SI151A - Convex Optimization and its Applications in Information Science
Fall 2023, Final Exam Solutions

Note: We are interested in the reasoning underlying the solution, as opposed to simply the answer. Thus, solutions with the correct answer but without adequate explanation will not receive full credit; on the other hand, partial solutions with explanation will receive partial credit. Within a given problem, you can assume the results of previous parts in proving later parts (e.g., it is fine to solve part 3) first, assuming the results of parts 1) and 2)). Your use of resources should be limited to printed lecture slides, lecture notes, homework, homework solutions, general resources, class reading and textbooks, and other related textbooks on optimization. You should not discuss the final exam problems with anyone or use any electronic devices. Detected violations of this policy will be processed according to ShanghaiTech's code of academic integrity. Please hand in the exam papers and answer sheets at the end of exam.

I. Basic Knowledge

1. Prove that the positive semidefinite cone $\mathbb{S}_{++}^n$ is self dual. (10 points)

   Solution:

   By definition $\left(\mathbb{S}_+^n\right)^* = \left\{\boldsymbol{B} \in \mathbb{S}^n : \operatorname{tr}(\boldsymbol{AB}) \geq 0, \forall \boldsymbol{A} \in \mathbb{S}_+^n\right\}$.

   We first show that $\mathbb{S}_+^n \subseteq \left(\mathbb{S}_+^n\right)^*$. Assume $\boldsymbol{B}$ is positive semidefinite. The eigenvalue decomposition of $\boldsymbol{B}$ takes the form $\boldsymbol{B} = \sum_{i=1}^n \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$ where $\lambda_i \geq 0$ for $i = 1, \ldots, n$ and the $\boldsymbol{v}_i$ are the unit-normed eigenvectors of $\boldsymbol{B}$. Now for any $\boldsymbol{A} \in \mathbb{S}_+^n$ we have $\operatorname{tr}(\boldsymbol{AB}) = \sum_{i=1}^n \lambda_i \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{v}_i\boldsymbol{v}_i^\top\right) = \sum_{i=1}^n \lambda_i \boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_i$. Since $\boldsymbol{A} \in \mathbb{S}_+^n$ we have $\boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_i \geq 0$ for all $i = 1, \ldots, n$ and thus, since $\lambda_i \geq 0$ we get $\operatorname{tr}(\boldsymbol{AB}) \geq 0$. This shows $\mathbb{S}_+^n \subseteq \left(\mathbb{S}_+^n\right)^*$.

   To show the reverse inclusion, assume $B \in \mathbb{S}^n$ is such that $\operatorname{tr}(\boldsymbol{AB}) \geq 0$ for all $\boldsymbol{A} \in \mathbb{S}_+^n$. We want to show that $\boldsymbol{B}$ is positive semidefinite. By taking $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{x}^\top$ for any $\boldsymbol{x} \in \mathbb{R}^n$ we get that $\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{B}\right) = \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x} \geq 0$. This is true for all $\boldsymbol{x} \in \mathbb{R}^n$ and thus shows that $\boldsymbol{B}$ is positive semidefinite.

2. Determine the convexity (i.e., convex, concave, or neither) of the following function and provide your argument:

   $$f(x_1, x_2) = x_1^2 / x_2,$$

   where $(x_1, x_2) \in \mathbb{R} \times \mathbb{R}_{++}$. (10 points)

   Solution:

   f is quadratic-over-linear, and thus is convex.

