Homework 10

Self grades are due at 11 PM on April 13, 2023.

1. **Does strong duality hold?**

Consider

$$\min_{(x,y)\in\mathcal{D}} e^{-x} \tag{1}$$

$$\text{s.t. } x^2/y \le 0 \tag{2}$$

where $\mathcal{D} = \{(x,y) \mid y > 0\}$.

(a) Prove the problem is convex. Find the optimal value. *HINT: To prove the constraint function is convex, you will have to prove it is convex with respect to the vector* $\begin{bmatrix} x & y \end{bmatrix}^\top$. *Consider computing the Hessian of the constraint function, its determinant and trace, and show that it is PSD by analyzing signs of its eigenvalues.*

Solution: The second derivative of the objective function is e^{-x} which is non-negative thus the objective is a convex function. Furthermore, the constraint is jointly convex as can be verified by showing the Hessian is PSD (additionally one may notice this is the perspective function of x^2 which will be convex). The Hessian for $g(x,y) = \frac{x^2}{y}$ is given by,

$$\nabla^2 g(x,y) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix}. \tag{3}$$

Suppose λ_1, λ_2 are the eigenvalues of $\nabla^2 g(x,y)$. The determinant of the Hessian is 0 which gives us $\lambda_1\lambda_2 = 0$. Further trace of the Hessian is $\frac{2}{y} + \frac{2x^2}{y^3} > 0$ (since $y > 0$) which gives $\lambda_1 + \lambda_2 > 0$. Thus one eigenvalue must be positive and the other must be 0 which shows that the Hessian is positive semidefinite. Hence the problem is convex. Furthermore, the only feasible value of x is $x = 0$. Hence the optimal value is $e^{-0} = 1$.

(b) Next, we will proceed to find an optimal solution and an optimal value for the dual problem. The Lagrangian dual function $g(\lambda)$, can be written as:

$$g(\lambda) = \inf_{(x,y)\in\mathcal{D}} \left(e^{-x} + \lambda \frac{x^2}{y} \right). \tag{4}$$

Explain why $g(\lambda)$ is lower bounded by 0 for $\lambda \ge 0$. **Note: Here we are not dualizing the constraint $y > 0$ that is in the definition of $\mathcal{D}$ — this is only dualizing the other constraint.**

Solution: This is true since both terms in the sum are non-negative because $y > 0$.

(c) Show that $g(\lambda) = 0$ for $\lambda \ge 0$. *HINT: To show that the infimum in Equation (4) is 0, we want to show there exist (x,y) such that both e^{-x} and $\lambda\frac{x^2}{y}$ can get arbitrarily close to 0. HINT: Consider a sequence $\{x_k\}$ going to $+\infty$ and a sequence $\{y_k\}$ also going to $+\infty$ such that $\lim_{k\to\infty} \frac{x_k^2}{y_k} = 0$. Simply put, we want to drive x to infinity in order to drive e^{-x} to 0, while having y grow faster than x^2, so that the second term also goes to 0.*

Solution: To show that the infimum is 0, we pick any sequence $\{x_k\}$ going to $+\infty$ and pick a sequence $\{y_k\}$ also going to $+\infty$ such that $\lim_{k\to\infty} \frac{x_k^2}{y_k} = 0$. One example of such sequence pair is $y_k = x_k^4$ and $x_k = 2k$.
This gives $\lim_{k\to\infty} e^{-x_k} + \frac{x_k^2}{y_k} = 0$, so $g(\lambda) = \inf_{(x,y)\in\mathcal{D}} e^{-x} + \lambda\frac{x^2}{y} = 0$. Note that the (x_k, y_k) pair as $k \to \infty$ does not satisfy the constraint of the primal problem $x^2/y \leq 0$ but allows x^2/y to get arbitrarily close to 0.

(d) Now, write the dual problem and find an optimal solution λ^* and an optimal value d^* for the dual problem using the results above. What is the duality gap?

