DATA130026.01 Optimization Assignment 7

Due Time: at the beginning of the class, Apr. 24, 2024
姓名： 雍崔扬
学号： 21307140051

Problem 1

Consider the problem:

$$
(\mathrm{P}) \quad \begin{array}{ll} \displaystyle\min_{x\in\mathbb{R}, y>0} & e^{-x} \\ \text{s.t.} & \dfrac{x^2}{y} \le 0 \end{array}
$$

(1) Show that the problem is a convex optimization problem.
Find out the optimal solution and optimal value.
Does the Slater's Condition hold?

- **Solution:**
  The problem formulation can be standardized as follows:
  $$
  (\mathrm{P}) \quad \begin{array}{ll} \displaystyle\min_{x\in\mathbb{R}} & e^{-x} \\ \text{s.t.} & x = 0 \end{array}
  $$
  Upon examination, we notice that $x = 0$ is the only feasible point for the problem (P), making it the optimal solution.
  Consequently, we deduce that $\begin{cases} x^\star = 0 \\ p^\star = e^{-x^\star} = 1 \end{cases}$
  The Slater's Condition holds here, as there is no inequality constraint present.

(2) Derive the Lagrangian dual problem.
Find the optimal solution and optimal value of the dual problem and the duality gap.

- **(待纠正: 修改问题的形式，会造成对偶问题的变动，实际上强对偶性应该不成立)**
- **Solution:**
  The **Lagrangian** is given by:
  $L(x, \nu) = e^{-x} + \nu x$
  The second derivative with respect to x is given by:
  $\frac{\partial^2}{\partial x^2} L(x, \nu) = e^{-x}$
  Since $e^{-x} > 0$ for all $x \in \mathbb{R}$, the Lagragian is convex.

  The **dual function** is derived as follows:
  $$
  \begin{aligned}
  d(\nu) &= \inf_{x\in\mathbb{R}} L(x, \nu) \\
  &= \inf_{x\in\mathbb{R}} \{e^{-x} + \nu x\} \\
  &= \begin{cases} -\infty, & \text{if } \nu < 0 \\ 0, & \text{if } \nu = 0 \\ \nu(1 - \log(\nu)), & \text{if } \nu > 0 \end{cases} \\
  &= \begin{cases} -\infty, & \text{if } \nu < 0 \\ \nu(1 - \log(\nu)), & \text{if } \nu \ge 0 \end{cases}
  \end{aligned}
  $$
  Where we denote $0 \log(0) = 1$ for continuity at $\nu = 0$.

Thus, the **dual problem** can be formulated as:

$$(\mathrm{D})\begin{array}{ll}\max & \nu(1-\log(\nu)) \\ \text{s.t.} & \nu \geq 0\end{array}$$

The derivative of the dual function with respect to ν is:

$d'(\nu) = 1 - \log(\nu) - 1 = -\log(\nu)$

Setting $d'(\nu) = 0$ yields $\nu_0 = 1$

(Note that $d(\nu_0) = 1 > 0 = d(0)$)

Therefore the **dual optimum** and **dual optimal value** are $\begin{cases}\nu^\star = 1 \\ d^\star = d(\nu^\star) = 1\end{cases}$

The duality gap is computed by $p^\star - d^\star = 1 - 1 = 0$, demonstrating **strong duality**.

Problem 2

Consider the problem: $\min_x \sum_{i=1}^{n} \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2$

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

Then derive a dual problem.

- **Solution:**

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

  the problem can be reformulated as:

  $$(\mathrm{P})\begin{array}{ll}\min & \displaystyle\sum_{i=1}^{n} \|y_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2 \\ \text{s.t.} & y_i = A_i x + b_i,\ i = 1, 2, \ldots, n\end{array}$$

  The Lagrangian is given by:

  $$\begin{aligned} L(x, \nu_1, \ldots, \nu_n) &= \sum_{i=1}^{n} \|y_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2 + \sum_{i=1}^{n} \nu_i^T (y_i - A_i x - b_i) \\ &= \sum_{i=1}^{n} \{\|y_i\|_2 + \nu_i^T y_i\} + \frac{1}{2}\|x - x_0\|_2^2 - \sum_{i=1}^{n} \nu_i^T (A_i x + b_i) \end{aligned}$$

  - ① **Minimization over y_i:**

    - **Case 1:** $\|\nu_i\|_2 \leq 1$

      Utilizing the **Cauchy-Schwarz inequality**,

      we obtain $\nu_i^T y_i \geq -\|\nu_i\|_2 \|y_i\|_2 \geq -\|y_i\|_2$,

      which implies that $\|y_i\|_2 + \nu_i^T y_i \geq 0$

      Equality is achieved when $y_i = 0_{m_i}$

      Therefore, the infimum is zero for this case.

