Lecture 6: Generalizing Gurvits Machinery

Lecturer: Shayan Oveis Gharan *Jan 31st*

In the last lecture we saw a nice application of Gurvits' theorem in studying properties of Strongly Rayleigh distributions. It is natural to ask, if we can approximate sums of coefficients of a given real stable polynomial, as opposed to a single coefficient. This is the main topic of this lecture; we will prove a theorem proved in a joint work with Anari and we will see its applications in Nash welfare maximization problems.

Given two polynomials $p, q \in \mathbb{R}_{\geq 0}[z_1, \dots, z_n]$ we define the inner product of p, q as follows:

$$\langle p, q \rangle = q(\partial z)p|_{z=0} = p(\partial z)q|_{z=0} = \sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} c_p(\kappa)c_q(\kappa)\kappa!,$$

where $\kappa! = \prod_{i=1}^n \kappa_i!$. For example, suppose p, q are the generating polynomials of two k-DPPs, i.e., given two ensemble matrices K, L, we have

$$p = \sum_{S \in \binom{n}{k}} \det(K_{S,S})z^S, q = \sum_{S \in \binom{n}{k}} \det(L_{S,S})z^S.$$

Then,

$$\langle p, q \rangle = \sum_{S \in \binom{n}{k}} \det(K_{S,S}) \det(L_{S,S}),$$

is the correlation between these two DPPs. It is a fundamental open problem if in such a case $\langle p, q \rangle$ can be approximated within a constant factor, see [KT12, Open Problems Section] for further discussions.

For another example, say p is the bases generating polynomial of a matroid M_1 and q is the bases generating polynomial of another matroid M_2. Then, observe that $\langle p, q \rangle$ is exactly the number of common bases of M_1 and M_2. Again, it is a long-standing open problem to approximately count the number of bases at the intersection of two given matroids.

In this lecture we will sketch the proof of the following theorem.

Theorem 6.1. *For any two real stable polynomials $p, q \in \mathbb{R}_{\geq 0}[z_1, \dots, z_n]$,*

$$\sup_{\alpha \in \mathbb{R}^n_{\geq 0}} e^{-\alpha} f(p, q, \alpha) \leq \langle p, q \rangle \leq \sup_{\alpha \in \mathbb{R}^n_{\geq 0}} f(p, q, \alpha) \tag{6.1}$$

where

$$f(p, q, \alpha) := \inf_{y,z>0} \frac{p(y)q(z)}{(yz/\alpha)^\alpha}.$$

Furthermore, in the special case that p, q are multilinear we can improve $e^{-\alpha}$ to $(1-\alpha)^{1-\alpha}$.

Note that, as usual, $(yz/\alpha)^\alpha = \prod_{i=1}^n (y_i z_i/\alpha_i)^{\alpha_i}$. A few remarks are in order:

Convexification: We claim that we can cast this saddle-point max-min problem as a convex program. To see this, first, note that for $\alpha > 0$, with a change of variables $z \leftarrow z/\alpha$, we can write

$$\inf_{z>0} \frac{q(z)}{(z/\alpha)^\alpha} = \inf_{z>0} \frac{p(\alpha_1 z_1, \ldots, \alpha_n z_n)}{z^\alpha}$$

where we crucially use that $z > 0, \alpha > 0$ implies $z/\alpha > 0$. So, we write,

$$\sup_{\alpha>0} \inf_{y>0} \frac{p(y)}{y^\alpha} \inf_{z>0} \frac{q(\alpha z)}{z}$$

Second, as usual, with a change of variables $\tilde{y} = \log y$ and $\tilde{z} = \log z$ we can equivalently solve

$$\sup_{\alpha>0} \inf_{\tilde{y}} \log \frac{p(e^{\tilde{y}})}{e^{\langle \tilde{y}, \alpha \rangle}} \inf_{\tilde{z}} \log \frac{q(\alpha e^{\tilde{z}})}{e^{\langle \tilde{z}, \alpha \rangle}}$$

Note that for any fixed $\alpha > 0$, $\log p(e^{\tilde{y}}) - \langle \tilde{y}, \alpha \rangle$ is a convex function of $\tilde{y}$ and, similarly, for any fixed $\alpha > 0$, $\log q(\alpha e^{\tilde{z}}) - \langle \tilde{z}, \alpha \rangle$ is a convex function of $\tilde{z}$. Finally, we need to say both of these infimums are concave function of α. To see this, note the former is a linear function of α and infimum of linear functions is concave. The latter is a concave function of α (since q is real stable with non-negative coefficient, it is log-concave), and infimum of concave functions is concave.

