Exercises. Convex functions and convex sets.

(64) Show that if $f:\mathbb{R}^d \to \mathbb{R}$ is convex, then it satisfies Jensen's inequality; for all $n \geq 2$:

$$f(a_1x_1 + \dots + a_nx_n) \leq a_1 f(x_1) + \dots + a_n f(x_n)$$

for $a_1, \dots, a_n \in \mathbb{R}$ such that $a_i \geq 0$ for all i, and $\sum_{i=1}^{n} a_i = 1$.

Solution: We use the method of mathematical induction.

For $n=2$, this is the definition of a convex function.

Suppose that $f(b_2x_2 + \dots + b_nx_n) \leq b_2 f(x_2) + \dots + b_n f(x_n)$ holds when $b_i \geq 0$ and $\sum_{i=2}^{n} b_i = 1$.

Now let $a_1, \dots, a_n \in \mathbb{R}$, $a_i \geq 0$ and $\sum_{i=1}^{n} a_i = 1$, and $x_1, \dots, x_n \in \mathbb{R}^d$.

Then

$$f(a_1x_1 + \dots + a_nx_n) = f\left(a_1x_1 + (1-a_1)\left[\frac{a_2}{1-a_1}\cdot x_2 + \dots + \frac{a_n}{1-a_1}\cdot x_n\right]\right)$$

$$\leq a_1 f(x_1) + (1-a_1)\cdot f\left[\left(\frac{a_2}{1-a_1}\right)x_2 + \dots + \left(\frac{a_n}{1-a_1}\right)\cdot x_n\right]$$

Since $\frac{a_i}{1-a_1} \geq 0$ and $\sum_{i=2}^{n} \frac{a_i}{1-a_1} = \frac{1-a_1}{1-a_1} = 1$, we can apply the induction hypothesis and obtain that the above is

$$\leq a_1 f(x_1) + (1-a_1)\left[\left(\frac{a_2}{1-a_1}\right)f(x_2) + \dots + \left(\frac{a_n}{1-a_1}\right)f(x_n)\right]$$

$$= a_1 f(x_1) + a_2 f(x_2) + \dots + a_n f(x_n).$$

(This holds if $a_1 < 1$. If $a_1 = 1$ then $a_2 = \dots = a_n = 0$, and the result is trivial.)

(65) Show that for all $x_1, x_2, \dots, x_n > 0$ in $\mathbb{R}$, we have:

$$\frac{1}{n}(x_1 + \dots + x_n) \geq \sqrt[n]{x_1 \cdots x_n}.$$

Solution. This follows from Jensen's inequality and the fact that $\ln(x)$ is concave:

Use the function $\ln : \mathbb{R}_+ \longrightarrow \mathbb{R}$, then

$$\ln\left(\frac{1}{n}(x_1 + \dots + x_n)\right) \underset{\substack{\ln \text{ is} \\ \text{concave}}}{\geq} \frac{1}{n}(\ln(x_1) + \dots + \ln(x_n))$$

$$= \ln\left(\sqrt[n]{x_1 \dots x_n}\right)$$

Taking exponentials we get the desired result.

(66) Show that the functions $e^x, x^2, x^4, x^6, \dots, x^{2n}$ n natural number are all convex.

Solution We apply the criteria:

$$f \text{ is convex} \iff f''(x) \geq 0 \quad \forall x.$$

For e^x, we have $(e^x)'' = e^x > 0$.

For x^{2n}, we have $(x^{2n})'' = (2n)(2n-1)x^{2n-2} =$

$$= (2n)(2n-1)(x^{n-1})^2 \geq 0.$$

(67) $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = w^T x + b$, $w \in \mathbb{R}^n$, $b \in \mathbb{R}$

is convex and concave. ~~for all~~

Solution. We show that

$$f(tx + (1-t)y) = tf(x) + (1-t)f(y) \quad \text{for } t \in [0,1].$$

We have $f(tx + (1-t)y) = w^T(tx + (1-t)y) + b$

and $tf(x) + (1-t)f(y) = t(w^T x + b) + (1-t)(w^T y + b)$

$= w^T[tx + (1-t)y] + [tb + (1-t)b] = w^T(tx + (1-t)y) + b.$

Hence we get the equality.