Solution: The Lagrange dual problem is

$$d^* = \sup_{\lambda\geq 0} g(\lambda) \tag{5}$$

where $g(\lambda) = \inf_{(x,y)\in\mathcal{D}} \ e^{-x} + \lambda x^2/y$. Note $g(\lambda) = 0$ from previous part. It follows from the previous problems that $d^* = 0$ and any $\lambda \geq 0$ is optimal. The duality gap is 1. You can also write your answer using maximum instead of the supremum as:

$$d^* = \max_{\lambda\geq 0} g(\lambda) \tag{6}$$

(e) Does Slater's Condition hold for this problem? Does Strong Duality hold?

Solution: While the primal problem is convex, we cannot find a point that is strictly in the interior of the domain and satisfies the constraint as needed for Slater's condition. Specifically, for Slater's condition to hold we need the existence of an (x, y) pair such that $x^2/y < 0$. Note there is no such pair (x, y) since $y > 0$ and $x^2 \geq 0$. Hence Slater's condition does not hold for this problem.
From the previous parts we saw that $p^* \neq d^*$, and thus strong duality does not hold. Furthermore, the problem is convex. For convex problems we know that if Slater's condition holds then we must have strong duality (i.e Slater's is a sufficient condition). However since strong duality does not hold it implies that Slater's does not hold.
Note that this problem is an example illustrating that **convexity alone is not enough to guarantee strong duality for an optimization problem**.

2. LP at Boundary

Consider the LP:

$$\min_{x\in\mathbb{R}^n} \quad \vec{c}^\top \vec{x} \tag{7}$$
$$\text{s.t.} \quad A\vec{x} \leq \vec{b}. \tag{8}$$

Here we have non-zero $\vec{c} \in \mathbb{R}^n$, $\vec{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $\vec{b} \in \mathbb{R}^m$. The feasible set forms a polyhedron

$$\mathcal{P} = \{\vec{x} \in \mathbb{R}^n \mid \vec{a}_i^\top \vec{x} - b_i \leq 0,\ 1 \leq i \leq n\} \tag{9}$$
$$= \bigcap_{i=1}^{m} \{\vec{x} \in \mathbb{R}^n \mid \vec{a}_i^\top \vec{x} - b_i \leq 0\}. \tag{10}$$

In the second line, $\mathcal{P}$ is defined as the intersection of half-spaces. When this set is bounded, it is often referred to as a polytope instead.

In this problem we will prove facts about linear programs, including the crucial fact that the solution to an LP exists at the boundary of a polytope.

(a) First, consider the case where we have an unbounded polyhedron. For any such $\mathcal{P}$, assume there exists some $\vec{x}_0, \vec{v} \in \mathbb{R}^n$ with $\vec{x}_0 \neq \vec{v}$ such that $\mathcal{L} := \{\vec{x}_0 + \alpha\vec{v} \mid \alpha \in [0, \infty)\} \subset \mathcal{P}$. This is saying that there exists a line segment, unbounded on one side, that is contained within $\mathcal{P}$. Show that if $\vec{c}^\top \vec{v} < 0$, then $p^* = -\infty$.

Solution: If $\vec{c}^\top \vec{v} < 0$, then for any point $\vec{y}_\alpha = \vec{x}_0 + \alpha\vec{v} \in \mathcal{L}$, we have that

$$\vec{c}^\top \vec{y}_\alpha = \vec{c}^\top \vec{x}_0 + \alpha\vec{c}^\top \vec{v}. \tag{11}$$

If we take $\alpha \to \infty$,

$$\lim_{\alpha\to\infty} \vec{c}^\top \vec{y}_\alpha = \vec{c}^\top \vec{x}_0 + \lim_{\alpha\to\infty} \alpha\vec{c}^\top \vec{v} \tag{12}$$
$$= -\infty. \tag{13}$$

