Convex Optimization
Midterm Exam

Question 1 (Finite Cover of a Convex Set)

Suppose a compact convex set $C \in \mathbb{R}^d$ is covered by a family $\mathscr{F}$ of open half-spaces (in other words, $C \subseteq \cup_{H \in \mathscr{F}}(C \cap H)$). Then it is covered by $d+1$ or fewer of the half-spaces in $\mathscr{F}$.

(Note that for any given $H \in \mathscr{F}$, it may be the case that $H \cap C \neq \emptyset$ *and* $H \cap C^c \neq \emptyset$, where C^c denotes the complement of the set C.)

You may initially assume that $\mathscr{F}$ is a finite family – extra credit for showing that it works for an infinite family as well.

Solution (Helly's Theorem)

First, consider that since C is compact and $\mathscr{F}$ is an open cover of C, there is a finite subset of $\mathscr{F}$ that also covers C (by the properties of compactness). Thus, we can assume $\mathscr{F}$ is finite.

Consider the following family of sets

$$\mathscr{S} = \{C \cap H^c : H \in \mathscr{F}\}$$

Since $H \in \mathscr{F}$ is *open*, H^c is closed and so $C \cap H^c$ is the intersection of a closed set and a compact set. Thus, it is compact (and convex)

Now, the fact $\mathscr{F}$ covers C implies that

$$\bigcap_{S \in \mathscr{S}} S = \emptyset$$

Since $\mathscr{F}$ is finite (assume $|\mathscr{F}| = N$), we can write this as

$$\bigcap_{n=1}^{N} S_n = \emptyset$$

By (the converse of) Helly's Theorem[1], this means that there is a subset of $d+1$ sets in $\mathscr{S}$, $S_{i_1}, \cdots, S_{i_{d+1}}$ such that

$$\bigcap_{n=1}^{d+1} S_{i_n} = \emptyset$$

$$\bigcap_{n=1}^{d+1} (C \cap H_{i_n}^c) = \emptyset$$

[1] Recall that Helly's Theorem states that for a finite collection of convex subsets of $\mathbb{R}^d$, if the intersection of $d+1$ of these subsets is nonempty, then the whole collection has a nonempty intersection.

And this implies that

$$C \subseteq \bigcup_{n=1}^{d+1} H_{i_n}$$

As required.

Solution (Radon's Theorem)

First, consider that since C is compact and $\mathscr{F}$ is an open cover of C, there is a finite subset of $\mathscr{F}$ that also covers C (by the properties of compactness). Thus, we can assume $\mathscr{F}$ is finite.

Assume the smallest such cover has $k > d + 1$ elements. For each of those k sets, let $\boldsymbol{x}^{(i)}$ be a point in $C_i = F_i \cap C$. Each of those points must be unique (if not, we can remove the corresponding hyperplane and still maintain a cover, contradicting the assumption that this is smallest such cover).

Now, by Radon's Theorem[2], since $k > d+1$, there exists a partition (I, J) of the $\boldsymbol{x}^{(i)}$ such that the intersection of the convex hulls of points in I and J is non-empty. Let $\boldsymbol{z}$ be a point in the intersection, and assume without loss of generality that $\boldsymbol{z} \in J$. Clearly, $\boldsymbol{z} \in C$ and $\boldsymbol{z} \in C_z$ for some $z \in J$.

[2]Recall that Radon's Theorem states that for $n + 2$ points in $\mathbb{R}^n$, there exists a partition of these points such that the intersection of the convex hull of the two partitions is non-empty.

Now, since $z \in J$, $\boldsymbol{x}^{(i)} \notin C_z$ for all $i \in I$. Clearly, therefore

$$\{\text{conv}(\boldsymbol{x}^{(i)}), I \in i\} \cap C_z =$$

However, $\boldsymbol{z}$ is in both these sets. This is a contradiction.

■ □ ■

Question 2 (Separating Hyperplane Theorem)

Suppose $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Show that exactly one of the following two statements are true

1. There exists an $\boldsymbol{x} > \mathbf{0}$ such that $A\boldsymbol{x} = \boldsymbol{b}$.
2. There exists a $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that $A^\top \boldsymbol{\lambda} \geq \mathbf{0}$ (with $A^\top \boldsymbol{\lambda} \neq \mathbf{0}$) and $\boldsymbol{b} \cdot \boldsymbol{\lambda} \leq 0$.

Show how to computer the certifcate $\boldsymbol{\lambda}$, if it exists, via linear programming.

Solution (Method 1)

Let's prove this in two steps

2 ⇒ not 1 : Suppose a $\boldsymbol{\lambda}$ exists as described in step 2, and let $\boldsymbol{b} = A\boldsymbol{x}$. Then

$$\boldsymbol{b} \cdot \boldsymbol{\lambda} \leq 0 \Rightarrow \boldsymbol{x}^\top A^\top \boldsymbol{\lambda} \leq 0$$

However, we also know that $\boldsymbol{A}^\top \boldsymbol{\lambda} \geq \mathbf{0}$ with at least one strictly positive component. As such, at least one component of $\boldsymbol{x}$ must be negative for this inequality to hold, and $\boldsymbol{x} \nleq \mathbf{0}$. Thus, (1) above does not hold.

Not 1 ⇒ 2 : Consider the set

$$\mathscr{C} = \{\boldsymbol{\beta} : \boldsymbol{\beta} = A\boldsymbol{x}, \boldsymbol{x} > \mathbf{0}\}$$

And suppose $\boldsymbol{b} \notin \mathscr{C}$ (ie: suppose (1) is not true). Then by the separating hyperplane theorem, there exists a $\boldsymbol{\lambda}$ such that

$$\boldsymbol{\lambda} \cdot \boldsymbol{b} \leq \boldsymbol{\lambda} \cdot A\boldsymbol{x} = \boldsymbol{x} \cdot (A^\top \boldsymbol{\lambda}) \quad \forall \boldsymbol{x} > \mathbf{0} \tag{1}$$

Now, consider a particular $\boldsymbol{\lambda}$ that satisfies equation 1:

