Stanford University **Winter 2018**
Math 233A: Non-constructive methods in combinatorics **Instructor: Jan Vondrák**
Lecture date: 2/27/2018 **Scribe: Andy Tsao**

Lecture 15. Van der Waerden's conjecture and computing the permanent

Recall that the permanent of a matrix A is given by

$$\text{Per}(A) = \sum_{\pi \in S_n} \prod_{i=1}^{n} a_{i,\pi(i)}.$$

Consider the bipartite graph, where the two partitions each have n vertices labeled $1, \ldots, n$. The edge between vertex i and j, where i and j belong to different parts, is given weight a_{ij}. Then, the permanent of A can be written as follows:

$$\text{Per}(A) = \sum_{\text{matching } M} \prod_{(i,j) \in M} a_{i,j}.$$

For 0/1 matrices, the permanent is equivalent to counting the number of perfect matchings which is known to be #P-hard. This begs the question of whether or not we can efficiently approximate the permanent of a matrix (for now, we'll assume that the matrix has nonnegative entries).

We first look at the special case of doubly stochastic matrices. These are matrices $\{a_{ij}\}_{i,j=1}^{n}$ such that

$$\sum_{i=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ij} = 1$$

for all $1 \leq i, j \leq n$.

Theorem 15.1 (Birkhoff-Von Neumann) *Every doubly stochastic matrix is a convex combination of permutation matrices. That is,*

$$A = \sum_{\pi \in S_n} \lambda_\pi P_\pi$$

for some $\lambda_\pi \geq 0$ *such that* $\sum_{\pi \in S_n} \lambda_\pi = 1$.

Remark 15.2 *The Birkhoff-Von Neumann theorem immediately implies that the permanent of any doubly stochastic matrix is strictly positive.*

15.1 Examples

Two simple examples of doubly stochastic matrices are permutation matrices P_π and the matrix where all entries are equal to $\frac{1}{n}$ (denoted $\frac{1}{n}J$). It is easy to compute

$$\text{Per}(P_\pi) = 1$$
$$\text{Per}\left(\frac{1}{n}J\right) = \frac{n!}{n^n} \approx e^{-n}.$$

It is quite easy to see that the permanent of a doubly stochastic matrix cannot be more than 1. In 1926, Van der Waerden conjectured that $\mathrm{Per}(\frac{1}{n}J) = \frac{n!}{n^n}$ is the other extreme.

Conjecture 1 *Any doubly stochastic matrix A satisfies*

$$Per(A) \geq \frac{n!}{n^n}.$$

This result was proved to be correct in 1981. More recently, Gurvits found a proof using the machinery of stable polynomials which is rather elegant. Before we dive into the proof, we make a few auxiliary definitions.

Define the polynomial $p(x_1, \ldots, x_n) = \prod_{i=1}^n \sum_{j=1}^n a_{ij} x_j$. This is a polynomial where every monomial has degree exactly n. The permanent of the matrix $A = \{a_{ij}\}$ corresponds to the coefficient of the $x_1 \cdots x_n$ monomial. Thus we may write

$$\mathrm{Per}(A) = \frac{\partial^n p}{\partial x_1 \cdots \partial x_n}\Big|_{x=0}.$$

(In fact, the derivative is constant, so setting $x = 0$ is unnecessary, but convenient.) We also define the *capacity* of p to be

$$\mathrm{cap}(p) = \inf_{x>0} \frac{p(x_1, \ldots, x_n)}{x_1 \cdots x_n}.$$

First, consider univariate polynomials. In general, there is no relation between $\mathrm{cap}(R)$ and $R'(0)$ (consider for example $R(x) = Cx^2 + 1$ for $C > 0$ large). However, for real-rooted polynomials there is a relationship.

Lemma 15.3 *For any univariate real-rooted polynomial $R(x)$ of degree k with nonnegative coefficients,*

$$cap(R) \leq \left(\frac{k}{k-1}\right)^{k-1} R'(0)$$

for $k \geq 2$, and

$$cap(R) = R'(0)$$

for $k = 1$.

