Lecture 2 - Convex functions

- convex function definition
- first order condition
- second order condition
- examples of convex functions
- operations that preserve convexity.

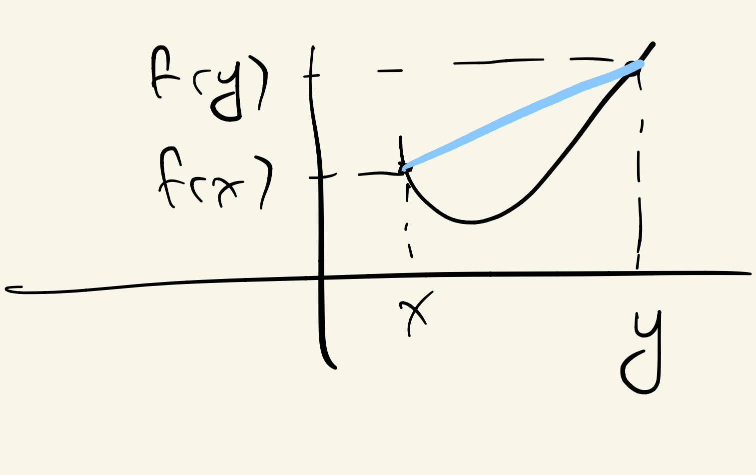
Definition of convex function

A function $f: \mathbb{R}^n \to \mathbb{R}$ if:

1. $\mathrm{dom}\, f$ is convex
2. $\forall x, y \in \mathrm{dom}\, f$, $\theta \in [0,1]$

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

// The graph of f "between" x and y lies below the line that connects $(x, f(x))$ and $(y, f(y))$

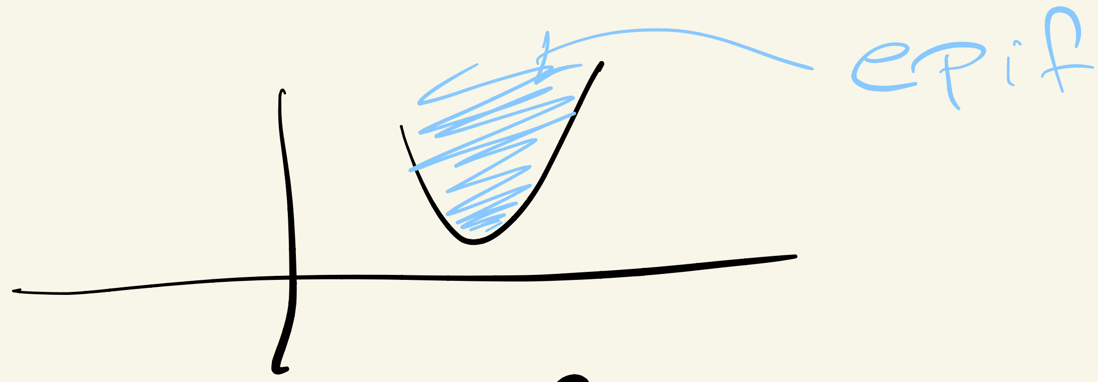


strictly convex $\leadsto$ $\cancel{\leq}$ $<$

concave $\leadsto$ $\cancel{\leq}$ $\geq$

Epigraph (the connection between convex functions and convex sets)

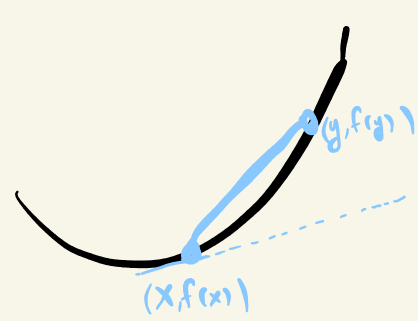
$$\text{epi} f := \{ (x,t) : x \in \text{dom} f,\ t \geq f(x) \}$$



Lemma: f is convex iff $\text{epi} f$ is a convex set.

