Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
Department of Mathematics

Convex Optimization

Retake Exam

September 28$^{\text{th}}$, 2022

Student ID:	

General remarks.

- TODO: Check the below.
- Place your student identity card on the table.
- This is a closed book exam. No extra material is allowed (except for a subject-neutral dictionary).
- Use a pen to write down your solutions. Pencils and red pens are not allowed.
- Turn off all your technical devices and put them away.
- Write your solutions directly after each exercise, if you need more space use the extra sheets at the end. Extra paper will be distributed if needed.
- You have 3 hours to solve the exam. 15 minutes before the end of the exam, no premature submission is allowed anymore.
- For all problems, provide a complete solution in English, including all explanations and implications in a mathematically clear and well-structured way. Please hand in a readable and clean solution. Cross out any invalid solution attempts.
- Unless you are explicitly asked to prove them, you may use results from the lecture without proof (provided that you reference or state them clearly). Results from the problem sets may only be used if you prove them.
- Ask any question that you might have immediately and during the exam.

Good luck!

Problem:	1	2	3	4	5	6	7	Total
Points:	/ 10	/ 6	/ 10	/ 10	/ 8	/ 10	/ 8	/ 62

Exercise 1 (10 Points)

For each of the following statements decide whether they are true (T) or false (F). If a question has multiple statements, points are distributed uniformly. Wrong answers lead to a deduction of the corresponding points while blank answers do not change the score. The minimum number of points achievable for Exercise 1 is zero. No negative points will be carried over.

T	F	
☐	☐	Consider the convex optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \ & f(x) \\ s.t. \ & g(x) - 1 \leq 0, \end{aligned} \tag{P}$$

where

$$\begin{aligned} f(x) &:= \frac{1}{\sqrt{2\pi}} e^{\frac{\|x\|^2}{2}} \\ g(x) &:= \left\| x - \begin{pmatrix} 0 \\ 2 \end{pmatrix} \right\|_2^2 . \end{aligned}$$

Then strong duality holds for (P). **(2 Points)**.

☐ ☐ The function $f \colon \mathbb{R}_{>0} \to \mathbb{R},\ f(x) := tx\log(x) - \log(x)$, where $t > 0$, is self concordant **(2 Points)**.

☐ ☐ Consider the following minimization problem

$$\begin{aligned} \min \ & x^2 + y^2 \\ \text{s.t. } & x \leq 0 \\ & -x + y \leq 1 \\ & 2x + y = 1. \end{aligned} \tag{P}$$

The central path C of (P) is well defined and has at least 2 points in it (i.e. $|C| \geq 2$) **(2 Points)**.

(2 Points) Consider $f \colon \mathbb{R} \to \mathbb{R},\ f(x) = e^{-x}$.

☐ ☐ The function f is strongly convex.

☐ ☐ The function f is strictly convex.

For each of the following four programs decide whether they are convex in the form given.

☐ ☐

$$\begin{aligned} \max \quad & x^2 - y^4 \\ s.t. \quad & x - y \leq 0, \\ & x \geq 0. \end{aligned}$$

☐ ☐

$$\begin{aligned} \min \quad & e^{x+y} \\ s.t. \quad & x^2 \leq 0, \\ & y \geq 0. \end{aligned}$$

☐ ☐

$$\begin{aligned} \min \quad & \log\left(e^x + e^{2y} + e^{3z}\right) \\ s.t. \quad & x + y + z = 0, \\ & x^2 - y - z \leq 0. \end{aligned}$$

☐ ☐

$$\begin{aligned} \min \quad & xyz \\ s.t. \quad & xy + yz + zx \leq 1, \\ & \frac{xy}{z} = 1, \\ & x, y, z > 0. \end{aligned}$$

Exercise 2 (6 Points)

Give a short proof or counterexample for the following statements. If you give a counterexample you also have to show how it contradicts the statement.

(a) **(3 Points)** Let $C \subseteq \mathbb{R}^n$ be a convex set and $g\colon C \to \mathbb{R}$ convex. Further let $I = [a, b] \subseteq \mathbb{R}$ be an interval such that $g(C) \subseteq I$ and let $f\colon I \to \mathbb{R}$ be convex. Then $f \circ g\colon C \to \mathbb{R}$, $x \mapsto f(g(x))$ is convex.

