CSE 203B W23 Midterm

Part I: Problem 1

Set $\{(x, y)|x^2 + y^2 \geq 1, x, y \in R\}$ is convex.

[Solution] False. Complement of the interior of the unit circle in R-2

Part I: Problem 2

Given cone $K = \{x|a_1^T x \geq 0, a_2^T x \leq 0$, where $a_1, a_2, x \in R^n\}$, its dual cone is $K^* = \{a_1\theta_1 + a_2\theta_2|\theta_1 \geq 0, \theta_2 \leq 0$, and $\theta_1, \theta_2 \in R\}$.

[Solution] True. The dual cone of the set $\{x|Ax \geq 0\}$ is $\{A^\top x|X \geq 0\}$ with $A = [a_1, -a_2]^\top$.

Part I: Problem 3

The **dual of the dual** cone $K = \{x|\ ||Ax + b||_2 < c^T x + d\}$ is itself, where $A \in R^{mn}, x, b, c \in R^n$ and $d \in R$.

[Solution] False. K is not closed.

Part I: Problem 4

Given a function $f(x) = 1.4x^{1.3} + 2.2x^{2.5}$, where $x \in R_+$. The function is convex.

[Solution] True. By composition rules of second order characterization of convex functions.

Part I: Problem 5

Function $g(x) = max_y f(x, y)$ is a convex function for any $f(x, y) \in R$, where $x, y \in R^n$.

[Solution] False. Consider the counter example $f(x, y) = x^3$.

Part I: Problem 6

Given function $f(x) = x^T Ax + b^T x + c$, where $x, b \in R^n, A \in R^{nn}$, and $c \in R$. Suppose that matrix A is not positive semidefinite. We can claim that the conjugate function, $f^*(y) = \infty$, for all $y \in R^n$.

[Solution] True. A not psd implies $\exists v$ s.t. the quadratic $v^\top Av$ is unbounded in the direction of v.

Part I: Problem 7

Given a differentiable and convex function $f(x)$, where $x \in R^n$, and a fixed point $\bar{x} \in R^n$. Suppose that in a $n + 1$ dimension space $[x^T, t]^T$, we draw the supporting hyperplane

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

We can claim that the supporting hyperplane intersects the t axis at its conjugate function i.e. $[x^T = \underline{0}^T, t = -f^*(y)]^T$ where $y = \nabla f(\bar{x})$, and $\underline{0}$ is a vector of 0.

[Solution] True. The conjugate is maximized for x^* s.t. $y = \nabla f(x^*)$. $f^*(y) = \nabla f(x^*)^\top x^* - f(x^*)$.

Part I: Problem 8

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 and sufficient condition for $\bar{x}$ to be an optimal solution.

[Solution] False. The condition is necessry but not sufficient. (Check KKT conditions.)

Part I: Problem 9

In the textbook subsection (5.1.1), we have the problem formulation (5.1) and its Lagrangian $L(x, \lambda, \nu)$. The variable x of the Lagrangian has to be feasible, i.e. x satisfies the constraint in formulation (5.1).

[Solution] False. The Lagrangian dual function is defined over the intersection of the domains of the constraints and objective.

Part I: Problem 10

Given a function $f(x, y) = x^T Ay$, we have $x, y \in R^n$ and matrix $A \in R^{nn}$. We can claim that the equality,
$min_x max_y f(x, y) = max_y min_x f(x, y)$ is true when there is a bounded (not infinite) solution.

[Solution] True. Check Theory of Games (by John Von Neumann and Oskar Morgenstern)

Part II: Problem 1

Problem 1. Dual Cone: Find the dual cone of the following cones. (20 pts)

1.1 $K = \{\begin{bmatrix} x \\ t \end{bmatrix} \mid ||Ax||_2 \leq t\}$, where $A \in R^{nn}, x \in R^n$, and $t \in R_+$. Matrix A is nonsingular. (hint: If you have no clue, start with a small but nontrivial case, e.g. n=1 and/or 2.

1.2 $K = \{\begin{bmatrix} x \\ t \end{bmatrix} \mid ||Ax||_p \leq t\}$, where $A \in R^{nn}, x \in R^n$, $p \geq 1$, and $t \in R_+$. Matrix A is nonsingular.

