. So the maximum value this can have is this period, and that is with the choice U equal $B^T F$ divided by norm $B^T F$. That's the worst thing that can happen. So these are all if and only ifs. I haven't drawn them, but it's that. Okay.

And then you have that, and that's it. We're done. Okay, because now this is tractable. That is a convex function of B . This is a convex function of something happened there – D , thank you. That's a linear function of D . The whole thing – this is clearly convex and B, D . Okay, so that's the condition and that's what we have here.

Maybe I have it wrong here, or did I assume B is symmetric – yeah, sorry. I assumed B is symmetric, so the transpose isn't needed here. Okay, so I get that. By the way, just let me point out, you're used to it. I don't know – we're six weeks into the class. There's something like that. It's not a big deal to look at that and say that's a convex problem in B and D .

These are things that if you don't know this material, that looks like a hopelessly complicated problem, I mean hopelessly complicated just because, you know, it's got the log of the determinant and all that. You wouldn't know anything. This is non-differentiable; you know that. The norm – that's not norm squared; that's norm, right, but you look at that and say that's like a second order of constraint. It's nothing, the second order of constraint.

So all right, this allows you to calculate, to compute efficiently the following. This says that if you have a polyhedron described this time not by the vertices, but by the inequalities, so basically you have these things, $Ax \leq b$, like that. You can now calculate efficiently the maximum volume of ellipsoid that sits inside it; okay.

And that you can do efficiently. Now, interestingly, if I take a polyhedron described by inequalities and I ask you to calculate the minimum volume ellipsoid that goes, you know, the minimum volume that covers it, that's MP hard, again. So you have to watch out with these things. They're not – I'm just casually mentioning them here. This is not simple. Things that are – you know, easy problems and just minor variations on them are very hard.

Question?

Student: Is the MP hardness because you're using the representation of the polygon, which doesn't actively constraint?

Instructor (Stephen Boyd): Something like that.

Student: [Inaudible].

Instructor (Stephen Boyd): Yeah, MP hardness is actually very easy to see. It turns out, you wanna hear something hard, is just the following. Forget finding the minimum volume of ellipsoid that covers a polyhedron described by inequalities. They this much easier problem. Suppose someone just asked you the following.

Here is a specific ellipsoid. Can you please tell me if it covers my polyhedron? That's the feasibility – not even the feasibility problem. That's not finding the polyhedron that's feasible. It's basically check if a specific polyhedron is feasible.

That's MP hard. Turns out it's maximizing a convex quadratic function over an ellipsoid. And I can easily embed that to all sorts of MP hard problems.

Yeah, but I don't know that I really answered your question. I don't know that I really have a good answer for why is it easy to go from – if the data structure for the ellipsoid is vertices, why is it easy to calculate the minimum volume in covering ellipsoid, but hard to calculate the maximum volume 1 inside. And why does it reverse here? I don't know; I don't really have a good story for that, so I don't know. Question?

Student: So given this description, for example, we have this [inaudible] can you actually envision and find vertices?

Instructor (Stephen Boyd): Excellent question. Now, of course, if you could efficiently go, if you could transform between these two data structure to describe a polyhedron, then of course, right here, right now, we would have shown $P = NP$. Now that – I'm not ruling it out. Could happen someday, maybe around Week 8 or something like – no. But it – no, I'm glad you pointed that out because in fact, there's a catch there.

So the question is given a polyhedron described by inequalities, can you describe it by its vertices, and the other way around? And in fact, these are just two different data structures, to describe the polyhedron, right. That's no problem. Now it's interesting

because, for example, ellipsoids, we were just talking about multiple data structures describing ellipsoid. Those are all equivalent. They're all about the same size. They all involved like some sort of a matrix and they involve a vector.

You know, roughly, you know, in that – you know, transforming from any one to any other is just linear algebra. That's all very straightforward and everything is polynomial. It's all like N^3 or whatever, right.

So those are sort of equivalent, they're different, but equivalent parameters; they're polynomially equivalent parameterization. Now, in general, there's some pretty shocking things, I'll just mention some of them. If you have a polyhedron described by inequalities, it is possible to list the vertices. It's possible. There's algorithms that just work, just lists them.

Here's the problem. The number of vertices is exponential in the problem size. Okay, so even, you know, if you have a – I mean things you can just write down in a minute, you could just write down, for example, take 100 dimensions, 200 linear inequalities, okay.

