Problem 1

For each of the following prove whether it is convex, concave or neither.

a). $f(x,y) = \min\{\sqrt{x}, \sqrt{xy}\}$ with $\operatorname{dom} f \in \mathbb{R}^2_{++}$.

Pointwise maximum: If $f_1, \dots, f_m$ are convex then $\max\{f_1(x) \dots f_m(x)\}$ is convex. From pg. 81 in the textbook we see this extends to pointwise infimum for concave functions.

Clearly, $\sqrt{x}$ is concave on $\operatorname{dom} f$.

Now, want to show $\sqrt{xy}$ is concave. Take (x_1, y_2) and (x_2, y_2) then

$$f(\theta(x_1,y_1) + (1-\theta)(x_2,y_2)) \geq \theta f((x_1,y_1)) + (1-\theta) f((x_2,y_2))$$

by squaring

$$(\theta x_1 + (1-\theta)x_2)(\theta y_1 + (1-\theta)y_2) \geq (\theta\sqrt{x_1y_1} + (1-\theta)(\sqrt{x_2y_2}))^2$$

which holds by Cauchy-Schwarz.

Hence f is concave because it is the pointwise minimum of concave functions.

b) $f(x,y) = \frac{1}{xy}$ with $\text{dom} f = \{(x,y) \in \mathbb{R}^2_{++}\}$.

Take the Hessian:

$$\nabla^2 f(x,y) = \begin{bmatrix} \frac{2}{x^3 y} & \frac{1}{x^2y^2} \\ \frac{1}{x^2y^2} & \frac{2}{xy^3} \end{bmatrix} = \frac{1}{xy}\begin{bmatrix} \frac{2}{x^2} & \frac{1}{xy} \\ \frac{1}{xy} & \frac{2}{y^2} \end{bmatrix}$$

For $x, y \geq 0$ and $a, b \in \mathbb{R}$ we have

$$2g(a,b) = (a \ b)\begin{bmatrix} \frac{2}{x^2} & \frac{1}{xy} \\ \frac{1}{xy} & \frac{2}{y^2} \end{bmatrix}\begin{pmatrix} a \\ b \end{pmatrix} =$$

$$= 2\left(\frac{a^2}{x^2} + \frac{b^2}{y^2} + \frac{ab}{xy}\right) \geq 0.$$

let $v = ax^{-1}$ and $u = by^{-1}$ then

$$g(a,b) = v^2 + u^2 + vu.$$

WLOG, assume $v < 0 < u$ then $-v \leq u$ gives $-vu \leq u^2$ or $u^2 + vu \leq 0$. Similarily, $v^2 + vu > 0$.

Therefore, we have $g(a,b) \geq 0$ for all $a, b \in \mathbb{R}$.

Hence, the Hessian is PSD and thus $\frac{1}{xy}$ is convex.

c) $f(x,y) = x\log y$ with $\operatorname{dom} f = \{(x,y) \in \mathbb{R}^2 \mid x \geq 0, y > 0\}$.

Take the Hessian

$$\nabla^2 f(x,y) = \begin{bmatrix} 0 & \frac{1}{y} \\ \frac{1}{y} & -\frac{x}{y^2} \end{bmatrix}$$

Find eigenvalues:

$$\begin{bmatrix} -\lambda & \frac{1}{y} \\ \frac{1}{y} & -\lambda - \frac{x}{y^2} \end{bmatrix} = (-\lambda)\cdot\left(-\lambda - \frac{x}{y^2}\right) - \left(\frac{1}{y}\right)\left(\frac{1}{y}\right)$$

$$= \lambda^2 + \frac{\lambda x}{y} - \frac{1}{y^2}$$

$$\det \nabla^2 f(x,y) = \frac{\lambda^2 y^2 + \lambda x - 1}{y^2}.$$

Then,

$$\lambda_1 = -\frac{x + \sqrt{x^2 + 4y^2}}{2y^2}, \quad \lambda_2 = \frac{-x + \sqrt{x^2 + 4y^2}}{2y^2}$$

plug $(1,1)$

$$\lambda_1 = -\frac{1 + \sqrt{5}}{2}, \quad \lambda_2 = \frac{-1 + \sqrt{5}}{2}$$

Since we have positive and negative eigenvalues the Hessian at $(1,1)$ is a saddle point.

