Math 270: Geometry of Polynomials **Fall 2015**

Lecture 5: Lowerbound on the Permanent and Application to TSP

Lecturer: Zsolt Bartha, Satyaki Mukherjee *Scribe: Yumeng Zhang*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Given an $n \times n$ matrix $\mathbf{A} = (a_{ij})$, the permanent of $\mathbf{A}$ is defined as

$$\operatorname{per}\mathbf{A} = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)},$$

where S_n is the symmetric group of permutations on $\{1, 2, \ldots, n\}$. Unlike the determinant, computing the exact value of the permanent of a given matrix is known to be #P-hard. Thus, most research has focused on finding approximation algorithms for and proving upper and lower bounds for restricted classes of matrices.

A sharp upper-bound on the permanent the case of $0-1$ matrices was proved by Brégman:

Theorem 5.1 (Brégman). *If $\mathbf{A}$ is a $0-1$ matrix, then*

$$\operatorname{per}\mathbf{A} \le \prod_{i=1}^{n} (r_i!)^{1/r_i},$$

where r_i is the i'th row sum.

In this lecture we are interested in lower-bounding the permanent in different special case of doubly stochastic matrices. A (real-valued) matrix $\mathbf{A}$ is called *doubly stochastic* if each of its entry satisfies $a_{ij} \ge 0$ and $\sum_i a_{ij} = \sum_j a_{ij} = 1$. The main result is the following theorem first conjectured by Van der Waerden in 1926 and first completely proved in 1981.

Theorem 5.2 (Main Result). *For any $n \times n$ matrix $\mathbf{A}$, if $\mathbf{A}$ is doubly stochastic, then*

$$\operatorname{per}\mathbf{A} \ge n!/n^n.$$

Equality is achieved uniquely at the matrix J_n in which every entry is $1/n$.

The original proof of this theorem, due to Erorychev [Ero81] and Falikman [Fal81] was based on a deep inequality of Alexandrov and Fenchel concerning mixed volumes of convex bodies.

In this lecture we give a short proof to this famous conjecture using real stable polynomials, following the paper of Gurvits in 2008 [Gur08]. We begin by motivating this lower bound by showing how it yields a polynomial time approximating algorithm for the traveling salesman problem [Vis12] on regular graphs.

The traveling salesman problem can be defined as the following.

Definition 5.3 (TSP)**.** Given an undirected graph $G = (V, E)$ and cost function $c : E \to \mathbb{R}_{\geq 0}$, find a closed walk (tour) of minimum cost that connects all the vertices and returns to the origin.

The traveling salesman problem in its most general form is NP-hard. It is also NP-hard in general to find a tour that approximates the minimum cost with arbitrarily small error. Thus we can only hope to find such an algorithm for special cases.

An instance of TSP is called GRAPHTSP if it has a cost function of constant value $c \equiv 1$. In the case of GRAPHTSP on regular graphs, Vishnoi showed the following result.

Theorem 5.4 ([Vis12])**.** *There exists a randomized polynomial-time algorithm that given any simple, connected, k-regular graph on n vertices, finds a tour of length $\ell \leq (1 + \frac{64}{\log k} n)$ with probability $1 - \frac{1}{n}$.*

The algorithm is built on two results about permanents:

1. The lower bound on permanent: Let $\mathbf{A}$ be the adjacency matrix of a bipartite graph G. The permanent of $\mathbf{A}$ equals to the number of perfect matchings in graph G. Thus if G is k-regular, then
$$\#\text{perfect matchings} = \operatorname{per} \mathbf{A} \geq \frac{n! k^n}{n^n}.$$
2. The polynomial time algorithm of approximately sampling perfect matchings by Jerrum-Sinclair-Vigoda:

   Theorem 5.5 ([JSV04])**.** *There exists a randomized algorithm to almost uniformly sample perfect matchings in a bipartite graph of n vertices with running time $O(|n| \log(1/\epsilon))$, where ϵ is the approximation error.*

Given these two results, the general idea of the algorithm is to do the following:

1. Pick a cycle-cover[1] C in G.
2. Pick a minimum spanning tree T of the graph obtained by contracting all cycles in C to single vertices.
3. Output the tour $C \cup T \cup T'$, where T' is an identical disjoint copy of T, and each tree edge is taken to be traversed once in each direction.

The construction above yields a tour of length upper-bounded by $n + 2(|C| - 1)$, where $|C|$ is the number of cycles in C. Thus to get a shorter path, we are interested in sampling small cycle covers efficiently. For that purpose we translate the problem of finding cycle covers into sampling perfect matchings of bipartite graphs, after which we can apply Jerrum-Sinclair-Vigoda's algorithm to sample a random cycle cover.

