

## ConvexOptimizationI-Lecture03

**Instructor (Stephen Boyd):**All right. Now I guess I'll take that as the – as a sign that we're on. So actually, first of all, embarrassingly, I guess I have to introduce myself even though it's the second week. And I have to apologize for not being here last week. I've – actually, for the record, I've never done that before.

I've never missed an entire first week, and I promise I won't miss any more classes, maybe, mostly, it's – anyway, I'll make a very – if I miss anymore classes, it'll be for a really, really good reason.

Anyway, you were in very good hands with Jacob, so that's okay. A couple of announcements [inaudible], I'll start with a bunch of administrative things. The first is that we have two more TA's. This is Argerio Zimnus, who is sleeping, as most senior PhD students would be at this hour, and Cwong Mo Co is over here.

What we'll be doing is adding more office hours both for them. We now have – since we now have essentially like an army of TA's, we'll – we can have – we were joking about just having office hours like 24/7. I don't think we can pull that off yet, but we're going to do something like that, and in fact, I do have some questions. Also I'm going to add office hours for myself, and we don't know – we're trying to figure out what – I don't know, if people have opinions about where we should distribute the office hours, so I don't know. We have our own conjectures, but if anyone has any other?

For example, in my case I might add them Tuesday and/or Thursday afternoons. What do you think? Yeah? Oh, that was a yes, oh, okay. Okay, all right, fine. Then I'll do that. I'll add office hours for myself, and for the TA's.

Now this week, of course, is chaos week, otherwise known as the equals' week. If you're not in this department you'll – you actually – you cannot even imagine what this week is like, but just look around you and you'll get the idea.

If you're in this department, you know full well. So those office hours, at least in my case will kick in – have to kick in next week, but look to the website for that. Let's see, a couple of more administrative things. Here's one, and I – it has to do with the homework.

What we're going to do is we actually did a little bit of planning for the homework, and the way it's going to work, and actually it's a good excuse for me to say a little bit about how the course is gonna go for a while, and this is actually substantive.

Homework 1 and Homework 2, Homework 1, which you've already started – yes? Okay, so Homework 1 and Homework 2 are going to be just these big long slogs through – I don't know, it's – you'll have nightmares about it, and people coming up to you on the street and saying, is this set convex, is this function convex, and things like that for years.

But the nightmares will subside eventually. They'll always be there to some extent, so there'll be little triggers later in your life that will set it off again. Someone will say something like, which of the following, and you'll – something will remind you of it and you'll get all on edge.

But seriously, what we're going to do is this; we're trying to go fast over the first material, which is actually a little bit, I mean, it is interesting. We're going to use all of it eventually, but it's sorta the mathematical basis of what we're gonna do, and so the idea is to actually not cover it as well as we should.

Meaning, we're going to go fast. Faster than you – there's no way you could kind of absorb every subtlety. So what we'd like you to do is read the book. I mean, that goes without saying. Read the book, you'll do the homework, listen to the lectures, and you'll get maybe 75 percent of it, something like that.

Actually, that's enough because what's going to happen then is in – by Homework 3, it'll actually get interesting and we'll be doing a lot of applications from then on, kinda mixed in with new material, and at that point it gets interesting. So the idea is to avoid having something which for four weeks, non-stop, is essentially just a math course, so that's sorta the rational behind this.

On your part what it means is this; if you think we're going fast, you're right, we are. But it also means don't worry about getting absolutely everything. All the things we're gonna see you're gonna get lots more chances, and it'll be more interesting later because it's gonna be in an applied context.

Having said that though, this is your time to kinda get down some of the basics. Okay, so that's our plan for how the course is gonna go. I should also say that there was one change. I don't know if it was announced yet, and that is that the homework we're gonna try to have graded a little more seriously.

So the – then for example 263; in 263, the homework's were graded on a scale of 0 to 4, very crudely. That's nominally two bits. The actual amount of information from the entropy was on the – of .2 bits. We haven't worked out actually what the entropy is, but it was something like that.

So we're going to try to do – at least give you a little more feedback and make this 0 to 10. It's going to be imperfect, but we're going to attempt that. So it's going to look something like that. The idea is if it's perfect, that's a 10, if you – anyway, we'll – when we – we'll get into that later.

Another admin – maybe just two more admin things and then we'll move on to real material. The other is that we posted another reading assignment, and so right now you should be reading Chapter Two, and when you finish Chapter Two, keep reading, just go right on to Chapter Three, and four for that matter after that.

And then let's see, one more admin announcement. This is a bit specific, but it's a good excuse to mention something. We got several questions about what do inf and sup mean, and that's a good excuse for me to say a little something.

So, you are welcome – if you don't know what these mean, you're welcome to treat these as min and max, okay? So, two – one thing is actually for questions like that. When we get more than 10 or 15 questions, we're going to start a FAQ, on the website, so please let – what? Unknown Speaker:

It's already there.

**Instructor (Stephen Boyd):** Sorry, it's already there. Okay, it wasn't there two minutes ago, but anyway. So there's a – there's going to be a FAQ section in the website. Already pushed over? Okay. So there is a FAQ section in the website. And so please check there if you have some sort of basic question first. And this is anything ranging from, I can't get this software to work, to what do inf and sup mean, or when is Homework 3 due, or I don't – I guess that's posted on the website, but that kind of stuff. We'll just post stuff there.

And if you have suggestions, actually, for us to add stuff there, just send us an email. Okay. And on this topic, I should say a little bit; if you have had analysis, of course on mathematical analysis, then you know what infimum and supremum mean. If you haven't, I would simply substitute min and max.

And the way the course is going to work is something like this; the book, by the way, is not – I mean, there are many more mathematical books, so if that's your interest on your website as pointers to other books, you can go insane. And it can be as formal as you like.

By the way, outside the lectures you're happy to hunt me down. I'll tell you what I know, if this is your interest. Frankly that's not really the whole point of this class, but I'm happy to answer any question about that.

And I'm talking about things like detailed conditions, is this open, is this closed, what happens if the boundary – all these kinds of things. You know, what are the exact conditions under which this holds and things like that. In lecture I'll be very cavalier about these things.

So boundary conditions don't hold, everything I say is modulo, some technical conditions. In the book we, generally speaking, don't lie. If you look in the book there's just a few things that are wrong. Well, you can treat it as true, the book. But in lecture I'll just be happy to just go broad brush over these things.

And I think the class, by the way, is – just from past experience, if you've had – if you have a deeper understanding of all the analysis and stuff like that, great. But I haven't noticed that it makes any difference in the experience. You don't have to know it; you can work with an intuitive idea of these things. Not really worry too much about

boundary issues and things like that and you'll be just fine and totally effective when – at using these methods.