Hence f is neither concave or convex.

Problem 2

a) Let f^* denote the conjugate function. Show that ∇f and ∇f^* are inverse mappings, i.e., $(\nabla f)^{-1}(x) = \nabla f^*(x)$ for any x.

The convex conjugate is given by

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

Let's say that the sup is achived at x^* (we can say this because we know there is one global optimum to the sup) so,

$$f^*(y) = \langle x^*, y \rangle - f(x^*)$$

$$x^* - \nabla f(y) = 0$$

$$x^* = \nabla f^*(y)$$

Also, $\nabla_y (y^T x - f(x^*))$

$$= \nabla_y (y^T (x^*) - f(x^*))$$

$$= \nabla_y (y^T (x^*)) - \nabla_y \cancel{f(x^*)}^{\,0}$$

$$= x^*$$

$$= \nabla_y f^{-1}(y)$$

Showing that $(\nabla f)^{-1}(y) = x^* = \nabla f^*(y)$.

b) Prove that if f is $\frac{1}{L}$-strongly convex, then its conjugate f^* is L-smooth.

want to show $\|f^*(y) - f^*(x)\|_2 \leq L\|y-x\|_2$

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|y-x\|_2^2$$

$$f(\nabla f^*(y)) \geq f(\nabla f^*(x)) + \underbrace{\nabla f(\nabla f^*(x))}_{x}(\nabla f^*(y) - \nabla f^*(x)) + \frac{1}{2L}\|y-x\|_2^2$$

$$f(\nabla f^*(y)) \geq f(\nabla f^*(x)) + x^T(\nabla f^*(y) - \nabla f^*(x)) + \frac{1}{2L}\|\nabla f^*(y) - \nabla f^*(x)\|_2^2$$

$$f(\nabla f^*(x)) \geq f(\nabla f^*(y)) + y^T(\nabla f^*(x) - \nabla f^*(y)) + \frac{1}{2L}\|\nabla f^*(y) - \nabla f^*(x)\|_2^2$$

$$\cancel{f(\nabla f^*(x))} + \cancel{f(\nabla f^*(y))} \geq \cancel{f(\nabla f^*(x))} + \cancel{f(\nabla f^*(y))} + x^T(\nabla f^*(y) - \nabla f^*(x)) + y^T(\nabla f^*(x) - \nabla f^*(y))$$
$$+ \frac{1}{L}\|\nabla f^*(y) - \nabla f^*(x)\|_2^2$$

$$\Rightarrow (x-y)^T(\nabla f^*(y) - \nabla^* f(x)) \geq \frac{1}{L}\|\nabla f^*(y) - \nabla f^*(x)\|_2^2$$

cauchy-schwarz $\longrightarrow$ $\|x-y\|_2 \, \|\nabla f^*(y) - \nabla^* f(x)\|_2 \geq \frac{1}{L}\|f^*(y) - f^*(x)\|_2^2$

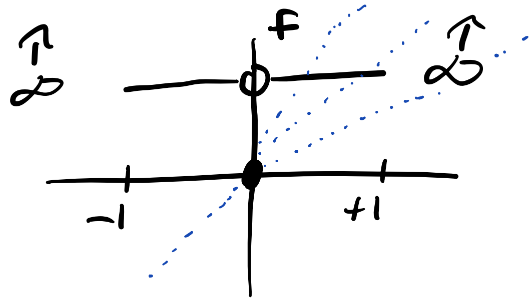
$\blacksquare$

Problem 3

Let card(x) denote the number of nonzero entries in $x \in \mathbb{R}^n$.

$$f(x) = \begin{cases} \text{card}(x) & \|x\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

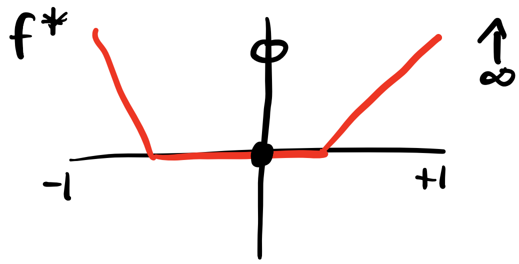
a) Find f^*, the conjugate of f. Give a simple sketch of f and f^*.



