Partition Functions, Polynomials and Optimization Indian Institute of Science, Bangalore, Jan 13, 2017 Speaker: Nisheeth Vishnoi Real Stable Polynomials and Gurvits' Theorem

# 1 Outline

Recall from the last talk the definition of capacity which was out proxy for computing the permanent of a non-negative matrix.

**Definition 1.1.** Let  $p \in \mathbb{R}[z_1, ..., z_n]$  be any *n*-variate real polynomial. We define its capacity to be

$$\operatorname{Cap}(p) = \inf_{z_1, \dots, z_n > 0} \frac{p(z)}{\prod_{i=1}^n z_i}.$$

In this talk we will prove the following theorem:

**Theorem 1.2** (Gurvits '08). Let A be an  $n \times n$  matrix with non-negative entries. Let  $p_A$  be the corresponding product polynomial and  $(p_A)_{[n]}$  be the coefficient of  $\prod_{i=1}^{n} z_i$ . It holds

$$(p_A)_{[n]} \leq \operatorname{Cap}(p_A) \leq \frac{n^n}{n!} (p_A)_{[n]}.$$

The proof is non-trivial and very general – it relies on the theory of *real stable polynomials* – which we develop first.

# 2 Real Stable Polynomials

We are primarily concerned with univariate and multivariate polynomials  $f(z_1, ..., z_n) \in \mathbb{R}[z_1, ..., z_n]$ . On occasion we may run into polynomials with complex coefficients. Of interest will be zeros of such a polynomial which is always a subset of  $\mathbb{C}^n$ . For a number  $z \in \mathbb{C}$ , its real part is denoted by  $\Re(z)$  and its imaginary part by  $\Im(z)$ . Let  $\mathcal{H} \stackrel{\text{def}}{=} \{z \in \mathbb{C} : \Im(z) > 0\}$  denote the upper-half complex plane.

#### 2.1 Stability

A polynomial  $f(z_1, ..., z_n)$  is said to be *stable* with respect to (w.r.t.) a region  $\Omega \subseteq \mathbb{C}^n$  if no root of f lies in  $\Omega$ . Of particular interest is the region

$$\mathcal{H}^n = \{(z_1, \ldots, z_n) \in \mathbb{C}^n : \forall i, \Im(z_i) > 0\}$$

and polynomials with no root in this region will be referred to as  $\mathcal{H}$ -stable or simply stable. To emphasize the fact that the coefficients of f are all real numbers, we often call such polynomials *real stable*.

### 2.2 Univariate Polynomials – Real Rootedness

When *f* is a univariate polynomial, real stability amounts to saying that all the roots of *f* are real, or *f* is *real-rooted*. This is because of the following simple lemma which states that the complex roots of a univariate polynomial with real coefficients appear as pairs and, hence, if there is a complex root, there would be one with a positive imaginary part.

**Lemma 2.1.** *For*  $f \in \mathbb{R}[z]$ *, if for*  $a, b \in \mathbb{R}$  f(a + ib) = 0*, then* f(a - ib) = 0*.* 

Proof. 
$$f(a+1b) = \sum_i a_i(a+1b)^i = 0 = \sum_i a_i \overline{(a+1b)^i} = \sum_i a_i \overline{(a+1b)^i} = \sum_i a_i (a-1b)^i = f(a-1b).$$

The benefits of being real-rooted. As a simple example of how real-rootedness helps we derive an interesting property for probability distributions whose generating functions are real-rooted. For a probability distribution  $a_0, a_1, \ldots, a_d$  over  $\{0, 1, \ldots, d\}$ , its generating function is defined to be the degree d polynomial  $g(z) \stackrel{\text{def}}{=} \sum_{i=0}^{d} a_i z^i$ . Suppose g(z) is real-rooted. What does this say about the probability distribution itself? Start by observing that if g(z) is real-rooted, then all its roots have to be non-positive as  $a_i \ge 0$  for all i. Thus,  $g(z) = a_d \prod_{i=1}^{d} (z + \alpha_i)$  for nonnegative  $\alpha_i$ s. Let  $p_i \stackrel{\text{def}}{=} \frac{1}{1+\alpha_i}$  so that  $\alpha_i = \frac{1-p_i}{p_i}$ . Since  $\alpha_i \ge 0$ ,  $0 < p_i \le 1$  for all i. Thus,

$$g(z) = a_d \prod_{i=1}^d \left( z + \frac{1 - p_i}{p_i} \right) = a_d \sum_{S \subseteq [d]} z^{|S|} \prod_{i \notin S} \frac{1 - p_i}{p_i} = \frac{a_d}{\prod_{i=1}^d p_i} \sum_{S \subseteq [d]} z^{|S|} \prod_{i \notin S} (1 - p_i) \prod_{i \in S} p_i$$