So that's nothing, right. You can solve an LP of that size, well, in three weeks, you'll write a code for it, and that can be solved literally in milliseconds. Okay. So you know, everyone has the right visualization skills. When I say minimize [inaudible] where A is 200 by 100, we're all visualizing the same thing. We're visualizing a polyhedron and a direction and you kind of find the vertex or whatever that's most in that direction. Everybody got that?

Now, here's the problem, the number of vertices of a polyhedron in R^{100} defined by 200 linear inequalities is absolutely immense. I mean it's some number that, you know, it's on the order of like 100 factorial or something like that. I don't know what it is, and by the way, if I'm not quite right, and that's not equal approximately to the number of subatomic particles in the universe, I have to change those numbers very slightly and it will be. Trust me, okay.

So that's the problem. It's the going from inequalities to vertices can lead to a huge increase in the size of the problem. That's the first problem. Now, by the way, for certain cases it's not. I mean if everything is an R^2 , anybody can figure out a way to transform between these two representations, vertex versus face or in that case, edge.

Anybody can do it in R^3 , as well, and if you go to Google and type things like transform vertex face blah, blah, blah, you'll find tons of papers, but it'll be for specific cases like R^2 and R^3 is what you're gonna find.

You'll even find for example, in some RKs that they'll be polynomial algorithm. But K has to be fixed, the dimension, so you'll even see that. By the way, the other way has the same problem going from one to the other. So anyway, let me summarize all of this because there was a weird, random walk, so let me summarize it.

The change in data structure to describe the different parameterizations of ellipsoids, that's casual, it's linear algebra. It is just linear algebra. It has taken, you know, some square roots, a couple of singular value, inverses, nothing, okay.

Changing the data structure described a polyhedron from vertex to inequality description, that's not casual. So that's a very important thing to note, I mean unless you're in $\mathbb{R}^2$. In $\mathbb{R}^2$, there doesn't seem to be any real problem anyway because your eyeball can solve all of them and it hardly matters; okay.

So I didn't even consider those to be real problems.

Student:[Inaudible] the problem we just solved earlier before we got this example, not correlate with what you're just saying about covering [inaudible] finding the minimum [inaudible].

Instructor (Stephen Boyd):In the problem before, you're given the vertices and you're asked to find the minimum one. That's easy. In this problem, you're given the faces and you're asked to calculate like the minimum, maximum volume of ellipsoid that sits inside it.

Those two are the easy problems. You flip those two, or re-match them and they're actually essentially infinitely hard, not only infinitely hard, but they're basically you don't even get to this point. Just if someone just alleges an ellipsoid covers a polyhedron, you can't even verify it.

You wouldn't know, let alone optimize over such an ellipsoids. So there's a – I mean these are subtle and you have to look at them carefully, think about them. Okay, now there's an amazing fact here and it's actually beautiful – it's the sort of fact about geometry.

It's absolutely beautiful and it's this. If take a convex bounded on [inaudible] interior set and compute the Lowner John ellipsoid so that might be say this polyhedron here and you compute the Lowner John ellipsoid, so you find the minimum volume ellipsoid that covers it, okay. And that's this one.

The following is a fact, comes right out of duality. I won't derive it, there are derivations in the book, comes right out of duality. If you take that ellipsoid and you shrink it by a factor of N , then you're guaranteed to be inside the set.

Okay, by the way the same is true here. If you take the maximum volume of ellipsoid that fits in a set and you grow it by N , it covers the set. Okay, now, you might say N is a big factor. Well, yeah, but at least it's a factor.

And it basically says the following, by the way, if you split the difference, I can say it this way, I can approximate any convex set, you know, bounded [inaudible] interior and that kind of stuff, any convex set in $\mathbb{R}^N$ within square root N .

That's what it means because I would split the difference here, right, I'd shrink this ellipsoid by square root N , and then I'd get one where if you scale it by square root N , you cover, you shrink it by square root N , you're inside.

By the way, if this set is also symmetric, you can sharpen it to square root N , in which case I can even say I can get it into the one quarter approximation. The importance of this is extreme, both theoretically and practically and let me say why. This basically says that ellipsoids are universal geometric approximators of convex sets.

That's what it says because I can bound sort of – you know, I can get a bound and there's a fixed factor whereby I can shrink and expand and fit inside and outside the set. Everybody got that?

Now, that's gonna come up; it's gonna have lots of implications, but what about this. What is someone says, how about a ball. A ball doesn't have that property obviously because for example, here's a convex set. I can make it as skinny as I like, but sort of the minimum size ball that fits this is quite big and the minimum sized ball is like that and there's no bound on these.