3. Prove that for convex problems, any locally optimal point is globally optimal. (10 points)

   Solution:

   Let $\boldsymbol{x}^*$ be a local minimizer of f_0 on the set $\mathcal{X}$, and let $\boldsymbol{y} \in \mathcal{X}$. By definition, $\boldsymbol{x}^* \in \operatorname{dom} f_0$. We need to prove that $f_0(\boldsymbol{y}) \geq f_0\left(\boldsymbol{x}^*\right) = p^*$. There is nothing to prove if $f_0(\boldsymbol{y}) = +\infty$, so let us assume that $\boldsymbol{y} \in \operatorname{dom} f_0$. By convexity of f_0 and $\mathcal{X}$, we have that for $\boldsymbol{x}_\theta := \theta\boldsymbol{y} + (1-\theta)\boldsymbol{x}^* \in \mathcal{X}$,

   $$f_0\left(\boldsymbol{x}_\theta\right) - f_0\left(\boldsymbol{x}^*\right) \leq \theta\left(f_0(\boldsymbol{y}) - f_0\left(\boldsymbol{x}^*\right)\right).$$

   Since $\boldsymbol{x}^*$ is a local minimizer, the left-hand side in this inequality is nonnegative for all small enough values of $\theta > 0$. We conclude that the right hand side is nonnegative, i.e., $f_0(\boldsymbol{y}) \geq f_0\left(\boldsymbol{x}^*\right)$, as claimed.

II. Convex Problem

Determine whether these following problems are convex or not respectively.

– If yes, equivalently reformulate the original problem into a standard convex optimization form, i.e., Linear Programming (LP), Second-Order Cone Programming (SOCP), and Semidefinite Programming (SDP).

– If no, relax the original problem to a convex problem at first, and then equivalently reformulate the relaxed problem into a standard convex optimization form, or provide a successive convex approximation-based algorithm to solve it.

Hint: Standard convex optimization form of Linear Programming (LP):

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d \\ \text{subject to} \quad & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned} \tag{1}$$

Standard convex optimization form of Second-Order Cone Programming (SOCP):

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{f}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & \|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\| \leq \boldsymbol{c}_i^\mathsf{T}\boldsymbol{x} + d_i,\ i = 1, \dots m \\ & \boldsymbol{F}\boldsymbol{x} = \boldsymbol{g}. \end{aligned} \tag{2}$$

Standard convex optimization form of Semidefinite Programming (SDP):

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & x_1\boldsymbol{F}_1 + x_2\boldsymbol{F}_2 + \ldots + x_n\boldsymbol{F}_n \preceq \boldsymbol{G} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned} \tag{3}$$

1. Consider the following rank minimization problem

$$\begin{aligned} \underset{\boldsymbol{X}}{\text{minimize}} \quad & \text{rank}(\boldsymbol{X}) \\ \text{subject to} \quad & \boldsymbol{X} \in \mathcal{C}, \end{aligned}$$

where $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ is the optimization variable and $\mathcal{C}$ is a convex set. (10 points)

Solution:

This is not a convex problem. We relax it as the nuclear norm minimization problem, i.e.,

$$\begin{aligned} \underset{\boldsymbol{X}}{\text{minimize}} \quad & \|\boldsymbol{X}\|_* \\ \text{subject to} \quad & \boldsymbol{X} \in \mathcal{C}. \end{aligned}$$

Note that we have the fact that the nuclear norm is the dual norm of spectral norm, i.e.,

$$\|\boldsymbol{X}\|_* = \max\{\langle \boldsymbol{X}, \boldsymbol{W} \rangle \mid \|\boldsymbol{W}\| \leqslant 1\}.$$

Then,

$$\|\boldsymbol{X}\|_* = \max\{\langle \boldsymbol{X}, \boldsymbol{W} \rangle \mid \|\boldsymbol{W}\| \leqslant 1\} = \max\left\{\langle -\boldsymbol{X}, \boldsymbol{W} \rangle \mid \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W} \\ \boldsymbol{W}^\top & \boldsymbol{I} \end{bmatrix} \succeq \boldsymbol{0}\right\}.$$