Note that, since,  $\sum_{i=1}^{d} a_i = 1$ ,

$$g(1) = a_d \prod_{i=1}^d (1 + \alpha_i) = a_d \prod_{i=1}^d \frac{1}{p_i} = 1.$$

Hence,

$$g(z) = \sum_{S \subseteq [d]} z^{|S|} \prod_{i \notin S} (1 - p_i) \prod_{i \in S} p_i = \sum_{k=0}^d z^k \sum_{S \subseteq [d]:|S|=k} \prod_{i \notin S} (1 - p_i) \prod_{i \in S} p_i.$$

Consider a sequence of independent random variables  $Y_1, \ldots, Y_d$  such that each  $Y_i$  is 1 with probability  $p_i$  and 0 with probability  $1 - p_i$ . Further, let  $X \stackrel{\text{def}}{=} \sum_{i=0}^{d} Y_i$  denote the number of 1s obtained if we sample from each  $Y_i$ . Then,  $\Pr[X = k] = \sum_{S \subseteq [d]: |S| = k} \prod_{i \notin S} (1 - p_i) \prod_{i \in S} p_i$ . Thus,

$$g(z) = \sum_{k=0}^{d} z^k \Pr[X = k].$$

In other words, if the generating function of a probability distribution is realrooted, then the distribution corresponds to a sum of independent Bernoulli random variables. Thus, for a real-rooted polynomial with non-negative coefficients, its coefficients are unimodal; this is the content of one of the problems of this week's homework.

Now we present a bit more non-trivial application of real-rootedness which lies at the core of the proof of Gurvits' theorem.

**Lemma 2.2.** Let 
$$f(z) = \sum_{i=0}^{d} a_i z^i$$
 with  $a_i \ge 0$  for all  $i$ , then  $f'(0) \ge \left(\frac{d-1}{d}\right)^{d-1} \inf_{t>0} \frac{f(t)}{t}$ 

*Proof.* Recall that in the univariate case, if the polynomial has coefficients that are non-negative, then all its roots have to be non-positive. Thus, if  $f(z) = \sum_{i=0}^{d} a_i z^i$  is real-rooted with  $a_i \ge 0$  for all i, it can be written as  $a_d \prod_{i=1}^{d} (z + \alpha_i)$  where  $\alpha_i \ge 0$ . For a moment, assume that  $\alpha_i > 0$  for all i. Thus, at any positive z = t, using the AM-GM inequality and the fact that  $\prod_i \alpha_i = \frac{a_0}{a_d}$ , we obtain

$$f(t) = a_d \prod_{i=1}^d (t + \alpha_i) = a_0 \prod_{i=1}^d \left( 1 + \frac{t}{\alpha_i} \right) \le a_0 \left( 1 + \frac{t}{d} \sum_i \frac{1}{\alpha_i} \right)^d = a_0 \left( 1 + \frac{tf'(0)}{a_0 d} \right)^d.$$

Optimizing for *t* via a simple calculus exercise then shows that

$$f'(0) \ge \left(\frac{d-1}{d}\right)^{d-1} \inf_{t>0} \frac{f(t)}{t}$$

Now, if some  $\alpha_i = 0$ , since  $f'(0) = a_1 \ge 0$ , then f(0) = 0, and the above holds trivially.

When is a polynomial real-rooted? Given a polynomial, it is not obvious by looking at its coefficients if it is real-rooted or not. So how would we ever *know* for a polynomial if it is real-rooted or not? How robust is real-rootedness? For instance, if f is real-rooted, so are the polynomials f(cz) for a real number c, and  $z^d \cdot f(1/z)$ . A bit more non-trivially, so is the derivative of f: f'(z). To see this, recall from calculus that between any two roots of f there is exactly one root of f'. Thus, if all the d roots of f are real, then so are all the d - 1 roots of f'. This latter fact is a manifestation of the more general Gauss-Lucas theorem which states that the convex hull of the set of roots of a real (or complex) polynomial f contains the set of roots of f'.

