Counting and Sampling	**Fall 2017**

Lecture 19: Log Concavity in Optimization

Lecturer: Shayan Oveis Gharan — *December 6th*

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

In the last lecture we discussed applications of log concavity of real rooted polynomials in designing a deterministic approximation algorithm for permanent.

We also introduced the class $\mathbb{H}$-stable polynomial. Recall that a polynomial $p(z_1, \dots, z_n)$ is $\mathbb{H}$ -stable if $p(z_1, \dots, z_n) \neq 0$ whenever $\Im z_i > 0$ for all i.

In the last lecture we saw that any univariate $\mathbb{H}$-stable polynomial (with real coefficients) is real rooted. We also showed that any real rooted polynomial with non-negative coefficients is log-concave. Therefore, univariate $\mathbb{H}$-stable polynomials with nonnegative coefficients is log-concave in its variable.

It turns out that there is a bigger story. In fact, *any* $\mathbb{H}$-stable polynomial $p(z_1, \dots, z_n)$ with nonnegative coefficients is log concave in $z_1, \dots, z_n$.

Lemma 19.1. *Any $\mathbb{H}$ stable polynomial $p \in \mathbb{R}_+[z_1, \dot{,} z_n]$ with nonnegative coefficients is log concave in its variables.*

The above fact is a *generalization* of the well-known fact that for PSD matrices X, $\det(X)$ is log-concave. As we will see below, Theorem 19.5, determinant of sum of PSD matrices give us $\mathbb{H}$-stable polynomials.

In this lecture and the next one we will see applications of the above lemma in optimization and counting.

19.1 Estimating the largest Coefficients of a $\mathbb{H}$-Stable Polynomial

We say a polynomial p is d-homogeneous if the degree of every monomial of p is d. In this section we prove the following theorem which is based on an extension of a result due to Nikolov [Nik15]:

Theorem 19.2. *There is a deterministic polynomial time algorithm that given evaluation oracle to a $\mathbb{H}$-stable homogeneous multilinear polynomial $p(z_1, \dots, z_n) = \sum_S a_S z^S$ compute $\max_S a_S$ within a e^d error where d is the degree of p.*

Consider the following convex program:

$$\begin{aligned} \max \quad & \log p(z_1, \dots, z_n) \\ \text{s.t.} \quad & \sum_{i=1}^n z_i = d \\ & z_i \geq 0 \qquad \forall i. \end{aligned} \tag{19.1}$$

Note that since p is log-concave, using ellipsoid method, we can (approximately) solve the above program given an evaluation oracle access to p.

We claim that the above program is a *relaxation* to the coefficient maximization problem. That is the the optimum of the above program is an upper bound of $\max_S a_S$. This is because letting $z_i = \mathbb{I}[i \in S]$ for all i, we get $p(z) = a_S$ and all constraints are satisfied.

Next, we describe a rounding algorithm. Consider the following probability distribution μ: For every i, $\mu(i) = z_i/d$. Note that since $\sum_i z_i = d$ this is indeed a probability distribution. To find a set S we sample k elements independently (with repetition) from μ. Call these elements $e_1, \dots, e_d$. Fix any set S with $|S| = d$; we claim that

$$\mathbb{P}\left[\{e_1, \dots, e_d\} = S\right] \geq z^S \frac{d!}{d^d} \approx e^{-d}. \tag{19.2}$$

Note that this completes the proof of the above theorem because expected value of the algorithm is at least

$$\mathbb{E}\left[ALG\right] = \sum_{S:|S|=d} a_S \mathbb{P}\left[\{e_1, \dots, e_k\} = S\right] \sum_{S:|S|=d} a_S z^S e^{-d} = \frac{p(z_1, \dots, z_n)}{e^d}.$$

It remains to prove Equation 19.2. In order to get $\{e_1, \dots, e_d\} = S$ we need to sample each element of S exactly once. Recall that for any $i \in S$, $\mu(i) = z_i/d$. Therefore, this probability is z^S/d^d. But, note that we can choose the elements of S in $d!$ possible orders. Therefore, we gain a factor of $d!$. Equation (19.2) and Theorem 19.2 follows.

Remark 19.3. *Note that the above algorithm gives a e^{-d} approximation in expectation. However, it could be that this happens with exponentially small probability. In this case we can turn the above proof into an algorithm using the conditional expectation method. We leave the proof as an exercise and we refer interested readers to [Nik15] for details.*

We conclude this section by discussing an application of the above theorem by designing an approximation algorithm for the largest volume parallepiped among a set of vectors. Let $v_1, \dots, v_n \in \mathbb{R}^d$ be a set of vectors. For a set of indices $S \subseteq [n]$, define

$$\text{vol}(S) = \det([\{\langle v_i, v_j \rangle]_{i,j \in S}).$$

be the square of the volume of the parallelepiped spanned by the vectors in S. Note that for any set $S \subseteq [n]$ the above quantity is nonnegative. This is because the gram-matrix with i, j-th entry $\langle v_i, v_j \rangle$ is a PSD matrix and the determinant of any PSD matrix is non-negative.

