Lecture 7: Stochastic Block Model 2

Scribe: Antares Chen 10/24/2018

7.1 Introduction

In this note, we continue our discussion of recovering planted partitions in a stochastic block model $\mathcal{G}_{n,p,q}$ by presenting an SDP based algorithm. Our analysis will use SDP duality and a version of complementary slackness to show that when p and q are sufficiently separated (or rather when the average internal degree is sufficiently different from the average external degree), the SDP is uniquely solved by a matrix encoding the planted partition thus the algorithm exactly recovers the planted partition.

7.2 An Algorithm for Exact Recovery

Recall that the problem of recovering a planted partition given $G \sim \mathcal{G}_{n,p,q}$, tasks us with finding the hidden communities V_1 and V_2. The spectral based algorithm we previously presented approximately recovered V_1 and V_2 provided $p - q \geq O\left(\sqrt{\frac{p+q}{n}}\right)$. For this lecture, we will derive a semi-definite programming based algorithm that will *exactly* recover the partitions.

Similar to last week, we will need to assume that edges within communities are sampled at a higher probability than edges across communities. Instead of parameterizing over p and q, we will use the average internal degree $a = \frac{np}{2}$, the expected number of edges a vertex has to others in its own community, and the average external degree $b = \frac{nq}{2}$, the expected number of edges a vertex has to others in the other partition.

Provided that $a \gg b$ (exact condition to be provided later), a calculation last time demonstrated that the bisecting cut with the fewest edges maximizes the likelihood that it is the hidden cut separating the partitions in G. The min-bisection problem is in general NP-Hard and so a natural place to start building our SDP based algorithm for recovering the planted partition is to write a relaxation for min-bisection.

7.2.1 A Semi-definite Program

Let V and E be the set of vertices and edges in G. As we have done for max-cut and planted clique, we will first write a quadratic program, convert it to only contain quadratic terms of form $z_i z_j$, then replace each quadratic term with the inner product of two vectors $\langle v_i, v_j \rangle$, and finally replace $\langle v_i, v_j \rangle$ with X_{ij} for some $X \succeq 0$ since if $X \succeq 0$, there are real vectors $v_1, \ldots, v_n$ such that $X_{ij} = \langle v_i, v_j \rangle$.

Given graph G, we want to recover a cut V_1 and $V_2 = V - V_1$. To derive the quadratic program, let's start by looking at what we would want our solution to be. In particular we would want z_i to be defined by

$$z_i = \begin{cases} +1 & \text{if } i \in V_1 \\ -1 & \text{if } i \in V_2 \end{cases}$$

for each $i \in V$. The number of edges that are cut by (V_1, V_2) is then given by

$$\sum_{(i,j)\in E} \frac{1}{4}(z_i - z_j)^2$$

For example, if $i, j \in V_1$ then $(z_i - z_j)^2 = 0$ while if $i \in V_1$ and $j \in V_2$ then $(z_i - z_j)^2 = 4$. We now have our objective: minimize the above quantity counting the number of edges cut by (V_1, V_2). However, we also want $z_i = \pm 1$ for all i and our cut to be a bisection. The former is satisfied by constraining

$$z_i^2 = 1 \qquad \forall i \in V$$

To enforce that our cut is a bisection, we want to ensure that the number of z_i's set to $+1$ is the same as the number of z_i's set to -1. This should do the trick

$$\sum_{i\in V} z_i = 0$$

because if the number of $+1$'s is the same as the number of -1's then the sum should exactly be 0. To summarize, our quadratic program is the following.

$$\begin{aligned} \text{minimize} \quad & \sum_{(i,j)\in E} \frac{1}{4}(z_i - z_j)^2 \\ \text{subject to} \quad & \sum_{i\in V} z_i = 0 \\ & z_i^2 = 1 \qquad\qquad \forall i \in V \end{aligned}$$

Next, write the objectives and constraints of the QP to be sums of quadratic terms $z_i z_j$. For the objective function, consider the following.

$$\sum_{(i,j)\in E} \frac{1}{4}(z_i - z_j)^2 = \sum_{(i,j)\in E} \frac{1}{4}\left(z_i^2 - 2z_i z_j + z_j^2\right) = \sum_{(i,j)\in E} \frac{1}{4}\left(2 - 2z_i z_j\right)$$

where the center substition occurs because we constrain $z_i^2 = 1$ for all $i \in V$. Now observe that it is equivalent to optimize among the following.

