Abstract. These are condensed notes for the course, updated as the course progresses. While the main content of the lectures is reflected here, some details and, occasionally, topics might be missing.

1. Matching polynomial

(1.1) Weighted graphs, matchings and matching polynomials. We consider a (finite, undirected, without multiple edges or loops) weighted graph \( G = (V, E; a) \) with set \( V \) of vertices, set \( E \) of edges and non-negative real weights \( a : E \to \mathbb{R}_+ \) on the edges. A matching in \( G \) is a set \( e_1, \ldots, e_k \) of pairwise disjoint edges of weight \( a(e_1) \cdots a(e_k) \). We consider the empty set a matching of weight 1. The matching polynomial is the univariate polynomial

\[
\text{Mat}_G(\lambda) = 1 + \sum_{\text{matching}} \lambda^k a(e_1) \cdots a(e_k).
\]

Together with \( \text{Mat}_G(\lambda) \) it is convenient to consider the matching defect polynomial

\[
q_G(\lambda) = \lambda^{|V|} \text{Mat}_G \left( -\frac{1}{\lambda^2} \right).
\]

It is easy to see that \( q_G(\lambda) \) is indeed a monic polynomial of degree \( |V| \).

One gets a lot of mileage from the following simple recurrence:

\[
(1.1.1) \quad \text{Mat}_G(\lambda) = \text{Mat}_{G-v}(\lambda) + \lambda \sum_{u \in V : \{v,u\} \in E} a_{vu} \text{Mat}_{G-v-u}(\lambda),
\]

where \( v \in V \) is a vertex, \( G - v \) is the graph obtained from \( G \) by deleting \( v \) and all incident edges, \( G - v - u \) is the graph obtained from \( G \) by deleting vertices \( v \) and \( u \)
and all incident edges and $a_{uv}$ is the weight on the edge $\{v, u\}$. Consequently, for the matching defect polynomial, we get

$$q_G(\lambda) = \lambda q_{G-v}(\lambda) - \sum_{u \in V : \{v, u\} \in E} a_{uv} q_{G-v-u}(\lambda).$$

(1.1.2)

Our first main result is the Heilmann - Lieb Theorem (1972).

(1.2) Theorem. The roots of $\text{Mat}_G(\lambda)$ are negative real.

The proof uses the concept of interlacing polynomials.

(1.3) Interlacing polynomials. Let $f$ be a polynomial of degree $n$ and $n$ distinct real roots $\alpha_1 < \alpha_2 < \ldots < \alpha_n$ and let $g$ be a polynomial of degree $n-1$ and $n-1$ distinct real roots $\beta_1 < \beta_2 < \ldots < \beta_{n-1}$. We say that $g$ interlaces $f$ if

$$\alpha_i < \beta_i < \alpha_{i+1} \quad \text{for} \quad i = 1, \ldots, n-1.$$ 

A canonical example is supplied by the Rolle’s Theorem: $f$ is a polynomial with distinct real roots and $g = f'$. Here is a useful lemma.

(1.4) Lemma.

1. Let $f$ and $g_1, \ldots, g_k$ be real polynomials such that each $g_k$ interlaces $f$ and the highest degree terms of $g_1, \ldots, g_k$ have the same sign. Let $a_1, \ldots, a_k$ be non-negative real, not all equal to 0, and let $g = a_1 g_1 + \ldots + a_k g_k$. Then $g$ interlaces $f$.

2. Suppose that $g$ interlaces $f$ and the highest terms of $f$ and $g$ have the same sign. Then for any real $a$, the polynomial $f$ interlaces the polynomial $h(\lambda) = (\lambda - a)f(\lambda) - g(\lambda)$.

Sketch of Proof. Let $\alpha_1 < \alpha_2 < \ldots < \alpha_n$ be the roots of $f$. To prove Part (1), we note that each $g_k$ changes its sign once on each interval between consecutive roots of $f$ and all $g_k$ change the sign in the same way. Hence $g$ also changes its sign once on each interval between consecutive roots of $f$ and $g$ interlaces $f$.

To prove Part (2) we note that $h$ changes its sign once on each interval between consecutive roots of $f$ and also once on the interval $(-\infty, \alpha_1)$ and once on the interval $(\alpha_n, +\infty)$.

Now we are ready to prove the Heilmann - Lieb Theorem.

(1.5) Sketch of proof of Theorem 1.2. It suffices to prove that the roots of $q_G(\lambda)$ are real. Since the roots of a non-zero polynomial depend continuously on the polynomial, it suffices to prove that the roots of $q_G(\lambda)$ are real, assuming that $G$ is the complete graph on $n$ vertices and that $a_{uv} > 0$ for all $\{u, v\} \in E$.

We use (1.1.2) to prove that by induction on $n$ that the roots of $q_G(\lambda)$ are real and, moreover, for every $v \in V$, the polynomial $q_{G-v}(\lambda)$ interlaces $q_G(\lambda)$. This is checked immediately if $|V| = 2$. Suppose that $|V| > 2$ and let $v \in V$ be a vertex. By the induction hypothesis, for any $u \in V$ such that $\{v, u\} \in E$ the polynomial $q_{G-v-u}(\lambda)$ interlaces $q_{G-v}(\lambda)$. Using (1.1.2) and Lemma 1.4, we conclude that $q_{G-v}(\lambda)$ interlaces $q_G(\lambda)$.
(1.6) Examples (Problems). In what follows, \( a(e) = 1 \) for all \( e \in E \).

(1.6.1) Interval. Let \( V = \{1, \ldots, n\} \) and let \( E \) consist of all pairs \( \{i, i + 1\} \) for \( i = 1, \ldots, n - 1 \). Then \( q_G(\lambda) = U_n(\lambda/2) \), where \( U_n \) is the Chebyshev polynomial of the second kind, defined by

\[
U_n(x) = \frac{\sin((n+1)\theta)}{\sin\theta} \text{ where } \cos \theta = x.
\]

(1.6.2) Circle. Let \( V = \{1, \ldots, n\} \) and let \( E \) consist of all pairs \( \{i, i + 1\} \) for \( i = 1, \ldots, n - 1 \) and \( \{n, 1\} \). Then \( q_G(\lambda) = 2T_n(\lambda/2) \), where \( T_n \) is the Chebyshev polynomial of the first kind defined by

\[
T_n(x) = \cos n\theta \text{ where } \cos \theta = x.
\]

(1.6.3) Complete graph. Let \( V = \{1, \ldots, n\} \) and let \( E \) consist of all pairs \( \{i, j\} \) for \( 1 \leq i < j \leq n \). Then \( q_G(\lambda) = H_n(\lambda) \), where \( H_n \) is the Hermite polynomial defined by

\[
H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}.
\]

(1.6.4) Complete bipartite graph. Let \( V = \{1, \ldots, 2n\} \) and let \( E \) consist of all pairs \( \{i, j\} \) where \( 1 \leq i \leq n \) and \( n + 1 \leq j \leq 2n \). Then \( q_G(\lambda) = (-1)^n n! L_n(\lambda^2) \), where \( L_n \) is the Laguerre polynomial defined by

\[
L_n(\lambda) = \frac{e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^n).
\]

(1.7) Bounds on the roots (Problems). Given non-negative weights \( a : E \to \mathbb{R}_+ \) on the edges of \( G \), let us define

\[
\beta_G = \max_{v \in V} \sum_{u \in V} a_{vu}.
\]

If \( q_G(\lambda) = 0 \) then \( |\lambda| \leq 2\sqrt{\beta_G} \) and if \( \text{Mat}_G(\lambda) = 0 \) then \( \lambda \leq -\frac{1}{4\beta_G} \). The bounds are pretty useful, though not optimal.

Suppose that \( a_e = 1 \) for all \( e \in E \) and that \( \Delta(G) \) is the largest degree of a vertex of \( G \), so that \( \beta_G = \Delta(G) \). If \( q_G(\lambda) = 0 \) then \( |\lambda| \leq 2\sqrt{\Delta(G) - 1} \) and if \( \text{Mat}_G(\lambda) = 0 \) then \( \lambda \leq -\frac{1}{4(\Delta(G) - 1)} \). These bounds are asymptotically optimal.

Here is a useful property (log-concavity) of the coefficients of a real polynomial with all roots real.
(1.8) **Theorem.** Suppose that the roots of the real polynomial

\[ p(x) = \sum_{k=0}^{n} a_k x^k \]

are also real. Let

\[ b_k = \frac{a_k}{\binom{n}{k}} \quad \text{for} \quad k = 0, \ldots, n. \]

Then

\[ b_k^2 \geq b_{k+1} b_{k-1} \quad \text{for} \quad k = 1, \ldots, n - 1. \]

If \( a_k \geq 0 \) for \( k = 0, \ldots, n \) then the above inequalities imply that

\[ a_k^2 \geq a_{k+1} a_{k-1} \quad \text{for} \quad k = 1, \ldots, n - 1. \]

**Sketch of Proof.** By the repeated application of the Rolle’s Theorem, the roots of the polynomial

\[ q(x) = \frac{d^{k-1}}{dx^{k-1}} p(x) \]

of \( \deg q \leq n - k + 1 \) are real. Then the roots of the polynomial

\[ r(x) = x^{n-k+1} q \left( \frac{1}{x} \right) \]

of \( \deg r \leq n - k + 2 \) are real. Then the roots of the quadratic polynomial

\[ s(x) = \frac{d^{n-k-1}}{dx^{n-k-1}} r(x) \]

are real. Hence the discriminant of \( s(x) \) is non-positive, which implies the desired inequality. \( \Box \)

2. **Correlation decay**

(2.1) **The probability space of matchings.** Let \( G = (V, E) \) be a graph (for now, we let \( a(e) = 1 \) for all edges) and let \( \lambda \geq 0 \) be a parameter. We consider the set of all matchings in \( G \) as a (finite) probability space, with probability of a matching of \( k \) edges equal to \( \lambda^k / \operatorname{Mat}_G(\lambda) \) for \( k = 0, 1, \ldots. \) We are interested in the following question: given a vertex \( v \) of \( G, \) what is the probability that a random matching contains \( v? \)

