

ConvexOptimizationI-Lecture02

Instructor (Jacob Mattingley): Okay. Good morning and welcome to EE364-a. This is the second lecture, and in case you haven't guessed, I am not Professor Boyd. He is currently on his way back from India, and he'll be here in time for the lecture on Tuesday. You don't need to worry. It will be a proper professor from Tuesday onwards.

A couple of things before we move on. First of all, the first couple of weeks are going to be relatively dry mathematical material. There's a bit of ground that we have to cover before we can do the really interesting parts, so some of that starts today. As part of that, we're going to be going quite fast indeed, so hopefully lots of you have read or scanned part of chapter two, because we're hoping to get through most of that if not all of that today.

If you do find that I'm going quite fast today, that's hopefully because I am. You might find that it's helpful if you go back and read chapter two once you're done with this lecture. Before we dig right in, another thing is that we have posted homework one. It's not due until the 17th, which is next Thursday. Homework 1's already out. Office hours for TAs have been posted, and there will probably be more to come.

Today, we're going to be talking about convex sets. I'll start off by just describing what an affine set is and then a convex set. We'll go through some examples, then we'll talk about how you make more complicated convex sets from simpler ones and talk about a few other topics towards the end. This should be review to start off. We can form from any two points – we can parameterize the line going through those two points using a parameter such as θ . In this example, θ can vary over the whole real line, and it should trace out the whole line.

If we've got a point X_1 here and a point X_2 – when θ is zero, we're sitting right at X_2 . When θ is one, we're sitting at X_1 . It traces out the whole line like this. More generally, an affine set can be described as a set where any two distinct points and the line through them are all inside the set. So take any two points from an affine set, trace the line through them, and that line better be in the same set.

One example of an affine set is a solution set of linear equations, which you've all seen by now. There are lots of different ways of writing it. We can write it with a [inaudible] parameter. We can write it using an equation on the left-hand side here, but at the end of the day, this is one of the ways you can represent any affine set. A convex set, which is going to be of considerable interest to us this quarter, is one where instead of being able to vary θ over the entire real line, we're instead constrained to having θ be between zero and one. It can take on any value between zero and one.

If we find two points in our set, instead of before where we could have any point on the line through these two points, the line segment described with θ here has θ varying over zero, one. So start off with, line segment θ between zero and one and these are the points traced out by θ as we vary it.

The convex set is a set which takes any two points in the set – so take X_1 and X_2 from your convex set, and the line segment between those two points had better be in the set. That means that an affine set is, in fact, a convex set, because an affine set contains the entire line through X_1 and X_2 . Of course, it does contain the line segment between X_1 and X_2 . Our first example of a convex set is the affine set from before.

This is just a formal way of expressing it here. We take X_1 and X_2 out of this convex set C . We vary λ between zero and one, and the convex combination of points here has to lie in the set C . Here are some examples. The set here – obviously, if I had picked these two points here, every point along this line lies inside the set. This here is a problem, because if I pick these two points here, all of these points here do not lie in the convex set. That clearly violates this set membership here, and this cannot be a convex set.

This one here is a more technical example. This one looks like it should probably be convex but for some of the points where the boundary is missing. If I had picked a point here and a point here, this line actually is not inside the set, because it's only got a partial boundary around here. These are some very simple examples described graphically constrained to just our two. You'll see quite quickly that convex C can make sense in quite bizarre spaces.

Following on from convex combinations, before, we had just two points involved in our determination of whether a set was convex. The thing is that we can actually extend this concept to a convex combination. If we picked points X_1 through X_K to start with, we can describe a convex combination of those points as being a set where we weight each of the points, X_1 through X_K , and the weights all have to be nonnegative and they have to add up to one. This is just a definitional thing at this stage. We define a convex combination as something like this.

If we had three points like this, we'll see that the set of all convex combinations actually looks something like this. If you do the math, you'll find out that as you vary λ_1 and λ_2 and λ_3 , and you always have them nonnegative and you vary them so that they sum to one, the set of points that the sum traces out is called the convex hull of the set. We can take any set of points or just a set and we can form the convex hull by finding all of the convex combinations of points within that set.

Let's take these black points here. Supposed someone just handed you these black points and they said what is the convex hull of these points? At that stage, that sounds very unlikely, but later on, we'll see why you might be interested. We take these points and we want to know all of the convex combinations. This is the smallest convex set that contains all of those points. It's formed by basically taking the points at the edges of the set and adding the line segments involved in those points and try to make them as straight as possible. Then, you fill in the rest of the set.

So if you asked, for example, how I got to this point here, you could form it as a convex combination between these three points here. Or, in fact, this one lines just directly on a line between these two. For any set, we can find the convex hull of that set. Here's

another example. The kidney shaped figure that we had before – to find the convex hull of this set, we need to just basically fill in this region here. That's the problematic region before. If we found two points like this, the line segment did not lie entirely within the set. Any questions so far on convex combinations and convex hulls?

The main difference is on what you can do with λ . An affine combination of points that we had before is you have three variables. A linear combination – this would be a linear combination if we didn't have this restraint here. This thing makes it a convex combination instead of linear. Any other questions?

That's exactly right. If you already have a convex set and you want to know the convex hull, that's simply the convex set itself, because that set is required to contain all convex combinations.

