

ConvexOptimizationI-Lecture04

Instructor (Stephen Boyd): Great, I think we're on. Let me start today with a couple of announcements. Let's see, the first one – you can go down to the pad, is that we have posted homework 2. So – which I'm sure you'll be very happy to hear about.

And homework 2 by the way is – I promise, sort of the last one like homework 1 that's this kind of slog through interminable math. So it'll be the last one. And then after – and starting homework 3 I actually promise it'll start getting interesting. So that's our promise.

Let's see, the other announcement – I guess this one's important; next Monday there is a holiday. When – that's normally when the section is. There's gonna be a section and we are trying to figure out when it's gonna be.

We've been offered two, like not optimal times. I think one is like Tuesday – next Tuesday at 11:00 a.m.

Student: [Inaudible]

Instructor (Stephen Boyd): Wait till you hear the other one. Wednesday – oh, I don't remember this one – I don't know, I can't remember anymore. Kwangmoo, do you remember? Wednesday at 2:00 p.m.?

Student: 2:15.

Instructor (Stephen Boyd): 2:15.

Student: Tuesday.

Instructor (Stephen Boyd): Now Tuesday, what do you mean?

Student: [Inaudible]

Instructor (Stephen Boyd): Wait, well we have random – okay, does anyone have like, I don't know, violent or other – can articulate a very good reason why it shouldn't be either Tuesday or Wednesday? Our thought was that Wednesday was pretty close to the homework due time or something like that.

So Tuesday – Tuesday at a crappy time. It's also in a crappy room by the way also, so Tuesday crappy time, crappy room. Sure, okay. Hopefully that hasn't been taken now, so we'll – this obviously we'll announce – we'll announce this on the email list, and we'll also of course put it on the website, so when we – when that's fixed, but there will be a section that would have been Monday, it'll be Tuesday.

No, wait a minute; I can announce it in class, so – which is strange. We're gonna repeat that when there's another Monday holiday later in the quarter, and we don't feel like missing sections, so whatever happens next week is going to happen one other time during the quarter.

Okay, another – another announcements, my office hours, I had promised to expand them, modulo quals, and in fact I – it looks like, you have to check though, it looks like I actually will have some time today after class, and then also after 3:30 today, so I'll be around if people want to come by.

And then sometime of the weekend, or whatever, I'll figure out when my actual office hours will be. But I'll put something – there'll be something in the afternoon, like maybe Tuesday afternoon or something like that, and I'll coordinate that with the TA office hours.

And then that'll of course be announced. Okay, any questions about last time, our whirlwind tour through – or actually really like a sprint through convex functions? If not, we'll finish that up today. Is it – a ton of material here and the only – as I said last time, the important part is that you should get a – just sort of the basics, get the basics down and then some of the more advanced stuff you'll get when you're forced to use it.

Student:I have a question.

Instructor (Stephen Boyd):Yes?

Student:When we were talking about matrix space and restricting it to a line to determine the [inaudible] –

Instructor (Stephen Boyd):Right.

Student:How do you ensure that when you add a multiple, to make sure you can stay in that same space?

Instructor (Stephen Boyd):Oh, well the – right, so the question had to do – we were analyzing convexity of a function of matrices, which I forget what it was. I could look back here, but I won't. Of symmetric matrices, and we had to generate an arbitrary line in symmetric matrix space. And we did that by picking a symmetric base point, that was some X_0 or – I don't know how I labeled it, and then a symmetric direction, and that was this; $X_0 + t V$.

Now it's a sub-space, or quite clearly if you add a symmetric matrix to a scalar multiple [inaudible] you get a symmetric matrix. So this is a symmetric matrix. Now there we also assumed, without loss generality, that X_0 the base point, was positive-definite because I think it was the log get maybe we were looking at, or something like that.

This was assumed to be positive-definite, but V was not, V was a direction. I don't know if I answered your question. I think I just said what I said last time. Are you – how's your cognitive state? Is it improved?

Student: So what you're saying is when you add matrices of the same type as that one, you stay in this?

Instructor (Stephen Boyd): It's a sub-space, absolutely. A set of symmetric matrices is a sub-space, and a linear combination of – a linear – a multiple of one to another, you're in the sub-space. Okay, let's move on. And let me just remind you where we are – what the big picture is.

The big picture comes down to this; you're gonna want to – well two things, you're gonna want to construct convex functions, but you're also gonna want to deconstruct convex functions. And what I mean by that is that there'll be some function that's gonna come up in some problem, I don't know what it is, you know, medical imaging or optical control.

It doesn't really matter. And you're gonna want to know is it convex or not. And in very rare cases, you will be forced to do this by hand. So by hand means you whip out your theta between 0 and 1, two arbitrary points, and try to argue this inequality.

That's gonna occur occasionally. If the function is twice differentiable, the equivalent of doing it sort of by hand – and by the way, every now and then you do have to do things by hand. So to – you will actually have to work out the hessian, and it might be easy, it might be not easy to establish that the hessian is positive-semi-definite.

But in fact, the way you really want to do things is reserve these two for when there is no other option. What you really want to use is the calculus of convex functions to establish convexity. And for that matter, to decompose a function in front of you into a sequence of operations with the leaves of the tree sort of being adams, and adam is a function known to be convex or concave or something like that.

And here you want to know about the operations that preserve convexity, and we look at sort of the really obvious simple ones, you know, nonnegative weighted sum, that's kind of obvious. Composition if an affine function, that's also very easy to show. Pointwise maximum supremum, we looked at examples too, and by the time you get to sort of supremum, you can actually start showing convexity of some things that at first glance look pretty complicated.

I mean, even with maximum, for example, to show that the sum of the five largest absolute values of entries of a vector is convex. It's not – first of all, it's not an obvious fact, it's not that easy to show. It's like one line if you use a pointwise maximum argument.

Then we look at composition. That's where we are now and I'll finish that off now. And I pointed out that in fact, these are – this is not a minimal list. And in fact, the minimal list, as far as I know, of the things covered here, this is basically like two. There are two operations that are used because composition actually subsumes nonnegative weighted sum and pointwise maximum supremum.

All right, so let's look at composition. The composition, the most basic case, it's the only one I actually remember, is the scalar case, so here you compose a function with another, and the question is, when does that preserve convexity?