[1] A cycle cover in G is a collection of vertex-disjoint cycles, possibly containing length 2 cycles obtained by picking an edge twice, which cover all the vertices

For any graph G, let $G' = (V', E')$ be the (natural) bipartite graph constructed from G with

$$V' = \{u^{\mathrm{R}}, u^{\mathrm{L}} : u \in V\},$$
$$E' = \{(u^{\mathrm{L}}, v^{\mathrm{R}}), (u^{\mathrm{R}}, v^{\mathrm{L}}), (u, v) \in E\}.$$

It is easy to see that G' is k-regular and bipartite. Moreover, for every perfect matching M of G', the natural projection of M back to G gives a cycle cover $C(M)$ of G. (See Figure 5.1 for illustration)

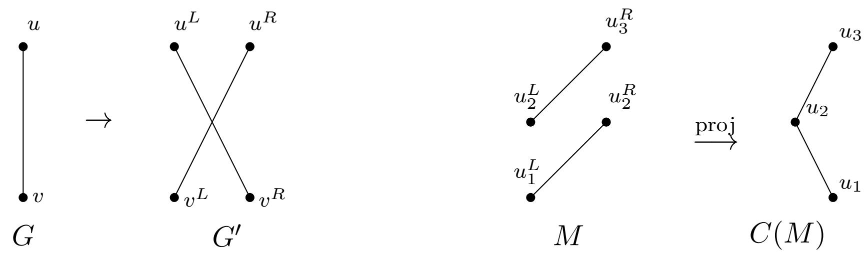

Figure 5.1: Constructing cycle covers from perfect matchings

Proof of Theorem 5.4. Let $\mathbb{P}(\cdot)$ denote the probability distribution induced on the set of cycle covers of G by the uniform measure on the perfect matchings of G'. The lower bound on permanent yields an uniform upper bound on the probability of every single cycle cover:

$$\mathbb{P}(C) = \frac{1}{\text{per}\,\mathbf{A}} 2^{\#\text{ of cycles of length } \geq 3} \leq \frac{2^{n/3}}{n!k^n/n^n} \leq 2^{n/3} \big(\frac{e}{k}\big)^n,$$

where in the second step we use the fact that there are at most n/ℓ disjoint cycles of length ℓ.

Note that in our construction of cycle covers, every vertex belongs to at most two cycles of C (corresponding to the two matchings it contains). Thus

$$\begin{aligned} P(\#\text{ of cycles of length } \leq \ell) \geq \gamma_n) &\leq P(\exists 2\gamma_n \text{ vertices covered by cycles of length } \leq \ell) \\ &= \sum_{t=2\gamma_n}^{n} P(\#\text{ of such vertices} = t) \\ &\leq \sum_{t=2\gamma_n}^{n} \binom{n}{t} \cdot k^{(1-\ell^{-1})t} \cdot k^{n-t} \max_C \mathbb{P}(C), \end{aligned}$$

where the $k^{(1-\ell^{-1})t}$ term comes from the number of cycle covers on those t vertices using cycles of length $\leq \ell$ and we have $\mathbb{P}(C) \leq 2^{n/3}\big(\frac{e}{k}\big)^n$. On the complement of the described event, the cost function satisfies

$$\text{cost} \leq n + 2(\#\text{short cycles} + \#\text{long cycles}) \leq n + 2(\frac{n}{\gamma_n} + \ell).$$

By choosing the parameter l, γ_n properly, we can show that the desired upper-bound on the cost of output tour holds with probability $1 - \frac{1}{n}$. □

Remark 5.6. The regularity of G is required to ensure that the adjacency matrix $\mathbf{A}$ is a multiple of a doubly stochastic matrix. The result does not hold non-regular graphs. In fact, perfect matchings might not even exist on graphs with very unbalanced degree distribution.

5.1 Proof of Theorem 5.2

In this section we give a proof of Theorem 5.2. We first give an alternative formulation of the permanent as a high derivative of a simple real stable polynomial.

Definition 5.7 (Alternative Definition)**.** Given $n \times n$ matrix $\mathbf{A}$, define polynomial $p_{\mathbf{A}}(x_1, \ldots, x_n) = \prod_i (\sum_j a_{ij} x_j)$, then

$$\frac{\partial^n p_{\mathbf{A}}(0, \ldots, 0)}{\partial x_1 \cdots \partial x_n} = \operatorname{per} \mathbf{A}.$$

We will be restricting our attention doubly stochastic $\mathbf{A}$ so $p_{\mathbf{A}}$ will always be real stable, as it is a product of linear polynomials with nonnegative coefficients. The following quantity was introduced by Gurvits in the context of quantum information theory [Gur03] and may be seen as a convex relaxation of the derivative at zero of a polyomial.