Proof: Since R is real-rooted and has nonnegative coefficients, all of its roots must lie on the non-positive real axis. If $k = 1$, then $R(x) = ax + b$; it is easy to see that

$$\mathrm{cap}(R) = \inf_{x>0} \frac{R(x)}{x} = \lim_{x\to\infty} \frac{ax+b}{x} = a = R'(0).$$

For $k \geq 2$, we write

$$R(x) = \prod_{i=1}^{k} (x + \alpha_i),$$

where $\alpha_i \geq 0$. If $R(0) = 0$, then $\text{cap}(R) = R'(0)$, since R is a convex function for $x > 0$. Otherwise, $R(0) > 0$, and we may assume that $\alpha_i > 0$ for all i. Letting $a_i = \frac{1}{\alpha_i}$, we can rewrite R as follows:

$$\begin{aligned} R(x) &= \prod (x + \alpha_i) \\ &= \prod \left(x + \frac{1}{a_i} \right) \\ &= \prod \frac{1}{a_i} \prod (a_i x + 1), \\ \frac{R(x)}{x} &= \frac{1}{x} \prod \frac{1}{a_i} \prod (a_i x + 1). \end{aligned}$$

By the AM-GM inequality, we have $\prod_{i=1}^k (a_i x + 1) \leq \left(\frac{1}{k} \sum_{i=1}^k (a_i x + 1) \right)^k$, yielding the inequality

$$\frac{R(x)}{x} \leq \frac{1}{x} \frac{1}{\prod_{i=1}^k a_i} \left(1 + \frac{x}{k} \sum_{j=1}^k a_j \right)^k.$$

Letting $x = \frac{k}{k-1} \cdot \frac{1}{\sum_{j=1}^k a_j}$ (which is the choice minimizing the RHS) yields

$$\begin{aligned} \inf_{x>0} \frac{R(x)}{x} &\leq \frac{k-1}{k} \sum_{j=1}^k a_j \cdot \frac{1}{\prod_{i=1}^k a_i} \left(1 + \frac{1}{k-1} \right)^k \\ &= \frac{\sum_{j=1}^k a_j}{\prod_{i=1}^k a_i} \left(\frac{k}{k-1} \right)^{k-1}. \end{aligned}$$

It remains to notice that $R(0) = \prod_{i=1}^k \frac{1}{a_i}$ and $R'(0) = R(0) \sum_{j=1}^k a_j = \frac{\sum_{j=1}^k a_j}{\prod_{i=1}^k a_i}$. □

Theorem 15.4 (Gurvits) *For any stable polynomial $p(x_1, \ldots, x_n)$ of total degree n with nonnegative coefficients,*

$$\mathit{cap}(p) \leq \frac{n!}{n^n} \frac{\partial^n p}{\partial x_1 \cdots \partial x_n}\Big|_{x=0}.$$

Proof: The proof is by induction on n. For the base case, consider Lemma 15.3 with $k = 1$: in this case, indeed we have $\mathit{cap}(p) = p'(0)$.

For $n > 1$, let

$$r(x_2, \ldots, x_n) = \frac{\partial p}{\partial x_1}\Big|_{x=(0,x_2,\ldots,x_n)}.$$

By Lemma 15.3, since p has degree at most n in x_1, we know that

$$\inf_{x_1>0} \frac{p(x_1, \ldots, x_n)}{x_1} \leq \left(\frac{n}{n-1} \right)^{n-1} r(x_2, \ldots, x_n).$$

Also, r has total degree at most $n-1$, so the induction hypothesis for polynomials of $n-1$ variables gives us

$$\inf_{x_2,\ldots,x_n>0} \frac{r(x_2, \ldots, x_n)}{x_2 \cdots x_n} \leq \frac{(n-1)^{n-1}}{(n-1)!} \frac{\partial^{n-1} r}{\partial x_2 \cdots \partial x_n}\Big|_{(x_2,\ldots,x_n)=0}.$$

Putting these two inequalities together yields

$$\inf_{x_1,\dots,x_n>0} \frac{p(x_1,\dots,x_n)}{x_1x_2\cdots x_n} \le \left(\frac{n}{n-1}\right)^{n-1} \inf_{x_2,\dots,x_n>0} \frac{r(x_2,\dots,x_n)}{x_2\cdots x_n}$$
$$\le \frac{n^{n-1}}{(n-1)!}\cdot \frac{\partial^n p}{\partial x_1\cdots\partial x_n}\Big|_{x=0} = \frac{n^n}{n!}\cdot\frac{\partial^n p}{\partial x_1\cdots\partial x_n}\Big|_{x=0}.$$

This completes the proof. □

Now, to complete the proof of Conjecture 1, we must show that $\mathrm{cap}(p) \ge 1$. Indeed, for any $x_1,\dots,x_n > 0$, we have

$$p(x_1,\dots,x_n) = \prod_{i=1}^n \sum_{j=1}^n a_{ij}x_j \overset{\text{AM-GM}}{\ge} \prod_{i=1}^n\prod_{j=1}^n x_j^{a_{ij}} = \prod_{j=1}^n x_j^{\sum_{i=1}^n a_{ij}} = \prod_{j=1}^n x_j.$$

Combining the two results gives us $\mathrm{Per}(A) \ge \frac{n!}{n^n}\mathrm{cap}(p) \ge \frac{n!}{n^n}$.