And the answer is, and this is the only one I remember, is a convex increasing function of a convex function is convex. And here I should – I do have warn you, you should – I mean you're gonna have to know that there is a subtlety. This is not a subtlety interesting for its mathematical interest or anything like that. This is a subtlety that comes up in practice, and will mess you up if you're not aware of it.

And the subtlety is this; the – when you say an increasing convex function, the increasing actually refers to the extended value function. Okay? For example, you have to distinguish between for example, X^2 – now X^2 is clearly not – X^2 on ours, obviously not increasing.

However, X^2 on \mathbb{R}_+ , if you restricted to \mathbb{R}_+ you would say, well that's increasing. And in fact, any normal person would say it. And it is indeed increasing, however; remember that if you simply – if you define it to be undefined for $-X$ then the extended valued version actually looks like this.

And it actually assigns a value minus infinity. At this point, this function is by no means increasing, I mean, not remotely. It starts out at $+\infty$, and then dives down to zero. So that's non-increasing. Now there is another version of this, and that's this; you take the square, like this, and then you assign it to be zero below for X^- , okay?

This thing in increase – this thing is actually increasing on – the extended valued version is increasing, this function. So these are things unfortunately you do have to know, that do come up, you can get deeply confused, and it can actually have real consequences if you don't – if you're not sensitive to this.

There's some irritating things here, let me just show you one. Here's one; let's suppose we want to look at something like the norm of $AX - B$ to the 1.62. Don't ask me why, but we do. And we want to argue that that's a convex function of X , okay?

Well you can work from the inside out and you can say, look, that's an affine function. That's a convex function of an affine function, and therefore norm $AX - B$ is convex. Now what we want to do is say something like well this is a convex function to the 1.62 power. And I'd like it to say a convex function of the 1.62 power is convex. That's kind – it's sort of true, but sort of like not true, or something like – you have to be very careful now.

All right, now I can't just define – let's get my notation consistent, H . I can't just say $H(Z) = Z$ to the 1.62 for $Z+$. I mean, I can say that, but the problem is this function here – I mean, this is by the way, slang or – not quite slang, it's an informal way to say that the domain is non-negative Z . That's what this is.

You cannot apply the composition theorem, absolutely cannot. So the correct way to do it is to say this, there you go, is to define the function that looks like that. So that's the – and I know this is a little bit silly, but this is very important, and it will mess you up if you don't – if you're not sensitized to these things.

This is the function which is 0, if you pass a negative argument to it, and it's the number to the – it's the argument to the 1.62 power if you pass a positive argument to it. That function is actually increasing and convex on the extent – its extension. Its extension is. And now I can apply this to this and then I can make another argue – I can say, oh hey, look at that.

What I'm passing to the function H is actually always nonnegative, it's a norm. But if you + -- if you add – therefore, this part, this case, or this switch in the definition of H never occurs. And so indeed, we find that this here is convex. So I don't know if that makes sense, but that's it.

So I didn't mention that last time, but I thought I'd just quickly mention it now. So let's look at composition; the sophisticated version is the vector composition virgin – version here. So if you have vector composition, it means the argument G is actually multiple functions.

So let's say it's G_1 through G_k , so I wanna – I'm gonna take a vector function, it takes many arguments of a bunch of functions, and again, the base composition theorem, or the canonical one that I remember is a convex increasing function of a convex function is convex.

And by increasing I just mean non-decreasing, so it's just informal. So a convex increasing function of a con – of convex functions, I should say, is convex, so that looks like this. And the way – now when you prove these, making no – I mean its three lines, right. You make no assumptions whatsoever about differentiability. But the way to sorta understand what happens is to write out the formulas.

The same for the composition case, the scalar composition. So here we will take $N=1$, so the second derivative of F is = to – this is just a chain rule for the second derivative of this thing, it's the – its G prime, that's of course – that's DGD_X which is a vector times then the hessian evaluated G of X times G prime, so that's a quadratic form, and then plus this thing.

And that's nothing but sum. This is partial H , partial whatever, it's I component, evaluated at G of X and then times G prime prime I of X like that, okay? And now you can see for exam – we can check things out.

If G is convex this first term is nonnegative no matter what, and then the second term to be convex you need – sorry, that if H is convex, you need the G – if all the G 's are convex, it means this vector is an element wise nonnegative vector that interproducted with – I don't know if that's a verb, but it is now. You interproduct that with the gradient of H .

If H is increasing that's a nonnegative vector, and so this thing is positive too. But now you can see that there's actually other ways to constructing. So for example, a more sophisticated version would go something like this; let's see if I can get it right. A convex function of a bunch of others is convex provided – okay, now let me see if I get right, I'll start at this thing, provided for each argument either the argument itself is convex and H is increasing in that argument, or the function – the argument is concave and H is decreasing.

Did that – I think that came out right. By the way, there are people who know these things and memorize them – not memorize them, but they have very good mnemonic or something like that. I'm not one of them, I just redo this. Whenever I see it – whenever I have to deal with it I go somewhere quiet, or find a little piece of a whiteboard or paper and just write out formulas like this or something.

So this is vector composition. And we looked at some examples the last time. One more, and this would pretty much finish off the basic operations that preserve convexity of functions, and this is minimization.

Now by the way, we've already seen that max – there's a sense in which maximization – there's not a sense, sorry, maximization preserved convexity. And we've already seen that. That's the maximum pointwise supremum. Notice that in the case of maximization, there is no requirement of any kind on how the variable you're maximizing over enters into the function. Everyone know what I'm talking about here?

The point is that if I have a function of two variables, F of XY and I want to look at something like this, let's say, and define that – I'll try to get my notation right, as G of X , this thing, the only requirement here for convexity of G is the following; is that for each Y this function is a convex function of X , period.

That's the on – and for that matter, this thing does not have to be convex, it doesn't even have to – this doesn't even have to be like an \mathbb{R}^n or even be a vector space, it could be an arbitrary set. So supremum over a completely arbitrary family – so if you maximize, well assuming you can – when you maximize there's no condition – I mean, nothing matters as to how Y enters into this function.

Now minimize also works, and that's a bit a weird because generally what you're gonna find out with all these things are they weird asymmetries. In other words, it's – that you can do one thing but not the other. The cool part – well why we have this class is it turns out often the thing you can do is the thing you often want to do.