$$\begin{aligned} \text{minimize} \quad \sum_{(i,j)\in E} \frac{1}{2} - \frac{1}{2} z_i z_j \qquad & \iff \qquad \text{minimize} \quad \sum_{(i,j)\in E} \frac{1}{2} - \frac{1}{2} z_i z_j \\ & \iff \qquad \text{maximize} \quad \sum_{(i,j)\in E} z_i z_j \\ & \iff \qquad \text{maximize} \quad \sum_{i,j\in V} A_{ij} z_i z_j \end{aligned}$$

where A is the adjacency matrix for G. For the constraint $\sum_{i \in V} z_i = 0$, observe that

$$\sum_{i \in V} z_i = 0 \qquad \iff \qquad \left(\sum_{i \in V} z_i = 0\right)^2 = 0 \qquad \iff \qquad \sum_{i,j \in V} z_i z_j = 0$$

Our final quadratic program is

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j \in V} A_{ij} z_i z_j \\ \text{subject to} \quad & \sum_{i,j \in V} z_i z_j = 0 \\ & z_i^2 = 1 \qquad \forall i \in V \end{aligned}$$

Introduce $v_1, \ldots, v_n \in \mathbb{R}^n$ and replace $z_i z_j$ with $\langle v_i, v_j \rangle$ to derive the vector programming relaxation for min-bisection.

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j \in V} A_{ij} \langle v_i, v_j \rangle \\ \text{subject to} \quad & \sum_{i,j \in V} \langle v_i, v_j \rangle = 0 \\ & \langle v_i, v_i \rangle = 1 \qquad \forall i \in V \\ & v_i \in \mathbb{R}^n \qquad \forall i \in V \end{aligned}$$

Finally, replace $\langle v_i, v_j \rangle$ with X_{ij} and enforce $X \succeq 0$ to derive the semi-definite programming relaxation.

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j \in V} A_{ij} X_{ij} \\ \text{subject to} \quad & \sum_{i,j \in V} X_{ij} = 0 \\ & X_{ii} = 1 \qquad \forall i \in V \\ & X \succeq 0 \end{aligned} \tag{7.1}$$

7.2.2 The Algorithm

Now that we have the SDP relaxation for min-bisection equation 7.1, it is clear what our algorithm should be: solve the SDP and then apply some rounding procedure to extract the vector encoding our partition.

SDP Algorithm
Given $G \sim \mathcal{G}_{n,p,q}$, recover the partition via the following:

1. Solve SDP equation 7.1 for the solution matrix X.
2. Compute the top eigenvector $\boldsymbol{v}$ of X.
3. Return $V_1 = \{i : \boldsymbol{v}_i > 0\}$ and $V_2 = \{i : \boldsymbol{v}_i \leq 0\}$.

To show that this algorithm exactly recovers the planted partitions V_1, V_2 provided a, b are well-separated, we will demonstrate the following theorem.

Theorem 7.1. *There exists a universal constant $c > 0$ such that if $a - b > c\sqrt{\log n}\sqrt{a+b}$, then the SDP relaxation 7.1 possesses a unique optimal solution of $X^* = \boldsymbol{w}\boldsymbol{w}^\top$ where $\boldsymbol{w}_i = 1$ if $i \in V_1$ and otherwise $\boldsymbol{w}_i = -1$*

In particular, theorem 7.1 implies that the algorithm exactly recovers the planted partition. Any solution to the SDP will (approximately due to numerical considerations) be a rank-one product of an indicator vector for the hidden communities because it is the *only* solution! One thing to note is that the SDP algorithm rounds the top eigenvector in the same way that the spectral algorithm did. In some sense, we have significant freedom in how we want to round our solution. Because the SDP is uniquely solvable by $\boldsymbol{w}\boldsymbol{w}^\top$ provided $a - b > c\sqrt{\log n}\sqrt{a+b}$, any reasonable rounding procedure that recovers an integral decomposition of X would work.