- Let $\boldsymbol{x}' = \alpha\mathbf{1}$ and put this into equation 1. Letting $\alpha \to 0$, the RHS of equation 1 tends to 0[3] and so we find that
$$\boldsymbol{b} \cdot \boldsymbol{\lambda} \leq 0$$
- Let $\boldsymbol{x}' = \alpha\boldsymbol{e}_i$ for every i successively. Letting $\alpha \to \infty$, we find that
$$A^\top \boldsymbol{\lambda} \geq \mathbf{0}$$
If this were not the case, the RHS would shoot to $-\infty$, and since the LHS is finite, this would result in a contradiction.
- Finally, we need to show that $A^\top \boldsymbol{\lambda} \neq \mathbf{0}$. This is not difficult – we can assume, without loss of generality, that A has full row-rank (if it doesn't, simply remove any rows that are linear combination of the others and resolve the problem in a lower dimension). $A^\top$ therefore has full column rank. And therefore,
$$\boldsymbol{\lambda} \neq \mathbf{0} \Rightarrow A^\top \boldsymbol{\lambda} \neq \mathbf{0}$$

[3]Because

$$\begin{aligned} \boldsymbol{x}' \cdot (A^\top \boldsymbol{\lambda}) &\leq \|\boldsymbol{x}'\| \cdot \|A^\top \boldsymbol{\lambda}\| \\ &= \alpha \|A^\top \boldsymbol{\lambda}\| \\ &\to 0 \end{aligned}$$

Though I wasn't so picky in requiring this exact logic.

To find the certificate using linear programming, we need to solve the LP

$$\begin{aligned} \min \quad & \boldsymbol{b} \cdot \boldsymbol{\lambda} \\ \text{s.t.} \quad & A^\top \boldsymbol{\lambda} \geq \mathbf{0} \\ & \mathbf{1} \cdot (A^\top \boldsymbol{\lambda}) = 1 \end{aligned}$$

The equality constraint simply forces $A^\top \boldsymbol{\lambda}$ to be strictly positive. If the LP has a zero or negative optimal solution, the certificate exists

Solution (Method 2)

A few of you provided the following solution (usually liberally embellished). This method also uses the separating hyperplane theorem, but instead considers the following two sets

$$A = \{A\boldsymbol{x} : \boldsymbol{x} > 0\}$$

$$B = \mathbb{R}^n_{++}$$

Consider a $\boldsymbol{b}$ and choose an $\boldsymbol{x}$ such that $A\boldsymbol{x} = \boldsymbol{b}$ (as discussed above, A can be assumed to have full-rank, so only one $\boldsymbol{x}$ satisfies this).

Now,

$$\boldsymbol{b} \notin A \Leftrightarrow \boldsymbol{x} \notin B$$

Therefore, if $\boldsymbol{b} \notin A$, there is therefore a separating hyperplane between $\boldsymbol{x}$ and $\mathbb{R}^n_{++}$. In other words, there exists a $\boldsymbol{c} \neq \mathbf{0}$

$$\boldsymbol{c} \cdot \boldsymbol{x} \leq \boldsymbol{c} \cdot \boldsymbol{p} \quad \forall \boldsymbol{p} \in \mathbb{R}^n_{++} \tag{2}$$

Letting $p_i \to \infty$ implies that $\boldsymbol{c} \leq \mathbf{0}$. Furthermore, since A has full rank, we can find a $\boldsymbol{\lambda}$ such that $\boldsymbol{c} = A^\top \boldsymbol{\lambda}$. Thus, we have found a $\boldsymbol{\lambda}$ such that

$$\boldsymbol{c} = A^\top \boldsymbol{\lambda} \leq \mathbf{0}$$

$$\boldsymbol{c} = A^\top \boldsymbol{\lambda} \neq \mathbf{0}$$

Finally, putting $\boldsymbol{c} = A^\top \boldsymbol{\lambda}$ into 2, we get

$$\boldsymbol{x}^\top A^\top \boldsymbol{\lambda} \leq \boldsymbol{p}^\top \boldsymbol{A}^\top \boldsymbol{\lambda} \quad \forall \boldsymbol{p} \in \mathbb{R}^n_{++}$$

$$\boldsymbol{b} \cdot \boldsymbol{\lambda} \leq \boldsymbol{p}^\top A^\top \boldsymbol{\lambda} \quad \forall \boldsymbol{p} \in \mathbb{R}^n_{++}$$

Letting the RHS tend to 0 as (in the first solution), we obtain our required result.

■ □ ■

Question 3 (Monotone Extension of a Convex Function)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex. A function $h : \mathbb{R}^n \to \mathbb{R}$ is monotone nondecreasing if $h(\boldsymbol{x}) \geq h(\boldsymbol{y})$ whenever $\boldsymbol{x} \geq \boldsymbol{y}$. The monotone extension of f is defined as

$$g(\boldsymbol{x}) = \inf_{\boldsymbol{z} > \boldsymbol{0}} f(\boldsymbol{x} + \boldsymbol{z})$$

(We will assume $g(\boldsymbol{x}) > -\infty$).

Show that g is convex and monotone non-decreasing, and that $g(\boldsymbol{x}) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x}$.

Let h be any other monotone non-decreasing and convex function, with $h(\boldsymbol{x}) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x}$. Show that $h(\boldsymbol{x}) \leq g(\boldsymbol{x})$ for all $\boldsymbol{x}$. Thus, g is the maximum convex monotone underestimator of f.

Solution (Convexity)

Operations that preserve convexity : Seven people solved the problem as follows. $\boldsymbol{x} + \boldsymbol{z}$ is an affine function of both $\boldsymbol{x}$ and $\boldsymbol{z}$. Thus, $f(\boldsymbol{x} + \boldsymbol{z})$ is the composition of a convex function and an affine function, and is jointly convex in $\boldsymbol{x}$ and $\boldsymbol{z}$.

$g(\boldsymbol{x})$ is the partial minimization of this function over the convex set $\boldsymbol{z} \geq \boldsymbol{0}$ and is therefore convex.