The use of a convex hull – if you already have a convex set, the convex hull is useless to you. But sets are by no means likely to be convex in some sense, so it's useful when we want to make a convex set. But certainly if you call some function f and you [inaudible] on a convex set, it would simply return the set without trouble. Yes, a convex hull has to be convex by that definition.

Let's have a look at convex cones now. Let's introduce a conic combination of points. This is again a specialized version of a linear combination of points. We take two points. We can have more than two. We have nonnegative weights on these points. That's described as a conic combination of these two points, X_1 and X_2 . Here, we have a figure where we have a point X_1 and a point X_2 , and this is the set of points $\lambda_1 X_1$ plus $\lambda_2 X_2$ where λ_1 and λ_2 are nonnegative.

Obviously, it's a cone, and that's why we chose the name convex cone. It also has to be convex. The reason for that is if we look back to our definition of convex, there was a special case of these. We added another constraint on top of this. If we merely wanted to insure convexity, we'd only have to check that X was in this cone when λ_1 and λ_2 added up to one. Any convex cone has to be a convex set.

Now we have a look at hyperplanes and halfspaces. Hopefully, you've seen these many times before. We start off with a definition. A hyperplane is a set where $A^T X$ is equal to some constant B . A is nonzero, because otherwise, we could have any point in the whole series. This is an example hyperplane. We parameterize it with a point, say, X_0 , and then we need to know that $A^T X_0$ is equal to B , and then any other point that also has an inner product with A of B is also lying on this hyperplane.

This is a hyperplane in two space, which is a line. In three space, you can think of it as looking something like this [inaudible] here. This would be a hyperplane in three space. It need not go through the origin, by the way. Following on from that, we have a halfspace, and this is similar to a hyperplane, but we just take the area on one side of that plane. We can have it either open or closed. This can be a strict inequality or a non-strict inequality.

This is an example here – same hyperplane, and it's the half space on the lower, I guess, in this case.

Likewise, the non-shaded region you could describe as another halfspace with $A^T x \geq b$. The direction of this inequality indicates what side it falls on. A few things – halfspaces are not vector spaces. If you take two points from here, it should be closed. If it's a vector space, it would be closed under scale of multiplication. That's not going to be the case for halfspace, so a halfspace is not a vector space. People often get that confused, so don't fall into that trap.

That's a hyperplane and a halfspace. They are both convex. A hyperplane is also affine, which ensures that it's convex, but a halfspace is merely convex and it's not affine. It's not affine because we would have to be able to contain the line through any two points. Let's pick these two points here and then draw a line between them.

Student: Is there displacement from the origin and A and B that are written?

Instructor (Jacob Mattingley): There is, and it's basically if you find the point that's nearest the origin – I think this is part of the homework exercise. If you found that x_1 was the point nearest the origin, then you could use A to find out – I'm going to leave that to you, because I think it is part of the homework. There is a relationship. Those are some very simple convex sets.

Some more sets that we're interested in – first of all, a Euclidean ball. At the moment, Euclidean is going to mean that we use the two norm in the ball description. We parameterize a ball by picking a center point x_c and a radius for the ball of lowercase R . It's the set of any point that is [inaudible] in the Euclidean sense to x_c then R . Here's one way of expressing that. Here's another way of expressing that. This is a free parameter representation where you have a ball U and it's just a unit ball. You scale it by R , and then you center it around x_c .

This is the free parameter. Here are two [inaudible] descriptions of the same thing. We can get more interesting than a Euclidean ball and take an ellipsoid, which is a generalization of the Euclidean ball, and that's points – we take a symmetric positive definite matrix P and we use this expression here. You can actually go back and forth between this representation for an ellipsoid and this representation for an ellipsoid. This is just a sample ellipsoid. Ellipsoids and Euclidean balls are convex sets. One of the ways you could prove it is you could find two points and show that if those two points are in the set, then their convex combinations have to be in the set as well.

Again, if you haven't read chapter two and you haven't seen these things before, I'm going to be going quite fast. More generally, we're seeing two norms a lot. We're seeing the Euclidean norm. But more generally, we can define a norm as being any function that satisfies these things here. First of all, this is something like definiteness and non-negativity. We need to know that – these are just a bunch of things that you should look at that define a norm.

You can verify quite quickly that the Euclidean norm does in fact satisfy these, and there are various other ones as well. Just a little point is that this notation here is [inaudible] of the absolute symbol, which is a norm in the space \mathbb{R} . If you take the absolute value of a number, it's going to satisfy all of these things in the space \mathbb{R} .

In 364a, quite different from 263, if we simply have the norm symbol without any subscript, it does not necessarily mean the two norm, and in fact, it's usually used to represent a general norm. We've kind of moved up a level of sophistication, if you like, that no longer is it simply enough to just say this is the Euclidean norm of X , but we want to specifically say we are after the two norm. You'll find and use often different kinds of norms in this course.

We can define as we saw previously similar to the Euclidean ball here, for general norms, we can define a norm ball. This ball doesn't necessarily look like a ball. It's a ball in the general sense. We define them very much the same as we define a Euclidean ball, but they're going to look quite different. Let's take an example. We have the one norm, which is the sum of absolute values. Can anyone tell me what the norm ball with the one norm looks like in \mathbb{R}^2 ?

Student: Square.

Instructor (Jacob Mattingley): Okay. It looks similar to a square. It's actually a diamond like this if I recall correctly. The one that is a square, however, is the infinity norm. That's the maximum absolute value. That does indeed look something like this in \mathbb{R}^2 . These are two other norm balls. They don't look much like balls, but that's their name. You could call this the one norm ball and this the infinity norm ball. There are others.