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

$$\text{card}(x) = \sum_{i=1}^{n} \mathbb{1}(x_i)$$

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



$$= \sup_{x \in \text{dom} f} \sum_i y_i x_i - \sum_i \mathbb{1}(x_i)$$

$$= \sup_{x_1, \dots, x_m} \sum_{i=1}^{n} (y_i x_i - \mathbb{1}(x_i))$$

$$f^* = \sum_{i=1}^{n} \sup (y_i x_i - \mathbb{1}(x_i))$$

we can make the last equality because each x_i is independent

b) Find the conjugate function of f^*, i.e., f^{**}.

$$f^{**}(x) = \sup_{z \in \mathbb{R}^n} x^T z - f^*(z)$$

$$= \sup_{z \in \mathbb{R}^n} \sum_{i=1}^n x_i z_i - \sum_{i=1}^n \max\{|z_i| - 1, 0\}$$

$$= \underbrace{\sup_{z \in \mathbb{R}^n} \sum_{i=1}^n}_{*} x_i z_i - \max\{|z_i| - 1, 0\}$$

$$= \sum_{i=1}^n \sup_{z_i \in \mathbb{R}} x_i z_i - \max\{|z_i| - 1, 0\}$$

$$= \sum_{i=1}^n \sup_{z_i \in \mathbb{R}} \left(x_i z_i - \begin{cases} |z_i| - 1 & \text{if } |z_i| > 1 \\ 0 & \text{otherwise} \end{cases} \right)$$

* works since they are independent at each i.

If $|z_i| < 1$ then: $\sup_{|z_i| \le 1} x_i z_i = |x_i|$

if $|z_i| > 1$ then: $\sup_{|z_i| \ge 1} x_i z_i - |z_i| + 1$

↳ if $|x_i| > 1$ then its unbounded.

↳ if $|x_i| < 1$ then $\sup_{|z_i| \ge 1} x_i z_i - |z_i| + 1$

$z_i < 0 \Rightarrow \sup_{z_i \le -1} z_i(x+1) + 1$, $z_i > 0 \Rightarrow \sup_{z_i \ge 1} x_i z_i - z_i + 1$

$z = -1$

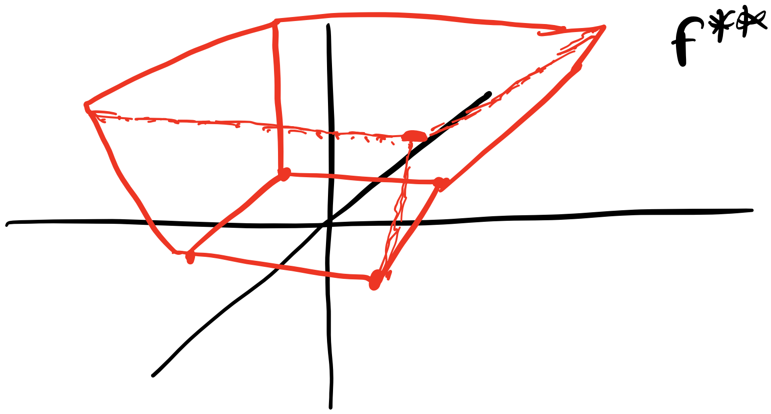
$= \sup z_i(x_i - 1) + 1$

$\Rightarrow z = 1$.