Okay, so that's all my admin stuff, any other questions? Otherwise, we'll jump in. So today we'll cover convex functions. And let me say a little bit about how this is going to work. If you go down to the pad here, that would be great. There we go.

So we're going to look at convex functions. First of all, just define what one is, and then look at various aspects of it. And let me say operationally why you need to know this; because one of the things you're going to need to do is actually determine if a function is convex because to use this material, that's like the most basic skill.

So that's why we want to – that's why you want to look at all this and see how this works. Okay. So let's just jump in. Well a function is convex if its domain is convex. That's the first requirement. And the second is that it satisfies this inequality for theta between zero and one.

So this says that if you evaluate  $F$  at – this point you saw in the convex sets lecture. This is a convex mixture, a convex combination of  $X$  and  $Y$ . So geometrically it's a point on the line segment between  $X$  and  $Y$ .

This says if you evaluate a function at a point on a line segment between  $X$  and  $Y$ , the result is actually less than the same mixture of the values of the end points. Or in terms of the graph, it says that if you take two points on the graph of the function and then draw the straight line that connects them, and I guess an old word for that is the cord, or something like this. Right, so that's a very old word, is the cord, and it says – this says the cord lies above the graph.

So that's what – that's what this inequality says. And actually you'll get a very good idea of what this means eventually. Another way to say it just very roughly is upward curvature. So it just means curves up, that's all. And by the way, of course, a convex function can look like that. It can be monotonic decreasing. Nevertheless, the curvature is upward for a function like that.

It doesn't have to be bowl shaped, but it should have positive curvature. Okay, now you say a function in concave is  $-F$  is convex. That means negative curvature, downward curvature, something like that. And it's strictly convex if it's convex. And not only that, but this inequality here holds with strict inequality provided data is strictly between zero and one. So that's strict convexity. And you have strict concavity too.

So let's look at some examples; well – and these are just on  $R$ . So these are just functions you can draw. So the first is just an affine function, so that's linear plus a constant. That's – it has zero curvature, so it's convex. And in fact what happens for a function that is affine is the following; is that in effect you have equality here. For always, for an affine function that's exactly what it means.

It means that if you do a linear interpolation between two points, you actually get the exact function value. So that's essentially the boundary of a set of convex functions. Exponential, doesn't matter what the coefficient is in here, this is convex. So if A is positive, it's increasing, but the curvature's upwards. If it's negative, it's a decreasing function, but it's convex.

Powers separate out. It depends on the values of the exponent. If the exponent's one or bigger, or if it's negative, or – well zero [inaudible] that's just a constant one, in that case it's convex. I mean, these are things you can just draw and see. So I think the question is to whether or not a function on R is convex, is a non-issue.

Here's how you check; you draw it, and you use your eyeball to see if it curves up. So there's really no issue there, okay? So these – we'll find formal ways to verify all these, but I think in terms of a function on R, there is no issue. You just – you draw, does it curve upward or not, that's the question.

And you have things like power of absolute value would be another one. Negative entropy is  $X \log X$ . That's something that goes like this; it's got this infinite slope here and then goes like that. Something like that, and – but at point as it curves upwards, that's negative entropy.

So entropy, which is minus  $X \log X$  is going to be concave. Okay? Now in concave functions examples would be like an affine function, and in fact, of course, if a function is both convex and concave, then it's affine. And that's clear because it says basically; this inequality holds one way and the other. That means its equality. This in fact implies the function is affine.

Okay? This power's in the range between zero and one, you know, like square root for example, you just draw this, and it's clearly concave. Log rhythm, it's another famous example. Okay. A little more interesting example on RN and RM by N, that's the set of M by N real matrices.

So here's – of course you have an affine function on RN, that's [inaudible] plus X plus B. That's a general linear function plus a constant. So this is the form of a general linear function, affine function on RN. And that's going to be convex, it's also concave.

Norms; so any norm is convex. That follows actually from triangle and equality, or – I mean, that's part of the definition of being a norm. And examples are things like this; the so-called p-norms, which is the sum of the absolute value of  $X_I$  and then to the one over P. Now for P equals one, that's the sum of the absolute values for P equals two. It's usually Euclidean norm, but for example, for P equals three, it's the three norm, which is the cube root of the sum of the cubes of the absolute values of a vector. Yes?

**Student:** Is there a half norm?

**Instructor (Stephen Boyd):** That's a very good question, is there a half norm? So some people would – let me answer it first – well let me just first give the answer. The answer is no because the square of the sum of the square roots is actually not a convex function, and therefore it can't be a norm because norms have to be – all norms are convex. It's actually concave.

Now having said that, it is currently very popular in statistics and a bunch of other areas to use a – as a eristic for finding a sparse solution that involves – we'll see this later, by the way, using things like the so called p-norm for P less than one, but it's not a norm, so it's weird.

**Student:** Why do all the norms have to be convex?

**Instructor (Stephen Boyd):** That actually follows from the definitions of – norm has to satisfy a triangle in equality and a homogeneity property. And then from those two you can establish it has to be convex. By the way, I should mention something here. It's not easy to show – if you put – just for general P, just take the three norm. It's not easy to show that's a convex function. It's not easy at all. So it's – I mean, it's not hard, but you have to know exactly just the right five or six steps to do it.

So this is maybe the first thing I've said that wasn't trivial, and it's not. Okay, let's look on matrices; actually, every now and then we're going to – there are going to be problems where the variables are matrices. Sometimes square, sometimes they'll be rectangular, but let's look at a couple right now.

What is an affine function on matrices? Well the general form looks like this. A trace of A transpose X plus B – by the way, when you see this you should read this as follows; this – by the way, some people write this as the inner product of A and X plus B. That's the standard inner product on matrices, is trace A transpose X.

If you work what it is entry by entry, it's just this. So it's as if – well it's this, let me write it another way, let's see if I can do this. If – that's not totally – that's mixing weird and – that mixing notation, but the point is this says that if you string out A as a vector, string out X as a vector and then calculate the ordinary inner product, well you would just get this, like that. Okay?

And that's the same as this thing. So this notation, you might as well get used it. When you see trace A transpose B – A transpose X, that really means you can think of it this way. And, by the way, another notation for this is A big dot X or something like that, so you'll see that.

Okay. Here's an interesting function, which is the norm of a matrix. So that's – and I'm talking about here, the spectral norm or the maximum singular value, or you can call it many other things, and there's not that many other – actually there's probably just a couple other names for it, the L2 induced norm and maybe the – now I'm running out of obvious ones.

So this is the square root of the largest [inaudible]  $X^T X$ . Now I want to point something out, that is very – that's a very complicated function of a matrix. So that's – it is not a simple – to take the matrix form  $X^T X$ , find the largest [inaudible] value of it then take the square root of that.