We introduce a metric on \( G, \) where \( \operatorname{dist}(u, v) \) is the smallest number of edges of \( G \) in a path connecting \( u \) and \( v \) (we allow \( \operatorname{dist}(u, v) = +\infty \) if \( u \) and \( v \) lie in different components of the graph). Let \( \Delta > 1 \) be an upper bound on the degrees of vertices of \( G. \) Our main goal is to establish the correlation decay result of Bayati, Gamarnik, Katz, Nair and Tetali (2007) that says that within an additive error \( 0 < \epsilon < 1, \) the probability that a random matching contains \( v \) depends only on the structure of \( G \) in an \( m\)-neighborhood of \( v \) for some \( m = m(\epsilon, \lambda, \Delta). \) One can choose, roughly,

\[ m = O \left( \lambda \Delta \ln \frac{1}{\epsilon} \right). \]
(2.2) Probabilities and recursions. For a set \( S \subset V \) and \( v \in V \setminus S \), let \( p_{S,v} = p_{S,v}(\lambda) \) denote the probability that a random matching does not contain \( v \) provided it contains no vertices from \( S \). Denoting by \( G - S \) the graph obtained from \( G \) by deleting all vertices in \( S \) together with incident edges, from (1.1.1) we get

\[
\text{Mat}_{G - S}(\lambda) = \text{Mat}_{G - S - v} + \lambda \sum_{u \in V \setminus S: \{v,u\} \in E} \text{Mat}_{G - S - v - u}(\lambda),
\]

which we rewrite as

\[
\frac{\text{Mat}_{G - S - v}(\lambda)}{\text{Mat}_{G - S}(\lambda)} = \left(1 + \lambda \sum_{u \in V \setminus S: \{v,u\} \in E} \frac{\text{Mat}_{G - S - v - u}(\lambda)}{\text{Mat}_{G - S - v}(\lambda)}\right)^{-1}
\]

and further interpret as

\[
(2.2.1) \quad p_{S,v} = \left(1 + \lambda \sum_{u \in V: \{v,u\} \in E} p_{S \cup v,u}\right)^{-1}.
\]

(2.3) Theorem. Let us consider the set of all non-negative vectors \( x = (x_{S,v}) \) with coordinates parameterized by pairs \( S \subset V \) and \( v \in V \setminus S \) and let \( T \) be a transformation of this set defined by

\[
y = T(x) \quad \text{where} \quad y_{S,v} = \left(1 + \lambda \sum_{u \in V \setminus S: \{v,u\} \in E} x_{S \cup v,u}\right)^{-1}.
\]

(1) If

\[
\frac{1}{1 + \lambda \Delta} \leq x_{S,v} \leq 1 \quad \text{for all} \quad S, v
\]

then

\[
\frac{1}{1 + \lambda \Delta} \leq y_{S,v} \leq 1 \quad \text{for all} \quad S, v;
\]

(2) Suppose that \( y' = T(x') \) and \( y'' = T(x'') \). Then

\[
\max_{S,v} \left| \ln y'_{S,v} - \ln y''_{S,v} \right| \leq \frac{\lambda \Delta}{1 + \lambda \Delta} \max_{S,v} \left| \ln x'_{S,v} - \ln x''_{S,v} \right|.
\]
Sketch of Proof. Part (1) is obvious and Part (2) follows since if we write $T$ in the logarithmic coordinates, that is, if $x = e^{-\xi}$ and $y = e^{-\eta}$ so that

$$\eta = \ln \left(1 + \lambda \sum_{u \in V \setminus S} e^{-\xi_{S \cup u,v}}\right)$$

then

$$\sum_{u \in V \setminus S} \left| \frac{\partial \eta_{S,v}}{\partial \xi_{S \cup v,u}} \right| \leq \frac{\lambda \Delta}{1 + \lambda \Delta}.$$ 

□

Theorem 2.3 asserts that $T$ is a contraction and while (2.2.1) implies that $(p_{S,v})$ is the (necessarily unique) fixed point of $T$. Let us define $x$ by $x_{S,v} = 1$ for all $S \subset V$ and $v \in V \setminus S$ and let $y = T^m(x)$. Then, by Theorem 2.3, we have

$$\max_{S,v} |\ln y_{S,v} - \ln p_{S,v}| \leq \left(\frac{\lambda \Delta}{1 + \lambda \Delta}\right)^m \ln(1 + \lambda \Delta).$$

In particular, if we choose

$$m \geq (1 + \lambda \Delta) \ln \frac{1}{\epsilon} + \ln \ln(1 + \lambda \Delta),$$

then

$$|\ln y_{S,v} - \ln p_{S,v}| \leq \epsilon \quad \text{for all} \quad S \subset V \quad \text{and} \quad v \in V \setminus S.$$ 

Next, we note that computing $y_{\emptyset,v}$, we need to access only the coordinates $(S, u)$, where the the vertices of $S$ and $u$ lie in the $m$-neighborhood of $v$. Hence we conclude:

(2.4) Corollary. Let $p_v$ be the probability that a random matching does not contain vertex $v$. For any

$$m \geq (1 + \lambda \Delta) \ln \frac{1}{\epsilon} + \ln \ln(1 + \lambda \Delta),$$

up to an additive error of $0 < \epsilon < 1$, the value of $\ln p_v$ is determined by the $m$-neighborhood of $v$ in $G$.

(2.5) Remarks. In fact, Bayati, Gamarnik, Katz, Nair and Tetali (2007) established a stronger bound, by proving that $T^2$ is a contraction with a factor of roughly

$$1 - \frac{1}{\sqrt{1 + \lambda \Delta}}.$$ 

We can incorporate non-negative weights $a_e$ on the edges of $G$, roughly by replacing $\Delta$ by

$$\max_{v \in V} \sum_{u \in V: \{v, u\} \in E} a_{vu}$$

throughout.
3. Graphs of large girth

In what follows, we consider the set of all matchings in a graph $G = (V, E)$ as a probability space, with probability of a matching with $k$ edges equal $\lambda^k / \text{Mat}_G(\lambda)$ for some fixed $\lambda \geq 0$. We are going to exploit the correlation decay phenomenon.

(3.1) Tree $\mathbb{T}^k_n$. Let $k \geq 2$. We define the rooted $k$-tree $\mathbb{T}^k_n$ with $n$ levels as follows: the root (at level 0) is connected to $k - 1$ vertices (descendants) at level 1, each vertex at level 1 is connected to the root and $k - 1$ vertices at level 2, etc., with each vertex at level $m$ connected to one vertex at level $m - 1$ (ancestor) and $k - 1$ vertices at level $m + 1$ (descendants). Finally, each vertex at level $n$ (leaf) is connected only to its ancestor at level $n - 1$.

(3.2) Lemma. Let $p_n$ be the probability that the root of $\mathbb{T}^k_n$ is not covered by a random matching. Then

$$\lim_{n \to \infty} p_n = \frac{\sqrt{1 + 4\lambda(k - 1)} - 1}{2\lambda(k - 1)}.$$

Sketch of Proof. That the limit, say $p$, exists follows from Corollary 2.4, since the $m$-neighborhoods of the root of $\mathbb{T}^k_n$ look the same for $n \geq m$. The recursion (2.2.1) becomes

$$p_n = (1 + \lambda(k - 1)p_{n-1})^{-1},$$

which gives the quadratic equation for $p$:

$$p = (1 + \lambda(k - 1)p)^{-1},$$

solving which we get the formula. □

(3.3) Definitions. A graph $G = (V, E)$ is called $k$-regular if every vertex of $G$ has $k$ neighbors. The girth of $G$ is the smallest integer $m \geq 3$ such that $G$ has an $m$-cycle $\{v_1, \ldots, v_m\}$, where $\{v_k, v_{k+1}\}$ for $k = 1, \ldots, m - 1$ and $\{v_m, v_1\}$ are edges of $G$. If $G$ has no cycles, we say that $\text{gr} G = +\infty$.

(3.4) Lemma. Let $G_n$ be a sequence of $k$-regular graphs such that $\text{gr} G_n \to +\infty$. Let us pick a vertex $v_n$ of $G_n$ and let $p_n$ be the probability that $v_n$ is not contained in a random matching. Then

$$\lim_{n \to \infty} p_n = \frac{2k - 2}{k\sqrt{1 + 4\lambda(k - 1)} + k - 2}.$$ 

Moreover, for any $\lambda_0 > 0$, the convergence is uniform over all $0 \leq \lambda \leq \lambda_0$ and vertices $v_n$ of $G_n$.

Sketch of Proof. That the limit, say $q$, exists follows from Corollary 2.4, since for any $m$, the $m$-neighborhood of $v_n$ looks identical for all sufficiently large $n$. By (2.2.1), we get the equation

$$q = (1 + \lambda kp)^{-1},$$

where $p$ is the limit in Lemma 3.2. □
(3.5) Lemma. Let \( G = (V, E) \) be a graph. For a vertex \( v \in V \), let \( p_v \) be the probability that \( v \) is not covered by a random matching. Then

\[
\lambda \frac{d}{d\lambda} \ln \operatorname{Mat}_G(\lambda) = \frac{1}{2} \sum_{v \in V} (1 - p_v).
\]

*Sketch of Proof.* Since \( 1 - p_v \) is the probability that a random matching contains \( v \), the right hand side of the formula is half of the expected number of vertices covered by a random matching, that is, the expected number of edges in a random matching.

On the other hand,

\[
\operatorname{Mat}_G(\lambda) = \sum_{k=0}^{\infty} (\text{the number of } k\text{-matchings in } G) \lambda^k,
\]

so that

\[
\lambda \frac{d}{d\lambda} \operatorname{Mat}_G(\lambda) = \sum_{k=0}^{\infty} k (\text{the number of } k\text{-matchings in } G) \lambda^k
\]

and

\[
\lambda \frac{d}{d\lambda} \ln \operatorname{Mat}_G(\lambda) = \frac{\lambda \operatorname{Mat}_G'(\lambda)}{\operatorname{Mat}_G(\lambda)} = \sum_{k=0}^{\infty} k (\text{the number of } k\text{-matchings in } G) \frac{\lambda^k}{\operatorname{Mat}_G(\lambda)},
\]

is also the expected number of \( k\)-matchings in \( G \).

(3.6) Theorem. Let \( G_n = (V_n, E_n) \) be a sequence of \( k\)-regular graphs such that \( \operatorname{gr} G_n \to +\infty \). Then

\[
\lim_{n \to \infty} \frac{1}{|V_n|} \ln \operatorname{Mat}_{G_n}(\lambda) = \frac{k}{2} \ln \left( \frac{1 + \sqrt{1 + 4\lambda k - 4\lambda}}{2} \right) - \frac{k - 2}{2} \ln \left( \frac{k\sqrt{1 + 4\lambda k - 4\lambda + k - 2}}{2k - 2} \right).
\]

*Sketch of Proof.* Combining Lemma 3.4 and Lemma 3.5, we conclude that

(3.6.1) \[
\lim_{n \to \infty} \frac{1}{|V_n|} \frac{d}{d\lambda} \ln \operatorname{Mat}_{G_n}(\lambda) = \frac{1}{2\lambda} - \frac{1}{\lambda} \frac{k - 1}{k\sqrt{1 + 4\lambda(k - 1) + k - 2}}
\]

and that the convergence is uniform on any interval \( \lambda_0 < \lambda < \lambda_1 \) for any \( 0 < \lambda_0 < \lambda_1 \). Moreover, the right hand side of (3.6.1) is regular at \( \lambda = 0 \), it is

\[
\frac{k}{2} - \frac{(2k - 1)k\lambda}{2} + O(\lambda^2) \quad \text{for } \lambda \approx 0,
\]

so we can integrate and obtain

\[
\lim_{n \to \infty} \frac{1}{|V_n|} \ln \operatorname{Mat}_{G_n}(\lambda) = \int_0^\lambda \left( \frac{1}{2\tau} - \frac{1}{\tau} \frac{k - 1}{k\sqrt{1 + 4\lambda(k - 1) + k - 2}} \right) d\tau,
\]

and the proof follows. \( \square \)
4. Lifts and bipartite graphs

Our next goal is to prove a theorem of Csikvári (2014):

**(4.1) Theorem.** Let $G = (V, E)$ be a $k$-regular bipartite graph. Then
\[
\frac{\ln \text{Mat}_G(\lambda)}{|V|} \geq \frac{k}{2} \ln \left( \frac{1 + \sqrt{1 + 4\lambda k - 4\lambda}}{2} \right) - \frac{k - 2}{2} \ln \left( \frac{k\sqrt{1 + 4\lambda k - 4\lambda} + k - 2}{2k - 2} \right)
\]
for $\lambda \geq 0$.

