Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-1

Lecture - 5: Convex sets and Convex functions

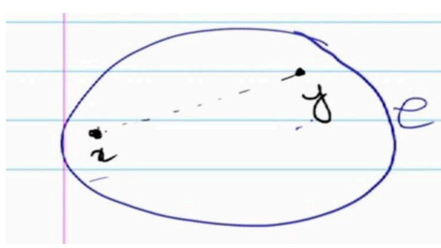
So, let us get back to studying convex functions and convex sets. So, let us say you have two points x and y in Rn ok. A convex combination of these points is defined to be of the form theta x plus one minus theta y. So, this is the convex combination of these points x

$$\theta x + (1-\theta) y \quad \forall \theta \in [0,1]$$

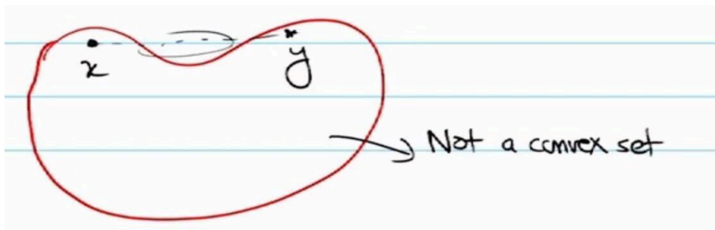
and y. So, you choose a theta a number between 0 and 1 including 0 and 1 right and you get another point z which is theta x plus 1 minus theta y and this z is said to be a convex combination of x and y.

So, what is the difference between convex combination, linear combination and affine combination So, what is what is linear combination of two points? So, this is convex combination. A linear combination would be some theta1 x plus theta2 y. I mean the theta 1 and theta 2 need not be related right? So, that becomes a linear combination and an affine combination you also allow theta to be beyond 0 and 1. I mean you have in that would be an affine combination in some sense that is usually called, but really I mean in this case we are interested in convex combination.

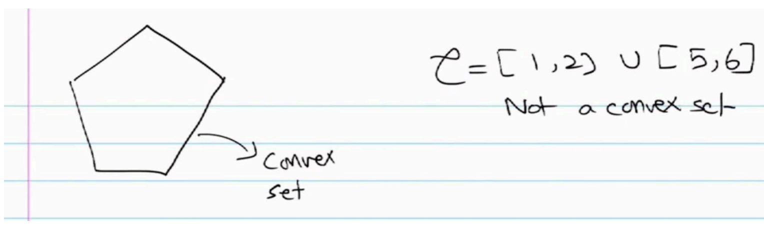
So, convex set is a set where a convex combination like if I end up finding a point which is a convex combination of two points. So, let us say I consider a set C and I choose two points x and y in C such that their convex combination also lies in C, then C is a convex



set ok. So, an example could So, a set of this form for instance if I choose any two points in the set let us say x and any other point y. So, convex combination first of all would lies on this line segment joining x and y right. So, does every point on this line segment lies in this in this set x in this set C in this example? right. So, and this is true not just for this and x and y, but you can choose any other x and y and this would be true right? So, this is an example of a convex set. So, a non-example would be something like this right. And now if I choose a point x somewhere over here and y somewhere over here and I draw the line segment joining x and y. I can get few points and more than few points such that it does not lie in C right.



So, this is a non-example. So, this is not a convex set. How about a set like this is this a convex set? Is this a convex set? Yes, right. So, this is a convex set. How about this set? So, C is defined to be this.

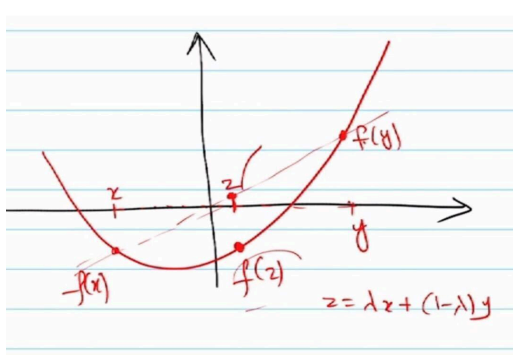


Is this a convex set? No, right. Why? So, if I choose a point like let us say I choose points 2 and 5 and I choose theta to be half. So, then 3.5 for instance does not lie in the set right. So, this is not a convex set ok.