Definition 5.8 (Capacity)**.** Given any polynomial p, the capacity of p is defined as

$$\operatorname{cap}(p) := \inf_{x_i > 0} \frac{p(x_1, \ldots, x_n)}{x_1 \cdots x_n}.$$

Remark 5.9. That the capacity is a convex relaxation is better appreciated after making the change of variables $x_i = e^{t_i}$ and observing that it is equivalent to minimizing $\log p(e^{t_1}, \ldots, e^{t_n})$ (which is convex for a polynomial with nonnegative coefficients) over the affine subspace $t_1 + t_2 + \ldots + t_n = 0$.

The next result relates the capacity of a real-rooted univariate polynomial to its derivative at 0.

Lemma 5.10. *Let R be a univariate, real-rooted polynomial of degree k with non-negative coefficients, then*

$$\operatorname{cap}(R) \leq \left(\frac{k}{k-1}\right)^{k-1} R'(0), \text{ for } k > 1.$$

In the case of $k = 1$, we have $\operatorname{cap}(R) = R'(0)$.

Remark 5.11. The definition of capacity can be interpreted as finding the "steepest" slope starting from the origin that lies under a given curve. In this way, it is intuitively true that the capacity of a polynomial can't be much bigger than the "slope" of the polynomial itself at the origin.

Proof of Lemma 5.10. For $k = 1$, there exist r_0, r_1 such that $R(t) = r_0 + r_1 t$ and we have that $\operatorname{cap}(R) = t_0 \wedge t_1 \leq r_1 = R'(0)$.

For $k \geq 2$, we first consider the case of $R(0) = 0$. If $R(0) = 0$, then we can write $R(t) = t(r_1 + r_2 t + \cdots + r_{k-1} t^{k-1})$. Thus $R(t)/t$ achieves its infimum at $t = 0$ and $\operatorname{cap}(R) = r_1 = R'(0)$.

When $R(0) \neq 0$, assume without loss of generality that R is monic. Then we can write $R(t)$ as $R(t) = \prod_{i=1}^k (t + \alpha_i)$ since $R(t)$ is real-rooted with non-negative coefficients. Letting $a_i = 1/\alpha_i$, we have

$$R(t) = \frac{\prod_{i=1}^k (a_i t + 1)}{\prod_{i=1}^k a_i} = R(0) \prod_{i=1}^k (a_i t + 1) \overset{\text{AM-GM}}{\leq} R(0) \left(\frac{1}{k} \sum_i (a_i t + 1)\right)^k = R(0) \left(\frac{\sum_i a_i}{k} t + 1\right)^k,$$

where in the third step we use the arithmetic mean-geometric mean inequality. Note that

$$R(t) = R(0)\prod_{i=1}^{k}(a_i t + 1) \Rightarrow R'(0) = R(0)\sum_{i=1}^{k} a_i.$$

Substituting $\sum_{i=1}^k a_i$ with $R'(0)/R(0)$ and dividing by t gives

$$\frac{R(t)}{t} \leq \frac{R(0)}{t}\Big(\frac{R'(0)}{kR(0)}t + 1\Big) = R'(0)\frac{1}{x}\Big(1 + \frac{x}{k}\Big)^{k-1}, \text{ where } x = \frac{R'(0)t}{R(0)}.$$

Taking an infimum over $x > 0$ gives (the infimum is achieved by $x = \frac{k}{k-1}$)

$$\mathrm{cap}(R) \leq \inf_{x>0} R'(0)\frac{1}{x}(1 + \frac{x}{k})^k = R'(0)(\frac{k}{k-1})^{k-1}.$$

□

The main result of [Gur08] is the following theorem which implies Theorem 5.2 as a corollary.