That's a chain of quite complicated operations. So that's a function which is not simple, but it's a norm and it's convex, okay? Now here's one extremely useful property for convex functions, is this; a function is convex if and only if it's convex when it's restricted to any line. That's very, very useful.

And in fact, it's one of the best ways to actually establish convexity of a function. And essentially it means the following; I already said that convexity of a function on  $\mathbb{R}$  is a non-issue. Plot and use your eyeball. This says in principle, convexity of any function is a non-issue. Now the only hitch is you have to unfortunately plot and use your eyeball on all possible lines that pass through the set of which there are [inaudible] number.

Okay, so that's the only – but conceptually it means that there's nothing subtle about convexity. It's basically if you take some complicated horrible function, multiple dimensions, and you take a line, then – and then view that function on that line, you should view something that looks like that. And if that happens no matter what line you choose, it's convex. That's what it is.

So this is not too hard to show, and may indeed be coming up on a homework or something for you. I mean, to establish it, this is the case. And let's look at an example; so here's an interesting function; if the log of the determinant of a positive definite matrix – by the way, that's a complicated function right there.

For example, if  $X$  were 10 by 10 and I wrote this out, it wouldn't – it would take like a book to write out what that is because you'd have 10 factorial terms in  $\log \det X$ . So this is a log of a polynomial of these 100 by – well okay, it's 10 by 10 so there's only 55 entries because the bottom is the same as the top or whatever. So there's 55 entries, you know, free variable.

So this is the log of a polynomial in 55 variables, and that polynomial probably would take a book. I'm just making a wide guess. By the way, it could be a whole lot bigger than a book too. I could be off a bit. But – so that's – this is really a fantastically complicated function. Actually, this function is concave. That's going to be very important. It's going to play all sorts of roles later.

So let's actually show that, that this is concave using this technique. This is – well it's the only technique you know except for applying the definition. Now – by the way, what does it mean to say its concave? Basically it says that if you have two positive definite matrices, and you evaluate the log and determinance of them, and then you form a blend of the two, let's say the average.

It says, the log of the determinant of that average is bigger than or equal to the average of the log of the determinants of the end points. Does everybody follow that? So that's – and this is not an obvious fact. I mean, once you know it it's obvious, but – well it's actually not obvious, let's just say that. Even when – once you know it it's obvious only because you know it.

Okay, so let's see how this works; so to establish this we have – what we have to do is we have to pick an arbitrary line in symmetric matrix space, okay? So what does a line look like in symmetric matrix space? Well it looks like this – and it's one that passes through the positive definite cone.

So it's going to look like this; without lost generality it looks like a point.  $X$ , which is positive definite, plus  $T$  times a direction  $V$ . Now this direction  $V$  is a symmetric matrix, but it does not have to be positive semi-definite, right, because that's a line in a direction – there's no reason the direction has to be positive – it does have to be symmetric. So this thing describes a function of a single variable  $T$ .

And we have to check that this is in fact a concave function of  $T$ , so that's what we have to do. By the way, if you're looking at some function, you have absolutely no idea if it's convex or concave. First thing to do when you're near a computer, sit down, generate arbitrary line and plot and look. Look at 10 or something like that.

And by the way, if you find one that doesn't curve up then you destroy all evidence that you did this and then comfortably say, obviously that's not convex, and then erase all the files that – it's three lines, but the point is then you erase the evidence.

If they keep coming up like this, then after 50 times it's like, well okay, here it goes. And then you roll up your sleeves and you move in to prove it. And it may not help you, but it's actually quite useful to do this. I've actually failed to follow this advice on several occasions, jumped in 45 minutes later, wandered – got tired, wandered over, typed in a few, quickly turned up, found a point where it had the wrong curvature and then realized I was just – I had just been completely wasting my time, so – okay, I just mentioned that.

Don't tell people where you learned that – where you heard that. All right, so let's work out what this is; well a right  $X$  – I'm going to write this – there's lots of ways to do it, but  $X$  is positive definite so it has a square root. So I'll take half out on each side, and this will look like this,  $T$ , it's gonna look like that, times  $X$  half.

That's what – this matrix is this. But the determinant of a triple product is the product of the determinants. And you take logs and it adds and all that, so you get this because the log of  $X$  half plus log – sorry, log det  $X$  half plus log det  $X$  half is log det  $X$ . So you get this thing here.

It's still not too obvious, but what we're going to do now is this; I'm going to write the – this thing is a symmetric, but not necessarily positive – it's certainly not positive semi-

definite or anything like – you don't know. Matrix, we'll take its eigen expansion. We'll write this as whatever,  $Q \lambda Q^{-1}$ , and you get this with a  $T$  there like this.

What I'll do now is I'll do the trick of writing  $I$  as  $QQ^{-1}$  transpose then pull it out on either side and I'll get  $\det Q$ . It doesn't –  $\det Q$  is either plus or minus one, but it doesn't matter. And then I'll end up with  $\det I + T \lambda$ . That's a diagonal matrix. I know what the determinant of a diagonal matrix is, it's a product of the entries, and so I get this thing, okay?

So I went over this a bit fast, but I promised I would go fast, so –yes?

**Student:** How did you choose the directional matrix  $V$ ?

**Instructor (Stephen Boyd):** It's extremely important that it's completely arbitrary. So the technical answer to your question, how did I choose the direction  $V$  is, I didn't. Or arbitrarily I suppose is the actual correct adverb or whatever, okay? That's – because if  $V$  is – if I chose  $V$  in any special way then my final conclusion is not gonna hold, it's not right.

To be convex has to be every line. It has to be convex when restricted to any line through the point. Okay. Now we're in pretty good shape, because I know what  $\log(1 + T \lambda)$  looks like. For any real number of  $\lambda$  – if  $\lambda$ 's positive it looks like one thing and it goes like this, or sorry, it goes like that. Nope, sorry, that was from my point of view. I did it right the first time, it goes like this. If  $\lambda$  has the other sign it goes like that; either way curvature is negative, and so this is concave.

Okay? By the way, this is gonna turn out to have sorts of implications. If you're in statistics, if you've taken information theory, communications, a lot of other things like that, actually, if you've done any computational geometry, it turns out some things you know are actually related to concavity of  $\log \det X$ .

Things involving like entropies of random variables and things like that. So actually throughout the class, when we get – when we've actually covered some material you'll see all sorts of things you know from other classes are going to start to connect to various things here.

Next topic is – this is just a bookkeeping thing. It's quite eloquent, but it – and it's actually something good to know about. It works like this; when you have a convex function, it turns out you can encode the – it's convenient to encode the domain of the function by simply assigning the value plus infinity when you're outside the domain.