That's – otherwise this would be interesting, but not useful. What's interesting here is that there – you actually – it's very rare that you can sort of both maximize and minimize something like that because they're sort of like – they have opposite flavors. But in fact you can, but it's a much stronger condition, here's the condition; you can minimize – some people call this partial minimization, but the way, because full minimization would be to minimize F of XY over X and Y .

You can minimize over a variable provided – and the result will be convex provided – this is very important – F is jointly convex in X and Y and the set C over which you minimize is convex. Okay? And then this is easy to show and all this sorta stuff.

Now let's look at some examples, here's a very basic one and it's one you may have seen before anyway. Let's take a general quadratic form and block it into two – block the variable into two groups, X and Y . So this is a – that's a general quadratic form. It's the quadratic form associated with this matrix $A \ B \ B^T \ C$ block.

All right, well we know that F is jointly convex in X and Y provided this full block two by two matrix is positive-semi-definite.

By the way, when is this function in X for any Y , what's the condition, on $A \ B \ C \ D$?

Student: A is positive [inaudible].

Instructor (Stephen Boyd): Precisely. It's only that A is positive-semi-definite. There's no condition whatever on B and C . So if I were to take this function here and maximize F of XY over Y , the result would be convex provided only that A is positive-semi-definite. Yes?

Student: [Inaudible] on what condition on A could you then state that that function was not convex?

Instructor (Stephen Boyd): On what condition on A – oh, which function? The –

Student: That one.

Instructor (Stephen Boyd): F , okay, in this case because it's a quadratic form, it's necessary and sufficient. So if this matrix here has one negative Eigen value or more, it's not convex period. And not only that, we can easily produce points – I can make you a line where instead of going like that it will go like that. Actually, you tell me the line? You tell me the line?

Student: It needs to go like that?

Instructor (Stephen Boyd): Yes. You have a quadratic form with one negative Eigen value.

Student:[Inaudible]

Instructor (Stephen Boyd): Yeah. Take the Eigen vector, you sail that along that line and it'll go like that, it'll go down. And that pretty much ends the discussion on convexity for that function. Now if I minimize this function over Y here – and what we'll do is – this is a technical condition. This actually works with C positive-semi-definite.

In fact, I'll tell you what happens when C is only positive-semi-definite. When C is positive-definite, you minimize over Y , yeah, you can do this analytically, because then you take the gradient and you set it equal to 0 and all that.

And you end up with this; you know, you find the optimal Y , you back substitute that into this form, and what you're gonna get is this thing. X transposed [inaudible] $A - B C$ inverse B transpose. Now this is the sure compliment, so – which you maybe have seen before, if not, I assure you on homework 3, is that right, yes, homework 3 you'll see it, because it's just a good thing to know about.

So that's called a sure compliment of – maybe of C in the big block matrix, so that's the sure compliment. And that's convex, and you know what that tells you, that tells you this matrix has to be positive-semi-definite. So you may have seen this somewhere in some other context, maybe estimation or – I don't where, but it would come up. It would come up in mechanics and come up I all sorts of things, this sure compliment here. Yes?

Student:[Inaudible] the graph is convex?

Instructor (Stephen Boyd): Joint convexity means that if you plot sort of X and Y , the whole – then its bowl shaped the graph.

Student: Okay.

Instructor (Stephen Boyd): Convexity in X means that if you fix Y and sort of look at a slice, it would go like this, it would go up, and it could be quite different, right, the two. Here's one; the distance to a convex set is gonna be – the distance to a convex set is convex. By the way, that's a rather complicated function. I mean, it's not simple.

If I draw some convex set like this, lets just get a rough idea of – I mean, the set is everything in here, let's work out what the distance to it is? What's the distance in here? 0, now I can draw the level curves of the distance, so for example, I might dra – I might end up – I won't do this right, but you know, you get the idea. It would look like – something like that, right?

These would be all of the points with a certain distance from here, and then as you move out, actually of course these get smoothed out, right, so if you have sharp corners in a set then the distance level curves get smoother. Like that, so you'd get things that would look like this, let's see if I can do it right, that's like actually circular. Is that it? Yeah. So I think that's it.

This is to give you a rough idea. So this function, though complicated, is convex. And the argument would go something like this; you'd say, well let's just take a look at this thing. I look at norm $X - Y$. Here I take norm $X - Y$, I have to make a – this is a very important argument, that's jointly convex in both X and Y .

That's the critical part that allows me to do this, okay. So that's gonna be convex.

Student:[Inaudible]

Instructor (Stephen Boyd):Oh, here, yeah, let me give you the full proof, okay? I mean, I could rip out – I could get out theta's and stuff and do it by hand, but it don't want to, show I'll show you the snappy proof. Snappy proof goes like this; I have two variables, X and Y . What kind of function is $X - Y$ of X , Y ?

Student:[Inaudible]

Instructor (Stephen Boyd):Yeah, it's affine, sure, but it's linear, which is stronger than affine, but affine is good enough, you're absolutely right. Therefore this is a convex function, the norm of an affine function of X and Y . That's better than getting out – whipping out the theta, wouldn't you say?

Student:[Inaudible]

Instructor (Stephen Boyd):What's that?

Student:[Inaudible]

Instructor (Stephen Boyd):It's actually linear, but affine is good enough to establish convexity. We'll look at the last example of – this one is sort of on the borderline between sort of basic methods and more advanced ones, although it actually comes up in more applications than you think, that's the problem with these things, is you look at something, think it's esoteric, and then by the time you look at enough problems you realize it actually does come up.

So that's the so-called perspective of a function. So the perspective of a function is a transformation. It does this; it takes a function from \mathbb{R}^N to \mathbb{R} and it's gonna produce a function which has one more argument, so it's gonna be a function from \mathbb{R}^N plus 1, so if you want to explicitly make it a pair, \mathbb{R}^N cross \mathbb{R} to \mathbb{R} , and it's this, it's G of X and T is T F of X over T .

So it does a weird scaling, it scales X first and then it multiplies by T , okay, so that's the – and that's a point – that's a function jointly of X and T . So obviously for any fixed positive T , this is obviously a convex function, because, you know, X over T is linear, therefore, affine. F is convex and then you multiply by a positive constant, that's clearly the case.