    - **Case 2:** $\|\nu_i\|_2 > 1$

      By setting $y_i = -t\nu_i$ where t approaches infinity, we analyze:

      $\|y_i\|_2 + \nu_i^T y_i = \|\nu_i\|_2 (1 - \|\nu_i\|_2) t \to -\infty \quad (t \to \infty)$

      Thus, $\|y_i\|_2 + \nu_i^T y_i$ is unbounded below for this case.

    In summary, the infimum of $\|y_i\|_2 + \nu_i^T y_i$ over $y_i \in \mathbb{R}^{m_i}$ can be described as:

    $$\inf_{y_i \in \mathbb{R}^{m_i}} \{\|y_i\|_2 + \nu_i^T y_i\} = \begin{cases} 0, & \text{if } \|\nu_i\|_2 \leq 1 \\ -\infty, & \text{otherwise} \end{cases}$$

- ② **Minimization over x:**
  To minimize with respect to x, we equate the gradient to zero:
  $\nabla_x\{\frac{1}{2}\|x-x_0\|_2^2-\sum_{i=1}^n \nu_i^T(A_ix+b_i)\}=(x-x_0)-\sum_{i=1}^n A_i^T\nu_i=0$
  This yields that $x=x_0+\sum_{i=1}^n A_i^T\nu_i$
  Substituting reveals that:
  $$\begin{aligned}\inf_x\{\frac{1}{2}\|x-x_0\|_2^2-\sum_{i=1}^n \nu_i^T(A_ix+b_i)\} &= \frac{1}{2}\|\sum_{i=1}^n A_i^T\nu_i\|_2^2-\sum_{i=1}^n \nu_i^T[A_i(x_0+\sum_{j=1}^n A_j^T\nu_j)+b_i] \\ &= -\frac{1}{2}\|\sum_{i=1}^n A_i^T\nu_i\|_2^2-\sum_{i=1}^n (A_ix_0+b_i)^T\nu_i\end{aligned}$$

By synthesizing ① and ②, we formulate the **dual function** as follows:
$$d(\nu_1,\dots,\nu_n)=\begin{cases}-\frac{1}{2}\|\sum_{i=1}^n A_i^T\nu_i\|_2^2-\sum_{i=1}^n (A_ix_0+b_i)^T\nu_i, & \text{if } \|\nu_i\|_2\le 1,\ i=1,\dots,n \\ -\infty, & \text{otherwise}\end{cases}$$
This leads to the following dual problem formulation:
$$\begin{aligned}\max \quad & -\frac{1}{2}\|\sum_{i=1}^n A_i^T\nu_i\|_2^2-\sum_{i=1}^n (A_ix_0+b_i)^T\nu_i \\ \text{s.t.} \quad & \|\nu_i\|_2\le 1,\ i=1,\dots,n\end{aligned}$$

Problem 3

(Lower bounds in Chebyshev approximation from least-squares)
Consider the Chebyshev or ℓ_∞-norm approximation problem:
$\min \|Ax-b\|_\infty$
where $A\in\mathbb{R}^{m\times n}$ and $\operatorname{rank}(A)=n$.

Let x_{ch} denote an optimal solution (there may be multiple optimal solutions; x_{ch} denotes one of them).
The Chebyshev problem has no closed-form solution, but the corresponding least-squares problem does. Define: $x_{\text{ls}}=\operatorname{argmin}\|Ax-b\|_2=\left(A^TA\right)^{-1}A^Tb$

We address the following question.
Suppose that for a particular A and b we have computed the least-squares solution x_{ls} (but not x_{ch}).
How suboptimal is x_{ls} for the Chebyshev problem?
In other words, how much larger is $\|Ax_{\text{ls}}-b\|_\infty$ than $\|Ax_{\text{ch}}-b\|_\infty$?