(b) Now, suppose our feasible set is defined by a bounded polytope (i.e. a polytope that contains its boundary). We say a point belongs to the interior of $\mathcal{P}$ (i.e. $\vec{x}_0 \in \text{Int}(\mathcal{P})$) if there exists some ball of radius $\epsilon > 0$ such that $\mathcal{B} := \{\vec{y} \in \mathbb{R}^n \mid \|\vec{y} - \vec{x}_0\|_2 \leq \epsilon\} \subset \mathcal{P}$. Show that the optimal point for the LP cannot be obtained in the interior of $\mathcal{P}$ and must be obtained on the boundary (that is, when for some $1 \leq i \leq n$, $\vec{a}_i^\top \vec{x} - b_i = 0$). For this, consider a proof by contradiction. Show that for any x_0 on the interior, there exists another point

$$\vec{x}_1 := \vec{x}_0 - \epsilon\frac{\vec{c}}{\|\vec{c}\|_2}, \tag{14}$$

such that $\vec{x}_1 \in \mathcal{P}$ and $\vec{c}^\top \vec{x}_1 < \vec{c}^\top \vec{x}_0$.

Solution: We begin by showing that $\vec{x}_1 \in \mathcal{P}$. This is done by showing that $\vec{x}_1 \in \mathcal{B} \subset \mathcal{P}$:

$$\|\vec{x}_1 - \vec{x}_0\|_2 = \left\|\epsilon\frac{\vec{c}}{\|\vec{c}\|_2}\right\|_2 \tag{15}$$
$$= \epsilon. \tag{16}$$

Since $\vec{x}_1$ is within an ϵ-ball of $\vec{x}_0$, we must have that $\vec{x}_1 \in P$.

Next, we show that $\vec{x}_1$ is more optimal:

$$\begin{aligned} \vec{c}^\top \vec{x}_1 &= \vec{c}^\top \vec{x}_0 - \epsilon \frac{\vec{c}^\top \vec{c}}{\|\vec{c}\|_2} && (17) \\ &= \vec{c}^\top \vec{x}_0 - \epsilon \|\vec{c}\|_2 && (18) \\ &< \vec{c}^\top \vec{x}_0. && (19) \end{aligned}$$

Note: it is also entirely possible to show that $\mathcal{B} \subset \text{Int}(\mathcal{P})$ in which case we can prove a stronger condition: namely, that given any point in the interior of the polytope, we can find another point, also in the interior, that is more optimal.

3. Sensitivity and Dual Variables

In this problem, we look into the interpretation of dual variables as sensitivity parameters of the primal problem. Recall the canonical, convex primal problem

$$\min_{\vec{x}} \quad f_0(\vec{x}) \tag{20}$$

$$\text{subject to} \quad f_i(\vec{x}) \leq 0, \quad i = 1, \ldots, m \tag{21}$$

$$h_j(\vec{x}) = 0, \quad j = 1, \ldots, p \tag{22}$$

where f_0, f_i are convex and h_j are affine (assume the problem has strong duality). Consider the *perturbed* problem

$$\min_{\vec{x}} \quad f_0(\vec{x}) \tag{23}$$

$$\text{subject to} \quad f_i(\vec{x}) \leq u_i, \quad i = 1, \ldots, m \tag{24}$$

$$h_j(\vec{x}) = v_j, \quad j = 1, \ldots, p \tag{25}$$

and define

$$p^\star(\vec{u}, \vec{v}) = \min\{f_0(\vec{x}) \mid \vec{x} \text{ such that } f_i(\vec{x}) \leq u_i \quad \forall i, h_j(\vec{x}) = v_j \quad \forall j\} \tag{26}$$

In words, $p^\star(\vec{u}, \vec{v})$ is the optimal value for the perturbed problem if it is feasible, and defined to be $+\infty$ (infeasible) otherwise. Note $p^\star(\vec{0}, \vec{0})$ is the original problem.

(a) Prove $p^\star(\vec{u}, \vec{v})$ is jointly convex in $(\vec{u}, \vec{v})$.

HINT: Consider $\mathcal{D} = \{(\vec{x}, \vec{u}, \vec{v}) \mid f_i(\vec{x}) \leq u_i \quad \forall i, h_j(\vec{x}) = v_j \quad \forall j\}$, which is the set of triples $(\vec{x}, \vec{u}, \vec{v})$ such that $\vec{x}$ is a feasible point for the perturbed problem with the perturbations $(\vec{u}, \vec{v})$. Is $\mathcal{D}$ convex? Also, define $F(\vec{x}, \vec{u}, \vec{v})$ to be a function that is equal to $f_0(\vec{x})$ on $\mathcal{D}$ and $+\infty$ otherwise.