We can also define a norm cone by adding another variable here and saying the norm cone is any point X, T where the norm of X is less than T . This is our favorite norm, the Euclidean norm, and it's the norm contrast out. It's got a couple of names. It's called an ice cream cone for obvious reasons, and it's also called a [inaudible] cone as well. This is actually a solid figure. It might look like a wire frame, but it's actually filled in. It's any point where X is in \mathbb{R}^2 here along the bottom and T is up here.

One of the things we know about norm balls and norm cones is that they're convex. Yet again, we're adding another building block, which will become very useful to us later. It doesn't necessarily mean a whole lot now, but just store away that norm balls, norm cones, some of these other things like halfspaces – they're all convex, and that's going to be very useful to us later.

Before I go on, I'll mention that if you want to know what the one norm cone and the infinity norm cones look like, you can think of these here as a cross section when you look down from the top of the cone. If you looked down on this cone, the level sets of the cone would look something like this. If you looked down on the level sets of these cones here, the cross sections are going to look like this. We could fill this in and similarly for

the infinity norm. You can think of them as a pyramid, and the one norm is rotated and the infinity norm is a [inaudible] pyramid.

Here, we have a slightly more interesting set of shapes that are also convex. A polyhedron is the solution set of finitely many linear inequalities and equalities. Here's one of the ways that we can represent it. We can say a polyhedron is any X such that AX is less than B and CX equals D . That would be a polyhedron. We note here that there's actually a slightly different symbol for less than here, and the reason for this, and you'll often see this both ways. You'll see an ordinary straight symbol here. You'll sometimes see this curvy one.

The reason is because this here is what we got from R . If we say three is less than or equal to five, this is what we use. Here, we're actually overloading this notation to say that AX , which is a vector, is component wise less than B , and to make this completely explicit, we're using this slightly different symbol here.

During this lecture and other times, I might get a little bit sloppy and the curves might not be as obvious, but the reason for this is because it's actually not quite the same as the ordinary R ordering, and we'll see a generalization of this quite soon. Note that it's different, and the reason is just to make it very clear that we're doing something different to what we were doing before in R .

Positive semidefinite we'll see in a second. This is one example of a polyhedron here. This is an intersection of a finite number of halfspaces and hyperplanes, and so in this particular case, we've got five halfspaces defined by these normal vectors A_1 through A_5 , and you can think of this here as the halfspace defined by A_1 . If you find the intersection of these five, you'll discover that it's this shape here.

A polyhedron is convex, and this is something that you need to prove. I'm not going to go through it at the moment, but you should be able to show fairly easily that when you find an intersection of a finite number of halfspaces and hyperplanes you get a convex set, and it's called a polyhedron. An affine set is a special case of a polyhedron, and so is a hyperplane from before.

Student: This one only uses the inequality, so could you give a quick example of one that actually uses equality?

Instructor (Jacob Mattingley): Okay. Here's one. Pretend that this is actually an R^3 , and we've got five inequalities like this. We've got an equality which constrains us to being in this plane. They can be anything like that.

Next, we have a look at the positive semidefinite cone, and that's what you were asking about before. We say that a matrix is – you should know about positive semidefinite matrices by now, but we define a couple more sets. First of all, S_n is the set of symmetric [inaudible] matrices. Let's have a quick look and see. Do we suppose this is a convex set? Any suggestions? Yes, it's a convex set. How would we go about proving it?

Let's use the definition. We'll pick an X , which is a symmetric matrix. We'll pick a Y , which is also a symmetric matrix. Our requirement for something to be convex – if this is a convex set, then if we take a α which is greater than or equal to zero and also less than or equal to one, and we take αX plus one minus αY , the set here is convex if each of these things here is also in the set. Is that true for a symmetric matrix if you take a sum of symmetric matrices? Is it still symmetric? Yes, it is. That means that this is a convex set.

It is, in fact, affine, yes. Again, you're quite right. The reason for that is because we never used the fact that α was between zero and one. α could be anything and we would still have a convex set. Yeah, it's all of these things. Let's look at a more interesting set. We'll define and use this notation extensively that S_{+}^N is the set of positive semidefinite N by N matrices. We'll assume and especially when we say is, we're assuming that they're symmetric.

This is a set like this. If you've forgotten what positive semidefinite means, it means that when we take a quadratic form such as $Z^T X Z$ is greater than or equal to zero for all Z . Here's the claim, that S_{+}^N is a convex cone. Looking back to convex cones, that means that if we take a nonnegative combination of two of these items from the set, it also needs to be in the set. You should be able to see quite easily that if we had X that was positive semidefinite and Y that was positive semidefinite, if we took a conic combination, $\alpha_1 X$ plus $\alpha_2 Y$, this, too, is going to be positive semidefinite. Can someone tell me how I can make that really explicit quite quickly?

Student: Z^T transpose on the left and [inaudible] right use the distributive property to just –

Instructor (Jacob Mattingley): Exactly. I'll do it like this. I'll say $Z^T (\alpha_1 X + \alpha_2 Y) Z$ like this, and then I'll expand it out and I'll get $\alpha_1 Z^T X Z$ plus $\alpha_2 Z^T Y Z$. We know that this is positive or nonnegative because X is positive semidefinite. We know the same about this. We know that α_1 and α_2 are both nonnegative, and so the whole must be nonnegative. I've just shown that if we take a conic combination of two positive semidefinite matrices, the result is positive semidefinite. This means that S_{+}^N is a convex cone and furthermore, it's also convex.

