Final

1. **Honor Code (0 pts)**

   Please copy the following statement in the space provided below and sign your name.

   As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I will follow the rules and do this exam on my own.

   If you do not copy the honor code and sign your name, you will get a 0 on the exam.

Solution:

2. **SID (3 pts)**

   When the exam starts, write your SID at the top of every page.

   No extra time will be given to complete this task.

3. **Favorites. Any answer, as long as you write it down, will be given full credit. (2 pts)**

   (a) (1 pts) What's something that made you happy this year?

   Solution: Any answer is fine.

   (b) (1 pts) What's your favorite number?

   Solution: Any answer is fine.

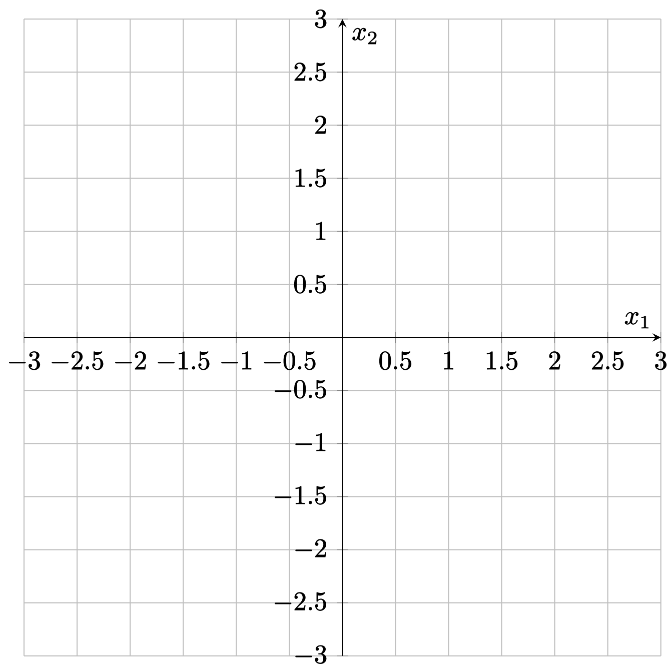
4. Linear Program (12 pts)

Consider the linear program

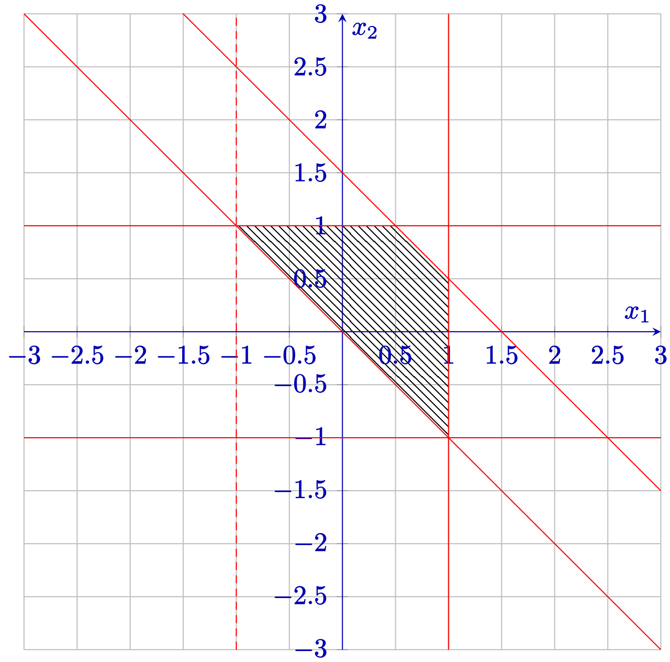
$$\begin{aligned} \min_{\vec{x}\in\mathbb{R}^2} \quad & \begin{bmatrix} 1 \\ -1 \end{bmatrix}^\top \vec{x} \\ \text{s.t.} \quad & \begin{bmatrix} -1 \\ -1 \end{bmatrix} \leq \vec{x} \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ & 0 \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix}^\top \vec{x} \leq 1.5. \end{aligned} \tag{1}$$

where $\vec{x} \doteq \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

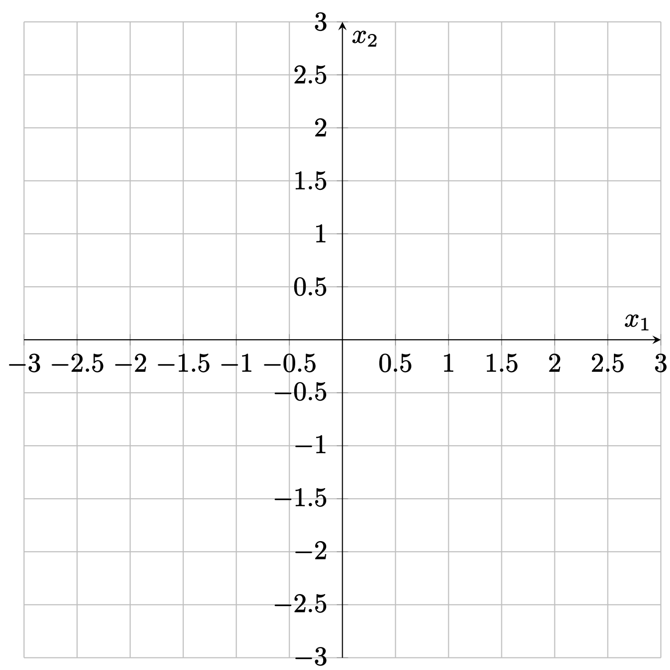
(a) (6 pts) **Draw the constraints on this problem and shade in the feasible region.**