(68) Show that if $A \in M_{n \times n}(\mathbb{R})$ is a symmetric positive semidefinite matrix, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

is convex.

Solution Let $t\in[0,1]$. By computation, we have:

$$(1-t)f(x)+tf(y)-f((1-t)x+ty)$$
$$=t(1-t)(x-y)^T A(x-y) \underset{\text{since } A \text{ is positive semidefinite.}}{\geq} 0$$

(69) If f_1 and f_2 are convex and $a_1, a_2 \geq 0$, then $f=a_1 f_1 + a_2 f_2$ is convex.

Solution For $t\in[0,1]$ we have:

$$f((1-t)x+ty) = a_1 f_1((1-t)x+ty) + a_2 f_2((1-t)x+ty)$$
$$\underset{(a_1,a_2\geq 0)}{\leq} a_1((1-t)f_1(x)+tf_1(y)) + a_2((1-t)f_2(x)+tf_2(y))$$
$$=(1-t)[(a_1f_1+a_2f_2)(x)] + t[(a_1f_1+a_2f_2)(y)]$$
$$=(1-t)f(x)+tf(y).$$

(70) Show that if f_1 and f_2 are convex, so is the max function:

$$f(x)=\max\{f_1(x), f_2(x)\}.$$

Deduce that $f(x)=|x|$ is convex.

Take $x,y\in\mathbb{R}^d$ and $t\in[0,1]$. Then $f((1-t)x+ty)$ is either $f_1((1-t)x+ty)$ or $f_2((1-t)x+ty)$. Suppose the first (the second is the same). Then:

$$f((1-t)x+ty) = f_1((1-t)x+ty) \underset{f_1 \text{ convex}}{\leq} (1-t)f_1(x)+tf_1(y)$$
$$\leq (1-t)f(x)+tf(y).$$

Since $|x|=\max\{x,-x\}$ and both are convex, so is $|x|$.

(71) Show that $f:\mathbb{R}^n_+ \to \mathbb{R}$, $f(x_1, \dots, x_n) = +\sum_{i=1}^n x_i \ln(x_i)$ is convex.

Solution We apply the criteria:

$$f \text{ convex} \iff \underset{\text{Hessian}}{H(f)} \text{ is positive semidefinite at each point.}$$

The gradient is $\nabla f(x) = \begin{bmatrix} \ln(x_1)+1 \\ \ln(x_2)+1 \\ \vdots \\ \ln(x_n)+1 \end{bmatrix}$

The Hessian is $H(f)(x) = \begin{bmatrix} 1/x_1 & & & 0 \\ & 1/x_2 & & \\ & & \ddots & \\ 0 & & & 1/x_n \end{bmatrix}$.

Since $x_1, x_2, \dots, x_n > 0$, the Hessian is positive definite, so $f(x)$ is convex $\forall (x_1, \dots, x_n) \in \mathbb{R}_+^n$.

(72) Let $f: \mathbb{R}^n \to \mathbb{R}$. The <u>epigraph</u> of f is:

$$E = \{ (x,s) \in \mathbb{R}^n \times \mathbb{R} \mid s \geq f(x) \}.$$

Show that E is convex $\iff$ f is convex.

<u>Solution</u> ($\Rightarrow$) Suppose that E is convex. Observe that $(x, f(x)), (y, f(y)) \in E$ for all $x, y \in \mathbb{R}^n$. Hence for all $r \in [0,1]$ we have

$$r(x, f(x)) + (1-r)(y, f(y)) \in E$$
$$\| $$
$$(rx + (1-r)y, \; rf(x) + (1-r)f(y))$$

By definition of E, $rf(x) + (1-r)f(y) \geq f(rx + (1-r)y)$.

This shows that f is convex.

($\Leftarrow$) Suppose that f is convex. Take $(x,s), (y,t) \in E$. Then $s \geq f(x)$ and $t \geq f(y)$. Let $r \in [0,1]$.

We <u>want</u> to show that $r(x,s) + (1-r)(y,t) \in E$

$$\| $$
$$(rx + (1-r)y, \; rs + (1-r)t)$$

that is, $rs + (1-r)t \overset{?}{\geq} f(rx + (1-r)y)$.