So I could hardly say that I cannot approximate a convex set in $\mathbb{R}^N$ by a ball. How about a bounding box? How about a rectangle? Can I do it by a rectangle? The answer there is N , sorry, that's got N parameters. The ball has only, well it's got N . A bounding box has like two or three N or something.

That's not enough either because in this case, for example, the bounding box can look like that. That's the bounding box, the smallest box that covers the set. The biggest box that fits inside is gonna be tiny period. And there's gonna be no bound on the ratio.

So ellipsoid is sort of the first time when you get enough sort of degrees of freedom to carry out universal approximation of a convex set and let me just ask you another one just for fun. How about simplexes? That's the convex hull of N plus one point. By the way, the answer is not obvious here, so you're just gonna have to guess, but I'd say you can do it, so how about simplexes, what do you think?

Can you approximate any convex set by a simplex; what do you think? You can, yeah, you can. There's enough – you have enough free parameters in a simplex to do that and it's easy because you approximate it by an ellipsoid, transform the ellipsoid to the unit ball or something like that and then stick a simplex around that and you can even bound the ratio or all that. It won't be N ; it'll be N -squared or something. It doesn't matter; it'll be something. It'll be bounded. Question?

Student:[Inaudible] rectangle instead of –

Instructor (Stephen Boyd):Oh, you mean a rectangle that can rotate? Yeah, then you can do it. Right. Then you can do it. Of course, we don't know how to calculate such a thing, but if you could do it, that gives you enough degrees of freedom to cover it. Okay.

So ellipsoids are not just sort of convenient things that describe shapes of things. It's actually very worthwhile to understand and remember that there's a sense in which they capture sort of the, let's say, yeah, something like the first order approximation of any convex set, as an ellipsoid.

By the way, let me mention one more property of this. Suppose you took, let's say, ee263. Let's just suppose you did. Okay, and someone says, "What are you doing?" Or you take a control class or for that matter you take a statistics class. And someone says, "What did you do all quarter?" Everything you did involved quadratic norms, right? You'd have quadratic functions.

In control you'd have X trans plus QX . In statistics, it would be hidden by something else or whatever, and someone would then ask a question like, they'd say, "What did you learn in two seasons?" "Well, I can minimize a norm." And they'd say, "A two norm, really and then they'd say, "Does the sum of the squares actually come up in your prop? Is that really what's important in your helicopter design?"

And you could look at them and see how gullible they are and if they look gullible you could try to go, "Oh, yeah, absolutely." Yep, sum of the squares of the vertical acceleration, pitch rate, absolutely." That can't be more than 1.3. If they go for it, the conversation is over. If they don't go for it, then you'd say, "All right, I'll tell you the truth. No. We don't care about sums of squares of things. It's just because we can do it and because that's the only class I took so far."

So that's generally the truth about [inaudible] squares. Okay, so all right. How does that have to do with this? Well, let's suppose that there is something you really care about like maneuverability or ride comfort of something. It doesn't really matter. And suppose what you do now is you go around and you run huge tests, wind tunnel tests. You have pilots write things out and say whether they like the ride or not.

And anyway you end up with some weird set that looks like, you know, some kind of weird convex set that looks like that. That's like nothing's happening and this is sort of a set of ones judged by, you know, wind tunnel simulations, actually getting pilots to ride it and all that and they'll say this is okay. If you're – this is of course, in a multi-dimensional space, but you know, and if you're in that set, you know, one of these things is like pitch rate, one's this, one's the RMS. It doesn't matter.

If you're in that set, everything's cool; that's it, okay. That's the real set, and this is not an ellipsoid. Let me tell you it is not an ellipsoid period. That's the real – this is the set, okay.

Now, this actually – now what you know now allows you to do something that would let you sort of go back and in a posterior way justify 263 material. For that matter, you could go back and justify ordinary regression and here's how you do it, like somebody suggested.

That's actually the set of acceptable ride qualities as judged by horrible long simulations, questionnaires with pilots and blah, blah, blah and whatever, you know, so all right.

This is where people like threw up and stuff over here. This is where they crashed actually. This is where they threw up, but anyway, so what do you do next. Precisely, you just make an ellipsoidal into an approximation. Probably in this case, you might want to make this one. Again, it's you know, you might want to make the inscribed ellipsoid.