Write the Lagrangian: with dual variable $\begin{bmatrix} \boldsymbol{Y} & \boldsymbol{M} \\ \boldsymbol{M}^\top & \boldsymbol{Y} \end{bmatrix} \succeq \boldsymbol{0}$

$$\begin{aligned} L(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{M}, \boldsymbol{W}) &= \langle -\boldsymbol{X}, \boldsymbol{W} \rangle + \left\langle \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{M} \\ \boldsymbol{M}^\top & \boldsymbol{Z} \end{bmatrix}, \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W} \\ \boldsymbol{W}^\top & \boldsymbol{I} \end{bmatrix} \right\rangle \\ &= \langle -\boldsymbol{X}, \boldsymbol{W} \rangle + 2\langle \boldsymbol{M}, \boldsymbol{W} \rangle + \text{tr}(\boldsymbol{Y}) + \text{tr}(\boldsymbol{Z}) \\ &= \langle 2\boldsymbol{M} - \boldsymbol{X}, \boldsymbol{W} \rangle + \text{tr}(\boldsymbol{Y}) + \text{tr}(\boldsymbol{Z}) \end{aligned}$$

Maximizing L over $\boldsymbol{W}$ to get the dual function:

$$g(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{M}) = \max_{\boldsymbol{W}} L(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{M}, \boldsymbol{W}) = \begin{cases} \text{tr}(\boldsymbol{Y}) + \text{tr}(\boldsymbol{Z}), & 2\boldsymbol{M} = \boldsymbol{X} \\ +\infty, & 2\boldsymbol{M} \neq \boldsymbol{X} \end{cases}$$

Hence, nuclear norm can be written as the dual problem of the dual characterization, i.e.,

$$\begin{aligned} \|\boldsymbol{X}\|_* = \underset{\boldsymbol{Y}, \boldsymbol{Z}}{\text{minimize}} \quad & \tfrac{1}{2}\text{tr}(\boldsymbol{Y} + \boldsymbol{Z}) \\ \text{subject to} \quad & \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{Z} \end{bmatrix} \succeq \boldsymbol{0}. \end{aligned}$$

Hence the original nuclear norm minimization problem can be written as

$$\begin{array}{ll} \underset{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}}{\text{minimize}} & \frac{1}{2}\operatorname{tr}(\boldsymbol{Y}+\boldsymbol{Z}) \\ \text{subject to} & \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{Z} \end{bmatrix} \succeq \mathbf{0} \\ & \boldsymbol{X} \in \mathcal{C}. \end{array}$$

2. Consider the problem of projecting a point $\boldsymbol{a} \in \mathbb{R}^n$ on a set:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \|\boldsymbol{x}-\boldsymbol{a}\|_2 \\ \text{subject to} & \boldsymbol{x} \in \operatorname{conv}\left(\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_m\right), \end{array}$$

where $\operatorname{conv}(\cdot)$ is defined by

$$\operatorname{conv}(\mathcal{E}) = \left\{ \sum_{i=1}^n \lambda_i \boldsymbol{x}_i \mid n \in \mathbb{N}, \sum_{i=1}^n \lambda_i = 1, \lambda_i \geqslant 0 \text{ and } \boldsymbol{x}_i \in \mathcal{E} \text{ for } \forall i \in \{1,\dots,n\} \right\},$$

and $\mathcal{E}_1, \dots, \mathcal{E}_m$ are m ellipsoids in $\mathbb{R}^n$ defined as

$$\mathcal{E}_i = \left\{ \boldsymbol{A}_i \boldsymbol{u} + \boldsymbol{b}_i \mid \|\boldsymbol{u}\|_2 \le 1 \right\}, \quad i = 1, \dots, m,$$

with $\boldsymbol{A}_i \in \mathbb{R}^{n\times n}$ and $\boldsymbol{b}_i \in \mathbb{R}^n$. (15 points)

Solution:

Formulate this as a second order cone program. First express the problem as

$$\begin{array}{ll} \underset{\boldsymbol{x},\{\theta_i\}_{i=1}^m,\{\boldsymbol{u}_i\}_{i=1}^m}{\text{minimize}} & \|\boldsymbol{x}-\boldsymbol{a}\|_2 \\ \text{subject to} & \boldsymbol{x} = \sum_{i=1}^m \theta_i \left(\boldsymbol{A}_i \boldsymbol{u}_i + \boldsymbol{b}_i\right) \\ & \theta_i \ge 0, \quad i = 1, \dots, m \\ & \sum_{i=1}^m \theta_i = 1 \\ & \|\boldsymbol{u}_i\|_2 \le 1, \quad i = 1, \dots, m. \end{array}$$