Any questions on convex sets all right? So what are convex functions? So what is our definition of a convex function mathematical definition? So a function is said to be convex if f of lambda x plus one minus lambda y is less than or equal to lambda f of x plus one minus lambda f of y for all lambda in zero to one. So, what does this say? So,

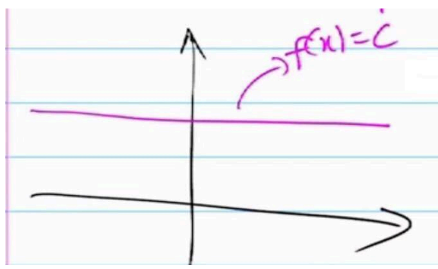
$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)$$
$$\forall \lambda \in [0,1] \text{ and } \forall x,y \in \mathrm{dom}(f)$$

suppose I take two points x and y which are in the domain of f ok. And I look at their convex combination of these two points lambda x and 1 minus lambda y. So, the function evaluated at the convex combination of these two points is basically less than or equal to the convex combination of the function values at those points.

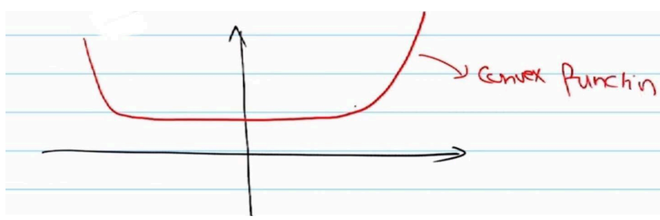


So, what does it really mean? So, let me let me draw it. Now, suppose I choose my x this is my x and let us say this this is my y. So, this point would be f of y. and this point is f of x. Now, let me like if I consider this convex combination of points x and y, and let us say I choose somewhere over here that is my z right. So, this z turns out to be lambda x plus one minus lambda y. Now, the function evaluated at z is over here this is f of z. Now the same convex combination if I look at, if I draw a line, so the function is evaluated at z, so this is f of z and this is the convex combination of f of x and f of y, so this exceeds the value f of z. right, and this is this is the definition of a convex function.

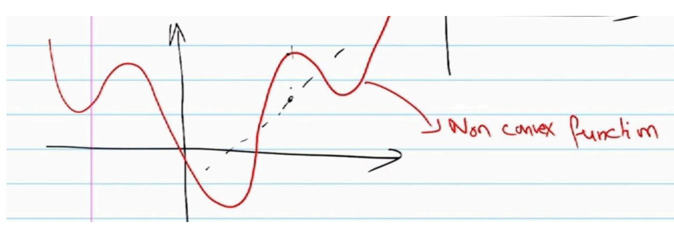
So, when you I mean this need not be always the case that I mean. So, for instance, if I consider I mean a good a good way to sort of view a convex function is I mean it should look like an upward facing parabola as long as you have that kind of function I mean that is a I mean that is a picture that you should keep in mind when you think of when I say convex function a picture that you should keep in mind is it should be an upward facing kind of parabola. That is not always true. So, for instance, if I look at a constant function right which is f of x is constant I mean this is trivially satisfied with equality I mean. So, this satisfies the definition.



So, the constant function is also a convex function ok. So, this is f of x equal to c this is also a convex function So, a constant function is always a convex function. A function of



this form for instance which looks something like this is also convex even though it is constant in this region I mean it acts like an upward-facing kind of parabola. So, this is also a convex function, but at the same time if I look at a function that behaves something like this. So, this is an example of a non-convex function why So, suppose I now if consider two points and draw a line.