(b) **(3 Points)** For every $n \in \mathbb{N}_{\geq 1}$ let $\mathcal{S}^n = \left\{ A \in \mathbb{R}^{n \times n} \mid A^\top = A \right\}$ be the set of symmetric matrices and $\mathcal{S}^n_{++} = \left\{ A \in \mathcal{S}^n \mid A \succ 0 \right\}$ the set of positive definite matrices. Then $\mathcal{S}^n_{++}$ is a proper cone for every $n \in \mathbb{N}_{\geq 1}$.

Exercise 3 (10 Points)

You are tasked to model a wheel chair ramp. The ramp consists of a wooden prism with a rectangular triangle as basis. On the side of the ramp given by the hypotenuse a metal plate is added.
The ramp has to satisfy several constraints. It shall have height at least 15cm. The slope of the ramp shall not exceed 20%. For space reasons no dimension (width, height or depth) of the ramp may exceed 1.5m. To handle the weight of the wheel chairs the metal plate has to be wider than long (here the length is taken along the slope).
Finally the production cost is given by a price C_m for a square meter of metal plate and a price C_w per cubic meter of wood.

- **(7 Points)** Formulate the problem of finding the minimal cost ramp - encoded by its width, height and depth - satisfying the constraints as a GP in general form.
- **(3 Points)** Give a convex program that is equivalent to your program.

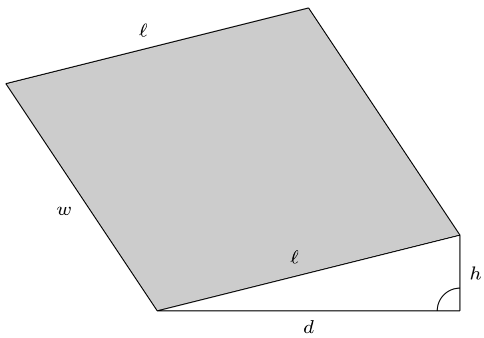

Figure 1: Sketch of the ramp to be designed. The grey side is where the metal plate is added. The width, depth, height and length are marked with w, d, h and ℓ respectively.

Exercise 4 (10 Points)

For given q, c consider the parabola equation $f(x) = -x^2 + q$ and the rotated parabola $g(x) = \sqrt{c - x}$. You are asked for an axis parallel rectangle which interior lies completely under the graphs of f and g, completely above the x-axis and completely to the right of the y-axis. Furthermore under all these rectangles you are tasked to find that one with maximal area.
Give a convex program that computes the height and the length of this maximal rectangle. Explain why your program is convex and why it does solve the problem at hand.

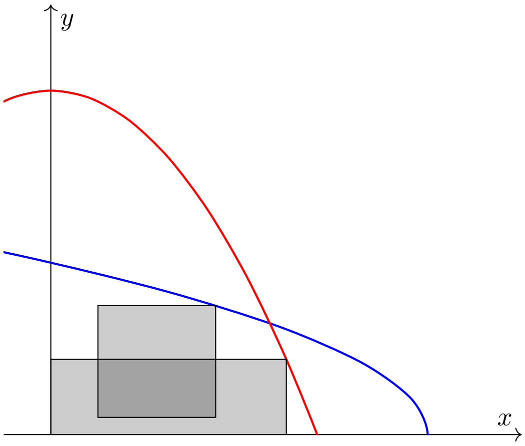

Figure 2: Example instance of two curves together with two feasible solutions.

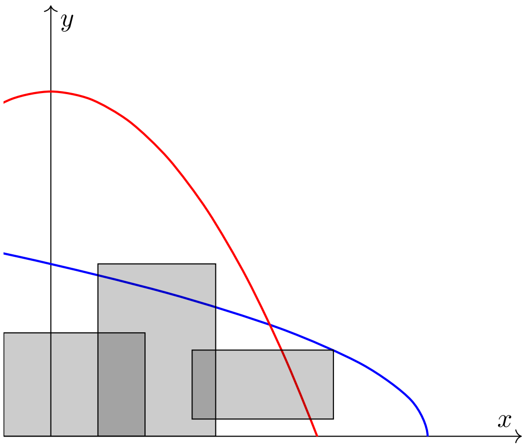

Figure 3: The same two curves but now with three infeasible rectangles.

