EECS 127/227AT **Fall 2020**
Homework 9
Homework 9 is due on Gradescope by Friday 11/13 at 11.59 p.m.

1 Proving convexity via duality

This question is connected with the use of duality in SOCP. The relevant sections of the textbooks are Sec. 10.1 of the textbook of Calafiore and El Ghaoui and Sec. 4.4.2 of the textbook of Boyd and Vandenberghe.

Consider the function $f : \mathbb{R}^n_{++} \to \mathbb{R}$, with values

$$f(x) = 2 \max_t \; t - \sum_{i=1}^n \sqrt{x_i + t^2}. \tag{1}$$

(a) Explain why the problem in (1) that defines f is a convex optimization problem (in the variable t). That is, explain why the problem $\max_t \; t - \sum_{i=1}^n \sqrt{x_i + t^2}$ is a convex one.

As a reminder, an optimization problem involving the maximization of a concave function is called convex, since it is equivalent to minimizing a convex function.

(b) Formulate the optimization problem in (1) as an SOCP.

Hint: Introduce a slack variable z_i for each term in the summation, and express the condition $z_i \geq \sqrt{x_i + t^2}$ as a second-order cone constraint.

(c) Is f convex? (Recall that f is a function of $x \in \mathbb{R}^n_{++}$.)

(d) For $y \in \mathbb{R}^n_{++}$, let

$$g(y) := \max_{x>0} \; -x^\top y - f(x).$$

Namely, $g(y)$ is the value of the Fenchel conjugate of f at $-y$, where $y \in \mathbb{R}^n_{++}$. Show that we can write $g(y)$ as

$$g(y) = \min_t \; -2t + \sum_{i=1}^n \max_{x_i>0} \left(2\sqrt{x_i + t^2} - x_i y_i\right).$$

Remark: The interchange of $\min_t$ and $\max_{x>0}$ that is needed to do this can be justified because the function

$$-x^\top y - 2t + 2\sum_{i=1}^n \sqrt{x_i + t^2},$$

defined for $x \in \mathbb{R}^n_+$ and $t \in \mathbb{R}$, for fixed $y \in \mathbb{R}^n_{++}$, is concave in x for fixed t and convex in t for fixed x. You can take this for granted.

(e) Show that

$$g(y) = \sum_{i=1}^n \frac{1}{y_i} - \frac{1}{\sum_{i=1}^n y_i}.$$

Hence conclude that this function, defined on $\mathbb{R}^n_{++}$. is convex

Hint: First use calculus to solve the inner maximization problem in the expression you derived for $g(y)$ in the preceding part of the question, then use calculus solve the outer minimization problem.

2 Localization

This question describes how to express the problem of finding the location of an object, which shows up in many applications (such as GPS), as an SOCP. The relevant portions of the textbooks are Secs. 10.1 and 10.2 of the textbook of Calafiore and El Ghaoui and Sec. 4.4.2 of the textbook of Boyd and Vandenberghe.

We are given anchor positions $x_i \in \mathbb{R}^3$, and associated measurements of what appear to be distances R_i, $i = 1, \ldots, m$ from these anchor points to an unknown object. The problem is to estimate a position $x \in \mathbb{R}^3$ for the object, and an associated measure of uncertainty around the estimated point. For this, we take the measurement R_i to represent an upper bound for the true distance of the object from anchor point i. In the sequel, we denote by S_i, $i = 1, \ldots, m$ the spheres of center x_i and radius R_i, $i = 1, \ldots, m$. We are aiming at estimating the location of the object, which ought to be a point in the intersection of the spheres, $\bigcap_{i=1}^m S_i$, and at estimating the size of the intersection (Fig. 1) as a measure of the uncertainty in our estimate of the location of the object.