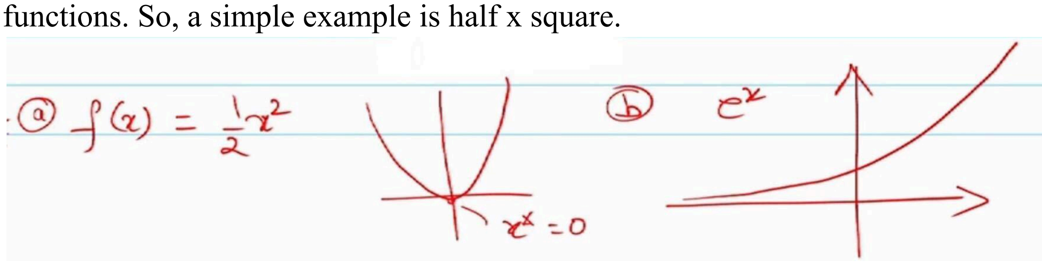
So, the function at any convex combination in this case. So, the function evaluates to a larger value then the convex combination of any two points right here. So, so this is this becomes an example of a non-convex function ok. So, and you can clearly see this function does not look like an upward-facing parabola it is specifically in this region it does not look like an in fact it is a downward-facing kind of parabola.

So, why do we care about convex functions and convex sets? So, convex function every I mean. So, first of all, I mean in the context of convex function why did we really need convex for this to be well defined right? So, we said what is the definition of a convex set if x and y lie in the set C then the convex combination should also belong to that set right. So, you want the function to be well-defined on a convex for a combination of two points x and y right. So, it makes sense to define these functions convex functions over a convex set.

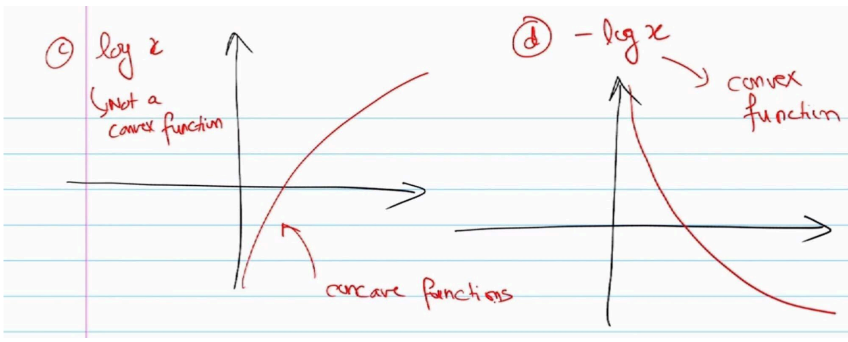
The other reason why is we study convex functions is suppose I find a. So, for a convex function suppose I end up finding a minima to it right. Let us say this is the local minima. This local minima also happens to be a global minima. There is no other minima which is better than this local minima right or smaller than this value.

Similarly over here if I end up finding a minima over here because this function is constant. So, everywhere it attains the same value, but then there is no other point at which it attains a value which is smaller than this. So, every local minima is a global minima. So, for a convex function every local minima is also a global minima. But can we say this about this non-convex function? No right. If I end up finding a local minima, I can always find a better local minima here right? So, not every local minima is a global minima. So, for non-convex function that is not true, but for convex functions you may have multiple local minima, but all of them are going to be global minima as well. Is this clear? And that's why it makes sense to find work with convex functions because the moment you find one of the minima you know that you have arrived at the globally optimal value and you cannot improve the function any or make it any smaller. So it's I mean you can terminate the algorithm pretty much. Whereas for a non-convex function if you arrive at a local minima is no way for you to know whether you want to terminate the algorithm and keep or keep looking for better solutions.

Even if you arrive at a solution right maybe a better solution may exist somewhere like far off from a particular local minima. So, when working with non-convex functions giving global guarantees is very difficult. You can only in most cases you can only talk about locally optimal solutions ok. Is this clear? So, let us give some example of convex functions. So, a simple example is half x square.