First order conditions (let's assume that everything is differentiable)

Let's first get an intuition from the one-dimensional case,



let $y > x$. In that case

$$f'(x) \leq \frac{f(y) - f(x)}{y - x}$$

otherwise by Taylor the first order approximation would be above the line connecting $(x, f(x))$ and $(y, f(y))$

Theorem

f is differentiable and ∇f exists at each point in $\text{dom} f$ which is open. Then

f is convex iff $\text{dom} f$ is convex

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \forall x, y \in \text{dom} f$$

proof $\longrightarrow$

$\Longrightarrow$

we will first prove it for one dimension and then move to higher dimensions

one-dimension

$\forall\, t \in [0,1]$, $x, y \in \text{dom} f$

$$f(ty+(1-t)x) \le t f(y) + (1-t) f(x) \Rightarrow$$

$$\Rightarrow f(x+t(y-x)) - f(x) \le t(f(y)-f(x))$$

$$\overset{t>0}{\Rightarrow} \frac{f(x+t(y-x)) - f(x)}{t} \le f(y) - f(x)$$

$$\xrightarrow{t \to 0^+} f'(x)(y-x) \le f(y) - f(x) \quad \boxtimes$$

multiply numerator and denominator with $y-x$

higher dimension

usefull lemma

let $f: \mathbb{R}^n \to \mathbb{R}$, $x, y \in \text{dom} f$ and

restriction to the line passing through x and y

$$g(t) = f(ty+(1-t)x),\ t \in [0,1] \qquad g(0) = f(x),\quad g(1) = f(y)$$

f convex $\Rightarrow$ g convex

proof

let $t_1, t_2 \in [0,1]$ and $\theta \in [0,1]$

$$g(\theta t_1 + (1-\theta) t_2) =$$

$$= f((\theta t_1 + (1-\theta) t_2) y + (1 - \theta t_1 - (1-\theta) t_2) x) =$$

$$= f\left(\theta(t_1 y + (1-t_1) x) + (1-\theta)(t_2 y + (1-t_2) x)\right)$$

f is convex

$$\leq \theta f(t_1 y + (1-t_1) x) + (1-\theta) f(t_2 y + (1-t_2) x)$$

$$= \theta g(t_1) + (1-\theta) g(t_2) \quad \boxtimes$$

applying the first order condition to g

we get

$$g'(t) = \frac{\partial f(x + t(y-x))}{\partial t} =$$

$$= \langle y-x, \nabla f(x + t(y-x)) \rangle$$

$t_y, t_x \in [0,1]$

$$g(t_y) \geq g(t_x) + g'(t_x)(t_y - t_x)$$

$t_y = 1 \Rightarrow g(1) = f(y)$

$t_x = 0 \Rightarrow g(0) = f(x)$

$\boxtimes$

$$g'(0) = \langle y-x, \nabla f(x) \rangle$$

$\Leftarrow$

Now we have that $\forall x,y \in \operatorname{dom} f$

$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$ and

we want to prove that $\forall \theta \in [0,1]$

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

how? Try to get an inequality of

the form $f(x) \geq f(\theta y + (1-\theta)x) + (\ldots)$

and $f(y) \geq f(\theta y + (1-\theta)x) + (\ldots)$

$$f(x) \geq f(\theta y + (1-\theta)x) + \langle \nabla f(\theta y + (1-\theta)x), x - \theta y - (1-\theta)x \rangle \Rightarrow$$

$$\Rightarrow f(x) \geq f(\theta y + (1-\theta)x) + \theta \langle \nabla f(\theta y + (1-\theta)x), x-y \rangle$$

$$f(y) \geq f(\theta y + (1-\theta)x) + \langle \nabla f(\theta y + (1-\theta)x), y - \theta y - (1-\theta)x \rangle \Rightarrow$$