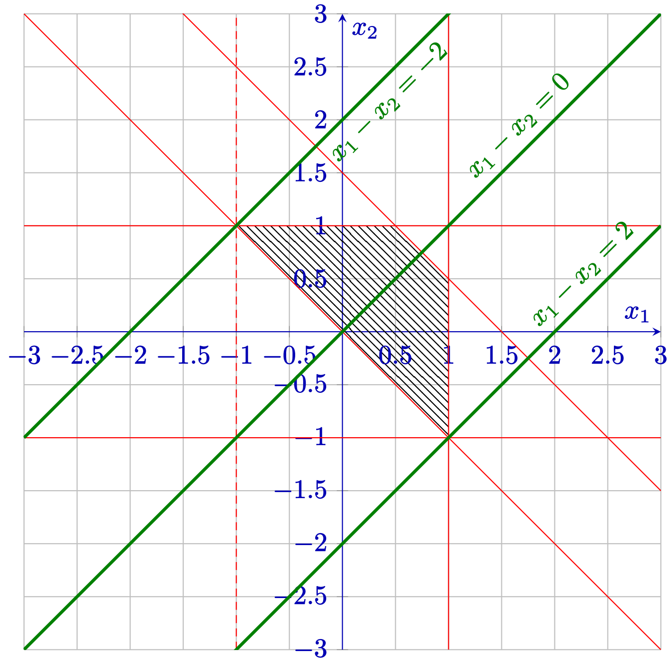
Solution:



(b) (3 pts) **Plot and label level sets of the objective**, i.e., $\begin{bmatrix} 1 \\ -1 \end{bmatrix}^\top \vec{x} = k$ for $k = \{-2, 0, 2\}$ on the figure below.



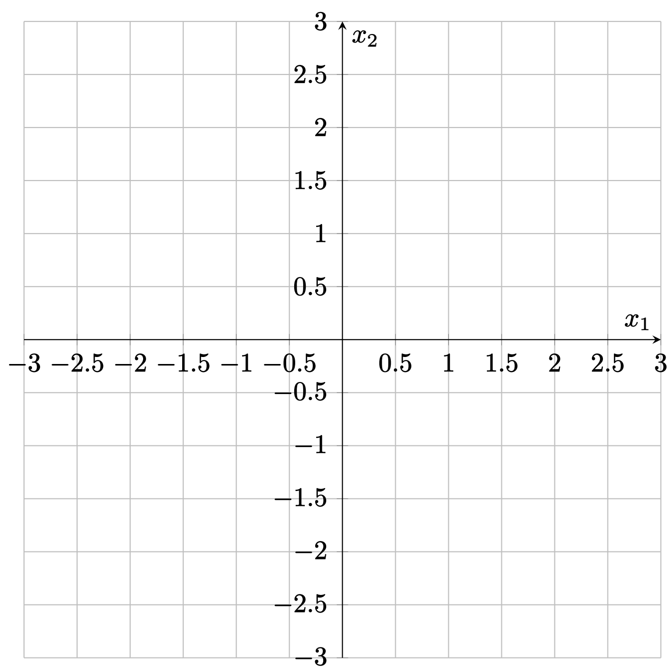
Solution:



(c) (3 pts) **Identify the optimal value $p^\star$ for the problem (1) and the vector $\vec{x}^\star$ which achieves it**. *You do not need to justify your answer.*

What are the active constraints at the optimal solution? *You do not need to justify your answer.*

NOTE: It may be helpful to draw the level sets and constraints on one plot. This plot below will not be graded; it is just there for your convenience.



Solution: Because the -2-level set only intersects the feasible set at one point, and no k-level set intersects the feasible set for $k < -2$, we have that $p^\star = -2$ and the optimal $\vec{x}^\star$ is the point of intersection of the -2-level set and the feasible set. This turns out to be $\vec{x}^\star = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

At this point, the active constraints are those that are met with equality. These turn out to be

- $x_1 \geq -1$;
- $x_2 \leq 1$;
- $x_1 + x_2 \geq 0$.

5. Weak vs Strong Duality (13 pts)

Consider the convex problem

$$p^\star = \min_{\vec{x}\in\mathbb{R}^2} \quad \frac{1}{2}(x_1+1)^2 + x_2^2 \qquad (2)$$
$$\text{s.t.} \quad x_1 = 0.$$

(a) (2 pts) **Find the primal optimum $p^\star$ in problem (2) by substituting the constraint $x_1 = 0$ into the objective function.** *You do not need to justify your answer.*

Solution: Substituting the constraint in we hav

$$p^\star = \min_{x_2\in\mathbb{R}} \left(\frac{1}{2}(0+1)^2 + x_2^2\right) = \frac{1}{2} + \min_{x_2\in\mathbb{R}} x_2^2 = \frac{1}{2}, \qquad (3)$$

with $\vec{x}^\star = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

(b) (3 pts) **Does Slater's condition hold for problem (2)? Does strong duality hold?** *Justify your answer.*

Solution: Yes, since the objective function is a convex function, and the single constraint is an affine equality constraint, Slater's condition must hold. Thus strong duality must hold.

(c) (8 pts) **Find the dual function $g(\nu)$ and the dual optimum $d^\star = \max_{\nu\in\mathbb{R}} g(\nu)$.** *Show your work.*

Solution: The Lagrangian is

$$L(\vec{x},\nu) = \frac{1}{2}(x_1+1)^2 + x_2^2 + \nu x_1. \qquad (4)$$

The dual function takes the form:

$$g(\nu) = \min_{\vec{x}\in\mathbb{R}^2} L(\vec{x},\nu). \qquad (5)$$

Since Equation (5) is a minimization problem whose objective function $L(\cdot,\nu)$ is convex in $\vec{x}$, we can find $\vec{x}^\star(\nu)$ by setting its gradient to 0. In particular, we have

$$\vec{0} = \nabla_{\vec{x}} L(\vec{x}^\star(\nu),\nu) \qquad (6)$$
$$= \begin{bmatrix} x_1^\star(\nu) + 1 + \nu \\ 2x_2^\star(\nu) \end{bmatrix} \qquad (7)$$
$$\implies \vec{x}^\star(\nu) = \begin{bmatrix} -1-\nu \\ 0 \end{bmatrix}. \qquad (8)$$

Thus

$$g(\nu) = L(\vec{x}^\star(\nu),\nu) \qquad (9)$$
$$= -\frac{1}{2}\nu^2 - \nu. \qquad (10)$$

To find $g^\star$, we need to solve the dual problem

$$g^\star = \max_{\nu\in\mathbb{R}} g(\nu) \qquad (11)$$

$$= \max_{\nu \in \mathbb{R}} \left(-\frac{1}{2}\nu^2 - \nu \right) \tag{12}$$

Since Equation (11) is a maximization problem and the objective function is concave in ν, we can find $\nu^\star$ by setting its gradient to 0. In particular, we have

$$\begin{aligned} 0 &= \nabla_\nu g(\nu^\star) \\ &= -\nu^\star - 1 \\ \implies \nu^\star &= -1. \end{aligned}$$

Thus

$$d^\star = g(\nu^\star) = \frac{1}{2}. \tag{13}$$

6. Transformations (12 pts)

For each of the below problems, assume:

- $\vec{x} \in \mathbb{R}^n$;
- $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$;
- $\mathcal{X} \subset \mathbb{R}^n$.

