Computational Method of Optimization

Convex Functions - Lecture III

Convex Functions

- Topics covered

- Basic properties and examples
- Operations that preserve convexity

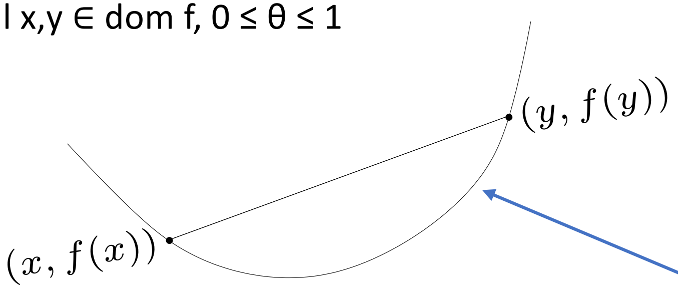
Source: Convex Optimization — Boyd & Vandenberghe

Definition

f: $\mathbf{R}^n \rightarrow \mathbf{R}$ is convex if **dom** f is a convex set and

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

for all x,y ∈ dom f, $0 \leq \theta \leq 1$



- f is concave if −f is convex
- f is strictly convex if dom f is convex and

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

for x, y ∈ dom f, x ≠ y, $0 < \theta < 1$

A function is convex if its domain is convex. That's the first requirement. And the second is that it satisfies this inequality for θ between 0 and 1. So this says that if you evaluate f at $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$ -- this point you saw in the convex sets lecture. This is a convex mixture, a convex combination of x and y. So geometrically it's a point on the line segment between x and y.

This says if you evaluate a function at a point on a line segment between x and y, the result is actually less than the same mixture of the values of the end points. Or in terms of the graph, it says that if you take two points on the graph of the function and then draw the straight line that connects them. this says the line lies above the graph.

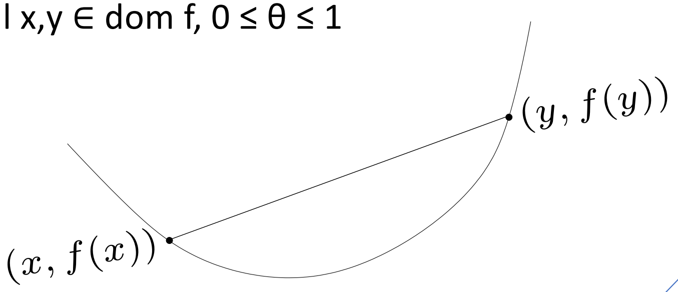
Another way to say it just very roughly is upward curvature. So it just means curves up, that's all. And by the way, of course, a convex function can look like that. It can be monotonic decreasing. Nevertheless, the curvature is upward for a function like that.

Definition

f: $\mathbf{R}^n \rightarrow \mathbf{R}$ is convex if **dom** f is a convex set and

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

for all x,y ∈ dom f, 0 ≤ θ ≤ 1



- f is concave if −f is convex
- f is strictly convex if dom f is convex and

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

for x, y ∈ dom f, x ≠ y, 0 < θ < 1

Now you say a function f in concave is –f is convex. That means negative curvature, downward curvature, something like that. And it's strictly convex if it's convex. And not only that, but this inequality here holds with strict inequality provided data is strictly between zero and one. So that's strict convexity. And you have strict concavity too.

Examples on R

Convex:
- affine: $ax+b$ on $\mathbf{R}$, for any $a,b \in \mathbf{R}$
- exponential: e^{ax}, for any $a \in \mathbf{R}$
- powers: x^{α} on $\mathbf{R}_{++}$,for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on $\mathbf{R}$, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

Concave:
- affine: $ax+b$ on R, for any $a, b \in \mathbf{R}$
- powers: x on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$

Some examples are just on R. So these are just functions you can draw. So the first is just an affine function, so that's linear plus a constant. That's – it has zero curvature, so it's convex. And in fact what happens for a function that is affine is the following; is that in effect you have equality here: $f(\theta x + (1 - \theta)y) = \theta f(x) + (1 - \theta)f(y)$
For always, for an affine function that's exactly what it means. Exact function value. So that's essentially the boundary of a set of convex functions.

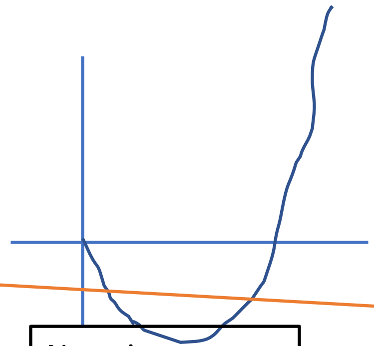
Exponential, doesn't matter what the coefficient is in here, this is convex. So if A is positive, it's increasing, but the curvature's upwards. If it's negative, it's a decreasing function, but it's convex.

