Lecture 7: Spectral methods for SBM and an introduction to Semidefinite Programming

In which we study a spectral algorithm that performs community detection in the stochastic block model, and we begin to work with semidefinite programming.

1 Spectral Community Detection

Given a graph $G = (V, E)$ sampled from the $SBM_{n,p,q}$ distribution, we consider the following algorithm:

- Let A be the adjacency matrix of G, and compute an eigenvector x of the largest eigenvalue of the matrix $A - \frac{p+q}{2} J$.
- Output the partition of V defined as $(\{v : x_v < 0\}, \{v : x_v \geq 0\})$

We will briefly sketch an analysis of the algorithm.

When we introduced the stochastic block model, we noted that, for a fixed partition, the expectation of A has the rank-two decomposition

$$\mathbb{E}\, A = \left(\frac{p+q}{2} \right) J + \frac{p-q}{2} \left(\begin{array}{c|c} \mathbf{1} & -\mathbf{1} \\ \hline -\mathbf{1} & \mathbf{1} \end{array} \right)$$

(Strictly speaking, the above decomposition holds for $\mathbb{E}\, A + pI$, but the term pI of spectral norm p will be dominated in subsequent computations by error terms of much bigger spectral norm). If A is such that

$$||A - \mathbb{E}\, A|| \leq \epsilon \cdot \frac{(p-q)}{2} \cdot n \tag{1}$$

then

$$\left\| \left(A - \left(\frac{p+q}{2} \right) J \right) - \frac{p-q}{2} \left(\begin{array}{c|c} \mathbf{1} & -\mathbf{1} \\ \hline -\mathbf{1} & \mathbf{1} \end{array} \right) \right\| \leq \epsilon \cdot \frac{(p-q)}{2} \cdot n$$

and the Davis-Kahan Theorem implies that if χ is the ± 1 indicator vector of the partition, that is, the vector such that $\chi \chi^T = \left(\begin{array}{c|c} \mathbf{1} & -\mathbf{1} \\ \hline -\mathbf{1} & \mathbf{1} \end{array} \right)$ and if x is an eigenvector of the largest eigenvalue of $A - \frac{p+q}{2} J$ then

$$|\sin(x, \chi)| \leq \frac{\epsilon}{1-\epsilon}$$

and then, by scaling x so that its inner product with χ is nonnegative and $||x||^2 = n$ we have

$$\langle x, \chi \rangle = ||x|| \cdot ||\chi|| \cdot \cos(x, \chi) = n \cdot \sqrt{1 - \sin^2(x, \chi)} \geq n \cdot \sqrt{1 - O(\epsilon^2)} \geq 1 - n \cdot O(\epsilon^2)$$

and so

$$||x - \chi||^2 = 2n - 2\langle x, \chi \rangle \leq n \cdot O(\epsilon^2)$$

which implies that the partition $\{v : x_v < 0\}, \{v : x_v \geq 0\}$ differs from the partition $\{v : \chi_v = -1\}, \{v : \chi_v = +1\}$ in at most $O(\epsilon^2 n)$ vertices, since each misclassified vertex adds at least 1 to the summation

$$\sum_v (x_v - \chi_v)^2 = ||x - \chi||^2 \leq O(\epsilon^2 n)$$

Multiplying x by a scalar does not change the partition $\{v : x_v < 0\}, \{v : x_v \geq 0\}$, so the error bound holds for every scalar multiple of x and hence for every eigenvector of the largest eigenvalue of $A - \frac{p+q}{2} J$.

It remains to understand under what conditions the spectral concentration bound (1) holds.

Matrix Chernoff bounds imply that, for every p, q, we have

$$||A - \mathbb{E}\, A|| \leq O(\sqrt{(p+q) n \log n})$$

and so the condition $(p-q)n >> \sqrt{(p+q) n \log n}$ suffices for approximate reconstruction.

It is possible to prove that, if $p + q >> \frac{\log n}{n}$ then we have the stronger bound

$$||A - \mathbb{E}\, A|| \leq O(\sqrt{(p+q) n})$$

and, in that regime, the condition $(p-q)n >> \sqrt{(p+q)n}$ suffices for approximate reconstruction.

In the regime in which p and q are $O(1/n)$, meaning that the graph has constant average degree, it is known that condition (1) is false with high probability.

2 Semidefinite Programming

We will now introduce the technique of semidefinite programming. Semidefinite programming is a general class of convex optimization problems, and it generalizes both eigenvalue computations and linear programming. Like linear programming, it has a theory of duality that helps reason about the optimality of solutions. Like spectral techniques, it allows to reason about the phenomenon of the random graphs being "concentrated" around their expectation.

For the sake of this course, it will suffice to introduce a couple of basic concepts and results.

We say that a real symmetric matrix M is Positive Semidefinite, which we abbreviate as PSD and write $M \succeq \mathbf{0}$, if all the eigenvalues of M are non-negative. We will use the partial order $A \succeq B$ among real symmetric matrices which holds iff $A - B \succeq \mathbf{0}$.

It follows from the variational characterization of eigenvalues of real symmetric matrices that a matrix M is PSD if and only if

$$\forall x \in \mathbb{R}^n : \quad x^T M x \geq 0$$

Another useful characterization of PSD matrices is that a matrix is PSD if and only if there are n vectors $x^{(1)}, \ldots, x^{(n)}$ such that

$$\forall i, j : \quad M_{i,j} = \langle x^{(i)}, x^{(j)} \rangle$$

When the above relation holds, we say that M is the *Gram matrix* of the vectors $x^{(1)}, \ldots, x^{(n)}$ and the vectors are called a *Cholesky decomposition* of the matrix. Note that the vectors need not be n-dimensional, and need not be unique.