Directly from definition : Five people solved the problem as follows. Let

$$\boldsymbol{x}^* = \text{argmin}_{\boldsymbol{z} \geq \boldsymbol{0}} f(\boldsymbol{x} + \boldsymbol{z})$$

Then we have that

$$g(\boldsymbol{x}) = f(\boldsymbol{x} + \boldsymbol{x}^*)$$

As such, for any $\boldsymbol{x}_1, \boldsymbol{x}_2$ and $\lambda \in [0, 1]$, $\bar{\lambda} = 1 - \lambda$,

$$\begin{aligned} g(\lambda \boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2) &= f\left[\lambda(\boldsymbol{x}_1 + \boldsymbol{x}_1^*) + \bar{\lambda}(\boldsymbol{x}_2 + \boldsymbol{x}_2^*)\right] \\ &\overset{4}{\leq} \lambda f(\boldsymbol{x}_1 + \boldsymbol{x}_2^*) + \bar{\lambda} f(\boldsymbol{x}_2 + \boldsymbol{x}_2^*) \\ &= \lambda g(\boldsymbol{x}_1) + \bar{\lambda} g(\boldsymbol{x}_2) \end{aligned}$$

[4] By convexity of f.

Incorrect! Three people attempted to solve the problem as follows. For any $\boldsymbol{x}_1, \boldsymbol{x}_2$ and $\lambda \in [0, 1]$, $\bar{\lambda} = 1 - \lambda$, we have

[5] Incorrect! This is simply not true. In moving from line 4 to line 5, what we're effectively doing is allowing *each* of the two *vecz*'s in the question to take a *different* value. This gives us *more* flexibility, and so our optimal value will go *down*. Thus, line 4 is $\geq$ line 5, not the other way round.

$$\begin{aligned} g(\lambda \boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2) &= \inf_{\boldsymbol{z}\geq \boldsymbol{0}} f(\lambda \boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2 + \boldsymbol{z}) \\ &= \inf_{\boldsymbol{z}\geq \boldsymbol{0}} f\left[\lambda(\boldsymbol{x}_1 + \boldsymbol{z}) + \bar{\lambda}(\boldsymbol{x}_2 + \boldsymbol{z})\right] \quad (3) \\ &\leq^{4} \inf_{\boldsymbol{z}\geq \boldsymbol{0}} \left[\lambda f(\boldsymbol{x}_1 + \boldsymbol{z}) + \bar{\lambda} f(\boldsymbol{x}_2 + \boldsymbol{z})\right] \quad (4) \\ &\leq^{5} \lambda \inf_{\boldsymbol{z}\geq \boldsymbol{0}} f(\boldsymbol{x}_1 + \boldsymbol{z}) + \bar{\lambda} \inf_{\boldsymbol{z}\geq \boldsymbol{0}} f(\boldsymbol{x}_2 + \boldsymbol{z}) \quad (5) \\ &= \lambda g(\boldsymbol{x}_1) + \bar{\lambda} g(\boldsymbol{x}_2) \end{aligned}$$

One person did manage to give a correct version of this answer by noticing, in line 3, that the two $\boldsymbol{z}$'s can be replaced by $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ and the $\inf_{\boldsymbol{z}\geq\boldsymbol{0}}$ changed to $\inf_{\boldsymbol{z}_1,\boldsymbol{z}_2\geq\boldsymbol{0}}$ without changing the problem. This 'decoupling' however, is no longer obviously possible in line 4.

Incorrect! One person tried to do this directly by considering the region in which f' is positive, negative and 0. This *would* work for the 1D case, but *not* for the general multi-dimensional case (for many reasons – among others, since $\nabla f(\boldsymbol{x})$ may have some positive and some negative components).

Solution (Monotone non-decreasing)

Every correct answer was some variation on the following solution. Take a $\boldsymbol{x} \geq \boldsymbol{y}$. Then we can write $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{d}$ where $\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{y} \geq \boldsymbol{0}$. We then have

$$\begin{aligned} g(\boldsymbol{x}) &= \inf_{\boldsymbol{z}\geq \boldsymbol{0}} f(\boldsymbol{x} + \boldsymbol{z}) \\ &= \inf_{\boldsymbol{z}\geq \boldsymbol{0}} f(\boldsymbol{y} + \boldsymbol{d} + \boldsymbol{z}) \\ &= \inf_{\hat{\boldsymbol{z}}\geq \boldsymbol{d}} f(\boldsymbol{y} + \hat{\boldsymbol{z}}) \quad (6) \\ &\geq^{6} \inf_{\hat{\boldsymbol{z}}\geq \boldsymbol{0}} f(\boldsymbol{y} + \hat{\boldsymbol{z}}) \quad (7) \\ &= g(\boldsymbol{y}) \end{aligned}$$

[6] This is true because the programs in lines 6 and 7 are identical except for their feasible region, and the feasible region in line 6 is smaller than that in line 7. Thus, the program in line 7 will have a 'better' (in this case smaller) solution.

Solution ($g(\boldsymbol{x}) \leq f(\boldsymbol{x})$)

Almost everyone got this (though some omitted it).

$$g(\boldsymbol{x}) = \inf_{\boldsymbol{z}\geq \boldsymbol{0}} f(\boldsymbol{x} + \boldsymbol{z}) \leq^{7} f(\boldsymbol{x} + \boldsymbol{0}) = f(\boldsymbol{x})$$

[7] Because $\boldsymbol{0} \in \boldsymbol{z} : \boldsymbol{z} \leq \boldsymbol{0}$, the feasible region of the program.

Solution (g is Maximum Understimators)

All solutions were a (usually much-lengthened!) version of the following argument. Since $h \leq f$, we have that for every $\boldsymbol{x}$ and $\boldsymbol{z}$

$$h(\boldsymbol{x} + \boldsymbol{z}) \leq f(\boldsymbol{x} + \boldsymbol{z})$$

Taking an infemum over $\boldsymbol{z} \geq \mathbf{0}$ on both sides

$$\inf_{\boldsymbol{z}\geq\mathbf{0}} h(\boldsymbol{x}+\boldsymbol{z}) \leq \inf_{\boldsymbol{z}\geq\mathbf{0}} f(\boldsymbol{x}+\boldsymbol{z})$$

The RHS is simply $g(\boldsymbol{x}$ (by definition) and the LHS is $h(\boldsymbol{x})$, since h is monotone non-decreasing.