Powers separate out. It depends on the values of the exponent. If the exponent's one or bigger, or if it's negative, or – well zero that's just a constant one, in that case it's convex.
And you have things like power of absolute value would be another one.

Examples on R

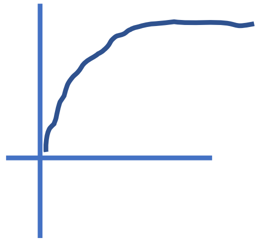
Convex:
- affine: ax+b on **R**, for any a,b ∈ **R**
- exponential: e^{ax}, for any a ∈ **R**
- powers: x^{α} on $\mathbf{R}_{++}$,for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^{p}$ on **R**, for $p \geq 1$
- negative entropy: x log x on $\mathbf{R}_{++}$

Concave:
- affine: ax+b on R, for any a, b ∈ **R**
- powers: x on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: log x on $\mathbf{R}_{++}$



Negative entropy

Negative entropy is x log x. That's something that goes like this; it's got this infinite slope here and then goes like that. Something like that, and – but at point as it curves upwards, that's negative entropy. So entropy, which is minus x log x is going to be concave.



Logarithm function

Now in concave functions examples would be like an affine function, and in fact, of course, if a function is both convex and concave, then it's affine. And that's clear because it says basically; this inequality holds one way and the other. That means its equality. This in fact implies the function is affine.

This power's in the range between zero and one, you know, like square root for example, you just draw this, and it's clearly concave. Logarithm, it's another famous example.

Examples on R^n and R^{mxn}

Affine functions are convex and concave;
all norms are convex

Example on R^n

- Affine function $f(x) = a^Tx+b$
- norms: $\|x\|_p = (\sum_{i=1}^{n} |xi|^p)^{1/p}$ for $p \geq 1$;
  $||x||_\infty \max_k = |x_k|$

Examples on R^{mxn}

- affine function
  $$f(X) = tr(A^TX)+b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}X_{ij} + b$$
- spectral (maximum singular value) norm
  $$f(X)=\|X\|_2 =\sigma_{max}(X) =(\lambda_{max}(X^TX))^{1/2}$$

An example on R^n and R mxn, that's the set of M by N real matrices. you have an affine function on RN, that's a transpose x plus b. That's a general linear function plus a constant. So this is the form of a general linear function, affine function on RN. And that's going to be convex, it's also concave.

Norms; so any norm is convex. That follows actually from triangle and equality, or –that's part of the definition of being a norm. And examples are things like this; the so-called p-norms, which is the sum of the absolute value of x_i and then to the one over p. Now for p equals one, that's the sum of the absolute values for P equals two.

Examples on R^n and R^{mxn}

Affine functions are convex and concave;
all norms are convex

Example on R^n

- Affine function $f(x) = a^Tx+b$
- norms: $\|x\|_p = (\sum_{i=1}^{n} |xi|^p)^{1/p}$ for $p\geq 1$;
  $||x||_\infty \max_k = |x_k|$

Examples on R^{mxn}

- affine function
  $$f(X) = tr(A^TX)+b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$
- spectral (maximum singular value) norm
  $$f(X)=\|X\|_2 = \sigma_{max}(X) = (\lambda_{max}(X^TX))^{1/2}$$

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

Here's an interesting function, which is the norm of a matrix. So that's –the spectral norm or the maximum singular value, or actually there's probably just a couple other names for it, the L2 induced norm.

So this is the square root of the largest eigen value ((λ_{max}) times X transpose X. Now I want to point something out, that is very – that's a very complicated function of a matrix. So that's – it is not a simple – to take the matrix form X transpose X, find the largest eigen value of it then take the square root of that.

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

Restriction of a convex function to a line

f: $\mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function g: $\mathbf{R} \to \mathbf{R}$,

$g(t) = f(x + tv)$,   **dom** g = $\{t \mid x + tv \in$ **dom** $f\}$

is convex (in t) for any $x \in$ **dom** f, $v \in \mathbf{R}^n$

can check convexity of f by checking convexity of functions of one variable

Example: f : $\mathbf{S}^n \to \mathbf{R}$ with f(X) = log detX, **dom f** = $\boldsymbol{S}^n_{++}$

$g(t) = \log\det(X + tV) = \log\det X + \log\det(I + t X^{-1/2} V X^{-1/2})$

$= \log\det X + \sum_{i=1}^{n} \log(1 + t\lambda_i)$

where, λ_i are the eigenvalues of $X^{-1/2} V X^{-1/2}$

g is concave in t (for any choice of $X \succ 0$, V); hence f is concave

Now here's one extremely useful property for convex functions, is this; a function is convex if and only if it's convex when it's restricted to any line.
It's basically if you take some complicated function, multiple dimensions, and you take a line, then – and then view that function on that line, you should view something that looks like that. And if that happens no matter what line you choose, it's convex. That's what it is.