(1) Prove the lower bound: $\|Ax_{\text{ls}}-b\|_\infty\le\sqrt{m}\|Ax_{\text{ch}}-b\|_\infty$
using the fact that for all $z\in\mathbb{R}^m$, $\frac{1}{\sqrt{m}}\|z\|_2\le\|z\|_\infty\le\|z\|_2$

- **Solution:**
  $$\begin{aligned}\|Ax_{\text{ch}}-b\|_\infty &\ge \frac{1}{\sqrt{m}}\|Ax_{\text{ch}}-b\|_2 \\ &\ge \frac{1}{\sqrt{m}}\|Ax_{\text{ls}}-b\|_2 \quad (\text{optimality: } x_{\text{ls}}=\operatorname{argmin}\|Ax-b\|_2) \\ &\ge \frac{1}{\sqrt{m}}\|Ax_{\text{ls}}-b\|_\infty\end{aligned}$$
  This set of inequalities culminates in the formulation:
  $\|Ax_{\text{ls}}-b\|_\infty\le\sqrt{m}\|Ax_{\text{ch}}-b\|_\infty$

(2) In example 5.6 (page 254) we derived a dual for the general norm approximation problem.

Example 5.6 *Norm approximation problem.* We consider the unconstrained norm approximation problem

$$\text{minimize} \quad \|Ax - b\|, \tag{5.63}$$

where $\|\cdot\|$ is any norm. Here too the Lagrange dual function is constant, equal to the optimal value of (5.63), and therefore not useful.

Once again we reformulate the problem as

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & Ax - b = y. \end{array}$$

The Lagrange dual problem is, following (5.61),

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & \|\nu\|_* \leq 1 \\ & A^T\nu = 0, \end{array} \tag{5.64}$$

where we use the fact that the conjugate of a norm is the indicator function of the dual norm unit ball (example 3.26, page 93).

Example 3.26 *Norm.* Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$, with dual norm $\|\cdot\|_*$. We will show that the conjugate of $f(x) = \|x\|$ is

$$f^*(y) = \begin{cases} 0 & \|y\|_* \leq 1 \\ \infty & \text{otherwise,} \end{cases}$$

i.e., the conjugate of a norm is the indicator function of the dual norm unit ball.

If $\|y\|_* > 1$, then by definition of the dual norm, there is a $z \in \mathbf{R}^n$ with $\|z\| \leq 1$ and $y^Tz > 1$. Taking $x = tz$ and letting $t \to \infty$, we have

$$y^Tx - \|x\| = t(y^Tz - \|z\|) \to \infty,$$

which shows that $f^*(y) = \infty$. Conversely, if $\|y\|_* \leq 1$, then we have $y^Tx \leq \|x\|\|y\|_*$ for all x, which implies for all x, $y^Tx - \|x\| \leq 0$. Therefore $x = 0$ is the value that maximizes $y^Tx - \|x\|$, with maximum value 0.

Applying the results to the ℓ_∞-norm (and its dual norm, the ℓ_1-norm),

we can state the following dual for the Chebyshev approximation problem:

$$\begin{array}{ll} & \max \quad b^T\nu \\ (\mathrm{D}) & \text{s.t.} \quad \|\nu\|_1 \leq 1 \\ & \qquad\ A^T\nu = 0 \end{array}$$

Any feasible ν corresponds to a lower bound $b^T\nu$ on $\|Ax_{\mathrm{ch}} - b\|_\infty$.

Denote the least-squares residual as $r_{\mathrm{ls}} = b - Ax_{\mathrm{ls}}$.

Assuming $r_{\mathrm{ls}} \neq 0$, show that $\begin{cases} \hat{\nu} = -\frac{r_{\mathrm{ls}}}{\|r_{\mathrm{ls}}\|_1} \\ \tilde{\nu} = \frac{r_{\mathrm{ls}}}{\|r_{\mathrm{ls}}\|_1} \end{cases}$ are both feasible in (D).

By duality $b^T\hat{\nu}$ and $b^T\tilde{\nu}$ are lower bounds on $\|Ax_{\mathrm{ch}} - b\|_\infty$.

Which is the better bound?

How do these bounds compare with the bound derived in part (1)?

