IE531: Algorithms for Data Analytics
Spring, 2023
Homework 2: Semidefinite Programming and Data Analytics
Due Date: February 17, 2023

Instructions

1. You will submit a PDF-version of your answers on Compass on-or-before midnight of the due date.

Instructions

In Section 12 of Lesson 1, you were introduced to the canonical *Semidefinite Program* (SDP)

$$\left(\begin{array}{c} min : \mathbf{C} \circ \mathbf{X} \\ \mathbf{A}_i \circ \mathbf{X} = b_i, i \in \{1, 2, \ldots, m\} \\ \mathbf{X} \succeq \mathbf{0} \end{array}\right) \equiv \left(\begin{array}{c} min : trace(\mathbf{C}^T \mathbf{X}) \\ trace(\mathbf{A}_i^T \mathbf{X}) = b_i, i \in \{1, 2, \ldots, m\} \\ \mathbf{X} \succeq \mathbf{0} \end{array}\right) \equiv \left(\begin{array}{c} min : trace(\mathbf{C}\mathbf{X}) \\ trace(\mathbf{A}_i \mathbf{X}) = b_i, i \in \{1, 2, \ldots, m\} \\ \mathbf{X} \succeq \mathbf{0} \end{array}\right)$$

where $\mathbf{C}$ and $\{\mathbf{A}_i\}_{i=1}^m$ are symmetric matrices that are known, and $\mathbf{X}$ is a symmetric matrix of unknown variables. I showed you how to use the package `cvxpy` to solve instance of an SDP, as well. In this homework, you are going to show that a some of the archetypal problems in Data Analytics can be cast as an instance of an SDP (and effectively solved using an SDP-solver).

We will pay attention to a particular problem called the *Norm Minimization* problem in this HW. We have a r-many, d-dimensional data vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r \in \mathcal{R}^d$. We wish to find a d-dimensional vector that "represents" this set of vectors, by solving

$$\min_{\mathbf{y}} \left\{ \max_{i=1,2,\ldots,r} \|\mathbf{y} - \mathbf{b}_i\|_2 \right\}.$$

That is, $\mathbf{y} \in \mathcal{R}^d$ is a vector that minimizes the maximum norm/distance between itself and each of the r-many data vectors. You will show that this problem is essentially an instance of an SDP, which can be solved by using an SDP-solver.

Note

I covered this HW in Lecture 6. See 48:00 minute-marker of Lesson 6 on Echo360.org.

This will be a pattern in all of my lectures – you have to look for it in class, or in the taped-lectures. If you needed additional help, you could/should use the TA's Office Hours.

Note

1. (*20 points*) **Gershgorin's Theorem**

   (a) (*10 points*) Let $\mathbf{A}$ be an $n \times n$ matrix of real numbers. You can view the matrix as a collection of rows, or as a collection of columns. Let us start with a row-view of the matrix. For $i \in \{1, 2, \ldots, n\}$, let

$$R_i = \sum_{j=1; j \neq i}^{n} | a_{i,j} |$$

denote the sum of the absolute-values of the entries of $\mathbf{A}$ in the i-th row, not including the diagonal-element (i.e. $a_{i,i}$). Using this, define the interval $[a_{i,i} - R_i, a_{i,i} + R_i]$. These intervals are called the set of *Gershgorin Row Intervals*. Show that the eigenvalues of $\mathbf{A}$ lie in some Gershogorin Row Interval.

See this link for a formal proof.

(b) (*10 points*) As a follow on to problem 1a, define the Gershgorin Column Intervals

$$\{[a_{i,i} - C_i, a_{i,i} + C_i]\}_{i=1}^{n}, \text{ where } C_i = \sum_{j=1; j \neq i}^{n} | a_{j,i} |,$$

and show that the eigenvalues of $\mathbf{A}$ lie in some Gershgorin Column Interval.

See this link for a formal proof.