And let's look at an example; so here's an interesting function; if the log of the determinant of a positive definite matrix – by the way, that's a complicated function right there.
Basically, it says that if you have two positive definite matrices, and you evaluate the log and determinant of them, and then you form a blend of the two, let's say the average.

Restriction of a convex function to a line

f: $\mathbf{R}^n \rightarrow \mathbf{R}$ is convex if and only if the function g: $\mathbf{R} \rightarrow \mathbf{R}$,

$g(t) = f(x + tv)$, $\quad$ **dom** $g = \{t \mid x + tv \in$ **dom** $f\}$

is convex (in t) for any $x \in$ **dom** f, $\quad v \in \mathbf{R}^n$

can check convexity of f by checking convexity of functions of one variable

Example: $f : \mathbf{S}^n \rightarrow \mathbf{R}$ with $f(X) = \log \det X$, **dom f** = $\boldsymbol{S}^n_{++}$

$(\mathbf{I} + t\, \mathbf{Q}\, \boldsymbol{\Lambda}\, \mathbf{Q}^T)$

$g(t) = \log \det(X + tV) = \log \det X + \log \det(I + t X^{-1/2} V X^{-1/2})$

$X^{1/2}(I + t X^{1/2} V X^{1/2}) X^{1/2}$ $\qquad = \log \det X + \sum_{i=1}^{n} \log(1 + t\lambda_i)$

where, λ_i are the eigenvalues of $X^{-1/2} V X^{-1/2}$

g is concave in t (for any choice of $X \succ 0$, V); hence f is concave

And let's look at an example; so here's an interesting function; if the log of the determinant of a positive definite matrix – that's a complicated function right there. Actually, this function is concave.
What does it mean to say its concave? Basically, it says that if you have two positive definite matrices, and you evaluate the log and determinacies of them, and then you form a blend of the two, let's say the average. It says, the log of the determinant of that average is bigger than or equal to the average of the log of the determinants of the end points.
Let's see how this works; so, to establish this we have to pick an arbitrary line in symmetric matrix space. So, what does a line look like in symmetric matrix space? It's one that passes through the positive definite cone.
So it's going to look like this; without lost generality it looks like a point. X, which is positive definite, plus t times a direction V. Now this direction V is a symmetric matrix, but it does not have to be positive semi-definite, right, because that's a line in a direction – there's no reason the direction has to be positive – it does have to be symmetric. So this thing describes a function of a single variable t.

Restriction of a convex function to a line

f: $\mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function g: $\mathbf{R} \to \mathbf{R}$,

$g(t) = f(x + tv)$, **dom** g = $\{t \mid x + tv \in$ **dom** f$\}$

is convex (in t) for any $x \in$ **dom** f, $v \in \mathbf{R}^n$

can check convexity of f by checking convexity of functions of one variable

Example: f : $\mathbf{S}^n \to \mathbf{R}$ with $f(X) = \log \det X$, **dom f** = $\mathbf{S}^n_{++}$

$(I + t\,Q\,\Lambda\,Q^T)$

$g(t) = \log\det(X + tV) = \log\det X + \log\det(I + t\,X^{-1/2} V X^{-1/2})$

$X^{1/2}(I + t\,X^{1/2}\,V\,X^{1/2})\,X^{1/2}$ $= \log\det X + \sum_{i=1}^{n} \log(1 + t\lambda_i)$

where, λ_i are the eigenvalues of $X^{-1/2} V X^{-1/2}$

g is concave in t (for any choice of $X \succ 0$, V); hence f is concave

Let's work out what this is; well a right X – I'm going to write this – there's lots of ways to do it, but X is positive definite so it has a square root. So I'll take half out on each side, and this will look like this, T, it's going to look like that, times X half ($X^{1/2}(I + t\,X^{1/2}\,V\,X^{1/2})\,X^{1/2}$)
That's what – this matrix is this. But the determinant of a triple product is the product of the determinacies. And you take logs and it adds and all that, so you get this because log det X half plus log det X half is log det X. So you get this thing here.

It's still not too obvious, but what we're going to do now is this; I'm going to write the – this thing is a symmetric, but not necessarily positive – it's certainly not positive semi- definite or anything like – you don't know. Matrix, we'll take its eigen expansion. We'll write this as whatever, Q lambda Q transpose, and you get this with a t there like this.
What I'll do now is I'll do the trick of writing I as QQ transpose then pull it out on either side and I'll get det Q. det Q is either + or - one, but it doesn't matter. And then I'll end up with det I plus t lambda. That's a diagonal matrix. I know what the determinant of a diagonal matrix is, it's a product of the entries, and so I get this thing.

