Assignment 1

Convex Optimization SS 2018

Patrick Knöbelreiter, knoebelreiter@icg.tugraz.at

19.03.2018

Submission: Upload your solution as a PDF to the TeachCenter. You can also use your scanned hand-written manuscript.

Deadline: April 16th, 2018 at 23:55h.

Convex Sets (13 points)

1. Let $C \subseteq \mathbb{R}^n$ be a convex set, $x_1, \ldots, x_k \in C$ and let $\theta_1, \ldots, \theta_k \in \mathbb{R}$ be factors that satisfy $\theta_i \geq 0$ and sum up to one $\sum_{i=1}^k \theta_i = 1$. Show that any combination of points $\theta_1 x_1 + \cdots + \theta_k x_k$ is also element of the set C. Hint: Use induction on the definition of convexity.

2. Calculate the distance between two parallel hyperplanes $\{x \in \mathbb{R}^n \,|\, a^T x = b_1\}$ and $\{x \in \mathbb{R}^n \,|\, a^T x = b_2\}$. You can derive this geometrically. Draw a picture!

3. Prove or disprove the convexity of the following sets:
   a) The set $\{x \in \mathbb{R} \,|\, x \geq 1/2, x \leq -1/2\}$.
   b) The set of points closer to one set than another, i.e., $\{x \,|\, \mathrm{dist}(x, S) \leq \mathrm{dist}(x, T)\}$ with $S, T \subseteq \mathbb{R}^n$ (not necessarily convex) and $\mathrm{dist}(x, S) = \inf\{\|x - z\|_2 \,|\, z \in S\}$.
   c) A *hyperrectangle*: $\{x \in \mathbb{R}^n \,|\, a_i \leq x_i \leq b_i, i = 1, \ldots, n\}$.
   d) A *wedge*, i.e., a set of form $\{x \in \mathbb{R}^n \,|\, a_1^T x \leq b_1, a_2^T x \leq b_2\}$. What happens if $b_1 = b_2 = 0$?

4. Let $x_0, x_1, \ldots, x_k \in \mathbb{R}^n$. The set of points closer to a given point x_0 than the other points $x_1, \ldots, x_k$

$$\{x \in \mathbb{R} \,|\, \|x - x_0\|_2 \leq \|x - x_i\|_2, \quad i = 1, \ldots, k\} \tag{1}$$

   is called a *Voronoi region* around x_0 with respect to the points $x_1, \ldots, x_k$.
   a) Show that (1) is a convex set.
   b) Equation (1) can be expressed in terms of a polyhedron. Rewrite (1) in the form $\{x \,|\, Ax \preceq b\}$.
   c) Consider the case $k = 1$. Show that the set of all points closer to x_0 than to x_1 is a halfspace, $x_0 \neq x_1$. *Hint:* Recall the definition of a halfspace $\{x \,|\, a^T x \leq b\}$. Draw a picture!

Convex Functions (12 points)

1. Prove or disprove the convexity/concavity of the geometric mean

$$f(x) = \left(\prod_{i=1}^k x_i\right)^{\frac{1}{k}}$$

   on $\mathbb{R}^n_{++}$.

2. Let $f : \mathbb{R}^n \to [-\infty, +\infty]$. Show that the perspective of f defined via

$$g(x, y) = yf(x/y), \ x \in \mathbb{R}^n, \ y \in \mathbb{R}_+$$

is convex if and only if $f(x)$ is a convex function.

