Computational Method of Optimization

Vector Composition- Lecture IV

Vector Composition

- Agenda
- Vector composition
- Conjugate function
- quasiconvex functions
- Log-concave and log-convex functions
- Convexity with respect to generalized inequalities

Source: Convex Optimization — Boyd & Vandenberghe

Vector composition

Composition of g: $\mathbf{R}^n \rightarrow \mathbf{R}^k$ and h: $\mathbf{R}^k \rightarrow \mathbf{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

f is convex if

g convex, h convex, $\tilde{h}$ nondecreasing

g concave, h convex, $\tilde{h}$ nonincreasing

proof (for n = 1, differentiable g, h)

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

examples

$\sum_{i=1}^{m} \log g_i(x)$ is concave if g_i are concave and positive

Log $\sum_{i=1}^{m} \exp g_i(x)$ is convex if g_i are convex

There are then vector compositions ones and I have a function of a multi-argument function of a bunch of other functions.
you have something like this; x is convex if all of the functions are convex, and h is convex and non-decreasing in each argument.
Tilde is the extended value extension.

Let us 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 z's.
$h(z)=1^Tz$ and $h(z)=\max_i z_i$
It is certainly convex in z. It's also increasing in each argument, do you agree with that, because it just sums the 1. 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.

Lets try h of z is the max of z. The max function is itself convex. It's also – it's increasing in each argument.
If you increase any element of a vector the max doesn't go down, so it's non-decreasing.

Minimization

if f(x,y) is convex in (x,y) and C is a convex set, then

$g(x) = inf_{y \in C}\, f(x,y)$ Is convex

examples

$f(x,y)=x^T Ax + 2x^T By + y^T Cy$ with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succcurlyeq 0, \qquad C \succ 0$$

minimizing over y gives $g(x) = \inf_y f(x,y) = x^T (A-BC^{-1} B^T)x$

g is convex, hence Schur complement $A - BC^{-1} B^T \succcurlyeq 0$

distance to a set: $\mathrm{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex

We have a function of two variables, f of x,y and we define that 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.
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.

Now let's look at some examples. Let's take a general quadratic form and block it into two – block the variable into two groups, x and y. So that's a general quadratic form.
It's the quadratic form associated with this matrix A, B, B transpose and C. 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 ? It's only that A is positive-semi-definite. There's no condition whatever on B and C.

Minimization

if f(x,y) is convex in (x,y) and C is a convex set, then

g(x) = $inf_{y \in C}$ f(x,y) Is convex

examples

$f(x,y)=x^T Ax + 2x^T By + y^T Cy$ with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succcurlyeq 0, \qquad C \succ 0$$

minimizing over y gives $g(x) = \inf_y f(x,y) = x^T (A-BC^{-1} B^T)x$

g is convex, hence Schur complement $A - BC^{-1} B^T \succcurlyeq 0$

distance to a set: $dist(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex

what happens when C is only positive-semi-definite. When C is positive-definite, you minimize over y, 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 find the optimal y, you back substitute that into this form. X transposed times A – B C inverse B transpose. Now this is the Schur compliment.
So that's called a Schur compliment of C in the big block matrix.
And that's convex, and that tells you this matrix has to be positive-semi-definite.

Here's one; the distance to a convex set is going to be the distance to a convex set is convex.

Minimization

if f(x,y) is convex in (x,y) and C is a convex set, then

g(x) = $inf_{y \in C}$ f(x,y) Is convex

examples

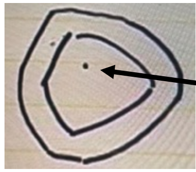
f(x,y)=$x^T Ax + 2x^T By + y^T Cy$ with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succcurlyeq 0, \qquad C \succ 0$$

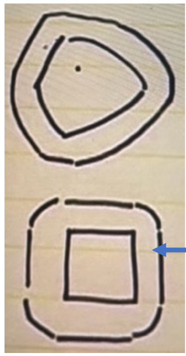
minimizing over y gives g(x) =$\inf_y f(x,y) = x^T (A-BC^{-1} B^T)x$

g is convex, hence Schur complement $A - BC^{-1} B^T \succcurlyeq 0$

distance to a set: dist(x, S) = $\inf_{y \in S} \|x - y\|$ is convex if S is convex



If we draw some convex set like this, lets just get a rough idea of –the set is everything in here, let's work out what the distance to it is? What's the distance in here? 0, now we can draw the level curves of the distance, so for example, we might draw we might end up – you get the idea. It would look something like that.