Restriction of a convex function to a line

f: $\mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function g: $\mathbf{R} \to \mathbf{R}$,

g(t) = f (x + tv), **dom** g = {t | x + tv ∈ **dom** f}

is convex (in t) for any x ∈ **dom** f, v ∈ $\mathbf{R}^n$

can check convexity of f by checking convexity of functions of one variable

Example: f : $\mathbf{S}^n \to \mathbf{R}$ with f(X) = log detX, **dom f** = $\boldsymbol{S}^n_{++}$

$(\mathbf{I} + \mathbf{t}\,\mathbf{Q}\,\boldsymbol{\Lambda}\,\mathbf{Q}^T)$

g(t) = log det(X + tV) = log det X +log det(I + t $X^{-1/2} V X^{-1/2}$)

$X^{1/2}(I + t X^{1/2} V X^{1/2}) X^{1/2}$ = log det X +$\sum_{i=1}^{n} \log(1 + t\lambda_i)$

where, λ_i are the eigenvalues of $X^{-1/2} V X^{-1/2}$

g is concave in t (for any choice of X ≻ 0, V); hence f is concave

How did you choose the directional matrix V?
It's extremely important that it's completely arbitrary. So the technical answer to your question, how did I choose the direction V is, I didn't. Or arbitrarily. That's – because if V is – if I chose V in any special way then my final conclusion is not going to hold, it's not right.

To be convex has to be every line. It has to be convex when restricted to any line through the point .
Now we're in pretty good shape, because I know what log 1 plus tλ looks like. For any real number of lambda – if λ's positive it looks like one thing and it goes like this, or it goes like that. If λ has the other sign it goes like that; either way curvature is negative, and so this is concave

This is going to turn out to have sorts of implications. If you're in statistics, if you've taken information theory, communications, a lot of other things like that, actually, if you've done any computational geometry, it turns out some things you know are actually related to concavity of log det X

Extended-value extension

extended-value extension $\tilde{f}$ of f is

$\tilde{f}$(x) = f(x), x ∈ **dom** f, $\tilde{f}$(x) = ∞, x ∉ **dom** f

often simplifies notation; for example, the condition

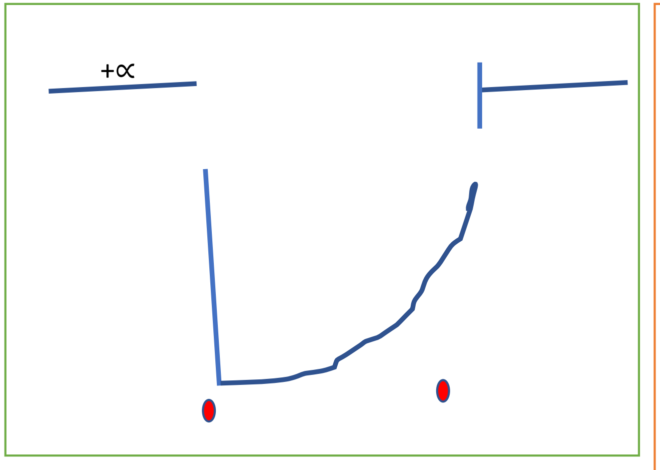
$0 \leq \theta \leq 1 \Longrightarrow \tilde{f}(\theta x + (1-\theta)y) \leq \theta \tilde{f}(x) + (1-\theta) \tilde{f}(y)$

(as an inequality in **R** ∪ {∞}), means the same as the two conditions

- **dom** f is convex
- for x,y ∈ **dom** f

$0 \leq \theta \leq 1 \Longrightarrow f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$

if you have a function f with some domain, then we define an extended valued extension as follows; if you're in the domain and we'll assign it infinity outside the domain. And, but technically there's a different between f and $\tilde{f}$



So if you have a function, sort of it looks like that, and its domain is from here to here, that's a convex function. What we simply do is we simply assign it the value sort of +∝ outside. So it looks like that, so you make it +∝. Actually everything works, including this inequality if you say – if you take this point and this point, everything works. If you draw this line will now have a slope going straight up and it all just works.

