DATA130026.01 Optimization Assignment 6

Due Time: at the beginning of the class, Apr. 17, 2024
姓名： 雍崔扬
学号： 21307140051

Problem 1

Find a dual problem to the convex problem $(\mathrm{P})\begin{array}{ll}\min & x_1^2+0.5x_2^2+x_1x_2-2x_1-3x_2\\ \text{s.t.} & x_1+x_2\le 2\end{array}$
Find the optimal solutions of both the dual and primal problems.

Solution:

- **(1) Establish the dual problem:**
  The Lagrangian is given by：
  $L(x,\lambda)=x_1^2+0.5x_2^2+x_1x_2-2x_1-3x_2+\lambda(x_1+x_2-2)$
  Since $\nabla^2_{xx}L(x,\lambda)\equiv\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}\succeq 0$, we deduce that $L(x,\lambda)$ is a convex quadratic function.
  By solving $\nabla_x L(x,\lambda)=\begin{bmatrix}2x_1+x_2-2+\lambda\\ x_2+x_1-3+\lambda\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$
  we have $\begin{cases}x_1=-1\\ x_2=4-\lambda\end{cases}$
  Thus, the dual function is given by:
  $$\begin{aligned}d(\lambda)&=\inf_{x\in\mathbb{R}^2}L(x,\lambda)\\ &=L(x,\lambda)|_{x=(-1,4-\lambda)}\\ &=1+\frac{1}{2}(4-\lambda)^2-(4-\lambda)+2-3(4-\lambda)+\lambda(1-\lambda)\\ &=-\frac{1}{2}\lambda^2+\lambda-5\end{aligned}$$
  Therefore the dual problem is：
  $$(\mathrm{D})\begin{array}{ll}\max & -\frac{1}{2}\lambda^2+\lambda-5\\ \text{s.t.} & \lambda\ge 0\end{array}$$
- **(2) Find the optimal solutions of both the dual and primal problems:**
  Given that $d(\lambda)=-\frac{1}{2}\lambda^2+\lambda-5=-\frac{1}{2}(\lambda-1)^2-\frac{9}{2}$
  it is easy to see that $\begin{cases}\lambda^\star=1\\ d^\star=d(\lambda^\star)=-\frac{9}{2}\end{cases}$
  Note that the primal problem is **convex** and the **Slater Condition** holds,
  hence the **strong duality** holds, i.e., $\begin{cases}p^\star=d^\star=-\frac{9}{2}\\ x^\star=(-1,4-\lambda^\star)=(-1,3)\end{cases}$

Problem 2

Find a dual problem to the following convex minimization problem:

$$(\mathrm{P}) \quad \begin{aligned} \min \quad & \sum_{i=1}^{n}(a_i x_i^2 + 2b_i x_i + e^{\alpha_i x_i}) \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1 \end{aligned} \qquad \text{where } a, \alpha \in \mathbb{R}^n_{++}, b \in \mathbb{R}^n$$

Solution:

The primal problem can be transformed into:

$$(\mathrm{NewP}) \quad \begin{aligned} \min \quad & \sum_{i=1}^{n}(a_i x_i^2 + 2b_i x_i + e^{\alpha_i y_i}) \\ \text{s.t.} \quad & y = x \\ & \sum_{i=1}^{n} x_i = 1 \end{aligned}$$

The Lagrangian function of (NewP) is given by:

$$\begin{aligned} L(x, y, \nu, \mu) &= \sum_{i=1}^{n}(a_i x_i^2 + 2b_i x_i + e^{\alpha_i y_i}) + \nu^T(y - x) + \mu(\sum_{i=1}^{n} x_i - 1) \\ &= \sum_{i=1}^{n}[a_i x_i^2 + (2b_i - \nu_i + \mu)x_i + e^{\alpha_i y_i} + \nu_i y_i] - \mu \end{aligned}$$

Denote $L_i(x_i, y_i, \nu, \mu) = a_i x_i^2 + (2b_i - \nu_i + \mu)x_i + e^{\alpha_i y_i} + \nu_i y_i \ \ (1 \le i \le n)$