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., that's like actually circular. Is

We look at norm X – Y. Here we take norm X-Y, that's jointly convex in both X and Y.

Perspective

The perspective of a function f: $\mathbf{R}^n \rightarrow \mathbf{R}$ is the function g: $\mathbf{R}^n \times \mathbf{R} \rightarrow \mathbf{R}$,

$g(x,t) = tf(x/t)$, $\textbf{dom}\ g = \{(x,t) \mid x/t \in \textbf{dom}\ f,\ t > 0\}$

g is convex if f is convex

Examples:

- $f(x)=x^T x$ is convex; hence $g(x,t)=x^T x/t$ is convex for $t > 0$

- negative logarithm $f(x) = -\log x$ is convex;

hence relative entropy

$g(x,t)=t \log t - t \log x$ is convex on $\boldsymbol{R}^2_{++}$

- if f is convex, then

$$g(x) = (c^T x + d)\, f((Ax + b) / (c^T x + d))$$

Is convex on $\{x \mid c^T x+d>0,\ (Ax+b)/(c^T x+d) \in \textbf{dom}\ f\}$

That's the so-called perspective of a function. So, the perspective of a function is a transformation. it takes a function from R^n to R and it's going to produce a function which has one more argument, so it's going to be a function from R^n plus 1, so if you want to explicitly make it a pair, R^n cross R to R, and it's this, it's g of x and t is t f of x over t.
So it scales x first and then it multiplies by t, so that's a function jointly of x and t. So obviously for any fixed positive t, this is a convex function, because x over t is linear, therefore, affine.
F is convex and then you multiply by a positive constant, that's clearly the case.

Look at some examples. Here, x transpose x, that's the sum of x squared. That's obviously convex, but that says that t times x over t transpose x over t, which when the t's cancel out is x transpose x over t is convex. So, quadratic over linear function is convex and we got it by a quick perspective operation on just the norm

Perspective

The perspective of a function f: $\mathbf{R}^n \rightarrow \mathbf{R}$ is the function g: $\mathbf{R}^n \times \mathbf{R} \rightarrow \mathbf{R}$,

$g(x,t) = tf(x/t)$, **dom** g = $\{(x,t) \mid x/t \in$ **dom** f, $t > 0\}$

g is convex if f is convex

Examples:

- $f(x)=x^T x$ is convex; hence $g(x,t)=x^T x/t$ is convex for $t > 0$
- negative logarithm $f(x) = -\log x$ is convex;

hence relative entropy

$g(x,t)=t \log t - t \log x$ is convex on $\boldsymbol{R}^2_{++}$

- if f is convex, then

$$g(x) = (c^T x + d) f((Ax + b) / (c^T x + d))$$

Is convex on $\{x \mid c^T x+d>0, (Ax+b)/(c^T x+d) \in \textbf{dom } f\}$

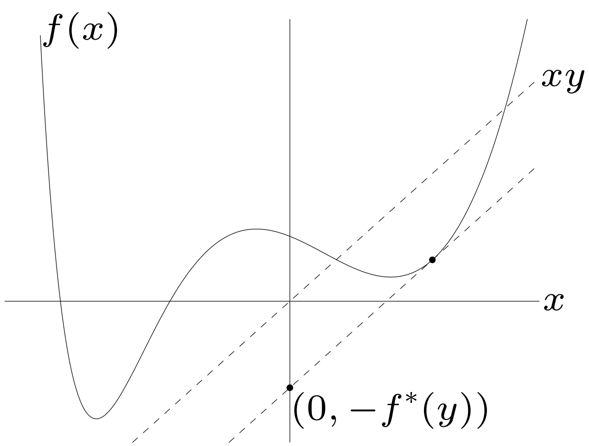
you start with negative log form t x – t f of x over t and you end up with t log t minus t log x and you get that this in convex on x and t. If you fix x for example, this is linear in t and you get the negative entropy over here.

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 that's where this is in the domain of that – or when c transpose x plus d is positive and this image point is in the domain of that, so that's going to a be convex.