First-order condition

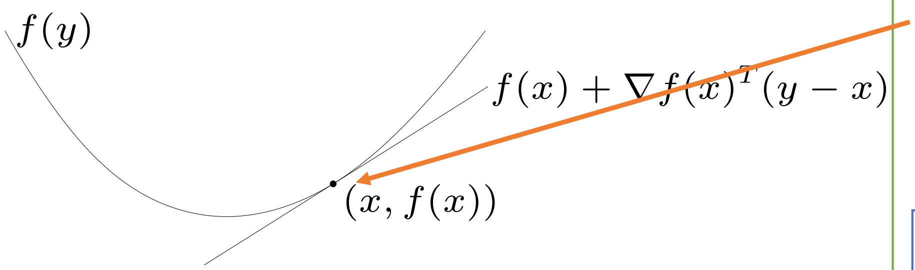
f is **differentiable** if **dom** f is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_1}, \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_2}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_n}\right)$$

exists at each $x \in$ **dom** f

1st odder condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)\,(y-x) \text{ for all } x, y \in \text{dom } f$$



First order approximation of f is global under estimator

We get to first order of condition. First of all, a function is differentiable if its domain is open. it's differentiable to point if the point is in the interior of the domain, but we'll just talk about being differentiable period.
So here's the first order condition for convexity; actually it's a hint as to why convex opposition works well. Here it is, it says the following; form that is the Taylor approximation. The first order Taylor approximation of f at x. Okay, that says as a function of y. That's the Taylor approximation.

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

So it's small – the error is small squared. So these two functions, one is this affine function and one is this one, is that the are very near as long as y is near x. For a convex function this thing is actually always an underestimator. It says that the Taylor expan – first order of Taylor expan is a global underestimator of function.

Second-order conditions

f is **twice differentiable** if **dom** f is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f}{\partial \boldsymbol{x}_i \, \partial \boldsymbol{y}_j} \qquad i,j = 1,2,\ldots.n$$

2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and if

$$\nabla^2 f(x) \succcurlyeq 0 \text{ for all } x \in \textbf{dom } f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in$ **dom** f, then f is strictly convex

Second order of condition; if it's twice differentiable, it says it's a hessian, which is the makers of partial derivatives. If that is – it says the following; it says it's convex if and only if that function is positive-semidefinite. So that's the condition.

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

Examples

quadratic function: $f(x) = (1/2)\, x^T P x + q^T x + r$ (with $P \in \mathbf{S}^n$)

$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$

convex if $P \succeq 0$

least-squares objective: $f(x) = \|Ax-b\|_2^2$

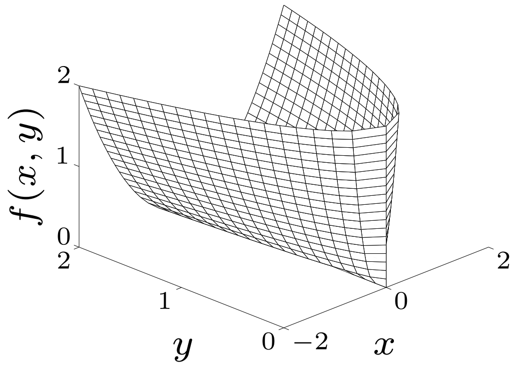
$\nabla f(x) = 2A^T(Ax-b) \qquad \nabla^2 f(x) = 2A^T A$

convex (for any A)

quadratic-over-linear: $f(x, y) = \frac{x^2}{y}$

$$\nabla^2 f(x,y) = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T$$

convex for y>0



Let's characterize all quadratic functions right now. So a quadratic function looks like this; it's a quadratic form, it's a p- asymmetric here, a linear function and a constant.
So the gradient of this function is Px plus q, and the hessian is P. So it's constant, it's got a constant hessian. So a quadratic function is convex if and only if P is bigger than or equal to zero, if it's a positive-semidefinite matrix. And an example would be least squares objective. That just – immediately. So here the hessian is A transpose A and that's going to be positive-semidefinite so that's always convex.

Here's a function that is not obvious, it's one – this is probably one of the first functions you encounter that is not obvious, it's not obvious that it's convex. Maybe other than minus log det x, but here's one; x squared over y just two variables. So it's – this is convex in x and y provided y is positive.

Examples

quadratic function: $f(x) = (1/2)\, x^T P x + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x)=Px+q, \qquad \nabla^2 f(x)=P$$

convex if $P \succcurlyeq 0$

least-squares objective: $f(x) = \|Ax-b\|_2^2$

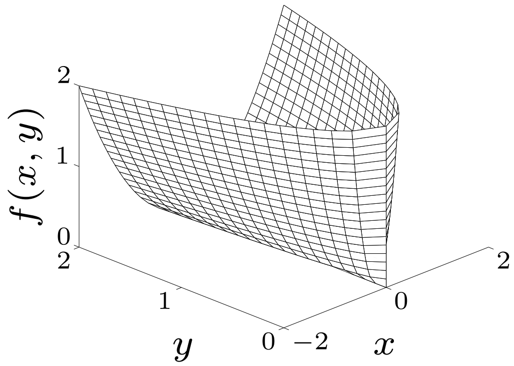
$$\nabla f(x)=2A^T(Ax-b) \qquad \nabla^2 f(x)= 2A^TA$$

convex (for any A)

quadratic-over-linear: $f(x, y) = \dfrac{x^2}{y}$

$$\nabla^2 f(x,y)= \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T$$

convex for $y>0$



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

Take a look at this function, here's a plot of it, so it looks like a boat or something like that anyway, so and if one slice it's quadratic and, if you slice it different directions you get obviously convex things. So lets work out the hessian, you calculate Partial squared f, partial x, partial y, and the diagonals, and you fill it out and you get a matrix which indeed is rank one, and positive-semidefinite. And here we're using the fact that Y is positive here, to ensure that this is positive semidefinite. So that's convex.