Maybe it's less obvious that when you fix X , $T^T F$ of some vector divided by a constant is convex, maybe that's less obvious, but – and showing this is not hard – by the way, the – as you might guess, the perspective of a function is related to the perspective of a set, so – and there is some connection between the epigraph of – this is sort of – you remember that one of the – probably the only maybe not obvious transformation of sets that preserves convexity was the perspective transformation.

So not surprisingly, the perspective of function is related through epigraphs to that, but I won't go into the details, we'll just look at some examples. Here, $X^T X$, that's the sum of X squared. Well that's obviously convex, but that says that $T^T X$ over $T^T X^T X$ over T , which when the T 's cancel out is $X^T X$ over T is convex, okay?

So there you go, your quadratic over linear function is convex and we got it by a quick perspective operation on just the norm. So, we went from – we took something that was completely obviously convex, and then it was something where this is less obvious.

Okay, so you start with negative log rhythm, and then you form $T^T X - T^T F$ of X over T and you end up with $T \log T$ minus $T \log X$ and you actually get that this is convex on X and T . Some of things you know, like if you fix X for example, this is linear in T so that hardly matters and you get the negative entropy over here.

It's not so obvious that this is jointly convex in both X and T , and by the way from this you can derive the various things like negative entropy, cobac libeler divergent and all that kind of stuff. Another example would be something like this; if F is convex then G which is an affine – this has to be positive – affine positive function multiplied by F of a general linear fractional transformation, but with the denominator required to be positive, that's going to be convex, so – and that's on the – that's where this is in the domain of that – or well, sorry, when $C^T X + D$ is positive and this image point is in the domain of that, so that's gonna be convex.

By the way, I recognize fully that none of this is motivated and all that kind of stuff, so don't think I don't. We'll look at a couple more concepts; one is the conjugant function which is going to get heavy use later – later in the class, so – but actually not for a couple of weeks, but then it will get very heavy use.

So this is just gonna be your first glimpse at it [inaudible]. Oh, I should add that all the things we're doing are complete – these are completely standard things, these are not weird esoteric things, they're – like the conjugant, everyone knows what it is. It's got slightly different names, some people call it the finshell conjugant and there's also some other names for it, depending on how fancy you want to sound when you use it, but these are just completely standard concepts here.

I'll let you know when we deviate into something that's not – so take any function, convex or not and you define the conjugant function, that's F^* of Y , and that's the

supremum of Y transpose X minus F of X over X . And so let's see if we can figure out what that looks like graphically.

It looks like this; you form the graph of F here, and then you form this function, YX – well we'll just do this in one variable. So, I form a function where the slope is Y , so that's Y , the slope here. And then it says I wanna maximize the deviation between these two functions.

Well, one side of deviation – anyway, I want to see how much smaller does F get than this line? And here you would see that that point occurs like right here. And that's gonna occur, by the way, when this is – if it's smooth, when this thing is tangent. Either way, it's gonna curve when it's tangent to this graph here, like that. And then this point will give you the conjugant function here.

So this gives you a very good way of working out the conjugant function. For example, if I decrease Y slightly, what will F star of Y do, or to be more convenient, let me ask you, what will minus F star of Y do for this function? This just requires you to visualize. If I decrease Y slightly, what happens here? Well I changed the slope, okay?

And then I go down here and I'll make a point of tangency, it'll shift down a little bit, right, or to the left – well down and to the left. And that – this point of intersection will creep up a little bit, everybody agree?

So in fact, minus F star will increase slightly, so F star would decrease. Now by the way, what happens as this slope changes? There's a critical moment here, and the critical moment occurs right there, everybody see that?

So at this very critical moment something interesting happens. I draw a line here and that would sort of give me that value of the conjugant there, negative value. Now as Y decreases further, what happens? This gets interesting. What happens is the point of contact leaves here, shifts over here like this, and this thing now starts going down. Everybody see this?

This is your first chance when we do duality, you'll get tons and tons of chances to do visual exercises like this, but this is your first exposure to this idea. So this is the conjugant – now here's an interesting about a conjugant function; it is convex – the F star of Y is convex in Y , completely independent of whether X is convex or not, I'm sorry, F is convex or not, totally independent.

And it's really stupid; look at this, what kind of function is what's uncovered there above my hand and to the left – right of my finger as a function of Y ? It's affine – that's affine in Y for each X . Hey, that's a supremum over affine functions. That's convex, period.

So this function is convex, okay? Actually I can tell you something fun – I'll go ahead and let you – no, I'll mention it now, its fun. Now you know usually – you've seen

conjugate's right? You've seen conjugate's for complex numbers, you've probably seen conjugate's in some other – I mean as a bunch of places work.

But, when someone uses in a mathematical context the word conjugate, and there's some sort of cultural background there. A conjugate is usually something which when applied twice does what?

I mean we're just talking roughly, right? If you do a conjugate twice, what do you get?

Student: The original.

Instructor (Stephen Boyd): Part of the original thing, or you know, something – and if you've seen this in a lot of different – other context, it might be something related to it like the closure or something like that, that would not be surprising, the conjugate.

So you might guess that the conjugate of a conjugate of a function would be an interesting thing, and you might even guess it was the function, but I'll tell you what it is and it's super cool. This is just for fun. I'll tell you what it is; there's a function, and the conjugate of the conjugate is not the function, it's something called the convex envelope of the function.

And it is the – let's see if I can say it right – it's the largest convex function that fits under this one. I think I said that right. So in this case here's what you do. In fact, I can describe them to you as epigraphs very easily, ready? It's this; we have to have a name for the convex envelope, so F^* is going to be the convex envelope – is going to be the function which is the convex envelope of F , and I think I can write it – I hope this is gonna work, there we go, it did work.

I don't know if that's clear. This says that you take the epigraph of a function and you form the convex hull of it and that's the epigraph of the envelope. So the – let's sketch here the convex envelope of this function. Well it's gonna look like this, it's the function itself out here.

It turns the corner and then it goes like this – ignore that, my little – and then it hugs the function again. Make sense? So that is what just if you're curious, this is. It'll come up, you'll see, later. So that's what it is. By the way, for a convex function, roughly speaking F^* star start is F .

Roughly speaking, means that there are some technical conditions and things like that. Okay. Let's look at some examples of the conjugate, and this is actually a calculation you're going to have to do a bunch of times. It's going to come up in duality and things like that, so you're gonna have to do conjugate calculations.