So if you have a function  $F$  with some domain, then we simply – we define an extended valued extension as follows; it's gonna agree with the function if you're in the domain and we'll assign it infinity outside the domain. And, I mean, let's not worry about this here, but technically there's a difference between  $F$  and  $\tilde{F}$ .

And the difference occurs when you call  $F$  at a point outside the domain. If you call  $F$  at a point outside the domain, you get the dreaded – what's returned as the dreaded NID token, otherwise known as ‘not in domain’. Of course, what would actually happen is some exception would be thrown at you or something like that, or a NAN would come back or something like that, but that's what that is.

Whereas  $F \tilde{}$  evaluated at a point outside the domain is of  $F$ , simply returns the token plus infinity. Okay? Now what happens is then everything kinda works. So if you have a function, sort of it looks like that, and its domain is from here to here, that's a convex function. What we simply do is we simply assign it the value sort of plus infinity outside; here, that's plus infinity, okay? So it looks like that, so you make it plus infinity.

Actually everything works, including this inequality if you say – if you take this point and this point, everything works. If you draw – everything will work because this line will now have a slope going straight up and it all just works.

So this is just something to know about, it's absolutely standard in convex analysis. So it's – you should know about it. Now for concave functions it's the opposite. You encode the domain – or I should say, you encode the not domain as minus infinity in the function value, okay?

Now we get to first order of condition. First of all, just to remind you, a function is differentiable if its domain is open. I mean you could also talk about being differentiable at a point, and you'd say that it's differentiable to point if the point is in the interior of the domain, but we'll just talk about being differentiable period.

So a point is differentiable if its domain is open and the gradient exists at each  $X$  in [inaudible]. Actually, this is a bit circular and what you really want to say – you really want to say this is not – this is informal so I'll just leave it that way.

So here's the first order condition for convexity; this is very important, and actually it's a hint as to why convex opposition actually works very, very well. Here it is, it says the following; form that is the Taylor approximation. The first order Taylor approximation of  $F$  at  $X$ . Okay, that says as a function of  $Y$ . That's the Taylor approximation.

So there's  $F$ , here is this Taylor approximation, of course this is drawn in  $R$ , but in general this is the Taylor approximation. And of course what the Taylor approximation is this; it – at the point in which you expanded, it's perfect. In other words, it coincides with the function. Nearby, so near  $X$ ,  $F$  – the Taylor expansion is very near, by which I mean to say, formally near squared.

So it's small – the error is small squared. So what the Newton tells you or calculus tells you is that these two functions, one is this affine function and one is this one, is that they are very near as long as  $Y$  is near  $X$ . For a convex function this thing is actually always an underestimator. So that's the important part.

It says that the Taylor expand – the first order of Taylor expansion is a global underestimator of the function. That's what it means. Now, what – let me – I mean, this is actually the key to everything. I mean it's – once you know all these it's kinda trivial, but this is the key to everything because let me explain what happens; later in the class we'll formulate all sorts of crazy problems as convex problems.

And we'll come back and we'll say, I've solved it, this is the solution. It will be some horrible complicate problem with hideous no non-linearity's, things that are non-differentiable, god knows what's in it. And someone will say, well how do you know that's the global solution? That problem is highly non-linear, it's got all – I mean, that's ridiculous.

I mean maybe it's a local solution, I'd believe that, but how could you possibly assert that in all over our 100 there's no point that does better than your point. Maybe there's some sick little region you haven't examined yet where the function miraculously does better.

Everybody see what I'm saying? And it's really a preposterous statement that you're making. I mean after a while you'll get used to it, but it's preposterous. And then you said, oh no, no, because this is a convex problem, I assure you this is the absolute best there is.

And someone will say, how can you say that, it's ridiculous. This is it, and the reason is this; from local information, that's a gradient, a gradient is local information. It says, from local information you get global conclusion, which is this inequality. So just – although the inequality is not a big deal, the fact that it says something that you evaluate a gradient somewhere, that's completely local.

To evaluate a gradient all you need to do is wiggle X around near that point and see how the function varies locally. You don't have to do any big exploration far away. What is says if that function is convex then from that little local exploration you can make global bounds on that function that hold arbitrarily far away. Everybody see what I'm saying?

So not a big deal, but this is – if you ever – what's going to happen is three weeks into the class things that are just preposterous will be asserted, you'll be asserting them in fact. And if you ever wanted to go back and say, well what – you know, but you know how it always is, right, there's just a string of little trivialities and the next thing you know you've said something quite deep, and then you always want to go back and say, where exactly did that happen?

Anyway, I say it happened here. That's what I say. Okay, second order of condition; I guess you've seen this. This you've probably seen in a calculus class or something like that. Usually it has to do with sort of this part, the suficion conditions or checking if something's a minimum or maximum or something like that.

And it basically says that the – it says, if it's twice differentiable, it says it's a hessian, which is the makers of partial derivatives. If that is – it says the following; it says it's convex if and only if that matrix is positive-semidefinite, okay? So that's the condition.

And then there's a gap in characterizing strict convexity. And the gap says something like this; it says, certainly if the hessian is positive-definite everywhere then the matrix – I'm sorry, and then the function is strictly convex. Actually the converse is false and an example would be  $S^4$  to the fourth on  $R$ . That's a strictly convex function;  $X^4$  to the fourth, but its second derivative at zero is zero, okay?

That's a case where the second derivative fails to be positive at the origin. So this is the condition. So this is useful sometimes. Actually, this is less useful, in fact, and also when – although you will have to – I mean, we'll see to it, let me put it that way, that you'll have to use this method to establish convexity of something. But as you'll see later, this is to be avoided because generally it's a point to show something as positive-semidefinite, unless it's really simple, the function.

Now we can knock off some examples real quick. Let's characterize all quadratic functions right now. So a quadratic function looks like this; it's a quadratic form, it's a p-asymmetric here, a liner function and a constant, okay?

So the gradient of this function is  $PX + Q$ , and the hessian is  $P$ . So that's the – it's constant, it's got a constant hessian. So a quadratic function is convex if and only if  $P$  is bigger than or equal to zero, if it's a positive-semidefinite matrix. And an example would be least squares objective. That just – immediately. So here the hessian is  $A^T A$  and that's going to be positive-semidefinite so that's always convex.

Here's a function that is not obvious, it's one – this is probably one of the first functions you encounter that is not obvious, it's not obvious that it's convex. Maybe other than  $-\log \det X$ , but here's one;  $X^2$  over  $Y$  just two variables. So it's – this is convex in  $X$  and  $Y$  provided  $Y$  is positive.

And let's take a look at that – first of all, let's just check a few things. If you fix  $Y$  it's convex in  $X$  that's clear. If you fix  $X$  it's convex and  $Y$  because it's a one over  $Y$  in that case, okay? Now by the way, I'm using there this idea that these are essentially lines. I mean that's a line, one is aligned with the  $X$  axis, one with the  $Y$  axis. So if a function of many variables is convex, it better be convex in each variable separately.