Shade in or circle "True" if the statement is **always** true. Otherwise, shade in or circle "False". *Ensure that the option you select is clear. You do not need to justify your answer. No partial credit will be awarded.*

(a) (3 pts) Suppose $\max_{\vec{x} \in \mathcal{X}} f(\vec{x}) < \infty$.

$$\max_{\vec{x} \in \mathcal{X}} f(\vec{x}) = -\left[\min_{\vec{x} \in \mathcal{X}} -f(\vec{x})\right]. \tag{14}$$

◯ True

◯ False

Solution: True.

(b) (3 pts) Suppose $\Omega \subseteq \mathcal{X}$, i.e., Ω is a subset of $\mathcal{X}$.

$$\max_{\vec{x} \in \mathcal{X}} f(\vec{x}) \leq \max_{\vec{x} \in \Omega} f(\vec{x}). \tag{15}$$

◯ True

◯ False

Solution: False, the relaxed problem will always achieve a solution at least as optimal as the constrained problem, so

$$\max_{\vec{x} \in \mathcal{X}} f(\vec{x}) \geq \max_{\vec{x} \in \Omega} f(\vec{x}). \tag{16}$$

(c) (3 pts) Suppose $\max_{\vec{x} \in \mathcal{X}} f(\vec{x}) < \infty$, $\max_{\vec{x} \in \mathcal{X}} g(\vec{x}) < \infty$, and both maxima are achieved.

$$\max_{\vec{x} \in \mathcal{X}} \left[f(\vec{x}) + g(\vec{x})\right] \leq \max_{\vec{x} \in \mathcal{X}} f(\vec{x}) + \max_{\vec{x} \in \mathcal{X}} g(\vec{x}). \tag{17}$$

◯ True

◯ False

Solution: True.

(d) (3 pts) Suppose $\max_{\vec{x} \in \mathcal{X}} f(\vec{x}) < \infty$ and the maximum is achieved at a unique maximizer.

$$\operatorname*{argmax}_{\vec{x} \in \mathcal{X}} e^{f(\vec{x})} = \operatorname*{argmax}_{\vec{x} \in \mathcal{X}} f(\vec{x}). \tag{18}$$

◯ True

◯ False

Solution: True, e^x is a monotonically increasing function and composition with a monotonically increasing function preserves order.

7. Low Rank Approximation (3 pts)

Let $A \in \mathbb{R}^{3\times 4}$ be a matrix whose full SVD is

$$A = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \end{bmatrix}}_{V^\top}. \tag{19}$$

Give the best rank-1 approximation to A, i.e., the solution to the problem

$$\operatorname*{argmin}_{\substack{B \in \mathbb{R}^{3\times 4} \\ \mathrm{rk}(B) \leq 1}} \|A - B\|_F^2 . \tag{20}$$

No justification is necessary. No partial credit will be awarded.

NOTE: Please leave your answer in terms of a matrix product.

Solution:

$$B^\star = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \end{bmatrix} \tag{21}$$

$$= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 7 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \end{bmatrix} \tag{22}$$

$$= 7 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \end{bmatrix}. \tag{23}$$

8. SOCP (12 pts)

Consider a matrix $A \in \mathbb{R}^{m\times n}$ and vectors $\vec{b} \in \mathbb{R}^m$, $\vec{c} \in \mathbb{R}^n$ and scalar $d \in \mathbb{R}$. Consider the problem

$$\min_{\vec{z}\in\mathbb{R}^n} (\|A\vec{z} - \vec{b}\|_2 - \vec{c}^\top \vec{z} - d)^2. \tag{24}$$

(a) (8 pts) Suppose $m = 1$ and $n = 1$. Then $\vec{z} = z$ is just a scalar, and $A, \vec{b}, \vec{c}$ are also just scalars. In particular, suppose $A = 1$, $\vec{b} = 1$, $\vec{c} = 1$, and $d = 1$. **For these values, is the optimization problem (24) convex?** *Justify your answer.*

HINT: First, rewrite the problem with the given values. Then, consider evaluating the objective function at $z = 0$ and $z = 2$.

Solution: The SOCP objective with the provided values is the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(z) \doteq (|z - 1| - z - 1)^2. \tag{25}$$

We check whether f is convex.

Let $z_1 = 0$ and $z_2 = 2$. Then

$$f(z_1) = (|z_1 - 1| - z_1 - 1)^2 = (|-1| - 1)^2 = (1 - 1)^2 = 0^2 = 0 \tag{26}$$

$$f(z_2) = (|z_2 - 1| - z_2 - 1)^2 = (|2 - 1| - 2 - 1)^2 = (1 - 2 - 1)^2 = (-2)^2 = 4. \tag{27}$$

Let $\lambda = \frac{1}{2}$. Then

$$f(\lambda z_1 + (1 - \lambda)z_2) = f(1) = (|1 - 1| - 1 - 1)^2 = (-2)^2 = 4. \tag{28}$$

Thus there exists $z_1, z_2 \in \mathbb{R}$ and $\lambda \in [0, 1]$ such that

$$\underbrace{f(\lambda z_1 + (1 - \lambda)z_2)}_{=4} > \underbrace{\lambda f(z_1) + (1 - \lambda)f(z_2)}_{=\frac{1}{2}\cdot 0+\frac{1}{2}\cdot 4=2} \tag{29}$$

which is a direct violation of Jensen's inequality for f. Thus f is not convex.