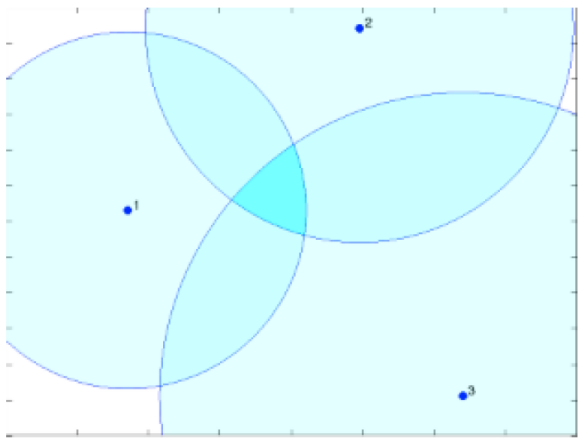

Figure 1: Localization problem

(a) We first focus on the "inner" approximation problem, of finding the largest radius of a sphere contained in the set of points that are consistent with the range measurements (Fig. 2). Show that the center of this sphere, x_0, and its radius, R_0, can be found by solving an SOCP, which you will describe.

(b) In some cases the measurements are not consistent, in the sense that the intersection between the spheres S_i, $i = 1, \ldots, m$ is empty. This may be due to a few faulty sensors. One way to try to identify the faulty sensors is to look for the smallest correction (increase) to the measured ranges, so that the measurements become consistent (Fig. 3). How would you compute those corrected range values?

Hint: Think about an l_1-norm heuristic applied to the range corrections.

(c) We now turn to the problem of estimating an "outer" approximation, again in the form of a sphere, to the intersection (Fig. 4). In the sequel, we assume that the measurements are consistent, and, for notational simplicity, that 0 is contained in the intersection, so that $z \geq 0$, where the m-vector z has components $z_i = R_i^2 - x_i^T x_i$, $i = 1 \ldots, m$. Explain why a sphere of center x_0 and radius R_0 is such an outer approximation if and only if

$$\|x - x_0\|_2^2 \leq R_0^2 \text{ for every } x \text{ such that } \|x - x_i\|_2^2 \leq R_i^2, \quad i = 1, \ldots, m.$$

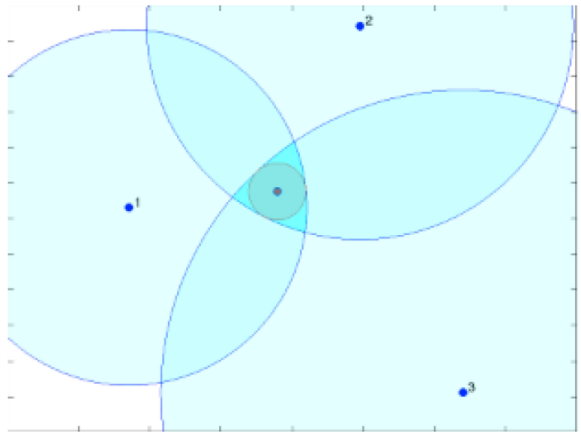

Figure 2: Localization problem: inner approximation

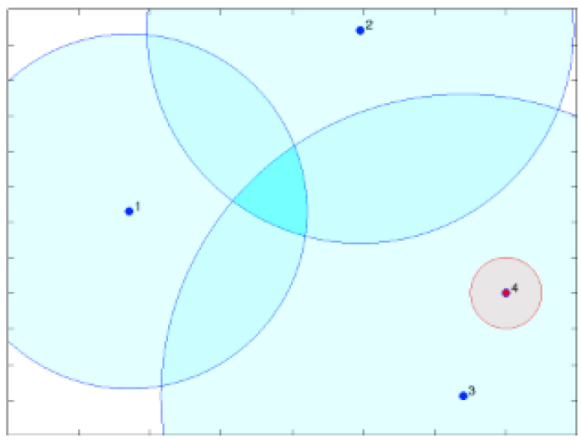

Figure 3: Localization problem with one faulty measurement

(d) Show that a sufficient condition for the above to hold is that there exists an m-vector $\lambda \geq 0$ such that

$$\forall\, x \in \mathbb{R}^3 \;:\; \|x - x_0\|_2^2 \leq R_0^2 + \sum_{i=1}^{m} \lambda_i \left(\|x - x_i\|_2^2 - R_i^2 \right), \tag{2}$$

and express the corresponding optimal radius, R_0^*, for the condition (2) as

$$(R_0^*)^2 = \min_{x_0, \lambda \geq 0} \; F(x_0, \lambda),$$

where

$$F(x_0, \lambda) \doteq \sum_{i=1}^{m} \lambda_i R_i^2 + \max_{x} \left(\|x - x_0\|_2^2 - \sum_{i=1}^{m} \lambda_i \|x - x_i\|_2^2 \right).$$