The converse of course is completely false; you can have a function convex in each variable separately, but not jointly convex. And the answer is – I mean, the distinction comes completely from the Hessian, right? So you have a, for example, a function of  $X$  and  $Y$  variables, you have a two by two hessian that is required to be positive-semidefinite.

To say that it's convex in X and Y, it says the diagonals of that hessian are non-negative. That's what it means. To – that tells me nothing about the off-diagonals, so there's a further condition linking the two and the cross condition. Question?

**Student:** Is there any way to prove convexity or show convexity from [inaudible] projecting it on to one dimension [inaudible] number of times and –

**Instructor (Stephen Boyd):** Be careful because we're not projecting onto a dimension, we're restricting to one dimension. And the answer is, yes. In principle if you restrict it to any line, any sort of one dimensional set that passes [inaudible] and the result is always convex, then it's convex, but it has to be understood that's in principle unless you're prepared to do something symbolic with every line like we just did with log det.

Okay. So let's take a look at this function, and let's show that it's con – by the way, here's a plot of it, so it looks – I don't know, some people have told me it looks like a boat or something like that, the bow of a boat or – I don't know, anyway, so and if you look in – I guess, one slice it's quadratic and it's – I guess it's – I don't know what it is anyway so, if you slice it different directions you get obviously convex things.

So lets work out the hessian, well I worked it out, and I'm not going to calculate in front of you all the partial derivatives of the things here, but you calculate partial squared F, partial X, partial Y, and the diagonals, and you fill it out and you get a matrix which indeed is rank one, and can – and positive-semidefinite because you write it this way.

And here we're using the fact that Y is positive here, to ensure that this is positive-semidefinite. So that's convex, so that's your – maybe one of your first non-obvious convex functions. Here's another on; the so called log sum X function. That's the log of a sum of exponentials of variables. Actually, before we go on let me just say a little bit about this function; it's – first of all it's sort of like a smooth max, or in fact some people call – there are fields where this is simply referred to as soft max.

So I know whole fields where this is universally called the soft max. The reason it's soft max is this; if you take a bunch of variables, X1 through XN, then X increases of course very quickly. So the biggest X is a lot bigger. If there's a good gap between X – the largest X and the next one, X will accentuate the spread.

Okay? If you add these up and then take the log – if for example one of the X's is kind of isolated and far away from the second – if the biggest is away from the second biggest then you can actually quickly see that basically this sum X is basically X of the largest. You take the log of that and you get the largest. So this is sort of a smooth – it's a soft max some people call it.

I should also mention this is – if you're in electrical engineering, this is the DB combining formula. So this is how you combine powers that add incoherently. So if you have for example, if you have a bunch of powers adding, and the powers come in at, minus 15 DPM, minus 22, minus 37 and minus 3DBM, everyone in here knows what the

power of the result is, except I forgot the numbers I just listed, but I think it's minus three with the largest one. The answer is it's minus three.

So, whatever that is, minus three DB reference to a millawatt. And let me explain what that is; X convert from decibels to power. This does incoherent power addition and log converts back to decibels. By the way, obviously if you're not in electrical engineering you don't have any idea what I'm talking about, and that's just fine.

So I'm just saying this is a function, you've seen it before and it comes up in statistics as well in exponential families, this comes up all the time. Okay. So, in a little statistical mechanics too, it's the denoninate – a fancy version of this is some normalization constant or whatever. I'm sure people here know a lot more about this than I do.

So that's a convex function, and the argument would go something like this; you take the hessian – by the way, if you're curious, no one can look at this and write this formula out for the hessian. Just, you know, so if you're looking at this and saying like, oh, that looks pretty easy, anyway, it's not.

You have to sit down and first you try to find some secret rules for – chain rules for calculating the hessian of some function, that's the first thing you do. Then that gives you a headache. Then I usually just basically – at some point you give up and you just go in there – you just calculate partial squared F, partial XI, partial XJ, there's just no way around it. You get this huge horrible mess then you have to go from your mess, your index by index representation to a matrix representation and then you're off and running.

Then you go back and erase all that and then write things like this. So – just like this and then that's the idea, but lets take a look at it and see what it says; now when you do this you look at it and you see well there's this first term, Z here is this X of X, so it's obviously – it's non-negative.

This think here, it's interesting. That's a non – that's a positive diagonal matrix so this is positive definite – positive-semidefinite – well it's positive-definite. And of course what you're hoping for is to add that to something positive definite at which point you – it's easy because you just say the sum of two positive-semidefinite matrices is positive-semidefinite, done.

And there's a little minor problem, which is that this sign goes the wrong way, and that could mean two things; it could mean either this is not convex, or you have to work harder to show this thing is positive-seimdefinite, and it's the latter here.

**Student:**Is Z a vector here?

**Instructor (Stephen Boyd):**Yep, Z's a vector.

**Student:**What is diag of [inaudible]?

**Instructor (Stephen Boyd):**What is diag of?

**Student:**[Inaudible]

**Instructor (Stephen Boyd):**Of Z here?

**Student:**Yeah?

**Instructor (Stephen Boyd):**What is diag of Z? Oh, diag of Z, you take a diag – this is actually entering – of course its mat lab slang or something. But it's entering – it's sort of entering mainstream mathematical notation.

**Student:**[Inaudible]

**Instructor (Stephen Boyd):**Exactly. Yeah. So diag of a vector imbeds the vector into the diagonal elements of a diagonal matrix. What department are you in?

**Student:**Double A.

**Instructor (Stephen Boyd):**What?

**Student:**Double A.

**Instructor (Stephen Boyd):**Double A, okay. Now wait a minute here, have you never used [inaudible]?

**Student:**A long time ago [inaudible].

**Instructor (Stephen Boyd):**It was a long time ago.

**Student:**Maybe.

**Instructor (Stephen Boyd):**Maybe, okay.

**Student:**Last year, I think.

**Instructor (Stephen Boyd):**Okay, last year. Okay, fine. Okay, that's fine, just checking. I thought you were going to say math or something like that.

**Student:**[Inaudible]

**Instructor (Stephen Boyd):**Okay, all right. So it was in a CS class last year, which – the memory of which – you didn't remember what class it was or – what was it?

**Student:**220.

**Instructor (Stephen Boyd):**220?

**Student:**Eight.

**Instructor (Stephen Boyd):**Eight, okay, I figured. Okay. So back to our thread here, it's not over yet, you still have to show this is positive-semidefinite. How do you show a matrix positive-semidefinite? You simply show that the associate quadratic form is always non-negative. So you put a V on the left to be transposed, a V on the right. You plug that in here and see what happens and you get this thing.