■ □ ■

Question 4 (Epigraph of K-Convex Functions)

Let $\mathscr{K} \subseteq \mathbb{R}^m$ be a proper convex cone with associated generalized inequality $\preceq_{\mathscr{K}}$, and let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$. The epigraph of $\boldsymbol{f}$, with respect to $\preceq_{\mathscr{K}}$, is defined as the set

$$\operatorname{epi}_{\mathscr{K}} \boldsymbol{f} = \left\{(\boldsymbol{x}, \boldsymbol{t}) \in \mathbb{R}^{n+m} : \boldsymbol{f}(\boldsymbol{x}) \preceq_{\mathscr{K}} \boldsymbol{t}\right\}$$

The function is called $\mathscr{K}$-convex if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\lambda \in [0,1]$ (and $\bar{\lambda} = 1 - \lambda$)

$$\boldsymbol{f}\left(\lambda\boldsymbol{x} + \bar{\lambda}\boldsymbol{x}\right) \preceq_{\mathscr{K}} \lambda\boldsymbol{f}(\boldsymbol{x}) + \bar{\lambda}\boldsymbol{f}(\boldsymbol{x})$$

Show that $\boldsymbol{f}$ is $\mathscr{K}$-convex if and only if $\operatorname{epi}_{\mathscr{K}}\boldsymbol{f}$ is a convex set.

Solution

Consider some $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ and $\lambda \in [0,1]$, with $\bar{\lambda} = 1 - \lambda$. The crux of this question lies in the quantity

$$\boldsymbol{f}(\lambda\boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2)$$

Now, consider

K-convexity ⇒ convex epigraph : Start from the expression above, and note that K-convexity implies that

$$\boldsymbol{f}(\lambda\boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2) \preceq_K \lambda\boldsymbol{f}(\boldsymbol{x}_1) + \bar{\lambda}\boldsymbol{f}(\boldsymbol{x}_2) \tag{8}$$

Now, consider $(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2) \in \operatorname{epi}\boldsymbol{f}$. This means that

$$\boldsymbol{f}(\boldsymbol{x}_1) \preceq_K \boldsymbol{y}_1 \qquad \boldsymbol{f}(\boldsymbol{x}_2) \preceq_K \boldsymbol{y}_2 \tag{9}$$

Feeding 9 into 8, we immediately get

$$\boldsymbol{f}(\lambda\boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2) \preceq_K \lambda\boldsymbol{y}_1 + \bar{\lambda}\boldsymbol{y}_2$$
$$\Rightarrow \lambda(\boldsymbol{x}_1, \boldsymbol{y}_1) + \bar{\lambda}(\boldsymbol{x}_2, \boldsymbol{y}_2) \in \operatorname{epi}\boldsymbol{f}$$

Since the choice of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ was arbitrary, $\operatorname{epi}\boldsymbol{f}$ is convex.

Convex epigraph $\Rightarrow$ K-convexity : Consider $(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2) \in \text{epi}\boldsymbol{f}$. Since we assume the epigraph is convex, $\left[\lambda(\boldsymbol{x}_1, \boldsymbol{y}_1) + \bar{\lambda}(\boldsymbol{x}_2, \boldsymbol{y}_2)\right] \in \text{epi}\boldsymbol{f}$, which means that

$$\boldsymbol{f}(\lambda \boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2) \preceq_K \lambda \boldsymbol{y}_1 + \bar{\lambda}\boldsymbol{y}_2$$

Choosing $\boldsymbol{y}_1 = \boldsymbol{f}(\boldsymbol{x}_1)$ and $\boldsymbol{y}_2 = \boldsymbol{f}(\boldsymbol{x}_2)$ (this is legal – the points still satisfy 9), we find

$$\boldsymbol{f}(\lambda \boldsymbol{x}_1 + \bar{\lambda}\boldsymbol{x}_2) \preceq_K \lambda \boldsymbol{f}(\boldsymbol{x}_1) + \bar{\lambda}\boldsymbol{f}(\boldsymbol{x}_2)$$

Thus, $\boldsymbol{f}$ is K-convex.

■ □ ■

Question 5 (Robust Linear Programming with Ellipsoidal Uncertainty Sets)

Consider the following robust linear program

$$\begin{array}{ll} \min & \max_{\boldsymbol{f}^{(0)} \in \mathscr{F}_0} (\boldsymbol{f}^{(0)} \cdot \boldsymbol{x}) \\ \text{s.t.} & \max_{\boldsymbol{f}^{(i)} \in \mathscr{F}_i} (\boldsymbol{f}^{(i)} \cdot \boldsymbol{x}) \end{array}$$

Where

$$\mathscr{F}_i = \left\{ \boldsymbol{f} : (\boldsymbol{f} - \boldsymbol{g}^{(i)})^\top V_i^{-1} (\boldsymbol{f} - \boldsymbol{g}^{(i)}) \leq 1 \right\}$$

and $V_i \succ 0$, $\boldsymbol{g}^{(i)} \in \mathbb{R}^n$.

Solution

The crux of this question was proving that[8]

$$\max_{\boldsymbol{f}^{(i)} \in \mathscr{F}_i} (\boldsymbol{f}^{(i)} \cdot \boldsymbol{x}) = \boldsymbol{g}^{(i)} \cdot \boldsymbol{x} + \left\| V_i^{1/2} \boldsymbol{x} \right\|_2 \qquad (10)$$

[8] $V^{1/2}$ is a matrix such that $V^{(1/2),\top} V^{(1/2)} = V$. The existence of such a matrix is implied by the fact $V \succ 0$.

Having done this, the program becomes

$$\begin{array}{ll} \min & \boldsymbol{g}^{(0)} \cdot \boldsymbol{x} + \|V_0^{1/2} \boldsymbol{x}\|_2 \\ \text{s.t.} & \boldsymbol{g}^{(i)} \cdot \boldsymbol{x} + \|V_i^{1/2} \boldsymbol{x}\|_2 \leq b_i \end{array}$$

Finally, we can use an epigraph formulation to write this as

$$\begin{array}{ll} \min & \beta_0 \\ \text{s.t.} & \|V_0^{1/2} \boldsymbol{x}\|_2 \leq \beta_0 - \boldsymbol{g}^{(0)} \cdot \boldsymbol{x} \\ & \|V_i^{1/2} \boldsymbol{x}\|_2 \leq b_i - \boldsymbol{g}^{(i)} \cdot \boldsymbol{x} \end{array}$$

This is an SOCP.

I saw two ways of proving equation 10. I think the first was rather nicer, but that's just me!