Exercise 5 (8 Points)

Show that the maximum of a convex function f over the polytope $\mathcal{P} = \mathrm{conv}\{v_1, v_2, \ldots, v_k\}$ is achieved at one of its vertices, i.e.,

$$\max_{x \in \mathcal{P}} f(x) = \max_{i \in [k]} f(v_i).$$

Exercise 6 (10 Points)

Let $n \in \mathbb{N}_{\geq 1},\ A_1, \dots, A_m, C \in \mathcal{S}^n$ and $c \in \mathbb{R}^n$. Consider the semidefinite program

$$\begin{aligned} &\min\ c^\top x \\ &\text{s.t. } A(x) \coloneqq \sum_{i=1}^m x_i A_i + C \succeq 0. \end{aligned}$$

The KKT-Conditions for this problem are given by

$$\begin{aligned} A(x) &\succeq 0, && && \text{(KKT-1)} \\ Y &\succeq 0, && && \text{(KKT-3)} \\ \operatorname{Tr}(A_i Y) &= c_i, && \text{for } i \in [n] \text{ and} && \text{(KKT-4)} \\ \operatorname{Tr}(A(x)Y) &= 0. && && \text{(KKT-5)} \end{aligned}$$

(a) **(4 Points)** Prove that the last condition $\operatorname{Tr}(A(x)Y) = 0$ is equivalent to $A(x)Y = 0$.
Hint. The term $Y^{1/2}A(x)Y^{1/2}$ might be helpful.

Now let $A \in \mathcal{S}^n$ be a symmetric matrix. The goal of the following exercises is to prove the identity

$$\max_{\substack{x \in \mathbb{R}^n \\ x^\top x = 1}} x^\top A x = \lambda_{\max}(A).$$

(b) **(2 Points)** Use duality to prove

$$\max_{X \succeq 0, \operatorname{Tr}(X) = 1} \operatorname{Tr}(AX) = \lambda_{\max}(A).$$

(c) **(4 Points)** Show

$$\max_{\substack{x \in \mathbb{R}^n \\ x^\top x = 1}} x^\top A x = \max_{X \succeq 0, \operatorname{Tr}(X) = 1} \operatorname{Tr}(AX)$$

and conclude the goal

$$\max_{\substack{x \in \mathbb{R}^n \\ x^\top x = 1}} x^\top A x = \lambda_{\max}(A).$$

Exercise 7   (8 Points)

Derive the explicit dual (without an infimum) of the following convex programs.

(a) **(4 Points)** The relative entropy between two vectors $x, y \in \mathbb{R}^n_{>0}$ is defined by

$$D_{KL}(x, y) \coloneqq \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right).$$

This function $D_{KL} \colon \mathbb{R}^{2n}_{>0} \to \mathbb{R}$ is convex.
Let $y \in \mathbb{R}^n_{>0}, A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ be given and consider the objective $f \colon \mathbb{R}^n_{>0} \to \mathbb{R}, f(x) \coloneqq D_{KL}(x, y)$ under equality constraints:

$$\begin{aligned} \min_{x \in \mathbb{R}^n_{>0}} \ & \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right) \\ \text{s.t. } & Ax = b \\ & \mathbb{1}^\top x = 1, \end{aligned}$$

where $\mathbb{1} \in \mathbb{R}^n$ is the vector with all entries being 1.

(b) **(4 Points)** Let $n, N \in \mathbb{N}_{\geq 1}$, $m_i \in \mathbb{N}_{\geq 1}, A_i \in \mathbb{R}^{m_i \times n}, b_i \in \mathbb{R}^{m_i}$ for $i \in [N]$ and $x_0 \in \mathbb{R}^n$. Derive the dual of

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} \|A_i x + b_i\|_2 + \frac{1}{2} \|x - x_0\|_2^2$$

by first introducing new variables $y_i \in \mathbb{R}^{m_i}$ and equality constraints $y_i = A_i x + b_i$.