The proof of Theorem 4.1 hinges on the construction of a 2-lift of a graph.

**(4.2) 2-lift of a graph.** Let $G = (V, E)$ be a graph. For every vertex $v \in V$, we introduce a pair $v^+$ and $v^-$ of vertices, and for every edge $\{u, v\} \in E$, we introduce either the pair $\{u^+, v^+\}$ and $\{u^-, v^-\}$ of edges, or the pair $\{u^+, v^-\}$ and $\{u^-, v^+\}$ of edges (we have a choice). The resulting graph $H$ is called a 2-lift of $G$. If $G$ is a bipartite $k$-regular graph then $H$ is a bipartite $k$-regular graph with twice as many vertices. Similarly, we define a 2-lift of a weighted graph, by copying the weight of each edge of $G$ on its liftings.

**(4.3) Lemma.** If $H$ is a 2-lift of a bipartite graph $G$ then
\[
\text{Mat}_H(\lambda) \leq \text{Mat}_G^2(\lambda) \quad \text{for all} \quad \lambda \geq 0.
\]
The same inequality holds for weighted graphs (with non-negative weights on the edges).

*Sketch of Proof.* Let $\hat{G}$ be the trivial 2-lift of $G$, consisting of two disconnected copies of $G$ (one with vertices $v^+$ and the other with vertices $v^-$ for $v \in V$). Then $\text{Mat}_{\hat{G}}(\lambda) = \text{Mat}_G^2(\lambda)$, so we need to prove that $\text{Mat}_H(\lambda) \leq \text{Mat}_G(\lambda)$. Let $e_1, \ldots, e_k$ be a matching in $H$ and let $W$ be the multiset of edges of $G$ obtained from $e_1, \ldots, e_k$ by the natural projection $v^+, v^- \mapsto v$ (as we can obtain the same edge more than once, $W$ is a multiset). It suffices to prove that given such a $W$, the contribution (sum of monomials) to $\text{Mat}_{\hat{G}}(\lambda)$ of all matchings in $\hat{G}$ that project onto $W$ is at least as large as the contribution of all matchings in $H$ that project onto $W$.

First, we note that it suffices to check the last statement when $W$ is connected. Second, each vertex belongs to at most two edges of $W$. Therefore, assuming that $W$ is connected, it is of the following 4 types:

a) $W$ is an edge, b) $W$ is a double edge, c) $W$ is a path, and d) $W$ is a cycle, in which case $W$ has to be an even cycle since $G$ is bipartite.

In a), there are two matchings of one edge each that project onto $W$, so the contributions to $\text{Mat}_H(\lambda)$ and $\text{Mat}_{\hat{G}}(\lambda)$ are equal. In b), there is one matching of two edges projected onto $W$, so the contributions to $\text{Mat}_H(\lambda)$ and $\text{Mat}_{\hat{G}}(\lambda)$ are
equal. In c), the inverse image of $W$ consists of two vertex-disjoint paths, so there are two matchings that project onto $W$ and hence the contributions to $\text{Mat}_H(\lambda)$ and $\text{Mat}_{\hat{G}}(\lambda)$ are equal. In d) the inverse image of $W$ consists of either two cycles of the same length $|W|$, in which case there are exactly two matchings projecting onto $W$ (recall that $|W|$ is even) or a cycle of length $2|W|$, in which case there are no matchings projecting onto $W$ (recall that $|W|$ is even). Hence we conclude that in all cases the contribution to $\text{Mat}_{\hat{G}}(\lambda)$ is at least as big as the contribution to $\text{Mat}_H(\lambda)$. □

The following lemma is due to Linial.

**Lemma.** Let $G$ be a graph. Then there exists a sequence $G_n$ of graphs such that $G_0 = G$, $G_{n+1}$ is a 2-lift of $G_n$ for $n \geq 0$ and $\text{gr } G_n \rightarrow +\infty$ as $n \rightarrow \infty$.

**Sketch of Proof.** Clearly, if $H$ is a 2-lift of $G$ then $\text{gr } G \geq \text{gr } G$. It suffices to prove that if $\text{gr } G = k$ and $\hat{G}$ contains $g$ cycles of length $k$ then there is a 2-lift of $H$ of $G$ which contains fewer than $g$ cycles of length $k$. Let us consider a random 2-lift: independently for each edge $\{u, v\}$ of $G$, we pick the pair $\{u^+, v^+\}$ and $\{u^-, v^-\}$ with probability $1/2$ or the pair $\{u^+, v^-\}$ and $\{u^-, v^+\}$ with probability $1/2$ as edges of $H$. We claim that the expected number of $k$-cycles in $H$ is $g$. Indeed, since the inverse image of every path in $G$ consists of two vertex-disjoint paths in $H$, with probability $1/2$, we have two $k$-cycles in $H$ projecting onto a given $k$-cycle in $G$. Since with positive probability we obtain a trivial lift $\hat{G}$ which has $2g$ cycles of length $k$, there ought to be a 2-lift that contains fewer than $g$ cycles of length $k$. □

**(4.5) Sketch of Proof of Theorem 4.1.** Using Lemma 4.4, we construct a sequence of graphs $G_n = (V, E_n)$, where $G_0 = G$, $G_{n+1}$ is a 2-lift of $G_n$ and $\text{gr } G_n \rightarrow +\infty$. We observe that each $G_n$ is bipartite, so by Lemma 4.3, for any $\lambda \geq 0$, the sequence

$$\frac{\ln \text{Mat}_{G_n}(\lambda)}{|V_n|}$$

is non-increasing. Besides, each $G_n$ is $k$-regular and the proof follows by Theorem 3.6. □

**(4.6) Perfect matchings.** A matching in $G = (V, E)$ is called perfect if it covers all vertices of $V$. We obtain the following corollary of Theorem 4.1:

**(4.7) Corollary.** Let $G = (V, E)$ be a $k$-regular bipartite graph with $|V| = 2n$ vertices. Then there are at least

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

perfect matchings in $G$. 

10
**Sketch of Proof.** Let \( N \) be the number of perfect matchings in \( G \). Assuming that \( N > 0 \), we get
\[
\text{Mat}_G(\lambda) = N\lambda^n (1 + o(1)) \quad \text{as} \quad \lambda \to +\infty.
\]
Therefore,
\[
\ln N = \lim_{\lambda \to +\infty} \ln \text{Mat}_G(\lambda) - n \ln \lambda,
\]
where the limit is \(-\infty\) if \( N = 0 \). The bound now follows from Theorem 4.1. \( \square \)

5. \( \mathbb{H} \)-stable polynomials and capacity

(5.1) **Definitions.** A complex polynomial \( p(z_1, \ldots, z_n) \) is called \( \mathbb{H} \)-stable if
\[
p(z_1, \ldots, z_n) \neq 0 \quad \text{whenever} \quad \Im z_1 > 0, \ldots, \Im z_n > 0.
\]
Given a polynomial \( p(x_1, \ldots, x_n) \) with non-negative real coefficients, we define its capacity by
\[
\text{cap} p = \inf_{x_1 > 0, \ldots, x_n > 0} \frac{p(x_1, \ldots, x_n)}{x_1 \cdots x_n}.
\]
Our goal is to prove the following theorem due to Gurvits (2008).

(5.2) **Theorem.** Let \( p(x_1, \ldots, x_n) \) be an \( \mathbb{H} \)-stable polynomial with non-negative real coefficients. For \( k = n, \ldots, 0 \), let us define polynomials \( p_k(x_1, \ldots, x_k) \) by
\[
p_n = p \quad \text{and} \quad p_k(x_1, \ldots, x_k) = \frac{\partial}{\partial x_{k+1}} p_{k+1}(x_1, \ldots, x_{k+1}) \bigg|_{x_{k+1} = 0}
\]
for \( k = n - 1, \ldots, 0 \). In particular, \( p_0 \) is the coefficient of \( x_1 \cdots x_n \) in \( p \). Suppose that the degree of \( p_k \) in \( x_k \) does not exceed \( d_k \) for some integer \( d_k \geq 0 \). Then
\[
p_0 \geq \left( \prod_{k=1}^n \left( \frac{d_k - 1}{d_k} \right)^{d_k - 1} \right) \text{cap} p,
\]
where we agree that
\[
\left( \frac{d_k - 1}{d_k} \right)^{d_k - 1} = 1 \quad \text{if} \quad d_k = 1 \quad \text{or} \quad d_k = 0.
\]

The proof of Theorem 5.2 hinges on the following three lemmas.

(5.3) **Lemma [Gauss-Lucas Theorem].** If \( p(z) \) is a non-constant complex polynomial with roots \( z_1, \ldots, z_d \in \mathbb{C}, d \geq 1 \), then every root \( w \) of \( p'(z) \) can be written as \( w = \alpha_1 z_1 + \ldots + \alpha_d z_d \) for some real \( \alpha_1, \ldots, \alpha_d \geq 0 \) such that \( \alpha_1 + \ldots + \alpha_d = 1 \).
Sketch of Proof. Without loss of generality, we assume that \( w \neq z_i \) for \( i = 1, \ldots, d \) and that \( f \) is monic, so that

\[
f(z) = \prod_{j=1}^{d} (z - z_j).
\]

Then

\[
0 = f'(w) = \sum_{j=1}^{d} \prod_{k \neq j} (w - z_k) \quad \text{and hence} \quad \sum_{j=1}^{d} \prod_{k \neq j} (w - z_k) = 0.
\]

Multiplying both sides of the last identity by \( f(w) \), we obtain

\[
\sum_{j=1}^{d} (w - z_j) \prod_{k \neq j} |w - z_k|^2,
\]

from which

\[
\alpha_j = \frac{\prod_{k \neq j} |w - z_k|^2}{\sum_{i=1}^{d} \prod_{k \neq i} |w - z_k|^2} \quad \text{for} \quad j = 1, \ldots, d.
\]

(5.4) Lemma.