And this turns out to be greater than or equal to zero using the Cauchy-Schwartz inequality applied here, okay? These are not – this is not obvious. Okay? This is not something that you just type in. This is like 30 minutes of thinking and breaking pencils and things like that.

Oh, you know, I just remembered something, speaking of that. I don't know that it's been announced yet, but I – it's just a suggestion. A suggestion actually it's not a bad thing to work on homework in possibly small groups of people depending on who you are. I mean people already probably have the sense to do this anyway, but you're being officially encouraged to work in small groups like that. In which case I could say so after 30 minutes of working this, breaking pencils, throwing pencils at the other people in the group and things like that.

Actually working in a group is very good because it's easy to think you understand something, very easy, when you're by yourself. And it turns out if you have to address like two or three other people and try to explain it – actually, you know while you're explaining it you'll realize that so your mouth is going like this, right, and there'll be something in the back of your head, I don't know what part of it is, will actually come out and say, by the way – also just looking at the others in the group you realize, this makes no sense whatsoever.

I mean, what you're saying makes no sense. So it's a very good [inaudible]. Whereas a lot of things you know by yourself can be totally obvious. Totally obvious, then you try to explain it to someone. Also, if you try to explain it and it's this long complicated – you go, look, it's super simple, it's so easy. You just – okay, you do this, but then it's, well hang on, I forgot to say this.

And then also, by the way, do you remember this and then you put that there – this is also a hint that this is way too complicated, your description of it, so you're encouraged to work in small groups. All right, let's move on. So a geometric [inaudible] is concave, so this is the product of a bunch of variables, these have to be positive, although actually this works for non-negative.

So it's a bunch of variables and then the  $n$ th root of that, so that's concave. And it's the same type of argument as for log sum expH, same type of thing. Okay, some other

connections, the idea of an epigraph and a sublevel set. So if you have any function then the sublevel – the alpha sublevel set is the set of points with  $F$  value less than alpha.

By the way, later in the course  $F$  will be an objective or it will be some type of objective, or something like that. And for example, if  $F$  is a design, then if  $F$  represents the power dissipated by some circuit design or something like this, this will be the set of designs that meet an alpha spec.

That's what this is going to mean, okay? That's what a sublevel set will often mean. By the way, if it's an estimation problem, an  $F$  is a measure of implausibility, like negative log likelihood in a statistical estimation problem. This will be the set of points, which are at least alpha plausible values, okay?

So something like that. All right, now if you have a convex function then the sublevel sets are convex. By the way, the converse of that is false, but it's a very important thing and we'll have a name for it very soon, a function whose sublevel sets is convex.

The real correct connection between convex function, convex set, because we've overloaded the word convex now to mean two things; it applies to sets and it applies to functions. The real connection is through something called the epigraph. So I guess epi means above, and so epigraph means everything about the graph.

The graph of a function of course, is the set of pairs,  $XY$ . So it's an RN plus one, it's the actual graph. The epigraph is everything above it. So if I have a function, the epigraph is this shaded region like this. And here's the real connection between convex set and functions.

A function is convex if and only if its epigraph is a convex set. So – and in fact, it turns out you should just be thinking about convex functions in terms of their epigraphs just always. So, for example, when I say here's a function like that, in fact let's put a domain restriction on it from there to there, you should just immediately visualize this set drawn in, and it's a convex set.

Actually this is going to be important because we're going to do a lot of calculus on – with convex functions and you want to think about what does it mean in terms of the sets? That's what you want to be doing. So this you should just be thinking of at all times, the epigraph.

Okay. Let's look at Jensen's inequality; so our basic inequality for convexity is this; it says if you take a point between zero and one, if you take two points and take some kind of weighted average of the two, that's what this is, and evaluate it, it's – so this says that  $F$  of the weighted average is less than the weighted average of  $F$ . I said it right. I always forget this.

I'll show you a mnemonic in a minute, which you can sneak off and do if you have something to draw on. You can even do it if you can't draw, but it takes me longer. Now

there's tons of extensions of this, for example, instead of just two points you could have a finit number of points and a bunch on fada I's that add up to one and are non-negative. That's a convex combination.

And the same inequality would be true thinking of an accountable infinite number of points, some combination that way, and then you can have an infinite one. The most general is this; if  $F$  is convex, then  $F$  of an expected value is less than or equal to expected value of  $F$  of  $X$ . Now that's called – that's Jensen's inequality.

And this is where  $Z$  is a random variable, however, which is in the domain of  $F$  almost surely. So that's this, and this is Jensen's inequality. And this basic case is nothing by this; it's a distribution on  $Z$ , extremely simple. It takes only two values,  $X$  and  $Y$  with probability theta and one minus theta, and you recover this thing.

So this is Jensen's inequality. I never remember Jensen's inequality, especially because whenever it comes up usually  $F$  is half the time convex, half concave. Of course if it's concave it goes the other way. And so I actually never remember it and I usually have to just draw a picture and remember this thing. There's another way to say it though, I know what that is; so if you want a general pneumonic about how it works, it basically says that for a convex function dithering actually hurts.

And let me explain what that means; it means that if I have a function – a point here, and here's a convex function, like that, okay? Let's imagine, that's actually my target point in some process, but now, actually when I manufacture it, I actually get a distribution of values like that, okay? Whose mean is this point, everybody cool on this?

So that's the – this happens, so this is – and then this tells you the cost, okay? And this could be the power, I don't really care, something like that. Or the speed of some – lets make it the power. So this is the power, and it basically says, you know, look, these points, that are where the manufacturing [inaudible] they were on your side. The threshold voltage went up or whatever, something happened, and it actually dissipates less power than your nominal design.

Everybody see that? And this is where it worked out badly. And now I ask, what's the expected value of the power? What is it? Well the first thing you would do is you'd say, well look, it's around here, because if I approximated this by an affine function and I propagated this distribution through an affine function, then its mean would be the same. So the first quick answer is something like this; well it's the same, if this is like one watt, and you have these manufacturing variations, sometimes you're less than a watt, sometimes you're more.

And it says that the average is going to be on the order of watt, you know, like a watt. Everybody got that? But the fact is its worse – the average is worse than a watt. It's like 1.05, it can't be better. And the reason is this; it's true that when you go up and down here,

The first order approximation, in that case sometimes you're less than a watt, sometimes more, but because the curvature is upward you get – let's see, you pay more when it's bad then you recover when it's good. Did that make any sense? And that's entirely due to the curvature. So that my pneumonic for Jensen's inequality. It basically says manufacturing variation generally isn't good. Yes?