Applications: In the following section we will see that without loss of generality we can assume $\alpha \in \text{Newt}(p) \cap \text{Newt}(q)$ and so $\|\alpha\| \leq \min\{degp, \deg q\}$. This implies that the above convex program gives $e^{\min\{\deg p, \deg q\}}$ approximation to $\langle p, q \rangle$. As immediate applications observe that when p, q are generating polynomial of two k-DPPs the above theorem gives an e^k approximation to correlation of these two DPPs. To this date, this is the best approximation algorithm known to estimate the correlation of two given DPPs.

When p, q are bases generating polynomials of two matroids and they are real stable (such as graphic matroid or partition matroid), then the above theorem gives an e^n approximation to the number bases at the intersection of the given matroids.

6.1 Intuitions

In this section, we justify the right hand side of (6.1); and in turn we justify why we get to such a complicated program. WLOG assume that polynomials p, q are generating polynomials of distributions μ_p, μ_q. In the last lecture, we explained the dual of the program $\inf_{z>0} \log \frac{p(z)}{z^\alpha}$ (after the change of variables $y_i = \log z_i$) is the max-entropy program:

$$\begin{aligned}
\max \quad & \sum_\kappa \nu(\kappa) \log \frac{\mu_p(\kappa)}{\nu(\kappa)} & \\
\text{s.t.,} \quad & \sum_\kappa \nu(\kappa)\kappa_i = \alpha_i & \forall i, \\
& \sum_\kappa \nu(\kappa) = 1 & \\
& \nu(\kappa) \geq 0 & \forall \kappa.
\end{aligned}$$

Note that since both μ_p and ν are distributions over the support of p, we can write the objective function as $-D_{KL}(\mu_p||\nu)$. In other words, the above program finds a distribution ν with marginals α that is the *closest* to p with respect the KL-divergence distance. As a special case, if α is the marginal vector of μ_p, then the optimum ν will be $\nu = \mu_p$ and the objective function will be 0.

Consequently, the optimum value of the following program is 0:

$$\sup_{\alpha>0}\inf_{z>0}\log\frac{p(z)}{z^\alpha}.$$

Note that if $\alpha\notin\mathrm{Newt}(p)$, then the infimum will be $-\infty$. Otherwise, the optimum is always negative unless α is the marginal vector of μ_p in which case, the optimum is 0.

This implies that

$$\sup_{\alpha>0}\inf_{y,z>0}\frac{p(y)q(z)}{(yz/\alpha)^\alpha}=\sup_{\alpha\in\mathrm{Newt}(p)\cap\mathrm{Newt}(q)}\inf_{y,z>0}\frac{p(y)q(z)}{(yz/\alpha)^\alpha}.$$

Since $\alpha\in\mathrm{Newt}(p)\cap\mathrm{Newt}(q)$ and for any $\kappa\in\mathrm{Newt}(p)$ we have $\|\kappa\|_1\le\deg p$ and for any $\kappa\in\mathrm{Newt}(q)$ we have $\|\kappa\|_1\le\deg q$, it follows that $\|\alpha\|_1\le\min\{\deg p,\deg q\}$.

Note that, in general, the intersection of two Newton polytopes may have new vertices which do not belong to the any of the two so just the fact that α is in the intersection is not enough to argue that α can be written as the convex combination of the vertices. But, it is well-known that the intersection of two matroid base polytopes is indeed the convex hull of the bases at the intersection. This fact generalizes to real stable polynomials and implies that we do not get new vertices.

The rest of the argument will be informal: Consider the distribution ν where for any $\kappa\in\mathrm{Newt}(p)\cap\mathrm{Newt}(q)$, $\nu(\kappa)=\frac{1}{Z}\mu_p(\kappa)\cdot\mu_q(\kappa)$, where Z is the normalizing constant. Indeed $Z=\langle p,q\rangle$. Let α^* be marginals of ν; then

$$\sup_{\alpha\in\mathrm{Newt}(p)\cap\mathrm{Newt}(q)}\inf_{y,z>0}\frac{p(y)q(z)}{(yz/\alpha)^\alpha}\ge\inf_{y,z>0}\frac{p(y)q(z)}{(yz/\alpha^*)^{\alpha^*}}$$