So I'll just do one and then I quit, and then you'll do some, I promise, and then that'll be enough. Then after that you'll do it on demand when you need to.

So the negative log rhythm is convex, let's calculate it's conjugate. Of course I can calculate the conjugate of the log rhythm if I want, I can [inaudible] of anything I like, but we're interest mostly in conjugates of convex functions.

So the – lets take a look at this function, so this looks like this, I guess that's the negative log, something like that. That's the negative log, and I want to work out the sup or maximum over X of XY plus $\log X$. Now $\log X$ of course is unbounded above. So obviously here if Y is 0 or positive, this thing is unbounded above. So the supremum is plus infinity, okay? So that much is clear.

Now let's assume then that Y is negative. This is interesting, because now it's a log that's growing, but then you subtract a linear thing and in fact a linear thing will win out on a log always, and they will – to find out exactly where they – it reaches it [inaudible], you just take the derivative and it's a – these are smooth things.

So you take the derivative of XY [inaudible] X and you get Y plus 1 over X is equal to 0 . That is the point X that would maximize XY plus $\log X$, provided Y is negative, okay? So this says that X is one over minus Y , okay? And that's a – and then you plug this X back into XY plus $\log X$ and you get minus 1 , that's the XY part, plus $\log 1$ over minus Y , okay? And that's this thing, that's the conjugate, and that's for Y .

And it's important when you work out a conjugate to get the domain correctly here, so okay. If you have a strictly convex quadratic, so that's one half X transpose QX – by the way, the one half is here because – the one half is there often because it simplifies some formulas or makes them look prettier.

Other cases where you're not like basically calculating gradients, you leave the one-half off, and – it's sorta you choice. Actually, this is one of those forea transform like things where you can stick the one over two pi anywhere you want, and of course everybody does it differently which is pain in the ass, so. So, it's one of those stories, but if you put the one half here you're gonna get something pretty.

So if you work out the supremum of Y trans – this is a quadratic function [inaudible] equal to 0 , you get X and all that, and you get a beautiful thing that the quadratic, the conjugate of a quadratic function associated with quadratic – with positive definite matrix Q is – is the quadratic for with positive definite matrix Q inverse.

So that's what that comes up in. Conjugate by the way has all sorts of interpretations, but I put them off until – like in economics and things like that. I put them off until we get to duality, and we get to cover everything on this.

Quasiconvex functions so I think we have only one more of these and then we're done – you done with your first dose of convex functions. Oh, I should let you know there's like a whole industry, in fact almost like a little subculture of – it's been going on for like 40 years, where people worked out little variations on convexities, so there's quasiconvexity, there's sudoconvexity, and I forget, that goes on and on and they – anyway, and – it

keeps otherwise dangerous people off the streets so it's – it provides a positive – the net benefit to society I believe is positive.

And it's fine or harmless people, they're nice enough, but just to warn you, you would find tons and tons – and they get of course all bent out of shape and they make little diagrams with – this show what implies what, and they go, did you know that a strictly quasiconvex, blah, blah, blah, is actually pseudonormal convex? Anyway, and they'll make – draw diagrams.

And if you don't adequately say, wow, that's really interesting, and then you'll be in trouble. So that's the best thing to do, that's my recommendation when you are confronted with someone. By the way, someone who pesters you and says did you quasiconvex or sudoquasinormal strictly convex.

I don't even know what they do; you can look on Google to find out all the different things. We have looked at all of these things, the 27 different extensions, and identified two that seemed above threshold. So I just want to let you – this is one of them, quasiconvexity.

So quasiconvex is this; its s function whose sublevel sets are convex. Now of course any convex function satisfies that, but that's not enough and here's an example which is obviously not convex, but every sublevel set is, so what happens is here, to get a sublevel set you just put a threshold. You know, I really should put it above like that, because you want to look below, I don't know.

You need a fishing line to put across here. And you pull it across here, and you can see for example, the α sublevel set is an interval the β sublevel set – I mean assuming that this continued indefinitely, is actually an half infinite interval, but it's convex, so this is – this is fine.

And you call something quasiconcave if it's – if the super level sets are all convex. So let's just look at some examples to see how this works. Squared the absolute value of X , I guess that's a function that looks like that; that's quasiconvex on \mathbb{R} , and it's obviously not convex.

It's quasiconvex because you – you just simply put a line anywhere and you get an interval. Here's an interesting one, the ceiling of – of a number. So the ceiling of a number is the – it's the smallest integer that's bigger than the number – bigger than or equal to the number, that's the ceiling.

Now this is pretty weird, because that's integer valued, okay? Now integer valued functions, for example – well let's have a discussion about integer valued convex functions. Who would like to start the discussion? What can you say?

Let's discuss the theory of integer valued convex functions. Do they exist?

Student:[Inaudible]

Instructor (Stephen Boyd):Constant, okay. For example here is a function, 3 convex, let's keep going, anything else?

Student:[Inaudible]

Instructor (Stephen Boyd):Four, thank you. It's going to be a long discussion at this rate. Okay. What do you think? I'm not asking – I don't want to prove – we're talking eyeballs here.

Student:[Inaudible]

Instructor (Stephen Boyd):Yeah, okay, so it can't – so the only – these are the only ones, right, the only integer valued – you couldn't have two values, like you couldn't have 3 at one place and 4 at another, okay. So that was a very – you are now in total command of the extensive theory of integer valued convex functions. Basically there hardly are any other than – they're known as constants, actually.

So that's why this is a bit interesting, that the ceiling is gonna – which is obviously integer valued, and it's going to be quasiconvex. And actually, what's kinda cool about it is we will be able to optimize things involving these. You will, I absolutely promise, there'll be problems involving like FIR filters, again, if you're in EE, and we'll ask questions like, what's the minimum order required to have one decibel ripple over this [inaudible] minimum 40 decibels rejection over here.

Now when you just hear that question the convex part of your brain is not lighting up, because you just hear – you heard order of a filter, that's an integer. Being familiar with the theory of integer valued convex functions and knowing that it's a very short theory, you might say, well there's no way you can optimize something involving an integer, well you can't, so we'll get to this.