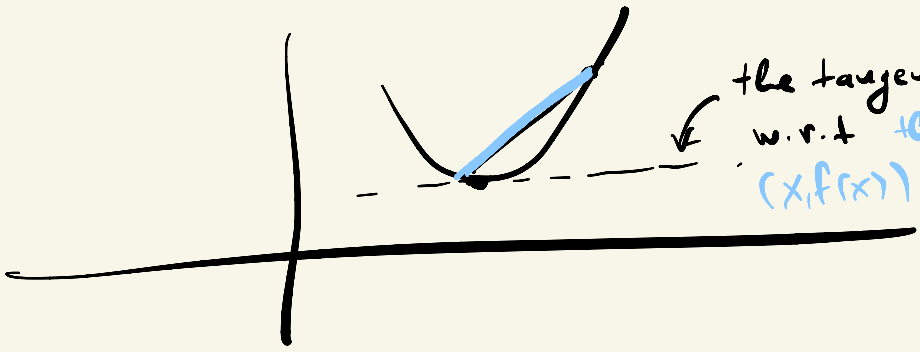
$$\Rightarrow f(y) \geq \quad // \quad + (1-\theta) \langle \nabla f(\theta y + (1-\theta)x), y-x \rangle$$

multiply first by $(1-\theta)$ } $\underline{\oplus}\!>$ add them ∎

second by θ

Second order condition

my intuition about the second order condition



the tangent slope is small w.r.t the line connecting $(x, f(x))$ and $(y, f(y))$

however since the line on the graph pass both through $(y, f(y))$ that means that the slope of the tangent has to increase

one-dimension $\Rightarrow f''(x) \geq 0 \quad \forall x \in \operatorname{dom} f$

higher dimensions $\Rightarrow \nabla^2 f(x) \succeq 0$

↙ hessian

$$(\nabla^2 f(x))_{ij} = \frac{\partial f}{\partial x_i \partial x_j}$$

classic Taylor $f: \mathbb{R} \to \mathbb{R}$

$$\tilde{f}(y) = f(x) + f'(x)(y-x) + \frac{f''(x)}{2!}(y-x)^2 + \dots$$

tensors ↘

Taylor for $f: \mathbb{R}^n \to \mathbb{R}$

$$\tilde{f}(y) = f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2!}(y-x)^T \nabla^2 f(x)(y-x) + \dots$$

Theorem (second order condition)

f is twice differentiable, $\nabla^2 f$ exists

at every point in dom f which is open

then

$$\nabla^2 f \succeq 0 \Rightarrow f \text{ is convex}$$

proof

one-dimension

$f''(x) \geq 0 \;\; \forall x \in \text{dom} f$

$\Rightarrow f'(y) \geq f'(x) \quad \forall y \geq x$

$$\Rightarrow f(y) - f(x) = \int_x^y f'(t)\,dt \geq \int_x^y f'(x)\,dt =$$

$$= (y-x) f'(x)$$

$\Rightarrow f(y) \geq f(x) + f'(x)(y-x)$ and by the first order condition we know that f is convex.

Before we used

usefull lemma

let $f: \mathbb{R}^n \to \mathbb{R}$, $x, y \in \operatorname{dom} f$ and

restriction to the line passing through x and y

$g(t) = f(ty + (1-t)x)$, $t \in [0,1]$ $\quad g(0) = f(x)$, $g(1) = f(y)$

f convex $\Rightarrow$ g convex

fortunately the lemma holds with iff.

Proof

fix $x, y \in \operatorname{dom} f$ we want to prove that $f(ty + (1-t)x) \leq$
$\leq t f(y) + (1-t) f(x)$

$$g(t) = g(t \cdot 1 + (1-t) \cdot 0) \overset{g \text{ convex}}{\leq}$$
$$t \cdot g(1) + (1-t) g(0) = t f(y) + (1-t) f(x)$$

$\square$

To prove the second order condition for higher dimensions it suffices to prove that g is convex.

