SI151A
Convex Optimization and its Applications in Information Science, Fall 2024
Homework 2

Due on Nov. 25, 2024, 11:59 PM

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points ($\leq 20\%$) of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Do your homework by yourself. Any form of plagiarism will lead to 0 point of this homework. If more than one plagiarisms during the semester are identified, we will prosecute all violations to the fullest extent of the university regulations, including but not limited to failing this course, academic probation, or expulsion from the university.
- If you have any doubts regarding the grading, you need to contact the instructor or the TAs within two days since the grade is announced.

I. Convex Optimization Problem

Consider the following compressive sensing problem via ℓ_1-minimization:

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{z}\|_1 \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}, \end{aligned}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times d}$, $\boldsymbol{z} \in \mathbb{R}^d, \boldsymbol{y} \in \mathbb{R}^m$.

Equivalently reformulate the problem into a linear programming problem. (20 points)

Solution:

Suppose the unknown signal is component-wise non-negative, the ℓ_1 minimization problem is just:

$$\min_{\boldsymbol{z}\in\mathbb{R}^d} \sum_{i=1}^{n} \boldsymbol{z}_i$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}, \quad \boldsymbol{z} \geq 0.$$

The general case of real-valued signals, the key trick is to add additional variables to "linearize" the non-linear objective function. Use $\boldsymbol{x}_i$ to represent $|\boldsymbol{z}_i|$, then we have:

$$\begin{aligned} &\min_{\boldsymbol{z},\boldsymbol{x}\in\mathbb{R}^d} \sum_{i=1}^{n} \boldsymbol{x}_i \\ &\text{subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}, \\ &\qquad \boldsymbol{x}_i = |\boldsymbol{z}_i|, \ i = 1, 2, \ldots, n. \end{aligned}$$

However, this problem is non-convex due to the second constraints. So we add "linear" inequalities, that is:

$$\boldsymbol{x}_i - \boldsymbol{z}_i \geq 0, \quad i = 1, 2, \ldots, n,$$
$$\boldsymbol{x}_i + \boldsymbol{z}_i \geq 0, \quad i = 1, 2, \ldots, n,$$

which is equivalent to:

$$\boldsymbol{x}_i \geq \max\{\boldsymbol{z}_i, -\boldsymbol{z}_i\} = |\boldsymbol{z}_i|, \quad i = 1, 2, \ldots, n.$$

Then we have the LP problem:

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{z}\in\mathbb{R}^d} \sum_{i=1}^{n} \boldsymbol{x}_i, \\ &\text{subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}, \\ &\qquad \boldsymbol{x}_i \geq \boldsymbol{z}_i, \quad \boldsymbol{x}_i \geq -\boldsymbol{z}_i, \quad i = 1, 2, \ldots, n. \end{aligned}$$

II. Second-order Cone Programming (SOCP)

Consider the following problem:

$$\min_{x\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 + \lambda\|\mathbf{D}\mathbf{x}\|_1,$$

where $\|\|_p$ is the L_p norm. Equivalently reformulate the problem into a SOCP. (20 points)

Solution:

It's not hard to see that this problem can be formulated as a SOCP

$$\begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}, s\in\mathbb{R}^p} \quad & t + \lambda\sum_{i=1}^{p} s_i \\ & \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq t, \\ & -s_i \leq (\mathbf{D}\mathbf{x})_i \leq s_i, \forall i = 1, \ldots, p. \end{aligned}$$

III. **Semidefinite Programming (SDP)**

Consider the following eigenvalue optimization problem:

$$\min_{\mathbf{w},\mathbf{S}} \ \lambda_{max}(\mathbf{S}) - \lambda_{min}(\mathbf{S}),$$
$$s.t.\ \ \mathbf{S} = \mathbf{B} - \sum_{i=1}^{k} w_i \mathbf{A}_i,$$

where $\mathbf{B} \in \mathbb{R}^{n\times n}$ and $\mathbf{A}_i \in \mathbb{R}^{n\times n}, i = 1, \ldots, k$ are given symmetric data matrices, $w_i \in \mathbb{R}, i = 1, \ldots, k$ are weights, and $\lambda(\cdot)$ means the eigenvalue. Equivalently reformulate the problem into a SDP. (20 points)

Solution:

$\mathbf{S}$ can be factored into $\mathbf{S} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$, where $\mathbf{Q}$ is an orthonormal matrix (i.e., $\mathbf{Q}^{-1} = \mathbf{Q}^T$) and $\mathbf{D}$ is a diagonal matrix consisting of the eigenvalues of $\mathbf{S}$. The conditions:

$$\lambda\mathbf{I} \preceq \mathbf{S} \preceq \mu\mathbf{I}$$

can be rewritten as:

$$\mathbf{Q}(\lambda\mathbf{I})\mathbf{Q}^T \preceq \mathbf{Q}\mathbf{D}\mathbf{Q}^T \preceq \mathbf{Q}(\mu\mathbf{I})\mathbf{Q}^T.$$