[Solution]
Properties that we will use in the proof:

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

Let $Ax = b$, since A is nonsingular, we have $x = A^{-1}b$, and K can be written as

$$K = \left\{ \begin{bmatrix} A^{-1}b \\ t \end{bmatrix} \middle| \quad \|b\|_p \leq t \right\}$$

Let

$$C = \left\{ \begin{bmatrix} b \\ t \end{bmatrix} \middle| \quad \|b\|_p \leq t \right\}$$

K can be rewritten as

$$K = \left\{ \begin{bmatrix} A^{-1}b \\ t \end{bmatrix} \Big| \quad \begin{bmatrix} b \\ t \end{bmatrix} \in C \right\}$$

According to the definition of dual cone, we have

$$\begin{aligned} K^* &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad y^T A^{-1} b + st \geq 0, \forall \begin{bmatrix} A^{-1}b \\ t \end{bmatrix} \in K \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad y^T A^{-1} b + st \geq 0, \forall \begin{bmatrix} b \\ t \end{bmatrix} \in C \right\} \end{aligned}$$

According to property 1, for $t \in \mathbb{R}_+$, we have

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

Thus

$$\begin{aligned} K^* &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad y^T A^{-1} b + st \geq 0, \forall \begin{bmatrix} \frac{b}{t} \\ 1 \end{bmatrix} \in C \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad y^T A^{-1} \frac{b}{t} + s \geq 0, \forall \begin{bmatrix} \frac{b}{t} \\ 1 \end{bmatrix} \in C \right\} \end{aligned}$$

Let $z = -\frac{b}{t}$, we have

$$K^* = \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad -y^T A^{-1} z + s \geq 0, \forall \begin{bmatrix} -z \\ 1 \end{bmatrix} \in C \right\}$$

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

$$\begin{aligned} K^* &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad -y^T A^{-1} z + s \geq 0, \forall \begin{bmatrix} z \\ 1 \end{bmatrix} \in C \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad y^T A^{-1} z \leq s, \forall \begin{bmatrix} z \\ 1 \end{bmatrix} \in C \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad \max_{\|z\|_p \leq 1} y^T A^{-1} z \leq s \right\} \\ &= \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad \max_{\|z\|_p \leq 1} ((A^{-1})^T y)^T z \leq s \right\} \end{aligned}$$

When $p \geq 1$, according to property 2, we have

$$K^* = \left\{ \begin{bmatrix} y \\ s \end{bmatrix} \Big| \quad \|(A^{-1})^T y\|_q \leq s \right\}$$

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

Note that K^* can be equivalently written as

$$K^* = \left\{ \begin{bmatrix} A^T y \\ s \end{bmatrix} \middle| \quad \|y\|_q \leq s \right\}$$

Part II: Problem 2

Problem 2. Conjugate Function: Find the conjugate function of the following functions. (20 pts)
2.1 $f(x) = -x^3 + 3x + 2$, where $x \in R$.

2.2 $f(x) = \frac{x_1^2}{x_2}$, where $x \in R^2_{++}$.

[Solution]

2.1

$$f^*(y) = \sup_x yx - f(x) = \sup_x yx + x^3 - 3x - 2$$

Let $g(x, y) = yx + x^3 - 3x - 2$,

$$\nabla_x g(x, y) = y + 3x^2 - 3$$

By setting $\nabla_x g(x, y)$ equal to 0, we have:

$$\hat{x} = \pm\sqrt{1 - \frac{y}{3}}$$

Note that $\nabla_{x^2} g(x, y) = 6x > 0$ when $x > 0$, which indicates $f^*(y)$ reaches only local maximum at $\hat{x}$, and the global maximum is reached when $x \to \infty$. Therefore, $f^*(y) = +\infty$.

2.2

(credited to Franklin Hunter Akins)

The conjugate function is

$$f^*(y) = \begin{cases} 0 \text{ if } y \in \{(y_1, y_2) y_1 \leq 0, y_2 \leq \text{ or } y_2 < 0, y_1^2/4 < -y_2\} \\ \infty \text{ otherwise} \end{cases}$$

and is shown schematically in Figure 1.

To derive it we consider a few cases.

Case 1: $y_1 < 0$ **and** $y_2 < 0$

If $y_1 < 0$ and $y_2 < 0$, then $g < 0$ for all $x \in R^2_{++}$. Taking $x_1 = x_2 = \alpha$, we have $g(\alpha) = \alpha(y_1 + y_2) - \alpha x_1^2$. Then taking the limit $\lim_{\alpha \to 0} g(\alpha) = 0$ shows that $\sup_x g = 0$ when y_1 and $y_2 < 0$.

Case 2: $y_2 < 0$ **and** $y_1^2/4 = -y_2$.