Then you take this and you go back to the intern who has only taken 263 and you say, "This is the norm you're going to use. And they, yeah it's a stupid 6 X 6 matrix or 8 X 8 matrix of a bunch of entries, and they go, "Where did all those come from?" And you say, "Last year's intern," right.

It was a lot of work to get it, but you see my point here, that you can actually – by the way, the ideas I'm talking about as far as I know, like no one actually uses them in practice, but everyone should use these ideas in practice, right. I mean this just comes up everywhere. I mean if you're in control, a lot of the methods actually fielded involve quadratic forms like Q and A: and S, then you ask the truth is where do they come from. People make them up.

They don't have to be made up. You can actually do something intelligent and actually get things to do shape the way you really – that do actually capture, not exactly, obviously not exactly. But crudely, they capture the shape of the set.

By the way, over here, this sufficiency was a factor of N. That's something that only matters to a theorist. Okay, and by the way, can anyone guess what's the worst set to ellipsoidally approximate? Just in R2.

Student: Something like –

Instructor (Stephen Boyd): Let's first talk about ellipsoidal approximations in R1. That's a short conversation. How does it go? How well can you approximate a convex set in R1 by an ellipsoid in R1?

Student: [Inaudible].

Instructor (Stephen Boyd): Oh, very well because they're both intervals. Okay, so that was the N equals 1. Let's do N equals 2. Well, you know it can never be worse than 2. Can it be as bad as 2, a factor of 2 – that's if theory says 2. And the answer is sure. A simplex is like the absolute worst thing you could be asked to approximate because in that case, your outer ellipsoid looks like that. Your inner one looks like that. They're 2 to 1. That's the end, and that's general.

So this N here cannot be made better. Now, a simplex is kind of a weird sick thing. It's got all sorts of corners, you know it's got corners on it. In fact, it's a sort of, it's as pointy as a convex set can be roughly, okay.

Now, most what [inaudible] a lot of convexes that come up in practice. The approximation number is definitely not N . It's often much, much less, and for example, if we actually went out and rode and worked out ride qualities or things like that for a vehicle or anything like that, you would find most sets that arise through natural causes – what I'm about to say is, of course, just total hand-waving, but I believe it to be the case.

Well, actually since I'm not making a statement, it cannot be disproved; however as a rough idea, I will say this, most of the convex sets that come up through natural causes can actually be approximated by ellipsoids stunningly well. Nothing close to square – to N . Okay, so I just mentioned that. That's for those of you who are worried about that factor of N ; you needn't be. Okay, so that's our discussion of that.

Centering – that's an interesting thought. So in centering, it works like this. You have a set and you want to find a point that's deep inside the set. By the way, we've already seen one application of this, which is design centering, which is yield maximization.

So in yield maximization, this set describes the set of acceptable values. In other words, the point that you're looking for is what you tell people to shoot for and manufacture. That's what you want to do.

I mean you can also think of less socially positive applications. This could be the range of points where if you're within there, you take out a target. And then you want to ask somebody you know, "Where do you put your sight?" Right, and that's another question. The answer is you don't put it there, and you know, you put it right in the center, what center. You want to maximize the probability now that you're in the set.

But we'll go back to manufacturing, okay. So here, now the simplest, actually the yield [inaudible], we already talked about that, that's actually a convex problem provided the probability distribution is log concave. Doesn't mean you can solve it. You can, but not by methods from this quarter, but you can solve it. Abstractly it's a convex problem.

So a lot of people use a various eurisitic for that. They work unbelievably well. One is this – you find the center of the largest ball that fits inside the set depending on what the set is, I mean if it's a polyhedron, for example, we already looked at that like Day 1 for linear programming.

This is a linear program to calculate the maximum volume, but the largest ball that fits inside the polyhedron. That's an LP. By the way, I can say largest because you would only say largest ball if this were – you'd only use the word largest if there were linear total ordering.

For balls, there is because – at least if you're talking about the size of a ball, it's the radius and so the totally ordered. You would never say what's the largest ellipsoid inside a set because that makes no sense. You have to put in something like largest volume ellipsoid or something like that.

Okay, so that's an LP, and we just worked out this.

Calculating the maximum volume ellipsoid in a set is also a tractable problem and this is called the maximum volume ellipsoid centers. So it's called XMVE. This is X [inaudible], and you can think of lots of others. One, this one has a very important property.

It's affine invariance. That affine invariance means if I transform the whole problem by, for example, scaling the coordinates, if I stretch it some way, that will change radically the [inaudible] center.