Examples

log-sum-exp: f (x) = $\log\sum_{k=1}^{n} \exp xk$ is convex

$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z}$ **diag** (z) - $\frac{1}{(\mathbf{1}^T z)2} zz^T$ ($z_k = \exp xk$)

to show $\nabla f(x) \succcurlyeq 0$, we must verify that $v^T\nabla^2 f(x)v \geq 0$ for all v:

$$v^T\nabla^2 f(x)v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

Since $(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$
(from Cauchy-Schwarz inequality)

Geometric mean: $f(x) = (\prod_{k=1}^{n} x_k)^{1/n}$ on R_{++}^{n} is concave

(similar proof as for log-sum-exp)

Here's another on log sum x function. That's the log of a sum of exponentials of variables. First of all it's sort of like a smooth max, or in fact some people call – there are fields where
this is simply referred to as soft max. This is universally called the soft max. The reason it's soft max is this; if you take a bunch of variables, x_1 through x_n, then x increases of course very quickly. So the biggest x is a lot bigger. If there's a good gap between x – the largest x and the next one, x will accentuate the spread.
If you add these up and then take the log – if for example one of x's is kind of isolated and far away from the second – if the biggest is away from the second biggest then basically this sum x is basically x of the largest. You take the log of that and you get the largest. So this is sort of a smooth – it's a soft max some people call it.

So that's a convex function, and you take the hessian, calculate partial squared f, partial x_i, partial x_j, when you do this you look at it and you see well there's this first term, z here is this x of x, so it's obviously – it's non-negative. That's a positive diagonal matrix
so this is positive definite

Examples

log-sum-exp: f (x) = $\log\sum_{k=1}^{n} \exp xk$ is convex

$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z}$ **diag** (z) - $\frac{1}{(\mathbf{1}^T z)2} zz^T$ ($z_k = \exp xk$)

to show $\nabla f(x) \succcurlyeq 0$, we must verify that $v^T \nabla^2 f(x) v \geq 0$ for all v:

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

Since $(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$
(from Cauchy-Schwarz inequality)

Geometric mean: f(x) =$(\prod_{k=1}^{n} x_k)^{1/n}$ on R_{++}^{n} is concave

(similar proof as for log-sum-exp)

You still have to show this is positive-semidefinite. How do you show a matrix positive-semidefinite? You simply show that the associate quadratic form in always non-negative. So you put a v on the left to be transposed, a v on the right. You plug that in here and see what happens and you get this thing. And this turns out to be greater than or equal to zero using theCauchy-Schwartz inequality applied here,

Geometric mean is concave, so this is the product of a bunch of variables, these have to be positive, although actually this works for non-negative.

So it's a bunch of variables and then the nth root of that, so that's concave. And it's the same type of argument as for log sum exp x_k, same type of thing.

Epigraph and sublevel set

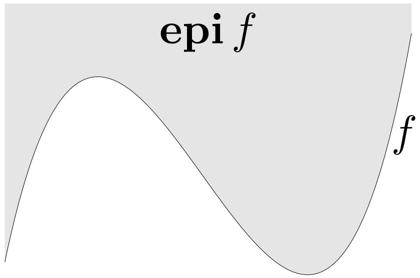
α-**sublevel set** of $f : \mathbf{R}^n \rightarrow \mathbf{R}$:

$$C_\alpha = \{x \in \text{dom } f \mid f(x) \leq \alpha\}$$

sublevel set of convex functions are convex (converse is false)

Epigraph of $f : \mathbf{R}^n \rightarrow \mathbf{R}$

$$\textbf{epi } f = \{(x,t) \in \mathbf{R}^{n+1} \mid x \in \textbf{dom } f, f(x) \leq t\}$$



f is convex if and only if **epi** f is a convex set

The idea of an epigraph and a sublevel set. So, if you have any function then the sublevel – the alpha sublevel set is the set of points with f value less than α.
And for example, if f is a design, then if f represents the power dissipated by some circuit design or something like this, this will be the set of designs that meet an α spec.
That's what a sublevel set will often mean. By the way, if it's an estimation problem, an f is a measure of implausibility, like negative log likelihood in a statistical estimation problem. This will be the set of points, which are at least alpha plausible values
If you have a convex function, then the sublevel sets are convex. The connection between convex function, convex set, it applies to sets, and it applies to functions. The real connection is through something called the epigraph. So, epi means above, and so epigraph means everything about the graph.
The graph of a function of course, is the set of pairs, x,t. So it's an $\mathbf{R}^{n+1}$, it's the actual graph. The epigraph is everything above it. So if I have a function, the epigraph is this shaded region like this. And here's the real connection between convex set and functions.