The variables are $\boldsymbol{x}, \theta_1, \dots, \theta_m, \boldsymbol{u}_1, \dots, \boldsymbol{u}_m$. This problem is not convex because of the products $\theta_i \boldsymbol{u}_i$. It becomes convex if we make a change of variables $\theta_i \boldsymbol{u}_i = \boldsymbol{y}_i$:

$$\begin{array}{ll} \underset{\boldsymbol{x},\{\theta_i\}_{i=1}^m,\{\boldsymbol{y}_i\}_{i=1}^m}{\text{minimize}} & \|\boldsymbol{x}-\boldsymbol{a}\|_2 \\ \text{subject to} & \boldsymbol{x} = \sum_{i=1}^m \left(\boldsymbol{A}_i \boldsymbol{y}_i + \theta_i \boldsymbol{b}_i\right) \\ & \theta_i \ge 0, \quad i = 1, \dots, m \\ & \sum_{i=1}^m \theta_i = 1 \\ & \|\boldsymbol{y}_i\|_2 \le \theta_i, \quad i = 1, \dots, m. \end{array}$$

To cast in the standard SOCP form, we also need to make the objective linear by introducing a variable t :

$$\begin{array}{ll} \underset{\boldsymbol{x},t,\{\theta_i\}_{i=1}^m,\{\boldsymbol{y}_i\}_{i=1}^m}{\text{minimize}} & t \\ \text{subject to} & \|\boldsymbol{x}-\boldsymbol{a}\|_2 \le t \\ & \boldsymbol{x} = \sum_{i=1}^m \left(\boldsymbol{A}_i \boldsymbol{y}_i + \theta_i \boldsymbol{b}_i\right) \\ & \theta_i \ge 0, \quad i = 1, \dots, m \\ & \sum_{i=1}^m \theta_i = 1 \\ & \|\boldsymbol{y}_i\|_2 \le \theta_i, \quad i = 1, \dots, m. \end{array}$$

III. Lagrange Duality

Find the Lagrange dual problem of the following problems.

1. Consider the standard conic programming problem

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \langle \boldsymbol{c}, \boldsymbol{x} \rangle \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \in \mathcal{K}, \end{array}$$

where $\boldsymbol{c} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^m, \boldsymbol{A} \in \mathbb{R}^{n\times m}, \langle \cdot, \cdot \rangle$ denotes the standard Euclidean inner product and $\mathcal{K} \subset \mathbb{R}^n$ is a convex cone. (10 points)

Solution:
We will treat the problem as a constrained optimization problem in order to derive the dual. Let $\boldsymbol{y}$ be the Lagrange multiplier associated with the constraint $\boldsymbol{Ax} = \boldsymbol{b}$. Then the Lagrangian is

$$\begin{aligned} L(\boldsymbol{y}) &= \inf_{\boldsymbol{x}\in\mathcal{K}} \langle \boldsymbol{c}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{b} - \boldsymbol{Ax}\rangle \\ &= \langle \boldsymbol{y}, \boldsymbol{b}\rangle + \inf_{\boldsymbol{x}\in\mathcal{K}} [\langle \boldsymbol{c}, \boldsymbol{x}\rangle - \langle \boldsymbol{y}, \boldsymbol{Ax}\rangle] \\ &= \langle \boldsymbol{y}, \boldsymbol{b}\rangle + \inf_{\boldsymbol{x}\in\mathcal{K}} \left\langle \boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{y}, \boldsymbol{x}\right\rangle \end{aligned}$$