Theorem 19.4. *There is a polynomial time algorithm that for any given set of vectors $v_1, \dots, v_n \in \mathbb{R}^d$ and any $1 \leq k \leq d$, finds a set S such that*

$$\text{vol}(S) \geq e^{-k} \max_T \text{vol}(T).$$

The above theorem simply follows Theorem 19.2 and the fact that for any $1 \leq k \leq d$ the following polynomial is $\mathbb{H}$-stable.

$$p(z_1, \dots, z_n) = \sum_{S \in \binom{n}{k}} \text{vol}(S) z^S. \tag{19.3}$$

So, to prove the theorem we need to show that the above polynomial is $\mathbb{H}$-stable. First of all, there is a very general technique to construct a $\mathbb{H}$-stable polynomial.

Theorem 19.5 (Mother of all $\mathbb{H}$-stable polynomials)**.** *For any PSD matrices $A_1, \dots, A_n \in \mathbb{R}^{n \times n}$ and any symmetric matrix $B \in \mathbb{R}^{n \times n}$ we have*

$$p(z_1, \dots, z_n) = \det(B + z_1 A_1 + \dots + z_n A_n)$$

is $\mathbb{H}$-stable.

Proof. The proof simply follows from the fact that the eigenvalues of a symmetric matrix are real. We use the following fact:

Fact 19.6. *A polynomial $p(z_1, \ldots, z_n)$ is $\mathbb{H}$-stable if and only for any two vectors $a, b \in \mathbb{R}^n$ where $a > 0$ we have the univariate polynomial $p(at + b)$ is real rooted.*

Proof. We prove one side of this fact as it is the side that we will use. That is if $p(at+b)$ is real rooted for all $a, b \in \mathbb{R}^n$ where $a > 0$ then p is $\mathbb{H}$-stable. We prove by contradiction. Suppose p is not $\mathbb{H}$-stable. Then, there exists $z_1, \ldots, z_n \in \mathbb{H}$ such that $p(z_1, \ldots, z_n) = 0$. Since each for each j, $\Im z_j > 0$ we can write $z_j = a_j i + b_j$ for $a_j > 0$. But then i is a root of the polynomial $p(at + b)$ for the corresponding vectors a, b. □

Now, we get back to prove the theorem. Let $a, b \in \mathbb{R}^n$ and $a > 0$. We show that the polynomial

$$\det(B + (a_1 t + b_1)A_1 + \cdots + (a_n t + b_n)A_n) = \det(t \sum_j a_j A_j + (B + \sum_j b_j A_j)$$

is real rooted. We assume that A_j matrices are all positive definite. The proof of the claim for the PSD matrices just follows from a limiting argument. Let $\bar{A} = \sum_j a_j A_j$ and $\bar{B} = B + \sum_j b_j A_j$. Since a_j's are positive, we have $\bar{A} \succ 0$. Therefore, it is invertible, and furthermore $\bar{A}^{-1/2}$ is well-defined. It follows that

$$\begin{aligned} \det\left(t \sum_j a_j A_j + \left(B + \sum_j b_j A_j\right)\right) &= \det(t\bar{A} + \bar{B}) \\ &= \det(\bar{A}(tI + \bar{A}^{-1/2}\bar{B}\bar{A}^{-1/2}) \\ &= \det(\bar{A}) \det(tI + \bar{A}^{-1/2}\bar{B}\bar{A}^{-1/2}). \end{aligned}$$

But $\det(\bar{A})$ is just a number, and the polynomial $\det(tI + \bar{A}^{-1/2}\bar{B}\bar{A}^{-1/2})$ is real rooted. The latter is because the matrix $\bar{A}^{-1/2}\bar{B}\bar{A}^{-1/2}$ is a symmetric matrix and it has real eigenvalues and that $\det(tI + \bar{A}^{-1/2}\bar{B}\bar{A}^{-1/2})$ is just the characteristic polynomial of $\bar{A}^{-1/2}\bar{B}\bar{A}^{-1/2}$. □

In fact, almost all of the $\mathbb{H}$-stable polynomials that we are aware off are constructed by instantiating the above theorem and then employing *stability preserving operators*. We will discuss these operators in the next lecture.

To prove that the polynomial of (19.1) is $\mathbb{H}$-stable, first note that by the above theorem the polynomial

$$\det(xI + z_1 v_1 v_1^T + \cdots + z_n v_n v_n^T)$$

is $\mathbb{H}$-stable by the above theorem. By Couchy-Binet inequality we can rewrite the above polynomial as follows:

$$\sum_{k=0}^{n} x^{n-k} \sum_{S \in \binom{n}{k}} \mathrm{vol}(S) z^S.$$

Observe that the polynomial in (19.1) is the coefficient of x^{n-k} in the above polynomial. So, to obtain that polynomial first we differentiate the above polynomial k times with respect to x, and then we set $x = 0$. This completes the proof of Theorem 19.4.

We note that Theorem 19.4 has interesting applications in machine learning in studying determinantal point processes; we refer interested readers to [KT12].

References

[KT12] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *arXiv preprint arXiv:1207.6083*, 2012. 19-3

[Nik15] Aleksandar Nikolov. Randomized rounding for the largest simplex problem. In *STOC*, pages 861–870, 2015. 19-1, 19-2