(b) (4 pts) The problem can be reformulated as

$$\min_{\vec{x}\in\mathbb{R}^{n+1}} \begin{bmatrix}\vec{0}\\ 1\end{bmatrix}^\top \vec{x} \tag{30}$$

$$\text{s.t.} \quad \left\| \begin{bmatrix}A & \vec{0}\end{bmatrix} \vec{x} - \vec{b} \right\|_2 - \begin{bmatrix}\vec{c}\\ 1\end{bmatrix}^\top \vec{x} - d \leq 0 \tag{31}$$

$$\left\| \begin{bmatrix}A & \vec{0}\end{bmatrix} \vec{x} - \vec{b} \right\|_2 - \begin{bmatrix}\vec{c}\\ -1\end{bmatrix}^\top \vec{x} - d \geq 0. \tag{32}$$

where $\vec{0}$ is the all-zeros vector of the appropriate dimension. **Which constraint should be dropped to make the problem an SOCP?** *Justify your answer.*

Solution: We should drop the constraint (32) to make the problem an SOCP in standard form; this is the case because SOCP constraints are affine (i.e., $A\vec{x} = \vec{b}$) or second-order cone (i.e., $\|F\vec{x} - \vec{g}\|_2 - \vec{h}^\top\vec{x} - k \leq 0$).

Note that the original rpoblem was not convex, but the new problem without constraint (32) is convex, so the two problems are not equivalent.

9. Power Purchase (20 pts)

This problem describes a simplified version of the optimization problem on the power grid. Consider a consumer who wants to consume d units (in kilo-Watt-hour (kWh)) of electrical energy. The consumer can purchase this energy from a combination of electricity generators. Consider n generators indexed by $i = 1, 2, ..., n$, and denote the energy output of the i^{th} generator by x_i. The cost of generating x_i units of energy is given by a generator specific cost $f_i(x_i)$:

$$f_i(x_i) = a_i x_i^2 + b_i x_i \tag{33}$$

where $a_i \geq 0$. Each generator has a constraint on the maximum energy it can produce, and this cap is denoted by m_i, i.e., $0 \leq x_i \leq m_i$.

The consumer tries to purchase d units of energy at minimum cost by **optimizing the amount of energy** x_i purchased from each generator by solving:

$$\begin{aligned} \min_{x_1, x_2, \ldots, x_n} \quad & \sum_{i=1}^{n} f_i(x_i) \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = d \\ & 0 \leq x_i \leq m_i \quad i = 1, ..., n. \end{aligned} \tag{34}$$

(a) (3 pts) **Choose the smallest class that problem (34) belongs to (LP/QP/SOCP/etc.).** *You do not need to justify your answer.*

Solution: This problem is a quadratic program (QP), since the objective is a quadratic function of $\vec{x}$, and the constraints are affine functions of $\vec{x}$.

(b) (6 pts) We consider a specific case of the optimization problem (34):

$$\begin{aligned} \min_{x_1, x_2} \quad & \left(x_1^2 + 2x_1\right) + \left(\frac{3}{2}x_2^2\right) \\ \text{s.t.} \quad & x_1 + x_2 = 1 \\ & 0 \leq x_1 \leq 2 \\ & 0 \leq x_2 \leq 1. \end{aligned} \tag{35}$$

This problem is a specific instance of (34) with $n = 2$, $d = 1$, $m_1 = 2$, and $m_2 = 1$. **What are the optimal values of x_1, x_2?** *Show your work.*

HINT: Try eliminating x_2 by replacing it with $1 - x_1$, solve the unconstrained optimization problem and check back to see if the constraints are satisfied.

Solution: As notation, define

$$f(\vec{x}) \doteq x_1^2 + 2x_1 + \frac{3}{2}x_2^2. \tag{36}$$

Replace $x_2 = d - x_1 = 1 - x_1$. With this substitution, the objective function becomes

$$\widetilde{f}(x_1) \doteq f(x_1, 1 - x_1) = \frac{5}{2}x_1^2 - x_1 + \frac{3}{2}. \tag{37}$$

The problem then becomes

$$\min_{x_1 \in \mathbb{R}} \quad \frac{5}{2}x_1^2 - x_1 + \frac{3}{2} \tag{38}$$

$$\begin{aligned}\text{s.t.}\quad & 0 \leq x_1 \leq 2 \\ & 0 \leq 1 - x_1 \leq 1.\end{aligned}$$

The corresponding unconstrained problem is

$$\min_{x_1 \in \mathbb{R}} \widetilde{f}(x_1) = \min_{x_1 \in \mathbb{R}} \left(\frac{5}{2}x_1^2 - x_1 + \frac{3}{2} \right). \tag{39}$$

The objective function is convex, so its minimum occurs when the gradient is 0:

$$0 = \nabla_{x_1} \widetilde{f}(x_1^\star) \tag{40}$$

$$= 5x_1^\star - 1 \tag{41}$$

$$\implies x_1^\star = \frac{1}{5} \tag{42}$$

$$\implies x_2^\star = 1 - x_1^\star = \frac{4}{5}. \tag{43}$$

These values of $(x_1^\star, x_2^\star)$ satisfy the original problem's constraints, so they solve the original constrained problem as well.

(c) (6 pts) Consider a similar problem as in the previous subpart (b), this time with $d = 2$:

$$\begin{aligned}\min_{x_1, x_2}\quad & \left(x_1^2 + 2x_1\right) + \left(\frac{3}{2}x_2^2\right) \\ \text{s.t.}\quad & x_1 + x_2 = 2 \\ & 0 \leq x_1 \leq 2 \\ & 0 \leq x_2 \leq 1.\end{aligned} \tag{44}$$

What are the optimal values of x_1, x_2? *Show your work.*

HINT: Try eliminating x_2 by replacing it with $2 - x_1$. Check the function value at the critical points of the problem, i.e., the points where the gradient is zero or undefined, points on the boundaries, and $\pm\infty$.

Solution: We used this idea of checking critical points in Discussion 9.