1. Let \( f_m : \mathbb{C}^n \to \mathbb{C}, m = 1, 2, \ldots \) be a sequence of \( \mathbb{H} \)-stable polynomials and let \( f : \mathbb{C}^n \to \mathbb{C} \) be yet another polynomial such that \( f_m \to f \) uniformly on compact subsets of \( \mathbb{C}^n \). Then \( f \) is either \( \mathbb{H} \)-stable or identically 0.

2. Let \( f(z_1, \ldots, z_n) \) be an \( \mathbb{H} \)-stable polynomial and let

\[
g(z_1, \ldots, z_{n-1}) = f(z_1, \ldots, z_{n-1}, 0).
\]

Then \( g \) is either \( \mathbb{H} \)-stable or identically 0.

3. Let \( f(z_1, \ldots, z_n) \) be an \( \mathbb{H} \)-stable polynomial and let

\[
g(z_1, \ldots, z_n) = \frac{\partial}{\partial z_n} f(z_1, \ldots, z_n).
\]

Then \( g \) is either \( \mathbb{H} \)-stable or identically 0.

Sketch of Proof. The proof of Part (1) follows from a theorem of Hurwitz, which claims that if \( \Omega \subset \mathbb{C}^n \) is a connected open set and \( f_m \) is a sequence of functions analytic in \( \Omega \) such that \( f_m(z) \neq 0 \) for all \( z \in \Omega \) and all \( m \) and \( f_m \to f \) uniformly on compact subsets of \( \Omega \) then either \( f(z) \neq 0 \) for all \( z \in \Omega \) or \( f \equiv 0 \) on \( \Omega \). The proof of Hurwitz theorem reduces to that for \( n = 1 \) (consider a section of \( \Omega \) through
a given point by a complex line identified with \( \mathbb{C} \), while for \( n = 1 \) it follows from
the Rouche Theorem.

To prove Part (2), consider a sequence
\[
g_m(z_1, \ldots, z_{n-1}) = f(z_1, \ldots, z_{n-1}, im^{-1}).
\]
Then \( g_m \) are \( \mathbb{H} \)-stable and \( g_m \to g \) uniformly on compact sets in \( \mathbb{C}^{n-1} \), so the
proof follows from Part (1).

To prove Part (3), let \( d \) be the degree of \( f \) in \( z_n \). If \( d = 0 \) then \( g \equiv 0 \). Hence we
assume that \( d > 0 \), in which case we write
\[
f(z_1, \ldots, z_n) = \sum_{k=0}^{d} z_k h_k(z_1, \ldots, z_{n-1}),
\]
where \( h_d(z_1, \ldots, z_{n-1}) \neq 0 \). Let us consider a sequence \( f_m \) of polynomials,
\[
f_m(z_1, \ldots, z_n) = m^{-d} f(z_1, \ldots, z_{n-1}, m z_n).
\]
Then \( f_m \) are \( \mathbb{H} \)-stable and
\[
f_m(z_1, \ldots, z_n) \to z_d h_d(z_1, \ldots, z_{n-1})
\]
uniformly on compact subsets of \( \mathbb{C}^n \). It follows from Part (1) that the polynomial
\( z_d h_d(z_1, \ldots, z_{n-1}) \) is \( \mathbb{H} \)-stable and hence
\[
h_d(z_1, \ldots, z_{n-1}) \neq 0 \quad \text{whenever} \quad \Im z_1 > 0, \ldots, \Im z_{n-1} > 0.
\]
Let us fix arbitrary \( z_1, \ldots, z_{n-1} \) such that \( \Im z_1 > 0, \ldots, \Im z_{n-1} > 0 \) and let us
consider a univariate polynomial
\[
p(z) = f(z_1, \ldots, z_{n-1}, z).
\]
Hence \( p(z) \) is a polynomial of degree \( d > 0 \). Since \( f \) is \( \mathbb{H} \)-stable, we have \( \Im z \leq 0 \)
for every root \( z \) of \( p \). By Lemma 5.3, for every root \( w \) of \( p' \) we have \( \Im w \leq 0 \) so that
\[
p'(z) = g(z_1, \ldots, z_{n-1}, z) \neq 0 \quad \text{whenever} \quad \Im z > 0,
\]
which completes the proof of Part (3). \( \square \)

(5.5) Lemma. Suppose that \( R(t) \) is a polynomial with non-negative real coefficients and real roots such that \( \deg R \leq d \). Then
\[
R'(0) \geq \left( \frac{d - 1}{d} \right)^{d-1} \inf_{t > 0} \frac{R(t)}{t},
\]
where we agree that
\[
\left( \frac{d-1}{d} \right)^{d-1} = 1 \quad \text{if} \quad d = 1 \quad \text{or} \quad d = 1.
\]

Sketch of Proof. It is not hard to see that the function
\[
x \mapsto \left( \frac{x-1}{x} \right)^{x-1} \quad \text{for} \quad x \geq 1
\]
is decreasing, so without loss of generality, we may assume that \( \deg R = d \).
If \( \deg R(t) \leq 1 \) then \( R(t) = a + bt \) for some \( a, b \geq 0 \) and
\[
R'(0) = b = \inf_{t > 0} \frac{R(t)}{t},
\]
so the desired inequality holds. Hence we assume that \( d \geq 2 \).
If \( R(0) = 0 \) then
\[
\inf_{t > 0} \frac{R(t)}{t} = R'(0),
\]
(the infimum is obtained as \( t \to 0^+ \)) and the desired inequality holds as well.
Hence we assume that \( R(0) > 0 \). Scaling, if necessary, we assume that \( R(0) = 1 \).
Then we can write
\[
R(t) = \prod_{i=1}^{d} \left( 1 - \frac{t}{\alpha_i} \right),
\]
where \( \alpha_1, \ldots, \alpha_d < 0 \) are the roots of \( R(t) \). Denoting \( a_i = -\alpha_i^{-1} \), we obtain
\[
R(t) = \prod_{i=1}^{d} (1 + a_i t) \quad \text{for some} \quad a_1, \ldots, a_d > 0.
\]
We have
\[
R'(0) = \sum_{i=1}^{d} a_i
\]
and using the arithmetic-geometric mean inequality, we obtain
\[
R(t) \leq \left( 1 + \frac{a_1 + \ldots + a_d t}{d} \right)^d = \left( 1 + \frac{R'(0)}{d} t \right)^d \quad \text{for} \quad t > 0.
\]
Therefore,
\[
(5.5.1) \quad \inf_{t > 0} \frac{R(t)}{t} \leq \inf_{t > 0} t^{-1} \left( 1 + \frac{R'(0)}{d} t \right)^d.
\]
Since \( R'(0) > 0 \) and \( d \geq 2 \), the infimum in the right hand side of (5.5.1) is attained at a critical point of \( t \). It is not hard to show that

\[
t_0 = \frac{d}{(d-1)R'(0)}
\]

is the unique critical point and hence

\[
\inf_{t>0} \frac{R(t)}{t} \leq t_0^{-1} \left( 1 + \frac{R'(0)}{d-t_0} \right)^d = \left( \frac{d}{d-1} \right)^{d-1} R'(0)
\]

and the proof follows.

\( \square \)

(5.6) **Sketch of proof of Theorem 5.2.** It suffices to prove that

\[
p_{k-1}(x_1, \ldots, x_k) \geq \left( \frac{d_k - 1}{d_k} \right)^{d_k-1} \inf_{x_k>0} \frac{p_k(x_1, \ldots, x_{k-1}, x_k)}{x_k}
\]

for all \( x_1, \ldots, x_{k-1} > 0 \),

since combining (5.6.1) for \( k = 1, 2, \ldots, n \), we get the desired inequality. By Parts (2) and (3) of Lemma 5.4, each polynomial \( p_k \) is either \( \mathbb{H} \)-stable or identically 0. If \( p_k \equiv 0 \) then \( p_{k-1} \equiv 0 \) and (5.6.1) follows. Hence we assume that \( p_k \) is \( \mathbb{H} \)-stable.

Given \( x_1, \ldots, x_{k-1} > 0 \), let us define a univariate polynomial \( R(t) \) by

\[
R(t) = p_k(x_1, \ldots, x_{k-1}, t).
\]

Hence \( R(t) \) is a polynomial with non-negative real coefficients with \( \deg R \leq d_k \) and we claim that all roots of \( R(t) \) are real. Indeed, suppose that \( \alpha \pm \beta i \) is a pair of complex conjugate roots of \( R(t) \) for some \( \beta > 0 \). By continuity of roots, we conclude that for a sufficiently small \( \epsilon > 0 \), the univariate polynomial

\[
z \mapsto p_k(x_1 + \epsilon i, \ldots, x_k + \epsilon i, z)
\]

has a root with \( \Im z > 0 \), which contradicts the \( \mathbb{H} \)-stability of \( p_k \).

Now, by Lemma 5.5, we have

\[
R'(0) \geq \left( \frac{d_k - 1}{d_k} \right)^{d_k-1} \inf_{t>0} \frac{R(t)}{t},
\]

which is (5.6.1).  \( \square \)
6. COROLLARIES FOR THE PERMANENT

(6.1) Permanent. Let $A = (a_{ij})$ be an $n \times n$ matrix (real, complex, or over an arbitrary field). The permanent of $A$ is

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

where $S_n$ is the symmetric group of the permutations of $\{1, \ldots, n\}$. If $a_{ij} \in \{0, 1\}$ for all $i$ and $j$ then per $A$ is the number of perfect matchings in the bipartite graph on $n + n$ vertices with biadjacency matrix $A$: we have $a_{ij} = 1$ if and only if vertex $i$ on one (left) side of the graph is connected to vertex $j$ on the other side (right) of the graph by an edge.

Our goal is to prove the following result known as the van der Waerden conjecture (originally proved by Falikman and Egorychev in 1981). The proof below is due to Gurvits (2008).

(6.2) Theorem. Let $A = (a_{ij})$ be an $n \times n$ doubly stochastic matrix, that is, non-negative real matrix with all row and column sums equal to 1. Then

$$\text{per } A \geq \frac{n!}{n^n}.$$

Sketch of Proof. Given an $n \times n$ matrix $A = (a_{ij})$, we define a polynomial

$$p_A(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij}x_j \right).$$

It is not hard to see that the coefficient $p_0$ of $x_1 \cdots x_n$ in $p_A$ is per $A$. Moreover, if $A$ is a non-negative real matrix with no zero rows then $p_A$ is $\mathbb{H}$-stable: if $\Im z_1 > 0, \ldots, \Im z_n > 0$ then

$$\Im \left( \sum_{j=1}^{n} a_{ij}z_j \right) = \sum_{j=1}^{n} a_{ij}\Im z_j > 0$$

and hence $p_A(z_1, \ldots, z_n) \neq 0$.