The conjugate function

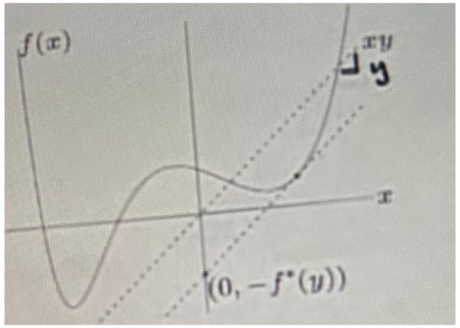
the **conjugate** of a function f is

$f^*(y) = sup_{x \in dom\ f}\ (y^Tx - f(x))$

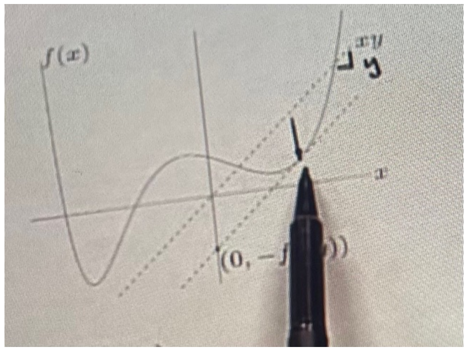


f^* is convex (even if f is not)

Take any function, convex or not and you define the conjugant function, that's star of Y, and that's the supremum of y transpose x minus f of x over x. And let's see if we can figure out what that looks like graphically.



you form the graph of f here, and then you form this function, yx –we'll just do this in one variable. So, we form a function where the slope is y, so that's y, the slope here. And then it maximize the deviation between these two functions.

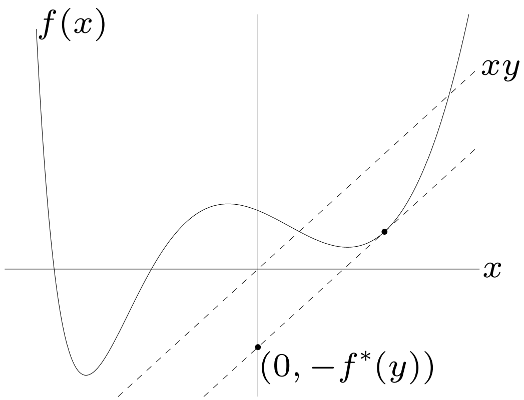


For example, if we decrease Y slightly, what will F start of Y do, or to be more convenient, what will minus f star of y do for this function? If we decrease y slightly, what happens here? We changed the slope? Then we go down and we'll make a point of tangency, it'll shift down a little bit, right, or to the left – well down and to the left. And this point of intersection will creep up a little bit,

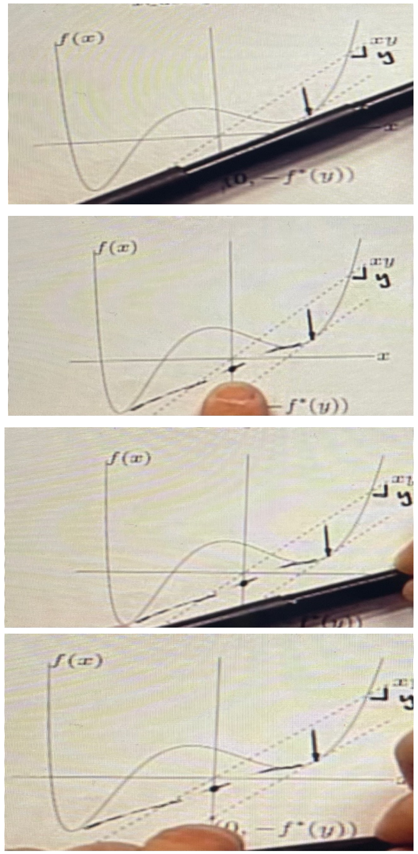
The conjugate function

the **conjugate** of a function f is

$f^*(y) = sup_{x \in dom\ f}\ (y^T x - f(x))$



f^* is convex (even if f is not)



in fact, minus F star will increase slightly, so F star would decrease. Now by the way, what happens as this slope changes? We draw a line here and that would sort of give us that value of conjugant there, negative value

Now as Y decreases further? What happens is the point of contact leaves here, shifts over here like this, and this thing now starts going down.