We can do a similar thing for positive definite matrices, and here, we just add an extra plus to indicate that it's strictly positive definite and not just positive semidefinite. Here's an example of what the cone looks like in S_{++}^2 . This is a cone of positive semidefinite matrices. You do write something like this. Exactly how you do it in MatLab, I'm not sure, but if you want, you could pick a point in the space – say we pick this point here.

You could find out its X , Y , and Z coordinates, plug it into the matrix and you see what the minimum [inaudible] value is. If it's greater than or equal to zero, then it's in this cone. If it's less than or equal to zero, it's – well, if it's less than zero, it's outside the cone. If you did a grid search in here, you could build up a set of points which are inside the cone.

Student: Can you have a positive definite matrix that's not symmetric?

Instructor (Jacob Mattingley): Yes, you can. This particular picture, it does depend on it because we've explicitly said that. Basically, it's bad form to give somebody a non symmetric matrix. This proof here neither used the fact that that was symmetric, so yes, it's true that the matrices don't have to be symmetric, but there's never really any reason why we would [inaudible]. There's no harm in taking any matrix said to be positive semidefinite and immediately symmetrizing it without loss of generality.

We could actually find an explicit form for the determinate of this matrix. It also needs the fact – what's that? Anyway, we can generate this plot. Feel free to look into how you do it. I'm going to move on for now because we do have a lot to get through.

Why did we use this particular sign here rather than an ordinary one? That's a good question. We have specifically used this curvy sign, and that is again because we're not using the ordinary ordering on \mathbb{R} . This is now a matrix that we're saying is greater than or equal to zero in some sense, and it's the sense of positive semidefiniteness. We have again very deliberately chosen to use this different sign to avoid confusion with the pure, ordinary symbol like this. You will find some people who use that symbol, and that's okay.

So now we've looked at some sets that are by themselves convex sets, but we also want to know how to make convex sets out of other convex sets. There are certain operations that when we apply them to the set, they will preserve convexity. What it means is that you'll be able to build up more interesting sets from these simple ones that we started with. If we were constrained to one set, we could explicitly write equations and explicitly prove we're convex. We'd have a lot of work to do if we cannot with an unusual set.

First of all, let's have a look at some of these things. One of the ways we can see if a given set C is convex is we can just apply the definition. Sometimes, that'll be easy. Sometimes, it won't be easy. Another way we can do it is we can take operations that are known to preserve convexity. We can take these little [inaudible] that we know are convex, and we can put them together and quite quickly make a fairly interesting looking set. Here's one. It's an intersection.

This is actually any number of intersections, and I think we look at that on the next slide. An affine function will preserve convexity. If we take a set C , put it through a function F which is affine, it's going to preserve convexity. A perspective function and linear-fractional – we'll see those in just a few moments.

Here's another way. This is the crude and nasty MatLab approach. Say I want to know if a set is convex, and say I've got some oracle which tells me whether a given point is in the set or not in the set. Here's what I can do. I can pick a couple of points, say, X_1 and X_2 completely at random from the space that I'm in at the time. Then, I can test my oracle and I can see if X_1 and X_2 are in the convex set. Suppose they are. The

requirement of convexity is that when I find this convex combination of these two points, that point itself has to lie in the convex set.

If somebody gives you some random set and you just want a very first pass of is this even remotely likely to be convex, here's what you do. You make a tiny little window in the corner of your screen. You run MatLab and you start generating random points from the set. You take a convex combination and you test to see if it's still in the set. Now, as soon as you find some λ which is between zero and one where $\lambda X_1 + (1 - \lambda) X_2$ is not in the set and X_1 and X_2 are, you immediately return and say it's obvious. Of course it's not convex. You can see that when you're looking at it. It's quite easy to do that.

Here's another thing, which I won't go into at length. Basically, midpoint convexity where you just take λ equal to 0.5. That's actually enough for some technical conditions. If you want to run a whole bunch of tests, you could just find X_1 and X_2 and take their average. You test to see if the average is in the set. If it isn't, you return and say no way it's convex. If it is, you don't necessarily know much. It's only useful if it's not a convex set.

But if you did, say, 100,000 points, you could give a very rough first pass. Note the quotation marks – it's very rough. It's a heuristic method. Nevertheless, it can be a very fast way of finding out whether something is convex. It fails in situations where – it's to do with boundary issues. I'm not going to talk about these much, and the focus of this class is not really on the technicality, but if we took these points here – in this particular case, it would work. Say that point there was in the set. Then these other points wouldn't be anything [inaudible].

Let's have a look in more detail at intersections. Here's the claim – the intersection of any number of convex sets is convex. Fortunately, I can just refer you to your homework, because I think you have to show this in your homework for a couple of sets. When I say any number, that means finite, infinite and unaccountably infinite. This is a very robust way of finding out whether something's convex.

Here's an example. Here, it's an unaccountably infinite set, I think. We just take this random thing here and we know that the individual elements are convex. We put them all together and we put up with this shape here. I don't want to dwell on the shape. It's colored in lots of detail in the book, and in fact, let me just take a moment and mention the book.