After premultiplying the above $\mathbf{Q}^T$ and postmultiplying $\mathbf{Q}$, these conditions become:

$$\lambda\mathbf{I} \preceq \mathbf{D} \preceq \mu\mathbf{I},$$

which are equivalent to:

$$\lambda \le \lambda_{\min}(\mathbf{S}) \quad \text{and} \quad \lambda_{\max}(\mathbf{S}) \le \mu.$$

Therefore, the original problem can be written as:

$$\begin{aligned} EOP: \quad & \min_{\mathbf{w},\mathbf{S},\mu,\lambda} \mu - \lambda \\ & \text{s.t. } \mathbf{S} = \mathbf{B} - \sum_{i=1}^{k} w_i \mathbf{A}_i, \\ & \lambda\mathbf{I} \preceq \mathbf{S} \preceq \mu\mathbf{I}. \end{aligned}$$

This last problem is a semidefinite program.

IV. **Duality**

Derive a dual for the problem

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & -\mathbf{c}^T\mathbf{x} + \sum_{i=1}^{m} y_i \log y_i \\ \text{s.t.} \quad & \mathbf{P}\mathbf{x} = \mathbf{y} \\ & \mathbf{x} \succeq 0, \quad \mathbf{1}^T\mathbf{x} = 1, \end{aligned}$$

where $\mathbf{P} \in \mathbb{R}^{m\times n}$ has nonnegative elements, and its columns add up to one (i.e., $\mathbf{P}^T\mathbf{1} = \mathbf{1}$). The variables are $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$. (For $c_j = \sum_{i=1}^{m} p_{ij} \log p_{ij}$, the optimal value is, up to a factor $\log 2$, the negative of the capacity of a discrete memoryless channel with channel transition probability matrix P; see exercise 4.57 in Boyd S, et al. **Convex optimization** (you can find it in blackboard).)

Simplify the dual problem as much as possible.(20 points)

Solution:

The Lagrangian is

$$\begin{aligned} L(x, y, \lambda, \nu, z) &= -c^T x + \sum_{i=1}^{m} y_i \log y_i - \lambda^T x + \nu(1^T x - 1) + z^T(Px - y), \\ &= (-c + \lambda + \nu 1 + P^T z)^T x + \sum_{i=1}^{m} y_i \log y_i - z^T y - \nu. \end{aligned}$$

The minimum over x is bounded below if and only if

$$-c - \lambda + \nu 1 + P^T z = 0.$$

To minimize over y, we set the derivative with respect to y_i equal to zero, which gives $\log y_i + 1 - z_i = 0$, and conclude that

$$\inf_{y_i \geq 0} (y_i \log y_i - z_i y_i) = -e^{z_i - 1}.$$

The dual function is

$$g(\lambda, \nu, z) = \begin{cases} -\sum_{i=1}^m e^{z_i - 1} - \nu, & -c - \lambda + \nu 1 + P^T z = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

The dual problem is

$$\begin{aligned} &\text{maximize} \quad -\sum_{i=1}^m \exp(z_i - 1) - \nu \\ &\text{subject to } P^T z - c + \nu 1 \succeq 0. \end{aligned}$$

This can be simplified by introducing a variable $w = z + \nu 1$ (and using the fact that $1 = P^T 1$), which gives

$$\begin{aligned} &\text{maximize} \quad -\sum_{i=1}^m \exp(w_i - \nu - 1) - \nu \\ &\text{subject to } P^T w \succeq c. \end{aligned}$$

Finally, we can easily maximize the objective function over ν by setting the derivative equal to zero (the optimal value is $\nu = -\log\left(\sum_{i=1}^m e^{1-w_i}\right)$), which leads to

$$\begin{aligned} &\text{maximize} \quad -\log\left(\sum_{i=1}^m \exp(w_i)\right) - 1 \\ &\text{subject to } P^T w \succeq c. \end{aligned}$$

V. **Convex Problem Applications in Power Allocation**

Consider the following power allocation problem

$$\begin{aligned} \underset{p_1, \ldots, p_K}{\text{maximize}} \quad & \sum_{k=1}^K \ln\left(1 + \frac{p_k |h_k|^2}{N_0}\right) \\ \text{subject to} \quad & \sum_{k=1}^K p_k = P_{\max} \\ & p_k \geq 0, k = 1, \ldots, K, \end{aligned}$$

where $N_0 > 0$.

1. Determine that this problem is convex or not, and provide your argument. (5 points)
2. Write down the dual problem. (5 points)
3. Derive the KKT conditions. (5 points)
4. Derive the expression of the optimal solution to the problem above. (5 points)

Solution:

1. This problem is convex.

   Objective function: Since the $\ln(\cdot)$ is convex function, and the $1 + \frac{p_k |h_k|^2}{N_0}$ is affine function, the $\ln\left(1 + \frac{p_k |h_k|^2}{N_0}\right)$ is convex. Additionally, since the sum of convex function is convex, the $\sum_{k=1}^K \ln\left(1 + \frac{p_k |h_k|^2}{N_0}\right)$ is convex.