Note that $\begin{cases} \frac{\partial}{\partial x_i} L = \frac{\partial}{\partial x_i} L_i \\ \frac{\partial}{\partial y_i} L = \frac{\partial}{\partial y_i} L_i \end{cases} \ (1 \le i \le n)$

- **① First, we compute the infimum of $a_i x_i^2 + (2b_i - \nu_i + \mu)x_i$:**
  Since $a \in \mathbb{R}^n_{++}$, we deduce that for every $1 \le i \le n$, it holds that:
  $\inf_{x_i}\{a_i x_i^2 + (2b_i - \nu_i + \mu)x_i\} = -\frac{1}{4a_i}(2b_i - \nu_i + \mu)^2$

- **② Then we explore the infimum of $e^{\alpha_i y_i} + \nu_i y_i$:**
  $\frac{\partial}{\partial y_i} L_i = \frac{\partial}{\partial y_i}(e^{\alpha_i y_i} + \nu_i y_i) = \alpha_i e^{\alpha_i y_i} + \nu_i$
  We observe that the equation $\alpha_i e^{\alpha_i y_i} + \nu_i = 0$ has a solution, specifically $y_i = \frac{1}{\alpha_i}\log(-\frac{\nu_i}{\alpha_i})$), if and only if $\nu_i < 0$.
  Hence we have $\inf_{y_i}\{e^{\alpha_i y_i} + \nu_i y_i\} = \begin{cases} -\infty, & \text{if } \nu_i > 0 \\ 0, & \text{if } \nu_i = 0 \\ -\frac{\nu_i}{\alpha_i} + \frac{\nu_i}{\alpha_i}\log(-\frac{\nu_i}{\alpha_i}), & \text{if } \nu_i < 0 \end{cases}$

By combining ①②, we deduce that:

$$\begin{aligned} d_i(\nu, \mu) &\overset{\Delta}{=} \inf_{x_i, y_i} L_i(x_i, y_i, \nu, \mu) \\ &= \inf_{x_i, y_i}\{a_i x_i^2 + (2b_i - \nu_i + \mu)x_i + e^{\alpha_i y_i} + \nu_i y_i\} \\ &= \inf_{x_i}\{a_i x_i^2 + (2b_i - \nu_i + \mu)x_i\} + \inf_{y_i}\{e^{\alpha_i y_i} + \nu_i y_i\} \\ &= \begin{cases} -\frac{1}{4a_i}(2b_i - \nu_i + \mu)^2 - \frac{\nu_i}{\alpha_i} + \frac{\nu_i}{\alpha_i}\log(-\frac{\nu_i}{\alpha_i}), & \text{if } \nu_i < 0 \\ -\frac{1}{4a_i}(2b_i + \mu)^2, & \text{if } \nu_i = 0 \end{cases} \end{aligned}$$

where $\mathrm{dom}(d_i) = \{(\nu, \mu) \in \mathbb{R}^{n+1} : \nu_i \le 0\}$

Therefore, the dual function can be expressed as:

$$
\begin{aligned}
d(\nu,\mu) &= \inf_{x,y} L(x,y,\nu,\mu) \\
&= \sum_{i=1}^{n} \{\inf_{x_i,y_i} L_i(x_i,y_i,\nu,\mu)\} - \mu \\
&= \sum_{i=1}^{n} d_i(\nu,\mu) - \mu
\end{aligned}
$$

where $\mathrm{dom}(d) = \bigcap_{i=1}^{n} \mathrm{dom}(d_i) = \{(\nu,\mu) \in \mathbb{R}^{n+1} : \nu \preceq 0_n\}$

Consequently, the dual problem is formulated as follows:

$$
(\mathrm{D}) \begin{array}{ll} \max\limits_{\nu,\mu} & \sum\limits_{i=1}^{n} d_i(\nu,\mu) - \mu \\ \text{s.t.} & \nu \preceq 0_n \end{array}
$$

Problem 3

Consider the following optimization problem in the variables $\alpha \in \mathbb{R}$ and $q \in \mathbb{R}^n$:

$$
(\mathrm{P}) \begin{array}{ll} \min & \alpha \\ \text{s.t.} & Aq = \alpha f \ \text{ where } A \in \mathbb{R}^{m\times n}, f \in \mathbb{R}^m, \varepsilon \in \mathbb{R}_{++} \\ & \|q\|_2^2 \le \varepsilon \end{array}
$$

Assume in addition that the rows of A are linearly independent.

