CSE 203B Midterm

Part I: True or False: Explain your answer with one sentence (27 pts)

I.1 (convex set): Set $\{(x^2, y^2)|x + y \geq 4, x, y \in R_+\}$ is convex.
True: Note that the set $\mathcal{C} = \{(x^2, y^2)|x + y \geq 4, x, y \in R_+\}$ is equivalent to $\{(w, z)|\sqrt{w} + \sqrt{z} \geq 4, w, z \in R_+\}$. By concavity of $f(x) = \sqrt{x}$ for $x \geq 0$, any linear combination (w_3, z_3) of any $(w_1, z_1), (w_2, z_2) \in \mathcal{C}$ has $\sqrt{w_3} + \sqrt{z_3} \geq 4$.

I.2 (dual cone): Given cone $K = \{x|Ax \geq 0, x \in R^n\}$, where $A \in R^{m\times n}$ its dual cone is $K^* = \{y|A^T y \geq 0, y \in R^m\}$.
False: The dual cone is given by $K^* = \{A^T y : y \geq 0, y \in R^m\}$.

I.3 (Convex Function): Given a function $f(x) = \log(e^{a_1x_1} + e^{a_2x_2} + e^{a_3x_3})$ with domain $D = \{x|x \in R^3\}$, we can show that $f(x)$ is a convex function for every arbitrary vector $a \in R^3$.
True: For any $x = (x_1, x_2, x_3), y = (y_1, y_2, y_3) \in R^3$ Consider the linear combination $z = (z_1, z_2, z_3) = (x_1, x_2, x_3)/2 + (y_1, y_2, y_3)/2$.

$$\begin{aligned}
f(z) &= \log(e^{a_1z_1} + e^{a_2z_2} + e^{a_3z_3}) \\
&= \log(e^{a_1x_1/2}e^{a_1y_1/2} + e^{a_1x_2/2}e^{a_1y_2/2} + e^{a_3x_3/2}e^{a_1y_3/2}) \\
&\overset{(a)}{\leq} \log\Big((e^{a_1x_1} + e^{a_2x_2} + e^{a_3x_3})^{1/2}(e^{a_1y_1} + e^{a_2y_2} + e^{a_3y_3})^{1/2}\Big) \\
&= \frac{1}{2}\cdot\Big(\log(e^{a_1x_1} + e^{a_2x_2} + e^{a_3x_3}) + \log(e^{a_1y_1} + e^{a_2y_2} + e^{a_3y_3})\Big),
\end{aligned}$$

where (a) follows from the Holder inequality (select $p = q = 2$ in the inequality). The 1/2 linear combination suffices to show convexity since the domain is dense. Therefore f is convex.

I.4 (Conjugate Function): Given function $f(x) = x_1^2 - 4x_1x_2 + x_2^2$, where $x \in R^2$, then the conjugate of the conjugate function, $f^{**}(x)$, is equal to itself, i.e., $f^{**}(x) = f(x)$.
False: The relationship holds iff f is convex. But f is not convex it can be verified from its Hessian:

$$\begin{bmatrix} 2 & -4 \\ -4 & 2 \end{bmatrix}.$$

I.5 (Convex Function): Function $g(x) = min_y f(x, y)$ is convex function, if $f(x, y)$ is a convex function with respect to variable x.
False: Consider $f(x, y) = xy$ where $x \in R, y \in [-1, 1]$. For any fixed value of $y = c$ clearly $f(x, c)$ is convex. However $g(0) = 0 \geq -1 = (-1 - 1)/2 = (g(-1) + g(1))/2$ and therefore g is not convex.

I.6 (Convex Function): Given a differentiable but nonconvex function $f(x)$, where $x \in R^n$, and a fixed point $\bar{x} \in R^n$, the hyperplane

$$[\nabla f(\bar{x})^T, -1](\begin{bmatrix} x \\ t \end{bmatrix} - \begin{bmatrix} \bar{x} \\ f(\bar{x}) \end{bmatrix}) = 0$$

is a supporting hyperplane of epigraph, **epi** $f = \{\begin{bmatrix} x \\ t \end{bmatrix} |f(x) \leq t\}$.
False: A non-convex f may not have a supporting hyperplane at all points $\bar{x}$. A simple graphical

counter example suffices.

I.7 (Problem Formulation): For every convex optimization problem defined as eq. (4.1) in textbook, where all functions are convex, there is always an optimal solution.
False: Since the region can be unfeasible, e.g. consider the linear programming convex formulation that may not have an optimal solution when the feasible region is empty.