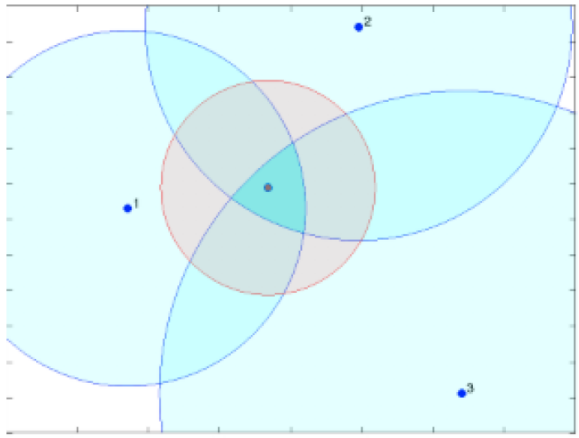

Figure 4: Localization problem: outer approximation

(e) Show that

$$F(x_0, \lambda) = x_0^T x_0 + \lambda^T z + \begin{cases} \frac{\|X\lambda - x_0\|_2^2}{S-1} & \text{if } S := \sum_{i=1}^m \lambda_i > 1, \\ 0 & \text{if } S = 1, \quad x_0 = X\lambda, \\ +\infty & \text{otherwise,} \end{cases}$$

where $z_i := R_i^2 - x_i^T x_i$, $i = 1 \ldots, m$, and $X := [x_1, \ldots, x_m]$.

(f) Express the problem above for determining the optimal radius for the condition (2), i.e.

$$\begin{aligned} \min_{x_0, \lambda} \ & F(x_0, \lambda) \\ \text{s.t.} \ & \lambda \geq 0, \end{aligned}$$

as an SOCP.

Hint: As is the case in many SOCP problems, you may find it convenient to use the fact a constraint of the form

$$v^T v \leq 2yz, \ \ y \geq 0, \ z \geq 0,$$

which is called a *hyperbolic constraint* (see Sec. 10.1.1. of the textbook of Calafiore and El Ghaoui) can be equivalently expressed as

$$\left\| \begin{bmatrix} v \\ \frac{1}{\sqrt{2}}(y - z) \end{bmatrix} \right\|_2 \leq \frac{1}{\sqrt{2}}(y + z).$$

3 Capacity of a discrete memoryless channel with bounded expected input cost

This question is connected with geometric programming, but is basically a problem about duality. For the geometric programming aspects the relevant portions of the textbooks are Sec. 9.7 of the textbook of Calafiore and El Ghaoui and Sec. 4.5 of the textbook of Boyd and Vandenberghe. For duality the relevant portions of the textbooks are Sec. 8.5 of the textbook of Calafiore and El Ghaoui and Secs. 5.1 and 5.2 of the textbook of Boyd and Vandenberghe.

Background: In communications a *discrete memoryless channel* takes an input from the finite set $\{1, \ldots, n\}$ and produces an output from the finite set $\{1, \ldots, m\}$. You can think of the inputs as signals that a transmitter can send out and the outputs as possible decisions made by the receiver after receiving the noise-corrupted transmitted signal. The channel is characterized by its *channel matrix*, which is an $n \times m$ matrix with entries p_{ij}, which satisfy

$$\begin{aligned} p_{ij} &\geq 0, \text{ for all } 1 \leq i \leq n, 1 \leq j \leq m, \\ \sum_{j=1}^{m} p_{ij} &= 1, \text{ for all } 1 \leq i \leq n. \end{aligned}$$

The interpretation is that p_{ij} represents that probability that the receiver will make the decision j when it receives the noise-corrupted version of the signal i sent by the transmitter.

We let r_i denote the entropy of the output when the input is i. Thus

$$r_i := -\sum_{j=1}^{m} p_{ij} \log p_{ij}, \; i = 1, \ldots, n.$$

We are also going to assume that there is a *cost* s_i to transmit the signal i, for $1 \leq i \leq n$.