$\nabla^2 f \succeq 0$

$$g'(t) = \langle y - x, \nabla f(ty + (1-t)x) \rangle$$

because

$$g''(t) = (y-x)^T \cdot \nabla^2 f(ty + (1-t)x)(y-x) \geq 0$$

g is convex $\Rightarrow$ f is convex $\quad \square$

Theorem

f is twice continuously differentiable, $\mathrm{dom} f$ is open

f convex $\Rightarrow \nabla^2 f(x) \succeq 0 \ \forall x \in \mathrm{dom} f$

proof

suppose $\nabla^2 f(x) \not\succeq 0 \Rightarrow \exists v: v^T \nabla^2 f(x) v < 0$

now consider $x + \varepsilon v$ for sufficiently small ε. Since $\mathrm{dom} f$ is open $x + \varepsilon v \in \mathrm{dom} f$

since $\nabla^2 f$ is continuous $v^T \nabla^2 f(x + \varepsilon v) v < 0$

By Taylor $f(x + v\varepsilon) = f(x) + \langle \nabla f(x), v\varepsilon \rangle$

$$+ (\varepsilon v)^T \cdot \nabla^2 f(x + \varepsilon' v) \varepsilon v$$

for $0 \leq \varepsilon' \leq \varepsilon$

$\Rightarrow f(x + v\varepsilon) < f(x) + \langle \nabla f(x), v\varepsilon \rangle$

which violates the first order condition

$\Rightarrow f$ is not convex $\boxtimes$

examples of convex functions

- $e^{ax} \rightsquigarrow (e^{ax})'' = a^2 e^{ax} \geq 0 \quad \forall x \in \mathbb{R}$

- $x\log x \rightsquigarrow (x\log x)'' = (1+\log x)' = 1/x > 0 \quad \forall x \in \mathbb{R}_+$

- $\|\cdot\|$ a norm function is convex because of the triangle inequality

$$\|ty+(1-t)x\| \leq \|ty\| + \|(1-t)x\| = t\|y\| + (1-t)\|x\|$$

Quadratic over linear

$$f(x,y) = x^2/y \quad , y > 0$$

$$\nabla^2 f = \begin{bmatrix} 2/y & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} =$$

$$= \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y & -x \end{bmatrix} \succeq 0$$

positive semidefinite

$M = r^T \cdot r$

Log-sum-exponential

$$f(x) = \log \sum e^{x_i}$$

softmax

$\Rightarrow \max\{x_1, \dots, x_n\} \le f(x_1, \dots, x_n) \le \max\{x_1, \dots, x_n\} + \log n$

proof

let $x^* = \max\{x_1, \dots, x_n\}$

$$x^* = \log e^{x^*} \le \log \sum e^{x_i}$$

$$\log \sum e^{x_i} \le \log(n \cdot e^{x^*}) = \log n + \log e^{x^*} =$$

$$= \log n + x^*$$

$$\frac{\partial f(x)}{\partial x_i \partial x_j} = \begin{cases} \dfrac{e^{x_i}}{\sum e^{x}} - \dfrac{(e^{x_i})^2}{(\sum e^{x})^2} & i = j \\[2ex] -\dfrac{e^{x_i} e^{x_j}}{(\sum e^{x_k})^2} & i \neq j \end{cases}$$

$$v^T \cdot \nabla^2 f \, v = \frac{1}{(\sum e^{x})^2} \left(\left(\sum_i v_i^2 e^{x_i} \right) \left(\sum e^{x} \right) - \sum_{i,j} v_i v_j e^{x_i} e^{x_j} \right)$$

$$= \frac{1}{(\sum e^{x})^2} \left(\left(\sum_i v_i^2 e^{x_i} \right) \left(\sum e^{x} \right) - \left(\sum_i v_i e^{x_i} \right)^2 \right)$$

$$= \frac{1}{(\sum e^{x})^2}\left(\left(\sum_i v_i^2 e^{x_i}\right)\left(\sum e^{x}\right) - \left(\sum_i v_i e^{x_i}\right)^2\right)$$

let $a \in \mathbb{R}^n$ s.t $a_i = v_i \cdot \sqrt{e^{x_i}}$

let $b \in \mathbb{R}^n$ s.t $b_i = \sqrt{e^{x_i}}$

$$\frac{1}{(\sum e^{x})^2}\left(\|a\|^2 \cdot \|b\|^2 - (\langle a \cdot b \rangle)^2 \right) \geq 0$$