HINT: Prove and use the following lemma:

Let S_1, S_2 be convex sets with a function $f : S_1 \times S_2 \to \mathbb{R}$ which is jointly convex in both arguments. Define $g(x) = \min_{y \in S_2} f(x, y)$. Then $g(x)$ is convex in $x \in S_1$.

Solution: By checking the definition of convexity and using the fact that f_i are convex and h_i are affine, we have that $\mathcal{D}$ is a convex. We also have that $f_0(\vec{x})$ is jointly convex in $(\vec{x}, \vec{u}, \vec{v})$, since it is convex in $\vec{x}$ and does not depend on the other two variables $\vec{u}, \vec{v}$. Hence F is convex. Finally

$$p^\star(\vec{u}, \vec{v}) = \min_{\vec{x}} F(\vec{x}, \vec{u}, \vec{v}) \tag{27}$$

is the pointwise minimization of a convex function.

Now, we prove the lemma mentioned in the hint. For the functions $f(x, y), g(x)$ defined in the hint, consider

$$\begin{aligned}
&g(\theta x_1 + (1-\theta)x_2) \leq f(\theta x_1 + (1-\theta)x_2, y) \ \forall \ y \in S_2 \\
\implies &g(\theta x_1 + (1-\theta)x_2) \leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \ \forall \ y_1, y_2 \in S_2 \\
\implies &g(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\
\implies &g(\theta x_1 + (1-\theta)x_2) \leq \theta \min_y f(x_1, y) + (1-\theta) \min_y f(x_2, y) = \theta g(x_1) + (1-\theta) g(x_2)
\end{aligned}$$

where $\theta \in [0, 1]$, which shows that $g(x)$ is indeed convex. In our case, $p^\star(\vec{u}, \vec{v})$ is analogous to $g(x)$, and $\min_{\vec{x}} F(\vec{x}, \vec{u}, \vec{v})$ is analogous to $\min_y f(x, y)$. Hence $p^\star(\vec{u}, \vec{v})$ is convex in $(\vec{u}, \vec{v})$.

(b) Assume that strong duality holds, and that the dual optimum is attained. Let $(\vec{\lambda}^\star, \vec{\nu}^\star)$ be the optimal dual variables for the dual of the unperturbed primal problem (20). Show that for any point $\hat{\vec{x}}$ that is feasible for the perturbed problem, we have

$$f_0(\hat{\vec{x}}) \geq p^\star(0,0) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\nu}^\star \tag{28}$$

HINT: Consider the Lagrangian for the original problem $L(\vec{x}, \vec{\lambda}, \vec{\nu})$ and $g(\vec{\lambda}, \vec{\nu}) = \min_{\vec{x}} L(\vec{x}, \vec{\lambda}, \vec{\nu})$. Now relate this to the perturbed problem.

Solution: Let $L(\vec{x}, \vec{\lambda}, \vec{\nu})$ be the Lagrangian for the original problem, and $g(\vec{\lambda}, \vec{\nu}) = \min_{\vec{x}} L(\vec{x}, \vec{\lambda}, \vec{\nu})$ be the dual function.

First, consider a $\hat{\vec{x}}$ which is feasible for the perturbed problem, i.e.,

$$f_i(\hat{\vec{x}}) \leq u_i, \quad i = 1, \ldots, m \tag{29}$$
$$h_j(\hat{\vec{x}}) = v_j, \quad j = 1, \ldots, p. \tag{30}$$

Now, consider the unperturbed problem, whose optimal solution is given by $p^\star(\vec{0}, \vec{0})$. By strong duality,

$$p^\star(\vec{0}, \vec{0}) = d^\star = g(\vec{\lambda}^\star, \vec{\nu}^\star) \leq \mathcal{L}(\vec{x}, \vec{\lambda}^\star, \vec{\nu}^\star) \qquad \forall \vec{x}. \tag{31}$$