3. Prove or disprove the convexity/concavity of the following functions:
   a) $f(x) = \log(x)$ for $x \in \mathbb{R}_{++}$
   b) $f(x) = |a|x| - x|^2$ for $a, x \in \mathbb{R}$
   c) $f(x_1, x_2, x_3) = -e^{(-x_1+x_2-2x_3)^2}$ for $x \in \mathbb{R}^3$
   d) $f(x) = \ln(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$ for $x \in \mathbb{R}^n$

Convex Sets

1) Proof: by induction:

$n=1$: $\theta_1 \cdot x_1 \overset{!}{\in} C \qquad \sum_i \theta_i = 1$

$1 \cdot x_1 \in C$ ✓

$n=2$: $\theta_1 x_1 + \theta_2 x_2 \overset{!}{\in} C \qquad \sum_i \theta_i = 1$

$\theta_1 x_1 + (1-\theta_1) x_2 \in C$ ✓

$n=k$: $\theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k \in C$

Suppose, we know have $k-1$ elements in a set

$\lambda = \sum_{i=1}^{k-1} \theta_i$ Therefore, we need to introduce λ.

$\theta_i' = \frac{\theta_i}{\lambda}$

$p := \sum_i^{k-1} \theta_i' x_i \in C$ Point p is still in the convex set.

Thus, by convexity of C:

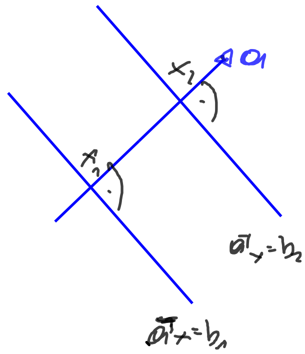
$\lambda \cdot p + (1-\lambda) x_k \in C$

$= \cancel{\lambda} \cdot \sum_{i=1}^{k-1} \frac{\theta_i}{\cancel{\lambda}} x_i + \theta_k x_k$

$= \sum_{i=1}^{k} \theta_i x_i \quad \blacksquare$

2) Distance between 2 hyperplanes: $\{x \in \mathbb{R}^n \mid a^T x = b_1\}$
$\{x \in \mathbb{R}^n \mid a^T x = b_2\}$

If I draw a line orthonormal through one hyperplane, then it will also intersect the 2nd hyperplane orthonormal.
→ Distance is minimal, if the line intersects orthonormal.



$a^T x_i = b_i \mid a \quad \{i = 1..2\}$
$a^T a x_i = b_i a \quad \mid : a^T a \Leftrightarrow \|a\|_2^2$
$x_i = \frac{b_i}{\|a\|_2^2} a$

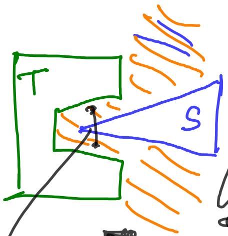
$$\boxed{\|x_1 - x_2\|_2 = \frac{|b_1 - b_2|}{\|a\|_2}}$$

3) a) set $\{x \in \mathbb{R} \mid x \geq \frac{1}{2}, x \leq -\frac{1}{2}\}$

Proof: by definition → convex

Answer: An empty set is a convex set.

b) $\{x \mid dist(x, S) \leq dist(x, T)\}$ with $S, T \subseteq \mathbb{R}^n$.



$dist(x \mid S) = \inf\{\|x - z\|_2 \mid z \in S\}$

Geometric Proof: counter example !

Concave !

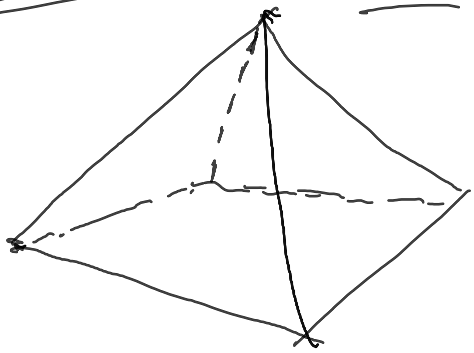
Analytical Proof:

$\{x \mid dist(x, S) \leq dist(x, T)\}$ with $S = \{-1, 1\}$
and $T = \{0\}$
$\Rightarrow$ Is not convex $\forall x$, e.g. $x \in \mathbb{R} : x \leq -\frac{1}{2}$