**Theorem 2.3** (Gauss-Lucas). Let  $f \in \mathbb{C}[z]$ , then all the roots of f'(z) can be written as a convex combination of the roots of f(z).

*Proof.* Write  $f(z) = a_d \prod_i (z - \alpha_i)$ . Thus,

$$\frac{f'(z)}{f(z)} = \sum_{i} \frac{1}{z - \alpha_i}.$$

Thus, if  $\beta \in \mathbb{C}$  is such that  $f'(\beta) = 0$  and  $f(\beta) \neq 0$ , then  $\sum_i \frac{1}{\beta - \alpha_i} = 0$ . This implies that  $\sum_i \frac{\overline{\beta} - \overline{\alpha_i}}{|\beta - \alpha_i|^2} = 0$ . Thus, separating  $\overline{\beta}$  out and conjugating, we obtain  $\sum_i p_i \alpha_i$  where  $p_i \stackrel{\text{def}}{=} \frac{\frac{1}{|\beta - \alpha_i|^2}}{\sum_j \frac{1}{|\beta - \alpha_i|^2}}$ . The Gauss-Lucas theorem follows by noticing that  $p_i \geq 0$  for all i and  $\sum_i p_i = 1$ .

#### 2.3 Multivariate Polynomials

Recall that  $f(z_1, ..., z_n)$  is said to be real stable if  $f \in \mathbb{R}[z_1, ..., z_n]$  and no root of it lies in  $\mathcal{H}^n$ . It seems harder to show that a multivariate polynomial is real stable. The first lemma reduces the problem of checking real stability of a multivariate polynomial to checking real-rootedness of a set of univariate polynomials, and turns out to be quite effective.

**Lemma 2.4.** A multivariate polynomial  $f(z_1,...,z_n) \in \mathbb{R}[z_1,...,z_n]$  is stable if and only if for all  $v \in \mathbb{R}^n$  and all  $u \in \mathbb{R}^n_{>0}$ , the univariate polynomial f(v + tu) is real-rooted.

*Proof.* Suppose that f(v + tu) is real-rooted for all  $v \in \mathbb{R}^n$  and all  $u \in \mathbb{R}^n_{>0}$ , but f is not real stable. The latter implies that there is an  $a = (a_1, \ldots, a_n) \in \mathcal{H}^n$  such

that f(a) = 0. Let  $v \stackrel{\text{def}}{=} \Re(a)$  and  $u \stackrel{\text{def}}{=} \Im(a)$ . Since  $a \in \mathcal{H}^n$ ,  $u_i > 0$  for all *i*. But then  $f(a) = f(v + \iota u) = 0$  and, hence,  $\iota$  is a root of  $f(v + \iota u)$  which contradicts the real-rootedness of  $f(v + \iota u)$ .

For the other direction, suppose that there are  $v \in \mathbb{R}^n$  and  $u \in \mathbb{R}^n_{>0}$  and a  $t = t_1 + \iota t_2$  such that f(v + tu) = 0. Since complex roots of a univariate polynomial appear in conjugates (Lemma 2.1), we may assume that  $t_2 > 0$ . Thus, the imaginary part of each component of v + tu is strictly positive contradicting the fact that f is real stable.

Using the lemma above, several multivariate polynomials can be shown to be real stable. Perhaps the simplest such polynomial (which can be seen to be real stable without appealing to the lemma above) is  $\sum_i a_i z_i$  when  $a_i \ge 0$  for all *i*. Since a root of a polynomial that is a product of two polynomials is a root of at least one of those two polynomials, the polynomial  $\prod_j \sum_i a_{ij} z_i$  is also real stable. A bit more non-trivially, the following important class of polynomials arising from determinants can be shown to be real stable.

**Lemma 2.5.** Let  $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$  be positive definite matrices<sup>1</sup> and B be a symmetric  $m \times m$  real matrix. Then the polynomial  $f(z_1, \ldots, z_n) \stackrel{\text{def}}{=} \det(z_1A_1 + \cdots + z_nA_n + B)$  is real stable.

We will derive the proof in one of the exercises. The above lemma can be established in the setting when  $A_i$ s are positive semi-definite (PSD) as opposed to being positive-definite. This is quite useful for applications. However, extending Lemma 2.5 requires the following theorem from complex analysis whose proof is beyond the scope of the talk.

**Theorem 2.6** (Hurwitz). Let  $\{f_k\}_{k\geq 0}$  be a sequence of  $\Omega$ -stable polynomials over  $z_1, \ldots, z_n$  for a connected and open set  $\Omega$  that uniformly converge to a polynomial f over compact subsets of  $\Omega$ . Then f is  $\Omega$ -stable.