7.3 Analyzing the Algorithm

To prove theorem 7.1, we need to complete two tasks: (1) certify that $\boldsymbol{w}\boldsymbol{w}^\top$ is an optimal solution for the semi-definite program 7.1 and (2) show that $\boldsymbol{w}\boldsymbol{w}^\top$ is the unique optimal solution. When we need to certify previously that some matrix was an optimal solution to an SDP, we demonstrated that there was a feasible solution to the SDP's dual that achieves the same objective value. Let's begin by deriving the dual program to SDP 7.1.

7.3.1 A Dual Semi-definite Program

Since our primal program 7.1 is a maximization problem, our dual should be a minimization problem which computes the lowest upper-bound to the objective value of the primal SDP 7.1. First observe that the primal SDP can be rewritten in the following form.

$$\begin{array}{ll} \text{maximize} & \langle A, X\rangle \\ \text{subject to} & \langle J, X\rangle = 0 \\ & \langle E_{ii}, X\rangle = 1 \quad \forall i \in V \\ & X \succeq 0 \end{array} \tag{7.2}$$

where J is the $n \times n$ all-ones matrix and E_{ii} is the indicator on the (i,i)-th entry $[E_{ii}]_{ii} = 1$. Now, rewrite this SDP as an infinite linear program in X by replacing $X \succeq 0$ with the constraints

$$v^\top X v = \langle vv^\top, X\rangle \geq 0 \qquad \Longleftrightarrow \qquad -\langle vv^\top, X\rangle \leq 0 \qquad\qquad \forall v \in \mathbb{R}^n$$

The directions of the inequalities are swapped to get a lower-bound. Next, multiply non-negative multipliers to each constraint.

$$y_0 \;:\; y_0\langle J, X\rangle = 0 \qquad\qquad y_i \;:\; y_i\langle E_{ii}, X\rangle = y_i \qquad\qquad c_v \;:\; -c_v\langle vv^\top, X\rangle \leq 0$$

Summing all the constraints together, we get the following inequality

$$y_0\langle J, X\rangle + \sum_{i\in V} y_i \langle E_{ii}, X\rangle - \sum_{v\in\mathbb{R}^n} c_v \langle vv^\top, X\rangle \leq \sum_{i=1}^n y_i$$
$$\left\langle y_0 J + \mathrm{diag}(y) - \sum_{v\in\mathbb{R}^n} c_v vv^\top, X\right\rangle \leq \sum_{i=1}^n y_i$$

where $\mathrm{diag}(y)$ denotes the diagonal matrix with $y_1, \ldots, y_n$ placed along the diagonal. Now if we constrain $y_0 J + \mathrm{diag}(y) - \sum_{v\in\mathbb{R}^n} c_v vv^\top = A$, we get that the above inequality upper-bounds the primal objective function! Since we want the lowest upper-bound, our dual programs takes the following preliminary form.

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^n y_i \\ \text{subject to} \quad & y_0 J + \mathrm{diag}(y) - \sum_{v\in\mathbb{R}^n} c_v vv^\top = A \\ & y_0, y_i, c_v \geq 0 \qquad\qquad \forall i \in V \ \forall v \in \mathbb{R}^n \end{aligned}$$

But with $c_v \geq 0$, the sum $\sum_{v\in\mathbb{R}^n} c_v vv^\top$ is some matrix Z where $Z \succeq 0$. Thus the first dual constraint becomes $y_0 J + \mathrm{diag}(y) - Z = A$ for some $Z \succeq 0$. Rewriting the preliminary dual program

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^n y_i \\ \text{subject to} \quad & y_0 J + \mathrm{diag}(y) \succeq A \\ & y_0, y_i \geq 0 \qquad\qquad \forall i \in V \end{aligned} \tag{7.3}$$

With the dual to our SDP relaxation for min-bisection, we are now ready to provide a dual certificate for the solution $X^* = \boldsymbol{w}\boldsymbol{w}^\top$.

7.3.2 Proving Optimality

We demonstrate that $X^* = \boldsymbol{w}\boldsymbol{w}^\top$ is an optimal solution by providing a dual certificate or a dual feasible solution that achieves the same objective value as X^*. Recall this is sufficient because weak duality for a maximization problem states that the dual objective always upper-bounds the primal objective. Think about how we constructed dual program 7.3! Now, if our primal solution of interest achieves the same objective as some dual feasible solution, then there can be no other primal solution achieving a greater primal objective value without contradicting weak duality. Consequently, our solution must be an optimal solution.