Cauchy-Schwartz : Note that the program on the LHS of equation 10 is

$$\begin{array}{ll} \max & \boldsymbol{f} \cdot \boldsymbol{x} \\ \text{s.t.} & (\boldsymbol{f}-\boldsymbol{g})^\top V_i^{-1}(\boldsymbol{f}-\boldsymbol{g}) \leq 1 \end{array}$$

Now consider that the constraint can be re-written as

$$\left\| V_i^{-1/2}(\boldsymbol{f}-\boldsymbol{g}) \right\|_2^2 \leq 1$$

substituting $\boldsymbol{u} = V_i^{-1/2}(\boldsymbol{f}-\boldsymbol{g})$, our program becomes

$$\begin{array}{rl} \boldsymbol{g} \cdot \boldsymbol{x} + \max & (V_i^{1/2}\boldsymbol{u}) \cdot \boldsymbol{x} \\ \text{s.t.} & \|\boldsymbol{u}\|_2^2 \leq 1 \end{array}$$

The objective can be re-written as follows

$$(V_i^{1/2}\boldsymbol{u}) \cdot \boldsymbol{x} = \boldsymbol{u}^\top V_i^{1/2}\boldsymbol{x} = \boldsymbol{u} \cdot (V_i^{1/2}\boldsymbol{x})$$

Now consider that by the Cauchy-Schwartz Inequality,

$$(V_i^{1/2}\boldsymbol{x}) \cdot \boldsymbol{u} \leq \|\boldsymbol{u}\|_2 \cdot \|V_i^{1/2}\boldsymbol{x}\|_2$$

This inequality is tight – in other words, the maximum is attained when the two vectors are aligned. Since $\|\boldsymbol{u}\|_2 \leq 1$, we find that the maximum ends up being

$$\boldsymbol{g} \cdot \boldsymbol{x} + \|V_i^{1/2}\boldsymbol{x}\|_2$$

KKT Conditions

■ □ ■

Question 6 (Optimization over Polynomials)

The Markov-Lucacs Theorem states that the polynomial

$$p(t) = p_0 + p_1 t + \cdots + p_{2k} t^{2k}$$

is non-negative on $\mathbb{R}$ if and only if it is the sum of the squares of two or more polynomials of degree k or less – ie:

$$\begin{aligned} p(t) &= r(t)^2 + s(t)^2 \\ &= (\boldsymbol{r} \cdot \boldsymbol{t})^2 + (\boldsymbol{s} \cdot \boldsymbol{t})^2 \\ &= \boldsymbol{t}_k^\top \left(\boldsymbol{r}\boldsymbol{r}^\top + \boldsymbol{s}\boldsymbol{s}^\top \right) \boldsymbol{t}_k \end{aligned}$$

Where $\boldsymbol{t}_k = (1, t, \cdots, t^k)$ and $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^{k+1}$.

Part A

Use the representation above to show that $p(t)$ is non-negative on $\mathbb{R}$ if and only if

$$p_i = \sum_{m+n=i+2} Y_{mn}$$

for some $Y \in \mathbb{S}_+^{k+1}$ (ie: Y is a positive semidefinite $(k+1) \times (k+1)$ matrix.

Solution

Most solutions here followed the same (correct) route. The first step involves some book-keeping. We need to show that

$$p(t) = \boldsymbol{t}_k^\top Y \boldsymbol{t}_k = \sum_{i=0}^{2k} p_i t^i$$

with

$$p_i = \sum_{m+n=i+2} Y_{mn}$$

This is tedious but not particularly difficult to do. Here's the cleanest method I saw (though I gave credit to all valid methods)[9]

[9] In the below, I use the notation $\boldsymbol{Y}_{\Diamond j}$ to denote the j^{th} row of Y.

$$\begin{aligned}
\boldsymbol{t}_k^\top Y \boldsymbol{t}_k &= \boldsymbol{t}_k^\top \left(\textstyle\sum_{j=1}^{k+1} \boldsymbol{Y}_{\Diamond j} t^{j-1} \right) \\
&= \sum_{i=1}^{k+1} \left[t^{i-1} \left(\textstyle\sum_{j=1}^{k+1} Y_{ij} t^{j-1} \right) \right] \\
&= \sum_{i,j=1}^{k+1} Y_{ij} t^{i+j-2} \\
&= \sum_{i=0}^{2k} \left[\left(\sum_{m+n=i+2} Y_{mn} \right) t^i \right]
\end{aligned}$$

Having does this, let's do what the question asks:

$\boxed{p(t) = \boldsymbol{t}_k^\top Y \boldsymbol{t}, Y \succeq 0 \Rightarrow p(t) \geq 0}$: This is not difficult. If $Y \succeq 0$, then for any $\boldsymbol{t}_k$ (ie: for any t) the quadratic form is positive (and therefore so is $p(t)$.

$\boxed{p(t) \geq 0 \Rightarrow \exists Y \succeq 0 \textbf{ s.t. } p(t) = \boldsymbol{t}_k^\top Y \boldsymbol{t}}$: To prove this direction, we need to use the ML Theorem to write $p(t)$ as

$$p(t) = \boldsymbol{t}_k^\top \left(\boldsymbol{r}\boldsymbol{r}^\top + \boldsymbol{s}\boldsymbol{s}^\top \right) \boldsymbol{t}_k$$

This is in the form required, with $Y = \boldsymbol{r}\boldsymbol{r}^\top + \boldsymbol{s}\boldsymbol{s}^\top$. All that remains to do is that this Y is positive semidefinite. To do this, consider that for any $\boldsymbol{x}$

$$\begin{aligned}\boldsymbol{x}^\top Y \boldsymbol{x} &= \boldsymbol{x}^\top \left(\boldsymbol{r}\boldsymbol{r}^\top + \boldsymbol{s}\boldsymbol{s}^\top\right)\boldsymbol{x} \\ &= \|\boldsymbol{r}^\top \boldsymbol{x}\|_2^2 + \|\boldsymbol{s}^\top \boldsymbol{x}\|_2^2 \\ &\geq 0\end{aligned}$$

Thus, Y must be positive semidefinite. Note that I was fairly strict in requiring *some* argument to show that Y is positive definite to get full credit.[10]

[10] An incorrect argument I sometimes saw was that since $p(x) = \boldsymbol{t}_k^\top Y \boldsymbol{t}_k \geq 0$, Y must be positive semidefinite. This, however, is not quite right, because the inequality $\boldsymbol{t}_k^\top Y \boldsymbol{t}_k \geq 0$ does not hold for *every* vector $\boldsymbol{t}_k$ – it only holds for vectors of the form $(1, t, \cdots, t^k)$, and therefore proves nothing about the positive semidefiniteness of Y. Indeed, if that was enough of an argument, then the whole sum-of-squares theorem wouldn't be needed to prove this!