The Hessian of g is

$$H = \begin{bmatrix} -2/x_2 & 2x_1/x_2^2 \\ 2x_1/x_2^2 & -2x_1^2/x_2^3 \end{bmatrix} .$$

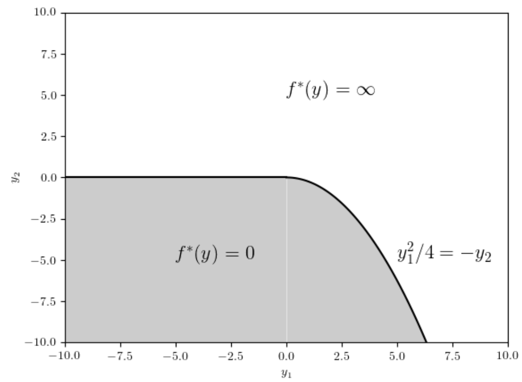

Figure 1: Schematic description of conjugate function.

The determinant of the Hessian is 0. The principal minors $-2/x_2$ and $-2x_1^2/x_2^3$ are both negative for $(x_1, x_2) \in R_{++}^2$. The determinant is 0. Therefore $H \preceq 0$ and any x with $g = 0$ is a maximum.

The gradient of g is

$$ {}_x g = y + \begin{bmatrix} -2x_1/x_2 \\ x_1^2/x_2^2 \end{bmatrix} , $$

which only depends on the ratio x_1/x_2. Setting this to zero leads to the system of equations $2x_1/x_2 = y_1$, $-x_1^2/x_2^2 = y_2$, which has a solution if and only if $y_1^2/4 = -y_2$. Setting $x_1 = y_1 x_2/2$, $g(x_1, x_2) = 0$, and so the conjugate function is $f^*(y) = 0$.

Case 3: $y_2 > 0$
To analyze the problem in this case, we look at the behavior of $g(x, y)$ along a curve $x_2 = f(x_1)$. Along the curve, the function g takes on the value

$$ g(x, y) = x_1 y_1 + f(x_1) y_2 - \frac{x_1^2}{f(x_1)} . $$

We can see that if $y_2 > 0$,we can take $x_2 = f(x_1) = x_1^n$ with $n \geq 2$ and the function will become asymptotically $y_2 x_1^n$ which is unbounded, regardless of y_1.

Case 4: $y_2 < 0$**,** $y_1^2/4 < -y_2$
If $y_2 < 0$, then both the second and third terms are negative. If $x_2 = f(x_1) = x_1^n$, then if $n > 1$, the expression is asymptotically dominated by y_2 for large x_1 and asymptotically dominated by $x_1 y_1$ for small x_1. If $0 < n < 1$, then the expression is asymptotically dominated by $-x_1^{2-n}$ and the function g goes to $-\infty$ as $\alpha \to \infty$.

To balance them, we can pick $f(x_1) = \alpha x_1$ for $\alpha \in R_{++}$. Then the function is linear in x_1: $g(x_1) = x_1(y_1 + \alpha y_2 - 1/\alpha)$. For g to be positive (for any x_1), we need $y_1 + \alpha y_2 - 1/\alpha > 0$, or

$$ y_2 \alpha^2 + y_1 \alpha - 1 > 0. $$

Since $y_2 < 0$ this is a downward facing parabola. The set is feasible for α if there is a real root that is not a double, or

$$ y_1^2 > -4y_2 $$

since

$$\alpha_* = \frac{-y_1 - \sqrt{y_1^2 + 4y_2}}{2y_2} > 0$$

since $y_2 < 0$ and $y_1 > 0$. Picking an $\alpha = \alpha_* - \epsilon$ will then satisfy $y_1 + \alpha y_2 - 1/\alpha > 0$. Therefore the conjugate function $f^*(y) = \infty$ when $y_2 < 0$ and $y_1^2 > -4y_2$ (since I can then take $x_1 \to \infty$).

Case 4: $y_2 < 0$, $y_1^2/4 < -y_2$

If $y_2 < 0$ and $y_1^2 < -4y_2$, then there is no value of α that makes g positive for any x_1. Therefore the function is bounded above by 0. Taking the limit $x_1, x_2 \to 0$ achieves that bound, so the function is maximized by taking the limit $x_1, x_2 \to 0, 0$ and $f^*(y) = 0$ in this case.

Part III: Problem 3

Recall that given a connected, undirected graph $G = (V, E)$ with n_0 vertices, the graph embedding problem is to assign coordinates in $\mathbb{R}^m$ to each vertex $v \in V$. Let $A \in \{0,1\}^{n_0 \times n_0}$ be the symmetric adjacency of G, and let D be the corresponding diagonal degree matrix such that $D_{ii} = \sum_j A_{i,j}$. The *graph Laplacian* is defined to be $L_0 = D - A$.