Suppose that over a long sequence of transmissions the transmitter transmits the signal i a fraction p_i of the time, where $p_i \geq 0$ for all $1 \leq i \leq n$ and $\sum_{i=1}^{n} p_i = 1$. Then the average cost per transmission is $p^\top s$, where $p \in \mathbb{R}^n$ is the column vector of the p_i and $s \in \mathbb{R}^n$ is the column vector of the costs. Let us also define

$$q^\top := p^\top P.$$

Then $q \in \mathbb{R}^m$ has entries q_j equal to the expected fraction of times the receiver decides j. Note that $q_j \geq 0$ for all $1 \leq j \leq m$ and $\sum_{j=1}^{m} q_j = 1$.

A theorem of Shannon tells us that the rate at which we can communicate over such a discrete memoryless channel (with the transmitter choosing input i a fraction p_i of the time) is given by

$$-\sum_{j=1}^{m} q_j \log q_j - p^T r.$$

Namely it is difference between the entropy of the output distribution and the term $p^T r$, which is called the *conditional entropy* of the output given the input. This difference is in fact called the *mutual information* between the input and the output and can be shown to be nonnegative (simply by using the convexity of the function $x \log x$). ■

In this question we are going to study the problem of maximizing the rate at which we can communicate over the channel subject to a bound on average cost per transmission. The variable over which we will optimize will be choice of signalling strategy at the transmitter, namely p.

Suppose the bound on average cost per transmission is S. Convince yourself that the problem can be posed as

$$\begin{aligned}
\max_{p,q} \quad & -\sum_{j=1}^{m} q_j \log q_j - p^T r \\
\text{s.t.} \quad & p^\top P = q, \\
& p^\top s \leq S, \\
& p^\top \mathbb{1} = 1, \\
& p \geq 0,
\end{aligned}$$

where $\mathbb{1}$ denotes the vector of ones and the last inequality is interpreted coordinatewise. Note that the domain of the objective function is the set of pairs (p, q) where $p \in \mathbb{R}^n$ and $q \in \mathbb{R}^m_{++}$.

In this question, we will show that the Lagrangian dual of this problem can be written as the GP

$$\begin{aligned}
\min_{w,z} \quad & w^S \sum_{j=1}^{m} z_j \\
\text{s.t.} \quad & w^{s_i} \prod_{j=1}^{m} z_j^{P_{ij}} \geq e^{-r_i}, \; i = 1, \ldots, N, \\
& w \geq 1.
\end{aligned}$$

Remark: To check that this is a GP, just note that the first set of inequalities is the same as

$$e^{-r_i} w^{-s_i} \prod_{j=1}^{m} z_j^{-P_{ij}} \leq 1, \; i = 1, \ldots, N,$$

and the last inequality is the same as

$$w^{-1} \leq 1.$$

Further, the objective function is the posynomial $\sum_{j=1}^{m} w^S z_j$. Also, recall that in a GP it is implicit that all the variables have to be strictly positive.

Remark: Since we are used to computing duals for minimization problems, we will consider the problem

$$\begin{aligned}
\min_{p,q} \quad & \sum_{j=1}^{m} q_j \log q_j + p^\top r \\
\text{s.t.} \quad & p^\top P = q, \\
& p^\top s \leq S, \\
& p^\top \mathbb{1} = 1, \\
& p \geq 0,
\end{aligned} \tag{3}$$

and dualize it to get a maximization problem. We will then convert that maximization problem into a minimization problem of the desired form. It is not necessary to go through this if one does the Lagrangian calculation directly in a way that is suitable for maximization problems, but it may be less confusing to do it this way because it fits the pattern we have become familiar with.

(a) Write the Lagrangian for the primal problem in (3) and, after eliminating the dual variables for the inequality constraints $p \geq 0$ in (3), show that Lagrangian dual of the problem in (3) can be written as

$$\begin{aligned} \max_{\gamma,\nu,\mu} \quad & -\sum_{j=1}^{m} e^{\nu_j - 1} + \mu - \gamma S \\ \text{s.t.} \quad & r - P\nu - \mu\mathbb{1} + \gamma s \geq 0, \\ & \gamma \geq 0, \end{aligned}$$

Hint: In computing the dual objective function you will need to minimize $\sum_{j=1}^{m} q_j \log q_j + \nu^\top q$ over $q \in \mathbb{R}^m_{++}$. where ν denotes the vector of dual variables for the equality constraints $p^\top P = q$. This can be done coordinate by coordinate and will involve the Fenchel dual of the convex function $q_j \log q_j$ with domain $\mathbb{R}_{++}$, evaluated at $-\nu_j$.