First exposure to this idea. So, this is the conjugant –it is convex – the f star of y is convex in y completely independent of whether x is convex or not, f is convex or not, totally independent. A function of Y is affine – that's affine in Y for each X. That's a supremum over affine functions. That's convex

The conjugate function

Examples

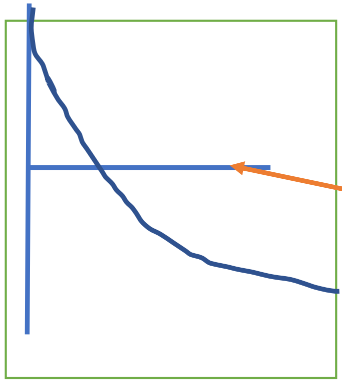
- negative logarithm f (x) = − log x

$$f^*(y) = sup_{x>0}\,(xy + \log x)$$

$$= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & otherwise \end{cases}$$

- strictly convex quadratic $f(x) = (\frac{1}{2})\, x^T Q\, x$
- with $Q \in \boldsymbol{S}^n_{++}$

$$f^*(y) = sup_x\,(y^T x - (\tfrac{1}{2})\, x^T Q\, x)$$
$$= \tfrac{1}{2}\, y^T Q^{-1}\, y$$



Negative logarithm is convex, let's calculate it's conjugate.
Let's 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 we want to work out the sup or maximum over x of xy plug 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. Now let's assume then that y is negative. Take the derivative of (xy + log x) that gives y+1/x which gives x=1/(-y). you plug this x back into xy+ log x. You get -1 for xy part plus log(1/-y), that's the conjugate, and that's for y.

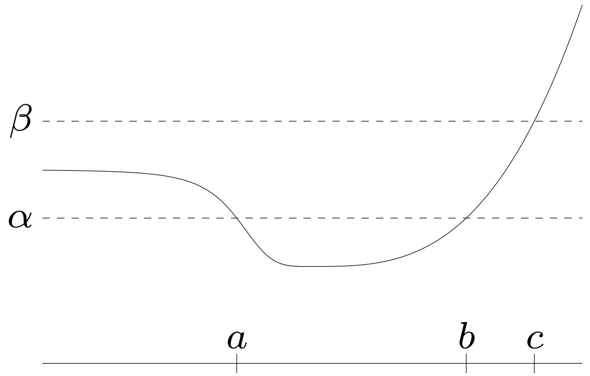
If you have a strictly convex quadratic, so that's one-half x transpose Qx – the one half is here because – the one half is there often because it simplifies some formulas. So if you work out the supremum of y transpose this is a quadratic function equal to 0, you get x and you get quadratic, the conjugate of a quadratic function associated with quadratic – with positive definite matrix Q is quadratic for with positive definitive matrix Q inverse.

Quasiconvex functions

f: $\mathbf{R}^n \rightarrow \mathbf{R}$ is quasiconvex if dom f is convex and the sublevel sets

$$S = \{x \in \mathbf{dom}\ f \mid f(x) \leq \alpha\}$$

are convex for all α



- f is quasi-concave if −f is quasi-convex
- f is quasilinear if it is quasi-convex and quasi-concave

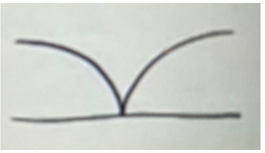
Quasiconvex is this; its s function whose sublevel sets are convex. Now of course any convex function satisfies that, but that's not enough and here's an example which is obviously not convex, but every sublevel set is to get a sublevel set you just put a threshold.

you can see for example, the **α**-sublevel set is an interval the **β**- sublevel set –assuming that this continued indefinitely, is actually an half infinite interval, but it's convex, so this is fine.

And you call something quasiconcave if the super level sets are all convex.

Examples

- $\sqrt{|\ x\ |}$ is quasiconvex on **R**
- ceil (x) =inf {z ∈ **Z** | z ≥ x} is quasilinear
- log x is quasilinear on $\mathbf{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$ is a quasiconcave on $\boldsymbol{R}^2_{++}$
- Linear fractional function
- $f(x)=\frac{a^T x+b}{c^T x+d}$, **dom** f ={x | $c^T x + d > 0$}
  is quasilinear
- Distance ratio
  $f(x)=\frac{\|x-a\|2}{\|x-b\|2}$, **dom** f ={x | $\|x-a\|2 < \|x-b\|_2$}
  is quasiconvex



look at some examples to see how this works. Squared the absolute value of x. that's a function that looks like that; that's quasiconvex on R, and it's obviously not convex.