Since f is convex we have:

$$f(rx + (1-r)y) \leq rf(x) + (1-r)f(y) \underset{\substack{r,\,1-r \geq 0 \\ f(x) \leq s,\; f(y) \leq t}}{\leq} rs + (1-r)t.$$

73 Suppose that $a_1, a_2, \ldots, a_n > 0$. Show that the ellipsoid $S = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : a_1 x_1^2 + a_2 x_2^2 + \ldots + a_n x_n^2 \leq 1\}$ is a convex set.

<u>Solution</u> We use the criterion:

$f : \mathbb{R}^n \to \mathbb{R}$ convex $\Rightarrow S = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid f(x) \leq 0\}$ is a convex subset of $\mathbb{R}^n$.

Since $S = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : a_1 x_1^2 + \ldots + a_n x_n^2 \leq 1\}$

$= \{(x_1, \ldots, x_n) \in \mathbb{R}^n : f(x) \leq 0\}$

for $f(x) = f(x_1, \ldots, x_n) = a_1 x_1^2 + \ldots + a_n x_n^2 - 1$, we only have to show that $f(x_1, \ldots, x_n) = a_1 x_1^2 + \ldots + a_n x_n^2 - 1$ is convex.

This follows easily from the Hessian criteria:

$$H(f) = \begin{bmatrix} a_1 > 0 & & 0 \\ & a_2 > 0 & \\ 0 & & \ddots \; a_n > 0 \end{bmatrix}$$

so $H(f)$ is definite positive and f is even strictly convex.

74 Consider whether the following statements are true or false:

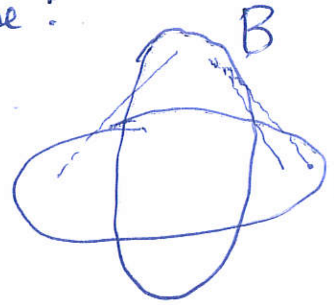
(a) The intersection of two convex sets is convex.

(b) The union of two convex sets is convex.

(c) The difference of a convex set A from another convex set B is convex.

<u>Solution</u> (a) is true (indeed for any family of convex sets)

(b) and (c) are false:



$A \cup B$ is not convex

$A \setminus B =$ 

$A \setminus B$ is not convex.

Optimization. Duality.

(75) Express the following optimization problem (*) as a standard linear program in matrix notation.

Solution: The standard form is:

$$\boxed{\min_{x \in \mathbb{R}^d} c^T x \quad \text{subject to } Ax \le b}$$

Now the problem we consider is:

$$(*) \quad \max_{x \in \mathbb{R}^2, \xi \in \mathbb{R}} p^T x + \xi$$

subject to the constraints $\xi \geq 0,\ x_0 \le 0,\ x_1 \le 3$.

The standard form is:

$$\min_{\begin{bmatrix} x_0 \\ x_1 \\ \xi \end{bmatrix} \in \mathbb{R}^3} \begin{bmatrix} -p \\ -1 \end{bmatrix}^T \begin{bmatrix} x_0 \\ x_1 \\ \xi \end{bmatrix}$$

$$\text{subject to} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \xi \end{bmatrix} \le \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$$

(76) Consider the linear program

$$\min_{x \in \mathbb{R}^2} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\text{subject to} \quad \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \le \begin{bmatrix} 33 \\ 8 \\ -5 \\ -1 \\ 8 \end{bmatrix}$$

Derive the dual program using Lagrange duality.

Solution For a general LP

$$\min_{x\in\mathbb{R}^d} c^T x \quad \text{subject to } Ax \le b$$

$A \in M_{m\times d}(\mathbb{R})$, $b \in \mathbb{R}^m$,

we consider the Lagrangian

$$\mathcal{L}(x,\lambda) = c^T x + \lambda^T(Ax-b) \qquad (\lambda \geqslant 0)$$

$$= (c + A^T\lambda)^T x - \lambda^T b$$

The dual Lagrangian is

$$\mathcal{D}(\lambda) = \min_{x\in\mathbb{R}^d} \mathcal{L}(x,\lambda) = -b^T\lambda .$$

and the dual program is:

$$\max_{\lambda\in\mathbb{R}^m} -b^T\lambda$$

$$\text{subject to } \begin{cases} c + A^T\lambda = 0 \\ \lambda \geqslant 0 \end{cases}.$$