Again, define

$$f(\vec{x}) = x_1^2 + 2x_1 + \frac{3}{2}x_2^2. \tag{45}$$

Replace $x_2 = d - x_1 = 2 - x_1$. With this substitution, the objective function becomes

$$\widetilde{f}(\vec{x}) \doteq f(x_1, 2 - x_1) = \frac{5}{2}x_1^2 - 4x_1 + 6 \tag{46}$$

The problem then becomes

$$\begin{aligned}\min_{x_1}\quad & \frac{5}{2}x_1^2 - 4x_1 + 6 \\ \text{s.t.}\quad & 0 \leq x_1 \leq 2 \\ & 0 \leq 2 - x_1 \leq 1.\end{aligned} \tag{47}$$

Combining the constraints gives us the problem

$$\min_{x_1 \in \mathbb{R}} \quad \frac{5}{2}x_1^2 - 4x_1 + 6 \tag{48}$$

$$\text{s.t.} \quad 1 \leq x_1 \leq 2.$$

The objective function $\widetilde{f}$ is convex; setting its to 0 and solving gives us the point $(x_1, x_2) = (\frac{4}{5}, \frac{6}{5})$. However, this violates the constraints.

We now check the function value at the critical points $x_1 \in \{1, 2\}$, which will give us the answer since the minimization problem is convex and the unconstrained minimization solution is infeasible. At $x_1 = 1$ we have $\widetilde{f}(x_1) = \widetilde{f}(1) = \frac{9}{2}$. At $x_1 = 2$ we have $\widetilde{f}(x_1) = \widetilde{f}(2) = 8$. Thus $p^\star = \frac{9}{2}$ and $\vec{x}^\star = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(d) (5 pts) Consider an instance of (34) where $n = 2$, two generators generate energy x_1, x_2 to fulfill demand d, with associated costs $f_1(x_1)$ and $f_2(x_2)$ and capacities m_1, m_2:

$$\begin{aligned} \min_{x_1, x_2} \quad & f_1(x_1) + f_2(x_2) \\ \text{s.t.} \quad & x_1 + x_2 = d \\ & 0 \leq x_i \leq m_i, \quad i = 1, 2. \end{aligned} \tag{49}$$

Consider the dual variables corresponding to the constraints as below:

Dual Variable	Constraint
ν	$x_1 + x_2 = d$
λ_1	$0 \leq x_1$
λ_2	$x_1 \leq m_1$
λ_3	$0 \leq x_2$
λ_4	$x_2 \leq m_2$

Write the complementary slackness conditions for this problem.

Suppose you solve the above problem (49), and find that your solution is such that $0 < x_1^\star < m_1$ and $0 < x_2^\star = m_2$. **Which dual variables are necessarily equal to zero?**

You may assume (without needing to prove) that strong duality holds.

Solution: The complementary slackness conditions are:

$$0 = \lambda_1^\star(-x_1^\star) \tag{50}$$

$$0 = \lambda_2^\star(x_1^\star - m_1) \tag{51}$$

$$0 = \lambda_3^\star(-x_2^\star) \tag{52}$$

$$0 = \lambda_4^\star(x_2^\star - m_2). \tag{53}$$

Since strong duality holds for this problem, the KKT conditions hold for $(\vec{x}^\star, \vec{\lambda}^\star, \nu^\star)$ where $\vec{x}^\star$ is optimal for the primal problem (49) and $\vec{\lambda}^\star$ is optimal for the dual problem of (49).

By complementary slackness, we must have:

- $-x_1^\star < 0$ so $\lambda_1^\star = 0$.
- $x_1^\star - m_1 < 0$ so $\lambda_2^\star = 0$.
- $-x_2^\star < 0$ so $\lambda_3^\star = 0$.

- $x_2^\star - m_2 = 0$ so it is *not necessarily* true that $\lambda_4^\star = 0$.

Finally, we remark that there are no slackness conditions for $\nu^\star$, so it is not necessarily true that $\nu^\star = 0$.

10. Proximal Operator (8 pts)

For a function $h\colon \mathbb{R}^n \to \mathbb{R}$, define its *proximal operator* $\mathsf{prox}_h\colon \mathbb{R}^n \to \mathbb{R}^n$ as

$$\mathsf{prox}_h(\vec{x}) \doteq \operatorname*{argmin}_{\vec{u}\in\mathbb{R}^n} \left(h(\vec{u}) + \frac{1}{2}\|\vec{x}-\vec{u}\|_2^2 \right). \tag{54}$$

Consider a special case where $n = 1$, i.e., we have a scalar problem, and $h(x) \doteq \lambda|x|$ for a constant $\lambda \geq 0$. Let $x_0 \in \mathbb{R}$. Prove that

$$\mathsf{prox}_h(x_0) \doteq \operatorname*{argmin}_{u\in\mathbb{R}} \left(\lambda|u| + \frac{1}{2}(x_0-u)^2 \right) \tag{55}$$

is the soft-thresholding function that we've seen in the context of the LASSO problem, i.e.,

$$\operatorname*{argmin}_{u\in\mathbb{R}} \left(\lambda|u| + \frac{1}{2}(x_0-u)^2 \right) = \begin{cases} x_0+\lambda, & \text{if } x_0 < -\lambda \\ 0, & \text{if } -\lambda \leq x_0 \leq \lambda \\ x_0-\lambda, & \text{if } x_0 > \lambda. \end{cases} \tag{56}$$

Solution: We aim to solve the problem

$$\min_{u\in\mathbb{R}} \left(\lambda|u| + \frac{1}{2}(x_0-u)^2 \right). \tag{57}$$

For notational convenience, define $f, f_+, f_- \colon \mathbb{R} \to \mathbb{R}$ by

$$f_+(u) \doteq \lambda u + \frac{1}{2}(x_0-u)^2 \tag{58}$$

$$f_-(u) \doteq -\lambda u + \frac{1}{2}(x_0-u)^2 \tag{59}$$

$$f(u) \doteq \begin{cases} f_+(u) & u \geq 0 \\ f_-(u) & u < 0. \end{cases} \tag{60}$$

Then $f(u)$ is the objective function of the proximal minimization problem. Note that f is not differentiable at $u = 0$, but is differentiable everywhere else.