If $\boldsymbol{c} - \boldsymbol{A}^\top\boldsymbol{y} \in \mathcal{K}^*$, then $\left\langle \boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{y}, \boldsymbol{x}\right\rangle \geq 0$, so $\inf_{\boldsymbol{x}\in\mathcal{K}} \left\langle \boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{y}, \boldsymbol{x}\right\rangle = 0$. If $\boldsymbol{c} - \boldsymbol{A}^\top\boldsymbol{y} \notin \mathcal{K}^*$, then $\inf_{\boldsymbol{x}\in\mathcal{K}} \left\langle \boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{y}, \boldsymbol{x}\right\rangle = -\infty$. Thus, $L(\boldsymbol{y})$ is given by

$$L(\boldsymbol{y}) = \begin{cases} \langle \boldsymbol{y}, \boldsymbol{b}\rangle, & \boldsymbol{c} - \boldsymbol{A}^\top\boldsymbol{y} \in \mathcal{K}^*, \\ -\infty, & \boldsymbol{c} - \boldsymbol{A}^\top\boldsymbol{y} \notin \mathcal{K}^*. \end{cases}$$

Let $\boldsymbol{s} = \boldsymbol{c} - \boldsymbol{A}^\top\boldsymbol{y} \in \mathcal{K}^*$. Then the Lagrangian dual problem is to maximise $L(\boldsymbol{y})$, which can be written as the given form:

$$\begin{aligned} &\underset{\boldsymbol{y}\in\mathbb{R}^m}{\text{maximize}} \;\; \langle \boldsymbol{b}, \boldsymbol{y}\rangle \\ &\text{subject to } \boldsymbol{A}^\top\boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c} \\ &\qquad\qquad\quad \boldsymbol{s} \in \mathcal{K}^*. \end{aligned}$$

2. Consider the following relative entropy minimization problem

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \sum_{k=1}^n x_k \log(x_k/y_k) \\ \text{subject to} & \boldsymbol{Ax} = \boldsymbol{b} \\ & \mathbf{1}^\top \boldsymbol{x} = 1, \end{array}$$

where the parameters $\boldsymbol{y} \in \mathbb{R}^n_{++}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, and $\boldsymbol{b} \in \mathbb{R}^m$ are given. (15 points)
Solution:
The Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{z}, \mu) = \sum_k x_k \log(x_k/y_k) + \boldsymbol{b}^\top\boldsymbol{z} - \boldsymbol{z}^\top\boldsymbol{Ax} + \mu - \mu\mathbf{1}^\top\boldsymbol{x}.$$

Minimizing over x_k gives the conditions

$$1 + \log(x_k/y_k) - \boldsymbol{a}_k^\top\boldsymbol{z} - \mu = 0, \quad k = 1, \ldots, n,$$

where $\boldsymbol{a}_k$ is the k th column of $\boldsymbol{A}$, with solution

$$\boldsymbol{x}_k = \boldsymbol{y}_k e^{\boldsymbol{a}_k^\top\boldsymbol{z}+\mu-1}.$$

Plugging this into L gives the Lagrange dual function

$$g(\boldsymbol{z}, \mu) = \boldsymbol{b}^\top\boldsymbol{z} + \mu - \sum_{k=1}^n y_k e^{\boldsymbol{a}_k^\top\boldsymbol{z}+\mu-1}$$

and the dual problem

$$\underset{\boldsymbol{z},\mu}{\text{maximize}} \; \boldsymbol{b}^\top\boldsymbol{z} + \mu - \sum_{k=1}^n y_k e^{\boldsymbol{a}_k^\top\boldsymbol{z}+\mu-1}.$$

This can be simplified a bit if we optimize over μ by setting the derivative equal to zero:

$$\mu = 1 - \log\sum_{k=1}^n y_k e^{\boldsymbol{a}_k^\top\boldsymbol{z}}.$$

After this simplification the dual problem, i.e.,

$$\underset{\boldsymbol{z}}{\text{maximize}} \quad \boldsymbol{b}^\top\boldsymbol{z} - \log\sum_{k=1}^n y_k e^{\boldsymbol{a}_k^\top\boldsymbol{z}}.$$

IV. **KKT Conditions**

1. Consider the following problem of projection onto unit ℓ_2-norm ball,

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \|\boldsymbol{x}-\boldsymbol{a}\|_2^2 \\ \text{subject to} & \|\boldsymbol{x}\|_2^2 \leq 1, \end{array}$$

where $\boldsymbol{a} \in \mathbb{R}^n$ is given.