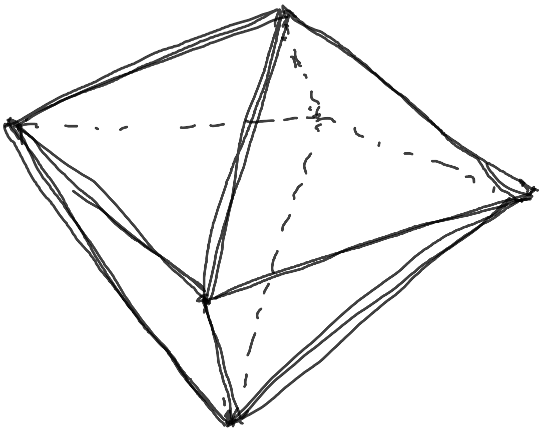
c) hyper rectangle $\{x \in \mathbb{R}^n \mid a_i \le x_i \le b_i, i \ldots n\}$

- Rectangle for higher dimensions.
- Cartesian product of intervalls.

$n > 2$



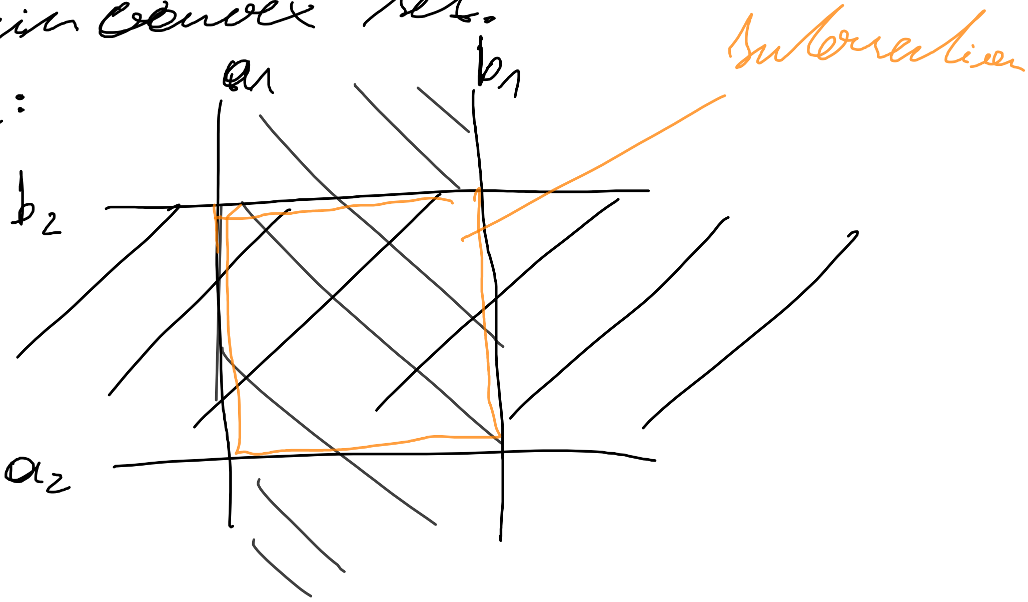
or:



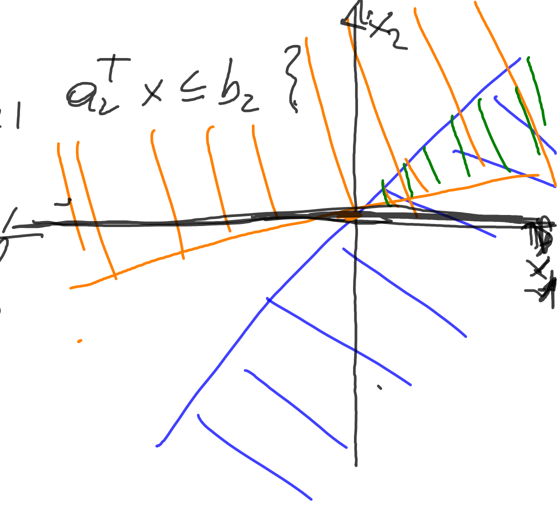
Rectangle is a convex set by definition, whereas a rectangle of dimensions $n>2$ is called hyperrectangle

Halfspaces are convex by definition. If we intersect halfspaces, the intersection is again convex set.