The utility of these results are that if the i-th row- or column-sum is small, then one of the eigenvalues of $\mathbf{A}$ will be more-or-less near the value of $a_{i,j}$. For diagonal matrices, this result says the entries in the diagonal are its eigenvalues. We will use the above two results to establish the positive semidefiniteness of a matrix (i.e. all its eigenvalues are non-negative).

2. *(30 points)* **Second-Order Cone Programming**: The *Second-Order Cone* (also called *Lorenz Cone*, *Ice-Cream Cone*, etc) is the set of n-dimensional vectors, $\mathcal{Q}_n$, where

$$\mathcal{Q}_n := \left\{ \mathbf{x} \in \mathcal{R}^n \,\middle|\, \mathbf{x} = \begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix}, y_0 \geq \|\mathbf{y}\|_2, y_0 \in \mathcal{R}, \mathbf{y} \in \mathcal{R}^{n-1} \right\}.$$

Show that

(a) $\mathcal{Q}_n$ is convex, and

$$\mathbf{x}_1 = \begin{pmatrix} y_0^1 \\ \mathbf{y}_1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} y_0^2 \\ \mathbf{y}_2 \end{pmatrix} \in C_n \Rightarrow \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 = \begin{pmatrix} \lambda y_0^1 + (1-\lambda) y_0^2 \\ \lambda \mathbf{y}_1 + (1-\lambda)\mathbf{y}_2 \end{pmatrix} \in C_n,$$

this is because

$$\|\lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2\|_2 \leq \lambda\|\mathbf{y}_1\|_2 + (1-\lambda)\|\mathbf{y}_2\|_2 \leq \lambda y_0^1 + (1-\lambda) y_0^2.$$

If you need help with showing the convexity of the 2-norm/Euclidean-distance, see this link.

(b)

$$(\mathbf{x} \in \mathcal{Q}_n) \Leftrightarrow \left(\begin{pmatrix} y_0 & \mathbf{y}^T \\ \mathbf{y} & y_0 \times \mathbf{I} \end{pmatrix} \geq \mathbf{0} \right)$$

The matrix

$$\begin{pmatrix} y_0 & \mathbf{y}^T \\ \mathbf{y} & y_0 \times \mathbf{I} \end{pmatrix}$$

is the Schur Complement, and is symmetric, with $y_0 \geq 0$ along the diagonal; and the only non-diagonal, non-zero entries are in the first-row and first-column of the matrix. The Gershgorin intervals $[y_0 - y_i, y_0 + y_i]$ will never have any negative-values (why?).

If $\mathbf{x} \in Q_n$, then (using the notation from earlier), $y_0 \geq \|\mathbf{y}\|_2$. This means, $\mid \mathbf{y}_i \mid \leq y_0$ for any $i \in \{1, 2, \ldots, n-1\}$ (why?). Using Gershogorin's Theorem we know that none of the eigenvalue of the matrix shown above can be negative. The reverse implication will require the use of results about Diagonally Dominant matrices (link).

3. (*30 points*) **Norm Minimization**: We have a r-many, d-dimensional data vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r \in \mathcal{R}^d$. We wish to find a $\mathbf{y} \in \mathcal{R}^d$ that solves

$$\min_{\mathbf{y}} \left\{ \max_{i=1,2,\ldots,r} \|\mathbf{y} - \mathbf{b}_i\|_2 \right\}.$$

Show that

$$\left(\begin{array}{c} \min t \\ \|\mathbf{y} - \mathbf{b}_i\| \leq t \end{array} \right) \Leftrightarrow \left(\begin{array}{c} \max -t \\ \left(\begin{array}{c} 0 \\ \mathbf{b}_i \end{array} \right) - \left(\begin{array}{c} -t \\ \mathbf{y} \end{array} \right) \in Q_{d+1} \text{ for } i \in \{1, 2, \ldots, r\} \end{array} \right)$$

Straightforward. Ask the TA for help if you are lost, if that did not work out, come see me after class.