So, this is the most commonly used example of a convex function and if you look at how this function looks like simply looks like this right with x star equal to 0 being the optimal solution. this function is minimized at x equal to 0. How about e to the x? Is this a convex function e to the x?. So, the graph for e to the x looks something like this. So, again it looks like an almost looks like an upward-facing kind of parabola. So, this is also an example of a convex function ok. what about log x? Is log x a convex function? So, what is the graph for log x look like? Right. So, does this look like an upward-facing parabola? No, right. So, log x is not a convex function.



So, not a convex function. But at the same time if I look at minus log x, so minus log x would have a graph and this looks like an upward-facing parabola and this is a convex function. So, functions so, these type of functions which rather look like a downward-facing parabola instead of upward-facing parabola such functions are called concave functions. And we talk about minimizing convex function, we talk about maximizing concave functions right because these are sort of why because if f is a concave function. implies minus of f is a convex function, right? So, log x is concave, but minus log x is convex.

So, if it is a concave function then minus f should be a convex function. So, the question is every non-convex function concave and that is not the case right? So, if I invert particular function I mean. So, this is a non-convex function this is this is neither convex nor concave.you can say that this is locally convex here, locally concave here, locally convex here again locally concave locally convex, but not every. So, function is concave if it is if minus of f is like a is convex yeah. So, in real optimization problems if you if you arrive at these kind of objective functions. all you can say is that you have arrived at a locally optimal solution that's all you can say for general class of functions which look something like this when you are solving this in practice I mean you cannot guarantee that you have converged to a globally optimal solution in most cases you can provide local guarantees you can say that you have locally converged to one of the optimal solutions it may be the best it may not be the best maybe you will try a different initial initialization and try to converge to better local minima, but I mean you cannot provide guarantees beyond saying that we have converged to a locally optimal solution.

So, there are there can be certain functions which are both convex and concave right like constant functions I mean you can view them as concave you can view because if you maximize it it will give you the same value if you minimize it you will give you the same value right. I mean you can have like both ways that is fine. I mean the definition of concavity is basically related to convexity. So, this class of function is also non-concave this is not just non-convex it is also non-concave.

So, let us look at a few operations that preserve convexity. So, we are going to use capital gamma to denote the class of all convex functions. The first such operation that preserves convexity is non-negative weighted sum. So, that means, if I have f1 and f2 that are convex.

So, that means, they lie they belong to this class capital gamma and I consider weights w1 w2 greater than equal to 0, then w1 f1 plus w2 f2 that also belongs to gamma. So, non-negative weighted sum of convex functions is also a convex function. Is this clear? Can you think of an example? So, some common examples are when we look at L1 penalty right. So, norm of x one norm of x right by definition this is defined to be and mod x we know has is a convex function it looks like an upward-facing parabola kind of thing right? So, summation of so, this is this is a non-negative weighted sum of mod xi's.

① Non-negative weighted sum: $f_1, f_2 \in \Gamma$

$w_1, w_2 \geq 0$

$\Rightarrow w_1 f_1 + w_2 f_2 \in \Gamma$

① l_1-penalty

$\|x\|_1 = \sum_{i=1}^{n} |x_i|$

convex function

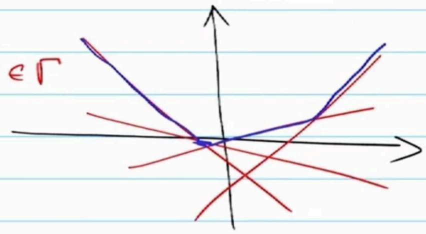
② $\left(\frac{1}{2}x^2 + \frac{1}{4}x^4\right)$

So, this is also a convex function So, when we look at pi's which are numbers between 0 and 1. So, if I consider those pi's, so these are non-negative weighted sum of negative log pi right and negative log x we know is a convex function. So, this is another example and so on like. So, if you have x square and x to the power 4 right. So, something like this. of x square plus one-fourth x to the four. Is this a convex function? Yes right because this is convex this is convex. In this case both have the same optimal solution as well right? And sometimes you will see and we will also look at some augmented Lagrangian methods later instead of working with these class of function which are relatively difficult to optimize you also add something to those functions which are relatively easier to optimize maybe also share the same optimal solution and that way this augmented sum or this this kind of function you can also I mean this you can try to optimize it much better than just trying to optimize the original function that is one particular operation that preserves convexity.