That's quasi [inaudible] and it's easy enough to check. These are just stupid things, every ceiling, the sublevel set of a ceiling function is just an interval, and so it's very easy. Log is quasilinear – product $X_1 X_2$, that's a quadratic form that is neither convex nor concave. It's the quadratic form associated with that matrix.

That matrix, just take a look at it, it's not positive definite and it's not negative definite, it's got split Eigen values, so it's neither. It's certainly neither.

Student:[Inaudible]

Instructor (Stephen Boyd):Quasilinear it means quasiconvex and quasiconcave. And by the way, these functions do not have to be linear, not even remotely. A linear fractional function is quasilinear, meaning it both quasiconvex and – this is the most famous by the way, because this function here is absolutely not convex.

If you have a linear fractional function it looks something like this – well there might be a place where it goes into minus infinity or something like that, but it's gonna look like this – like that. I drew the wrong one; it doesn't have to look like that. All right, I take that back, it can be, depending on A , B , C and D , sorry. But in general it doesn't matter.

A linear fractional function is gonna be quasiconvex and quasiconcave. There are famous problems that have this form. Here's an interesting one; distance ratio, so the ratio of your distance to two points, on the set of points where you're closer to one than the other. This is a half space by the way, which I think – it's a half space. So as long as you're closer to one thing than another, then your ratio of your distance to that then the other is actually gonna be quasiconvex.

But the way, all you need to do is take this function and put a gamma – take it to the gamma, like if you're in wireless and you have here the SIR from a good transmitter and a bad one. And it says, -- well you have the inverse SIR, and it says that the inverse SIR is convex, quasiconvex, provided you're closer to the transmitter you want to hear than to the interferer.

Sorry, I have to throw these in, because I'm bored. Sorry, go on?

Student: Is it the same of convexity preserving operations that work on these quasiconvex [inaudible]?

Instructor (Stephen Boyd): No, that's what cool, it's not. It's a more general set. For example, a monotone function of a quasiconvex function is quasiconvex, because quasiconvex just basically says, you know, all that matters is your sublevel sets. So if I put some monotone transformation it works just as well. So it's a different set.

Oh, and it's neither a restriction nor a generalization of the others. For example, the sum of two quasiconvex functions is generally not quasiconvex. And let's just look at a quick example to show that; here's a quasiconvex function, here's another one, and the sum, for sure, is not.

Oh, there's another name for quasiconvex, it's called unimodal, that's another name. What reminded me was I added these two together in my head and that's what we would call not unimodal.

Okay, so let's look at one that's not nontrivial, just for fun, it's the internal rate of return. So here I have a cash flow sequence, so X_0 up to X_N and then X_I is a payment in period I , and positive means to us.

So, we'll assume that X_0 is negative, that means that we pay out at period $T=0$ and we'll assume that the sum of all of these things is positive, so I guess that's what we'd call a good investment. The sum of – if these are positive it means it's a payment we receive, and this is the sum total of everything and it's balanced out by our investments.

So minus X_0 is the investment. Then if I work out the present value of this cash flow stream, with an interest rate R then that's nothing more than this, I sum from $I=0$ to N X_I with $1+R$ to the minus I .

So that's my present value assuming an interest rate of R . Now the internal rate of return is defined – usually completely, incorrectly, and ambiguously in – in fact essentially every text I look on in finance. It actually said it's the interest rate that makes the present value 0. Well let's see, I would say somewhere in middle school you would learn that – for example, a function like that could actually have a 0 at – there are many values of R for which it can be 0, but this apparently hasn't stopped people from defining it as V .

So the correct one is it's the smallest interest rate which makes the present value 0, so that's the definition of the internal rate of return. For example, if your rate of return is 8% means that your cash flow stream has a present – or let's say the income derived from the cash flow stream is exactly the same as had you invested it at an interest rate of 6% or 8% or whatever I said. So that's the idea.

Now this is a rather complicated function of X , so in fact even – I guess not of a particular one, but if you had a whole bunch of different cash flows and things like that, it's quite complicated to visualize what – and the way you visualize what the internal rate of return of a cash flow stream is – well you really have to just form this sum.

That's a function of R ; note it's a function of powers of 1 over 1 plus R , so it's actually a polynomial in 1 over 1 plus R . Then you'd find the roots of the polynomial. These would be the values of the interest rate at which that investment is neutral and then you'd find the smallest of those.

This is complicated; it's like an Eigen value or something. This is not simple. It's quasiconcave. By the way, this is your first hint that something cool is happening, I'm going to mention it; because this is the first time we ever hit it.

We're going to find out things like quasiconcave functions, we're going to be able to maximize. You'll be able to maximize this, five lines of code with complicated constraints, so you'll be able to maximize this. We can minimize convex functions and quasiconvex functions, we can maximize quasiconcave ones.

Now it turns out the – it's hard to minimize internal rate of return, that's a very difficult problem. What this says is it's actually easy at least algorithmically and mathematically to maximize internal rate of return.

Now if you think about it carefully in practice, we're not neutral. Problems of maximizing and minimizing internal rate of return do not occur in equal – with equal frequencies, right? Actually pretty much you never want to minimize internal rate. You want to maximize internal rate of return, and you'll wow, look at that, the thing we want to do turns out to be the thing we can do.

I just thought I'd mention this, because it comes out – I mean it's kinda stupid, you don't mention this in like 263, you say, isn't that amazing, we can minimize AX minus B . Every now and then we want to maximize it. Just though I'd mention it.

So what is – how do you show this? To show something as quasiconcave you look at the super level sets. By the way, what you're doing is you're saying please let's study the set of cash flow streams that have an internal rate of return exceeding 8%. That's what you – that's a rather complicated set.

That's like the set of symmetric matrices whose Eigen values are less than – whose maximize Eigen is less than 7. I mean it's not impossible to imagine what it is, but it's quite complicated. What is it? To say that your internal rate of return is bigger than R , that means that for all interest rates up to R your internal rate of return – I'm sorry, your present value is actually positive.

So, that's this. That's an infinite set of inequalities, one for each interest rate up to capital R . I'll let you check this, but it's a quick way to see this. By the way, if you tried to do this in cases of a sequence of 2 or 3 using the formula for the roots of a polynomial or something, this would not be fun.

Now for quasiconvex functions there's a modified Jensen's inequality, and let me just draw it, because it's better to draw it than to – than – I don't know what's here, so. Jensen's inequality says something like this; you have two points like this, and let me just draw a convex function first.