But it won't change the maximum volume ellipsoid thing because everything will transform by the fact the determinant of T where T is the linear part of the transformation. So this is affine invariant. Now as to which is better; which is not.

It just means that if someone's calculating in [inaudible] center, you should ask them the following question. How well do you trust your choice in coordinates, you know, have you scaled everything properly? You know, this kind of thing. Is it true that X_3 being on the order of 1 is about the same as X_2 being on the order of 1.

If they don't immediately answer that question with, "Oh, yes, I've been very careful to scale everything here, very carefully so they're all on the same order," they don't immediately answer it that way, then you need to poke them and bug them and say you better check your scaling, because it matters here. Here it doesn't matter at all, those affine invariant.

So, okay. Another center is the so-called analytic center of a set of inequalities [inaudible]. That's a typo. So this we're gonna look at in great detail in two weeks from now so when we actually talk about [inaudible] point methods and how do you solve all these problems. So we'll save that for then mostly.

But it's this. What you do is you set up a problem. You have some equality constraints and you have some inequalities, and these don't really matter. You want the margin here is actually minus FI of X . So for example, if minus FI of X is 0, you'd say it's tight. Minus FI of X is 1, would say you have a slack or margin of 1 and you kind of want to minimize the margin.

And there's lots of ways you could maximize the minimum margin, but an interesting one is to maximize the product of the margins, which is the same as minimizing the negative sum of the logs of the margins, okay. So that's another weapon.

It sounds odd, but we will get to what it means later. We'll also see that although this looks rather complicated, it turns out it's shockingly low complexity, so we're talking 15 lines. Indeed 15 lines that you will write soon. So, and it turns out here by the way in this case, there's also an ellipsoid inequality about if you calculate this you'll get an inner and outer ellipsoid. Yeah.

Student:[Inaudible].

Instructor (Stephen Boyd):Yeah.

Student:How would you extract that center?

Instructor (Stephen Boyd):That's part of the data. I mean I forget how we parametrize it with B or D ; does anyone remember. Yeah, so that's easy. Yeah, you might have to do some calculations, but whatever it is it's very, very straightforward. You just transform it. I don't know where I was – no, forget it; it's easy enough. It's a quick linear algebra calculation to get the center.

Okay. So here's an example – this will be important later, so I don't mind going over it. Also, by the way, just the idea of an analytic center is something that in many applications should be propagated. You'll see lots about this in the next two or three weeks and related problems, but it's something that should just be propagated because anytime you find a problem where someone says, "Here my specifications. Here are my inequalities. Pick me a point in them."

And you say, "Really, any point? Like what if I picked a point just barely in them," and they go, "No, if you're going to pick a point in them, you might as well get one that satisfies a bunch of them, you know or something like that." Analytic center is actually probably a good choice and it actually comes up in a lots and lots of already comes up in things like maximum [inaudible] estimation and we'll see all those connections; okay.

So percent of linear inequalities, we have a polyhedron, and these are the level curves of this – this is a so-called log barrier function for this thing, this set, and you can see the level curves here. When you get really high curves, they kind of hug the shape of the polyhedron.

At the minimum, it's a smooth convex function. So it's got a minimizer here. That's the analytic center and the analytic center here, this is a smooth convex function near that minimum, this thing looks like a quadratic period. Therefore, the levels sets near the analytic center are ellipsoids.

That ellipsoid is a pretty good approximator of the shape of this set and indeed there's a bound. Now, the bound is interesting. The bound, so you can puff it up by, well a factor of M , that would be M -squared. You could puff up the ellipsoid by a factor of M , but M is not the space dimension, it's the number of inequalities, so it's a worse than the maximum volume ellipsoid or Lowner John ellipsoid. So that's it.

We'll get to these ideas later. Okay. Let's look at another topic is the idea of discrimination classification. It's got lots of names, and let's see how that works and so you want to separate. Here I have a bunch of points in $\mathbb{R}^N$, and they're labeled. They have binary labels. They're binary classification, so I can think of it this way.

These, I'm labeling them actually by the symbol, so X is one set; Y is another. This could be a bunch of vectors of something where there actually was something present and this could be a bunch of vectors where something was not present or something like that, but I know which is which here. I'm told which are which.

And what I want to do is this. I mean the oldest problem and by the way, linear discrimination, this goes back easily to MIT in the '40s and probably earlier than that, so this has a long, long history of linear discrimination.