Extra sheets you may need

Exercise 1 Solution (10 Points)

(a) *True.* This immediately follows from Slater's condition with e.g. $x^* = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$.

(b) *True.* We calculate

$$
\begin{aligned}
f'(x) &= -\frac{1}{x} + t + t\log(x), \\
f''(x) &= \frac{1}{x^2} + \frac{t}{x}, \\
f'''(x) &= -\frac{2}{x^3} - \frac{t}{x^2}.
\end{aligned}
$$

This yields

$$
\frac{|f'''(x)|}{f''(x)^{3/2}} = \frac{2/x^3 + t/x^2}{\left(1/x^2 + t/x\right)^{3/2}} = \frac{2 + tx}{(1+tx)^{3/2}}.
$$

Now define $h(a) := \frac{2+a}{(1+a)^{3/2}}$ so that

$$
h(tx) = \frac{|f'''(x)|}{f''(x)^{3/2}}.
$$

By definition we get $h(0) = 2$ hence, if we can show that h is decreasing for $a > 0$, we get the claim. Therefore we will show $h'(a) < 0$ for $a > 0$:

$$
\begin{aligned}
h'(a) &= \frac{(1+a)^{3/2} - 3/2\,(1+a)^{1/2}\,(2+a)}{(1+a)^3} \\
&= -\frac{2 + a/2}{(1+a)^{5/2}} < 0.
\end{aligned}
$$

This shows the claim.

(c) *False.* The IPM can only be applied if there are strictly feasible points, since otherwise the log would not be defined. Hence it cannot be applied for this example as the only feasible point it $x = 0, y = 1$ at which both inequality constraints are active. Therefore the central path is not well defined.

(d) f is strictly but not strongly convex. This can be seen by noting that the derivative is positive but converges towards zero as x tends to infinity.

(e) The first program is not convex as the objective function is not convex (its second derivative is $2 - 12y^2$. The second one is convex is all involved functions are convex. The third one is convex as well. The objective is convex as a log-sum which is known from the lecture. As the equality constraint is affine it is convex as is the inequality. The final one is not convex as for example the second constraint is not affine. It is a GP though not in convex form.

Exercise 2 Solution (6 Points)

(a) *Counterexample.* Let $n = 1$, $C = I = \mathbb{R}$ and $g(x) = x^2, f(x) = -x$. Then $(f \circ g)(x) = -x^2$ which is not convex.

(b) *Counterexample.* Let $n = 1$. Then $\mathcal{S}^n_{++} = (0, \infty)$ and hence open. Therefore it is not proper.

Exercise 3 Solution (10 Points)

When implementing the constraints one after the other in the straight forward way we get the following program, where l is the length of the hypothenuse of the base triangle:**(2 Points)**

$$\begin{aligned}
\min \quad & C_m wl + \frac{1}{2} C_w hdw \\
\text{s.t.} \quad & w, h, d \leq 1.5 \\
& d \geq 5h \\
& l = \sqrt{d^2 + h^2} \\
& w \geq l \\
& h \geq 0.15 \\
& w, d \geq 0.
\end{aligned}$$

The length of the hypothenuse is $\sqrt{d^2 + h^2}$, but the constraint $l = \sqrt{d^2 + h^2}$ is not given by a posynomial. Squaring gives the equivalent formulation $l^2 = d^2 + h^2$, but this is still not feasible as we are only allowed to upper bound posynomials against monomials in a geometric program. So we only lower bound l via $l^2 \geq d^2 + h^2$ or equivalently $\frac{d^2}{l^2} + \frac{h^2}{l^2} \leq 1$ which now is a valid constraint in a geometric program.
Now we have to check that loosening the equality to an inequality does not change the result. As l only appears in the constraint in which it is upper bounded by w, no triple (w, h, d) can become feasible for a relaxed constraint on l if it is not already feasible with $l^2 = d^2 + h^2$. And the objective value only worsens when increasing l so we did not change the optimal value, nor the set of (w, h, d) that are part of some optimal solution.
So in total the geometric program in general form is given by:

$$\begin{aligned}
\min \quad & C_m wl + \frac{1}{2} C_w hdw \\
\text{s.t.} \quad & w, h, d \leq 1.5 \\
& \frac{5h}{d} \leq 1 \\
& \frac{d^2}{l^2} + \frac{h^2}{l^2} \leq 1 \\
& \frac{l}{w} \leq 1 \\
& \frac{0.15}{h} \leq 1 \\
& w, d \geq 0.
\end{aligned}$$