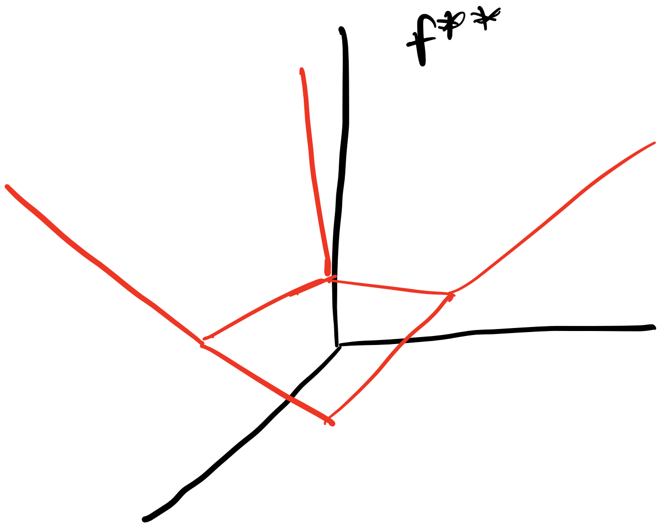
In all cases we get $|x_i|$ from the supremum.

Therefore, f^{**} is the l_1 norm $\|x\|_1 = \sum_{i=1}^{n} |x_i|$.

sketch:



also,



* The argument for showing these sketches in $n=1$ is because f, f^*, f^{**} are independent and are just a summation and independent at each $i = 1, \dots, n$.

Problem 4

Consider the following functions of a variable $x \in \mathbb{R}^n$:

$$f_0(x) = \max\{A_0 x + c_0, 0\}$$
$$f_i(x) = \max\{A_i x + B_i f_{i-1}(x) + c_i, 0\}$$
for $i = 1, 2, \dots M$.

This function is similar to that of a ReLU function, i.e, $f(x) = \max\{x, 0\}$ which we know is convex and nondecreasing. These are interestingly used in convex NNs like in the paper "Input Convex Neural Networks" by Brandon Amos, et al.

We show that $f_M(x)$ is convex by induction

base case $i = 0$,

$$f_0(x) = \max\{A_0 x + c_0, 0\}$$

- The function $A_0 x + c_0$ is affine and hence convex.
- pointwise maximum of convex function is convex.
- Thus $f_0(x)$ is convex.

inductive step. Take $f_i(x) = \max\{A_i x + B_i f_{i-1}(x) + c_i, 0\}$

where B_i is element-wise nonnegative

- The function $A_i x + B_i f_{i-1}(x) + c_i$ is affine and hence convex
- pointwise maximum of an affine function is convex
- Thus $f_i(x)$ for $i = 1, 2, \ldots, M$ is convex.

Hence we have proven that $f_M(x)$ is convex.

Problem 5

For each of the following functions, prove whether it is convex, concave, or neither.

a) $f(x,y) = \sqrt{x \min\{y,1\}}$ with $\text{dom} f = \mathbb{R}^2_{++}$.

$f(x,y)$ can be written as

$$= \begin{cases} \sqrt{x}, & y > 1 \\ \sqrt{xy}, & y \leq 1 \end{cases}$$

$$= \min\{\sqrt{x}, \sqrt{xy}\}.$$

we proved in problem 1 that this is concave.

$\rightarrow$ min of concave functions is concave.

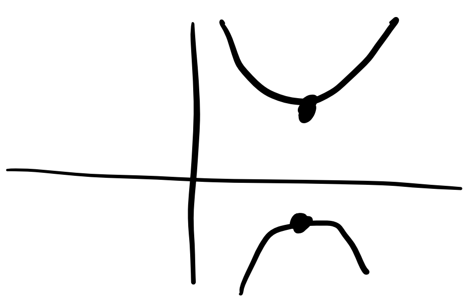
b) The difference between the maximum and minimum value of a polynomial on a given interval, as a function of its coefficients:

$$f(x) = \sup_{t \in [a,b]} p(t) - \inf_{t \in [a,b]} p(t)$$

where $p(t) = x_1 + x_2 t + x_3 t^2 + \dots + x_n t^{n-1}$.

a, b are real constants $a < b$.

using this fact: $\max(f(x)) = -\min(-f(x))$



we will write $f(x)$ as

$$f(x) = \underbrace{\sup_{t\in[a,b]} p(t)}_{\text{convex}} + \underbrace{\sup_{t\in[a,b]} -p(t')}_{\text{convex}}$$

- $p(x,t) = \sum_{k=1}^{n} x_k t^{k-1}$ is linear in x.
- $\sup_t p(t)$ is convex since pointwise maximum of a convex function is convex.
- Sum of convex functions is convex.