Looking back at the notes for review session 3, though, I found that my solution there was very ambiguous and might have led you to believe that this was the correct answer. I therefore didn't penalize anyone who gave that justification.

Part B

For a polynomial of the form above, reformulate the optimization

$$\min_{x \in \mathbb{R}} p(x)$$

as a semidefinite program.

Solution

Consider that $\min_{x \in \mathbb{R}} p(x)$ is equivalent to

$$\begin{aligned}\max_{(x,t)} \quad & t \\ \text{s.t.} \quad & t \leq p(x)\end{aligned}$$

(The best way to see this is to consider a line $y = t$. This program is equivalent to pushing this line up until it touches $p(x)$. This clearly finds the minimum of $p(x)$).

Re-arranging the inequality, we find that this is equivalent to

$$\begin{aligned}\max_{t} \quad & t \\ \text{s.t.} \quad & p(x) - t \geq 0 \quad \forall x\end{aligned}$$

Re-stated, this constraint basically requires the polynomial $q(x) = p(x) - t$ to be positive for all $\mathbb{R}$. The coefficients of this polynomial are

$$q_0 = p_0 - t$$

$$q_i = p_i \qquad \forall i = 1, \cdots, 2k$$

Using part A, the statement that q should be positive is equivalent to the statement that these coefficients can be assembled into a positive semidefinite matrix. Thus, this program is equivalent to[11]

[11] Incorrect variations I saw included putting $\boldsymbol{t}_k^\top Y \boldsymbol{t}_k$ in the objective (this is not linear and therefore not an SDP – and in fact, it's not even quadratic; it's a polynomial of order $2k$!) or maximizing t with respect to the constraint $\boldsymbol{t}_k^\top Y \boldsymbol{t}_k - t$ (though the objective is now linear, this constraint is neither linear nor quadratic – it's of order $2k$). The key error of these solutions was keeping $\boldsymbol{t}_k$ as a decision variable, which is bound to fail because any such formulation must include terms of the form t^{2k}.

$$\begin{aligned} \max_{t,Y} \quad & t \\ \text{s.t.} \quad & Y_{11} = p_0 - t \\ & \left(\sum_{m+n=i+2} Y_{mn} \right) = p_i \quad \forall i \neq 0 \\ & Y \succeq 0 \end{aligned}$$

■ □ ■

Question 7 (Minimizing a Function Over the Probability Simplex)

Show that the necessary and sufficient condition to ensure that $\hat{\boldsymbol{x}}$ minimizes $f(\boldsymbol{x})$ over the probability simplex

$$\mathscr{P} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq \mathbf{0}, \mathbf{1} \cdot \boldsymbol{x} = 1\}$$

is given by

$$\exists \boldsymbol{g} \in \partial f(\hat{\boldsymbol{x}}) \text{ s.t. } \min_{1 \leq i \leq n} (g_i) \geq \boldsymbol{g} \cdot \hat{\boldsymbol{x}}$$

Solution

From the result developed in your homework, the necessary and sufficient condition can be written

$$\exists \boldsymbol{g} \in \partial f(\hat{\boldsymbol{x}}) \text{ s.t. } \boldsymbol{g} \cdot \boldsymbol{y} \geq \boldsymbol{g} \cdot \hat{\boldsymbol{x}} \quad \forall \boldsymbol{y} \in \mathscr{P}$$

This can be written as

$$\exists \boldsymbol{g} \in \partial f(\hat{\boldsymbol{x}}) \text{ s.t. } \min_{\boldsymbol{y} \in \mathscr{P}} (\boldsymbol{g} \cdot \boldsymbol{y}) \geq \boldsymbol{g} \cdot \hat{\boldsymbol{x}}$$

Now, consider the term $\min_{\boldsymbol{y} \in \mathscr{P}} (\boldsymbol{g} \cdot \boldsymbol{y})$ on the right. Here are three different ways to show that the optimal value of this LP is equal to the smallest component of $\boldsymbol{g}$. I was, once again, pretty strict in requiring a waterproof definition.

Extreme points : One person tried to give this argument but didn't quite get there.

The program is an LP. By standard results for linear programming, the solution of an LP occurs at one of the extreme points of the feasible set $\mathscr{P}$. In this case, the extreme points are the vectors $\boldsymbol{e}^{(i)}$, containing a 1 in position i and a zero everywhere else. Clearly, each of these vectors dotted with $\boldsymbol{g}$ produces $\boldsymbol{g}_i$. Thus, the optimal value of the minimization is the smallest component of $\boldsymbol{g}$, as stated in the question.

Duality : No-one tried this.

To show the above rigorously, we can use duality. Consider that the program $\min_{\boldsymbol{y}\in\mathscr{P}}(\boldsymbol{g}\cdot\boldsymbol{y})$ can be written as

$$\begin{aligned} \min \quad & \boldsymbol{g}\cdot\boldsymbol{y} \\ \text{s.t.} \quad & \mathbf{1}\cdot\boldsymbol{y} = 1 \\ & \boldsymbol{x} \geq 0 \end{aligned}$$

This is an LP, and its dual is

$$\begin{aligned} \max \quad & \lambda \\ & \lambda \leq g_i \quad \forall i \end{aligned}$$

Clearly, the optimal value of the program is the smallest component of $\boldsymbol{g}$. By strong duality for LPs, this is therefore also the optimal value of the primal, hence the result in the question.

Method 3 : Most people who got it right did it this way.

Showing

$$\min_{\boldsymbol{y}\in\mathscr{P}}(\boldsymbol{g}\cdot\boldsymbol{y}) \leq \min_{1,\cdots,n} g_i$$

is not hard, because the feasible region of the program on the right is smaller than that of the program on the left, so that on the left attains a "better" (ie: lower) value.