(b) If ν denotes the vector of dual variables for the equality constraints $p^\top P = q$ in (3) and μ denotes the dual variable for the equality constraint $p^\top \mathbb{1} = 1$ in (3), introduce the new variables $\alpha_j := -\nu_j - \mu$ for $1 \leq j \leq m$ and show that the dual can be written in the simplified form

$$\begin{aligned} \max_{\gamma,\nu,\mu} \quad & -e^{(\mu-1)} \sum_{j=1}^{m} e^{\alpha_j} + \mu - \gamma S \\ \text{s.t.} \quad & r + P\alpha + \gamma s \geq 0, \\ & \gamma \geq 0, \end{aligned}$$

(c) Eliminate the dual variable μ for the equality constraint $p^\top \mathbb{1} = 1$ in (3) to show that the dual can be further simplified to the form

$$\begin{aligned} \max_{\gamma,\nu} \quad & -\log\left(\sum_{j=1}^{m} e^{\alpha_j}\right) - \gamma S \\ \text{s.t.} \quad & r + P\alpha + \gamma s \geq 0, \\ & \gamma \geq 0, \end{aligned}$$

then show how this is equivalent to a GP.

Hint: Eliminating μ from the objective function is a matter of simple calculus, since μ is unconstrained and the objective function is concave in μ.

4 Spectrahedron

This question explores the structure of the feasibility set associated to a linear matrix inequality. The feasibility set of any semidefinite program is the feasibility set of a linear matrix inequality, so this question aimed at developing a better understanding of the feasibility sets of semidefinite programs.

The relevant sections of the textbooks are Secs. 11.2.2 and 11.2.3 of the textbook of Calafiore and El Ghaoui and Sec. 4.6.2 of the textbook of Boyd and Vandenberghe.

Given symmetric matrices $F_0, F_1, \ldots, F_m \in \mathbb{S}^n$, the set of symmetric matrices

$$\{F_0 + x_1 F_1 + \ldots + x_m F_m : x \in \mathbb{R}^m\},$$

where x denotes $\begin{bmatrix} x_1 & \ldots & x_m \end{bmatrix}^\top$, is called a *linear matrix pencil*. It is an affine subspace of the vector space $\mathbb{S}^n$.

We write $F(x)$ for $F_0 + \sum_{i=1}^m x_i F_i$.

The intersection of a linear matrix pencil with the cone of positive semidefinite matrices is called a *spectrahedron*. The condition that needs to be satisfied for this, namely

$$F(x) \succeq 0,$$

is called a *linear matrix inequality*. The term "spectrahedron" is also used to refer to

$$\{x \in \mathbb{R}^m : F(x) \succeq 0\},$$

in which case we think of it as a subset of $\mathbb{R}^m$. This set is also called the feasibility set of the LMI.

In this question we will study the spectrahedron associated to the linear matrix pencil

$$F(x, y) = \begin{bmatrix} 1 & 1-x & -x \\ 1-x & 1 & -y \\ -x & -y & 2y \end{bmatrix}.$$

We will think of this spectrahedron as a subset of $\mathbb{R}^2$, with the vectors in $\mathbb{R}^2$ being written as $\begin{bmatrix} x & y \end{bmatrix}^\top$.

(a) Find all the principal minors of $F(x, y)$.

Remark: For a square $n \times n$ matrix A, for each $1 \leq k \leq n$ there are $\binom{n}{k}$ principal k-minors. These are found by picking a subset $J \subset \{1, \ldots, n\}$ of size k and considering the $k \times k$ matrix one gets from A by erasing all the rows with index not in J and all the columns with index not in J and then taking the determinant of this $k \times k$ matrix.