Exercise 4 Solution (10 Points)

Observe that we can assume the lower left corner of the optimal rectangle to be at the origin. If not we can extend the rectangle, as both graph we have to lie under are monotone decreasing. Now assume the length of the rectangle in x direction is a and in y direction is b. **(2 Points)**
We begin with the first (non-convex) program. **(2 Points)**

$$\begin{aligned}
\max \quad & ab \\
\text{s.t.} \quad & b \leq -a^2 + q \\
& b^2 \leq c - a \\
& a, b \geq 0.
\end{aligned}$$

Now we rewrite $\max ab$ as $\max s^2$ with the added constraint $s^2 \leq ab$ which is a second order cone constraint **(2 Points)**. Then we see - as squaring is monotone on the non-negative reals - that we can replace this constraint with $\max s$ **(1 Point)**. Now we finally transform this problem to $\min -s$ **(1 Point)**. So overall the final problem is:**(1 Point)**

$$\begin{aligned}
\min \quad & -s \\
\text{s.t.} \quad & b + a^2 + q \leq 0 \\
& b^2 + a - c \leq 0 \\
& s^2 \leq ab \\
& a, b \geq 0.
\end{aligned}$$

The first two constraints are convex as positve squares are convex as well as linear functions. The third constraint is a SOC constraint and as such convex.**(1 Point)**

Exercise 5 Solution (8 Points)

First note that every polytope P can be expressed as a convex hull of its vertices $\{v_1, v_2, \ldots, v_k\}$. **(1 Point)**
Let f be a convex function that is being maximized over the polytope P. Assume by contradiction that the maximum value over P, denoted as f^*, is attained at a point x^* that is not a vertex of P and the value of the function attained at any vertex $i \in [k]$, denoted by f_i is strictly smaller than f^*. **(2 Points)**
Since P is a convex hull of its vertices, every point in P can be written as a convex hull of its vertices. **(1 Point)**
Since $x^* \in P$ there exist number $\lambda_i \in [0, 1]$ for $i \in [k]$ such that $\sum_{i=1}^{k} \lambda_i = 1$. **(2 Points)**
Thus $x^* = \sum_{i=1}^{k} \lambda_i v_i$ and by convexity of f

$$f^* = f(x^*) = f(\sum_{i=1}^{k} \lambda_i v_i) \leq \sum_{i=1}^{k} \lambda_i f(v_i) < f^* \sum_{i=1}^{k} \lambda_i = f^*$$

where the last inequality comes form the assumption that $f(v_i) < f^*$ for every $i \in [k]$, and the last equality comes from the fact that $\sum_{i=1}^{k} \lambda_i = 1$, which leads to a contradiction. **(2 Points)**
Note that the first inequality comes from the convexity of f, however applied to a combination of more than 2 points. This clearly holds by applying iteratively for $j \in \{2, \ldots, k\}$ the definition of convexity to a combination of 2 points, where the first one is a summation of points $\{1, \ldots, j-1\}$ and the second one is a j'th point. E.g., the first step, for $j = k$, goes as follows:

$$f\left(1 \cdot \sum_{i=1}^{k-1} \lambda_i v_i + \lambda_k v_k\right) \leq f\left(\sum_{i=1}^{k-1} \lambda_i f_i\right) + \lambda_k f(v_k)$$

(2 Points)

Exercise 6 Solution (10 Points)

(a) For easier notation, denote $A := A(x)$. By previous conditions we know $A, Y \succeq 0$ and hence Y has a psd square root $Y^{1/2}$. Denote $\tilde{A} := Y^{1/2} A Y^{1/2}$ which satisfies $0 = \operatorname{Tr}(AY) = \operatorname{Tr}(\tilde{A})$. Due to

$$x^\top \tilde{A} x = \left(Y^{1/2} x\right)^\top A \left(Y^{1/2} x\right) \geq 0,$$