The convex optimization book is available, and these slides are based very much on the book. Everything here is covered in detail in the book. The book is also correct from a mathematical point of view. If I'm tossing around boundary issues and there are little technical differences, we're not going to worry about them in lecture nearly so much. The book is correct. If you have done analysis and you are concerned about all kinds of things like closure and compactness, then you can go to the book and it will be correct. We're not going to worry so much about that now.

Here's a set where we have an infinite intersection of convex sets, and it is itself convex. Let's have another look at a way that we can preserve convexity. Let's have a look at affine functions. Suppose we have some function F that takes \mathbb{R}^N into \mathbb{R}^M , and it's affine. One way of looking at this is saying that if F of X is AX plus B , A is a matrix and B is a vector. This is actually not the most general way of expressing it, but this is pretty good.

The image of a convex set under the affine mapping F is convex. If you're given some set S which is just a set at this point and it's \mathbb{R}^N , we can use the mapping F and all of the points from X that are mapped into $F(X)$ – the new set is a convex set, and likewise for the inverse image. Just to make clear what the difference between these is – suppose we have some set S here, and this is a convex set. We take this under a mapping F to some other set here, this is the set F of S . This is a convex set if F is an affine mapping.

Likewise, we can do the same for an inverse mapping. If we had another set – these are not very interesting looking sets, but it could be any set. We can take this back via F and F is a convex set. After this mapping, it's also a convex set. If we have an affine function, we can actually create quite a lot of damage. When we have a known convex set, we can make other ones out of it.

Student: Which notion of adversity – does it have to be one to one?

Instructor (Jacob Mattingley): Actually, we don't really run into those issues. This is the definition here. It's simply the set of any points where the function value of that point matches to the set C . It's the inverse image of the set. The function – it is affine, so is it one to one? It probably is. It doesn't have to be. It's the inverse image, which is a well-defined thing. It's always convex if the set C is convex.

Some examples of this – we have simple scaling by two. We have translation where we move around, and we have prediction. Let's briefly mention that the converse is not true. We might have a set F of C that's convex when C wasn't convex. We're more interested in this, anyway. Here are some simple examples.

Prediction could be really simple. We might have a set X_1, X_2 . This might be a point in a convex set C , and if this is convex, then the set of X_1 given that the sum – the sum of X_1 – for sum X_2 , the set of just X_1 s might be a convex set. That's a very simple prediction. A more interesting one is what's called a linear matrix inequality or an LMI. This is an affine mapping. If we have a set of matrices A_i and they're all symmetric matrices, then the set X where this weighted sum of symmetric matrices and the positive semidefinite sense is less than B , that is also a convex set.

I'll talk about that one for a second. If we take the cone – this function F of X – we'll make it equal to B minus A of X where A of X is equal to X_1, A_1 up to X_M, A_M . This is an affine function, and it needs a little bit of hammering to get it in the same form as we saw above here. It is an affine function.

We already know that the positive semidefinite cone – we already know that this is a convex set. If we take the inverse image of the set here with this function F , what we get is that the inverse image is equal to any X and R_N such that $B - AX$ is in the positive semidefinite cone. This is just a fancy way of saying that $B - A^T X$ is less than or equal to zero in the positive semidefinite sense.

We've taken a very innocent looking set, S plus of N . We talked about how that was a convex set earlier. We took also a fairly innocent looking affine function, F of X , and immediately, we've got this fairly unusual looking LMI, which is actually a set of weighted symmetric matrices. That's what's on the left-hand side here. This is a convex set, and there were just two very simple ingredients. You can see that we can cause a lot of damage.

This comes up a lot in control, and it's a very useful set. Another one here is the hyperbolic cone. This one you can get in a similar way. We already saw that the set ZU , which is a cone when we restrict the U . We saw this set before. This is a norm cone, and we showed that this was a convex set. We know this is convex, and we want to get to something more interesting. You can see that if we apply the affine function F of X equal to $P - \frac{1}{2} X$ and $C^T X$, the inverse image of this mapping is also a convex set.

If you take this cone here and you map it through here, you get a hyperbolic cone. Again, that's something that will be interesting. You should verify the details and as always, details in the book if you need them as well. Here's some quite bizarre sets that are definitely not obviously convex, and you can show quite simply how they are convex if you know how to do it.

Departing even more from obvious sets, we've got a perspective function. This one's a little bit unusual to start with. If I have some function P that reduces $R_N + 1$ to R_N by taking the last component and dividing – you can think about it as taking some vector like this and we pick off the last component. We divide each of these elements by that component and then we throw this away. It's a somewhat unusual operation, but [inaudible] one, two, three, I'd pick the three, and my new vector would be one third two thirds. That's a perspective function.

The images and the inverse images of convex sets under the perspective function are convex. That also is going to be useful to us later. Remember, we're just building up a box of tricks that we're going to find useful later. If it's a bit dry, bear with us for the moment. There's a demand constraint here that this last entry has to be positive.

A generalization of the perspective function is when we have a linear-fractional function. If we have some matrix A , a vector C and vectors B and D , then again, under both the image and the inverse images of a convex set under this transformation are themselves convex. There is a domain constraint on positivity of the denominator here as well. Just to see one example of this because it's probably quite a mystifying concept if you haven't seen these before, we take just a set C here like this.

Here's a linear-fractional function, which you can think of as similar to a perspective function or a generalized perspective function. If we hold this figure on the side – if I hold it like this and I look at this figure from a different angle, I'll find that these parts here are closer to my eye. If I look at this, I'll find that these things loom closer in my vision than the other parts of the figure. This is quite a strange transformation, and C here is obviously not convex. In fact, why is C not convex?