To prove the other direction, note that for all $\boldsymbol{y}\in\mathscr{P}$

$$\begin{aligned} \boldsymbol{g}\cdot\boldsymbol{y} &= \sum_{i=1}^{n} y_i g_i \\ &\geq \sum_{i=1}^{n} y_i \min_i(g_i) \\ &= \min_i(g_i)\sum_{i=1}^{n} y_i \\ &\stackrel{12}{=} \min_i(g_i) \end{aligned}$$

Thus, every feasible point of the program on the LHS is greater or equal to a feasible point on the RHS. Thus, [13]

$$\min_{\boldsymbol{y}\in\mathscr{P}}(\boldsymbol{g}\cdot\boldsymbol{y}) \geq \min_{1,\cdots,n} g_i$$

■ □ ■

[12] Because $\boldsymbol{y}\in\mathscr{P}$, so $\sum_{i=1}^{n} y_i = 1$

[13] Or, a variation on this theme, if

$$g_i \geq \boldsymbol{g}\cdot\hat{\boldsymbol{x}} \quad \forall i$$

then multiplying both sides by y_i for $\boldsymbol{y}\in\mathscr{P}$ and summing over i gives

$$\boldsymbol{g}\cdot\boldsymbol{y} \geq \boldsymbol{g}\cdot\hat{\boldsymbol{x}} \quad \forall\boldsymbol{y}\in\mathscr{P}$$

Another variation on the same theme is simply to note that

$$\boldsymbol{y}\in\mathscr{P} \Rightarrow \boldsymbol{y}\in\text{conv}(\boldsymbol{e}^{(i)})$$

.

Question 8 (ℓ_1-minimization) ____________

Part A

Prove the following primal-dual relationship

$$\begin{array}{lll} \max & \boldsymbol{c}\cdot\boldsymbol{p}-\lambda\|\boldsymbol{p}-\boldsymbol{q}\|_1 \\ \text{s.t.} & \boldsymbol{p}\geq \boldsymbol{0} \\ & \boldsymbol{1}\cdot\boldsymbol{p}=1 \end{array} = \begin{array}{ll} \min & \boldsymbol{w}\cdot\boldsymbol{q}+\beta \\ \text{s.t.} & \boldsymbol{w}+\beta\boldsymbol{1}\geq \boldsymbol{c} \\ & \|\boldsymbol{w}\|_\infty \leq \lambda \end{array}$$

Where $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$ denotes the ℓ_1 norm of the vector $\boldsymbol{x}$ and the constant λ, the vector $\boldsymbol{q}\in\mathbb{R}^n_+$ and $\boldsymbol{c}\in\mathbb{R}^n$ are problem data. Assume that $\boldsymbol{q}$ is a probability mass function (ie: $\boldsymbol{1}\cdot\boldsymbol{q}=1$).

Solution (Method 1)

First, note that the problem is only convex for $\lambda \geq 0$.

Now, substitute $\boldsymbol{u}=\boldsymbol{p}-\boldsymbol{q}$. Bearing in mind that $\boldsymbol{1}\cdot\boldsymbol{q}=1$, the primal problem becomes

$$\begin{array}{ll} \max & \boldsymbol{c}\cdot\boldsymbol{u}+\boldsymbol{c}\cdot\boldsymbol{u}-\lambda\|\boldsymbol{u}\|_1 \\ \text{s.t.} & \boldsymbol{u}+\boldsymbol{q}\geq\boldsymbol{0} \\ & \boldsymbol{1}\cdot\boldsymbol{u}=0 \end{array}$$

The Lagrangian for this problem is (with $\boldsymbol{w}\geq\boldsymbol{0}$)[14]

$$\begin{aligned} \mathcal{L}(\boldsymbol{u},\boldsymbol{\mu},\beta) &= \boldsymbol{c}\cdot\boldsymbol{u}+\boldsymbol{c}\cdot\boldsymbol{q}-\lambda\|\boldsymbol{u}\|_1 \\ &\quad +\boldsymbol{\mu}\cdot(\boldsymbol{u}+\boldsymbol{q})-\beta\boldsymbol{1}\cdot\boldsymbol{u} \\ &= (\boldsymbol{c}+\boldsymbol{\mu})\cdot\boldsymbol{q} \\ &\quad +(\boldsymbol{c}+\boldsymbol{\mu}-\beta\boldsymbol{1})\cdot\boldsymbol{u}-\lambda\|\boldsymbol{u}\|_1 \\ &= (\boldsymbol{c}+\boldsymbol{\mu})\cdot\boldsymbol{q} \\ &\quad +\sum_{i=1}^n\left[(c_i+w_i-\beta)u_i-\lambda|u_i|\right] \end{aligned}$$

Now, consider maximizing this with respect to $\boldsymbol{u}$. $\mathcal{L}$ is basically a piecewise linear function of each component of $\boldsymbol{u}$, with a discontinuity at 0. We need *all* of the individual gradients to be positive for $[-\infty,0]$ and negative for $[0,\infty]$, or else we can make $\mathcal{L}$ shoot up to ∞. In other words, we need

$$-\lambda \leq \max_i\,[c_i+\mu_i-\beta] \leq \lambda$$

More tidily, this becomes

$$\|\boldsymbol{c}+\boldsymbol{\mu}-\beta\boldsymbol{1}\|_\infty \leq \lambda$$

[14] The minus sign in front of the β is perfectly legal, but unusual – it's there to make sure our dual ends up in the required form. We could just as easily use $+\beta$ and later substitute $\tilde{\beta}=-\beta$ (since β is unconstrained).

So the dual is

$$\begin{array}{rl} \min & (\boldsymbol{c}+\boldsymbol{\mu})\cdot \boldsymbol{q} \\ \text{s.t.} & \|\boldsymbol{c}+\boldsymbol{\mu}-\beta\mathbf{1}\|_\infty \le \lambda \\ & \boldsymbol{\mu} \ge 0 \end{array}$$

Substituting $\boldsymbol{w} = \boldsymbol{c}+\boldsymbol{\mu}-\beta\mathbf{1}$, and remembering that $\mathbf{1}\cdot\boldsymbol{q} = 1$, this becomes

$$\begin{array}{rl} \min & \boldsymbol{w}\cdot\boldsymbol{q}+\beta \\ \text{s.t.} & \|\boldsymbol{w}\|_\infty \le \lambda \\ & \boldsymbol{w}+\beta\mathbf{1} \ge \boldsymbol{c} \end{array}$$

The Slater Conditions are clearly met for this primal, so by Strong Duality, the optimal solutions of both problems are equal.