Since $F(x)$ is a 3×3 matrix, there will be three principal 1-minors (which are just the diagonal entries), three principal 2-minors, and one principal 3-minor (which is just the determinant). You are asked to find these.

(b) For a symmetric matrix $A \in \mathbb{S}^n$, it is known that A is positive semidefinite iff all its principal minors are nonnegative.

Based on this, characterize the spectrahedron defined by $F(x, y)$ as the intersection, over a finite collection of polynomials in the variables (x, y), of the sets where these polynomials are nonnegative.

A graph of the spectrahedron is shown in Figure 5.

(c) Show that if $F(x, y)$ is positive semidefinite then we must have $0 \leq x \leq 2$.

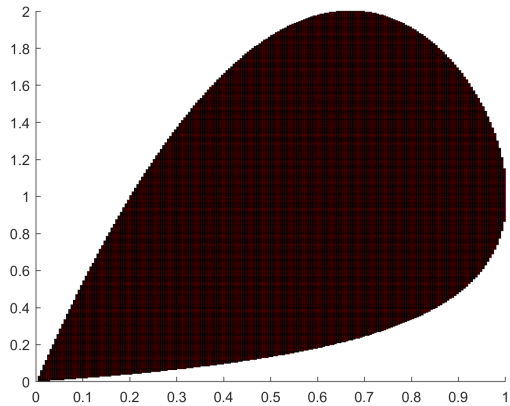

Figure 5: Spectrahedron. This is a convex set.

(d) Suppose $x = 0$. Show that the only value of y for which $F(0, y)$ is positive semidefinite is $y = 0$. What is the rank of $F(0, 0)$?

(e) Suppose $x = 2$. Show that there is no value of y for which $F(2, y)$ is positive semidefinite.

(f) What condition must the pair (x, y) satisfy for the rank of $F(x, y)$ to be strictly less than 3?

(g) Assume $0 < x < 2$. Verify that this makes the leading 2×2 block of $F(x, y)$ positive definite. Fixing x, apply the Schur complement criterion to show that $F(x, y)$ is positive semidefinite iff the determinant of $F(x, y)$ is nonnegative and positive definite iff the determinant of $F(x, y)$ is strictly positive.

Remark: From this part of the question we learn that the boundary of the spectrahedron in Figure 5 consists of pairs (x, y) where the rank of $F(x, y)$ is strictly less than 3.

(h) Show that for (x, y) in the spectrahedron, the rank of $F(x, y)$ is 1 precisely when $(x, y) = (0, 0)$.

Remark: From this part and the preceding part of the question we learn that the rank of $F(x, y)$ is 3 on the interior of the spectrahedron, is 2 at all points on the boundary of the spectrahedron except the point $(x, y) = (0, 0)$. where the rank of $F(0, 0)$ is 1.

5 Immersing a graph in Euclidean space with low distortion

This question is optional.

Many approximation algorithms for combinatorial problems in computer science start by first representing the combinatorial structure (such as a weighted graph) in some Euclidean space and then running an optimization problem in Euclidean space to get an approximation for the combinatorial quantity of interest (such as the maximum cut in a weighted graph). It is important that the mapping have *low distortion*, for instance that Euclidean distance between the vectors representing a pair of vertices of the graph is not very different from the weight of the corresponding edge in the underlying graph, if any. This question is motivated by such concerns.

The relevant portions of the textbooks are Secs. 11.2.1, 11.2.1 and 11.3.1 of the textbook of Calafiore and El Ghaoui and Sec. 4.6.2 of the textbook of Boyd and Vandenberghe.

Let $G = (V, E)$ be a graph. Here $V = \{1, \ldots, n\}$ is the vertex set and E is the set of edges, i.e. a subset of unordered pairs of vertices (i, j), $1 \leq i, j \leq n$. We assume that there are no self loops, i.e. if (i, j) is an edge then $i \neq j$.

Let m denote the number of edges. We assume that $m \geq 1$.

For each edge e of the graph, we are given two real numbers $\alpha_e \leq \beta_e$. You can think of the edge e as having some weight. α_e is a lower bound and β_e is an upper bound to the allowed approximations to that weight. We will use the term *capacitated graph* to refer to the graph together with the the pair of real numbers $\alpha_e \leq \beta_e$ for each edge.