Ceiling of a number is the – it's the smallest integer that's bigger than or equal to the number, that's the ceiling. Integer valued convex functions, Basically they're known as constants, like 3 or 4 etc. the ceiling is obviously integer valued, and it's going to be quasiconvex.

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.

$\begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}$ 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.

Examples

- $\sqrt{|\ x\ |}$ is quasiconvex on **R**
- ceil (x) =inf {z ∈ **Z** |z ≥ x} is quasilinear
- log x is quasilinear on $\mathbf{R}_{++}$
- $f(x_1, x_2) = x_1x_2$ is a quasiconcave on $\boldsymbol{R}^2_{++}$
- Linear fractional function
- $f(x)=\frac{a^Tx+b}{c^Tx+d}$, **dom** f ={x | $c^Tx + d > 0$}
  is quasilinear
- Distance ratio

  $$f(x)=\frac{\|x-a\|2}{\|x-b\|2}, \ \mathbf{dom}\ f = \{x \mid \|x-a\|2 < \|x-b\|_2\}$$

  is quasiconvex

Quasilinear it means quasiconvex and quasiconcave. And by the way, these functions do not have to be linear. A linear fractional function is quasilinear, meaning it both quasiconvex, because this function here is absolutely not convex.

If you have a liner fractional function it looks something like this – there might be a place where it goes into minus infinity or something like that. It can be, depending on a, b c and d.

A liner fractional function is going to be quasiconvex and quasiconcave. Here's a 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. So as long as you're closer to one thing than another, then your ratio of your distance to that then the other is going to be quasiconvex

Internal rate of return

- Cash flow $x=(x_0, \ldots, x_n)$; x is payment in period i
  (to us if x > 0)
- We assume $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$
- present value (PV) of cash flow x, for interest rate r:

  $$PV(x,r) = \sum_{i=0}^{n} (1+r)^{-i} x_i$$

- internal rate of return (IRR) is smallest interest rate for which PV(x, r) = 0

  $$IRR(x) = \inf \{r \geq 0 \mid PV(x,r) = 0\}$$

- IRR is quasiconcave: superlevel set is intersection of open halfspaces:

  $$PV(x,r) = \sum_{i=0}^{n} (1+r)^{-i} x_i > 0 \quad \text{for } 0 \leq r < R$$

Let's look at one that's not nontrivial, 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.

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 what we'd call a good investment. The sum of 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 we work out the present value (PV) of this cash flow stream, with an interest rate r then that's nothing more than this, we 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 (IRR) is defined for example, a function like that could 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 PV.

Internal rate of return

- Cash flow $x=(x_0, \ldots, x_n)$; x is payment in period i
  (to us if x > 0)
- We assume $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$
- present value (PV) of cash flow x, for interest rate r:

  $$PV(x,r) = \sum_{i=0}^{n} (1+r)^{-i} x_i$$

- internal rate of return (IRR) is smallest interest rate for which PV(x, r) = 0

  $$IRR(x) = \inf \{r \geq 0 \mid PV(x,r) = 0\}$$

- IRR is quasiconcave: superlevel set is intersection of open halfspaces:

  $$PV(x,r) = \sum_{i=0}^{n} (1+r)^{-i} x_i > 0 \quad \text{for } 0 \leq r < R$$

Now this is a rather complicated function of x, but if you had a whole bunch of different cash flows and 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.

We're going to find out things like quasiconcave functions, we're going to be able maximize.

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

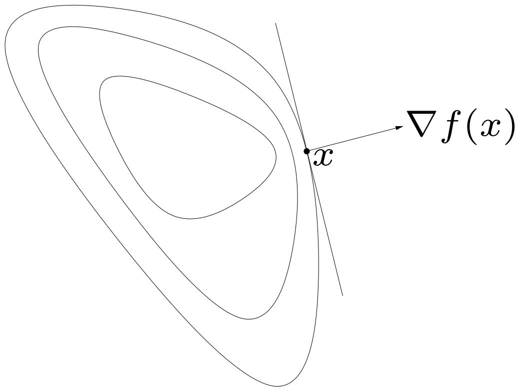
Properties

modified Jensen inequality: for quasiconvex f

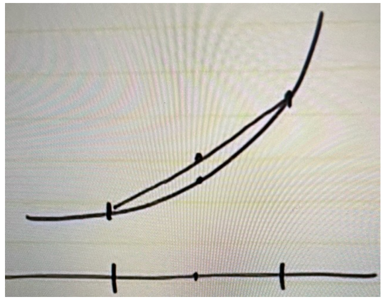
$$0 \le \theta \le 1 \Longrightarrow f(\theta x+(1-\theta)y) \le \max\{f(x),f(y)\}$$