So what you really want to do is find out is there a separating hyper plane for this data set, very basic question and if there is, it means is the following set of inequalities feasible, that the X is lying on one side and the Y is on the other. Now, the variables here is A and B .

Now, there is one difference here. These inequalities are strict; okay. In fact, we haven't dealt with strict inequalities yet, and we're going to now and you're going to see a trick. By the way, let's work out the non-strict classification – let's have that discussion right now. Under what circumstances can two sets of points in $\mathbb{R}^N$ be non-strictly separated? That means when does there exist an A and B for which let's say this is true. When?

Student: Always.

Instructor (Stephen Boyd): Always and by what choice of A and B ?

Student:

Zero.

Instructor (Stephen Boyd): That's right; zero. If I choose A equals 0 and A and B equals 0 and this always works, even if the points are all like messed up and next to each other. In fact, what if they're identical. So that's why it doesn't make any sense to look at that problem. We have to look at the strict problem. You haven't – we haven't looked at strict inequalities yet.

So now I'll just tell you the trick is very, very simple here. These inequalities – take a look at them. They are equivalent to a set of non-strict inequalities. They are equivalent to these. Right, put a 1 and a minus 1 there. Okay. This trick you will be carrying out this trick. You need to get it. Okay.

Here's the argument. Watch this. If there existed any A and B , where a strict inequality held here, and strict inequality held here, this is homogeneous in A and B . So is that. So suppose the A and B you give me that satisfies this has this. These are, you know, 10^{-5} minus 5 and these are -10^{-5} minus 5 . I can multiply those A and B by ten to the 5 and make the gap plus and minus 1. Okay, so I can make it this way.

Now, conversely, if A and B satisfies these inequalities, obviously these imply that. So that's how you do this, so these are tricky and you have to – and if you do these wrong or casually, just you know, normally, there's a lot of cases where it is strict inequality can just replace it with a non-strict inequality, and no one's gonna get hurt and in fact, the original one didn't even make any sense.

Here's an example. If someone says to you, I'm designing a circuit and the power can't be more than 50 milliwatts and you say, "Really," and then later you get a design document and it says as Item No. 1, P is less than 50 milliwatts, like that. It probably means whoever made this up hasn't thought very much or whatever, because you'd say, "Really, could it be exactly 50," anyway.

This is silly. But they probably meant that and there's no engineering difference between the two. Makes absolutely no sense because you'll judge this by spice or something like that and your error will be, you know, it won't be that good anyway.

Anyway if you're manufacture, it's gonna be plus or minus several percent anyway. Just change the room temperature and it's gonna change. So the point is, that's a case where the distinction between strict and non-strict is totally irrelative. This is just due to the ignorance of the person who wrote the specifications. They didn't know there was a difference, okay. So that's this one.

This is not one of those cases because if you just casually make these non –strict here, the answer is 0, okay. If you're using some tool to get the answer, you'll not get an answer, because if you give a problem to a solver like that, it'll be very, very happy to give you the answer and it won't be the one you want.

So there are cases where the strict versus non-strict inequalities actually matter in practice. They matter a great deal. I mean also this all matters conceptually, too. So you should not confuse the two ever frankly, but okay.

Now, to check if there's a set if there is a separating hyper plane is actually just a set of linear inequalities. That's this, okay – in A and B. So for us it's an LP feasibility problem. There's nothing more to say period. Okay, we will say more; we'll say it next week, but that's it.

Let me just mention something that you might use this for. You might use it this way. You might actually take a whole bunch of snapshots of stuff where something actually happened like a target was present or something. Or a disease was present or something. This can be gene expression data.

This could be the gene expression data, giant piles of gene expression data where disease was not present. Okay. These of course, would have a dimension. These Xs could be a million dimensional. For example, something like that. Okay, then what you want, I mean if, then supposed there were a separating hyper plane. There wouldn't be.

But let's imagine that there were one. If there were a separating hyper plane, you could actually now do something quite interesting. Somebody shows up, you do the gene expression or array. You get the data and you plug it in and this number here will either be positive or it'll be negative.

If it's minus 3, you might guess with some confidence, depending on if you believe any of this, that some disease is present. By the way, if it comes out about 0, what would you – that's a good time to say we don't know. So we'll get to the shades business. So that's the kind of thing you would use separating hyper planes for so it would be to basically predictions of new points where you don't know which of the two outcomes actually will happen you want to predict. Okay, so we'll quit here.

[End of Audio]

Duration: 77 minutes