Suppose $u^\star > 0$. Then the first-order condition for optimality reads

$$0 = \nabla_u f(u^\star) \tag{61}$$
$$= \nabla_u f_+(u^\star) \tag{62}$$
$$= \lambda - (u^\star - x_0) \tag{63}$$
$$\implies u^\star = x_0 - \lambda. \tag{64}$$

This shows that if $u^\star > 0$ then $u^\star = x_0 - \lambda$. Thus $u^\star > 0$ if and only if $x_0 > \lambda$.

Now suppose $u^\star < 0$. Then the first-order condition for optimality reads

$$0 = \nabla_u f(u^\star) \tag{65}$$
$$= \nabla_u f_-(u^\star) \tag{66}$$
$$= -\lambda - (u^\star - x_0) \tag{67}$$
$$\implies u^\star = x_0 + \lambda. \tag{68}$$

This shows that if $u^\star < 0$ then $u^\star = x_0 + \lambda$. Thus $u^\star < 0$ if and only if $x_0 < -\lambda$.

If $u^\star \neq 0$ then $|x_0| > \lambda$, so as a contrapositive, if $|x_0| \leq \lambda$ then $u^\star = 0$. Thus we have

$$\mathsf{prox}_h(x_0) = u^\star = \begin{cases} x_0 + \lambda, & \text{if } x_0 < -\lambda \\ 0, & \text{if } -\lambda \leq x_0 \leq \lambda \\ x_0 - \lambda, & \text{if } x_0 > \lambda, \end{cases} \tag{69}$$

as desired.

11. Modified SVM (31 pts)

Let $Z \in \mathbb{R}^{n \times d}$ be a constant known matrix, and let $C > 0$ be a scalar. Consider the problem

$$\begin{aligned} p^\star = \min_{\vec{w}, \vec{s}} \quad & \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \|\vec{s}\|_2^2 \\ \text{s.t.} \quad & \vec{s} \geq \vec{0} \\ & \vec{s} \geq \vec{1} - Z\vec{w}, \end{aligned} \tag{70}$$

where $\vec{w} \in \mathbb{R}^d$ and $\vec{s} \in \mathbb{R}^n$ are the optimization variables, $\vec{0} = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^\top \in \mathbb{R}^n$ is the vector of all zeros, $\vec{1} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^\top \in \mathbb{R}^n$ is the vector of all ones. Strong duality holds for this problem.

(a) (3 pts) **Choose the smallest class that problem (70) belongs to (LP/QP/SOCP/etc).** *You do not need to justify your answer.*

Solution: It is a QP – it has a quadratic objective and affine constraints.

(b) (4 pts) **Are the KKT conditions for problem (70) necessary, sufficient or both necessary and sufficient for global optimality?**

NOTE: You may use (without needing to prove) the fact that strong duality holds.

Solution: The objective function is a convex quadratic and the constraints are affine (hence convex) in $\vec{w}$ and $\vec{s}$, so the problem is convex.

Since the problem is convex, all functions involved are continuously differentiable, and strong duality holds, the KKT conditions are both necessary and sufficient for optimality; that is, they are equivalent to optimality conditions.

(c) (2 pts) **Let $\vec{\alpha}$ be the dual variable corresponding to the constraint $\vec{s} \geq \vec{0}$. What is the dimension (i.e., number of entries) of $\vec{\alpha}$?** *You do not need to justify your answer.*

Solution: $\vec{\alpha} \in \mathbb{R}^n$ since $\vec{s} \in \mathbb{R}^n$.

(d) (4 pts) **Show that the Lagrangian $L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta})$ of problem (70),** where $\vec{\alpha}$ is the dual variable corresponding to the constraint $\vec{s} \geq \vec{0}$, and $\vec{\beta}$ is the dual variable corresponding to the constraint $\vec{s} \geq \vec{1} - Z\vec{w}$, **is equal to**

$$L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta}) = \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \|\vec{s}\|_2^2 - \vec{s}^\top (\vec{\alpha} + \vec{\beta}) - \vec{w}^\top Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta}. \tag{71}$$

Solution: We have

$$\begin{aligned} L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta}) &= \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \|\vec{s}\|_2^2 + \vec{\alpha}^\top (-\vec{s}) + \vec{\beta}^\top (\vec{1} - Z\vec{w} - \vec{s}) & (72) \\ &= \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \|\vec{s}\|_2^2 - \vec{s}^\top (\vec{\alpha} + \vec{\beta}) - \vec{w}^\top Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta}. & (73) \end{aligned}$$

(e) (8 pts) **Write the KKT conditions for problem (70). Show that if $(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star)$ obey the KKT conditions for problem (70), then**

$$\vec{w}^\star = Z^\top \vec{\beta}^\star \qquad \text{and} \qquad \vec{s}^\star = \frac{\vec{\alpha}^\star + \vec{\beta}^\star}{C}. \tag{74}$$

HINT: For the first order/stationarity condition on the Lagrangian you will need to consider partial derivatives with respect to both $\vec{w}$ and $\vec{s}$.

Solution: Let $(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star)$ satisfy the KKT conditions. We have:

- Primal feasibility: $\vec{s}^\star \geq 0$ and $\vec{s}^\star \geq \vec{1} - Z\vec{w}^\star$.
- Dual feasibility: $\vec{\alpha}^\star \geq \vec{0}$, $\vec{\beta}^\star \geq \vec{0}$.
- Complementary slackness: $\alpha_i^\star s_i^\star = 0$ and $\beta_i^\star(1 - \vec{z}_i^\top \vec{w}^\star - s_i^\star) = 0$ for each i.
- Stationarity: $\nabla_{\vec{w}} L(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star) = \vec{0}$ and $\nabla_{\vec{s}} L(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star) = \vec{0}$. These become

$$\vec{0} = \vec{w}^\star - Z^\top \vec{\beta}^\star \tag{75}$$
$$\vec{0} = C\vec{s}^\star - (\vec{\alpha}^\star + \vec{\beta}^\star) \tag{76}$$

which rearrange to the claimed equalities.