**Student:** What [inaudible]?

**Instructor (Stephen Boyd):** That's your question? So this is a good time for me to announce this. There will be periodically times, I'd say multiple times per lecture where I'll go in something and I'll make no sense whatsoever. Its best – if this starts happening like three or four times a lecture then you can let me know, but that's good feedback.

It's best at least for now in the spirit of just rushing forward, we'll just move on. Did anyone understand what I said? A handful of people, they're just probably just being polite. Okay. You'll learn to just move on. Now, let me – actually, let me – this is a good point to kind of explain – give a sign post and say where we are and how this works.

What you're gonna have to do is you're gonna have to look at functions and figure out if they're convex or not, that's what you have to do. So the methods I just looked at involving lines and all that, if you have to resort to that, if you actually write out a hessian, I mean this is to be avoided to be honest with you.

This is – you do this only in a last resort. Of course every now and then you have to; we'll arrange that you will have to because everyone should have to do this once or twice or something. But the point is the right way to do it is using a convex calculus. And so the right way to do it is this; what's gonna – this method three, just sort of just general method is gonna work like this, you'll learn a bunch of atoms.

You've already seen a bunch, affine function, powers, log sum exp,  $X$  squared over  $Y$ , norms, quadratic functions, thing like that, where once you know it, minus  $\log \det X$  you know it's convex, okay? Then we're going to look at a calculus. And a calculus meaning methods to combine these and rules for showing it's convex. And you saw this with convex sets last lecture.

But here there's gonna be a bunch – there are gonna be some simple ones here. Now in this rule set, these divide into what I would call sort of the really obvious basic ones and then there's sort of that intermediate tier, and then you get into the advanced ones and the ultra advanced ones and things like that, there's sort of no limit on these things.

So I will tell you when we move into different levels of esoteric on this, but – okay. So these are extremely basic sense of following. If you have a convex function and you scale it by a non-negative scalar, it's convex, that's totally obvious. If you add two functions that are convex, it's convex, okay? And that extends to adding five functions and it goes to an integral or even an expected value of convex functions would be convex.

Composition with an affine function; so if you pre-compose with an affine function, so in other words if you apply an affine function then a convex function, you get something that's convex, and it's – many ways to check this. Actually, just directly is simple enough, just with the old theta in there. Okay? Very simple to show these things.

Let's look at some examples; we can make some examples now.  $F(X) = \log(\sum_{i=1}^n \theta_i)$ , where  $\theta_i > 0$  for all  $i$ . This is defined on the interior of a polyhedron defined by  $A_i^T X < b_i$  for all  $i$ . The interiors where that's strictly less, and that's where this makes sense.

This is by the way called the log barrier for this polyhedron, and you have a minus sign here. So you know, that's kind of a very complicated highly non-linear function of  $X$ . You could work out the gradient; you could work out the hessian. This one wouldn't be too bad because if you work out the hessian, which I would not recommend doing by just getting in there and slogging it out, calculating partial squared  $F$ , partial squared  $X_i X_j$ , but using some of the rules for calculating Hessians, some of these are in the – these are in the appendix by the way.

It would work, but the easiest thing by far is to say this, here's my function, ready, I'm going to apply – here, I'm going to take this function which is  $f_i(Z) = \log(\sum_{j=1}^m Z_{ij})$ , okay? So that's it, it takes a bunch of variables, takes their logs, takes minus sum. That's a convex function. Why? Well each of these is a convex function and a sum is convex. That's a convex function. Again, nothing stunning yet.

Now we're going to simply pre-compose this function with this affine mapping. I'll call it  $B - AX$ , that's an affine mapping. And that gives you this function and then here, I just applied this, this composition rule. Here's another one; is the norm of any affine function, so  $\|AX + V\|$  is gonna be affine – convex, sorry. It's going to be a convex function of  $X$ .

Okay. Here's one that maybe not totally obvious; so here's one convex function and here's another one, like that. The point wise maximum is this function here; it looks like that, okay? So at each point it's the maximum of the two – of the functions, okay? It looks like that.

By the way, in terms of epigraphs, what does this correspond to? Precisely. So calculating point wise maximum of functions, in fact you can even write some silly formula for it, you know something like this;  $\text{epi } (\max_{i \in I} f_i) = \bigcap_{i \in I} \text{epi } f_i$ , something like that, okay?

So that's the correspondence here. That preserves convexity, okay? And you know, it means for example, here this function which is the max of a bunch of affine functions as a Piecewise linear function, that's convex. Obviously not all Piecewise linear functions are, but any Piecewise linear function expressed in this way is convex.

Here's one, this is again, not obvious. The sum of the R largest components of a vector. So take a vector in  $R^{50}$  and the sum of the top three elements. That's a very complicated Piecewise linear function, it's convex.

Why is it convex, because it's the maximum of  $A^T X$  or  $A -$  this is for the sum of the top three, is any vector with three one's and 47 zeros. Now there's a giant pile of those, there's 50, choose three. But the point is, it's the maximum of 50 choose three linear functions done. It's convex.

So you have to watch out here because you slowly sneak up on this and you'll find out after a while you've actually done something and some of these things are not obvious. Proving this by another method would really – could be very, very painful. I mean, this one, hessian doesn't even exist, it's not even differentiable, which would save you the horrible trouble of getting in there by hand and working out a hessian.

But just proving that directly would be a real pain. Now this maximum business extends, and it extends to an infinite max or a point wise supremum. So here's the statement; if you have a function of two variables,  $X$  and  $Y$ , and suppose it's convex in  $X$  for each  $Y$  in some set, you don't even know what the set is, totally irrelevant what the set is.

Finite, infinite, makes no difference whatsoever. Then it says that if you take – again, you can leave this as max if you haven't seen this before, this is simply – if you simply take the maximum over this possibly infinite collections of functions point wise you get a new function, that's going to be convex.

And here's some quick example; let's look at a couple of – let's start with this one, let's take the maximum eigen value of asymmetric matrix. Now we discussed that before, that is a really complicated function. For a matrix that is six by six or bigger, there is no formula for this, none, because there's no formula for the roots of a sixth degree and order and higher order polynomial.

I mean, you don't need to know that, but it's a good thing to know. Well it's not useful; it's just a cool thing to know. This is really a very complicated function, the maximum eigen value of a matrix, but watch this. If the supremum of  $Y^T X Y$  over all  $Y$  that have – over the units sphere. The units sphere is the set of all points whose norm is one.

Now let's check me and see how the argument goes. Look at  $Y^T X Y$ . Now by the way, when you look at that you are – at this point you are trained and wired, all by the quadratic part of your brain should be lighting up. This has been proved in FMRI studies, okay? But the point is actually here the variable is capital  $X$ . What kind of function of capital  $X$  is this? Remember, you have to suppress the flashing quadratic neurons. It's linear, that's a linear function of  $X$ . This is sum,  $\sum_{I,J} Y_I Y_J$ , it's linear in  $X$ .