- **(1) Explain why strong duality holds for problem.**
  - **Solution:**
    Note that (P) is a **convex problem**.
    We observe that $(0_n, 0)$ is strictly feasible.
    This is evident because $\begin{cases} A0_n = 0_m = 0 \cdot f \\ \|0_n\|_2^2 = 0 < \varepsilon \end{cases}$
    Consequently, **Slater's Condition** is satisfied, which implies that strong duality holds.

- **(2) Find a dual problem to problem (P).**
  (Do not assign a Lagrange multiplier to the quadratic constraint.)
  - **Solution:**
    The Lagrangian is given by:
    $$
    \begin{aligned}
    L(\alpha,q,\nu) &= \alpha + \nu^T(Aq - \alpha f) \\
    &= (1-\nu^T f)\alpha + \nu^T Aq \quad (\|q\|_2^2 \le \varepsilon)
    \end{aligned}
    $$
    - ① We note that $(1-\nu^T f)\alpha$ is bounded below if and only if $\nu^T f = 1$;
      and in such cases, the infimum is 0.
    - ② As for $\nu^T Aq$, we apply the **Cauchy-Schwarz Inequality** to deduce that:
      $\nu^T Aq \ge -\|A^T\nu\|_2 \cdot \|q\|_2 \ge -\sqrt{\varepsilon}\|A^T\nu\|_2$
      Evidently, equality in both inequalities is achieved when $A^T\nu = 0_n$.
      In the case where $A^T\nu \ne 0_n$, equality occurs when $q = -\sqrt{\varepsilon}\frac{A^T\nu}{\|A^T\nu\|_2}$.
      Therefore, for any given $\nu \in \mathbb{R}^m$, the infimum of $\nu^T Aq$ is $-\sqrt{\varepsilon}\|A^T\nu\|_2$.

    By synthesizing results ① and ②, the **dual function** can be formulated as:
    $$
    d(\nu) = \inf_{\alpha, \|q\|_2^2 \le \varepsilon} = \begin{cases} -\sqrt{\varepsilon}\|A^T\nu\|_2, & \text{if } \nu^T f = 1 \\ -\infty, & \text{otherwise} \end{cases}
    $$
    As a result, the **dual problem** is specified as:
    $$
    (\mathrm{D}) \begin{array}{ll} \max & -\sqrt{\varepsilon}\|A^T\nu\|_2 \\ \text{s.t.} & f^T\nu = 1 \end{array}
    $$

- **(3) Solve the dual problem obtained in part (2) and find the optimal solution of problem (P).**
  - **Solution:**
    The dual problem (D) can be reformulated into the following equivalent problem:
    $$(\mathrm{NewD})\begin{array}{ll}\min & \frac{1}{2}\|A^T\nu\|_2^2 \\ \text{s.t.} & f^T\nu = 1\end{array}$$
    This problem is convex and only includes an equality constraint.
    Consequently, according to Theorems 3.5 and 3.7 in Note 4,
    the **KKT conditions** are both necessary and sufficient for optimality.