15.2 Matrix scaling

Given a nonnegative matrix A, it is natural to ask whether the bounds in the previous section give us any control over the permanent of A. In this section, we will investigate deterministic algorithms that approximately compute the permanent. The idea is to split the computation into two steps. In the first, we do a series of row and column transformations to turn A into a (approximately) doubly stochastic matrix B. Then, the results from the previous section allow us to give an e^n approximation of $\mathrm{Per}(B)$. It is worth noting that the best known deterministic algorithm achieves a $(2-\epsilon)^n$ approximation. Moreover, it has been shown that any algorithm that achieves a $2^{n^{1-\epsilon}}$ approximation can be turned into a $(1+\epsilon)$ approximation. Randomized algorithms that achieve this are known (based on Monte Carlo Markov Chain sampling). However, for deterministic algorithms this is a big open question.

Our algorithm constructs matrices $A^{(1)}, A^{(2)}, \dots$ as follows:

- $A^{(1)}$ is constructed by normalizing the rows of A.
- For $k \ge 2$ even, $A^{(k)}$ is constructed by normalizing the columns of $A^{(k-1)}$.
- For $k \ge 2$ odd, $A^{(k)}$ is constructed by normalizing the rows of $A^{(k-1)}$.

First, we note that if $\mathrm{Per}(A) = 0$, then this algorithm will never yield a matrix that is approximately doubly-stochastic (because the permanent will remain 0). For all other matrices, we claim that this process will converge to a doubly stochastic matrix.

Suppose we are at stage k, and the matrix $A^{(k)}$ has normalized columns. To construct $A^{(k+1)}$, we multiply row i by μ_i, for all $1 \le i \le n$. Then,

$$\mathrm{Per}(A^{(k+1)}) = \mathrm{Per}(A^{(k)})\prod_{i=1}^n \mu_i.$$

Furthermore, letting $r_i = \sum_{j=1}^n a_{ij}$ be the i^{th} row sum, we note by AM-GM that

$$\prod_{i=1}^n r_i \leq \left(\frac{\sum_{i=1}^n r_i}{n}\right)^{1/n} = 1.$$

This means that $\prod_{i=1}^n \mu_i \geq 1$, so the permanent is a non-decreasing function of k. Moreover, the permanent will increase significantly, as long as we are far away from a doubly stochastic matrix, and also it is upper-bounded by 1, so this process cannot continue forever. The next lemma quantifies this fact.

Lemma 15.5 *If $A^{(k)}$ has rows satisfying $\sum_{i=1}^n (r_i - 1)^2 = \Delta$, then*

$$Per(A^{(k+1)}) \geq \frac{Per(A^{(k)})}{1 - \frac{\Delta}{2} + o(\Delta^{3/2})} \tag{15.1}$$

We skip the proof of this lemma. Finally, we want to claim that the permanent is not too small at the beginning. For this, we actually need to include an additional preprocessing step:

(1) Choose $\pi \in S_n$ to maximize $\prod_{i=1}^n a_{i,\pi(i)}$. Rearrange the rows and columns so that π corresponds to the identity.

(2) Scale the columns to make sure that $a_{ii} = \max_{1 \leq j \leq n} a_{ij}$. This is possible by solving an LP.

(3) Normalize the rows so that the diagonal entries are all at least $1/n$. This means that the $\text{Per}(A) \geq n^{-n}$.

(4) Run the above procedure until we are sufficiently close to a doubly stochastic matrix.

Finally, we have the following lemma which specifies how close we need to get to a doubly stochastic matrix.

Lemma 15.6 *If $A^{(k)} \geq 0$, with columns summing to 1, and $\sum (r_i - 1)^2 < \frac{1}{n^3}$, then $1 \geq Per(A^{(k)}) \geq \Omega(e^{-n})$.*

Therefore, at this point we can obtain an $O(e^n)$-approximation to the permanent of the original matrix (by tracking the updates throughout the process). Since at the beginning, we have $\text{Per}(A^{(1)}) \geq e^{-n}$, and in each step the permanent increases by a factor of $1 + O(n^{-3})$ (until $\sum (r_i - 1)^2 < \frac{1}{n^3}$), the process cannot take more than $O(n^4 \log n)$ steps.

More details can be found in [LSW00].

References

[LSW00] Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica*, 20(4), 2000.