1. Derive the KKT conditions for this problem. (10 points)
2. Solve the KKT system and obtain the optimal solution in closed form. (10 points)

Solution:

1. First, write the Lagrangian

$$L(\boldsymbol{x}, \lambda) = \|\boldsymbol{x}-\boldsymbol{a}\|_2^2 + \lambda\left(\|\boldsymbol{x}\|_2^2 - 1\right).$$

Then derive the KKT conditions:

* Stationarity $\nabla_{\boldsymbol{x}} L = 2(\boldsymbol{x}-\boldsymbol{a}) + 2\lambda\boldsymbol{x} = 0$
* Slackness $\lambda\left(\|\boldsymbol{x}\|_2^2 - 1\right) = 0$
* Primal feasibility $\|\boldsymbol{x}\|_2^2 - 1 \leq 0$
* Dual feasibility $\lambda \geq 0$

Simplifying the KKT conditions gives

(1) $(1+\lambda)\boldsymbol{x} = \boldsymbol{a}$

(2) $\lambda\left(\|\boldsymbol{x}\|_2^2 - 1\right) = 0$

(3) $\|\boldsymbol{x}\|_2^2 - 1 \leq 0$

(4) $\lambda \geq 0$

2. Conclusion: $\boldsymbol{x}^* = \frac{\boldsymbol{a}}{\max\left\{1, \|\boldsymbol{a}\|_2^2\right\}}$.

As we want to eliminate λ, we start with (4). We have two cases: $\lambda = 0$ or $\lambda > 0$.

(a) Cases $\lambda = 0$

(1) gives $\boldsymbol{x} = \boldsymbol{a}$ and (3) gives $\|\boldsymbol{x}\|_2^2 = \|\boldsymbol{a}\|_2^2 \leq 1$. This is the case that $\boldsymbol{a}$ inside the ball. Hence, $\boldsymbol{x}^* = \boldsymbol{a}$.

(b) Cases $\lambda > 0$

(2) gives $\|\boldsymbol{x}\|_2^2 = 1$, multiply (1) by $\boldsymbol{x}^\top$, use (2) we have

$$(1+\lambda)\underbrace{\boldsymbol{x}^\top\boldsymbol{x}}_{1} = \boldsymbol{x}^\top\boldsymbol{a} \Longrightarrow \lambda = \boldsymbol{x}^\top\boldsymbol{a} - 1 > 0$$

And $\boldsymbol{x}^\top\boldsymbol{a} - 1 > 0 \Longleftrightarrow \boldsymbol{x}^\top\boldsymbol{a} > 1 \Longleftrightarrow \underbrace{\|\boldsymbol{x}\|_2^2}_{1}\|\boldsymbol{a}\|_2^2 > 1$ so $\|\boldsymbol{a}\|_2^2 > 1$, this is the case $\boldsymbol{a}$ outside S. Furthermore, put $\lambda = \boldsymbol{x}^\top\boldsymbol{a} - 1$ to (1) gives $\left(\boldsymbol{x}^\top\boldsymbol{a}\right)\boldsymbol{x} = \boldsymbol{a}$ meaning $\boldsymbol{x}, \boldsymbol{a}$ are parallel so $\boldsymbol{x} = k\boldsymbol{a}$ for some scalar k, put $\boldsymbol{x} = k\boldsymbol{a}$ into $\left(\boldsymbol{x}^\top\boldsymbol{a}\right)\boldsymbol{x} = \boldsymbol{a}$ gives $k = \frac{1}{\|\boldsymbol{a}\|_2^2}$ so $\boldsymbol{x} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2^2}$.

Combine the 2 cases gives $\boldsymbol{x} = \frac{\boldsymbol{a}}{\max\left\{1, \|\boldsymbol{a}\|_2^2\right\}}$.