The line segment is very much not in C . If this were a convex set, though, under this transformation, it would remain convex. The reason it's not is because you wouldn't be able to see it too well if it wasn't convex. I'm actually not exactly sure what the dotted line is about. Actually, probably what it is – it's probably a line under the same transformation. I suspect that's what it is.

Another thing you can say about functions is if they transform line segments into line segments, it actually is a convex function. I suspect that's what it is, actually. You'd have to check me, but I think this line moves to this line, and because this line segment moves to this line segment, you can also say that it's a convex function. There are some slightly weird functions. Have a look at them later and put them away as things that we're going to find really useful to prove that sets are convex.

Let's have a look at generalized inequalities. Before, we saw moving from a simple inequality in \mathbb{R} to a component wise inequality. It turns out that it's quite useful to consider other kinds of inequalities as well. Let's introduce proper cones first. A proper cone is a convex cone with another bunch of constraints.

First of all, the cone has to be closed. I'll just mention that some of you will have an analysis background where you know exactly what closed means and you have read many things about closedness. Some of you don't, and that's fine for the purposes of this course. The five-second introduction to what closed means is if we have a shape and the set includes the border of that shape, then it's a closed set. If it doesn't include the boundary at every point, then it's not a closed set.

If you know what closed means, that was probably an insult to you, but that's actually probably enough. It's an intuitive idea of what closed means. It's probably enough for this course. That's the first requirement to make a convex cone into a proper cone. We also say that it's solid. It actually has a nonempty interior. Again, interior is not a concept that some of you will know precisely what it means. Others of you probably haven't heard it before. You can think of interior from an intuitive point of view as the inside of the set. The set not including its boundary.

Tell me about this here. If I just have this ray here, is that a proper cone? Why not? This has an empty interior even though it's getting thicker by the second, but if I just have a ray out from the origin, it has no interior, so it's not a proper cone. There's one more requirement, and that is that the cone cannot contain a line. That means that this set here – this is a cone, and it's also a convex cone, but it's not a proper cone because it contains this line here through the origin.

These are the requirements. It has to be convex, closed, solid and pointed, and that gives us a proper cone. We'll find that that's pretty useful. If we didn't have these constraints, when we go to generalized inequalities, it'd fall to pieces quite quickly. You're saying that any convex cone includes a line. So what about this cone here? This is a convex cone, and it does not include a line. It includes a ray out from the origin, but it doesn't include a line through the origin.

Here are some examples of proper cones. First of all, the one you're most familiar with is the nonnegative orthant. If I have \mathbb{R}^2 , for example, the nonnegative orthant is this portion of \mathbb{R}^2 . That's a proper cone. It's convex. It's closed because it includes these boundaries. It's solid because there's certainly plenty of stuff in the interior. It's pointed because it contains no line, and so it's a proper cone.

Here's another example. The positive semidefinite cone is [inaudible]. That's probably slightly more tricky to prove, but all of these concepts generalize into this cone here. They're not a problem. Here's another one, the set of nonnegative polynomials. We're going to find that this is extremely useful and used all the time. We're going to use this a lot. This one is a more interesting example, but we probably won't see it all that often. Still, it's a proper cone.

This one here? It's a cone because it contains – the definition of a cone was just that if I have a point X in the cone, αX for α positive has to be in the cone. I put any point in here, and I scale it with a linear vector, and it's still in the cone. I pick this point here. I can scale it positively and it's still in the cone. I take this point here. I scale it positively. I stay in this area. You can't find me a point which won't scale and remain in the cone.

This is not a convex cone. This isn't convex, and so neither is it a proper cone. Let's have a look now at generalized inequalities. Before, we had this symbol with the curvy lines. Now we add a K to it. If we take a proper cone K , we can parameterize a generalized inequality with this proper cone K . We describe it as being less than Y . If $Y - X$ is inside this cone K , we can [inaudible] strict inequalities in a similar way by saying that X is strictly this if $Y - X$ lies in the interior of the cone K .

Let's have a look at a quick example. The component wise inequality that we're familiar with – this says that X is less than Y in the component wise sense if each of the components X_i is less than or equal to Y_i . Let's take a couple of points – this one here and this one here. Let's call this X and this Y . X is described as being less than or equal to Y in this proper cone K if $Y - X$ lies in the cone. Let's find $Y - X$, and it's this thing here. $Y - X$ is not $-X - Y$. Is this backwards?

Okay. $Y - X$ – what's going on here? $Y - X$ is inside here. If I had a vector $3,3$ and a vector $1,1$. Subtract them and I get $2,2$. It points to Y , and this is inside this proper cone here. That's a component wise inequality. We can say the same about a matrix inequality. We've seen these before. $X \preceq Y$ is this PSD cone here means that $Y - X$ is positive semidefinite. These are so common that we drop the subscript. Not

only that, but we often also drop the special symbol and just use an ordinary symbol like this.

For the purposes of clarity, it's a good idea to stick with this. If it's obvious what you're talking about, you can drop the K and sometimes people even just use an ordinary symbol. As you'd expect when we're using the same symbol generalized, a lot of the things that we saw before carry over, and the properties are analogous. We have things like this here, and this property will carry over. Be careful, because not everything does.