4. (*20 points*) **SDP-formulation of the Norm-Minimization Problem** Using the above observations present an SDP-formulation for finding the vector $\mathbf{y} \in \mathcal{R}^d$ that solves the Norm-minimization problem.

Hint: Use your Google-skills to look for a formal treatment that Second-Order Cone Programming (SOCP) Problems are a special instance of Semidefinite Programming (SDP) Problems.

There are many references on the Web that show the SOCP-instance is a special case of SDP. For example, see the Wikipedia link on SOCP. All most all of them invoke the Schur Complement, introduced above in problem 2b. In essence, if $\mathbf{A}$ is a positive definite matrix (i.e. $\mathbf{A} \succ \mathbf{0}$).

Note, the symbol $\succ$ (resp. $\succeq$) is not be confused with $>$ (resp. $\geq$). The symbol $\succ$ (resp. $\succeq$) means all the eigenvalues of $\mathbf{A}$ are strictly-positive (resp. non-negative). For what it is worth, there is a "slight curve" on $\succ$ compared to $>$ (resp. there is a "slight curve" on $\succeq$ compared to $\geq$).

If we constructed a symmetric matrix $\mathbf{X}$, where

$$\mathbf{X} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix},$$

and if $\mathbf{S} = \mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B}$, then

$$(\mathbf{X} \succeq \mathbf{0}) \Leftrightarrow (\mathbf{S} \succeq \mathbf{0}).$$

That is, all the eigenvalues of $\mathbf{X}$ are greater-than-or-equal-to zero if and only if all the eigenvalues of $\mathbf{S}$ are greater-than-or-equal-to zero. You can find a proof of this claim on the Wikipedia page on the Schur Complement.

From Section 15 of Lesson 1of my notes, an instance of SOCP requires us to solve

$$\begin{gathered}\min_{\mathbf{x}} \mathbf{f}^T\mathbf{x}\\ \text{subject to:}\\ \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T\mathbf{x} + d_i, \forall i \in \{1, 2, \ldots, m\},\end{gathered}$$

where $\{\mathbf{A}_i \in \mathcal{R}^{k_i \times n}, \mathbf{b}_i \in \mathcal{R}^{k_i}, \mathbf{c}_i \in \mathcal{R}^n, d_i \in \mathcal{R}\}_{i=1}^m$, and $\mathbf{x} \in \mathcal{R}^n$ is the n-dimensional vector of unknowns (to be solved-for). Without loss of generality, we can assume $\mathbf{c}_i^T\mathbf{x} + d_i > 0$ (Why?). This would mean we can write $\|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T\mathbf{x} + d_i$ as

$$\begin{aligned}
\begin{pmatrix} (\mathbf{c}_i^T\mathbf{x} + d_i) \times \mathbf{I} & (\mathbf{A}_i\mathbf{x} + \mathbf{b}_i) \\ (\mathbf{A}_i\mathbf{x} + \mathbf{b}_i)^T & (\mathbf{c}_i^T\mathbf{x} + d_i) \times \mathbf{I} \end{pmatrix} & \underbrace{\succeq}_{\text{do not confuse it with } \geq} \mathbf{0} \\
\Leftrightarrow \quad & (\mathbf{c}_i^T\mathbf{x} + d_i) - \frac{(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i)^T(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i)}{(\mathbf{c}_i^T\mathbf{x} + d_i)} \geq 0 \\
\Leftrightarrow \quad & (\mathbf{c}_i^T\mathbf{x} + d_i)^2 \geq \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2^2 \\
\Leftrightarrow \quad & \underbrace{(\mathbf{c}_i^T\mathbf{x} + d_i)}_{>0} \geq \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2
\end{aligned}$$

We can equivalently use the positive semidefiniteness of the Schur Complement as a proxy for the constraints of the SOCP. This is why the SOCP is a special case of SDP.