And it basically says that if you draw this line, this cord lies above the graph. That's what Jensen's inequality says. It's says actually you do a better job – for example F at the midpoint is smaller than the midpoint in the average of F . That's basically what it says, that's convex.

Quasiconvex is cool. Let me draw – there's a nice quasiconvex function, and I'll pick some points. I'll pick this one and I'll pick that one. Oh, no – well the cord lies above it – it doesn't matter. I mean, in this case the cord would lie above it. But I'll draw the picture of what you have to have for a quasiconvex function.

Instead of the core, what you do is you draw this thing at the maximum of the two values. So it says the function has to lie below there, which it does. Another way to say it is if you have a quasiconvex function and you have one X and you have another Y , and you for a linear convex combination of X and Y , and you evaluate F , it's not worse than the worst of X and Y at the end points. That's a way to say it, if worse means large in F . Does that make sense?

So this is the picture, and the way you write that is this way, that along – when you form a blend and evaluate F , you're not worse than the – you couldn't be worse than F at the end points. That's the picture.

And you have some things like this, if F is differentiable, if it's quasiconvex – this occurs if there's a characterization here, and it basically says this, that if the – it says that the sublevel set of F lies on one side of the gradient picture. So if you take the gradient like this and form a half space, it says everything over here has a higher or equal value of F , that's what it says.

Or another way to say it is every – and for a quasiconvex function, every point that's better lies on this side of this thing, period.

Okay, that's another way to say it. By the way that's a hint that we can minimize it, because when you evaluate a gradient and it gives you half space and it says, I don't know where the minimum is, but it's got to be in this half space.

You can think of that vaguely as query – by evaluating a gradient you got – this is very vague and weird, but it basically says you got one bit of information, because it basically said it's in that half space, not that one.

We'll get to that later, but that's why this would work. And finally, in our whirlwind tour, the second of the two extensions and variations that were above threshold. This one is very simple, but it's important enough.

It's not like it's an extension or anything, it just happens a lot. And it comes up – it's log concavity. This is weird, because it turns out – to say something's log concave, just says $\log F$ is concave, that's all it says. And it's a weird thing, it says that if you evaluate the function at a point on the line say between X and Y , it's bigger than or equal to the weighted geometric mean.

It's the same as if you just take logs here, you get this. Now for some reason applications come up overwhelming log concave, not log convex. I don't know why. It doesn't matter, I'm just telling you. Log concave things come up all the time. Log convex things like never come or almost never come up.

It could be that in all the applications where log – where this thing comes up – which is statistics, it's everywhere in statistics. We'll see it's in lots of things. But everywhere where it comes up their conventions are that they want to maximize things, like you want to maximize likelihood ratios and maximize posterior probabilities and things like that.

So here's some examples; powers are log convex for positive exponents and log concave for negative ones. And this is very, very important, many, lots and lots of probability densities are log concave, okay? That's very important. And that's called a log concave probability distribution.

An obvious example would be a Gaussian, because if you take the log of this thing, and let's look as a function of X , you get a negative quadratic.

Another one would be like an exponential variable or a Laplasian variable with a two sided exponential distribution. But actually tons and tons of distributions are. So if you're in statistics or something like that, you know all – various families, all the ones you can probably think of right off the bat would be log concave.

Things like beta distribute – pretty much everything. Pi square, you know all of these things are gonna be log concave. The log concavity is going to be the basis of a whole bunch of bounds in statistics, so if you're in statistics this comes up and it may even – it might even have been mentioned explicitly, although I doubt it.

Actually who here – I know there's people in statistics here, they're not gonna self identify though. I'll find out. I can find out who's in – anyway. Has log concavity ever been mentioned in the statistics – yes, cool, what class?

Student:[Inaudible]

Instructor (Stephen Boyd):In seminars, oh, okay. Not in like – how about 310 – no, it wouldn't be in 310, it would be in the statistics stuff.

Student:In 300 it's mentioned in passing.

Instructor (Stephen Boyd):Cool, good, okay, there we go. Here's one, totally non-obvious, has amazing applications like this immediately. Is the cumulative distribution of a Galician. So it's got different names, I think urpsey, I think in ee and communications they decided it wasn't good enough [inaudible] and the called it the Q-function or something, is that correct? Something like – isn't that what that is?

Student:[Inaudible]

Instructor (Stephen Boyd):One minus – it's what?

Student:[Inaudible]

Instructor (Stephen Boyd):One minus five, okay great, that's good for my colleagues to introduce a new notation for something that's been used for hundreds of years by many other people. Sorry, let's move on.

So the log of accumulative distribution, and I suppose the log of a q-function is long concave, okay. And you know, this is not so obvious. The cumulative distribution looks like this of a Galician, it goes like that and then goes up like that. It looks like that. If you take – it's obviously not concave.

You take the log, and the part that's 1 – well that's 0 obviously [inaudible] and this part curves down like that. And it's asymptotic in fact to something like a square here, which is the asymptotic of the Galician, the tail of a – small values here are well approximated by Galician itself.

Student:[Inaudible]

Instructor (Stephen Boyd):No, that's false. I'll tell you very shortly which are.

Student:[Inaudible]

Instructor (Stephen Boyd):All which?

Student:[Inaudible]

Instructor (Stephen Boyd):In one [Inaudible] that's true, because they're monetel. So they're definitely quasiconvex. We'll get to this. So let's look at some properties of these things. Well, just by working out the hessian of the log of a function is, it's something like that [inaudible] or something.

You get this, it says basic – it's actually quite interesting, we can look at it, it says that F of X – if you're talking about a log concave or a log convex function, obviously the function itself has to be positive, or non-negative at least. It can be 0 by the way, which in log will map to minus infinity which is perfectly okay, because for a concave function minus infinity is essentially our OOD token.

I gave it another name the other day that was out of domain. What did I call it before, NID. Anyway, one of those tokens – it's the token for being out of the domain, so minus infinity. So this is non-negative, and this is actually quite interesting. I guess you could put the F of X over here or something like that.

This says to be concave is to say that the hessian is less than or equal to 0. This says actually the hessian can be less than or equal to. That's a rank 1 – that's a dyad. So this actually says, you're actually allowed one positive Eigen value in the hessian of a log concave function.