I.8 (Problem Formulation/Duality): Given a convex programming problem:
minimize $f_0(x)$, subject to $Ax \leq b$, $x \in R^n$, $A \in R^{m\times n}$, $b \in R^m$,
where $f_0(x)$ is a differentiable convex function, we can claim that
$\nabla f_0(\bar{x}) \in \{-A^T\theta | \theta \in R^m_+\}$
is a necessary condition for $\bar{x}$ to be an optimal solution.
True: Since $Ax \leq b$ we obtain the cone $K = \{x| -Ax \geq 0\}$ and the gradient at the optimal solution, $\nabla f_0(\bar{x})$, should fall within its dual cone $K^* = \{-A^T\theta|\theta \in R^m_+\}$.

I.9 (Duality): Given a function $f(x, y)$, the inequality
$\min_x \max_y - f(x, y) \geq \max_y \min_x - f(x, y)$
is always true.
True: Follows from applying the min-max theorem to $g(x, y) = -f(x, y)$.

Part II: Problem 1
[Solution]
Properties that we will use in the proof:

1. If $p > 0$, then $\|tx\|_p = |t|\|x\|_p$, for any $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$.
2. If $0 < p < q$, then for any $x \in \mathbb{R}^n$, $\|x\|_p < \|x\|_q$.
3. When $p \geq 1$, $\|y\|_q = \max_{\|x\|_p \leq 1} y^T x$ is the dual norm of $\|x\|_p$, where $\frac{1}{p} + \frac{1}{q} = 1$. Specifically, when $p = 1$, $q = \infty$.

According to the definition of dual cone, we have

$$K^* = \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \middle| \quad y^T x + st \geq 0, \forall \begin{bmatrix} x \\ t \end{bmatrix} \in K \right\}$$

According to property 1, for $t > 0$, we have

$$\begin{aligned} \|x\|_p &\leq t \\ \frac{1}{t}\|x\|_p &\leq 1 \\ |\frac{1}{t}|\|x\|_p &\leq 1 \\ \|\frac{1}{t}x\|_p &\leq 1 \end{aligned}$$

Thus

$$\begin{aligned}K^* &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; y^T x + st \geq 0, \forall \begin{bmatrix} \frac{x}{t} \\ 1 \end{bmatrix} \in K \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; y^T x + s \geq 0, \forall \begin{bmatrix} x \\ 1 \end{bmatrix} \in K \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; -y^T x + s \geq 0, \forall \begin{bmatrix} -x \\ 1 \end{bmatrix} \in K \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; y^T x \leq s, \forall \begin{bmatrix} -x \\ 1 \end{bmatrix} \in K \right\}\end{aligned}$$

According to property 1, $\|x\|_p = \|-x\|_p$, therefore

$$\begin{aligned}K^* &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; y^T x \leq s, \forall \begin{bmatrix} x \\ 1 \end{bmatrix} \in K \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; \max_{\|x\|_p \leq 1} y^T x \leq s \right\}\end{aligned}$$

1. When $p \geq 1$, according to property 3, we have

$$K^* = \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; \|y\|_q \leq s \right\}$$

   where $\frac{1}{p} + \frac{1}{q} = 1$.

2. When $0 < p < 1$, according to property 2, for any $x \in \mathbb{R}^n$, $\|x\|_p < \|x\|_1$ and $C = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \,\middle|\, \; \|x\|_1 \leq t \right\}$ is the conic hull of K. This yields a tight upper bound of $\max_{\|x\|_p \leq 1} y^T x$:

$$\max_{\|x\|_p \leq 1} y^T x < \max_{\|x\|_1 \leq 1} y^T x$$

$$K^* = \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; \max_{\|x\|_1 \leq 1} y^T x \leq s \right\}$$

   According to property 3, $\|y\|_\infty = \max_{\|x\|_1 \leq 1} y^T x$, therefore,

$$K^* = \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \,\middle|\, \; \|y\|_\infty \leq s \right\}$$

Part II: Problem 2
[Solution]

1. $$f^*(y) = \sup_{x \in \mathbb{R}^2_+} \left(y^T x - x_1 x_2 \right) = \sup_{x \in \mathbb{R}^2_+} \left(y_1 x_1 + y_2 x_2 - x_1 x_2 \right)$$

- **Case 1** ($\exists k,\ y_k > 0$): Set $x_k = t > 0$, and $x_{i\neq k} = 0$. Then:

$$\lim_{t\to\infty} (y_k x_k + y_i x_i - x_k x_i) = \lim_{t\to\infty} y_k t \to \infty$$

  Therefore, ($\exists k,\ y_k > 0$) is not in $\textbf{dom} f^*$.