$\tilde{A}$ is psd and hence $\operatorname{Tr}\left(\tilde{A}\right) = 0$ implies $\tilde{A} = 0$ **(2 Points)**. Next we show that $\tilde{A} = 0$ implies $AY = 0$. Therefore first use the SVD $Y = V^\top D V$ where V is an invertible matrix. This allows us to assume without loss of generality that Y is diagonal (since $Vx = 0 \Leftrightarrow x = 0$ for invertible matrices). Further we group the strictly positive eigenvalues and get

$$Y = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}, \qquad A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{1,2}^\top & A_{2,2} \end{pmatrix},$$

where $\Lambda \succ 0$. Here $A_{1,1}$ is of the same size as Λ and the other dimensions follow. Finally the condition $\tilde{A} = 0$ becomes

$$0 = \begin{pmatrix} \Lambda^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{1,2}^\top & A_{2,2} \end{pmatrix} \begin{pmatrix} \Lambda^{1/2} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \Lambda^{1/2} A_{1,1} \Lambda^{1/2} & 0 \\ 0 & 0 \end{pmatrix}.$$

Since $\Lambda \succ 0$, we obtain $A_{1,1} = 0$ and hence the Schur complement implies $A_{1,2} = 0$. Therefore the claim

$$AY = \begin{pmatrix} 0 & 0 \\ 0 & A_{2,2} \end{pmatrix} \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} = 0$$

follows **(2 Points)**.

(b) By the duality theory from the lecture we get the dual

$$\begin{aligned} &\min\ y \\ &\text{ s.t. } yI_n \succeq A \end{aligned}$$

(1 Point). Using the SVD $A = VDV^\top$ we get that the inequality constraint can be rewritten to

$$\sum_{i=1}^{n} (y - \lambda_i)\, v_i v_i^\top \succeq 0$$

and hence to

$$y - \lambda_i \geq 0, \qquad i \in [n].$$

This yields the dual optimal variable $y^* = \lambda_{\max}(A)$. Since the positive semidefinite cone is proper, the slater condition holds and hence the claim follows **(1 Point)**.

(c) We use $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ and $x^\top x = \mathrm{Tr}(xx^\top)$ to get

$$\max_{\substack{x \in \mathbb{R}^n \\ x^\top x = 1}} x^\top A x = \max_{\substack{x \in \mathbb{R}^n \\ \mathrm{Tr}(xx^\top) = 1}} \mathrm{Tr}(Axx^\top).$$

Since every rank one matrix X can be written as $X = xx^\top$ for some X we further get

$$\max_{\substack{x \in \mathbb{R}^n \\ \mathrm{Tr}(xx^\top) = 1}} \mathrm{Tr}(Axx^\top) = \max_{\substack{X \in \mathbb{R}^{n\times n} \\ \mathrm{rank}(X) = 1 \\ \mathrm{Tr}(X) = 1}} \mathrm{Tr}(Axx^\top).$$

Since rank one matrices only have one eigenvalue, $\mathrm{Tr}(X) = 1$ implies $X \succeq 0$ and hence we can add the constraint without changing the solution **(1 Point)**. Therefore we are done if we can show that the optimal value of the right hand side is attained by a matrix with rank one. Since both, the primal

$$\begin{aligned} \max\ &\mathrm{Tr}(AX) \\ &\mathrm{Tr}(X) = 1 \\ &X \succeq 0 \end{aligned}$$

and dual

$$\begin{aligned}&\min y\\ &\text{s.t. } yI_n \succeq A\end{aligned}$$

are strictly feasible, they attain there optimum at a primal-dual pair (X^*, y^*) that satisfies the KKT-Conditions **(1 Point)**. By (a), these are equivalent to

$$yI_n \succeq A \tag{KKT-1}$$
$$X \succeq 0 \tag{KKT-3}$$
$$\mathrm{Tr}(X) = 1 \tag{KKT-4}$$
$$(yT_n - A)\, X = 0. \tag{KKT-5}$$

From (KKT-5) we hence get that every column x of X satisfies $(yT_n - A)\, x = 0$. Now let x be such nonzero column and normalize it such that $\|x\|_2 = 1$. Then $\mathrm{Tr}(xx^\top) = 1,\ (yT_n - A)\, xx^\top = 0$ and hence $(xx^\top, y^*)$ satisfies the KKT conditions as well. By strict feasibility, $(xx^\top, y^*)$ attains the optimum and hence the claim follows. The overall objective is now achieved by combining (b) and (c) **(2 Points)**.