- **Solution:**
  - **① Demonstrating** $\begin{cases} \hat{\nu} = -\frac{r_{\mathrm{ls}}}{\|r_{\mathrm{ls}}\|_1} \\ \tilde{\nu} = \frac{r_{\mathrm{ls}}}{\|r_{\mathrm{ls}}\|_1} \end{cases}$ **are both feasible in** (D)**:**

- We observe that $\begin{cases} \|\hat{\nu}\|_1 = \| - \frac{r_{\text{ls}}}{\|r_{\text{ls}}\|_1} \|_1 = \frac{\|r_{\text{ls}}\|_1}{\|r_{\text{ls}}\|_1} = 1 \\ \|\tilde{\nu}\|_1 = \| \frac{r_{\text{ls}}}{\|r_{\text{ls}}\|_1} \|_1 = \frac{\|r_{\text{ls}}\|_1}{\|r_{\text{ls}}\|_1} = 1 \end{cases}$
  hence both vectors meet the the l_1-norm constraint $\|\nu\|_1 \leq 1$.
- Since $A^T r_{\text{ls}} = A^T(b - Ax_{\text{ls}}) = A^T(b - A(A^T A)^{-1} A^T b) = A^T b - A^T b = 0$
  we deduce that $\begin{cases} A^T \hat{\nu} = -\frac{1}{\|r_{\text{ls}}\|_1} A^T r_{\text{ls}} = 0 \\ A^T \tilde{\nu} = \frac{1}{\|r_{\text{ls}}\|_1} A^T r_{\text{ls}} = 0 \end{cases}$
  hence both vectors satisfy the orthogonality condition $A^T \nu = 0$

Therefore, $\hat{\nu}$ and $\tilde{\nu}$ are both dual feasible.

- **② Comparing the dual objective values of $\hat{\nu}$ and $\tilde{\nu}$:**
  We can calculate the dual objective values at $\hat{\nu}$ and $\tilde{\nu}$ as follows:
  $\begin{cases} b^T \hat{\nu} = \frac{(-b)^T r_{\text{ls}}}{\|r_{\text{ls}}\|_1} = \frac{(-Ax_{\text{ls}} - r_{\text{ls}})^T r_{\text{ls}}}{\|r_{\text{ls}}\|_1} = -\frac{\|r_{\text{ls}}\|_2^2}{\|r_{\text{ls}}\|_1} \\ b^T \tilde{\nu} = \frac{b^T r_{\text{ls}}}{\|r_{\text{ls}}\|_1} = \frac{(Ax_{\text{ls}} + r_{\text{ls}})^T r_{\text{ls}}}{\|r_{\text{ls}}\|_1} = \frac{\|r_{\text{ls}}\|_2^2}{\|r_{\text{ls}}\|_1} \end{cases}$
  The comparison between the dual objective values shows $b^T \tilde{\nu} \geq b^T \hat{\nu}$,
  indicating that $\tilde{\nu}$ gives a better lower bound than $\hat{\nu}$.

- **③ Comparing the lower bound given by $\tilde{\nu}$ and the bound in (1):**
  The lower bound provided by (1) is $\frac{1}{\sqrt{m}} \|b - Ax_{\text{ls}}\|_\infty = \frac{1}{\sqrt{m}} \|r_{\text{ls}}\|_\infty$
  Consider:

$$\begin{aligned} b^T \tilde{\nu} &= \frac{\|r_{\text{ls}}\|_2^2}{\|r_{\text{ls}}\|_1} \\ &= \frac{\|r_{\text{ls}}\|_2}{\|r_{\text{ls}}\|_1} \|r_{\text{ls}}\|_2 \\ &\geq \frac{1}{\sqrt{m}} \cdot \|r_{\text{ls}}\|_\infty \end{aligned}$$

  where the final step utilizes the norm inequalities: $\begin{cases} \|x\|_1 \leq \sqrt{m} \|x\|_2 \\ \|x\|_2 \leq \|x\|_\infty \end{cases} (\forall\ x \in \mathbb{R}^m)$

  - **Proving the first norm inequality $\|x_1\| \leq \sqrt{m} \|x\|_2$:**
    For every given $x \in \mathbb{R}^m$, let $y = [|x_1|, \ldots, |x_m|]$.
    Using the Cauchy-Schwarz inequality on vectors y and 1_m, we have:
    $\|x\|_1 = \sum_{i=1}^{m} |x_i| = \sum_{i=1}^{m} |x_i| \cdot 1 = |1_m^T y| \leq \|1_m\|_2 \|y\|_2 = \sqrt{m} \|x\|_2$
  - The second norm inequality $\|x\|_2 \leq \|x\|_\infty$ was previously provided in part (1).

THE END