Since the inequality holds for all $\vec{x}$ (feasible or not), it holds for all $\hat{\vec{x}}$ which are feasible for the perturbed problem. Then we have

$$\begin{aligned} p^\star(\vec{0}, \vec{0}) &= g(\vec{\lambda}^\star, \vec{\nu}^\star) && (32)\\ &\leq L(\hat{\vec{x}}, \vec{\lambda}^\star, \vec{\nu}^\star) && (33)\\ &= f_0(\hat{\vec{x}}) + \sum_{i=1}^{m} \lambda_i^\star f_i(\hat{\vec{x}}) + \sum_{j=1}^{p} \nu_j^\star h_j(\hat{\vec{x}}) && (34)\\ &\leq f_0(\hat{\vec{x}}) + \sum_{i=1}^{m} \lambda_i^\star u_i + \sum_{j=1}^{p} \nu_j^\star v_j. && (35) \end{aligned}$$

Rearranging terms, we get

$$f_0(\hat{\vec{x}}) \geq p^\star(\vec{0}, \vec{0}) - \sum_{i=1}^{m} \lambda_i^\star u_i - \sum_{j=1}^{p} \nu_j^\star v_j \tag{36}$$

(c) Using the previous subpart, show that for all $\vec{u}, \vec{v}$, we have

$$p^\star(\vec{u}, \vec{v}) \geq p^\star(\vec{0}, \vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\nu}^\star. \tag{37}$$

Solution: From this previous subpart, we have,

$$f_0(\hat{\vec{x}}) \geq p^\star(\vec{0}, \vec{0}) - \sum_{i=1}^{m} \lambda_i^\star u_i - \sum_{j=1}^{p} \nu_j^\star v_j \tag{38}$$

This statement is true for any $\hat{\vec{x}}$, i.e., for any feasible point for the perturbed problem. The right hand side, however, does not depend on $\hat{\vec{x}}$, and in fact is a constant (with respect to $\hat{\vec{x}}$, though certainly not u and v). Thus, we may as well take the minimum on both sides with respect to all feasible $\hat{\vec{x}}$, so as to get the tightest possible bound for the right-hand side. This yields

$$p^\star(u, v) = \min_{\substack{\hat{\vec{x}} \\ f_i(\hat{\vec{x}}) \leq u_i, i=1,\ldots,m \\ h_j(\hat{\vec{x}}) = v_j, j=1,\ldots,p}} f_0(\hat{\vec{x}}) \tag{39}$$

$$\geq \min_{\substack{\hat{\vec{x}} \\ f_i(\hat{\vec{x}}) \leq u_i, i=1,\ldots,m \\ h_j(\hat{\vec{x}}) = v_j, j=1,\ldots,p}} \left(p^\star(\vec{0},\vec{0}) - \sum_{i=1}^{m} \lambda_i^\star u_i - \sum_{j=1}^{p} \nu_j^\star v_j \right) \tag{40}$$

$$= p^\star(\vec{0},\vec{0}) - \sum_{i=1}^{m} \lambda_i^\star u_i - \sum_{j=1}^{p} \nu_j^\star v_j \tag{41}$$

$$= p^\star(\vec{0},\vec{0}) - \vec{u}^\top \vec{\lambda}^\star - \vec{v}^\top \vec{\nu}^\star \tag{42}$$

which is what we wanted to show.

(d) Suppose we only have 1 equality and 1 inequality constraint (that is u, v are scalars). For each of the following situations, argue whether the value of $p^\star(u,v)$ increases or decreases as compared to $p^\star(0,0)$ or whether we are unsure as to whether it increases or decreases.

i. If $\lambda^\star$ is large and we pick $u < 0$.

ii. If $\lambda^\star$ is large and we pick $u > 0$.

iii. If $\nu^\star$ is large and positive (resp. negative) and we pick $v < 0$ (resp. $v > 0$).

Solution:

i. We see that $-\lambda^\star u \geq 0$ and hence $p^\star(u,v)$ increases greatly since $\lambda^\star$ is large.

ii. We cannot say anything since $-\lambda^\star u \leq 0$ and since the relationship derived in (b) is a lower bound we cannot say with certainty that $p^\star(u,v)$ will decrease.

iii. Similar to case 1, $p^\star(u,v)$ increases greatly.