To begin, what's the primal objective value for our solution of interest $X^* = \boldsymbol{w}\boldsymbol{w}^\top$? Observe that for $i, j \in V$, we have that X^*_{ij} is the following:

$$X^*_{ij} = \begin{cases} 1 & \text{if } i, j \in V_1 \text{ or } i, j \in V_2 \\ -1 & \text{otherwise} \end{cases}$$

Define a_i, b_i to be the internal and external degree of $i \in V$ and notice that the objective value is

$$\langle A, X^* \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} X^*_{ij} = \sum_{i=1}^{n} \left(\sum_{j: i,j \text{ in same partition}} A_{ij} - \sum_{j: i,j \text{ in different partitions}} A_{ij} \right) = \sum_{i=1}^{n} \big(a_i - b_i\big)$$

We just need an assignment of y_0 and y_i's such that $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \big(a_i - b_i\big)$ to certify $X^* = \boldsymbol{w}\boldsymbol{w}^\top$ is optimal. Why not have

$$y_i = a_i - b_i \qquad y_0 = \frac{a+b}{n}$$

where we choose y_0 to ultimately ensure that the assignment of y_i's is dual feasible. To actually show feasibility, we need to demonstrate that it satisfies the constraints of dual 7.3. Readily $y_0, y_i \geq 0$ so the brunt of our work will be in demonstrating $y_0 J + \operatorname{diag}(y) - A \succeq 0$. Let $M = y_0 J + \operatorname{diag}(y) - A$ and note that M is a random matrix since A the adjacency matrix of $G \sim \mathcal{G}_{n,p,q}$ is random.

As we have done previously, we will compare the behavior of M to its expectation $\mathbb{E}[M]$. Let's start by analyzing the spectrum of $\mathbb{E}[M]$ by decomposing it into the following.

$$\mathbb{E}[M] = y_0 J + \mathbb{E}[\operatorname{diag}(y)] - \mathbb{E}[A]$$

Recall from last lecture that $\mathbb{E}[A] = \frac{p+q}{2} J + \frac{p-q}{2} \boldsymbol{w}\boldsymbol{w}^\top$ which parameterized by a and b is

$$\mathbb{E}[A] = \frac{a+b}{n} J + \frac{a-b}{n} \boldsymbol{w}\boldsymbol{w}^\top$$

Meanwhile, a_i and b_i are random variables as well. Thus $\big[\mathbb{E}[\operatorname{diag}(y)]\big]_{ij} = \mathbb{E}[y_i] = \mathbb{E}[a_i] - \mathbb{E}[b_i] = a - b$ when $i = j$ and $\big[\mathbb{E}[\operatorname{diag}(y)]\big]_{ij} = 0$ otherwise. We have that $\mathbb{E}[M]$ is the following.

$$\mathbb{E}[M] = (a-b)I - \left(\frac{a-b}{n}\right) \boldsymbol{w}\boldsymbol{w}^\top$$

Observe that $\boldsymbol{w}$ is an eigenvector of $\mathbb{E}[M]$ with eigenvalue 0 since

$$\mathbb{E}[M]\boldsymbol{w} = (a-b)\boldsymbol{w} - \left(\frac{a-b}{n}\right) \boldsymbol{w}\boldsymbol{w}^\top \boldsymbol{w} = (a-b)\boldsymbol{w} - (a-b)\boldsymbol{w} = 0$$

and for any other eigenvector $\boldsymbol{w}' \perp \boldsymbol{w}$, its corresponding eigenvalue is $a - b$. Now that we understand the spectrum of $\mathbb{E}[M]$, we want to say that the spectrum of M will be relatively close to $\mathbb{E}[M]$. The following concentration statement will be useful.

Lemma 7.2. *Given $G \sim \mathcal{G}_{n,p,q}$, then with high probability*

$$\|M - \mathbb{E}[M]\| \leq O(\sqrt{\log n}\sqrt{a+b})$$

where $\|\cdot\|$ denotes the spectral norm.

This can be proven using the matrix Bernstein inequality, a matrix analog of Chernoff bounds. With this lemma, we are finally ready to show that our assignment of y's is feasible.