- **Case 2** ($y \preceq 0$): There exists no way to make any term of the inside of the supremum positive within the domain. Therefore, the supremum occurs at $x = 0$ which implies $f^*(y) = 0$ in this region.

To summarize:

$$f^*(y) = \begin{cases} 0 & y \preceq 0 \\ \infty & \text{otherwise} \end{cases}$$

2.

$$f^*(y) = \max\left\{ \sup_{||x||_p \le a} \left(y^T x - ||x||_p^p\right),\ \sup_{||x||_p > a} \left(y^T x - a^{1-\frac{1}{p}} ||x||_p\right) \right\}$$

Let $g(y,x) = y^T x - ||x||_p^p$ and $h(y,x) = y^T x - a^{1-\frac{1}{p}}||x||_p$.

- **Case 1** ($p = 1$): In this case, the partial function is continuous, and $g(y,x) = h(y,x) = y^T x - ||x||_1$.
  - **Case 1.1** ($\exists k,\ |y_k| > 1$): $y_k > 1$, for any k. Set $x_k = t > 0$ and $x_{i\neq k} = 0$. Then:

$$\lim_{t\to\infty} \left(y^T x - ||x||_1\right) = \lim_{t\to\infty} (y_k t - t) = \lim_{t\to\infty} (y_k - 1)t \to \infty$$

    Similarly, in the case where $y_k < -1$, for any k, we can set $x_k = -t < 0$ and $x_{i\neq k} = 0$. Then:

$$\lim_{t\to\infty} \left(y^T x - ||x||_1\right) = \lim_{t\to\infty} (-y_k t - t) = \lim_{t\to\infty} (-y_k - 1)t \to \infty$$

  - **Case 1.2** ($||y||_\infty \le 1$): In this case, recall Hölder's inequality: $y^T x \le ||y||_q ||x||_p$, where p-norm and q-norm are duals. Recall that the dual of 1-norm is the ∞-norm. Since we don't have constraints on what values $||x||_1$ can take for $p = 1$, we know that equality is trivially achievable for this inequality. Then:

$$\begin{aligned} y^T x - ||x||_1 &\le ||y||_\infty ||x||_1 - ||x||_1 \\ y^T x - ||x||_1 &\le (||y||_\infty - 1)\,||x||_1 \\ \sup_x \left(y^T x - ||x||_1\right) &= 0, \text{ for } x = 0 \text{ since } ||y||_\infty \le 1 \end{aligned}$$

- **Case 2** ($p > 1$): Let $\frac{1}{p} + \frac{1}{q} = 1 \implies q = \frac{p}{p-1}$.
  - **Investigation of** $\sup_{||x||_p \le a} g(y,x)$: This is a bounded case since we don't have any

norm division terms in g. We can take the derivative to see where the optimum is:

$$\begin{aligned}
\nabla_{x_i} g(y,x) = y_i - p|\hat{x}_i|^{p-1} \cdot \textbf{sign}(\hat{x}_i) &= 0 \\
|y_i| &= p|\hat{x}_i|^{p-1} \\
|y_i|^q &= p^q |\hat{x}_i|^p \\
\sum_i |y_i|^q &= p^q \sum_i |\hat{x}_i|^p \\
||y||_q^q &= p^q ||\hat{x}||_p^p \\
\frac{||y||_q^q}{p^q} &= ||\hat{x}||_p^p \\
p^{\frac{-1}{p-1}} ||y||_q^{\frac{1}{p-1}} &= ||\hat{x}||_p, \text{ with } ||\hat{x}||_p \leq a \\
g(\hat{x}, y) = y^T \hat{x} - ||\hat{x}||_p^p = ||y||_q ||\hat{x}||_p - ||\hat{x}||_p^p \\
g(\hat{x}, y) &= \left(p^{\frac{-1}{p-1}} - p^{\frac{-p}{p-1}}\right) ||y||_q^q, \text{ for } ||y||_q \leq a^{p-1} p
\end{aligned}$$

This optimum is nonnegative, as $p^{\frac{-1}{p-1}} > p^{\frac{-p}{p-1}}$ for $p > 1$. It is also a maximum since $\nabla_{xx} g(y,x) = \nabla^2 \left(-||x||_p^p\right) = \nabla^2 \left(-\sum_i |x_i|^p\right) \preceq 0$ (diagonal matrix with nonpositive diagonal entries.)