Next, we claim that if $A$ is a doubly stochastic matrix then $\text{cap } p_A = 1$. Indeed,

$$p_A(1, \ldots, 1) = \prod_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij} \right) = 1.$$
since the row sums of $A$ are 1 and hence $\text{cap} p_A \leq 1$. On the other hand, for any $x_1, \ldots, x_n > 0$, applying the arithmetic-geometric mean inequality, we have

$$p_A (x_1, \ldots, x_n) = \prod_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij} x_j \right) \geq \prod_{i=1}^{n} \left( \prod_{j=1}^{n} x_j^{a_{ij}} \right) = \prod_{j=1}^{n} x_j,$$

since the column sums of $A$ are 1. Hence $\text{cap} p_A \geq 1$.

Starting with $p_n = p_A$, let us define polynomials $p_k$ as in Theorem 5.2. Then $\deg p_k \leq k$, so we can choose $d_k = k$. Applying Theorem 5.2, we obtain

$$\text{per} A \geq \left( \prod_{k=1}^{n} \left( \frac{k-1}{k} \right)^{k-1} \right) \text{cap} p_A = \frac{n!}{n^n}.$$

In fact, if $A = (a_{ij})$ is an $n \times n$ doubly stochastic matrix and $\text{per} A = n!/n^n$ then $a_{ij} = 1/n$ for all $i, j$ (exercise).

As Gurvits (2008) noticed, the bound can be sharpened if $A$ has few non-zero entries.

(6.3) Theorem. Let $A = (a_{ij})$ be an $n \times n$ doubly stochastic matrix with at most $m_k \geq 1$ non-zero entries in the $k$-th column for $k = 1, \ldots, n$. Then

$$\text{per} A \geq \prod_{k=1}^{n} \left( \min \{m_k, k\} - 1 \right)^{\min \{m_k, k\} - 1}$$

with the usual agreement that the $k$-th factor is 1 if $\min \{m_k, k\} = 1$.

Sketch of Proof. We construct the polynomial $p_A$ and polynomials $p_k$ as in the proof of Theorem 6.2. Then the degree of $p_k$ in $x_k$ does not exceed $\min \{m_k, k\}$ and the proof follows from Theorem 5.2. □

(6.4) Perfect matchings in regular bipartite graphs. If $G = (V, E)$ is a $k$-regular bipartite graph with $n + n$ vertices then the number of perfect matchings in $G$ is $\text{per} A = k^n \text{per}(k^{-1}A)$, where $A$ is the biadjacency matrix of $G$. We note that $k^{-1}A$ is a doubly stochastic matrix and hence by Theorem 6.3 the number of perfect matchings in $G$ is at least

$$k^n \left( \frac{k-1}{k} \right)^{(k-1)(n-k+1)} \prod_{j=1}^{k-1} \left( \frac{j-1}{j} \right)^{j-1},$$

which is a stronger bound than the one of Corollary 4.7. In particular, if $k = 2$ then (6.4.1) transforms into

$$2^n \cdot \frac{1}{2^{n-1}} = 2,$$

which is of course sharp, and if $k = 3$ then (6.4.1) transforms into

$$6 \left( \frac{4}{3} \right)^{n-3},$$

which was established first by Voorhoeve (1979) by a different method.
7. Ramifications

We prove yet another theorem of Gurvits (2015).

(7.1) Theorem. Let \( p(x_1, \ldots, x_n) \) be an \( \mathbb{H} \)-stable polynomial with non-negative real coefficients. Suppose that the degree of \( p \) in \( x_k \) does not exceed \( d_k \) for \( k = 1, \ldots, n \), where \( d_k \) are positive integers. Let \( r_1, \ldots, r_n \) be positive integers such that \( r_k \leq d_k \) for \( k = 1, \ldots, n \). Then the coefficient of \( x_1^{r_1} \cdots x_n^{r_n} \) in \( p \) is at least as big as

\[
\left( \prod_{k=1}^{n} \frac{r_k^{r_k} (d_k - r_k)^{d_k - r_k} d_k!}{r_k! (d_k - r_k)! d_k!} \right) \inf_{x_1>0, \ldots, x_n>0} \frac{p(x_1, \ldots, x_n)}{x_1^{r_1} \cdots x_n^{r_n}}.
\]

We note that the coefficient of \( x_1^{r_1} \cdots x_n^{r_n} \) in \( p \) trivially does not exceed

\[
\inf_{x_1>0, \ldots, x_n>0} \frac{p(x_1, \ldots, x_n)}{x_1^{r_1} \cdots x_n^{r_n}}.
\]

Sketch of Proof of Theorem 7.1. We deduce the result from Theorem 5.2. Let us introduce a polynomial \( q \) in variables \( y_{11}, \ldots, y_{1r_1}, \ldots, y_{21}, \ldots, y_{2r_2}, \ldots, y_{n1}, \ldots, y_{nr_n} \) by

\[
q(\ldots, y_{k1}, \ldots, y_{kr_k}, \ldots) = p\left(\ldots, \frac{y_{k1} + \ldots + y_{kr_k}}{r_k}, \ldots\right)
\]

that is, \( q \) is obtained from \( p \) by the substitution

\[
(7.1.1) \quad x_k = \frac{y_{k1} + \ldots + y_{kr_k}}{r_k} \quad \text{for} \quad k = 1, \ldots, n.
\]

Clearly, the coefficients of \( q \) are non-negative real and \( q \) is \( \mathbb{H} \)-stable. Moreover, the coefficient of \( x_1^{r_1} \cdots x_n^{r_n} \) in \( p \) is the coefficient of \( \cdots y_{k1} \cdots y_{kr_k} \cdots \) in \( q \) multiplied by \( \prod_{k=1}^{n} \frac{r_k^{r_k}}{r_k!} \). Next, we observe that the degree of \( q \) in variables \( \{y_{k1}, \ldots, y_{kr_k}\} \) does not exceed \( d_k \) and that the degree of

\[
\frac{\partial^j}{\partial y_{k1} \cdots \partial y_{kj}} q
\]

in \( \{y_{k1}, \ldots, y_{kr}\} \) does not exceed \( d_k - j \). Finally, we observe that

\[
(7.1.2) \quad \inf_{\ldots y_{k1}>0, \ldots y_{kr_k}>0} \frac{q(\ldots, y_{k1}, \ldots, y_{kr_k}, \ldots)}{\cdots y_{k1} \cdots y_{kr_k} \cdots} \geq \inf_{x_1>0, \ldots, x_n>0} \frac{p(x_1, \ldots, x_n)}{x_1^{r_1} \cdots x_n^{r_n}}.
\]

18
Indeed, given ... $y_k > 0$, ..., $y_{kr_k} > 0$, ..., we define $x_k$ by (7.1.1) and notice that by the arithmetic-geometric mean inequality, we have

$$x_k^{r_k} \geq y_k \cdots y_{kr_k},$$

from which (7.1.2) follows.

Combining all the above observations, we deduce from Theorem 5.2 that the coefficient of $x_1^{r_1} \cdots x_n^{r_n}$ in $p$ is at least as big as

$$\left( \prod_{k=1}^{n} \frac{r_k}{r_k!} \right) \left( \prod_{k=1}^{n} \prod_{j=1}^{r_k} \left( \frac{d_k - j}{d_k - j + 1} \right)^{d_k - j} \right) \inf_{x_1 > 0, \ldots, x_n > 0} \frac{p(x_1, \ldots, x_n)}{x_1^{r_1} \cdots x_n^{r_n}},$$

as required. \(\square\)

(7.3) Counting subgraphs with prescribed degrees. Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be positive integer vectors and let $\Sigma(R, C)$ be the set of all $m \times n$ matrices $D = (d_{ij})$ with row sums $R$, column sums $C$ and 0-1 entries. Gale-Ryser Theorem (1957) states a convenient necessary and sufficient condition for $\Sigma(R, C)$ to be non-empty: assuming that

$$m \geq c_1 \geq c_2 \geq \ldots \geq c_n > 0 \quad \text{and} \quad n \geq r_i > 0 \quad \text{for} \quad i = 1, \ldots, m,$$

we must have

$$\sum_{i=1}^{m} \min \{r_i, k\} \geq \sum_{j=1}^{k} c_j \quad \text{for} \quad k = 1, \ldots, n$$

and

$$\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j.$$

Given a non-negative real $m \times n$ matrix $W = (w_{ij})$, we define a polynomial

$$\text{Fl}(R, C; W) = \sum_{D \in \Sigma(R, C)} \prod_{D=(d_{ij})} \prod_{i,j} w_{ij}^{d_{ij}},$$

where we agree that $0^0 = 1$, so that Fl is indeed a polynomial in $w_{ij}$ (and remains a continuous function as $w_{ij} \to 0+$. If $w_{ij} \in \{0, 1\}$ then Fl($R, C; W$) is the number of subgraphs with degrees $r_1, \ldots, r_m, c_1, \ldots, c_n$ of vertices in the bipartite graph on $m + n$ vertices with biadjacency matrix $W$, or, equivalently, the number of 0-1 flows in the bipartite network with biadjacency matrix $W$, demands $r_1, \ldots, r_m$ and supplies $c_1, \ldots, c_n$.

Let us define a polynomial

$$p_W(x_1, \ldots, x_m; y_1, \ldots, y_n) = \prod_{1 \leq i \leq m, 1 \leq j \leq n} (x_i + w_{ij}y_j).$$
Then \( \text{Fl}(R, C; W) \) is the coefficient of \( x_1^{n-r_1} \cdots x_m^{n-r_m} y_1^{c_1} \cdots y_m^{c_m} \) in \( p_W \) and hence

\[
\text{Fl}(R, C; W) \leq \inf_{x_1, \ldots, x_m > 0 \atop y_1, \ldots, y_n > 0} \frac{p_W(x_1, \ldots, x_m; y_1, \ldots, y_n)}{x_1^{n-r_1} \cdots x_m^{n-r_m} y_1^{c_1} \cdots y_m^{c_m}}.
\]

If \( \exists x_i, \exists y_j > 0 \) then \( \exists (x_i + w_{ij} y_j) > 0 \), from which it follows that \( p_W \) is \( \mathbb{H} \)-stable. Applying Theorem 7.1, we get

\[
\text{Fl}(R, C; W) \geq \left( \prod_{i=1}^m \frac{r_i! (n-r_i)!}{r_i! (n-r_i)! n!} \right) \left( \prod_{j=1}^n \frac{c_j! (m-c_j)!}{c_j! (m-c_j)! m^m} \right) \times \inf_{x_1, \ldots, x_m > 0 \atop y_1, \ldots, y_n > 0} \frac{p_W(x_1, \ldots, x_m; y_1, \ldots, y_n)}{x_1^{n-r_1} \cdots x_m^{n-r_m} y_1^{c_1} \cdots y_m^{c_m}}.
\]

### 8. Capacity, Convexity, (Log) Concavity

**8.1 Capacity.** Given a polynomial \( p(x_1, \ldots, x_n) \) with non-negative real coefficients and a multi-index \( R = (r_1, \ldots, r_n) \) of non-negative integers, we define

\[
\text{cap}_R(p) = \inf_{x_1 > 0, \ldots, x_n > 0} \frac{p(x_1, \ldots, x_n)}{x_1^{r_1} \cdots x_n^{r_n}}.
\]

This extends our Definition 5.1 of capacity.