(f) (5 pts) **Compute the dual function of problem (70) as**

$$g(\vec{\alpha}, \vec{\beta}) \doteq L(\vec{w}^\star(\vec{\alpha}, \vec{\beta}), \vec{s}^\star(\vec{\alpha}, \vec{\beta}), \vec{\alpha}, \vec{\beta}) \tag{77}$$

where from the previous part we have that

$$\vec{w}^\star(\vec{\alpha}, \vec{\beta}) = Z^\top \vec{\beta} \qquad \text{and} \qquad \vec{s}^\star(\vec{\alpha}, \vec{\beta}) = \frac{\vec{\alpha} + \vec{\beta}}{C}. \tag{78}$$

Your final expression for $g(\vec{\alpha}, \vec{\beta})$ should not contain any maximizations, minimizations or terms including $\vec{w}$, $\vec{s}$, $\vec{w}^\star$, or $\vec{s}^\star$. It should only contain $\vec{\alpha}$, $\vec{\beta}$, C, Z, and numerical constants. *Show your work.*

Solution: The dual function is

$$\begin{aligned}
g(\vec{\alpha}, \vec{\beta}) &= L(\vec{w}^\star(\vec{\alpha}, \vec{\beta}), \vec{s}^\star(\vec{\alpha}, \vec{\beta}), \vec{\alpha}, \vec{\beta}) && (79)\\
&= \frac{1}{2}\left\|\vec{w}^\star(\vec{\alpha}, \vec{\beta})\right\|_2^2 + \frac{C}{2}\left\|\vec{s}^\star(\vec{\alpha}, \vec{\beta})\right\|_2^2 - \vec{s}^\star(\vec{\alpha}, \vec{\beta})^\top(\vec{\alpha} + \vec{\beta}) - \vec{w}^\star(\vec{\alpha}, \vec{\beta})^\top Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta} && (80)\\
&= \frac{1}{2}\left\|Z^\top \vec{\beta}\right\|_2^2 + \frac{C}{2}\left\|\frac{\vec{\alpha} + \vec{\beta}}{C}\right\|_2^2 - \left(\frac{\vec{\alpha} + \vec{\beta}}{C}\right)^\top (\vec{\alpha} + \vec{\beta}) - \vec{\beta}^\top Z Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta} && (81)\\
&= -\frac{1}{2}\vec{\beta}^\top Z Z^\top \vec{\beta} - \frac{1}{2C}\left\|\vec{\alpha} + \vec{\beta}\right\|_2^2 + \vec{1}^\top \vec{\beta}. && (82)
\end{aligned}$$

(g) (5 pts) Let $\vec{\alpha}^\star$ and $\vec{\beta}^\star$ be optimal dual variables that solve the problem

$$d^\star \doteq \max_{\vec{\alpha}, \vec{\beta} \geq \vec{0}} g(\vec{\alpha}, \vec{\beta}). \tag{83}$$

It turns out that $\vec{\alpha}^\star$ can also be obtained by solving the quadratic program:

$$\begin{aligned}
\min_{\vec{\alpha}} \quad & \left\|\vec{\alpha} + \vec{\beta}^\star\right\|_2^2 \\
\text{s.t.} \quad & \vec{\alpha} \geq \vec{0}.
\end{aligned} \tag{84}$$

Solve this quadratic program (84) directly and find $\vec{\alpha}^\star$. *Show your work.*

HINT: The duality or KKT approaches are not recommended. Consider $\vec{\alpha} = \begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^\top$, and use the components of $\vec{\alpha}$ to decompose the problem into n separate scalar problems. Solve *each one by checking critical points; that is, points where the gradient is* 0*, the boundary of the feasible set, and* $\pm\infty$*.*

Solution: We have that

$$\left\|\vec{\alpha} + \vec{\beta}^{\star}\right\|_2^2 = \sum_{i=1}^{n}(\alpha_i + \beta_i^{\star})^2. \tag{85}$$

Also, the $\vec{\alpha} \geq \vec{0}$ constraint is n separate constraints of the form $\alpha_i \geq 0$. Thus, we can solve for each α_i separately as

$$\alpha_i^{\star} \in \operatorname*{argmin}_{\alpha_i \geq 0}(\alpha_i + \beta_i^{\star})^2. \tag{86}$$

This problem is convex and so we can solve it by checking the critical points.

- The gradient (w.r.t. α_i) is 0 if and only if $\alpha_i = -\beta_i^{\star}$. If $\beta_i^{\star} > 0$ then this solution is infeasible, and if $\beta_i^{\star} = 0$ then $\alpha_i = 0$.
- The constraint boundary is $\alpha_i = 0$; this solution is feasible with objective value $(\beta_i^{\star})^2$.
- The limit $\alpha_i \to +\infty$ makes the objective value arbitrarily large, much larger than $(\beta_i^{\star})^2$. The limit $\alpha_i \to -\infty$ makes the solution infeasible.

Thus the optimal solution for each scalar problem is $\alpha_i^{\star} = 0$. Thus $\vec{\alpha}^{\star} = \vec{0}$.

12. Levenberg-Marquardt Regularization for Newton's Method (26 pts)

Newton's method often suffers from non-invertibility of the Hessian $H(\vec{x}) = \nabla^2 f(\vec{x})$. One solution is to use the modified Hessian $H(\vec{x}) + \mu I$, which is known as the *Levenberg-Marquardt regularized Hessian.* In this problem, we explore how to find an appropriate μ given certain conditions on f. This problem doesn't depend on an understanding of Newton's method, other than the first part.

(a) (3 pts) Let $f: \mathbb{R}^n \to \mathbb{R}$ be twice differentiable with Hessian $H(\vec{x})$. Let $\vec{x}_k$ be the k^{th} iterate of Newton's method on f. **Write the Newton's method step for $\vec{x}_{k+1}$ in terms of $\vec{x}_k$.** You may assume $H(\vec{x}_k)$ is invertible.

