Lecture 11

1 Markov chains and Cheeger's inequality

Sampling a random element from a set $S \subset \{0,1\}^n$ is a fundamental computational problem that arises in a number of contexts. The sampling problem is closely related to a seemingly different computational task - approximating the cardinality of S within a multiplicative factor of $1+\epsilon$. Here are some examples of these tasks:

(i) The elements of S are grid points inside the convex body $P \in \mathbb{R}^d$. A special case is the problem of numerically integrating a convex function of d variables.

The convex body P might be specified by a collection of linear inequalities. There is a natural graph associated with S: the neighbors of a vertex v are the grid points nearest to it. If a neighboring grid point does not satisfy an inequality defining P then a self loop is added to v. The graph is $2d$ regular, and it is easy to generate the neighbors of v by looking at the nearest grid points and checking whether they belong to P. We therefore have a graph $G(S,E)$ defined on an exponentially large set S, together with an efficient procedure to reconstruct the edges of G locally.

A natural algorithm for picking a random element of S is to start from a particular vertex of G and repeatedly move to a random neighbor. Stop after t steps and output the current vertex. The big question is this: is it possible that the output vertex is close to random after polynomially many steps t, even though the size of the graph is exponentially large? This is the question we will address using the connections between sparse cuts and spectral gaps.

(ii) Given a graph $G(V,E)$, consider the problem of finding a random spanning tree in the graph. The number of spanning trees in G can be counted exactly using the properties of electrical networks. The set S of spanning trees of G can be exponentially large in $n = |V|$. Once again, we may define a natural graph associated with S where two spanning trees are adjacent if they differ in exactly two edges. The graph is connected and given a spanning tree $s \in S$, all the neighbors of s can be found efficiently.

(iii) More generally, the elements of the set S may have positive weights associated with them. The sampling problem now asks for an element of S sampled with probability proportional to its weight. A fundamental model in statistical physics is a system of particles located on a two dimensional (or three dimensional) grid. Assume that each particle can be in state ± 1. If there are n particles, a configuration of the system is an element of $\{+1,-1\}^n$. The energy of the system is determined by nearest neighbor interactions - for example, a simple rule might say $E = \sum_{(i,j)} -\sigma_i\sigma_j$. This is the so-called ferromagnetic regime, where configurations with aligned spins have low energy and are favored.

The equilibrium distribution at a given temperature T is the Gibbs distribution, where a configuration having energy E occurs with probability proportional to $e^{-\frac{E}{kT}}$. Statistical physics tells us that the properties of the material can be determined by sampling from the Gibbs distribution.

The metropolis process is a natural Markov chain for sampling from weighed sets. The probability of making a transition from $x \to y$ is proportional to $\min(1, \frac{w(y)}{w(x)})$ for the Metropolis process. For sampling from the Gibbs distribution we choose a random particle and flip its state with probability $\min(1, e^{-\frac{\Delta E}{kT}})$ where ΔE is the change in energy due caused due to flipping.

1.1 Markov Chains and random walks

A random walk on a graph G is following process: a particle starts from a fixed vertex (or a distribution over vertices) and moves to a random neighbor at every step. We are interested in the probability distribution that describes its position after t steps. If G is d-regular, connected and non-bipartite then as t tends to infinity, the probability distribution tends to the uniform distribution. In the case of a bipartite graph, if the particle starts on the left side of the bipartition, it will be on the right side at odd numbered steps and left side at even numbered steps, and therefore there is no limiting distribution.

The random walk on G can be described by its normalized adjacency matrix M as follows: let v_i be the n-dimensional vector, whose j-th component is the probability that the particle is at vertex j after i steps of the random walk. It is easy to see that $v_{i+1} = Mv_i$ as the action of M averages over neighbors.

Wlog we assume that the walk starts at a fixed vertex, the starting distribution can be taken to be e_1. The behavior of the walk can be understood by decomposing $e_1 = \sum \alpha_i v_i$ in the spectral basis for M, we note that $\alpha_1 = \frac{1}{\sqrt{n}}$. The matrix M is diagonal in the spectral basis so after k steps the walk evolves to,

$$M^k e_1 = \frac{1}{\sqrt{n}} v_1 + \sum_{i>1} \lambda_i^k \alpha_i v_i \tag{1}$$

If all the eigenvalues of M are positive, then after the random walk converges to the uniform distribution $\frac{1}{\sqrt{n}} v_1$ and the rate of convergence is controlled by the spectral gap $\mu = 1 - \lambda_2$. However, all we know is that $-1 \leq \lambda_i \leq 1$ for eigenvalues of M, negative eigenvalues are a real issue, for example $\lambda_{min} = -1$ if the graph is bipartite and in this case the random walk does not converge to the uniform distribution. If there are eigenvalues close to -1 the rate of convergence is not controlled by the spectral gap.

These issues can be overcome if we modify the walk slightly and consider the lazy random walk $(M + I)/2$ instead. The eigenvalues of $(M + I)/2$ are $\frac{1+\lambda_i}{2}$ and therefore lie in the interval $[0, 1]$, the spectral gap for the lazy random walk is $\frac{1-\lambda_2}{2} = \mu/2$.