For $||y||_q > a^{p-1}p$, we always have $\nabla_x g(y,x) > 0$, so the best possible value that $\sup_{||x||_p \leq a} g(y,x)$ can provide is at the boundary, $||x||_p = a$. This value is $a||y||_q - a^p$, with x tuned to satisfy equality in Hölder's inequality ($\forall i,\ |y_i|^q = k|x_i|^p$ for some constant k.) Therefore:

$$\sup_{||x||_p \leq a} g(y,x) = \begin{cases} \left(p^{\frac{-1}{p-1}} - p^{\frac{-p}{p-1}}\right) ||y||_q^q & ||y||_q \leq a^{p-1}p \\ a||y||_q - a^p & ||y||_q > a^{p-1}p \end{cases}$$

– **Investigation of** $\sup_{||x||_p > a} h(y,x)$: For $||y||_q > a^{1-\frac{1}{p}}$, let $||y||_q = y^T z$ for some vector z with $||z||_p = 1$, as per the definition of the dual norm. Then:

$$\begin{aligned}
||y||_q &> a^{1-\frac{1}{p}} \\
y^T z &> a^{1-\frac{1}{p}} ||z||_p \\
y^T z - a^{1-\frac{1}{p}} ||z||_p &> 0
\end{aligned}$$

Set $x = tz$, $t > a$. Evaluate the limit of $h(y,x)$ as $t \to \infty$:

$$\lim_{t \to \infty} \left(y^T x - a^{1-\frac{1}{p}} ||x||_p\right) = \lim_{t \to \infty} \left(y^T z - a^{1-\frac{1}{p}} ||z||_p\right) t \to \infty$$

For $||y||_q \leq a^{1-\frac{1}{p}}$, use Hölder's inequality, $y^T x \leq ||y||_q ||x||_p$, with equality achievable if $\forall i,\ |y_i|^q = k|x_i|^p$ for some constant k (the fact that we're restricted to the $||x||_p > a$

case does not make this impossible for any y.) Then:

$$y^T x - a^{1-\frac{1}{p}}||x||_p \leq ||y||_q||x||_p - a^{1-\frac{1}{p}}||x||_p$$
$$y^T x - a^{1-\frac{1}{p}}||x||_p \leq \left(||y||_q - a^{1-\frac{1}{p}}\right)||x||_p$$
$$\sup_{||x||_p > a}\left(y^T x - a^{1-\frac{1}{p}}||x||_p\right) = a||y||_q - a^{2-\frac{1}{p}},\ \text{for } ||x||_p \to a \text{ since } ||y||_q \leq a^{1-\frac{1}{p}}$$

Therefore:

$$\sup_{||x||_p > a} h(y,x) = \begin{cases} a||y||_q - a^{2-\frac{1}{p}} & ||y||_q \leq a^{1-\frac{1}{p}} \\ \infty & ||y||_q > a^{1-\frac{1}{p}} \end{cases}$$

Now we have to consider the maximum between $\sup_{||x||_p \leq a} g(y,x)$ and $\sup_{||x||_p > a} h(y,x)$. Obviously, this is going to depend on regions of y, but also the relationship between p and a. Based on all of the above, two cases exist:

- **Case 2.1** ($a^{1-\frac{1}{p}} \leq a^{p-1}p$): In this case, since ∞ is larger than any finite value, $||y||_q > a^{1-\frac{1}{p}}$ gives an infinite supremum. Otherwise, $||y||_q \leq a^{1-\frac{1}{p}}$ stays within the bounds of $||y||_q \leq a^{p-1}p$, so the supremum in this region becomes the supremum of $||x||_p < a$, since:

$$\begin{aligned} ||y||_q \leq a^{1-\frac{1}{p}} &\implies ||y||_q \leq a^{1-\frac{1}{p}} + \left(p^{\frac{-1}{p-1}} - p^{\frac{-p}{p-1}}\right)\frac{||y||_q^q}{a} \\ &\implies a||y||_q \leq a^{2-\frac{1}{p}} + \left(p^{\frac{-1}{p-1}} - p^{\frac{-p}{p-1}}\right)||y||_q^q \\ &\implies a||y||_q - a^{2-\frac{1}{p}} \leq \left(p^{\frac{-1}{p-1}} - p^{\frac{-p}{p-1}}\right)||y||_q^q \end{aligned}$$

- **Case 2.2** ($a^{1-\frac{1}{p}} > a^{p-1}p$): We have one extra region to check compared to **Case 2.1**, $a^{p-1}p < ||y||_q \leq a^{1-\frac{1}{p}}$. This is still the finite region of the $||x||_p > a$ supremum, so we don't have to worry about infinity. Notice that:

$$a^{1-\frac{1}{p}} > a^{p-1}p \implies a^{2-\frac{1}{p}} > a^p p > a^p \implies a||y||_q - a^p > a||y||_q - a^{2-\frac{1}{p}}$$