To use this theorem for a matrix  $A_i$  which is just guaranteed to be PSD one approximates each  $A_i$  by a sequence of matrices  $A_i + \frac{1}{2^k}I$  which are positive definite and converge to  $A_i$  as k goes to infinity. One can ask if all real stable polynomials arise from such determinants. This is the content of the Generalized Lax Conjecture.

<sup>&</sup>lt;sup>1</sup>Semi-positive definite and positive definite matrices over reals are symmetric.

## 2.4 Closure Properties

What makes the stability theory particularly powerful is that many of the *closure* properties discussed in Section 2.1 hold in the multivariate setting as well. Thus, we can start with real stable polynomials and prove stability for new ones. For us, the key closure properties will be closure under *inversion*, *specialization* and *differentiation*.

**Inversion.** If  $f(z_1, ..., z_n)$  is real stable with the degree of  $z_i$  in f being  $d_i$ , then the polynomial  $f(1/z_1, ..., 1/z_n) \prod_{i=1}^n z_i^{d_i}$  is also real stable. Suppose  $(a_1, ..., a_n) \in \mathcal{H}^n$  be such that  $f(1/a_1, ..., 1/a_n) = 0$ . Since the coefficients of f are real, if  $\overline{f(1/a_1, ..., 1/a_n)} = f(\overline{1/a_1}, ..., \overline{1/a_n}) = 0$ . Since, if the imaginary part of  $a_i$  is positive that of  $1/a_i$  is negative, the imaginary part of  $\overline{1/a_i}$  is positive for each i. This contradicts the real stability of f and establishes our first closure result.

**Specialization.** It is easy to see that if  $f(z_1, ..., z_n)$  is a stable polynomial, then  $f(a, z_2, ..., z_n)$  is also stable if  $\Im(a) > 0$ . However, if  $f(z_1, ..., z_n)$  is real,  $f(a, z_2, ..., z_n)$  may have complex coefficients and, thus, may not be real stable. The following lemma, which relies on Hurwitz's theorem (Theorem 2.6), shows that if  $a \in \mathbb{R}$  then  $f(a, z_2, ..., z_n)$  is real stable.

**Lemma 2.7.** If  $f(z_1, ..., z_n)$  is real stable, then for all  $a \in \mathbb{R}$  (the closure of  $\mathcal{H}$ ),  $f(a, z_2, ..., z_n)$  is also real stable.

*Proof.* If  $a \in \mathcal{H}$ , then the proof follows from the discussion above. Thus, it is sufficient to prove this lemma for  $a \in \mathbb{R}$ . We will only sketch a proof here. Suppose, for sake of contradiction, that  $f(a, a_2, \ldots, a_n) = 0$  with some  $a_j$  such that  $\Im(a_j) > 0$ . It follows from the definition that  $f_k \stackrel{\text{def}}{=} f(a + \iota 2^{-k}, z_2, \ldots, z_n)$  is stable for any k > 0. The lemma now follows from Hurwitz's theorem since  $\lim_{k\to\infty} f_k = f(a, z_2, \ldots, z_n)$  is stable and, being real, is real stable.

**Differentiation.** The next crucial closure property is closure of real stability under taking partial derivatives. Following is some basic notation for partial derivatives of multivariate polynomials. Let  $\partial_i \stackrel{\text{def}}{=} \partial/\partial z_i$ . For a tuple  $\alpha : [n] \mapsto \mathbb{Z}_{\geq 0}$ , let  $\partial^{\alpha} \stackrel{\text{def}}{=} \prod_{i=1}^{n} \partial_i^{\alpha_i}$ .

**Lemma 2.8.** Let f be real stable. Then,  $\partial_1 f$  is also real stable.

Since the choice of the first variable is arbitrary, for any  $\alpha : [n] \mapsto \mathbb{Z}_{\geq 0}$ ,  $\partial^{\alpha} f$  is real stable. Thus, the real stability of  $\partial^{\alpha} f$  follows by an inductive application of this lemma.