Jensen's inequality

basic inequality: if f is convex, then for $0 \le \theta \le 1$,

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

extension: if f is convex, then

$f(\mathbf{E}z) \le \mathbf{E}\, f(z)$

for any random variable z

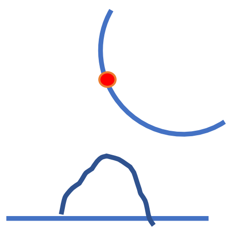
basic inequality is special case with discrete distribution

prob $(z = x) = \theta$, **prob** $(z = y) = 1 - \theta$

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

If f is convex, then f of an expected value is less than or equal to expected value of f of x. Now that's called Jensen's inequality. And this is where Z is a random variable, however, which is in the domain of f almost surely. So that's this, and this is Jensen's inequality. And it's a distribution on Z, extremely simple. It takes only two values, x and y with probability theta and one minus theta, and you recover this thing (i.e., $f(\theta x + (1 - \theta)y) \le \theta f(x) + (1 - \theta)f(y)$)

And let me explain what that means; it means that if I have a function – a point here, and here's a convex function, like that, okay? Let's imagine, that's actually my target point in some process, but now, when I manufacture it, I actually get a distribution of values like that? Whose mean is this point, So that's the – this happens, so this is – and then this tells you the cost? And this could be the power, something like that.



Operations that preserve convexity

practical methods for establishing convexity of a function

1. verify definition (often simplified by restricting to a line)

2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$

3. show that f is obtained from simple convex functions by operations that preserve convexity

- composition with affine function
- pointwise maximum and supremum
- nonnegative weighted sum
- minimization
- composition
- perspective

Look at functions and figure out if they're convex or not, that's what you have to do. So the methods I just looked at involving lines and all that, if you have to resort to that, if you actually write out a hessian, you do this only in a last resort.

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

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

Positive weighted sum & composition with affine function

nonnegative multiple: αf is convex if f is convex, $\alpha \geq 0$

sum: $f_1 + f_2$ convex if f_1 , f_2 convex (extends to infinite sums, integrals)

composition with affine function: f (Ax + b) is convex if f is convex

examples

- log barrier for linear inequalities

  $f(x) = -\sum_{i=1}^{m} \log(bi - a_i^T x)$ dom f ={x| $a_i^T x < b_i$, i=1,...,m}

- (any) norm of affine function: f (x) = ‖Ax + b‖

If you have a convex function and you scale it by a non-negative scalar, it's convex, that's totally obvious. If you add two functions that are convex, it's convex. And that extends to adding five functions and it goes to an integral or even an expected value of convex functions would be convex.

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

Positive weighted sum & composition with affine function

nonnegative multiple: αf is convex if f is convex, $\alpha \geq 0$

sum: $f_1 + f_2$ convex if f_1 , f_2 convex (extends to infinite sums, integrals)

composition with affine function: f (Ax + b) is convex if f is convex

examples

- log barrier for linear inequalities

  $f(x) = -\sum_{i=1}^{m} \log(bi - a_i^T x)$ dom f $= \{x \mid a_i^T x < b_i, i=1,\dots,m\}$

- (any) norm of affine function: $f(x) = \|Ax + b\|$

Let's look at some examples; we can make some examples now. f of x is minus log sum b_i minus a_i transpose x, and this is defined on the region where these are strictly positive. That's the interior of a polyhedron. So I have a polyhedron defined by a_i transpose x less than b_i. The interiors where that's strictly less, and that's where this makes sense.

Now we're going to simply pre-compose this function with this affine mapping. I'll call it b minus ax, that's an affine mapping. And that gives you this function and then here, I just applied this, this composition rule. Here's another one; is the norm of any affine function, so norm ax plus b is going to be affine – convex. It's going to be a convex function of x.