first-order condition: differentiable f with cvx domain is quasiconvex iff

$$f(y) \le f(x) \Longrightarrow \nabla f(x)^T (y-x) \le 0$$



sums of quasiconvex functions are not necessarily quasiconvex



Now for quasiconvex functions there's a modified Jensen's inequality, and let us just draw it. 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 for example, F at the midpoint is smaller than the midpoint in the average of F. That's convex.

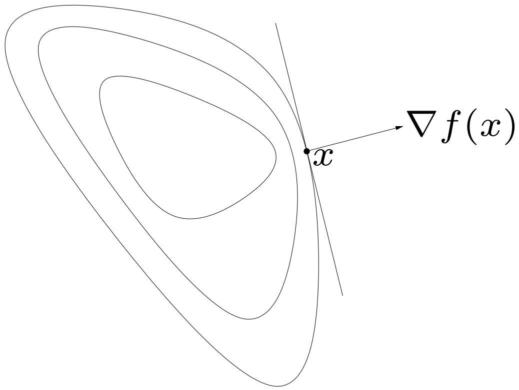
Properties

modified Jensen inequality: for quasiconvex f

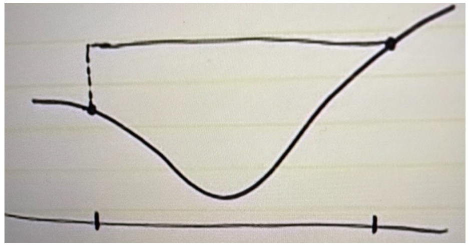
$$0 \le \theta \le 1 \Longrightarrow f(\theta x+(1-\theta)y) \le \max\{f(x),f(y)\}$$

first-order condition: differentiable f with cvx domain is quasiconvex iff

$$f(y) \le f(x) \Longrightarrow \nabla f(x)^T (y-x) \le 0$$



sums of quasiconvex functions are not necessarily quasiconvex



Quasiconvex. Let us draw – Quasiconvex function, and I'll pick some points. I'll pick this one and I'll pick that one. The cord lies above it. 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.

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 F at the end points. That's the picture.

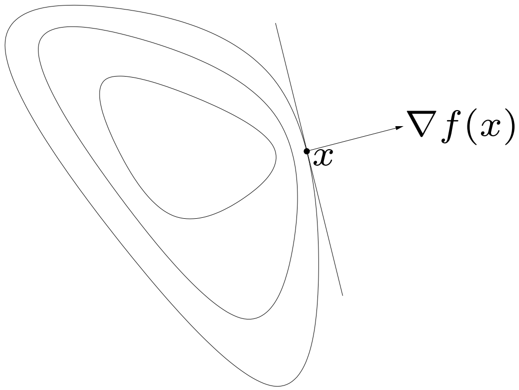
Properties

modified Jensen inequality: for quasiconvex f

$$0 \leq \theta \leq 1 \implies f(\theta x+(1-\theta)y) \leq \max\{f(x), f(y)\}$$

first-order condition: differentiable f with cvx domain is quasiconvex iff

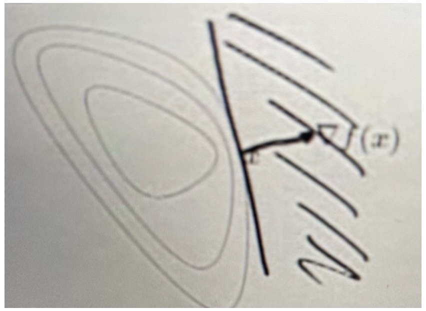
$$f(y) \leq f(x) \implies \nabla f(x)^T (y-x) \leq 0$$



sums of quasiconvex functions are not necessarily quasiconvex

If f is differentiable, if it's quasiconvex – this occurs, if there's a characterization here, and it basically 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.