Therefore $f(x)$ is convex in x with some $t \in [a,b]$, $a<b$.

c) The function

$$f(x) = \inf_{\alpha > 0} \frac{g(y+\alpha x) - g(y)}{\alpha}$$

g is convex and $y \in \operatorname{dom} g$.

Theorem: if $f(x,y)$ is jointly convex in (x,y) and C is a convex set, then

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

is convex.

Then we need to prove the following $\tilde{g}$ is convex for $f(x)$ to be convex.

$$\tilde{g}(x,\alpha) = \frac{g(y+\alpha x) - g(y)}{\alpha}$$

We will apply perspective to $\tilde{g}$ which perserves convexity.

$$f(x,t) = t\, g\left(\frac{x}{t}\right) \longrightarrow f(x,\alpha) = \frac{1}{\alpha}\, g(\alpha x),\ t = \frac{1}{\alpha}$$

$$\tilde{g} = \frac{h(\alpha x)}{\alpha}$$

$$h = g(y+x) - \cancel{g(y)}$$

h is convex: $h(\theta x_1 + (1+\theta) x_2) \leq \theta h(x_1) + (1-\theta) h(x_2)$

$$h(y + \theta x_1 + (1-\theta) x_2) \leq \theta h(y + x_1) + (1-\theta) h(y + x_2)$$

$$= \theta(y) + (1-\theta) y$$

- we have the form $f(x) = \inf_{\alpha > 0} \tilde{g}(x,\alpha)$
- partial minimization over a perspective function perserves convexity.
- thus $\tilde{g}(x,\alpha)$ is convex.
- Hence f is convex.

d)

$$f(x) = \left(\sum_{i=1}^{n} 1/x_i\right)^{-1} \quad \text{with } \mathbf{dom} f = \mathbb{R}^n_{++}$$

want to show the Hessian is negative semidefinite.

f is known as the harmonic mean:

$$f(x) = \frac{1}{1/x_1 + \ldots + 1/x_n}, \quad \mathbf{dom} f = \mathbb{R}^n_{++}$$

If a function f is a positive, differentiable function and log f is concave it implies that f is concave.

- If log f is concave, then f is concave.

Consider,

$$h(x) = \log f(x) = -\log\left(\left(\sum_{i=1}^{n} x_i\right)^{-1}\right)$$

$$= -\log(1/x_1 + \ldots + 1/x_n)$$

Take the first and second derivative,

$$\frac{\partial h(x)}{\partial x_i} = \frac{1/x_i^2}{1/x_1 + \ldots + 1/x_n}$$

$$\frac{\partial^2 h(x)}{\partial x_i^2} = \frac{-2/x_i^3}{1/x_1 + \ldots + 1/x_n} + \frac{1/x_i^4}{(1/x_1 + \ldots + 1/x_n)^2}$$

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} = \frac{1/(x_i^2 x_j^2)}{(1/x_1 + \cdots + 1/x_n)^2} \quad , i \neq j$$

show that $v^T \nabla^2 h(x) v \geq 0$ for all $v \neq 0$:

$$\left(\sum_{i=1}^{n} v_i/x_i^2\right)^2 < 2\left(\sum_{i=1}^{n} 1/x_i\right)\left(\sum_{i=1}^{n} y_i^2/x_i^3\right)$$

This follows by Cauchy-Schwarz

$$(a^T b)^2 \leq \|a\|_2^2 \, \|b\|_2^2$$

where $a_i = \frac{1}{\sqrt{x_i}}$, $b_i = \frac{y_i}{x_i\sqrt{x_i}}$.

Thus $h(x)$ is negative semi-definite

Hence $\log f$ is concave $\rightarrow$ f is concave on $\mathrm{dom} f = \mathbb{R}_{++}^n$.