Lemma 7.3. *Let y_0, y_i for all $i \in V$ be assigned via the discussion above. Then the following hold*

(1) $M\boldsymbol{w} = 0$ with probability 1.

(2) There is an absolute constant c such that if $a - b > c\sqrt{\log n}\sqrt{a+b}$ then with high probability $x^\top M x > 0$ for all $x \perp \boldsymbol{w}$.

Proof. We first demonstrate claim (1). Observe

$$\boldsymbol{w}^\top M \boldsymbol{w} = y_0 \cdot \boldsymbol{w}^\top J \boldsymbol{w} + \boldsymbol{w}^\top \mathrm{diag}(y)\boldsymbol{w} - \boldsymbol{w}^\top A \boldsymbol{w}$$

Since $\boldsymbol{w}$ is the indicator of a balanced cut, we have that $J\boldsymbol{w} = 0$. Next, we have

$$[\mathrm{diag}(y)\boldsymbol{w}]_i = y_i \boldsymbol{w}_i = \big(a_i - b_i\big)\boldsymbol{w}_i$$

If we can show that $[A\boldsymbol{w}]_i = \big(a_i - b_i\big)\boldsymbol{w}_i$, then claim (1) holds. Indeed

$$[A\boldsymbol{w}]_i = \sum_{j=1}^{n} A_{ij}\boldsymbol{w}_j = \boldsymbol{w}_i \sum_{j=1}^{n} A_{ij}\boldsymbol{w}_i\boldsymbol{w}_j = \boldsymbol{w}_i \sum_{j=1}^{n} A_{ij}X_{ij} = \big(a_i - b_i\big)\boldsymbol{w}_i$$

where the second equality holds as $\boldsymbol{w}_i^2 = 1$. As a remark, certainly $\boldsymbol{w}$ should be in the null space of $\mathbb{E}[M]$ if it is in the null space of M with probability 1! To demonstrate claim (2), consider any vector $x \perp \boldsymbol{w}$ and decompose it as $x = \boldsymbol{w}' + \alpha\boldsymbol{w}$ where $\boldsymbol{w}' \perp \boldsymbol{w}$. We then have that

$$x^\top M x = (\boldsymbol{w}' + \alpha\boldsymbol{w})^\top M(\boldsymbol{w}' + \alpha\boldsymbol{w}) = \boldsymbol{w}'^\top M\boldsymbol{w}' + 2\alpha\boldsymbol{w}'^\top M\boldsymbol{w} + \alpha^2\boldsymbol{w}^\top M\boldsymbol{w} = \boldsymbol{w}'^\top M\boldsymbol{w}'$$

since $\boldsymbol{w}$ is in the null space of M. Meanwhile as $\boldsymbol{w}'$ is orthogonal to $\boldsymbol{w}$, we have that

$$\boldsymbol{w}'^\top \mathbb{E}[M]\boldsymbol{w}' = (a-b)\|\boldsymbol{w}'\|_2^2 - \left(\frac{a-b}{n}\right)\big(\boldsymbol{w}^\top\boldsymbol{w}'\big)^2 = (a-b)\|\boldsymbol{w}'\|_2^2$$

Now by definition of the spectral norm and because $\boldsymbol{w}' \neq 0$

$$\frac{\boldsymbol{w}'^\top(\mathbb{E}[M] - M)\boldsymbol{w}'}{\|\boldsymbol{w}'\|_2^2} \leq \|M - \mathbb{E}[M]\| \qquad \iff \qquad \boldsymbol{w}'^\top M\boldsymbol{w}' \geq \boldsymbol{w}'^\top\mathbb{E}[M]\boldsymbol{w}' - \|M - \mathbb{E}[M]\| \cdot \|\boldsymbol{w}'\|_2^2$$

But this is equivalent to

$$\boldsymbol{w}'^\top M\boldsymbol{w}' \geq \big((a-b) - \|M - \mathbb{E}[M]\|\big) \cdot \|\boldsymbol{w}'\|_2^2$$

However lemma 7.2 states that with high probability $\|M - \mathbb{E}[M]\| \geq O(\sqrt{\log n}\sqrt{a+b})$. There must be a sufficiently large constant c such that if $a - b \geq c\sqrt{\log n}\sqrt{a+b}$, we have that

$$x^\top M x = \boldsymbol{w}'^\top M\boldsymbol{w}' \geq \big((a-b) - \|M - \mathbb{E}[M]\|\big) \cdot \|\boldsymbol{w}'\|_2^2 > 0 \qquad \square$$

Lemma 7.3 implies that $x^\top M x = 0$ for all $x \in \mathrm{span}(\boldsymbol{w})$ and $x^\top M x > 0$ otherwise. In all cases $x^\top M x \geq 0$ hence $M \succeq 0$ thus our assignment is feasible and $X^* = \boldsymbol{w}\boldsymbol{w}^\top$ is optimal.