4. KKT with circles

Consider the problem

$$\min_{\vec{x}\in\mathbb{R}^2} \quad x_1^2 + x_2^2 \tag{43}$$
$$\text{s.t.} \quad (x_1-1)^2 + (x_2-1)^2 \leq 4 \tag{44}$$
$$(x_1-1)^2 + (x_2+1)^2 \leq 4 \tag{45}$$

where $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top \in \mathbb{R}^2$.

(a) Sketch the feasible region and the level sets of the objective function. Find the optimal point $\vec{x}^*$ and the optimal value p^*.

Solution:

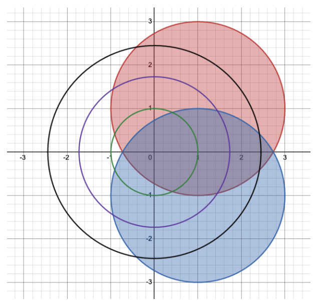

Figure 1: Feasible region and level sets corresponding to $\min_{\vec{x}\in\mathbb{R}^2} \quad x_1^2 + x_2^2$.

The feasible region is given by the purple area in Figure 1. The optimal solution is the closest point to the origin inside the feasible region. Since the origin is an element of this feasible region, we have $\vec{x}^* = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$, and $p^* = 0$.

(b) Does strong duality hold?

Solution: The problem is convex (i.e., the objective function and the feasible set are both convex). The feasible set contains interior points (e.g., $\vec{x} = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$), so Slater's condition is satisfied and thus strong duality holds.

(c) Write the KKT conditions for this optimization problem. Do there exist Lagrange multipliers λ_1^* and λ_2^* that prove the optimality of $\vec{x}^*$?

Solution: The Lagrangian is given by

$$\mathcal{L}(x,\lambda) = x_1^2 + x_2^2 + \lambda_1[(x_1-1)^2 + (x_2-1)^2 - 4] + \lambda_2[(x_1-1)^2 + (x_2+1)^2 - 4]. \tag{46}$$

We can write the KKT conditions as follows:

i. Stationarity:

$$x_1^* + (\lambda_1^* + \lambda_2^*)(x_1^* - 1) = 0, \tag{47}$$
$$x_2^* + \lambda_1^*(x_2^* - 1) + \lambda_2^*(x_2^* + 1) = 0. \tag{48}$$

ii. Primal feasibility:

$$(x_1^* - 1)^2 + (x_2^* - 1)^2 - 4 \leq 0, \tag{49}$$
$$(x_1^* - 1)^2 + (x_2^* + 1)^2 - 4 \leq 0. \tag{50}$$

iii. Dual feasibility:

$$\lambda_1^* \geq 0,\ \lambda_2^* \geq 0. \tag{51}$$

iv. Complementary slackness:

$$\lambda_1^*[(x_1^* - 1)^2 + (x_2^* - 1)^2 - 4] = 0, \tag{52}$$
$$\lambda_2^*[(x_1^* - 1)^2 + (x_2^* + 1)^2 - 4] = 0. \tag{53}$$

From the stationarity conditions (along with dual feasibility), we can conclude that

$$x_1^* = 0,\ x_2^* = 0 \Rightarrow \lambda_1^* = \lambda_2^* = 0. \tag{54}$$

Since in part b we already show this is a convex problem with differentiable objective and constraint functions, and that Slater's condition holds, we know KKT conditions provide necessary and sufficient conditions for optimality. For these values $\vec{x}^*$ and $\vec{\lambda}^*$ that satisfy the KKT conditions, we can conclude that $\vec{x}^*$ is primal optimal (and, additionally, that $\vec{\lambda}^*$ is dual optimal).

5. Water Filling

Consider the following problem:

$$\begin{aligned} \text{minimize} \quad & -\sum_{i=1}^{n} \log(\alpha_i + x_i) && (55) \\ \text{subject to} \quad & \vec{x} \geq 0,\ \vec{1}^\top \vec{x} = c, && (56) \end{aligned}$$

where $\alpha_i > 0$ for each $i = 1, \ldots, n$.