The supremum of $||x||_p = a$ dominates the finite value. We can also simplify $a^{1-\frac{1}{p}} > a^{p-1}p$:

$$a^{1-\frac{1}{p}} > a^{p-1}p$$
$$a^{2-\frac{1}{p}-p} > p$$
$$a^{\frac{-(p^2-2p+1)}{p}} > p$$
$$a < p^{\frac{-p}{(p-1)^2}}$$

In summary (with $q = \frac{p}{p-1}$):

$$f^*(y) = \begin{cases} 0 & p = 1 \wedge ||y||_\infty \leq 1 \\ \left(p^{\frac{-1}{p-1}} - p^{\frac{-p}{p-1}}\right)||y||_q^q & p > 1 \wedge ||y||_q \leq \min\left\{a^{1-\frac{1}{p}},\ a^{p-1}p\right\} \\ a||y||_q - a^p & p > 1 \wedge a < p^{\frac{-p}{(p-1)^2}} \wedge a^{p-1}p < ||y||_q \leq a^{1-\frac{1}{p}} \\ \infty & \text{otherwise} \end{cases}$$

Part II: Problem 3
[Solution]

1. For d free components, we denote the 1s vector in $\mathbb{R}^d$ with $\mathbf{1}$. Write the Lagrangian:

$$L(x, y, \nu_1, \nu_2) = x^\top L' x + y^\top L' y + b^\top x + d^\top y + \nu_1 \mathbf{1}^\top x + \nu_2 \mathbf{1}^\top y$$

   Let $g(\nu_1, \nu_2) = \inf_{x,y} L(x, y, \nu_1, \nu_2)$. Note that $L(x, y, \nu_1, \nu_2)$ is convex in x and y. The infimum can be recovered by solving for the first-order condition. The gradient of g with respect to x is $\nabla_x L(x, y, \nu_1 \nu_2) = 2L'x + b + \nu_1 \mathbf{1}$ (likewise for y). Equating to zero and plugging back in yields the dual problem:

$$\max_{\nu_1, \nu_2 \in \mathbb{R}} -\frac{1}{4}\left[(x + \nu_1 \mathbf{1})^\top L'^{-1}(x + \nu_1 \mathbf{1}) + (y + \nu_2 \mathbf{1})^\top L'^{-1}(y + \nu_2 \mathbf{1})\right]$$

   The corresponding closed form solutions for ν_1, ν_2 is then given by

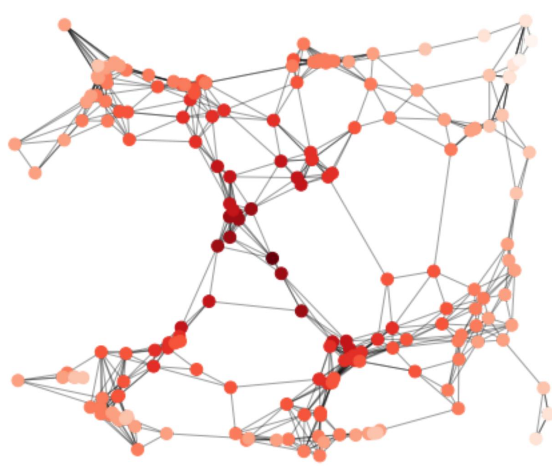
$$\nu_1 = -\frac{\mathbf{1}^\top L^{-1} b}{\mathbf{1}^\top L^{-1} \mathbf{1}} \quad \nu_2 = -\frac{\mathbf{1}^\top L^{-1} d}{\mathbf{1}^\top L^{-1} \mathbf{1}}$$

   Note that assuming G is connected, and that there are more than one fixed node, L' is positive definite. The principle minors of L are positive, as are the eigenvalues of L'. Alternatively, can also show that L' is *weakly chained diagonally dominant* and therefore nonsingular. I mainly looked for a correct Lagragian and dual function. The dual problem should not contain primal variables.

2. Correct implementations of either the closed form or primal / dual problems were accepted. Note that the constraints of the primal should be something like:

```
cp.sum(x) == 0
```

   The objective is the same as the one used in homework 4. The minimum of the primal / dual objectives is around 49.19. The solution should look reasonably close to:



3. Can demonstrate optimality by deriving (1.) closed form solution (2.) citing duality—e.g. showing solution to the dual problem yields the same numerical result up to numerical error (3.) showing solution satisfies kkt conditions up to numerical error.