We say that this capacitated graph can be *immersed* in the Euclidean space $\mathbb{R}^n$ if we can find vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ such that for each edge $e = (i, j)$ we have

$$\alpha_{(i,j)} \leq \|v_i - v_j\|_2^2 \leq \beta_{(i,j)}.$$

In this question we will show how to pose the decision problem of whether a given capacitated graph on n vertices can be immersed in $\mathbb{R}^n$ as a feasibility problem for an SDP in standard form.

Remark: Recall that an SDP is standard form looks like:

$$\begin{aligned} \min_X \ & \text{trace}(CX) \\ \text{s.t.} \ & \text{trace}(A_i X) = b_i, \ i = 1, \ldots, m, \\ & X \succeq 0. \end{aligned}$$

Here the minimization is over matrices $X \in \mathbb{R}^{l \times l}$. The matrices $C, A_1, \ldots, A_m \in \mathbb{S}^l$ as well as the vectors $b_1, \ldots, b_m \in \mathbb{R}^l$ are given. The constraint $X \succeq 0$ is the constraint that X should be positive semidefinite.

Also recall that to pose a minimization problem as a feasibility problem, we can just take the objective to be the constant 0 (so the question then just becomes whether the value of the problem is 0, in which case the problem is feasible, or ∞, in which case the problem is infeasible). For an SDP in standard form to be a feasibility problem, therefore, we could just take C to be the zero matrix.

Remark: Note that the dimension of the Euclidean space into which we are seeking to immerse the capacitated graph equals the number of vertices.

(a) Let

$$V := \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix},$$

and

$$Z := V^T V.$$

Note that $Z \in \mathbb{S}^n$. Show that it is possible to find matrices $A^{ij} \in \mathbb{S}^n$ for $(i,j) \in E$ such that the constraints

$$\alpha_{(i,j)} \leq \|v_i - v_j\|_2^2 \leq \beta_{(i,j)}$$

can be written as

$$\alpha_{(i,j)} \leq \text{trace}(A^{ij} Z) \leq \beta_{(i,j)}.$$

Hint: Expand out $||v_i - v_j||_2^2$ by using the fact that $||x||_2^2 = x^T x$ for any vector x. Then use the fact that $\mathrm{tr}(A^{ij} Z) = \sum_{k=1}^n \sum_{l=1}^n A_{kl}^{ij} Z_{kl}$, with $Z_{kl} = v_k^T v_l$, to explicitly determine the matrix entries A_{kl}^{ij}. Many of them will be zero.

(b) Introduce slack variables representing $\text{trace}(A^{ij} Z) - \alpha_{(i,j)}$ and $\beta_{(i,j)} - \text{trace}(A^{ij} Z)$ for each $(i,j) \in E$. Define a block diagonal matrix $X \in \mathbb{S}^{(n+2m)}$ whose top $n \times n$ block is Z and whose bottom $2m \times 2m$ block is a diagonal matrix having the first m slack variables and then the next m slack variable. Show that it is possible to define matrices $U^{ij} \in \mathbb{S}^{(n+2m)}$ and $V^{ij} \in \mathbb{S}^{(n+2m)}$ for each $(i,j) \in E$ such that the conditions

$$\alpha_{(i,j)} \leq \text{trace}(A^{ij} Z) \leq \beta_{(i,j)}$$

are equivalent to the two conditions

$$\text{trace}(U^{ij} X) = \alpha_{(i,j)}$$

and

$$\text{trace}(V^{ij} X) = \beta_{(i,j)}.$$

(c) Show that it is possible to force the desired block diagonal structure on X and the diagonal structure on the bottom $2m \times 2m$ block of X by introducing a family of matrices $W^{ab} \in \mathbb{S}^{(n+2m)}$ for some family $\mathcal{A}$ of pairs (a,b) with $1 \leq a < b \leq n + 2m$ and forcing the conditions

$$\text{trace}(W^{ab} X) = 0, \quad (a,b) \in \mathcal{A}.$$

At this point you should have the desired formulation as an SDP feasibility problem in standard form of the decision problem of whether it is possible to immerse the capacitated graph in Euclidean space of dimension equal to the number of vertices of the graph.