The speed of convergence to the uniform distribution π is measured in terms of the ℓ_1 distance $d_1(v_i, \pi) = \sum_i |v_i - \pi_i|$, the Markov chain is said to be rapidly mixing if $d_1(v_i, \pi) \leq \epsilon$ for $i = poly(\log N, \frac{1}{\epsilon})$. It is convenient to bound the squared ℓ_2 distance $d_2(v_i, \pi) = \sum_i (v_i - \pi_i)^2$ in terms of the spectral gap, this translates to a bound on the ℓ_1 distance via the Cauchy Schwarz inequality $\sqrt{N} d_2(v_i, \pi)^2 \geq d_1(v_i, \pi)^2$.

If the spectral gap $\mu = \frac{1}{p(n)}$ where n is the degree of a node and $p(n)$ is a polynomial, then after $k = O(p(n) \log N)$ steps,

$$d_2(v_k, \pi) = \left| \left(\frac{I + M}{2} \right)^k e_1 - \frac{1}{\sqrt{n}} v_1 \right|^2 \leq \left(\frac{1 + \lambda_2}{2} \right)^k = \left(1 - \frac{1}{2p(n)} \right)^k \leq \frac{1}{\text{poly}(N)} \tag{2}$$

After $O(p(n)\log N)$ steps of the random walk we have $\frac{1}{\text{poly}(N)} \geq d_1(v_i,\pi)^2$. If the spectral gap is $1/p(n)$, the Markov chain is rapidly mixing and converges to the uniform distribution in $O(p(n)\log N)$ steps.

1.2 Volume estimation

Recall the random walk defined on the grid points inside a convex body defined earlier. The edge expansion $h(S) = \frac{E(S,\overline{S})}{d|S|}$ is approximately equal to the ratio of the surface area of S to its volume. The approximation is valid only if the grid size is sufficiently small compared to the size of the convex body. The discussion on the convergence of the Markov chain here is just a sketch, for example if the convex object is an elongated parallelogram the graph on grid points inside it will not even be connected.

The hard part of Cheeger's inequality $h(G) \leq \sqrt{2(1-\lambda_2)}$ is useful for showing that the spectral gap is large, if we can bound the edge expansion $h(G)$. The following isoperimetric inequality for convex objects $P \in \mathbb{R}^n$ yields bounds on the edge expansion for the graph on grid points inside P,

$$Vol_{n-1}(S,\overline{S}) \geq \frac{\min(Vol(S), Vol(\overline{S}))}{diam(P)} \tag{3}$$

For example a bound of $1/n^3$ for edge expansion obtained using the isoperimetric inequality translates to a bound of $O(n^6 \log N)$ on the convergence time of the Markov chain using Cheeger's inequality. The isoperimeteric inequality shows that the random walk on grid points inside P converges to the uniform distribution over grid points in polynomial time if the diameter of P is $poly(n)$ and P is 'sufficiently nice'.

1.3 Counting partial orders

A good illustration of the Markov chain method is Khachiyan's solution to the problem of sampling from the uniform distribution on the total orderings consistent with a given partial order on elements $x_1, \cdots, x_n$. The partial order is specified by inequalities of the form $x_i \geq x_j$ among the elements x_i and the task is to estimate the number of total orderings compatible with the constraints.

It is easy to find one consistent ordering by running a topological sort algorithm on the directed graph representing the constraints. The neighbors of a consistent total ordering are generated by by exchanging elements x_i, x_j in the order if the resulting order is consistent, and adding a self loop otherwise. Each vertex has degree $\binom{n}{2}$ in the resulting graph.

Khachiyan observed that a constraint $x_i < x_j$ is a halfspace that divides the n dimensional hypercube $[0,1]^n$ into two equal parts. The hypercube gets partitioned into $n!$ equal parts by the $\binom{n}{2}$ hyperplanes corresponding to the results of all possible comparisons between elements (x_i, x_j). The order polytope is the intersection of hyperplanes corresponding to the partial order constraints and has volume $\frac{K}{n!}$ where K is the number of consistent total orderings.

Estimating the number of consistent total orders is equivalent to estimating the volume of order polytope. The walk on orderings can be viewed as a walk on the $n!$ simplices that partition the hypercube, so the isoperimetric inequality (3) can be used to bound the edge expansion.

The diameter of the order polytope is $\sqrt{n}$ as it is contained in the hypercube. The volume of a cut $(S, \overline{S})$ in the order polytope is $\frac{|S|}{n!}$. Each edge in the graph connects two unit simplices whose boundary is an $n-1$ dimensional unit simplex. The surface area per edge is $1/(n-1)!$ and the total surface area is at least $\frac{|S|}{n!\sqrt{n}}$. The total number of edges across the cut S must be at least $|S|/n^{3/2}$, yielding a bound on the edge expansion,

The isoperimetric inequality (3) applied to the order polytope yields,

$$Vol_{n-1}(S, \overline{S}) = \frac{E(S, \overline{S})}{(n-1)!} \geq \frac{|S|}{n!\sqrt{n}}$$

This yields the following bound on the edge expansion as the degree d is $O(n^2)$,

$$h(S) = \frac{|E(S, \overline{S})|}{d|S|} \geq \frac{1}{n^{7/2}}$$

Going through the calculation using Cheeger's inequality, the Markov chain converges to the stationary distribution in time $O(n^7 \log N)$. Here N is the total number of orderings which is upper bounded by $N!$, so the convergence time is $O(n^8 \log n)$. A sharper analysis of the convergence time is possible, our goal here was to show convergence in polynomial time.