7.3.3 Proving Uniqueness

Our statement of this lemma 7.3 is actually stronger than what we needed to show optimality. We required only $x^\top Mx \geq 0$, but ended with a more granular characterization of when $x^\top Mx > 0$ and when $x^\top Mx = 0$. Indeed, this fact is what enables us to demonstrate that X^* is the unique solution to SDP 7.1. We are now ready to prove our main theorem 7.1.

Proof of theorem 7.1. For any primal feasible X and dual feasible $y_0, y_1, \dots, y_n$, the dual constraint $y_0 J + \operatorname{diag}(y) \succeq A$ is satisfied. This dictates $x^\top \big(y_0 J + \operatorname{diag}(y)\big)x \geq x^\top Ax$ for any $x \in \mathbb{R}^n$. We thus have

$$\langle A, X\rangle \leq \langle \operatorname{diag}(y) + y_0 J, X\rangle = \langle \operatorname{diag}(y), X\rangle + y_0\langle J, X\rangle = \sum_{i\in V} y_i X_{ii} + y_0 \sum_{i,j\in V} X_{ij}$$

However, the primal constraints ensure that $X_{ii} = 1$ for every $i \in V$ and $\sum_{i,j\in V} X_{ij} = 0$. This means

$$\langle A, X\rangle \leq \sum_{i\in V} y_i$$

Now suppose if X is optimal, then the above inequality is tight. Recall we had just shown for $X^* = \boldsymbol{w}\boldsymbol{w}^\top$ that $\langle A, X^*\rangle = \sum_{i\in V} y_i = \sum_{i\in V} \big(a_i - b_i\big)$. This implies

$$0 = \langle \operatorname{diag}(y) + y_0 J, X\rangle - \langle A, X\rangle = \langle M, X\rangle$$

By lemma 7.3, we have that $\boldsymbol{w}$ is in the null space of M readily implying that $\langle M, X^*\rangle = 0$. We now claim that X^* is the only such matrix that satisfies $\langle M, X\rangle = 0$. Take any feasible solution X and write its spectral decomposition. We then have that

$$\langle M, X\rangle = \left\langle M, \sum_i \lambda_i z_i z_i^\top \right\rangle = \sum_i \lambda_i \langle M, z_i z_i^\top\rangle = \sum_i \lambda_i z_i^\top M z_i$$

There can only be two cases, $z_i \in \operatorname{span}(\boldsymbol{w})$ or $z_i \perp \boldsymbol{w}$. In the former, we have that $z_i^\top M z_i = 0$. However as $a - b \geq c\sqrt{\log n}\sqrt{a+b}$, lemma 7.3 states with high probability that $z_i^\top M z_i > 0$ in the former. Because $X \succeq 0$, it must be that $\lambda_i > 0$ and if $z_i^\top M z_i > 0$ implying $\langle M, X\rangle \neq 0$! It must be that $z_i \in \operatorname{span}(\boldsymbol{w})$. Consequently, $X = X^*$ is the unique solution to SDP 7.1 as required. □

As a final remark, our proof actually hinges on an example of complementary slackness. Roughly, complementary slackness states that if the primal optimal solution has the same objective value as the dual optimal solution, then the inner product of a particular primal constraint value and its corresponding dual multiplier is 0. Imagine taking the dual of dual SDP 7.3. We would then recover the primal SDP 7.1 with matrix X as a multiplier for constraint $y_0 J + \operatorname{diag}(y) - A \succeq 0$. Since lemma 7.3 demonstrates that the primal and dual optimal solutions have the same objective value, complementary slackness would immediately imply $\langle M, X\rangle = 0$.