This problem arises in information theory, in allocating power to a set of n communication channels. The variable x_i represents the transmitter power allocated to the ith channel, and $\log(\alpha_i + x_i)$ gives the capacity or communication rate of the channel, so the problem is to allocate a total power of c to the channels, in order to maximize the total communication rate.

(a) Verify that this is a convex optimization problem with differentiable objective and constraint functions. Find the domain $\mathcal{D}$ of the objective function $-\sum_{i=1}^{n} \log(\alpha_i + x_i)$ where it is well defined.

Solution: The objective function is convex with domain $\{\vec{x} : x_i + \alpha_i > 0 \forall i\}$, which is an open set in $\mathbb{R}^n$, and the objective function is differentiable on its domain. The n inequalities are determined by the function $-x_i$, $1 \leq i \leq n$, each of which is convex with domain $\mathbb{R}^n$ (note that these have to be thought of as function on $\mathbb{R}^n$) and is differentiable on its domain. There is one equality constraint, determined by the function $\vec{1}^\top \vec{x} = c$, which is convex with domain $\mathbb{R}^n$ and is differentiable on its domain.

The domain $\mathcal{D}$ of the optimization problem is the intersection of the domains of the objective and the constraint functions. Hence we have

$$\mathcal{D} = \{\vec{x} : x_i + \alpha_i > 0 \forall i\} \quad (57)$$

(b) Let $\lambda \in \mathbb{R}^n$ and $\nu \in \mathbb{R}$ be the dual variables corresponding to the constraints $x_i \geq 0, i = 1, \ldots, n$ and $\vec{1}^\top \vec{x} = c$, respectively. Write a Lagrangian for the optimization problem based on these dual variables.

Solution: We can write the Lagrangian as:

$$\mathcal{L}(\vec{x}, \lambda, \nu) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \sum_{i=1}^{n} \lambda_i x_i + \nu(\vec{1}^\top \vec{x} - c). \quad (58)$$

The domain of the Lagrangian is $\mathcal{D} \times \mathbb{R}^n \times \mathbb{R}$.

(c) Write the KKT conditions for the problem.

Solution: The KKT conditions are the following:

i. Stationarity:

$$-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0, \quad i = 1, \ldots, n \quad (59)$$

ii. Primal feasibility:

$$\vec{x}^* \geq 0, \quad i = 1, \ldots, n, \quad \vec{1}^\top \vec{x}^* = c \quad (60)$$

iii. Dual feasibility:

$$\lambda^* \geq 0, \quad i = 1, \ldots, n \quad (61)$$

iv. Complementary slackness:

$$\lambda_i^* x_i^* = 0, \quad i = 1, \ldots, n \tag{62}$$

(d) Since our problem is a convex optimization problem with differential objective and constraint functions, the KKT conditions provide sufficient conditions for optimality. Hence, we know that if we can find $\vec{x}^*$ and (λ^*, ν^*) that verify the KKT conditions, then $\vec{x}^*$ will be a primal optimal point, (λ^*, ν^*) will be dual optimal. We therefore attempt to find solutions for the KKT conditions. As a first step, show how to simplify the KKT conditions so that they are expressed in terms of only $\vec{x}^*$ and ν^*, i.e. we show how λ^* can be eliminated from these conditions.

Solution: The Lagrangian stationary conditions tell us how to write λ_i^* in terms of ν^* and x_i^*. However, the corresponding dual feasibility conditions, which require $\lambda_i^* \geq 0$, will manifest themselves as a new constraint on the pair (ν^*, x_i^*) when we do this. The corresponding complementary slackness condition will also now need to be expressed in terms of the pair (ν^*, x_i^*).
Eliminating λ^* in this way leads to the following simplified set of KKT conditions.

$$\vec{1}^\top \vec{x}^* = c \tag{63}$$

$$x_i^* (\nu^* - \frac{1}{\alpha_i + x_i^*}) = 0, \quad 1 \leq i \leq n, \tag{64}$$

$$x_i^* \geq 0, \quad 1 \leq i \leq n, \tag{65}$$

$$\nu^* \geq \frac{1}{\alpha_i + x_i^*}, \quad 1 \leq i \leq n. \tag{66}$$

(e) Solve for $x_i^*, 1 \leq i \leq n$, in terms of ν^* from the simplified KKT conditions derived in the preceding part of this question.