2D case:



d) wedge: $\{x \in \mathbb{R}^n \mid a_1^T x \le b_1,\ a_2^T x \le b_2\}$
What happens $b_1 = b_2 = 0$?



- A wedge is is an intersection of 2 halfspaces ⟶ convex.

- $b_1 = b_2 = 0 \implies$ cone, that the intersection is at the origin.

4) Voronoi $\quad \{x \in \mathbb{R} \mid \|x - x_0\|_2 \le \|x - x_i\|_2,\ i = 1 \dots k\}$ (1)

$x_0, \dots, x_k \in \mathbb{R}^n$ closer to x_0 than $x_1, \dots, x_k$

a) At (b) we showed it can be a Polyhedron, which is convex by definition.

b) Rewrite in the form of $\{x \mid Ax \le b\}$

Rule: $\|x\|_2 = \sqrt{x^T x}$

$$\|x - x_0\|_2 \le \|x - x_i\|_2$$

$$\iff \sqrt{(x-x_0)^T (x-x_0)} \le \sqrt{(x-x_i)^T (x-x_i)}$$

$$\iff (x-x_0)^T (x-x_0) \le (x-x_i)^T (x-x_i)$$

umformen & kürzen

$$\iff \cancel{x^T x} - x^T x_0 - x_0^T x + x_0^T x_0 \le \cancel{x^T x} - x^T x_i - x_i^T x + x_i^T x_i$$

$$\iff -2x^T x_0 + x_0^T x_0 \le -2x^T x_i + x_i^T x_i$$

$$\iff -2x^T x_0 + 2x^T x_i \le x_i^T x_i - x_0^T x_0$$

$$\iff \underbrace{2x^T (x_i - x_0)}_{A} \le \underbrace{x_i^T x_i - x_0^T x_0}_{b}$$

$$A = 2 \begin{bmatrix} x_1 - x_0 \\ x_2 - x_0 \\ \vdots \\ x_k - x_0 \end{bmatrix} \qquad b = \begin{bmatrix} x_1^T x_1 - x_0^T x_0 \\ x_2^T x_2 - x_0^T x_0 \\ \vdots \\ x_k^T x_k - x_0^T x_0 \end{bmatrix}$$

c) Case: $k=1$. Set of all points closer to x_0 than to x_1 is a halfspace, $x_0 \neq x_1$

halfspace: $\{x \mid a^T x \leq b\}$

$a^T \in \mathbb{R}^{k\times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^k$

$$\begin{array}{c|c} & n \\ \hline k & \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \end{array} \cdot \overset{n}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} \leq \overset{k}{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}$$

Any point $x_0 \in Ax < b$ we can choose the mirror image of x_0 w.r.t. $\underbrace{x_i = x_0 + \lambda a_i}_{\text{Geradengleichung}}$

λ must be chosen, so that the distance of x_i to the hyperplane fulfills: $a_i^T x = b_i \iff$ distance of x_0 to the hyperplane

$$b_i - a_i^T x_0 = a_i^T x_i - b_i$$

$$b_i - a_i^T x_0 = a_i^T (x_0 + \lambda a_i) - b_i$$

$$b_i - a_i^T x_0 = a_i^T x_0 + a_i^T \lambda a_i - b_i$$

$$2b_i - a_i^T x_0 = a_i^T x_0 + a_i^T a_i \lambda$$

$$2(b_i - a_i^T x_0) = a_i^T a_i \lambda$$

$$\lambda = \frac{2(b_i - a_i^T x_0)}{\|a_i\|_2^2}$$

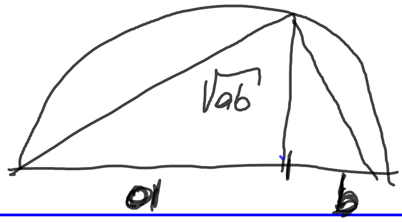
Set λ in the begining term:

$$x_i = x_0 + \frac{2(b_i - a_i^T x_0)}{\|a_i\|_2^2} a_i$$

Convex functions

1) Proof by concavity of geometric mean

Properties: • geometric mean $\leq$ arithmetic mean
• ratio



Prove: 2nd derivative of $f(x)'' > 0$

Hessian $\nabla^2 f(x)$ is given:

$$\frac{\partial^2}{\partial^2 x_k}\left(\prod_{i=1}^{k} x_i\right)^{\frac{1}{k}} = \frac{1}{k}\left(\frac{1}{k}-1\right)\left(\prod_{i=1}^{k} x_i\right)^{\frac{1}{k}-2}\cdot\left(\prod_{\substack{i=1\\ i\neq j}}^{k} x_j\right)^2$$

$$= \frac{1}{k^2}(1-k)\left(\prod_{i=1}^{k} x_i\right)^{\frac{1}{k}}\cdot\frac{1}{x_k^2}$$

$$\frac{\partial}{\partial x_k} = \frac{1}{k}\left(\prod x_i\right)^{\frac{1}{k}}\cdot\frac{1}{x_k}$$

$$\frac{\partial}{\partial x_e \partial x_e} = \frac{1}{k^2}\left(\prod x_i\right)^{\frac{1}{k}}\cdot\frac{1}{x_k x_e}$$

$$H = \begin{bmatrix} \frac{\partial^2}{\partial^2 x_e} & \frac{\partial^2}{\partial x_e \partial x_e} \\ \frac{\partial^2}{\partial x_e \partial x_e} & \frac{\partial^2}{\partial x_k \partial x_e} \end{bmatrix} \Rightarrow v^T \nabla^2 f(x) v \leq 0$$

$$H = -\frac{\prod x_i^{\frac{1}{k}}}{k^2}\cdot\left(n\cdot\begin{bmatrix} \frac{1}{x_1^2} & & \\ & \frac{1}{x_2^2} & \\ & & \ddots & \\ & & & \frac{1}{x_k^2}\end{bmatrix} - \begin{bmatrix}\frac{1}{x_i} \\ \vdots \\ \frac{1}{x_k}\end{bmatrix}\boxed{\begin{bmatrix}\frac{1}{x_i} & \cdots & \frac{1}{x_k}\end{bmatrix}}\right)$$

$$\left(\begin{matrix} \frac{1}{x_i^2} & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \frac{1}{x_k^2}\end{matrix}\right)$$

To show that, we use $v^T H v \leq 0$, $\forall v$.

$$v^T H v = \frac{-\prod x_i^{\frac{1}{k}}}{k^2} \left(n \sum_{i=1}^{n} 1^2 \left(\frac{v_i}{x_i}\right)^2 - \left(\sum_{i=1}^{n} 1 \cdot \frac{v_i}{x_i} \right)^2 \right) \leq 0$$

Hölder's Inequality: $\left(\sum^{n} a_k b_k \right)^2 \leq \left(\sum_{k=1}^{n} a_k^2 \right) \left(\sum_{k=1}^{n} b_k^2 \right).$

$a = 1$

$b = \frac{v}{x}$

$\Rightarrow$ This $f(x)$ is concave.

2) $f : \mathbb{R}^n \to [-\infty, +\infty]$ // Note: $\mathbb{R}$ is usually exclusive "∞" by definition.