One of the things that doesn't carry over is we don't have a linear or a total ordering. That means that if we find X and Y , X may be neither less than Y or greater than Y . Here's another way of expressing that. This generalized inequality here is not in general a linear ordering, so like I just said, we may have neither of these holding. You can see that quite easily in our usual case here. If I have a point here and a point here, then this point in the sense of component wise is neither less than this point nor is it more than this point. It's not a total ordering.

We now introduce another concept in this lecture, that X in a set is described as being the minimum element with respect to this generalized inequality if any other point inside the set is actually more than X in this case. That's one concept, and we'll have a look at a picture in a second. A second concept is a minimal element. We describe an element X as being minimal if any point that is less is actually the same point. That's a little bit bizarre to start with. Let's have a look at that.

We're going to use the cone K is equal to \mathbb{R} plus to the 2, so component wise inequality in \mathbb{R}^2 . This point X_1 here is the minimum element of S_1 . What that means is if I picked any other point Y in S , then I can explicitly say that X_1 is less than or equal to Y . That will hold for any point in here. Let's have a look at the difference here in S_2 . What we need here is that I picked X_2 and S_2 . X_2 is a minimal element. That means that if I pick some other point here, I know that if Y is less than or equal to X_2 , then it means that Y is equal to X_2 .

Here's a point, for example. Make that Y . X_2 is not the minimum element, because for this way, I cannot write X_2 less than [inaudible] to Y . I can't actually do that because it's not a linear ordering. That's not allowed. But I can say that any point that is less is actually the same point. That's because – here's another way of thinking about it. If I take this cone K , this set here is the set X_1 plus K , and that's the set of all points that are unambiguously more than X_1 . They can always be described in reference to the cone K as being more than X_1 . X_1 is unambiguously least.

In this set, if I drew the same figure here, these are all of the points that are unambiguously more than X_2 . This point here is not – this point Y is not in the set, so I can't say that X_2 is unambiguously less than Y . What I can say is I can form this other set X_2 minus K , and this is the set of points that are unambiguously less than X_2 . Any point in here, I can say with full authority – this is a point that's less than X_2 . That means

because I find that there are no points inside the set S_2 that are unambiguously less, I can describe X_2 as being a minimal element.

It's a slightly confusing terminology here, but here's how it goes. A minimum is something that you cannot under any circumstances claim that there's a least point. All points are more is what you say for minimum. It's very confusing, even when you've seen it before. Minimum – all points are more. Minimal – no points are less in this unambiguous sense.

We could have a different cone. We could have this cone here. This is a proper cone. We could call this K_2 . Then, we could again look at this and we could say – we could talk about this relative to K_2 . How would you describe X_1 relative to the generalized inequality parameterized by the cone K_2 ? How would you describe X_1 in terms of this one? It's minimal, so it's no longer minimum because the set doesn't lie inside this cone. There are points that are incomparable, if you like. There are points that you could argue are actually no more than this in an unambiguous sense.

That's going to be useful to us. Let's introduce another concept while we're on a roll. This one is the separating hyperplane theorem, and it says that if we take two general convex sets, C and D , and if they're disjoint – they have no intersection and $C \cap D$ is the empty set. If C and D are disjoint, then there is some A , not zero, and there's a B so that the halfspace $A^T X$ in C lies in a halfspace $A^T X$ less than or equal to B . The opposite holds for X and D .

Here's a couple of sets. Here is a hyperplane between the sets, and if F is in R^2 , then it's a line. You can also think about it as looking down – a prediction of R^3 . This says that we can find some hyperplane so that D lies entirely in this halfspace here, and C , if you like, lies entirely in this halfspace here. It's called a separating hyperplane for quite obvious reasons. This theorem, which I think is proven in the book, says that any two sets like this that are disjoint – there has to be a separating hyperplane.

Strict separation requires more, so I won't go into the details here, but strict separation would be the hyperplane that actually passes through neither set. We can use this separating hyperplane theorem and turn it into another one called the supporting hyperplane theorem. A supporting hyperplane we define as being one where – this is described as being a supporting hyperplane if the only point that it touches is – if it goes through X_0 and all of the set is on one side. You can say that if C is a convex set, then at every boundary point of C , there will be a supporting hyperplane.

Let's try to connect these two new ideas together. We have a separating hyperplane theorem which says if we have two sets and they're disjoint, we can find the hyperplane that separates them, and one which says that if we have a convex set and a boundary point, we can find a supporting hyperplane. We can actually use the separating hyperplane theorem to prove this one. We can actually say – we'll form a set X_0 , and we'll also form a set interior of C .

We take a convex set C . We'll assume that it's convex by looking at this part of it. We'll say that X_0 is one set and it's convex because it's just a single point. The interior of C is a convex set if C is convex. The interior of C and X_0 do not intersect if X_0 is on the boundary, and then we can find a separating hyperplane between X_0 and the interior of the set. That will be a supporting hyperplane.

How did I use the convexity? It's actually used in the proof. This is a non-convex set, and if I had a point here, I cannot find a supporting hyperplane, because any line through here is actually going to cut through the set. The theorem is that if C is a convex set, then at any boundary point, I can find a supporting hyperplane. The other thing I used is that the interior of C is convex if zero is convex, and so C and D are disjoint convex sets. I used convexity of those sets and the fact that they're disjoint to show that there is a separating hyperplane and I used that to show that a supporting hyperplane exists.