Solution: We can find the solution by considering two cases:

- If $\nu^* < 1/\alpha_i$, then we must have $x_i > 0$. as can be seen from the condition $\nu^* \geq 1/(\alpha_i + x_i^*)$. Then complementary slackness gives $\nu^* = 1/(\alpha_i + x_i^*)$, or equivalently, $x_i^* = -\alpha_i + 1/\nu^*$.
- If $\nu^* > 1/\alpha_i$, by complementary slackness, we must have $x_i^* = 0$. To see this, assume $x_i^* > 0$:

$$x_i^* > 0 \implies \nu^* = \frac{1}{\alpha_i + x_i^*} \implies \nu^* \leq \frac{1}{\alpha_i}, \tag{67}$$

which leads to a contradiction.

As a result, we can write the optimal solution $\vec{x}^*$ in terms of ν^* as:

$$x_i^* = \begin{cases} -\alpha_i + 1/\nu^* & \text{if } \nu^* < 1/\alpha_i, \\ 0 & \text{if } \nu^* \geq 1/\alpha_i, \end{cases} \tag{68}$$

$$= \max\{0, -\alpha_i + 1/\nu^*\} \tag{69}$$

Given $\hat{\nu} > 0$, define $\hat{x}_i := \max\{0, -\alpha_i + 1/\hat{\nu}\}$ for $1 \leq i \leq n$. It is straightforward to check that $\sum_{i=1}^n \hat{x}_i$ is continuous and strictly monotonically decreasing from ∞ to 0 for $0 < \hat{\nu} \leq 1/\alpha_{\text{min}}$, after which it equals 0 for $\hat{\nu} \geq 1/\alpha_{\text{min}}$. Here α_{min} denotes the minimum of the α_i, $1 \leq i \leq n$. This observation will be important in the proof of the uniqueness of ν^* in the next part of the question.

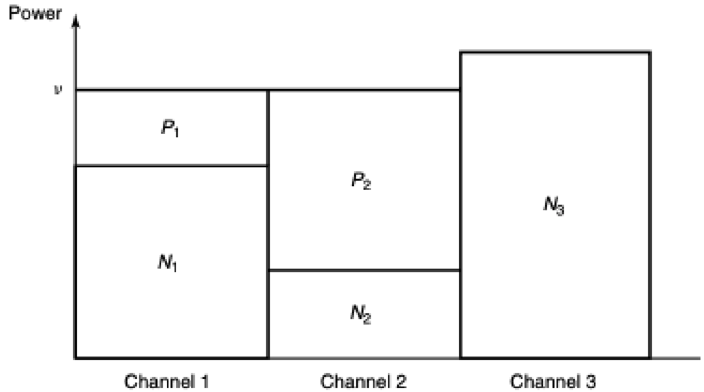

Figure 2: This graphic depicts a solution to the water-filling problem. On the x-axis we have n communication channels and on the y-axis we have the power in each channel. There is a base amount of noise N_i, which for us corresponds to α_i. Water-filling tells us that we should fill each channel until $\frac{1}{\nu^*}$, adding $\frac{1}{\nu^*} - \alpha_i$ power (in this graphic written as P_i), unless α_i already exceeds $\frac{1}{\nu^*}$. One algorithm for achieving this is to allot power to the channel with the least noise until it matches the channel with the second-least noise. Then we fill both simultaneously until they match the level of the channel with the third-least noise. Repeating this process until we run out of power to allot. This distribution of power is akin to filling connected basins with water, hence the name 'water filling'. Figure taken from *Elements of Information Theory* by Cover and Thomas.

6. **Homework Process**

   With whom did you work on this homework? List the names and SIDs of your group members.

   NOTE: If you didn't work with anyone, you can put "none" as your answer.