$f(x)$ is convex.

$$g(x) = y f\left(\frac{x}{y}\right) \quad x \in \mathbb{R}^n, y \in \mathbb{R}_{++}$$

Perspective of a function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, with domain $\mathbb{R}_{++}$ as $P(z,t) = z/t$

It scales or normalizes vectors so the last component is one, and removes the last component.

e.g.: Pinhole camera from 3D to 2D.

The dom g is a convex set for $(x,t), (y,s) \in \text{dom } g$, and $\theta \in [0,1]$, then

$$g(\theta x + (1-y), \theta t + (1-\theta)s) \leq \theta g(x,t) + (1-\theta) g(y,s).$$

~~Proof via epi:~~

~~For $t > 0$: we have~~

~~$(x,t,s) \in \text{epi } g \iff t f(x/t) \leq s \quad |:t$~~

~~$\iff f(x/t) \leq s/t$~~

~~$\iff (x/t, s/t) \in \text{epi } f$~~

~~epi g is the inverse mapping of epi f !~~

~~The epi g is convex, so the function g is convex.~~

~~Example $f(x) = x^T x$ on $\mathbb{R}^n$, convex in (x,t) for $t > 0$.~~

~~$$g(x,t) = t (x/t)^T (x/t) = \frac{x^T x}{t}$$, is convex for $t > 0$.~~

Ist nicht bewertet, bitte!

Proof via Jensen's inequality:

Suppose $\frac{s}{t} > 0$, $\frac{x}{t} \in \operatorname{dom} f$, $\frac{y}{s} \in \operatorname{dom} f$:

$$g(\theta x + (1-\theta) y, \theta t + (1-\theta) s)$$

$$= (\theta t + (1-\theta) s) f\left(\frac{(\theta x + (1-\theta) y)}{\theta t + (1-\theta) s}\right)$$

$$= (\theta t + 1-\theta s) f\left(\frac{(\theta t(\frac{x}{t}) + (1-\theta) s (\frac{y}{s}))}{\theta t + (1-\theta) s}\right)$$

$$\leq \theta t f\left(\frac{x}{t}\right) + (1-\theta) s f\left(\frac{y}{s}\right) \quad \square$$

We showed that $(x,t), (y,s) \in \operatorname{dom} g$ and g is convex set.

3) a) $f(x) = \log(x)$ for $x \in \mathbb{R}_{++}$

<u>Proof</u>: 2nd derivative must be > 0 → convex

$$f'(x) = \frac{1}{x} = x^{-1}$$

$$f''(x) = -1 \cdot x^{-1-1} = -x^{-2} = -\frac{1}{x^2} \dots \text{ is negative}$$

for all $x \in \mathbb{R}_{++}$ → <u>concave</u>!

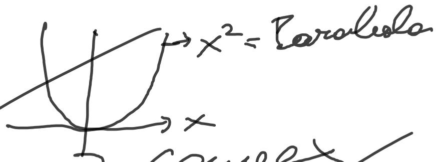
b) $f(x) = |a|x| - x|^2$ for $a, x \in \mathbb{R}$

$$f(x) = \begin{cases} |ax - x|^2 \Rightarrow |x(a-1)|^2 \Rightarrow x^2(a-1)^2 & x > 0 \\ |-ax - x|^2 \Rightarrow |x(-a-1)|^2 \Rightarrow x^2(-a-1)^2 & x < 0 \\ 0 & \text{else} \end{cases}$$

$$f'(x) = \begin{cases} 2x(a-1)^2 & x > 0 \\ 2x(-a-1)^2 & x < 0 \end{cases}$$

$$f''(x) = \begin{cases} 2(a-1)^2 & x > 0 \\ 2(-a-1)^2 & x < 0 \\ 0 & x = 0 \end{cases}$$

is convex, because $f''(x)$ is always ≥ 0.



<u>Case $a \in \mathbb{R} \setminus \{-1, 1\}$:</u>

Both cases are Parabola → <u>convex</u>.