Solution (Method 2)

Once again, note that the primal is only convex for $\lambda \ge 0$.

Now, consider that the problem can be written as

$$\max_{\boldsymbol{p}\ge 0, \mathbf{1}\cdot\boldsymbol{p}=1} \left\{ \boldsymbol{c}\cdot\boldsymbol{p} - \max_{\|\boldsymbol{w}\|_\infty\le\lambda} [\boldsymbol{w}\cdot(\boldsymbol{p}-\boldsymbol{q})] \right\}$$

Changing the inner max to a min (and noting that we can interchange $\boldsymbol{p}-\boldsymbol{q}$ with $\boldsymbol{q}-\boldsymbol{p}$ since the optimal value of $\boldsymbol{w}$ will take care of all relevant signs)

$$\max_{\boldsymbol{p}\ge 0, \mathbf{1}\cdot\boldsymbol{p}=1} \min_{\|\boldsymbol{w}\|_\infty\le\lambda} \{\boldsymbol{c}\cdot\boldsymbol{p} - \boldsymbol{w}\cdot(\boldsymbol{p}-\boldsymbol{q})\}$$

Now, note that the sets over which we are optimizing are compact and that the objective is bi-linear. We can therefore apply the min-max theorem and re-write this as

$$\min_{\|\boldsymbol{w}\|_\infty\le\lambda} \max_{\boldsymbol{p}\ge 0, \mathbf{1}\cdot\boldsymbol{p}=1} \{\boldsymbol{c}\cdot\boldsymbol{p} - \boldsymbol{w}\cdot(\boldsymbol{p}-\boldsymbol{q})\}$$

Collecting terms

$$\min_{\|\boldsymbol{w}\|_\infty\le\lambda} \left\{ \boldsymbol{w}\cdot\boldsymbol{q} + \max_{\boldsymbol{p}\ge 0, \mathbf{1}\cdot\boldsymbol{p}=1} [(\boldsymbol{c}-\boldsymbol{w})\cdot\boldsymbol{p}] \right\}$$

The inner term is an LP that clearly satisfies the Slater Conditions. Taking its dual, this becomes

$$\min_{\|\boldsymbol{w}\|_\infty\le\lambda} \left\{ \boldsymbol{w}\cdot\boldsymbol{q} + \min_{\beta, \beta\mathbf{1}+\boldsymbol{w}\ge\boldsymbol{c}} \beta \right\}$$

As required.

Solution (Method 3)

Just express the primal as an LP and find the dual!

Part B

Show that the dual can be simplified to the following 1-dimensional optimization problem

$$\min \left\{ \sum_{i=1}^{n} q_i \max(c_i - \beta + \lambda, 0) + \beta - \lambda : \beta \geq \max_{1 \leq i \leq n} (c_i) - \lambda \right\}$$

Show that this program can be solved in $\mathcal{O}(n \log n)$ time.

Solution

First, let $\boldsymbol{u} = \boldsymbol{w} + \lambda \mathbf{1}$. Since $\|\boldsymbol{w}\|_\infty \leq \lambda$, it follows that

$$\mathbf{0} \leq \boldsymbol{u} \leq 2\lambda \mathbf{1}$$

Furthermore, remembering that $\mathbf{1} \cdot \boldsymbol{q} = 1$,

$$\boldsymbol{w} \cdot \boldsymbol{q} + \beta = \boldsymbol{q} \cdot \boldsymbol{u} - (\mathbf{1} \cdot \boldsymbol{q})\lambda + \beta = \boldsymbol{q} \cdot \boldsymbol{u} - \lambda + \beta$$

As such, the dual is equivalent to the following LP

$$\begin{aligned} \min \quad & \boldsymbol{q} \cdot \boldsymbol{u} - \lambda + \beta \\ \text{s.t.} \quad & 0 \leq \boldsymbol{u} \leq 2\lambda \mathbf{1} \\ & \boldsymbol{u} + (\beta - \lambda)\mathbf{1} \geq \boldsymbol{c} \end{aligned}$$

Writing this component-wise, we have that

$$\begin{aligned} \min \quad & \boldsymbol{q} \cdot \boldsymbol{u} - \lambda + \beta \\ \text{s.t.} \quad & 2\lambda \geq u_i \geq \max(c_i - \beta + \lambda, 0) \end{aligned}$$

Now, consider that

- To ensure that this is feasible, we need

  $$c_i - \beta + \lambda \leq 2\lambda \quad \forall i$$

  or in other words

  $$\beta \geq \max_i (c_i) - \lambda$$

- Since every component of $\boldsymbol{q}$ is positive and the constraints on each component of u are independent, we will clearly choose the solution with the smallest $\boldsymbol{u}$ – ie: $u_i = \max(c_i, -\beta + \lambda, 0)$.

Thus, this LP is equivalent to

$$\begin{aligned} \min_{\beta} \quad & \left(\sum_{i=1}^{n} q_i \max(c_i - \beta + \lambda, 0) \right) - \lambda + \beta \\ \text{s.t.} \quad & \beta \geq \max_i (c_i) - \lambda \end{aligned}$$

Finally, note that the objective is a piecewise linear function of β. Thus, if we sort the vector $\boldsymbol{c}$ (average complexity $\mathcal{O}(n \log n)$), we immediately have a trivial bound on β and we can easily find the point at which the gradient of the line goes from negative to positive. Thus, sorting $\boldsymbol{c}$ gives us the solution.

It turns out that we can do even better. Indeed, since the q_i sum to 1, the $+\beta$ in the objective can be brought into the sum and then into the maximum, to obtain

$$\begin{aligned} \min_{\beta} \quad & \left(\sum_{i=1}^{n} q_i \max(c_i + \lambda, \beta) \right) - \lambda \\ \text{s.t.} \quad & \beta \geq \max_i (c_i) - \lambda \end{aligned}$$

In this form, it is clear the function is strictly increasing in β, and therefore the optimal solution is $\beta = \max_i(c_i) - \lambda$. This can be found in $\mathcal{O}(n)$ time – we simply need to look through all n items of $\boldsymbol{c}_i$ to find the maximum.

■ □ ■

Questions by Prof Garud Iyengar. Solutions under the license below.