Another operation is point-wise max . So, let us say you have functions f1 and f2 that belong to these functions right ah that belongs to set of all convex function.So, then max of f1 and f2 would also be convex. So if f1 and f2 are convex functions then max of f1 and f2 are also it is also going to be a convex function. Good way to visualize this is let us say you have lots of lots of functions like these. So in this case we are just considering linear functions and so on. Now I am doing point-wise maximization. So I am just so at this point this function sort of is maximum. So i select this at this point then we select this particular patch then this patch and then this patch right and you can see that this is a convex function so pointwise max is something that that is a convex function and we are

② Pointwise max: $f_1, f_2 \in \Gamma$

$$\Rightarrow \max\{f_1, f_2\} \in \Gamma$$



going to look at the application of it when we talk about lagrangian dual so pointwise max is another operation that preserves convexity.

Another common class of transformation that preserves convexity is affine transformation. So, if f of x is convex, then f of a x plus b is also convex. ok. So, under affine transformation, convexity is preserved. So, an example would be we know that this is a convex function right.

$$\frac{1}{2}\|x\|^2 \Rightarrow \frac{1}{2}\|Ax-b\|^2$$

So, this implies this would also be a convex function and this is an example where we try to find a solution to a system of equation Ax minus Ax equal to b right. So, this is a convex problem yeah. Yeah I mean that is a good question I mean if you are defining the two functions and like for instance if you are defining point wise maps or non-negative weighted sum I mean at least they should have overlapping domain if. So, if not then I mean basically you would restrict it to the domain basically the intersection of the two domains. this is one example another example would be log barrier function right.

② Log-barrier function $-\log(b - a^T x)$ is convex

$$a^T x \le b$$

So, minus log of x is convex. So, this would also be convex right and this is this is often the barrier method that we use when we want to enforce a transpose x less than equal to b. So, this is the barrier function that barrier function approach that we use right. So, we add this barrier to the objective function. So, that any violation is heavily penalized right. Because log anything less than 0 is going to be negative infinity right. So, log is not even defined. So, log 0 is negative infinity. So, if you start in a feasible set this barrier basically prevents you from getting out of that feasible set. and this is this barrier method is used when we when we try and solve constrained optimization problems in an unconstrained manner yeah all of this will be dealt I mean I am just giving you examples I mean as of now we are just looking at convex function. So, all of the examples that I am giving you will be used later when we try and develop algorithms ok.

So, we now look at, so how do we identify, I mean like if I give you a function, how do we identify that a particular function is convex? I mean one way is to use the definition that f of lambda x plus 1 minus lambda y is less than equal to lambda f x plus 1 minus lambda f y. And, but in certain cases, it becomes very tedious to do that, right.

So, the question is can we look at better ways to identify whether a particular function is convex? So, let us assume that function is continuously differentiable, right. Assume f is continuously differentiable. so that that that is to say that gradient of f of x is defined right. So, the first order condition necessary and sufficient condition for convexity is f is convex if and only if domain of f is convex and f of y is greater than equal to f of x plus gradient f of x transpose times y minus x. Is the statement clear? So, f is a convex function if and only first of all I mean as I said right it only makes sense to define convex function over a convex set.

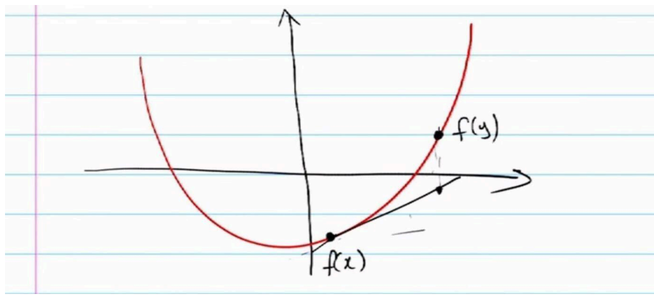
$$f \text{ is convex} \iff \text{dom } f \text{ is convex and}$$
$$f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

but because otherwise you I mean the convex combination of a two of two points may not even m and f may not even be defined there right. So, f is convex if and only if domain of f is convex and if you choose any two points x and y in the domain of f this inequality holds true. This is the first order this is called first order condition because of this first order I mean we use the first order derivative or the gradient of the function here. So, I will just prove this statement one way I think leave the other way as an exercise.