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.

Log-concave and log-convex functions

a positive function f is log-concave if log f is concave:

$$f(\theta x + (1-\theta)y) \geq f(x)^{\theta} f(y)^{1-\theta} \quad \text{for } 0 \leq \theta \leq 1$$

f is log-convex if log f is convex

- powers: x^a on $\mathbf{R}_{++}$ is log-convex for $a \leq 0$, log-concave for $a \geq 0$
- many common probability densities are log-concave, e.g., normal:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1}(x-\bar{x})}$$

- cumulative Gaussian distribution function Φ is log-concave

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

The log concavity because it turn out –
to say something's log concave, just says log f is concave, that's all it says. 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 log concave, not log convex. Log concave things come up all the time. Log convex things like never come or almost never come up.

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.
That's very important. And that's called a log concave probability distribution.

Log-concave and log-convex functions

a positive function f is log-concave if log f is concave:

$$f(\theta x + (1-\theta)y) \geq f(x)^{\theta} f(y)^{1-\theta} \quad \text{for } 0 \leq \theta \leq 1$$

f is log-convex if log f is convex

- powers: x^a on $\mathbf{R}_{++}$ is log-convex for $a \leq 0$, log-concave for $a \geq 0$
- many common probability densities are log-concave, e.g., normal:

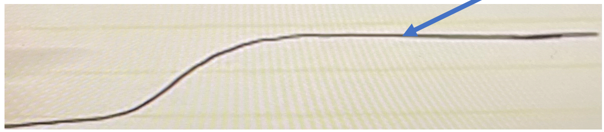
$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1}(x-\bar{x})}$$

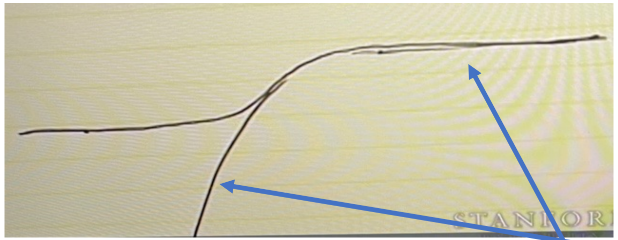
- cumulative Gaussian distribution function Φ is log-concave

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} \, du$$

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.

So the log of accumulative distribution, and I suppose the log of a q-function is long concave. The cumulative distribution looks like this of a Gaussian.It looks like that.





You take the log, and the part that's 1 – well that's 0 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 Gaussian, the tail of small values here are wel approximated by Gaussian itself.

Properties of log-concave functions

- twice differentiable f with convex domain is log-concave if and only if

$$f(x)\,\nabla^2 f(x) \preccurlyeq \nabla f(x)\,\nabla f(x)^T$$

$$\Rightarrow \qquad \nabla^2 f(x) \preccurlyeq \frac{\nabla f(x)\,\nabla f(x)^T}{f(x)}$$

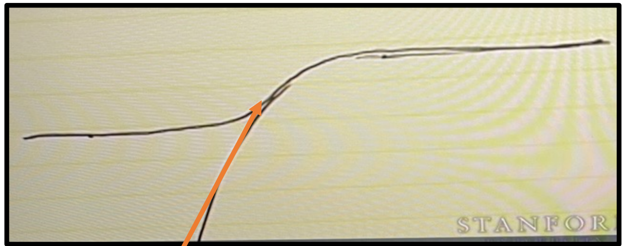
for all $x \in \text{dom } f$

- product of log-concave functions is log-concave
- Sum of log-concave functions is not always log-concave
- Integration: if $\mathbf{R}^n \times \mathbf{R}^m \rightarrow \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x, y)\, dy$$

- is log-concave (not easy to show)

Let's look at some properties of quasiconvex. Well, just by working out the hessian of the log of a function is. 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

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

Properties of log-concave functions

- twice differentiable f with convex domain is log-concave if and only if

$$f(x)\, \nabla^2 f(x) \preccurlyeq \nabla f(x)\, \nabla f(x)^T$$

$$\Rightarrow \qquad \nabla^2 f(x) \preccurlyeq \frac{\nabla f(x)\, \nabla f(x)^T}{f(x)}$$

for all $x \in \mathbf{dom}\, f$

- product of log-concave functions is log-concave
- Sum of log-concave functions is not always log-concave
- Integration: if $\mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x, y)\, dy$$

- is log-concave (not easy to show)

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 Gaussian, moved apart you get two bumps. 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 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.