    The Lagrangian for (NewD) is defined as:
    $L_d(\nu,\mu) = \frac{1}{2}\|A^T\nu\|_2^2 - \mu(f^T\nu - 1)$
    where μ represents the Lagrange multiplier.
    The corresponding KKT conditions are $\begin{cases}\nabla_\nu L_d(\nu,\mu) = AA^T\nu - \mu f = 0_m \\ f^T\nu = 1\end{cases}$
    Given that the rows of A are linearly independent, it follows that $AA^T \succ 0$.
    Solving the KKT system yields: $\begin{cases}\nu^\star = \frac{(AA^T)^{-1}f}{f^T(AA^T)^{-1}f} \\ \mu^\star = \frac{1}{f^T(AA^T)^{-1}f}\end{cases}$
    Therefore, the **optimal value** of the dual problem (D) is calculated as:
    $$\begin{aligned} d^\star &= -\sqrt{\varepsilon}\|A^T\nu^\star\|_2 \\ &= -\sqrt{\varepsilon}\frac{\|A^T(AA^T)^{-1}f\|_2}{f^T(AA^T)^{-1}f} \\ &= -\sqrt{\varepsilon}\frac{\|A^T(AA^T)^{-1}f\|_2}{\|A^T(AA^T)^{-1}f\|_2^2} \\ &= -\frac{\sqrt{\varepsilon}}{\|A^T(AA^T)^{-1}f\|_2}\end{aligned}$$
    By invoking **strong duality**, which is established in (1),
    we confirm that the **optimal value** of the primal problem (P) is
    $p^\star = d^\star = -\frac{\sqrt{\varepsilon}}{\|A^T(AA^T)^{-1}f\|_2}$
    Consequently, the **optimal solution** for (P) is given by:
    $$\begin{cases}\alpha^\star = p^\star = -\frac{\sqrt{\varepsilon}}{\|A^T(AA^T)^{-1}f\|_2} \\ q^\star = -\sqrt{\varepsilon}\frac{A^T\nu^\star}{\|A^T\nu^\star\|_2} = -\sqrt{\varepsilon}\frac{A^T(AA^T)^{-1}f}{\|A^T(AA^T)^{-1}f\|_2}\end{cases}$$

Problem 4

Consider the convex optimization problem:

$$(\mathrm{P})\begin{array}{ll}\min & \sum_{i=1}^n x_i\log(\frac{x_i}{c_i}) \\ \text{s.t.} & Ax \succeq b \\ & \sum_{i=1}^n x_i = 1\end{array} \qquad \text{where } c \succ 0_n, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$$

Find a dual problem and simplify it as much as possible.

Solution:

- Denote $A = [a_1, a_2, \dots, a_n]$
  The **Lagrangian** is given by:
  $$\begin{aligned} L(x, \lambda, \nu) &= \sum_{i=1}^{n} x_i \log(\frac{x_i}{c_i}) + \lambda^T(b - Ax) + \nu(\sum_{i=1}^{n} x_i - 1) \\ &= \sum_{i=1}^{n} \{x_i \log(\frac{x_i}{c_i}) - \lambda^T a_i x_i + \nu x_i\} + \lambda^T b - \nu \end{aligned}$$
  Thus $\frac{\partial}{\partial x_i} L(x, \lambda, \nu) = \log(\frac{x_i}{c_i}) + 1 - \lambda^T a_i + \nu$
  Setting this derivative to zero for optimization yields the **unique solution** $x_i = c_i e^{\lambda^T a_i - \nu - 1}$.
  It's important to note that the derivative, $\frac{\partial L}{\partial x_i}$, is **strictly increasing** as a function of x_i.
  Consequently, this observation ensures that the solution **minimizes** the Lagrangian function with respect to the variable x_i.

- As a result, the **dual function** can be formulated as:
  $$\begin{aligned} d(\lambda, \nu) &= \inf_x L(x, \lambda, \nu) \\ &= \inf_x \sum_{i=1}^{n} \{x_i \log(\frac{x_i}{c_i}) - \lambda^T a_i x_i + \nu x_i\} + \lambda^T b - \nu \\ &= -\sum_{i=1}^{n} c_i e^{\lambda^T a_i - \nu - 1} + \lambda^T b - \nu \\ &= -e^{-\nu - 1} \sum_{i=1}^{n} c_i e^{\lambda^T a_i} + \lambda^T b - \nu \end{aligned}$$
  Therefore, the **dual problem** is specified as:
  $$\text{(D)} \quad \begin{aligned} \max \quad & -e^{-\nu-1} \sum_{i=1}^{n} c_i e^{\lambda^T a_i} + \lambda^T b - \nu \\ \text{s.t.} \quad & \lambda \succeq 0_m \end{aligned}$$

- Given that ν is unconstrained and independent from λ,
  we can optimize ν to **simplify the problem formulation**.
  The partial derivative of the dual function $d(\lambda, \nu)$ with respect to ν is given by:
  $$\frac{\partial}{\partial \nu} d(\lambda, \nu) = e^{-\nu-1} \sum_{i=1}^{n} c_i e^{\lambda^T a_i} - 1$$
  Setting this derivative to zero for optimization yields the **unique solution**:
  $$\nu^\star = \log(\sum_{i=1}^{n} c_i e^{\lambda^T a_i}) - 1.$$
  Given that $c \succ 0$,
  it's important to note that the derivative, $\frac{\partial d}{\partial \nu}$, is **strictly decreasing** as a function of ν.
  Consequently, this observation ensures that $\nu^\star$ **maximizes** the Lagrangian function with respect to the variable ν.