So, I will just prove this implication one way. So, since f is a convex function right. So, we assume that f is convex. So, f of x plus lambda times y minus x this is same as f of lambda y plus 1 minus lambda and by definition this is less than equal to lambda times f y plus 1 minus lambda times f of x right. This is just using the definition of convexity from the definition of convexity. So, now if I subtract f of x on both sides is this clear and now I divide this by lambda and consider the limit when because this is true for every lambda in 0 to 1.

$$\Rightarrow: \quad f(x+\lambda(y-x)) = f(\lambda y + (1-\lambda)x) \quad \forall \lambda \in [0,1)$$
$$\leq \lambda f(y) + (1-\lambda) f(x) \quad [\text{From def}^n \text{ of convexity}]$$
$$f(x+\lambda(y-x)) - f(x) \leq \lambda(f(y) - f(x))$$
$$\lim_{\lambda \to 0} \frac{f(x+\lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x)$$
$$\nabla f(x)^T (y-x) \leq f(y) - f(x)$$
$$\text{or} \quad f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

So, this is true for all lambda in 0 to 1 right. So, I divide this by lambda and consider the case when lambda goes to zero right. So, this is what is this left-hand side now? This is nothing, but the derivative of f in the direction y minus x right which is to say that this is I mean this quantity is nothing, but gradient of f transpose y minus x. So, derivative of f in that direction that is less than equal to f y minus f of x and therefore, you recover the right-hand side or Is this clear? So, I will leave the other side as an exercise maybe it can be part of your homework problem we will see right, but what does this geometrically sort of represent right. So, if you have a convex function which looks something like this I choose a point y.



So, this is your f(y) and this is your f of x. So, if I define the tangent basically derivative is nothing, but the tangent defined at this particular point right. So, f of x plus this t basically y minus x times this. So, the function always sort of I mean you may be end up over here and this function always lies above this value. that is the sort of geometrical sort of representation of this particular first-order condition ok. So, this is what we I mean this is what this particular condition really meets and we will eventually use this in lot more context and we because this is directly in terms of the gradient.So, when we design optimization algorithm we will use this particular inequality lot more in that context.

Just as we have first-order conditions we also have second-order condition. for convexity. So, we now assume f is twice continuously differentiable right because we want Hessian

$$f \text{ is convex} \iff \nabla^2 f \text{ is positive semidefinite} + \text{dom } f \text{ is convex}$$

to exist or second-order derivative to exist. So, we assume f is twice continuously differentiable f is convex implies the hessian is positive semi-definite. This is the second-order condition for convexity. So, obviously f is convex. So, we have to have domain of f is also convex and the Hessian of f is positive semi-definite or to say that in scalar sense I mean f double prime is greater than or equal to 0. So, when what is the Hessian of let us say for this if I choose f of x to be half x square what is f double prime x 1 right which is greater than 0. So, this condition we know that half x square is convex. So, if f is twice differentiable f is convex if and only if hessian of f is positive semi-definite.

So, everyone knows what positive semi-definiteness is right. So, that means x transpose this should be greater than equal to 0 for every x which is the definition of positive

$$x^T \nabla^2 f(x) x \geq 0 \quad \forall x$$

semi-definiteness ok. what about if I consider f to be one-fourth x to the power 4 what about this function is this convex ? what is the second-order derivative of this function 3 x square right. So, if double prime x is 3 x square which we know is greater than equal to 0 it is not exactly strictly greater than 0 because we know that x at x equal to 0 this would be 0 but this is greater than equal to 0 and this is convex. Thank you.