Our first observation is that the capacity can be found by minimizing a convex function on \( \mathbb{R}^n \), which is easy. Making the substitution \( x_i = e^{t_i} \), we write

\[
(8.1.1) \quad \ln \text{cap}_R(p) = \inf_{t_1, \ldots, t_n} \ln p(e^{t_1}, \ldots, e^{t_n}) - \sum_{i=1}^n r_i t_i.
\]

**8.2 Lemma.** The function

\[
(t_1, \ldots, t_n) \mapsto \ln p(e^{t_1}, \ldots, e^{t_n})
\]

is convex.

**Sketch of Proof.** It suffices to prove that the restriction of the function onto any affine line \( t_i = a_i t + b_i \) is convex, that is, that the function \( t \mapsto \ln g(t) \), where

\[
g(t) = \sum_{i=1}^m \alpha_i e^{\lambda_i t}
\]

where \( \alpha_1, \ldots, \alpha_m > 0 \) is convex. Indeed,

\[
(\ln g(t))' = \frac{g'(t)}{g(t)} \quad \text{and} \quad (\ln g(t))'' = \frac{g''(t)g(t) - g'(t)g'(t)}{g^2(t)},
\]
where 

\[ g'(t) = \sum_{i=1}^{m} \alpha_i \lambda_i e^{\lambda_i t} \quad \text{and} \quad g''(t) = \sum_{i=1}^{m} \alpha_i \lambda_i^2 e^{\lambda_i t}, \]

so that

\[ g''(t)g(t) - g'(t)g'(t) = \sum_{i,j=1}^{m} \alpha_i \alpha_j \lambda_i^2 e^{(\lambda_i + \lambda_j)t} - \sum_{i,j=1}^{m} \alpha_i \alpha_j \lambda_i \lambda_j e^{(\lambda_i + \lambda_j)t} \]

\[ = \sum_{\{i, j\} \atop i \neq j} (\lambda_i - \lambda_j)^2 \alpha_i \alpha_j e^{(\lambda_i + \lambda_j)t} \geq 0 \]

and the proof follows. \[ \square \]

Our next observation is that the function \( R \mapsto - \log \text{cap } R(p) \) is concave.

\textbf{(8.3) Lemma.} Let \( R_1, \ldots, R_k; R \) be multi-indices of \( n \) non-negative integers such that

\[ R = \sum_{i=1}^{k} \alpha_i R_i \]

for some \( \alpha_1, \ldots, \alpha_k \geq 0 \) such that

\[ \sum_{i=1}^{k} \alpha_i = 1. \]

Then

\[ \text{cap}_R(p) \geq \prod_{i=1}^{k} (\text{cap}_{R_i}(p))^{\alpha_i}. \]

\textit{Sketch of Proof.} By (8.1.1), the function \( R \mapsto - \log \text{cap}_R(p) \) is the infimum of a family (indexed by \( t_1, \ldots, t_n \)) of affine functions in \( R \). Hence \( R \mapsto \log \text{cap}_R(p) \) is concave. \( \square \)

For \( R = (r_1, \ldots, r_n) \), let \( a_R \) be the coefficient of \( x_1^{r_1} \cdots x_n^{r_n} \) in the polynomial \( p(x_1, \ldots, x_n) \). Clearly, \( a_R \leq \text{cap}_R(p) \). As follows from Lemma 8.3, if \( \text{cap}_R(p) \) approximates \( a_R \) reasonably well for all possible multi-indices \( R \), then the function \( R \mapsto a_R \) is approximately log-concave (that is, the function \( R \mapsto \log a_R \) is approximately concave). The following lemma shows that the converse is approximately true.

\textbf{(8.4) Lemma.} Let \( p(x_1, \ldots, x_n) \) be a polynomial with non-negative coefficients. For a multi-index \( M = (m_1, \ldots, m_n) \) of non-negative integers, let \( a_M \) be the coefficient of \( x_1^{m_1} \cdots x_n^{m_n} \) in \( p \). Suppose that the function \( M \mapsto \log a_M \) is concave,
that is, whenever \( M = \alpha_1 M_1 + \ldots + \alpha_k M_k \) for some \( \alpha_1, \ldots, \alpha_k \geq 0 \) such that \( \alpha_1 + \ldots + \alpha_k = 1 \), we have

\[
a_M \geq \prod_{i=1}^{k} a_{M_i}^{\alpha_i}.
\]

Then for every multi-index \( R = (r_1, \ldots, r_n) \) of non-negative integers, one can find \( x_1, \ldots, x_n > 0 \) such that

\[
a_R x_1^{r_1} \cdots x_n^{r_n} \geq a_M x_1^{m_1} \cdots x_n^{m_n} \quad \text{for all} \quad M = (m_1, \ldots, m_n).
\]

In particular,

\[
a_R \geq \frac{\text{cap}_R(p)}{\text{number of monomials in } p}.
\]

**Sketch of proof.** Let us consider the set \( S \subset \mathbb{R}^{n+1} \) of points \((M, \ln a_M)\), where \( M \) ranges over all multi-indices of monomials in \( p \). Let us choose any \( \gamma > \ln a_R \) and consider a ray \( \{(R, \beta) : \beta \geq \gamma\} \) in \( \mathbb{R}^{n+1} \). We claim that the ray and the convex hull (the set of all convex combinations of points from \( S \)) are disjoint. Indeed, if

\[
R = \sum_{i=1}^{k} \alpha_i M_i \quad \text{where} \quad \alpha_1, \ldots, \alpha_k \geq 0 \quad \text{and} \quad \sum_{i=1}^{k} \alpha_i = 1,
\]

then

\[
\beta \geq \gamma > \ln a_R \geq \sum_{i=1}^{k} \alpha_i \ln a_{M_i}
\]

and the point \((R, \beta)\) is not a convex combination of points from \( S \).

Since the ray and the convex hull of \( S \) do not intersect, we can separate them by an affine hyperplane, that is, we can find real \( t_1, \ldots, t_n; t_{n+1} \) such that

\[
t_{n+1} \beta + \sum_{i=1}^{k} t_i r_i \geq t_{n+1} \ln a_M + \sum_{i=1}^{k} t_i m_i
\]

for \( R = (r_1, \ldots, r_n) \), all \( \beta \geq \gamma > \ln a_R \) and all \( M = (m_1, \ldots, m_n) \). Taking the separating hyperplane sufficiently generic, we can assume that \( t_{n+1} \neq 0 \), from which it follows that \( t_{n+1} > 0 \), from which we can further scale it to \( t_{n+1} = 1 \). Then we have

\[
\ln a_R + \sum_{i=1}^{k} t_i r_i \geq \ln a_M + \sum_{i=1}^{k} t_i m_i
\]

for \( R = (r_1, \ldots, r_n) \) and all \( M = (m_1, \ldots, m_n) \). Letting \( x_i = e^{t_i} \), we complete the proof.
9. Capacity and matrix scaling

The following theorem was proved by Sinkhorn (1964).

**9.1 Theorem.** Let \( A = (a_{ij}) \) be an \( n \times n \) matrix with positive entries. Then there exists an \( n \times n \) doubly stochastic matrix \( B = (b_{ij}) \) and positive \( \lambda_1, \ldots, \lambda_n; \mu_1, \ldots, \mu_n \) such that

\[
a_{ij} = \lambda_i \mu_j b_{ij} \quad \text{for all} \quad i, j.
\]

Given \( A \), the matrix \( B \) is unique and numbers \( \lambda_1, \ldots, \lambda_n; \mu_1, \ldots, \mu_n \) are unique up to a rescaling \( \lambda_i \rightarrow \lambda_i \tau, \mu_j \rightarrow \mu_j \tau^{-1} \) for some \( \tau > 0 \).

**Sketch of proof.** Let us consider a function

\[
g(X) = \sum_{i,j=1}^{n} x_{ij} \ln \frac{x_{ij}}{a_{ij}}
\]

on the set of all \( n \times n \) non-negative matrices \( X = (x_{ij}) \) and let \( P_n \) be the set of all \( n \times n \) doubly stochastic matrices \( X \). Then \( g \) attains its minimum on \( P_n \) at some doubly stochastic matrix \( B = (b_{ij}) \) (since \( g \) is strictly convex, the minimum \( B \) is unique). We observe that

\[
\frac{\partial g}{\partial x_{ij}}(X) = 1 + \ln \frac{x_{ij}}{a_{ij}},
\]

and if \( x_{ij} = 0 \) then the derivative is understood as the right derivative and its value is \(-\infty\). It follows then that we must have \( b_{ij} > 0 \) for all \( i, j \) since otherwise, letting \( B(t) = (b_{ij}(t)) \), where \( b_{ij} = (1 - t)b_{ij} + t/n \), we obtain \( g(B(t)) < g(B) \) for all sufficiently small \( t > 0 \). Since \( b_{ij} > 0 \) for all \( i, j \), we conclude that \( B \) is a critical point of \( g \) on the affine subspace of all \( n \times n \) matrices \( X = (x_{ij}) \) satisfying

\[
\sum_{j=1}^{n} x_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad \sum_{i=1}^{n} x_{ij} = 1 \quad \text{for} \quad j = 1, \ldots, n.
\]

Therefore, there exist Lagrange multipliers \( \alpha_1, \ldots, \alpha_n \) and \( \beta_1, \ldots, \beta_n \) such that

\[
\ln \frac{b_{ij}}{a_{ij}} = \alpha_i + \beta_j \quad \text{for all} \quad i, j.
\]

We let

\[
\lambda_i = e^{-\alpha_i} \quad \text{and} \quad \mu_j = e^{-\beta_j} \quad \text{for all} \quad i, j.
\]

The original Sinkhorn’s proof was based on iterated scaling of the matrix first to row sums 1, then to column sums 1, then to row sums 1 again, etc., and proving that the process converges.