- Substituting $\nu^\star = \log(\sum\limits_{i=1}^{n} c_i e^{\lambda^T a_i}) - 1$ into the dual function, we obtain:
  $$\begin{aligned}\sup_{\nu\in\mathbb{R}} d(\lambda,\nu) &= d(\lambda,\nu^\star) \\ &= -e^{-\nu^\star-1}\sum_{i=1}^{n} c_i e^{\lambda^T a_i} + \lambda^T b - \nu^\star \\ &= -1 + \lambda^T b - \log(\sum_{i=1}^{n} c_i e^{\lambda^T a_i}) + 1 \\ &= -\log(\sum_{i=1}^{n} c_i e^{a_i^T\lambda}) + b^T\lambda\end{aligned}$$
  Ultimately, this simplifies the dual problem, which can be reformulated as:
  $$(\text{SimpleD}) \begin{array}{ll} \max & -\log(\sum\limits_{i=1}^{n} c_i e^{a_i^T\lambda}) + b^T\lambda \\ \text{s.t.} & \lambda \succeq 0_n \end{array}$$

Problem 5

Weak duality for unbounded and infeasible problems.
The weak duality inequality, $d^\star \le p^\star$, clearly holds when $d^\star = -\infty$ or $p^\star = \infty$.
Show that it holds in the other two cases as well:

- **(1) If $p^\star = -\infty$, then we must have $d^\star = -\infty$**
  Solution:
  Consider the following general form of an optimization problem:
  $$(\text{P}) \begin{array}{ll} \min\limits_{x\in\mathcal{X}} & f(x) \\ \text{s.t.} & g(x) \preceq 0_m \\ & h(x) = 0_p \end{array}$$
  Denote the **feasible region** for primal problem as $\mathcal{F} = \mathcal{X} \cap \text{dom}(f) \cap \{x : \begin{cases} g(x) \preceq 0_m \\ h(x) = 0_p \end{cases}\}$
  The **Lagrangian** is given by:
  $$\begin{cases} L(x,\lambda,\nu) = f(x) + \lambda^T g(x) + \nu^T h(x) \\ \text{dom}(L) = \mathcal{X} \times \mathbb{R}^m_+ \times \mathbb{R}^p \end{cases}$$
  The **dual function** can be expressed as:
  $$\begin{cases} d(\lambda,\nu) = \inf\limits_{x\in\mathcal{X}} L(x,\lambda,\nu) = \inf\limits_{x\in\mathcal{X}}\{f(x) + \lambda^T g(x) + \nu^T h(x)\} \\ \text{dom}(d) = \{(\lambda,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p : d(\lambda,\nu) > -\infty\} \end{cases}$$
  The **dual problem** is formulated as:
  $$(\text{D}) \begin{array}{ll} \max & d(\lambda,\nu) \\ \text{s.t.} & (\lambda,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p \end{array}$$
  Given that $p^\star = -\infty$,
  indicating that the primal problem (P) is unbounded below,
  we can derive the following for any given $(\lambda,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$:
  $$\begin{aligned} d(\lambda,\nu) &= \inf_{x\in\mathcal{X}} L(x,\lambda,\nu) \\ &= \inf_{x\in\mathcal{X}}\{f(x) + \lambda^T g(x) + \nu^T h(x)\} \\ &\le \inf_{x\in\mathcal{X}}\{f(x) + I_{g(x)\preceq 0_m}(x) + I_{h(x)=0_p}(x)\} \\ &= \inf_{x\in\mathcal{F}} f(x) \\ &= -\infty \end{aligned}$$
  Therefore, $d(\lambda,\nu) = -\infty$ holds for all $(\lambda,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$,
  which implies that the domain of the dual function, $\text{dom}(d)$, is an **empty set**,

rendering the dual problem (D) **infeasible**.
Consequently, the optimal value of the dual problem $d^\star = -\infty$

