# MAT 585: Exact Recovery of the Semidefinite Relaxation for Stochastic Block Model

Afonso S. Bandeira

April 7, 2015

Today we consider a semidefinite programming relaxation algorithm for SBM and derive conditions for exact recovery. The main ingredient for the proof will be duality theory.

## 1 Quick Recap

- $n$ nodes, two clusters of equal size $\frac{n}{2}$, $p$ and $q$ probabilities of intra-cluster and inter-cluster edges.
- MLE approach:

$$\begin{aligned} \max & \sum_{i,j} A_{ij} x_i x_j \\ \text{s.t. } & x_i = \pm 1, \forall_i \\ & \sum_j x_j = 0, \end{aligned} \tag{1}$$

  where $A$ is the adjacency matrix of the graph $G$

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E(G) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

- MLE recovery condition

$$\frac{\alpha + \beta}{2} - \sqrt{\alpha\beta} > 1 \tag{3}$$

## 2 The algorithm

Note that if we remove the constraint that $\sum_j x_j = 0$ in (1) then the optimal solution becomes $x = \mathbf{1}$. Let us define $B = 2A - (\mathbf{1}\mathbf{1}^T - I)$, meaning that

$$B_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } (i,j) \in E(G) \\ -1 & \text{otherwise} \end{cases} \tag{4}$$

It is clear that the problem

$$\begin{aligned} \max \ & \sum_{i,j} B_{ij} x_i x_j \\ \text{s.t. } & x_i = \pm 1, \forall_i \\ & \sum_j x_j = 0 \end{aligned} \tag{5}$$

has the same solution as (1). However, when the constraint is dropped,

$$\begin{aligned} \max \ & \sum_{i,j} B_{ij} x_i x_j \\ \text{s.t. } & x_i = \pm 1, \forall_i, \end{aligned} \tag{6}$$

$x = \mathbf{1}$ is no longer an optimal solution. Intuitively, there is enough "$-1$" contribution to discourage unbalanced partitions. In fact, (6) is the problem we'll set ourselves to solve.

Unfortunately (6) is in general NP-hard (one can encode, for example, `Max-Cut` by picking an appropriate $B$). We will relax it to an easier problem by the same technique used to approximate the `Max-Cut` problem in the previous section (this technique is often known as *matrix lifting*). If we write $X = xx^T$ then we can formulate the objective of (6) as

$$\sum_{i,j} B_{ij} x_i x_j = x^T B x = \mathrm{Tr}(x^T B x) = \mathrm{Tr}(B x x^T) = \mathrm{Tr}(BX)$$

Also, the condition $x_i = \pm 1$ implies $X_{ii} = x_i^2 = 1$. This means that (6) is equivalent to

$$\begin{aligned} \max \quad & \text{Tr}(BX) \\ \text{s.t.} \quad & X_{ii} = 1, \forall_i \\ & X = xx^T \text{ for some } x \in \mathbb{R}^n. \end{aligned} \tag{7}$$

The fact that $X = xx^T$ for some $x \in \mathbb{R}^n$ is equivalent to $\text{rank}(X) = 1$ and $X \succeq 0$.This means that (6) is equivalent to

$$\begin{aligned} \max \quad & \text{Tr}(BX) \\ \text{s.t.} \quad & X_{ii} = 1, \forall_i \\ & X \succeq 0 \\ & \text{rank}(X) = 1. \end{aligned} \tag{8}$$

We now relax the problem and, as before, remove the non-convex rank constraint

$$\begin{aligned} \max \quad & \text{Tr}(BX) \\ \text{s.t.} \quad & X_{ii} = 1, \forall_i \\ & X \succeq 0. \end{aligned} \tag{9}$$

This is an SDP that can be solved (up to arbitrary precision) in polynomial time [2].

Since we removed the rank constraint, the solution to (9) is no longer guaranteed to be of rank-1. We will take a different approach from the one we used before to obtain an approximation ratio for `Max-Cut`, which was a worst-case approximation ratio guarantee. What we will show is that, for some values of $\alpha$ and $\beta$, with high probability, the solution to (9) not only satisfies the rank constraint but it coincides with $X = gg^T$ where $g$ corresponds to the true partition. After $X$ is computed, $g$ is simply obtained as its leading eigenvector.