As follows from Theorem 9.1, the function \( f(A) = \lambda_1 \cdots \lambda_n \mu_1 \cdots \mu_n \) is a well-defined function on positive \( n \times n \) matrices.
(9.2) Connections to capacity. Given an \( n \times n \) positive matrix \( A = (a_{ij}) \), we define the polynomial
\[
p_A (x_1, \ldots, x_n) = \prod_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij} x_j \right),
\]
as in the proof of Theorem 6.2. Let \( B = (b_{ij}) \) be an \( n \times n \) doubly stochastic matrix, and let \( \lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n \) be positive reals such that \( a_{ij} = \lambda_i \mu_j b_{ij} \). Then
\[
\text{cap} \ p_A = \inf_{x_1, \ldots, x_n > 0} \frac{p_A (x_1, \ldots, x_n)}{x_1 \cdots x_n} = \lambda_1 \cdots \lambda_n \mu_1 \cdots \mu_n \text{cap} \ p_B = \lambda_1 \cdots \lambda_n \mu_1 \cdots \mu_n,
\]
see the proof of Theorem 6.2.

Let us indeed define a function \( f \) on \( n \times n \) positive matrices \( A \) by \( f (A) = \lambda_1 \cdots \lambda_n \mu_1 \cdots \mu_n \). Here are some exercises:

Prove that the function \( f \) is homogeneous of degree \( n \), that is, \( f (tA) = t^n f (A) \) for all \( t > 0 \), that \( f^{1/n} \) is concave and that \( f \) is monotone: \( f (C) \leq f (A) \) provided \( 0 < c_{ij} \leq a_{ij} \) for all \( i, j \).

(9.3) Bounds on the permanents of doubly stochastic matrix. Suppose that \( A \) is obtained from \( B \) by scaling, \( a_{ij} = \lambda_i \mu_j b_{ij} \) for all \( i, j \). Then
\[
\text{per} \ A = \lambda_1 \cdots \lambda_n \mu_1 \cdots \mu_n \text{ per} \ B.
\]
In view on Theorem 9.1, it would be nice to get some bounds on the permanent of a doubly stochastic matrix \( B \). It is known that
\[
\prod_{i,j} (1 - b_{ij})^{1 - b_{ij}} \leq \text{per} \ B \leq 2^n \prod_{i,j} (1 - b_{ij})^{1 - b_{ij}}.
\]
The lower bound is due to Schrijver (1998). Lelarge (2015) proved the lower bound using the 2-lift construction, see Section 4. The upper bound is due to Gurvits and Samorodnitsky (2014), who also conjectured that \( 2^n \) can be replaced by \( 2^{n/2} \), after which the bound becomes sharp on block-diagonal matrices with diagonal \( 2 \times 2 \) blocks \( \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \).

10. Entropy

(10.1) The entropy function. Given \( x_1, \ldots, x_n \geq 0 \) such that \( x_1 + \ldots + x_n = 1 \), we define
\[
H (x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \ln \frac{1}{x_i}.
\]
where we agree that the $i$-th term is 0 if $x_i = 0$. Since $H$ is a (strictly) concave function, it attains its minimum on the simplex $x_1 + \ldots + x_n = 1$, $x_i \geq 0$ for $i = 1, \ldots, n$, at an extreme point (vertex), where one $x_i = 1$ and all other are 0. Therefore,

$$H(x_1, \ldots, x_n) \geq 0.$$ 

On the other hand, since $\ln x$ is a (strictly) concave function, we have

$$\sum_{i=1}^{n} x_i \ln \frac{1}{x_i} \leq \ln \left( \sum_{i=1}^{n} x_i \cdot \frac{1}{x_i} \right) = \ln n$$

and hence

$$H(x_1, \ldots, x_n) \leq \ln n,$$

with equality attained if and only if $x_1 = \ldots = x_n = 1/n$.

(10.2) The entropy of a partition. Let $\Omega$ be a probability space and let $\mathcal{F} = \{F_1, \ldots, F_n\}$ be a partition of $\Omega$ into pairwise disjoint events:

$$\Omega = \bigcup_{i=1}^{n} F_i \quad \text{and} \quad F_i \cap F_j = \emptyset \quad \text{for} \quad i \neq j.$$ 

We define the entropy of the partition by

$$H(\mathcal{F}) = H(p_1, \ldots, p_n) = \sum_{i=1}^{n} p_i \ln \frac{1}{p_i} \quad \text{where} \quad p_i = \Pr(F_i) \quad \text{for} \quad i = 1, \ldots, n.$$ 

(10.3) Conditional entropy. Let $\Omega$ be a probability space and let $\mathcal{F}$ and $\mathcal{G}$ be partitions of $\Omega$ into finitely many disjoint events. We say that $\mathcal{G}$ refines $\mathcal{F}$, if every event $F$ in $\mathcal{F}$ is a union of events in $\mathcal{G}$, in which case we write $\mathcal{F} \preceq \mathcal{G}$. For such partitions we define the conditional entropy

$$(10.3.1) \quad H(\mathcal{G}|\mathcal{F}) = \sum_{F \in \mathcal{F}} \Pr(F) \sum_{G \in \mathcal{G}} \Pr(G) \Pr(F) \ln \frac{\Pr(F)}{\Pr(G)}.$$ 

If $\Pr(F) = 0$ then the corresponding term in (10.3.1) is 0. In words: we consider each event $F$ in $\mathcal{F}$ (henceforth called block) as a probability space with conditional probability measure, compute the entropy of the partition of $F$ by events of $\mathcal{G}$ and average over all blocks $F$ of $\mathcal{F}$.

It follows that $H(\mathcal{G}|\mathcal{F}) \geq 0$ and that

$$H(\mathcal{G}) = H(\mathcal{F}) + H(\mathcal{G}|\mathcal{F}).$$
Telescoping, we get

\[ H(F_n) = H(F_0) + \sum_{k=1}^{n} H(F_k | F_{k-1}) \]

provided

\[ F_0 \preceq F_1 \preceq \ldots \preceq F_n. \]

Suppose that \( \Omega \) is finite and that every \( \omega \in \Omega \) is an event. Let \( F \preceq G \) be a pair of partitions of \( \Omega \). For \( \omega \in \Omega \), let \( F(\omega) \) be the block of \( F \) containing \( \omega \). Assuming that \( \Pr(F(\omega)) \neq 0 \), we consider \( F(\omega) \) as a probability space with conditional probability measure. Let \( F(\omega) \) be the partition of \( F(\omega) \) induced by events of \( G \). Using that

\[ \Pr(F(\omega)) = \sum_{\omega \in F} \Pr(\omega), \]

we can rewrite (10.3.1) as

\[ H(G|F) = \sum_{\omega \in \Omega} \Pr(\omega) H(F(\omega)). \]

If \( \Pr(F(\omega)) = 0 \), we just assume that the corresponding term in (10.3.3) is 0. Similarly, we can rewrite (10.3.2) as

\[ H(F_n) = H(F_0) + \sum_{\omega \in \Omega} \Pr(\omega) \sum_{k=1}^{n} H(F_{k-1}(\omega)), \]

where \( F_{k-1}(\omega) \) is the partition of the block of \( F_{k-1} \) containing \( \omega \) by events of \( F_k \).

11. THE BREGMAN-MINC INEQUALITY

Our goal is to prove the following result.

(11.1) Theorem. Let \( A = (a_{ij}) \) be an \( n \times n \) matrix with 0-1 entries and let \( r_i > 0 \) be the number of 1s in the \( i \)-th row of \( A \) for \( i = 1, \ldots, n \). Then

\[ \per A \leq \prod_{i=1}^{n} (r_i!)^{1/r_i}. \]

The inequality of Theorem 11.1 was conjectured by Minc and proved by Bregman (1973). We follow the entropy-based approach of Radhakrishnan (1997). The proof hinges on the following combinatorial lemma.
Lemma. Let $A$ be a matrix as in Theorem 11.1, let us fix an integer $1 \leq i \leq n$ and a permutation $\omega \in S_n$ such that $a_{i\omega(i)} = 1$. Let us choose a permutation $\tau \in S_n$ uniformly at random, find $1 \leq k \leq n$ such that $\tau(k) = i$ and cross out from $A$ the columns indexed by $\omega(\tau(1)), \ldots, \omega(\tau(k-1))$. Let $X$ be the number of 1s in the $i$-th row that remain after the columns are crossed, so $X = X(\tau)$ is a random variable on $S_n$. Then

$$\Pr(X = x) = \frac{1}{r_i} \quad \text{for} \quad x = 1, \ldots, r_i.$$ 

Sketch of Proof. Let $J$ be the set of indices $j$ of columns such that $a_{ij} = 1$. Hence $|J| = r_i$ and $\omega(i) \in J$. Let $I = \omega^{-1}(J)$, so that $|I| = r_i$ and $i \in I$. The number of 1s that remain in the $i$-th row is the number of indices in $\tau^{-1}(I)$ that are greater than or equal to $\tau^{-1}(i)$. Since $\tau$ is chosen uniformly at random, $\tau^{-1}(i)$ is equally likely to be the largest, the second largest, etc. element of $\tau^{-1}(I)$. □

(11.3) Sketch of Proof of Theorem 11.1. Let

$$\Omega = \{ \omega \in S_n : a_{i\omega(i)} = 1 \quad \text{for} \quad i = 1, \ldots, n \},$$

so that $\text{per } A = |\Omega|$. Without loss of generality, we assume that $\Omega \neq \emptyset$ and consider $\Omega$ as a finite probability space with uniform measure. For a permutation $\tau \in S_n$, we consider a family of partitions

$$\mathcal{F}_{\tau,0} \preceq \mathcal{F}_{\tau,1} \preceq \cdots \preceq \mathcal{F}_{\tau,n}$$

of $\Omega$, where a block of $\mathcal{F}_{\tau,k}$ consists of permutations $\omega \in \Omega$ with prescribed values of $\omega(\tau(1)), \ldots, \omega(\tau(k))$. In particular, $\mathcal{F}_{\tau,0}$ consists of a single block $\Omega$, while $\mathcal{F}_{\tau,n}$ is the partition of $\Omega$ into singletons. Applying (10.3.4), we obtain

$$\ln |\Omega| = \sum_{\omega \in \Omega} \Pr(\omega) \sum_{k=1}^{n} H(\mathcal{F}_{\tau,k-1}(\omega))$$

and averaging over $\tau \in S_n$, we get

(11.3.1) $$\ln |\Omega| = \sum_{\omega \in \Omega} \Pr(\omega) \left( \frac{1}{n!} \sum_{\tau \in S_n} \sum_{k=1}^{n} H(\mathcal{F}_{\tau,k-1}(\omega)) \right).$$

We fix an $\omega \in \Omega$, make a substitution $i = \tau^{-1}(k)$ and consider the sum

(11.3.2) $$\frac{1}{n!} \sum_{\tau \in S_n} \sum_{k=1}^{n} H(\mathcal{F}_{\tau,k-1}(\omega)) = \frac{1}{n!} \sum_{\tau \in S_n} \sum_{i=1}^{n} H(\mathcal{F}_{\tau,\tau^{-1}(i)-1}(\omega)).$$

27
Now, the space of $\mathcal{F}_{\tau,\tau^{-1}(i)-1}$ consists of the permutations $\sigma \in \Omega$ such that $\sigma(\tau(1)) = \omega(\tau(1)), \ldots, \sigma(\tau(k-1)) = \omega(\tau(k-1))$ further subdivided into blocks depending on the value of $\sigma(\tau(k))$ for $k = \tau(i)$. Hence the number of blocks in $\mathcal{F}_{\tau,\tau^{-1}(i)-1}(\omega)$ does not exceed the number $X$ of 1s in the $i$-th row of $A$, after the columns numbered $\omega(\tau(1)), \ldots, \omega(\tau(k-1))$ have been crossed out. Since

$$H(\mathcal{F}_{\tau,\tau^{-1}(i)-1}(\omega)) \leq \ln X,$$

from Lemma 11.2, we get

$$\frac{1}{n!} \sum_{\tau \in \mathcal{S}_n} H(\mathcal{F}_{\tau,\tau^{-1}(i)-1}(\omega)) \leq \frac{1}{\prod_{x=1}^{r_i}} \ln x = \frac{1}{r_i} (\ln r_i!).$$

Summing over $i = 1, \ldots, n$, we conclude from (11.3.1) and (11.3.2) that

$$\ln |\Omega| \leq \sum_{i=1}^{n} \frac{1}{r_i} \ln (r_i!),$$

which is the desired bound. $\square$

Here are some applications of Theorem 11.1.