To prove that this is a characterization, we see that if M is the Gram matrix of $x^{(1)}, \ldots, x^{(n)}$, then the quadratic form of M is a sum of squares and hence non-negative, meaning that M is PSD:

$$\forall y \in \mathbb{R}^n : \quad y^T M y = \sum_{i,j} M_{i,j} y_i y_j = \sum_{i,j} \sum_k x_k^{(i)} x_k^{(j)} y_i y_j = \sum_k \left(\sum_i y_i x_i^{(k)} \right)^2 \geq 0$$

For the other direction, if M is PSD and if $\lambda_1, \cdots, \lambda_n$ is the sequence of non-negative eigenvalues of M and $v^{(1)}, \ldots, v^{(n)}$ the corresponding sequence of orthnormal eigenvectors, then we can write

$$M = \sum_k \lambda_k v^{(k)} \left(v^{(k)} \right)^T$$

that is,

$$M_{i,j} = \sum_k \lambda_k v_i^{(k)} v_j^{(k)}$$

and we can verify that the collection of vectors $x^{(1)}, \ldots, x^{(n)}$ defined as

$$x_k^{(i)} = \sqrt{\lambda_k} v_i^{(k)}$$

is a Cholesky decomposition of M, because

$$\langle x^{(i)}, x^{(j)} \rangle = \sum_k x_k^{(i)} x_k^{(j)} = \sum_k \lambda_k v_i^{(k)} v_j^{(k)} = M_{i,j}$$

From these properties, we conclude that the set $\{M \in \mathbb{R}^{n \times n} : M \succeq \mathbf{0}\}$ of PSD matrices is convex, because if A and B are PSD matrices and $0 \leq \lambda \leq 1$ we have

$$\forall x : \quad x^T(\lambda A + (1-\lambda)A)x = \lambda x^T A x + (1-\lambda) x^T B x \geq 0$$

so that $\lambda A + (1-\lambda)B$ is also PSD.

Semidefinite programming (SDP) is the class of optimization problems in which the unknowns are the entries of a PSD matrix, and we want to optimize a linear functions of such unknowns under linear constraints. A generic form of a maximization semidefinite program is

$$\begin{aligned} & \max \sum_{i,j} c_{i,j} X_{i,j} \\ & \text{s.t.} \\ & \qquad \sum_{i,j} a_{i,j}^{(1)} X_{i,j} \leq b_1 \\ & \qquad \vdots \\ & \qquad \sum_{i,j} a_{i,j}^{(m)} X_{i,j} \leq b_m \\ & \qquad X \succeq \mathbf{0} \end{aligned}$$

where the coefficients $c_{i,j}$, $a_{i,j}^{(k)}$ and b_j are given. It is possible to define the problem as a minimization problem, and to have equality constraints, although such variants can be equivalently reduced to the normal form above.

Interior point algorithms for convex optimization apply to semidefinite programming, and can find, in polynomial time, a solution that is exponentially close to an optimal solution.

For notational convenience, we introduce the *Frobenious inner product* between matrices, defined as

$$A \bullet B := \sum_{i,j} A_{i,j} B_{i,j}$$

With this notation, given a matrix C of cost coefficients, and matrices $A^{(1)}, \ldots, A^{(m)}$ and scalars $b_1, \ldots, b_m$, an SDP in maximization normal form can be written as

$$
\begin{aligned}
& \max C \bullet X \\
& \quad \text{s.t.} \\
& \qquad A^{(k)} \bullet X \leq b_k \qquad\qquad k = 1, \ldots, m \\
& \qquad X \succeq \mathbf{0}
\end{aligned}
$$

Given the characterization of PSD matrices as Gram matrices of a collection of vectors, a generic SDP can also be written as the problem in which the variables are a collection $x^{(1)}, \ldots, x^{(n)}$ of vectors and the optimization problem has the form

$$
\begin{aligned}
& \max \sum_{i,j} c_{i,j} \langle x^{(i)}, x^{(j)} \rangle \\
& \quad \text{s.t.} \\
& \qquad \sum_{i,j} a^{(1)}_{i,j} \langle x^{(i)}, x^{(j)} \rangle \leq b_1 \\
& \qquad \vdots \\
& \qquad \sum_{i,j} a^{(m)}_{i,j} \langle x^{(i)}, x^{(j)} \rangle \leq b_m \\
& \qquad x^{(i)} \in \mathbb{R}^D \qquad\qquad i = 1, \ldots, n \\
& \qquad D > 0
\end{aligned}
$$

In the special case $D = 1$ in which the vectors are 1-dimensional, semidefinite programming becomes *quadratic programming*, the family of optimization problems in which the unknowns are real variables and the cost function and the constraints are homogeneous quadratic polynomials. Thus, semidefinite programming provides a generic way of constructing polynomial-time solvable convex relaxations of quadratic programming problems (the relaxation in SDP is that the dimension of the vectors is arbitrary), whereas quadratic programming is a family of non-convex optimization problems that do not admit polynomial time solvers (unless P=NP).

Both the max clique problem and the minimum balanced cut problem admit exact formulation as quadratic programming problems, and hence admit polynomial time solvable convex SDP relaxations. In the next few lectures, we will study the average-case behavior of such relaxations given graphs sampled from the planted clique distribution and the stochastic block model.