Now, to figure out $\inf p(y)/y^{\alpha^*}$ we need to find the distribution closest to μ_p with marginals α^*. It turns out that ν is the closest such distribution. So,

$$\inf_{y>0}\log\frac{p(y)}{y^{\alpha^*}}=\sum_{\kappa\in\mathrm{supp}(\nu)}\nu(\kappa)\log\frac{\mu_p(\kappa)}{\nu(\kappa)}=\sum_{\kappa\in\mathrm{supp}(\nu)}\nu(\kappa)\log\frac{Z}{\mu_q(\kappa)},$$

and, similarly $\inf_{z>0}\log\frac{q(z)}{z^\alpha}=\sum_{\kappa\in\mathrm{supp}(\nu)}\nu(\kappa)\frac{Z}{\mu_q(\kappa)}$. Therefore,

$$\inf_{y,z>0}\log\frac{p(y)q(z)}{(yz/\alpha^*)^{\alpha^*}}=Z+\log(\alpha^{*\alpha^*})+\sum_{\kappa\in\mathrm{supp}(\nu)}\nu(\kappa)\log\frac{1}{\nu(\kappa)}=Z+\log(\alpha^{*\alpha^*})+\mathcal{H}(\nu)\ge\langle p,q\rangle$$

where $\mathcal{H}(\nu)$ is the *entropy* of ν. The last inequality uses that $\log(\alpha^{*\alpha^*})\le\mathcal{H}$. We will prove this inequality in future lectures.

6.2 Nash Welfare Maximization Problem

In the Nash welfare maximization problem we have a set of n agents $[n]:=\{1,\dots,n\}$ and a set of m items $[m]:=\{1,\dots,m\}$ (where assume that $m\gg n$; we further assume that agent i has value $v_{i,j}\ge 0$ for item j. For an assignment of items to agents $x:[n]\times[m]\to\{0,1\}$ the value of i is defined as

$$v_i(x):=\sum_j x_{i,j}v_{i,j}.$$

The goal is to find an assignment to maximize the geometric mean of the values of all agents, i.e.,

$$\max_x\left(\prod_{i=1}^n v_i(x)\right)^{1/n}.$$

In this section we use Theorem 6.1 to give a constant factor approximation algorithm for this problem.

First, let me discuss a natural convex relaxation for this problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \log \sum_{j} v_{i,j} x_{i,j}, & \\ \text{s.t.,} \quad & \sum_{i=1}^{n} x_{i,j} = 1 & \forall 1 \leq j \leq m \\ & x_{i,j} \geq 0 & \forall i, j. \end{aligned} \tag{6.2}$$

Note that the optimum solution is obviously a feasible solution to the above program, so above is a convex relaxation of the problem. It turns out that the above program has an infinite integrality gap. For example, suppose we have n items where the first one is a BMW which has value 1 for everyone and the rest of them are chocolates and cookies which have value ϵ for everyone where $\epsilon > 0$ is very small. Obviously everyone very much like to get the BMW; but the integral only gets to assign BMW to one person and it gets a value of $\epsilon^{n-1/n}$. The CP fraction solutions assigns $1/n$ fraction of BMW to each agent and receives a value of at least $1/n$. As $\epsilon \to 0$ the value of integral optimum converges to 0.

Before discussing our convex program let us first discuss a natural rounding algorithm. Say $x : [n] \times [m] \to [0,1]$ is a fractional assignment. A natural rounding algorithm assigns item j to agent i with probability $x_{i,j}$; note that since $\sum_i x_{i,j} = 1$ we have a valid probability distribution for every item j. Let $V_i(x)$ be the random variable of the value that i receives under this assignment. Then,

$$\mathbb{E}\left[V_i(x)\right] = \sum_j x_{i,j} v_{i,j}.$$

But, the value of our algorithm is the sum of values of agents, it is rather the product.

Lemma 6.2. *Let*

$$p_x(z_1, \ldots, z_m) \quad = \quad \prod_{i=1}^{n} \sum_j v_{i,j} x_{i,j} z_j, \tag{6.3}$$

Then,

$$\mathbb{E}\left[\prod_i V_i(x)\right] = MAP(p)|_{z=0},$$

where recall that MAP is the operator that keeps all multi-affine monomials of p (and vanishes everything else).

Proof. The lemma simply follows by the fact that any monomial of p that has a square of say z_j (for some j) corresponds to assigning the j-th item to two or more agents. □

Let

$$q(z_1, \ldots, z_m) = e_n(z_1, \ldots, z_m) = \sum_{S \in \binom{m}{n}} z^S.$$

Observe that

$$\text{MAP}(p_x)|_{z=0} = \langle p_x, q \rangle.$$

Now, observe that not only we can use Theorem 6.1 to estimate $\langle p_x, q\rangle$, since $f(p,q,\alpha)$ is a convex function we can change also optimize over x such that $\langle p_x, q\rangle$ is maximized. We use the following program:

$$\begin{aligned} \max \quad & f(p_x, q, \alpha) \\ \text{s.t.,} \quad & \sum_i x_{i,j} = 1 \quad \forall 1 \le j \le m, \\ & x_{i,j} \ge 0 \qquad\qquad \forall i,j. \end{aligned} \tag{6.4}$$

Given an optimal solution x to this program we simply run the above randomized rounding algorithm; by Theorem 6.1 the objective function of the above program gives an $e^{\min\{\deg p, \deg q\}}$ approximation to OPT^n. Since $\deg p = \deg q = n$ we indeed obtain an e approximation to OPT.