That's cool. One positive Eigen value you're allowed. So it means you have one direction along with your curvature can actually be positive. We needed that because look at this, okay. But it's got to be overwhelmed by – it's gotta be no more than something having to do with the gradient here.

I mean I don't want to give too many interpretations of this, but. Here are some obvious things; product of log concave functions is log concave, totally obvious, because the product of two things, you take the log it's the sum. So that's the sum of convex functions.

Sum of log concave functions is not always log concave, that's easy. Add to make a mixture of Galician, moved apart you get tow bumps. Oh, by the way, if it's log concave its quasiconcave, because log is monotone. And you can't have two bumps on a quasiconcave function, because then a super level set will be two isolated intervals so it won't be convex, the set.

Now we get to something which is highly, highly, highly non-obvious, it's this; if I have a function of two variables and it is log concave and I integrate over one of the variables, the result is log concave, it's true.

For literally years I have looked for a simple proof of this, and simple means like simple, assume everything you want, X and Y can be one variable, F can be differentiable as many times as you like, so I'm talking about a simple proof that would please our careful mathematician friends, I'm talking for just a normal person where you just write out the first and second derivatives, look at some stuff and say – and you end up with this and say, there it is.

I'm just talking about something, a bunch of calculations that would establish it. I have failed, always. By the way, people – I've said this challenge out to a bunch of people, and several time people have come up with – by the way, I have a four page one which is very complicated and anyways.

So if anyone every finds one out – figures out a simple one for normal people that just uses calculus or something like that and a couple of arguments here and there, it'd be great. After which I will say that it was obvious, but at the moment it's obvious but unknown to us, this simple proof.

It's certainly not known to me, that's for sure. It's not obvious. The implications though are stunning, and I will show you some of them. Convulsion preserves long concavity. That gives you a lot of stuff right there. So if you have – by the way, convulsion it means if you have two random variables each with log concave density and you add them, the distribution of the sum obviously is gonna be log concave, so that's the kind of thing you would get just right off the bat.

So convulsion preserves log concavity and here's a fascinating app – is this, it says if you have a convex set and you have a random variable with log concave PDF, then the probability – this is a function of X , that X plus Y is [inaudible] is log concave.

And the way to show this is I just apply this one integral. Your integral thing is – that's the big hammer, you just reduce everything to the one as of now, non-obvious fact that you can integrate over one set of variables a log concave function and the result is log concave.

So roughly speaking, integration preserves log concavity. And you apply that – you'd make a function here, G of X plus Y , which is just the indicator function. It's 1 if you're in the set and 0 if you're outside and work out this integral and it's that. And this is very cool, this immediately has amazing applications which I will show you one right now.

So here's the way it works; I have some set of values like this in some multi-dimensional space like that, and these are – you should think of these as acceptable values of what happens. So these – each axis could be – I don't – it doesn't really matter. This could be

the – I'm setting the process recipe or the target process recipe or whatever for making IC's.

One of these could be threshold voltage which if it's too high the chip is too slow. If it's too low it dissipates too much power. By the way, how many people are in circuit design? Now come on. Does anyone care about it? Maybe they just sleep late. They were in the clean room until very late last night and – I know they're in this class.

But if no one is here then I will remove these examples from the – all right. So this is the acceptable [inaudible] and what's gonna happen is you're gonna set a target – you're gonna set the target value like that and you'll adjust all your manufacturing to hit that target value.

By the way, when you do manufacturing runs and you look what comes out on Monday, Tuesday, Wednesday, Thursday, and things are biased away from the point – if the mean of that is away from that point you need to get new people running the manufacturing, because you just – you would offset it.

So we can assume that when you hit the target point the distribution of values [inaudible] will come off in manufacturing will have that mean. I mean if it doesn't have that mean, then you shift it or whatever.

So you have a mean, and let's say that the manufacturings [inaudible] are given by this random variable W . It could be some horrible thing, some joint distribution, but we assume the following, its log concave PDF.

And Galician would work just instantly, but that's it. What's gonna happen now is this. It means that when you hit a target here, you're gonna get values around like that or whatever, every now and then you'll be out. If you're out then that's something that doesn't perform. That's something that's no good and that's a waste.

And your yield is the probability that you actually end up in the set. This all makes perfect sense. Now, imagine taking the target and moving it around. As you move it around your yield changes, obviously, right.

So for example, if you put the target here, what would the yield be? I'm not endorsing that as a choice by the way, but what would the yield be here? Would it be 0? No, probably not, because every now and then accidentally you would manufacture an acceptable part.

I mean if the distribution doesn't have fine eyed support, so every now and then accidentally. And this would be a poor choice, and it's totally obvious what a good choice is, you want something deep in the set, and that's got all sorts of names.

Its called design centering, is one of the names for it, and yield optimization, it's got all sorts of other stuff. Although, yield optimization means something different in other

contexts, so. Everybody got this? And now you should also be able to visualize also the yield as a function of the target point.

So it's high here, low here, less low here, and high is somewhere –at some point in the middle here. So, here's the point. That function is log concave. Just very roughly speaking, let me tell you what that means. It means we can optimize it. We can maximize it. So you put these two together, and what it says is you can actually do – you can actually directly maximize the yield.

Can you make a note? We'll make a homework problem on that. Not for homework 2, later, don't worry. We'll do a yield maximization problem. Okay, so that it. One last topic on convex functions is convexity with respect to generalized inequalities. Generalized inequalities allow you to overload ordinary inequality.

All you need to define convexity of functions is you need an id – you need the ability to talk about – well in the very general case you need to talk about the line segment between two points. We work in vector spaces, so for us that's fine. And then you need this inequality to be able to say that like $F(\theta X + (1 - \theta) Y)$ is less or equal to $\theta F(X) + (1 - \theta) F(Y)$ [inaudible].

And so you need that inequality. These can be vector valued and then you can talk about convexity with respect to a generalized inequality. The only case where this would really come up and is actually anything is with matrix convexity. And so there it just says the inequality holds, but the inequality has to be interpreted as in terms of matrices.

So – now I think we'll quit here. This sort of finishes our whirlwind tour through I guess the [inaudible] of convex sets and functions. Next time half the lecture will be a little bit boring, because it's just nomenclature of optimization, but by next week actually I think things will get interesting, so we'll quit here.

[End of Audio]

Duration: 77 minutes