One way to define the graph embedding problem (i.e. Laplacian Eigenmaps) is to solve the following problem:

$$\begin{aligned} &\min_{X \in \mathbb{R}^{n_0 \times m}} \langle X, L_0 X \rangle \\ &\text{s.t. } X^\top X = I, \ \mathbf{1}^\top X = 0 \end{aligned} \tag{1}$$

Where the inner product $\langle U, V \rangle$ is defined to be $\text{tr}(U^\top V)$.

(i) Problem (1) is nonconvex. Prove that the solution is given by m eigenvectors corresponding to the smallest nonzero m eigenvalues of L_0.

(ii) Derive the dual and KKT conditions of the semi-supervised problem (you may assume L is nonsingular).

$$\begin{aligned} &\min_{X \in \mathbb{R}^{n \times m}} \langle X, LX \rangle + \langle B, X \rangle \\ &\text{s.t. } \mathbf{1}^\top X = 0 \end{aligned} \tag{2}$$

(iii) Derive the dual and KKT conditions of the problem

$$\begin{aligned} &\min_{X \in \mathbb{R}^{n \times m}} \langle X, LX \rangle + \langle B, X \rangle \\ &\text{s.t. } \mathbf{1}^\top X = 0, \quad ||X||_1 \le c, \end{aligned} \tag{3}$$

where $||X||_1 = \sum_{ij} |X_{ij}|$ and c is some constant, say 1000. (Hint: if you get stuck, try to write the l1 ball in standard form as an lp). Use the framework[1] from homework 4 to implement this problem (where L and B are now given). Note that the random seed has been updated from homework 4 (different graph and fixed vertices).

[Solution]

3.1:

We graded this question leniently, looking for keywords including rayleigh quotient, conjugate dual,

[1] https://colab.research.google.com/drive/1suBO3RgKaqzJBh-tzpGLIlXn625hmzH2?usp=sharing

variational theorem. This is an instantiation of the min-max / variational theorem in the context of Laplacian matrices. Here's an informal / high-level argument. For simplicity, consider the case where $m = 1$. Note that L is Symmetric and PSD. It's eigenvalues $\lambda_i \geq 0$. Additionally, the rows and columns of L sum to zero. The unit vector with constant entries, $\bar{1}$, lies in the null-space of L—i.e. is an eigenvector associated with eigenvalue 0.

First, suppose $(\bar{1}, v_1, \ldots, v_n)$ is an orthogonal basis for $\mathbf{R}^n$ corresponding to eigenvalues $0 \leq \lambda_1 \leq, \ldots, \leq \lambda_n$. Then any $x \in \mathbb{R}^n$ s.t. $x \neq 0$, $x \perp \bar{1}$ can be expressed as a linear combination of $v_1, \ldots, v_n$; $a_1 v_1 + \ldots + a_n v_n$. Substituting the rayleigh quotient, we get

$$\frac{x^\top L x}{x^\top x} = \frac{(\sum_{i=1}^n a_i v_i)^\top L (\sum_{i=1}^n a_i v_i)}{(\sum_{i=1}^n a_i v_i)^\top (\sum_{i=1}^n a_i v_i)} = \frac{\sum_{i=1}^n \lambda_i a_i ||v_i||^2}{\sum_{i=1}^n a_i ||v_i||^2} \geq \lambda_1.$$

with minimizer $a = \mathbf{1}_{i=1}$

3.2: For this question and 3.3, I looked for a correct Lagrangian, and a statement about the first-order condition (and the other kkt conditions). Points were deducted for incorrect notation I could not follow the notation (e.g. confusing matrix multiplication)

$$\begin{aligned} &\min_{X \in \mathbb{R}^{n \times m}} \langle X, LX \rangle + \langle B, X \rangle \\ &\text{s.t. } \mathbf{1}^\top X = 0 \end{aligned} \tag{4}$$

Define the Lagrangian

$$\mathcal{L}(X, \nu) = \langle X, LX \rangle + \langle B, X \rangle + (\mathbf{1}_{n \times m}^\top X) \cdot \nu \tag{5}$$

The Lagrange dual function is then

$$g(\nu, \lambda_i) = \inf_X \mathcal{L}(X, \nu)$$

Solving for the first order conditions $\nabla_X L = 0$, where the gradients are given by

$$\nabla_X \mathcal{L}(X, \nu) = 2LX + B + \mathbf{1}_{n \times m} \nu^\top$$

So $Y := X^* = -L^{-1}(B + 1_{n \times m} \nu^\top)$ and $g(\nu) = \mathcal{L}(Y, \nu)$, so the dual problem is

$$\max_\nu g(\nu)$$

KKT Conditions:

- Primal feasibility: $\mathbf{1}^\top X = 0$
- Stationarity: $\nabla_X \mathcal{L}(X, \lambda_i, \nu) = 0$

3.3:

$$\begin{aligned} &\min_{X \in \mathbb{R}^{n \times m}} \langle X, LX \rangle + \langle B, X \rangle \\ &\text{s.t. } \mathbf{1}^\top X = 0, \quad ||X||_1 \leq c, \end{aligned} \tag{6}$$

Consider the equivalent problem

$$\begin{aligned} &\min_{X,Z \in \mathbb{R}^{n\times m}} \langle X, LX\rangle + \langle B, X\rangle \\ &\text{s.t. } \mathbf{1}^\top X = 0, \quad X_{ij} \le Z_{ij}, \quad -X_{ij} \le Z_{ij}, \quad \sum_{ij} Z_{ij} \le c \end{aligned} \tag{7}$$

Define the Lagrangian

$$\begin{aligned} \mathcal{L}(X, Z, \nu, \lambda_i) = &\underbrace{\langle X, LX\rangle + \langle B, X\rangle + (\mathbf{1}_{n\times m}^\top X)\cdot \nu}_{A} + \\ &\underbrace{\sum_{ij} \lambda_{1,ij}(X_{ij} - Z_{ij}) + \sum_{ij} \lambda_{2,ij}(-X_{ij} - Z_{ij}) + \lambda_3(\sum_{ij} Z_{ij} - c)}_{B} \end{aligned} \tag{8}$$

Where $\lambda_1, \lambda_2 \in \mathbb{R}^{n\times m}$, $\nu \in \mathbb{R}^m$, and $\lambda_3 \in \mathbb{R}$ and with complementary slackness conditions $\lambda_{1,ij}X_{ij} = \lambda_{1,ij}Z_{ij}$, $\lambda_{2,ij}X_{ij} = -\lambda_{2,ij}Z_{ij}$, $\lambda_3 \sum_{ij} Z_{ij} = \lambda_3 c$. The Lagrange dual function is then

$$g(\nu, \lambda_i) = \inf_{X,Z} \mathcal{L}(X, Z, \nu, \lambda_i)$$

Solving for the first order conditions $\nabla_X L = 0, \nabla_Z L = 0$, where the gradients are given by

$$\nabla_X \mathcal{L}(X, Z, \nu, \lambda_i) = 2LX + B + \mathbf{1}_{n\times m}\nu^\top + \lambda_1 - \lambda_2$$

$$\nabla_Z \mathcal{L}(X, Z, \nu, \lambda_i) = -\lambda_1 - \lambda_2 + \lambda_3 \mathbf{1}_{n\times m}$$

Solving for X yields $Y := X^* = \frac{1}{2}L^{-1}(-B - \nu\mathbf{1}_{n\times m} + 2\lambda + \lambda_3)$. Plugging back into g in conjunction with the condition that $\lambda_3\mathbf{1}_{n\times m} = \lambda_1 + \lambda_2$ yields

$$g(\nu, \lambda_i) = \langle Y, LY\rangle + \langle B, Y\rangle + (\mathbf{1}_{n\times m}^\top Y)\cdot \nu - \lambda_3 c + \sum_{ij}(\lambda_{3,ij} - \lambda_{2,ij}) \tag{9}$$

And the dual problem is

$$\max_{\nu,\lambda_i} g(\nu, \lambda_i) \quad \text{s.t. } \nu \in \mathbb{R}^m, \lambda_1, \lambda_2 \in \mathbb{R}_+^{n\times m}, \lambda_3 \in \mathbb{R}_+$$

KKT Conditions:

- Primal feasibility: $\mathbf{1}^\top X = 0, \quad X_{ij} \le Z_{ij}, \quad -X_{ij} \le Z_{ij}, \quad \sum_{ij} Z_{ij} \le c$
- Dual feasibility: $\lambda_i \ge 0$
- Complementary slackness: $\lambda_{1,ij}X_{ij} = \lambda_{1,ij}Z_{ij}$, $\lambda_{2,ij}X_{ij} = -\lambda_{2,ij}Z_{ij}$, $\lambda_3 \sum_{ij} Z_{ij} = \lambda_3 c$
- Stationarity: $\nabla_X \mathcal{L}(X, \lambda_i, \nu) = 0$, $\nabla_Z \mathcal{L}(X, \lambda_i, \nu) = 0$