(Cauchy-Swartz)

$$\Longrightarrow$$

Log-determinant $f(X) = \log\det(X)$

$, X \in S^n_{++}$

Important for maximizing the volume of an ellipsoid

Proof

trick restrict the function on a "line"

lemma

let $f : \text{dom} f \to \mathbb{R}$, $\text{dom} f \subseteq \mathbb{R}^n$. Then f is convex iff its restriction to the line is convex. More precisely

$\forall x_0 \in \text{dom} f, u \in \mathbb{R}^n$ then

$g(t) = f(x_0 + tu)$ has to be convex

over $\text{dom}\, g = \{t \in \mathbb{R} : x_0 + t \cdot u \in \text{dom} f\}$

$\Longrightarrow$

with this in mind, take

$Z \in S_{++}^n$ and $V \in S_{++}^n$

then we need to prove that

$g(t) = \log\det(Z + tV)$ is concave over

$\operatorname{dom} g = \{t : Z + tV \succ 0\}$

we check the second order condition

for g

$$g(t) = \log\det\left(Z^{1/2}\left(I + tZ^{-1/2}VZ^{-1/2}\right)Z^{1/2}\right)$$

$$= \log\det\left(Z\left(I + tZ^{-1/2}VZ^{-1/2}\right)\right) =$$

$$= \log\det(Z) + \log\det\left(I + tZ^{-1/2}VZ^{-1/2}\right) =$$

$$= \quad // \quad + \log\prod_{i=1}^{n}(1 + t\lambda_i)$$ λ_i eigenvalues of $Z^{-1/2}VZ^{-1/2}$

$$g'(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i} \quad \text{and} \quad g''(t) = \sum \frac{-\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0$$

g concave $\Rightarrow$ f concave ☒

operations that preserve convexity

- non-negative weight sums

$f = \sum w_i f_i \quad , w_i \geq 0 \quad , f_i$ is convex $\forall i$

- affine mapping : $g(x) = f(Ax+b)$

f convex $\Rightarrow$ g convex (assuming that after the affine transformation we are still in dom f)

proof

$g(\theta y + (1-\theta)x) = f(A(\theta y + (1-\theta)x) + b) =$

$= f(\theta(Ay+b) + (1-\theta)(Ax+b)) \leq$

$\leq \theta f(Ay+b) + (1-\theta) f(Ax+b) =$

$= \theta \cdot g(y) + (1-\theta) g(x)$

- pointwise maximum

$f(x) = \max_i \{f_i(x)\}$ f_i are convex $\Rightarrow g$ is convex

proof

$f(\theta y + (1-\theta)x) = \max_i \{f_i(\theta y + (1-\theta)x)\}$ i^* the argmax of this

$f_{i^*}(\theta y + (1-\theta)x) \overset{f_{i^*} \text{ is convex}}{\leq} \theta f_{i^*}(y) + (1-\theta) f_{i^*}(x) \leq$

$\leq \theta \max_i f_{i^*}(y) + (1-\theta) \max_i f_{i^*}(x) = \theta f(y) + (1-\theta) f(x)$

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

$\left.\begin{array}{l} g \text{ is convex} \\ h \text{ is convex and non-decreasing} \end{array}\right\} \Rightarrow f \text{ is convex}$

- restriction on lines:

$g_{xy}(t) = f(x + ty) \quad \forall\, x, y \in \operatorname{dom} f$

f is convex iff g_{xy} is convex $\forall\, x, y$