*Proof.* Assume on the contrary that  $\partial_1 f$  is not real stable and let  $a = (a_1, a_2, \dots, a_n) \in \mathcal{H}^n$  such that  $\partial_1 f(a_1, a_2, \dots, a_n) = 0$ . Let  $g(z) \stackrel{\text{def}}{=} f(z, a_2, \dots, a_n)$ . If  $g \equiv 0$ , then  $f(a_1, a_2, \dots, a_n) = 0$ , contradicting the stability of f. Hence,  $g \neq 0$ . Since f is real stable, so is g(z) by Lemma 2.7. By the Gauss-Lucas theorem, the roots of g'(z) are in the convex hull of the roots of g(z) and, hence, g'(z) is real stable. Since  $g'(z) = \partial_1 f(z, a_2, \dots, a_n)$ ,  $g'(a_1) = \partial_1 f(a_1, a_2, \dots, a_n) = 0$  by assumption. Thus,  $g'(a_1) = 0$  for  $a_1$  such that  $\Im(a_1) > 0$ , contradicting the stability of g'.

# **3** Proof of Gurvits' Theorem

In this section we prove the following theorem of Gurvits' – Theorem 3.1 follows as a corollary.

**Theorem 3.1** (Gurvits '08). Let  $p(z_1, ..., z_n) \in \mathbb{R}_+[z_1, ..., z_n]$  be a real-stable polynomial, with d being the maximum degree of any variable, and  $p_{[n]}$  be the coefficient of  $\prod_{i=1}^n z_i$  in p. Then

$$\operatorname{Cap}(p) \le \left(\frac{d}{d-1}\right)^{(d-1)n} p_{[n]}.$$

To see how Theorem 3.1 follows from this result, consider the polynomial  $p_A(z_1, ..., z_n) \stackrel{\text{def}}{=} \prod_{i=1}^n \sum_{j=1}^n a_{ij} z_j$ , and note that  $p_A$  is real stable, has degree d = n and

$$\left(\frac{n}{n-1}\right)^{(n-1)n} \le e^n$$

*Proof.* It follows from a repeated application of Lemma 2.7 and Lemma 2.8 that, for any  $1 \le i < n$ , the polynomial

$$g_i(z_1,\ldots,z_i) \stackrel{\text{def}}{=} \partial^{(i+1,\ldots,n)} p(z_1,\ldots,z_i,0,\ldots,0)$$

is real stable. Note that  $g_0 = \partial^{(1,...,n)} p(0,...,0) = p_{[n]}$ . Since all coefficients of p are nonnegative, it follows from Lemma 2.2 and Lemma 2.7 that, for any fixed positive  $b_1, \ldots, b_{i-1}$ ,

$$g_{i-1}(b_1,\ldots,b_{i-1}) = \partial_i g_i(b_1,\ldots,b_{i-1},0) \ge \left(\frac{d_i-1}{d_i}\right)^{d_i-1} \frac{g_i(b_1,\ldots,b_i)}{b_i},$$

where  $d_i$  is the degree of the polynomial  $g_i(b_1, ..., b_{i-1}, z_i)$ . Fixing  $s_1, s_2, ..., s_{i-1}$ , let  $s_i$  be defined to be

$$\arg\inf_{t>0}\frac{g_i(s_1,\ldots,s_{i-1},t)}{t}$$

Thus, applying the above inequality for i = 0 to n - 1 and letting  $d \stackrel{\text{def}}{=} \max_{i=1}^{n} d_i$ , we obtain that  $p_{[n]} = g_0$ , which is at least

$$\left(\frac{d-1}{d}\right)^{d-1} \frac{g_1(s_1)}{s_1} \ge \dots \ge \left(\frac{d-1}{d}\right)^{(d-1)n} \frac{g_n(s_1, \dots, s_n)}{\prod_{i=1}^n s_i} = \left(\frac{d-1}{d}\right)^{(d-1)n} \frac{p(s_1, \dots, s_n)}{\prod_{i=1}^n s_i}.$$
  
Since  $\frac{p(s_1, \dots, s_n)}{p(s_1, \dots, s_n)} \ge \inf_{i=1}^n s_i = \sum_{i=1}^n \frac{p(s_1, \dots, s_n)}{p(s_1, \dots, s_n)}$ , we obtain

Since  $\frac{p(s_1,...,s_n)}{\prod_{i=1}^n s_i} \ge \inf_{b_1 > 0,...,b_n > 0} \frac{p(b_1,...,b_n)}{\prod_{i=1}^n b_i}$ , we obtain

$$p_{[n]} \ge \left(\frac{d-1}{d}\right)^{(d-1)n} \operatorname{Cap}(p).$$

Rearranging, this completes the proof of the theorem.