## 3 The analysis

Without loss of generality, we can assume that $g = (1, \ldots, 1, -1, \ldots, -1)^T$, meaning that the true partition corresponds to the first $\frac{n}{2}$ nodes on one side and the other $\frac{n}{2}$ on the other.

### 3.1 Some preliminary definitions

Recall that the degree matrix $D$ of a graph $G$ is a diagonal matrix where each diagonal coefficient $D_{ii}$ corresponds to the number of neighbours of vertex $i$ and that $\lambda_2(M)$ is the second smallest eigenvalue of a symmetric matrix $M$.

**Definition 1** *Let $\mathcal{G}_+$ (resp. $\mathcal{G}_-$) be the subgraph of $G$ that includes the edges that link two nodes in the same community (resp. in different communities) and $A$ the adjacency matrix of $G$. We denote by $D_{\mathcal{G}}^+$ (resp. $D_{\mathcal{G}}^-$) the degree matrix of $\mathcal{G}_+$ (resp. $\mathcal{G}_-$) and define the Stochastic Block Model Laplacian to be*

$$L_{SBM} = D_{\mathcal{G}}^+ - D_{\mathcal{G}}^- - A$$

### 3.2 Convex Duality

A standard technique to show that a candidate solution is the optimal one for a convex problem is to use convex duality.

We will describe duality with a game theoretical intuition in mind. The idea will be to rewrite (9) without imposing constraints on $X$ but rather have the constraints be implicitly enforced. Consider the following optimization problem.

$$\max_{X} \min_{\substack{Z,\, Q \\ Z \text{ is diagonal} \\ Q \succeq 0}} \operatorname{Tr}(BX) + \operatorname{Tr}(QX) + \operatorname{Tr}\left(Z\left(I_{n\times n} - X\right)\right) \tag{10}$$

Let us give it a game theoretical interpretation. Suppose there is a primal player (picking $X$) whose objective is to maximize the objective and a dual player, picking $Z$ and $Q$ after seeing $X$, trying to make the objective as small as possible. If the primal player does not pick $X$ satistying the constraints of (9) then we claim that the dual player is capable of driving the objective to $-\infty$. Indeed, if there is an $i$ for which $X_{ii} \neq 1$ then the dual player can simply pick $Z_{ii} = -c\frac{1}{1-X_{ii}}$ and make the objective as small as desired by taking large enough $c$. Similarly, if $X$ is not positive semidefinite, then the dual player can take $Q = cvv^T$ where $v$ is such that $v^T X v < 0$. If, on the other hand, $X$ satisfies the constraints of (9) then

$$\operatorname{Tr}(BX) \leq \min_{\substack{Z,\, Q \\ Z \text{ is diagonal} \\ Q \succeq 0}} \operatorname{Tr}(BX) + \operatorname{Tr}(QX) + \operatorname{Tr}\left(Z\left(I_{n\times n} - X\right)\right),$$

since equality can be achieve if, for example, the dual player picks $Q = 0_{n\times n}$, then it is clear that the values of (9) and (10) are the same:

$$\max_{\substack{X,\\ X_{ii}\ \forall_i\\ X\succeq 0}} \operatorname{Tr}(BX) = \max_{X} \min_{\substack{Z,\ Q\\ Z \text{ is diagonal}\\ Q\succeq 0}} \operatorname{Tr}(BX) + \operatorname{Tr}(QX) + \operatorname{Tr}\left(Z\left(I_{n\times n} - X\right)\right)$$

With this game theoretical intuition in mind, it is clear that if we change the "rules of the game" and have the dual player decide her variables before the primal player (meaning that the primal player can pick $X$ knowing the values of $Z$ and $Q$) then it is clear that the objective can only increase, which means that:

$$\max_{\substack{X,\\ X_{ii}\ \forall_i\\ X\succeq 0}} \operatorname{Tr}(BX) \le \min_{\substack{Z,\ Q\\ Z \text{ is diagonal}\\ Q\succeq 0}} \max_{X} \operatorname{Tr}(BX) + \operatorname{Tr}(QX) + \operatorname{Tr}\left(Z\left(I_{n\times n} - X\right)\right).$$