Solution:

$$\vec{x}_{k+1} = \vec{x}_k - [H(\vec{x}_k)]^{-1}\nabla f(\vec{x}_k). \tag{87}$$

(b) (5 pts) **Show that for any symmetric matrix $A \in \mathbb{S}^n$ with $\lambda_i\{A\}$ as the i^{th} largest eigenvalue of A:**

$$\|A\|_F^2 = \sum_{i=1}^{n} \lambda_i\{A\}^2. \tag{88}$$

HINT: Consider using the eigendecomposition or SVD of A.

Solution: Let $A = U\Lambda U^\top$ be a spectral decomposition of A. Then

$$\|A\|_F^2 = \left\|U\Lambda U^\top\right\|_F^2 \tag{89}$$

$$= \|\Lambda\|_F^2 \tag{90}$$

$$= \sum_{i=1}^{n} \lambda_i\{A\}^2, \tag{91}$$

where in the second line we use the invariance of the Frobenius norm under multiplication by orthogonal matrices.

(c) (5 pts) Let $f: \mathbb{R}^n \to \mathbb{R}$ be twice differentiable with Hessian $H(\vec{x})$. Let $M > 0$ be such that $\|H(\vec{x})\|_F^2 \le M^2$ for all $\vec{x} \in \mathbb{R}^n$. **Use the result in part (b) to show that**

$$-M \le \lambda_{\min}\{H(\vec{x})\} \qquad \text{for all } \vec{x} \in \mathbb{R}^n, \tag{92}$$

where we let $\lambda_{\min}\{H(\vec{x})\}$ denote the smallest eigenvalue of the symmetric Hessian matrix $H(\vec{x})$. *HINT: $\lambda_{\min}\{H(\vec{x})\}^2 \le \sum_{i=1}^n \lambda_i\{H(\vec{x})\}^2$.*

Solution: We use the trace relation for the Frobenius norm and (b) to get

$$M^2 \ge \|H(\vec{x})\|_F^2 \tag{93}$$

$$= \sum_{i=1}^{n} \lambda_i\{H(\vec{x})\}^2 \tag{94}$$

$$\ge \lambda_{\min}\{H(\vec{x})\}^2. \tag{95}$$

Taking square roots shows that $\lambda_{\min}\{H(\vec{x})\} \in [-M, M]$, but we only need the lower bound to show the claim.

(d) (5 pts) Even though $H(\vec{x})$ may not be positive definite, we would like to find a small $\mu \geq 0$ such that $H(\vec{x}) + \mu I$ is positive definite for all $\vec{x} \in \mathbb{R}^n$. Fix $\epsilon > 0$ and define the optimization problem:

$$\mu^\star = \min_{\mu \geq 0} \quad \mu \qquad (96)$$
$$\text{s.t.} \quad \lambda_i\{H(\vec{x}) + \mu I\} \geq \epsilon, \quad \text{for all } i \in \{1, \ldots, n\} \text{ and } \vec{x} \in \mathbb{R}^n.$$

Show that the constraints in the above problem (96) are equivalent to

$$\lambda_{\min}\{H(\vec{x})\} \geq -\mu + \epsilon, \qquad \text{for all } \vec{x} \in \mathbb{R}^n. \qquad (97)$$

Solution: For the feasible region, we use the shift property of eigenvalues to get

$$\lambda_i\{H(\vec{x}) + \mu I\} \geq \epsilon \qquad \forall i \quad \forall \vec{x} \qquad (98)$$
$$\iff \lambda_i\{H(\vec{x})\} + \mu \geq \epsilon \qquad \forall i \quad \forall \vec{x} \qquad (99)$$
$$\iff \lambda_i\{H(\vec{x})\} \geq -\mu + \epsilon \qquad \forall i \quad \forall \vec{x} \qquad (100)$$
$$\iff \lambda_{\min}\{H(\vec{x})\} \geq -\mu + \epsilon \qquad (101)$$

where in the last step we use the fact that $\lambda_{\min}$ is the most negative eigenvalue.

(e) (8 pts) From part (d), we consider the optimization problem with $\epsilon > 0$:

$$\mu^\star = \min_{\mu \geq 0} \quad \mu \qquad (102)$$
$$\text{s.t.} \quad \lambda_{\min}\{H(\vec{x})\} \geq -\mu + \epsilon, \quad \text{for all } \vec{x} \in \mathbb{R}^n.$$

In class, you have seen how using slack variables can create an equivalent program:

$$\max_{\vec{x} \in \mathcal{X}} f(\vec{x}) \quad = \quad \min_{c \in \mathbb{R}} \quad c \qquad (103)$$
$$\text{s.t.} \quad f(\vec{x}) \leq c \quad \text{for all } \vec{x} \in \mathcal{X}.$$

Using this equivalence between formulations, solve for $\mu^\star$. You may assume that there exists $\vec{x}_0 \in \mathbb{R}^n$ such that $\lambda_{\min}\{H(\vec{x}_0)\} = -M$, i.e., the lower bound in part (c) is achieved with equality at some point $\vec{x}_0$.

Solution: We rearrange the constraints of problem to get

$$\mu^\star = \min_{\mu \geq 0} \quad \mu \qquad (104)$$
$$\text{s.t.} \quad \mu \geq \epsilon - \lambda_{\min}\{H(\vec{x})\}, \qquad \text{for all } \vec{x} \in \mathbb{R}^n.$$

By the problem statement, this is equivalent to the problem

$$\mu^\star = \max_{\vec{x} \in \mathbb{R}^n} [\epsilon - \lambda_{\min}\{H(\vec{x})\}] = \epsilon - \min_{\vec{x} \in \mathbb{R}^n} \lambda_{\min}\{H(\vec{x})\} = \epsilon - (-M) = \epsilon + M. \qquad (105)$$

With this $\mu^\star$, because it is feasible for problem (96), we have that all eigenvalues of $H(\vec{x}) + \mu^\star I \geq \epsilon$ for all $\vec{x}$. Thus all eigenvalues of $H(\vec{x}) + \mu^\star I$ are positive, and the matrix is symmetric; hence it is invertible, for every $\vec{x} \in \mathbb{R}^n$.