Consequences of integration property

- Convolution f $*$ g of log-concave functions, f, g is log concave

$$(f * g)(x) = \int f(x-y)g(y)dy$$

- if $C \subseteq \mathbf{R}^n$ convex and y is a random variable with log-concave pdf then

$$f(x) = \mathbf{prob}(x + y \in C)$$

is log-concave

proof: write $f(x)$ as integral of product of log-concave functions

$$f(x) = \int g(x+y)p(y)\, dy, \qquad g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C, \end{cases}$$

p is pdf of y

Convolution preserves long concavity. So if you Have convolution it means if you have two random variables each with log concave density and you add them, the distribution of the sum obviously is going to be log concave.

So convolution preserves log concavity and 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 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 of you're in the set and 0 if you're outside and work out this integral and it's that.

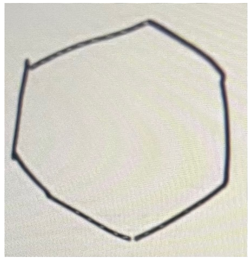
Example: yield function

$Y(x) = \textbf{prob}\,(x + w \in S)$

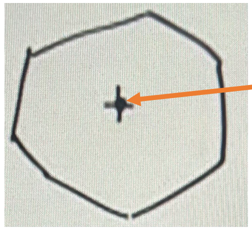
- $x \in \mathbf{R}^n$: nominal parameter values for product
- $w \in \mathbf{R}^n$: random variations of parameters in manufactured product
- S: set of acceptable values

If S is convex and w has a log-concave pdf, then

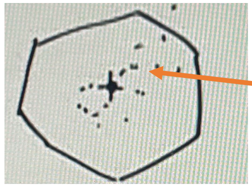
- Y is log-concave
- yield regions $\{x \mid Y(x) \geq \alpha\}$ are convex



If you have a mean, and let's say that the manufacturing products are given by this random variable w. It could be some joint distribution, but we assume the following, its log concave PDF.



And Gaussian would work just instantly. It means that when you hit a **target** here, you're going to 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.



And your yield is the probability that you 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.

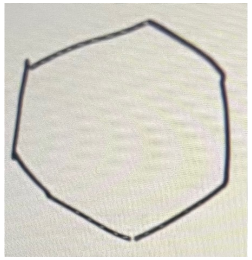
Example: yield function

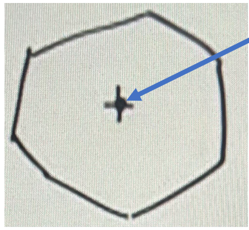
$$Y(x) = \textbf{prob}\,(x + w \in S)$$

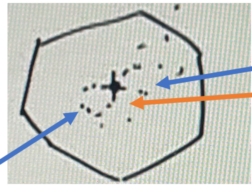
- $x \in \mathbf{R}^n$: nominal parameter values for product
- $w \in \mathbf{R}^n$: random variations of parameters in manufactured product
- S: set of acceptable values

If S is convex and w has a log-concave pdf, then

- Y is log-concave
- yield regions $\{x \mid Y(x) \geq \alpha\}$ are convex







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 accidently you would manufacture an acceptable part.

I mean if the distribution doesn't have fine support, 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.

It is called **design centering**, is one of the names for it, and yield optimization. Although, yield optimization means something different in other

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, you can directly maximize the yield.

Convexity with respect to generalized inequalities

$f : \mathbf{R}^n \to \mathbf{R}^m$ is K-convex if $\mathbf{dom}\, f$ is convex and

$$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta) f(y)$$

for x, $y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$

example $f : \mathbf{S}^m \to \mathbf{S}^m$, $f(X) = X^2$ is $\mathbf{S}^m_+$-convex

proof: for fixed $z \in \mathbf{R}^m$, $z^T X^2 z = \|Xz\|_2^2$ is convex in X, *i.e.*,

$$z^T(\theta X + (1-\theta)Y)^2 z \le \theta z^T X^2 z + (1-\theta) z^T Y^2 z$$

for $X, Y \in \mathbf{S}^m$, $0 \le \theta \le 1$

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 – in the very general case we 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 of theta x plus 1 minus theta y is less or equal to theta

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