Note that we can rewrite

$$\operatorname{Tr}(BX) + \operatorname{Tr}(QX) + \operatorname{Tr}\left(Z\left(I_{n\times n} - X\right)\right) = \operatorname{Tr}\left(\left(B + Q - Z\right)X\right) + \operatorname{Tr}(Z).$$

When playing:

$$\min_{\substack{Z,\ Q\\ Z \text{ is diagonal}\\ Q\succeq 0}} \max_{X} \operatorname{Tr}\left(\left(B + Q - Z\right)X\right) + \operatorname{Tr}(Z),$$

if the dual player does not set $B + Q - Z = 0_{n\times n}$ then the primal player can drive the objective value to $+\infty$, this means that the dual player is forced to chose $Q = Z - B$ and so we can write

$$\min_{\substack{Z,\ Q\\ Z \text{ is diagonal}\\ Q\succeq 0}} \max_{X} \operatorname{Tr}\left(\left(B + Q - Z\right)X\right) + \operatorname{Tr}(Z) = \min_{\substack{Z,\\ Z \text{ is diagonal}\\ Z-B\succeq 0}} \max_{X} \operatorname{Tr}(Z),$$

which clearly does not depend on the choices of the primal player. This means that

$$\max_{\substack{X,\\ X_{ii}\ \forall_i\\ X\succeq 0}} \operatorname{Tr}(BX) \le \min_{\substack{Z,\\ Z \text{ is diagonal}\\ Z-B\succeq 0}} \operatorname{Tr}(Z).$$

This is known as weak duality (strong duality says that, under some conditions the two optimal values actually match, see, for example, [2]). Also, the problem

$$\begin{array}{rc} \min & \mathrm{Tr}(Z) \\ \text{s.t.} & Z \text{ is diagonal} \\ & Z - B \succeq 0 \end{array} \tag{11}$$

is called the dual problem of (9).

The derivation above explains why the objective value of the dual is always greater or equal to the primal. Nevertheless, there is a much simpler proof (although not as enlightening): let $X, Z$ be respectively a feasible point of (9) and (11). Since $Z$ is diagonal and $X_{ii} = 1$ then $\mathrm{Tr}(ZX) = \mathrm{Tr}(Z)$. Also, $Z - B \succeq 0$ and $X \succeq 0$, therefore $\mathrm{Tr}[(Z - B)X] \geq 0$. Altogether,

$$\mathrm{Tr}(Z) - \mathrm{Tr}(BX) = \mathrm{Tr}[(Z - B)X] \geq 0,$$

as stated.

Recall that we want to show that $gg^T$ is the optimal solution of (9). Then, if we find $Z$ diagonal, such that $Z - B \succeq 0$ and

$$\mathrm{Tr}[(Z-B)gg^T] = 0, \quad \text{(this condition is known as complementary slackness)}$$

then $X = gg^T$ must be an optimal solution of (9). To ensure that $gg^T$ is the unique solution we just have to ensure that the nullspace of $Z - B$ only has dimension 1 (which corresponds to multiples of $g$). Essentially, if this is the case, then for any other possible solution $X$ one could not satisfy complementary slackness.

This means that if we can find $Z$ with the following properties:

1. $Z$ is diagonal
2. $\mathrm{Tr}[(Z - B)gg^T] = 0$
3. $Z - B \succeq 0$
4. $\lambda_2(Z - B) > 0$,

then $gg^T$ is the unique optima of (9) and so recovery of the true partition is possible (with an efficient algorithm).

$Z$ is known as the dual certificate, or dual witness.

### 3.3 Building the dual certificate

The idea to build $Z$ is to construct it to satisfy properties (1) and (2) and try to show that it satisfies (3) and (4) using concentration.