<u>Case $a \in \{-1, 1\}$:</u>

One part is 0 and the other part is a parabola. → f is <u>convex</u>.

c) $f(x_1, x_2, x_3) = -e^{(x_1 + x_2 - 2x_3)^2}$ for $x \in \mathbb{R}^3$.

<u>Proof</u>:

$$\nabla_x^2 f = \begin{bmatrix} \partial x_1 x_1 & \partial x_1 x_2 & \partial x_1 x_3 \\ & \partial x_2 x_2 & \partial x_2 x_3 \\ & & \partial x_3 x_3 \end{bmatrix} \geq 0 \text{ to be convex!}$$

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \frac{\partial f(x)}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2(-x_1+x_2-2x_3)\cdot e^{(-x_1+x_2-2x_3)^2} \\ -2(-x_1+x_2-2x_3)\cdot e^{(-x_1+x_2-2x_3)^2} \\ 4(-x_1+x_2-2x_3)\cdot e^{(-x_1+x_2-2x_3)^2} \end{bmatrix}$$

$$f(x_1,x_2,x_3) = -e^{\underbrace{(-x_1+x_2-2x_3)}_{\vartheta}^2} \qquad \| \; \vartheta = -x_1+x_2-2x_3$$

Auxilary calculation (AC):

$$\frac{\partial}{\partial x_1}\left(-e^{\vartheta^2}\right) = \frac{\partial}{\partial x_1}\vartheta^2\cdot\left(-e^{\vartheta^2}\right) = -2\vartheta(-1)e^{\vartheta^2} = \underline{\underline{2\vartheta e^{\vartheta^2}}}$$

$$\frac{\partial}{\partial x_2}\left(-e^{\vartheta^2}\right) = \quad = -2\vartheta(1)e^{\vartheta^2} = \underline{\underline{-2\vartheta e^{\vartheta^2}}}$$

$$\frac{\partial}{\partial x_3}\left(-e^{\vartheta^2}\right) = \quad = -2\vartheta(-2)e^{\vartheta^2} = \underline{\underline{4\vartheta e^{\vartheta^2}}}$$

$$\frac{\partial^2 f(x)}{\partial x^2} = H = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_3} \\ & \frac{\partial^2 f(x)}{\partial x_2^2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_3} \\ & & \frac{\partial^2 f(x)}{\partial x_3 \partial x_3} \end{bmatrix}$$

Hessian is symmetric!

AC:

$$\frac{\partial^2 f(x)}{\partial x_1^2} = 2\cdot(-1)\cdot e^{\vartheta^2} + 2\vartheta\cdot(2\vartheta(-1))\cdot e^{\vartheta^2} = -2e^{\vartheta^2} - 4\vartheta^2 e^{\vartheta^2}$$

$$\frac{\partial^2 f(x)}{\partial x_1 \partial x_2} = 2\cdot(1)\cdot e^{\vartheta^2} + 2\vartheta\cdot(2\vartheta\cdot 1 e^{\vartheta^2}) = 2e^{\vartheta^2} + 4\vartheta^2 e^{\vartheta^2}$$

$$\frac{\partial^2 f(x)}{\partial x_1 \partial x_3} = 2(-2)\cdot e^{\vartheta^2} + 2\vartheta\cdot(2\vartheta\cdot(-2)\cdot e^{\vartheta^2}) = -4e^{\vartheta^2} - 8\vartheta^2 e^{\vartheta^2}$$

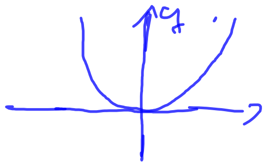
$$\frac{\partial^2 f(x)}{\partial x_3 \partial x_3} = 4\cdot(-2)\, e^{\vartheta^2} + 4\vartheta(2\vartheta\,(-2)\cdot e^{\vartheta^2}) = -8e^{\vartheta^2} - 16\vartheta^2 e^{\vartheta^2}$$

Man könnte Min/Max berechnen indem an $f'(x)=0$ setzt und nach x auflöst $\Rightarrow$ viel zu kompliziert.