Theorem 5.12. *If $p(x_1, \ldots, x_n)$ is a stable polynomial of degree n with non-negative coefficients, then*

$$\frac{\partial^n p(0, \ldots, 0)}{\partial x_1 \cdots \partial x_n} \geq \frac{n!}{n^n}\,\mathrm{cap}(p).$$

Proof of Theorem 5.12. Fix $x_2, \ldots, x_n$. Define $q = \frac{\partial}{\partial x_1}p$ and $r(x_2, \ldots, x_n) = q(0, x_2, \ldots, x_n)$ and observe that both q and r are real stable by the closure properties of Lecture 2. Lemma 5.10 implies that

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

Thus dividing both sides by $x_2 \cdots x_n$ and applying Lemma 5.10 inductively, we have

$$\begin{aligned}\inf_{x_1,\ldots,x_n>0}\frac{p(x_1, \ldots, x_n)}{x_1 \cdots x_n} &\leq \inf_{x_2,\ldots,x_n>0}\frac{r(x_2, \ldots, x_n)}{x_2 \cdots x_n}(\frac{n}{n-1})^{n-1} = \mathrm{cap}(r)(\frac{n}{n-1})^{n-1} \\ \text{(by induction)} &\leq \frac{(n-1)^{n-1}}{(n-1)!}\cdot(\frac{n}{n-1})^{n-1}\frac{\partial^{n-1}r(0, \ldots, 0)}{\partial x_1 \cdots \partial x_{n-1}} \\ &= \frac{n^n}{n!}\frac{\partial^n p(0, \ldots, 0)}{\partial x_1 \cdots \partial x_n}.\end{aligned}$$

□

Proof of Theorem 5.2. All we need to show is $\mathrm{cap}(p_{\mathbf{A}}) \geq 1$. Recall the definition of $p_{\mathbf{A}}(x_1, \ldots, x_n) = \prod_i(\sum_j a_{ij}x_j)$. Applying the power mean inequality (which amounts to concavity of the logarithm) and appealing to the double stochasticity of $\mathbf{A}$:

$$\sum_j a_{ij} = 1 \Rightarrow \sum_j a_{ij}x_j \geq \prod_j x_j^{a_{ij}}.$$

Thus,

$$p_{\mathbf{A}}(x_1, \ldots, x_n) \geq \prod_i\prod_j x_j^{a_{ij}} = \prod_j x_j^{\sum_i a_{ij}} = \prod_j x_j,$$

which in turn implies that $\mathrm{cap}(p_{\mathbf{A}}) = \frac{p_{\mathbf{A}}(x_1,\ldots,x_n)}{x_1\cdots x_n} \geq 1$. □

Remark 5.13. In the proof of this result, as well as in many other applications we will see in the future, real-rootedness allows one to use convexity inequalities such as the AM-GM inequality used in Lemma 5.10. Moreover, the closure of real-rootedness under coordinate conditioning and taking derivatives allows one to apply the same inequality repeatedly in an inductive way.

We conclude by briefly discussing an extension of the above result to arbitrary matrices with positive entries $(a_{ij}) > 0$. The crucial fact is that any such matrix admits a row-column scaling which makes it doubly stochastic: in particular, there are numbers $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$ such that the matrix

$$b_{ij} := \lambda_i \mu_j a_{ij}$$

is doubly stochastic, and $\mathrm{per}(B) = \prod_i \lambda_i \prod_j \mu_j \, \mathrm{per}(A)$. Moreover, these numbers can be efficiently computed in deterministic polynomial time [LSW98]. Since $\mathrm{per}(A)$ lies between 1 and $\frac{n!}{n^n} \geq e^{-n}$ by the above theorem, this implies that $\mathrm{per}(B)$ can be approximated up to a factor of e^n in deterministic polynomial time. This bound has very recently been improved to 2^n [GS14] and obtaining a better, perhaps $(1+\epsilon)^n$ bound, in deterministic polynomial time remains an outstanding open problem.

We remark that if randomization is allowed, the Markov chain based algorithm of [JSV04] gives a $(1+\epsilon)$ bound in the case of nonnegative matrices.

References

[Ero81] GP Erorychev. Proof of the van der waerden conjecture for permanents. *Siberian Mathematical Journal*, 22(6):854–859, 1981.

[Fal81] Dmitry I Falikman. Proof of the van der waerden conjecture regarding the permanent of a doubly stochastic matrix. *Mathematical Notes*, 29(6):475–479, 1981.

[GS14] Leonid Gurvits and Alex Samorodnitsky. Bounds on the permanent and some applications. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 90–99. IEEE, 2014.

[Gur03] Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 10–19. ACM, 2003.

[Gur08] Leonid Gurvits. Van der waerden/schrijver-valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. *Electron. J. Combin*, 15(1), 2008.

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004.

[LSW98] Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 644–652. ACM, 1998.

[Vis12] Nisheeth K Vishnoi. A permanent approach to the traveling salesman problem. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 76–80. IEEE, 2012.