In our particular case, we have:

$$\max_{\lambda\in\mathbb{R}^5} -\begin{bmatrix}33\\8\\-5\\-1\\8\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\\\lambda_5\end{bmatrix}$$

$$\text{subject to } \begin{cases} -\begin{bmatrix}5\\3\end{bmatrix} + \begin{bmatrix}2 & 2 & -2 & 0 & 0\\ 2 & -4 & 1 & -1 & 1\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\\\lambda_5\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \\ \begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\\\lambda_5\end{bmatrix} \geqslant 0 \end{cases}$$

(77) Consider the quadratic program

$$\min_{x\in\mathbb{R}^2} \frac{1}{2}\begin{bmatrix}x_1\\x_2\end{bmatrix}^T\begin{bmatrix}2&1\\1&4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}+\begin{bmatrix}5\\3\end{bmatrix}^T\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

subject to

$$\begin{bmatrix}1&0\\-1&0\\0&1\\0&-1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\le\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

Derive the dual quadratic program using Lagrange duality.

<u>Solution</u> Observe that $\begin{bmatrix}2&1\\1&4\end{bmatrix}$ is <u>positive definite</u> (eigenvalues $3\pm\sqrt{2}$)

so the objective function is convex.

In general a quadratic program is

$$\min_{x\in\mathbb{R}^d} \frac{1}{2}x^TQx + c^Tx$$

subject to $Ax\le b$.

[Q symmetric positive definite]

The Lagrangian is

$$\mathcal{L}(x,\lambda)=\frac{1}{2}x^TQx+c^Tx+\lambda^T(Ax-b) \qquad (\lambda\ge 0)$$

$$=\frac{1}{2}x^TQx+(c+A^T\lambda)^Tx-\lambda^Tb$$

Taking the derivative of $\mathcal{L}(x,\lambda)$ wrt. x and setting it to zero gives:

$$Qx+(c+A^T\lambda)=0$$

Since Q is invertible $x=-Q^{-1}(c+A^T\lambda)$

Substituting into the primal Lagrangian we get the dual Lagrangian.

$$\mathcal{D}(\lambda)=-\frac{1}{2}(c+A^T\lambda)^TQ^{-1}(c+A^T\lambda)-\lambda^Tb$$

Hence the dual optimization problem is:

$$\boxed{\max_{\lambda\in\mathbb{R}^m} -\frac{1}{2}(c+A^T\lambda)^TQ^{-1}(c+A^T\lambda)-\lambda^Tb \quad \text{subject to } \lambda\ge 0}$$

In our particular case, we have:

$$Q = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}, \text{ so } Q^{-1} = \frac{1}{7}\begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}, c = \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$$

Note that:

$$(c+A^T\lambda)^T Q^{-1}(c+A^T\lambda) = \lambda^T(AQ^{-1}A^T)\lambda + c^TQ^{-1}A^T\lambda$$
$$+ \lambda^T AQ^{-1}c + c^TQ^{-1}c = \lambda^T(AQ^{-1}A^T)\lambda + 2[AQ^{-1}c]^T\lambda + c^TQ^{-1}c$$

Observe that $AQ^{-1}A^T$ is a <u>positive semidefinite</u> 4×4 matrix but <u>not</u> positive definite.

Hence the dual program is:

$$\begin{cases} \max\limits_{\lambda \in \mathbb{R}^m} & -\frac{1}{2}\lambda^T(AQ^{-1}A^T)\lambda - [AQ^{-1}+b]^T\lambda - \frac{1}{2}c^TQ^{-1}c \\ \text{subject to} & \lambda \geq 0. \end{cases}$$

In our case, this gives:

$$c^TQ^{-1}c = \frac{88}{7}$$

$$AQ^{-1}c+b = \frac{1}{7}\begin{bmatrix} 24 \\ -10 \\ 8 \\ 6 \end{bmatrix}$$

$$AQ^{-1}A^T = \frac{1}{7}\begin{bmatrix} 4 & -4 & -1 & 1 \\ -4 & 4 & 1 & -1 \\ -1 & 1 & 2 & -2 \\ 1 & -1 & -2 & 2 \end{bmatrix}.$$ Hence we get:

<u>Dual Program</u>

$$\boxed{\begin{array}{l} \max\limits_{\lambda \in \mathbb{R}^4} \; -\frac{1}{2}\lambda^T \begin{bmatrix} 4/7 & -4/7 & -1/7 & 1/7 \\ -4/7 & 4/7 & 1/7 & -1/7 \\ -1/7 & 1/7 & 2/7 & -2/7 \\ 1/7 & -1/7 & -2/7 & 2/7 \end{bmatrix}\lambda - \frac{1}{7}\begin{bmatrix} 24 \\ -10 \\ 8 \\ 6 \end{bmatrix}\lambda - \frac{44}{7} \\ \\ \text{subject to } \lambda \geq 0 \end{array}}$$

(78) Consider the convex optimization problem

$$\min_{w \in \mathbb{R}^d} \frac{1}{2} w^T w$$

$$\text{subject to } w^T x \geq 1.$$

(Here $x \in \mathbb{R}^d$ is a fixed vector)

Derive the Lagrangian dual by introducing Lagrange multiplier λ.

<u>Solution</u>: Write it in the form:

$$\min_{w \in \mathbb{R}^d} \frac{1}{2} w^T w$$

$$\text{subject to } 1 - x^T w \leq 0$$

The Lagrangian is: $\mathcal{L}(w,\lambda) = \frac{1}{2} w^T w + \lambda(1 - x^T w)$

Hence $\mathcal{L}_D(\lambda) = \min_{w \in \mathbb{R}^d} \left(\frac{1}{2} w^T w + \lambda(1 - x^T w) \right) \quad (\lambda \geq 0).$

Equating the gradient to 0, we get $w - \lambda x = 0 \Rightarrow \boxed{w = \lambda x}$

Hence $\mathcal{L}_D(\lambda) = \frac{1}{2} \lambda^2 x^T x + \lambda(1 - \lambda x^T x)$

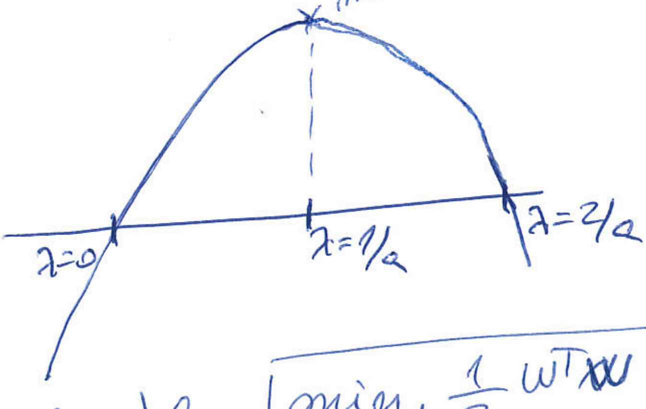
Hence $\boxed{\mathcal{L}_D(\lambda) = -\frac{1}{2} \lambda^2 x^T x + \lambda}$

The dual problem is

$$\boxed{\begin{array}{l} \max_{\lambda \in \mathbb{R}} \left[-\frac{1}{2} \lambda^2 x^T x + \lambda \right] \\ \text{subject to } \lambda \geq 0 \end{array}}$$

This can be solved, note that $-\frac{1}{2} \lambda^2 (x^T x) + \lambda$ is the parabola $-\frac{1}{2} a \lambda^2 + \lambda$, where $a = x^T x > 0$.

The graphic of the parabola is



$-\frac{1}{2} a \lambda^2 + \lambda = 0$ $\left\langle \begin{array}{l} \lambda = 0 \\ \lambda = 2/a > 0 \end{array} \right.$

<u>The maximum is attained at</u>

$\boxed{\lambda = 1/a} \Rightarrow w = \lambda x = \frac{1}{a} \cdot x$

$= \frac{x}{x^T x} \Rightarrow \boxed{w = \frac{x}{x^T x}}$

Hence the minimum value of the original problem $\boxed{\begin{array}{l} \min_{w \in \mathbb{R}^d} \frac{1}{2} w^T w \\ \text{subject to } w^T x \geq 1 \end{array}}$

is $\frac{1}{2} \left(\frac{x}{x^T x} \right)^T \left(\frac{x}{x^T x} \right) = \frac{1}{2} \cdot \frac{1}{x^T x}$