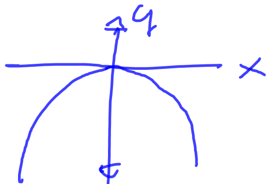
Besser: grafisch!

e^{ϑ^2} is always positive!

ϑ^2 is a parabel:



$| -\vartheta^2$:



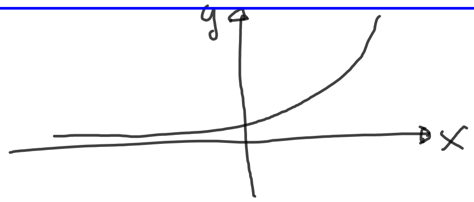
Result:

$$H = \begin{bmatrix} - & + & - \\ & - & + \\ & & - \end{bmatrix}$$

Concave! because not all entries are convex!

Notation: + convex, − concave

d) $f(x) = \ln(e^{x_1} + e^{x_2} + \dots + e^{x_n})$ for $x \in \mathbb{R}^n$

Def: e^x :  is convex.

Rule: $C_1 + C_2 = C \Rightarrow e^{x_1} + \dots + e^{x_n} = e^x \in C$

Simplify: $f(x) = \ln(e^x) = x$

Properties:
- $f(x) = \exp(ax)$ on $\mathbb{R}$, $a \in \mathbb{R}$ $\Rightarrow$ convex
- $f(x) = \log(x)$ on $\mathbb{R}^{++}$ $\Rightarrow$ concave

Proof by: $v^T H v \geq 0$ $v \neq 0$, $v \in \mathbb{R}^n$

$$\frac{\partial f(x)}{\partial x_i} = \frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}} \qquad i = 1, \dots, n$$

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \begin{cases} -\dfrac{e^{x_i} e^{x_j}}{\sum_{k=1}^{n} e^{x_k}} & i \neq j \\ -\dfrac{e^{x_i} e^{x_j}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2} + \dfrac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}} & i = j. \end{cases}$$

$$H = \begin{bmatrix} -\frac{e^{x_1} e^{x_1}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2} + \frac{e^{x_1}}{\sum_{k=1}^{n} e^{x_k}} & \cdots & -\frac{e^{x_1} e^{x_n}}{\sum_{k=1}^{n} e^{x_k}} \\ \vdots & \ddots & \vdots \\ -\frac{e^{x_n} e^{x_1}}{\sum_{k=1}^{n} e^{x_k}} & \cdots & -\frac{e^{x_n} e^{x_n}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2} + \frac{e^{x_n}}{\sum_{k=1}^{n} e^{x_k}} \end{bmatrix}.$$

$$H = \mathrm{diag}(u) - uu^T \qquad u = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

Now, we have to prove the positive definiteness.

$$v^T H v = \sum_{i=1}^{n} u_i v_i^2 - (v^T u)^2$$

Cauchy–Schwarz: of

$$s_i = \sqrt{u_i}\, v_i, \quad t_i = \sqrt{u_i} \quad i = 1..n$$

$$(v^T u)^2 = (s^T t)^2 \overset{(*)}{\leq} \|s\|^2 \|t\|^2 =$$

$$= \left(\sum_{i=1}^{n} u_i v_i^2\right) \underbrace{\left(\sum_{i=1}^{n} u_i\right)}_{1,\ \text{because } \frac{\sum_{i=1}^{n} e^{x_i}}{\sum_{k=1}^{n} e^{x_k}} = 1} = \sum_{i=1}^{n} u_i v_i^2 \qquad \square$$

H is <u>positive semidefinite</u>, since $v^T H v \geq 0$ and the inequality (*) holds for every $v \in \mathbb{R}^n$.

Thus $f(x)$ is <u>convex</u>.