If indeed $Z - B \succeq 0$ then (2) becomes equivalent to $(Z - B)g = 0$. This means that we need to construct $Z$ such that $Z_{ii} = \frac{1}{g_i} B[i, :]g$. Since $B = 2A - (\mathbf{1}\mathbf{1}^T - I)$ we have

$$Z_{ii} = \frac{1}{g_i}(2A - (\mathbf{1}\mathbf{1}^T - I))[i, :]g = 2\frac{1}{g_i}(Ag)_i + 1,$$

meaning that

$$Z = 2(D^+_{\mathcal{G}} - D^-_{\mathcal{G}}) + I$$

is our guess for the dual witness. As a result

$$Z - B = 2(D^+_{\mathcal{G}} - D^-_{\mathcal{G}}) - I - \left[2A - (\mathbf{1}\mathbf{1}^T - I)\right] = 2L_{SBM} + 11^T$$

It trivially follows (by construction) that

$$(Z - B)g = 0.$$

Therefore

**Lemma 2** *If*

$$\lambda_2(2L_{SBM} + 11^T) > 0, \tag{12}$$

*then the relaxation recovers the true partition.*

Note that $2L_{SBM} + 11^T$ is a random matrix and so this boils down to "an exercise" in random matrix theory.

### 3.4 Matrix Concentration

Clearly,

$$\mathbb{E}\left[2L_{SBM} + 11^T\right] = 2\mathbb{E}L_{SBM} + 11^T = 2\mathbb{E}D^+_{\mathcal{G}} - 2\mathbb{E}D^-_{\mathcal{G}} - 2\mathbb{E}A + 11^T,$$

and $\mathbb{E}D^+_{\mathcal{G}} = \frac{n}{2}\frac{\alpha \log(n)}{n}I$, $\mathbb{E}D^-_{\mathcal{G}} = \frac{n}{2}\frac{\beta \log(n)}{n}I$, and $\mathbb{E}A$ is a matrix such with 4 $\frac{n}{2} \times \frac{n}{2}$ blocks where the diagonal blocks have $\frac{\alpha \log(n)}{n}$ and the off-diagonal blocks have $\frac{\beta \log(n)}{n}$. We can write this as $\mathbb{E}A = \frac{1}{2}\left(\frac{\alpha \log(n)}{n} + \frac{\beta \log(n)}{n}\right) 11^T + \frac{1}{2}\left(\frac{\alpha \log(n)}{n} - \frac{\beta \log(n)}{n}\right) gg^T$

This means that

$$\mathbb{E}\left[2L_{SBM} + 11^T\right] = \left((\alpha-\beta)\log n\right) I + \left(1 - (\alpha+\beta)\frac{\log n}{n}\right) 11^T - (\alpha-\beta)\frac{\log n}{n} gg^T.$$

Since $2L_{SBM}g = 0$ we can ignore what happens in the span of $g$ and it is not hard to see that

$$\lambda_2\left[\left((\alpha-\beta)\log n\right) I + \left(1 - (\alpha+\beta)\frac{\log n}{n}\right) 11^T - (\alpha-\beta)\frac{\log n}{n} gg^T\right] = (\alpha-\beta)\log n.$$

This means that it is enough to show that

$$\|L_{SBM} - \mathbb{E}\left[L_{SBM}\right]\| < \frac{\alpha-\beta}{2}\log n, \tag{13}$$

which is a large deviations inequality. ($\|\cdot\|$ denotes operator norm)

We will not go into the details here but the idea is to write $L_{SBM} - \mathbb{E}\left[L_{SBM}\right]$ as a sum of independent random matrices and use Matrix Bernstein inequality (as in [1]). In fact, using this strategy one can show that, with high probability, 13 holds as long as

$$(\alpha-\beta)^2 > 8(\alpha+\beta) + \frac{8}{3}(\alpha-\beta). \tag{14}$$

### 3.5 Comparison with phase transition

To compare (14) with (3) we note that the latter can be rewritten as

$$(\alpha-\beta)^2 > 4(\alpha+\beta) - 4 \text{ and } \alpha+\beta > 2$$

and so the relaxation achieves exact recovery almost at the threshold, essentially only suboptimal by a factor of 2.

## References


[1] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–424, 2012.

[2] L. Vanderberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.