Okay? So for each  $Y$  that's a linear function. This is a supremum of an infinite collection of linear functions. In fact, there's one for every point on the units sphere in our end.

Supremum of a bunch of linear functions, linear functions are convex. Supremum over these things is convex, that's a convex function. And now you know something that's not totally obvious.

I mean, it's not that unobvious either, but it's – this would be – since that's like a two line proof or something like that, that's not bad. And of course it's a disaster if you actually try to write out what lambda max is. I mean, just – you write down square root of – or that's not lambda max, but if you write down the largest of the lambda I's for which the character [inaudible] vanishes, it's all over, you'll never recover, this is an example.

Let me hit the next one, is composition. So under certain circumstances composition preserved convexity. So let's see what that is; if you have  $H \circ G$  of  $X$ , -- so the rule goes like this, and I'll – the famous one is this, it says that a convex increasing function of a convex function is convex. Okay?

So if – this thing will be convex if you have a convex – if the outer function is convex and increasing, which – I mean non-decreasing, okay? So that's the condition. And the way to derive these as other ones, for example, it's convex if this thing is concave and  $H$  is convex and non – and decreasing, roughly is what it is.

Now the way to check these is simply to write out the chain rules. So if you take – you imagine that these functions are differentiable and of one variable and you work out the second derivative and you get something like this. And from things like this, this is how you read off the rules, by the way, these things can – unless you're doing this all day long these rules, you can easily forget them.

What you should remember is there are composition rules, and you have to go back and [inaudible]. Whenever I'm somewhere and I have to figure this out I actually quietly write this out and then figure out the rules myself. By the way, the rules hold even when these are non-differentiable. You don't need any derivatives. This is just to check.

And lets actually – maybe we can make up a new rule, let's make one up for fun, ready? All right, let's see, let's try here, lets say that  $G$  – I'm not gonna remember what I'm saying, so I'm gonna write it down. Suppose I told you – I have to get this right – suppose I told you that  $G$  prime is positive. So  $G$  is going to be increasing, and I told you – do I even need  $G$  prime? No, sorry. I take that back.

I'm going to tell you that  $G$  is concave. So that means this here, and lets make – lets see if I can get it right, and lets – let me ask, what are the conditions on  $H$  to make  $F$  concave? Does that make sense? So what we want is – we're gonna say that that's – we're gonna assume that that's less or equal to zero, and I need thing to be less than or equal to zero. Well that means this has to be positive. So I have to have  $H$  prime positive. So  $H$  has to be increasing.

And then I look over here, that's positive no matter what, so I need this to be less than or equal to zero, and I just made up my very own new composition rule, and it's this – I

hope I get this right. Okay, let me go very slowly, if  $H$  is concave and increasing and  $G$  is concave – okay, ready, so a concave increasing function of a concave function is concave. Isn't that right?

Anyway if I got it wrong, I don't care, you get the principle. Okay, it's a 300 level class, I can mess up minus signs all I like, that's your job to fix them and stuff like that. So I think I said that right. So let's look at some example – by the way, the only one I actually remember, to be honest with you is a convex increasing function of a convex function is convex.

Then I have – I remember one other thing, there are composition rules and then I have a note attached to that, there's lots of them and they get very confusing, although there are people, I've noticed who just know them immediately. They'll say, oh yeah, not that's a concave increasing function – no, decreasing function of a convex function, that's con – something or another and I'm like, really? I don't know. Then they have some internal pneumonic or something for doing this, but I don't know them just to tell you the truth.

Let's look at some examples; the  $\text{expt}H$  of a convex function is convex. And that's by the way is very interesting. It says that the exponential actually can only – it preserves positive of a curvature. If a function curves up, [inaudible] also curves up. That's what it's saying geometrically.

Inverse is interesting. It says the inverse of – it's not the inverse, it's the reciprocal, there you go, that's the English word, the reciprocal of a positive concave function is convex. So, for example, let's see if I can get an example of that. One over the square root is a – that would be a convex function, and indeed it is, it's  $X$  minus one half, and it goes kinda like that – sorry, for you, like that.

Actually, by the way, that's fine because flipping the axis doesn't change the curvature property, so I don't have to draw it for you. If it looks convex for me it looks convex for you. So that's the composition one. There are then vector compositions ones and I'll say – I'll give some examples here. So here you have vector composition, so here I have a function of – a multi-argument function of a bunch of other functions.

And now, of course, the possibilities just explode because you have horrible things where you have a function that's sort of increasing in some components, decreasing in others, the arguments themselves are either convex, concave, and it gets very complicated.

But here you have something like this;  $X$  is convex if all of the functions are convex, that's the easy case, and  $H$  is convex and decreasing in each argument – sorry, non-decreasing, roughly speaking increasing. By the way, there's one subtlety here that I want to point out, there a tilde here.

Tilde is the extended value extension. And I'm not going to spend time in lecture going over this, but you want to read that part of the book at some point about this tilde because that's not a typo, it's very important.

Let me give a couple of examples here. By the way, what this shows is that actually the earlier rules that you learned are actually – they can be derived from this. Let me give an example; how about the sum of convex functions? So here's a function  $H$ ,  $H$  of  $Z$  is one transpose  $Z$ , it's the sum of the  $Z$ 's. Okay?

This function is – well let's work it out. It is certainly convex, right, in  $Z$ ? It's also increasing in each argument, do you agree with that, because it just sums the  $I$ . So it's obviously increasing in each argument. Therefore, by this composition rule, it says that if I compose this with a bunch of functions, each of which is convex, the result is convex.

So I – so from this rule I've rederived the simpler rule which is the sum of convex functions is convex, everybody see that? Let's try one more. Lets try  $H$  of  $Z$  is the max of  $Z$ . Well the max function is itself convex. It's also – it's increasing in each argument, I mean that's clear, right?

If you increase any element of a vector the max doesn't go down, so it's non-decreasing. Therefore, this subsumes actually several of the earlier ones. Actually it turns out there's only two rules. There's this rule – in fact there are only two rules, there's this rule and there's the affine pre-composition.

But it's good for a human being at least to think of them as eight or 10 rules or something. Quick question?

**Student:** Increasing [inaudible] do you mean also that when you – that they increase from vector to vector, or just that – [Crosstalk]

**Instructor (Stephen Boyd):** Nope, I do not mean that. I mean this – I mean, hold all element fixed except one. Increase one the function cannot go down. That's what non-decreasing in each argument means. Okay, so I think we'll quick – this covers, these are the basic rules. By the way, this was very fast, [inaudible] we'll cement this, don't worry, and we'll continue next time. And then maybe even by next week it'll get interesting.

[End of Audio]

Duration: 77 minutes