Pointwise maximum

If $f_1,\dots,f_m$ are convex then $f(x) = \max\{f_1(x),\dots,f_m(x)\}$ is convex

Examples

- piecewise-linear function: $f(x) = \max_{i=1,..,m} (a_i^T x + b_i)$ is convex

- sum of r largest components of $x \in \mathbf{R}^n$:

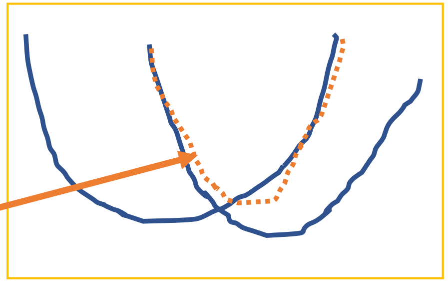
$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

Is convex ($x_{[i]}$ is the ith largest component of x)

Proof:

$f(x) = \max\{x_{i1} + x_{i2} + \cdots + x_{ir} \mid 1 \le i_1 < i_2 < \cdots < i_r \le n\}$

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



By the way, in terms of epigraphs, what does this correspond to? Precisely. So calculating point wise maximum of functions, in fact you can even write some formula for it, you know something like this; epi of max over i f_i is equal to the intersection over i of epi f_i, something like that.

$\operatorname{epi}(\max_i f_i) = \bigcap_i \operatorname{epi}(f_i)$

Pointwise maximum

If $f_1,\ldots,f_m$ are convex then $f(x) = \max\{f_1(x),\ldots,f_m(x)\}$ is convex

Examples

- piecewise-linear function: $f(x) = \max_{i=1,..,m}(a_i^T x + b_i)$ is convex

- sum of r largest components of $x \in \mathbf{R}^n$:

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

Is convex ($x_{[i]}$ is the ith largest component of x)

Proof:

$$f(x) = \max\{x_{i1} + x_{i2} + \cdots + x_{ir} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$$

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

Here's one, this is again, not obvious. The sum of the R largest components of a vector. So take a vector in R50 and the sum of the top three elements. That's a very complicated Piecewise linear function, it's convex.

Why is it convex, because it's the maximum of a transpose x or a – this is for the sum of the top three, is any vector with three one's and 47 zeros. Now there's a giant pile of those, there's 50, choose three. But the point is, it's the maximum of 50 choose three linear functions done. It's convex.

Pointwise supremum

if $f(x,y)$ is convex in x for each $y \in A$, then

$$g(x) = \sup_{y \in A} f(x, y)$$

Is convex

examples

- support function of a set C:

  $S_C(x) = \sup_{y \in C} y^T x$ is convex

- distance to farthest point in a set C:

  $f(x) = \sup_{y \in C} \|x - y\|$

- maximum eigenvalue of symmetric matrix:

for $X \in \mathbf{S}^n$

$$\lambda(X) = \sup_{\|y\|_2 = 1} y^T X y$$

So here's the statement; if you have a function of two variables, x and y, and suppose it's convex in x for each y in some set, you don't even know what the set is, totally irrelevant what the set is.
Then it says that if you take – again, you can leave this as max if you haven't seen this before, this is simply – if you simply take the maximum over this possibly infinite collections of functions point wise you get a new function, that's going to be convex.

Let's take the maximum eigen value of asymmetric matrix. That is a complicated function. For a matrix that is six by six or bigger, there is no formula for the roots of a sixth degree and order and higher order polynomial.
If the supremum of y transpose Xy over all y that have – over the units sphere. The units' sphere is the set of all points whose norm is one.
So, for each Y that's a linear function. This is a supremum of an infinite collection of linear functions. In fact, there's one for every point on the units' sphere in our end. Supremum of a bunch of linear functions, linear functions are convex. Supremum over these things is convex, that's a convex function.

Composition with scalar functions

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

$$f(x) = h(g(x))$$

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)

$$f''(x) = h''(g(x))\, g'(x) + h'(g(x))\, g''(x)$$

- note: monotonicity must hold for extended-value extension $\tilde{h}$

examples

- 1/g(x) is convex if g is concave and positive
- exp g(x) is convex if g is convex

if you have h of g of x, -- so the rule goes like this, and it says that a convex increasing function of a convex function is convex. So if this thing will be convex if you have a convex – if the outer function is convex and increasing, which non-decreasing. So that's the condition. And the way to derive these as other ones, for example, it's convex, if this thing is concave and H is convex and non decreasing.

Now the way to check chain rules. So, if you take – you imagine that these functions are differentiable and of one variable and you work out the second derivative and you get something like this.
G is concave. So that means this here, and what are the conditions on h to make fconcave? So–assume that $h'(g(x))\, g''(x)$ less or equal to zero, and I need thing $g'(x)$ to be less than or equal to zero. That means this $h''(g(x))\, g'(x)$ has to be positive. So, we have to have h'positive. So, h has to be increasing.

expt g of a convex function is convex. It says that the exponential preserves positive of a curvature.
One over the square root is a convex function, and it's x minus one half, and it goes kind a like that.