Finally, let me remark that $f(p_x, q, \alpha)$ is indeed a concave function (of α, x). This is because p_x is a log-concave function of x.

6.3 Ideas for the Hard Direction

The prove the left side of (6.1) we use the following observations:

Step 1: The statement on general real stable polynomials can be reduced to multilinear ones by the polarization technique. Given a polynomial $p \in \mathbb{R}[z_1, \dots, z_n]$, let $m \ge \max_i \deg_{z_j} p$. The polarization of p, $\pi_m(p) = \mathbb{R}[z_{1,1}, \dots, z_{1,m}, \dots, z_{n,m}]$ is the multilinear polynomial define as follows: For any $1 \le i \le n$ and $1 \le j \le \deg_{z_i} p$ we substitute any occurrence of z_i^j with $\frac{1}{\binom{m}{k}} e_k(z_{i,1}, \dots, z_{i,m})$. Note that this polynomial is *symmetric* under all clonings of z_i and if we specialize all $z_{i,j}$'s with z_i we get back the polynomial p.

Theorem 6.3 (Borcea, Brändén and Liggett [BBL09])**.** *$p \in \mathbb{R}[z_1, \dots, z_n]$ is stable iff $\pi_m(p)$ is stable for all $m \ge \max_i \deg_{z_i} p$.*

Step 2: For any two multilinear polynomials $p \in \mathbb{R}[y_1, \dots, y_n], q \in \mathbb{R}[z_1, \dots, z_n]$,

$$\langle p, q\rangle = \prod_{i=1}^n (1 + \partial_{y_i}\partial_{z_i}) pq|_{y=z=0}.$$

Step 3: Unfortunately, $1 + \partial_{y_i}\partial_{z_i}$ is not stability preserver. So, one can use the following idea: A polynomial $r \in \mathbb{R}[y_1, \dots, z_1, \dots, z_n]$ is bi-stable if $r[y_1, \dots, y_n, -z_1, \dots, -z_n)$ is stable. It is not hard to see if p, q are stable then $r = p \cdot q$ is bi-stable. This is because if $q \in \mathbb{R}_{\ge 0}[z_1, \dots, z_n]$ is stable then $q(-z_1, \dots, -z_n)$ is also stable.

Furthermore, observe that

$$(1 + \partial_{y_i}\partial_{z_i}) r(y_1, \dots, y_n, z_1, \dots, z_n)|_{y_i = z_i = 0} = (1 - \partial_{y_i}\partial_{z_i}) r(y_1, \dots, y_n, -z_1, \dots, -z_n)|_{y_i = z_i = 0}.$$

So, if r is bi-stable so is $(1 + \partial_{y_i}\partial_{z_i}) r|_{y_i = z_i = 0}$. Here is the main lemma:

Lemma 6.4. *For any multilinear bi-stable polynomial $p \in \mathbb{R}_{\ge 0}[y_1, \dots, y_n, z_1, \dots, z_n]$ and any $\alpha \in \mathbb{R}^n_{\ge 0}$,*

$$\prod_{i=1}^n (1 + \partial_{y_i}\partial_{z_i}) p|_{y=z=0} \ge (1-\alpha)^{1-\alpha} \inf_{y,z>0} \frac{p(y,z)}{(yz/\alpha)^\alpha}$$

Step 4: The proof follows of the above lemma by an induction argument similar to Gurvits. The main facts are closure of bi-stable polynomials under $1 + \partial_{y_i}\partial_{z_i}$ and substitution with real numbers. The main non-trivial part is the base case where the above lemma is proven when $n = 1$. In such a case we just need to use the special case of the Borcea-Brändén theorem that we proved in the last lecture that a bi-variate multilinear polynomial $ayz + by + cz + d$ is real stable iff $bc \geq ad$.

References

[BBL09] J. BORCEA, P. BRÄNDÉN, and T. M. LIGGETT. "NEGATIVE DEPENDENCE AND THE GEOMETRY OF POLYNOMIALS". In: *Journal of the American Mathematical Society* 22.2 (2009), pp. 521–567. URL: http://www.jstor.org/stable/40587241 (cit. on p. 6-5).

[KT12] A. Kulesza and B. Taskar. *Determinantal Point Processes for Machine Learning*. Hanover, MA, USA: Now Publishers Inc., 2012 (cit. on p. 6-1).