Exercise 7 Solution (8 Points)

(a) We get the Lagrangian

$$L\colon \mathbb{R}^n_{>0} \times \mathbb{R}^m \times \mathbb{R},\ L(x, \mu, \tilde{\mu}) = \sum_{i=1}^n x_i \log\left(\frac{x_i}{y_i}\right) + \tilde{\mu}^\top b - \tilde{\mu}^\top A x + \mu - \mu \mathbb{1}^\top x.$$

Since this function is still convex and differentiable in x **(1 Point)**, it suffices to set the gradient to zero which yields the equations

$$1 + \log\left(\frac{x_i}{y_i}\right) - \tilde{\mu}^\top a_i - \mu = 0, \qquad i \in [n],$$

where a_i is the i-th column of A. These equations have the solutions

$$x_i = y_i \exp\left(\tilde{\mu}^\top a_i + \mu - 1\right)$$

(1 Point) and hence the Lagrange dual function is given by

$$\widehat{L}\left(\tilde{\mu}, \mu\right) = \tilde{\mu}^\top b + \mu - \sum_{i=1}^n y_i \exp\left(\tilde{\mu}^\top a_i + \mu - 1\right).$$

Finally the lagrange dual problem is

$$\max_{\mu \in \mathbb{R},\ \tilde{\mu} \in \mathbb{R}^m} \tilde{\mu}^\top b + \mu - \sum_{i=1}^n y_i \exp\left(\tilde{\mu}^\top a_i + \mu - 1\right)$$

(2 Points) which can be simplified to

$$\max_{\tilde{\mu} \in \mathbb{R}^m} \tilde{\mu}^\top b - \log\left(\sum_{i=1}^n y_i \exp\left(\tilde{\mu}^\top a_i\right)\right)$$

by maximizing over μ (the last step was not expected).

(b) After introducing the equality constraint, the Lagrangian of

$$\begin{aligned} \min \ & \sum_{i=1}^{N} \|y_i\|_2 + \frac{1}{2} \|x - x_0\|_2^2 \\ \text{s.t. } & y_i = A_i x + b_i, \qquad \text{for } i \in [N] \end{aligned}$$

is given by

$$L(x, y, \nu_1, \dots, \nu_n) = \sum_{i=1}^{N} \|y_i\|_2 + \frac{1}{2} \|x - x_0\|_2^2 + \sum_{i=1}^{N} \nu_i^\top (A_i x + b_i - y_i).$$

To calculate the Lagrange dual function we start by minimizing of y_i and get

$$\inf_{y_i} \|y_i\|_2 - z_i^\top y_i = \inf_{y_i} \|y_i\|_2 + z_i^\top y_i = \begin{cases} 0, & \text{if } \|z_i\|_2 \leq 2 \\ -\infty, & \text{otherwise.} \end{cases}$$

To see this first suppose $\|\nu_i\|_2 > 1$ and set $y_i = -t\nu_i$ and let t tend to infinity. On the other hand, if $\|\nu_i\|_2 \leq 1$ the Cauchy Schwarz inequality yields $\|y_i\|_2 + \nu_i^\top y_i \geq 0$ **(2 Points)**. Since the Lagrangian is convex and differentiable with respect to x we can minimize over x by setting the gradient to zero **(1 Point)**. This yields

$$x = x_0 + \sum_{i=1}^{N} A_i^\top \nu_i$$

and hence the dual problem is given by

$$\begin{aligned} \max \ & \sum_{i=1}^{N} (A_i x_0 + b_i)^\top \nu_i - \frac{1}{2} \left\| \sum_{i=1}^{N} A_i^\top \nu_i \right\|_2^2 \\ \text{s.t. } & \|\nu_i\|_2 \leq 1, \qquad \text{for } i \in [N] \end{aligned}$$

(1 Point).