   Constrains: $\sum_{k=1}^K p_k = P_{\max}$ is affine set, and $p_k \geq 0$ is hyperspace. Affine set and hyperspace are convex set.

2. Introducing Lagrange multipliers $\lambda \in \mathbb{R}^K$ for the inequality constrains $p_k \geq 0$, and a multiplier $\nu \in \mathbb{R}$ for equality constrain $\sum_{k=1}^K p_k = P_{\max}$, we obtain Lagrangian

$$L(p, \lambda, \nu) = -\sum_{k=1}^{K} \ln\left(1 + \frac{p_k|h_k|^2}{N_0}\right) - \lambda^\top p + \nu\left(\sum_{k=1}^{K} p_k - P_{\max}\right),$$

where $p = [p_1, \ldots, p_K]^\top \in \mathbb{R}^K$ and $\lambda = [\lambda_1, \ldots, \lambda_K]^\top \in \mathbb{R}^K$. To find the dual function, we need to solve an unconstrained minimization of the Lagrangian. We set the gradient equal to zero,

$$\nabla_{p_k} L(p, \lambda, \nu) = \frac{|h_k|^2}{N_0 + p_k|h_k|^2} + \lambda_k - \nu = 0,\ k = 1, \ldots, K.$$

Then, we have

$$p_k = \frac{1}{(\nu - \lambda_k)} - \frac{N_0}{|h_k|^2},\ k = 1, \ldots, K.$$

Substitute into (7), we obtain the dual problem

$$\begin{aligned} \text{maximize}_{\lambda,\nu} \quad & \sum_{k=1}^{K}\left(1 - (\nu - \lambda_k)\frac{N_0}{|h_k|^2} - \ln\left(\frac{|h_k|^2}{N_0(\nu - \lambda_k)}\right)\right) - \nu P_{\max} \\ \text{subject to} \quad & \lambda_k \geq 0,\ k = 1, \ldots, K. \end{aligned}$$

3. The KKT conditions are:
Stationarity:

$$\frac{|h_k|^2}{N_0 + p_k|h_k|^2} + \lambda_k - \nu = 0,\ k = 1, \ldots, K$$

Complementary slackness:

$$\lambda_k p_k = 0,\ k = 1, \ldots, K$$

Primal feasibility:

$$\sum_{k=1}^{K} p_k = P_{\max},\ p_k \geq 0,\ k = 1, \ldots, K$$

Dual feasibility:

$$\lambda_k \geq 0,\ k = 1, \ldots, K$$

4. From the KKT conditions, we know that

$$\lambda_k = \nu - \frac{|h_k|^2}{N_0 + p_k|h_k|^2} \geq 0,\ k = 1, \ldots, K$$

and

$$\lambda_k p_k = \left(\nu - \frac{|h_k|^2}{N_0 + p_k|h_k|^2}\right) p_k = 0,\ k = 1, \ldots, K.$$

If $\nu < \frac{|h_k|^2}{N_0}$, the second equation can only hold if $p_k \geq 0$, which implies that $\nu = \frac{|h_k|^2}{N_0 + p_k|h_k|^2}$ by the first inequation. If $\nu \geq \frac{|h_k|^2}{N_0}$, then $p_k > 0$ is impossible, since it would imply $\nu \geq \frac{|h_k|^2}{N_0} > \frac{|h_k|^2}{N_0 + p_k|h_k|^2}$, which violates the complementary slackness condition. Therefore, we have

$$p_k = \begin{cases} \frac{1}{\nu} - \frac{N_0}{|h_k|^2}, & \text{if } \nu < \frac{|h_k|^2}{N_0} \\ 0, & \text{if } \nu \geq \frac{|h_k|^2}{N_0} \end{cases}$$

or, put more simply,

$$p_k = \max\left(0, \frac{1}{\nu} - \frac{N_0}{|h_k|^2}\right).$$

With the primal feasibility condition, we have

$$p_k = \max\left(0, \frac{1}{\nu} - \frac{N_0}{|h_k|^2}\right)$$

satisfying

$$\sum_{k=1}^{K} \max\left(0, \frac{1}{\nu} - \frac{N_0}{|h_k|^2}\right) = P_{\max}.$$

The left-hand side of $\sum_{k=1}^{K} \max\left(0, \frac{1}{\nu} - \frac{N_0}{|h_k|^2}\right) = P_{\max}$ is a piecewise-linear increasing function of $\frac{1}{\nu}$, with breakpoints at $\frac{N_0}{|h_k|^2}$, so the equation has a unique solution which is readily determined.