The square we saw before – its interior was convex, but nothing special. This holds for any convex set. If the set was the interior of C , all of these theorems would hold. It's just a convex set. In fact, if we look at this one here, I could have a hyperplane though here. This also is a separating hyperplane. Here's a convex shape, and this is a supporting hyperplane. So's this. These are not unique supporting hyperplanes and separating hyperplanes.

Here's another one. We are powering through the material, and again, I encourage you to have another look through chapter two. What we can do now is introduce a dual cone. Dual cones again are somewhat strange when you first see them, but they're going to be very useful to us later. You'll be seeing the world dual, duality and other things like that every week throughout the course. This is defined for any cone K – it doesn't have to be convex. It surely doesn't have to be proper. We define the dual cone, K^* , as equal to any point Y with inner product greater than or equal to zero with a point X in the cone K .

Let's have a look and see what that means. If I have a cone that looks something like this – I'll call this K . It happens to be proper, but that's fine. K^* in this case is any vector Y where $Y^T X$ for all X and K is nonnegative. Let's take this line here. Draw out a right angle here. That's a right angle right there. I take this one here and I do a similar thing. This set here is actually the dual cone of K . This one here is K^* . That's the set of – if I take any vector in here, its angle to all elements in the set here is less than or equal to 90 degrees.

If I look at this vector here or any vector in this dual cone K^* , its angle to all elements of K is less than or equal to 90 degrees. This is the dual cone. A few examples – if I look at the nonnegative orthant from before – that's the set here – the dual cone of the nonnegative orthant is the nonnegative orthant itself. It's described as being self-dual. You can see that quite easily if I have two nonnegative vectors Y and X , $Y^T X$ has to be less than or equal to zero, and you should be able to go through a proof of that quite easily. We could just do $Y^T X$, $X^T Y$ and so if the components are nonnegative, so is the product and so is the sum.

The positive semidefinite cone – that's also self-dual, so in this particular case, we actually use a slightly different definition of inner product, and again, details are in the book. If you find the set here, it's going to be the set itself. Euclidean norm cone – it's self dual, but finally, we get to one that's not self dual, and in fact, most of them aren't. A lot of the useful ones are, but here's one that isn't. It's the one norm cone. The dual cone of a one norm cone is actually the infinity norm cone.

The other thing that we note is that the dual cone of a dual cone is actually the cone – the initial cone that we were after – if it's a proper cone. That's with a big if, if K is proper. It may also hold if K is simply convex. You'd have to check me on that. What we can say now is that the dual cones of proper cones are proper, and hence, we can define generalized inequalities. If you give me some inequality with K , I can define a generalized inequality with K^* .

We talked about minimum and minimal elements before, and here is one of the ways that those come into play. We have some set S . We know nothing about the set. Let's say that X is the minimum element of S . That is true with respect to the generalized inequality. X is the minimum if and only if for all Z that are strictly in the dual cone if X is the unique minimizer of Z transpose over S , then X is the minimum element.

This is just touching on the very first parts of duality, which is going to come in lots in this course. It's a little mysterious at first, perhaps, but here's a definition. It's going to be quite fascinating what the Z means exactly. For now, that's what a minimum element is. A minimal element is slightly different.

If X minimizes Z transpose over S and there is some Z – if we can find any Z that's in the dual cone, then X is minimal. We'll see later how useful this is. Finally, we can say that if X is minimal, there is some Z such that Z transpose has a minimizer X . These things – I'm really starting to run out of time, so I'm just tossing them out there, and we'll see duality in lots more detail.

Last thing today is what we call the optimal production frontier. You've probably seen this before, especially in 263. We talk about it with a very simple production frontier. Here's what it says. P is a set of possible production methods. On this axis, we have fuel, and on this axis, we have labor. Say we're testing out different production methods in our factor. If a point is in the set P , it says that there is some production method, which, say, produces 100 milk bottles and takes this amount of labor and this amount of fuel.

If X is in P , there is some possible production method that uses that amount of labor and this amount of fuel. This is a very simple example in R2. This set we could find. Some marketing guy might tell us what products are available. We might do some research. We might do some testing. There are lots of different ways we could find this set. What we say is that an efficient method is one that is minimal with respect to R plus to the N .

First of all, what would it mean if we had a minimum element? Let's say we had X and P , and we're using this generalized inequality with R plus to the N . If X was a minimum in

the set, it would mean that if I had X here, say this was in the set. The whole of P would lie within this. It would say that any production method requires more labor, more fuel or both than X. That would be X being minimal. Minimum is a slightly different quality.

If I had X1 here, this is a minimum element. What it means is that there is no method that requires simultaneously less fuel and less labor than X1. It's efficient. Any vector along here – there is no method that is unambiguously better. There's none that simultaneous require less of both. Labor might be more expensive than fuel, so we might want to move along the curve and trade off between them, but we don't want to move off this curve. Likewise, this part of the curve is also efficient methods.

The problem with X5 is we can produce at X5, but we could produce at X2 and we would use less labor and simultaneously less fuel. There'd be no point in picking the method X5 if these were all the parameters involved. X1, X2, and X3 are efficient, and X4 and X5 are not. You can think about these as being minimal with respect to R plus to the N. There are other ways. We might be interested in a different cone, and there might, in fact, be a minimum for a different set, but this is just tying at least one of these things into a practical application.

Any final questions? We are done, and Professor Boyd will be back on Tuesday.

[End of Audio]

Duration: 77 minutes