- **(2) if $d^\star = \infty$, then we must have $p^\star = \infty$**
  Solution:
  Given that $d^\star = \infty$,
  indicating that the dual problem (P) is unbounded above,
  we can derive the following for any given $x \in \mathcal{F}$:
  $$\begin{aligned} f(x) &= f(x) + I_{g(x)\preceq 0_m}(x) + I_{h(x)=0_p}(x) \\ &\geq \sup_{(\lambda,\nu)\in\mathbb{R}_+^m\times\mathbb{R}^p} \{f(x) + \lambda^T g(x) + \nu^T h(x)\} \\ &\geq \sup_{(\lambda,\nu)\in\mathbb{R}_+^m\times\mathbb{R}^p} \inf_{x\in\mathcal{X}} \{f(x) + \lambda^T g(x) + \nu^T h(x)\} \\ &= d^\star \\ &= +\infty \end{aligned}$$
  Therefore, $f(x) = +\infty$ holds for all $x \in \mathcal{F}$,
  which implies that the **feasible region** of the primal function, $\mathcal{F}$, is an **empty set**,
  rendering the primal problem (P) **infeasible**.
  Consequently, the optimal value of the primal problem $p^\star = \infty$

Problem 6

Problems with one inequality constraint.
Express the dual problem of
$(\mathrm{P}) \begin{array}{ll} \min & c^T x \\ \text{s.t.} & f(x) \leq 0 \end{array}$ with $c \neq 0_n$, in terms of the conjugate f^*.
Explain why the problem you give is convex.
Note that we do not assume f is convex.

Solution:

- **(1) Express the dual problem in terms of the conjugate f^*:**
  The Lagragian is given by:
  $L(x,\lambda) = c^T x + \lambda f(x)$
  The **dual function** can be expressed using the conjugate function f^*:
  $$\begin{aligned} d(\lambda) &= \inf_{x\in\mathrm{dom}(f)} L(x,\lambda) \\ &= \inf_{x\in\mathrm{dom}(f)} \{c^T x + \lambda f(x)\} \\ &= \lambda \inf_{x\in\mathrm{dom}(f)} \{\frac{1}{\lambda} c^T x + f(x)\} \\ &= -\lambda \sup_{x\in\mathrm{dom}(f)} \{-\frac{1}{\lambda} c^T x - f(x)\} \\ &= -\lambda f^*(-\frac{c}{\lambda}) \end{aligned}$$
  Thus, the **domain** of the dual function is also expressed
  in terms of the domain of the conjugate function f^*:
  $$\begin{aligned} \mathrm{dom}(d) &= \{\lambda \in \mathbb{R}_+ : d(\lambda) > -\infty\} \\ &= \{\lambda \in \mathbb{R}_+ : f^*(-\frac{c}{\lambda}) < \infty\} \\ &= \{\lambda \in \mathbb{R}_+ : -\frac{c}{\lambda} \in \mathrm{dom}(f^*)\} \end{aligned}$$

Consequently, the dual problem can be formulated as:

$$(\mathrm{D}) \begin{array}{ll} \max & -\lambda f^*(-\frac{c}{\lambda}) \\ \text{s.t.} & \lambda \geq 0 \end{array}$$

- **(2) Explain why the dual function (D) is convex.**
  Solution:
  The **conjugate function** is defined as:
  $$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \{y^T x - f(x)\}$$
  This function is inherently **convex**
  because it is the pointwise supremum over $x \in \mathrm{dom}(f)$ of a family of functions $y^T x - f(x)$, each of which is affine with respect to y.

  Similarly, the **dual function** $d(\lambda) = \inf_{x \in \mathrm{dom}(f)} \{c^T x + \lambda f(x)\}$ is **concave**,
  because it is the pointwise infimum over $x \in \mathrm{dom}(f)$ of a family of functions $c^T x + \lambda f(x)$, each of which is affine with respect to λ.

  Therefore, the dual problem involves maximizing a concave function over a convex feasible region, which inherently characterizes the problem as **convex**.

THE END