(11.4) The number of perfect matchings in $k$-regular bipartite graphs. The number of perfect matchings in $k$-regular bipartite graph with $n + n$ vertices is the permanent of $n \times n$ matrix of 0s and 1s with exactly $k$ of 1s in every row and column. Hence by Theorem 11.1, this number does not exceed

$$\frac{(k!)^{n/k}}{n!} \leq k^n \left(\frac{k-1}{k}\right)^{(k-1)n}.$$

On the other hand, by Corollary 4.7 (see also Section 6.4), this number is at least

$$k^n \left(\frac{k-1}{k}\right)^{(k-1)n}.$$

Curiously, when $k$ is large enough, both (11.4.1) and (11.4.2) are roughly $k^n e^{-n}$ (in the logarithmic order, say).

(11.5) The number of Latin squares or 3-dimensional permutations. Let us consider $n \times n \times n$ arrays $A = (a_{ijk})$ of 0s and 1s, such that each of the $3n^2$ “skewers”, obtained by fixing some two coordinates and letting the remaining coordinate vary, contains exactly one “1”. Such “3-dimensional permutations” are in one-to-one correspondence with $n \times n$ “Latin squares”, consisting of $n \times n$ matrices such that each row and each column contains a permutation of $(1, \ldots, n)$. If we are to fill $A$ layer by layer, we can fill the 1st layer by choosing a permutation matrix, that is, a perfect matching in the complete bipartite graph on $n + n$ vertices. After
the first layer is filled, we fill the 2nd layer by choosing a perfect matching in the bipartite graph on \( n + n \) vertices obtained from the complete bipartite graph by deleting the perfect matching chosen at the first layer. In any case, at the 2nd layer we choose a perfect matching in an \((n-1)\)-regular bipartite graph on \( n + n \) vertices. Continuing in this way, to fill the \( k \)-th layer, we need to choose a perfect matching in the bipartite graph on \( n + n \) vertices obtained from the complete bipartite graph by deleting \( k - 1 \) pairwise edge-disjoint perfect matchings chosen at the previous \( k - 1 \) layers. Hence at the \( k \)-th layer we are choosing a perfect matching in an \((n-k)\)-regular bipartite graph on \( n + n \) vertices. From Section 11.4, the number of 3-dimensional permutation is between

\[
\prod_{k=1}^{n} k^n \left(\frac{k-1}{k}\right)^{(k-1)n} \quad \text{and} \quad \prod_{k=1}^{n} (k!)^{n/k},
\]

which can be written as

\[
\left(\frac{(1 + o(1)) \frac{n}{e^2}}{n} \right)^{n^2} \quad \text{as} \quad n \to \infty.
\]

**11.6 The number of \( d \)-dimensional permutations.** Similarly, one can define a “\( d \)-dimensional permutation” as a \( d \)-dimensional cubical array \( n \times \ldots \times n \) filled by 0s and 1s such that each of the \( d n^{d-1} \) “skewers” contains exactly one “1”. Using the entropy method, Luria and Linial (2011) proved an upper bound

\[
(11.6.1) \left( (1 + o(1)) \frac{n}{e^{d-1}} \right)^{n^{d-1}} \quad \text{as} \quad n \to \infty
\]

for the number of such \( d \)-dimensional permutations. For \( d > 3 \), no lower bound matching (11.6.1) is known.

Here is a heuristic argument purported to show that (11.6.1) is a plausible asymptotic formula. With the entries of the array we associate \( n^d \) independent Bernoulli random variables, each taking value 1 with probability \( 1/n \) and value 0 with probability \((n-1)/n\). Then the sum of the variables in a “skewer” is roughly a Poisson random variable with parameter 1, so the probability that a skewer contains exactly one “1” is roughly \( 1/e \). Assuming that the sums in skewers are independent (they are not!), we conclude that the probability that we get a \( d \)-dimensional permutation is roughly \( e^{-dn^{d-1}} \). Since the probability to get any particular permutation is

\[
\left( \frac{1}{n} \right)^{n^{d-1}} \left( 1 - \frac{1}{n} \right)^{n^d - n^{d-1}} \approx e^{-n^{d-1} \frac{n - n^d}{n}},
\]

we obtain (11.6.1).
(11.7) Permanents of doubly stochastic matrices. Let \( A = (a_{ij}) \) be an \( n \times n \) stochastic (i.e. non-negative, with row sums 1) matrix and suppose that
\[
a_{ij} \leq \frac{1}{r_i} \quad \text{for} \quad j = 1, \ldots, n
\]
and some positive integer \( r_i \). Then
\[
\text{per} A \leq \prod_{i=1}^{n} \frac{(r_i!)^{1/r_i}}{r_i}.
\]
Indeed, \( \text{per} A \) is a linear function in each row and hence the maximum value of \( \text{per} A \) as a function of the \( i \)-th row \( a_i = (a_{i1}, \ldots, a_{in}) \) on the polytope
\[
\sum_{j=1}^{n} a_{ij} = 1 \quad \text{and} \quad 0 \leq a_{ij} \leq \frac{1}{r_i} \quad \text{for} \quad j = 1, \ldots, n
\]
is attained at an extreme point, where \( a_{ij} \in \{0, 1/r_i\} \) for all \( j \) and then (11.7.1) follows from Theorem 11.1. This argument is due to Samorodnitsky (2000).

Assume now that \( A \) is a doubly stochastic matrix with all entries not exceeding \( \alpha/n \) for some constant \( \alpha \geq 1 \). Combining Theorem 6.2 and (11.7.1), we conclude that
\[
\text{per} A = e^{-n} n^{O(\alpha)}
\]
so the permanent of doubly stochastic matrices with small entries is strongly concentrated.

A positive matrix \( A \) is called \( \alpha \)-balanced if the ratio of any two entries in every row or column does not exceed \( \alpha \). Exercise: prove that if \( B \) is a doubly stochastic matrix obtained from an \( \alpha \)-balanced matrix by scaling then \( B \) is \( \alpha^2 \)-balanced and deduce that
\[
\text{per} A = e^{-n} n^{O(\alpha^2)} \text{cap} p_A \quad \text{where} \quad p_A(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij} x_j \right)
\]
and \( A \) is \( \alpha \)-balanced.

12. \( \mathbb{D} \)-stable polynomials

(12.1) \( \mathbb{D} \)-stable polynomials. Let
\[
\mathbb{D} = \{ z \in \mathbb{C} : |z| \leq 1 \}
\]
be the closed unit disc in the complex plain. A polynomial \( p(z_1, \ldots, z_n) \) is called \( \mathbb{D} \)-stable provided
\[
p(z_1, \ldots, z_n) \neq 0 \quad \text{whenever} \quad z_1, \ldots, z_n \in \mathbb{D}.
\]
Our main goal is to prove the following result of Hinkkanen (1997) who was building on results of Ruelle (1971) who was building on results of Asano (1970).
(12.2) Theorem. Suppose that
\[f(z_1, \ldots, z_n) = \sum_{S \subset \{1, \ldots, n\}} a_S \prod_{i \in S} z_i \quad \text{and} \quad g(z_1, \ldots, z_n) = \sum_{S \subset \{1, \ldots, n\}} b_S \prod_{i \in S} z_i\]
are \(D\)-stable. Then the polynomial
\[h(z_1, \ldots, z_n) = \sum_{S \subset \{1, \ldots, n\}} (a_S b_S) \prod_{i \in S} z_i\]
is also \(D\)-stable (we agree that for \(S = \emptyset\) we have \(\prod_{i \in S} z_i = 1\)).

The polynomial \(h\) is called the Hadamard product or the Schur product of polynomials \(f\) and \(g\) and denoted \(h = f * g\). The proof is based on the following lemma due to Asano (1970), known as the Asano contraction lemma.

(12.3) Lemma. Suppose that the bivariate polynomial
\[f(z_1, z_2) = a + bz_1 + cz_2 + dz_1 z_2\]
is \(D\)-stable. Then the univariate polynomial
\[g(z) = a + dz\]
is also \(D\)-stable.

Sketch of Proof. Since \(f\) is \(D\)-stable, we have \(a \neq 0\). Suppose that \(g\) is not \(D\)-stable. Then \(|d| \geq |a|\). Without loss of generality, we assume that \(|b| \geq |c|\). Let us find a \(z_2\) such that \(|z_2| = 1\) and
\[|b + dz_2| = |b| + |d| .\]
Note that
\[|a + cz_2| \leq |a| + |c| \leq |b| + |d| = |b + dz_2| .\]
Solving the equation \(f(z_1, z_2) = 0\) for \(z_1\), we obtain
\[z_1 = -\frac{a + cz_2}{b + dz_2},\]
so that
\[|z_1| = \frac{|a + cz_2|}{|b + dz_2|} \leq 1 ,\]
which contradicts the \(D\)-stability of \(f\).
\[\square\]
(12.4) Sketch of Proof of Theorem 12.2. We proceed by induction on $n$. If $n = 1$ then $f(z) = a_0 + a_1 z$ and $g(z) = b_0 + b_1 z$ where $|a_0| > |a_1|$ and $|b_0| > |b_1|$. Hence $h(z) = (a_0 b_0) + (a_1 b_1) z$ and $|a_0 b_0| > |a_1 b_1|$. Therefore $h$ is $D$-stable.

Suppose that $n > 1$. We write

$$ f(z_1, \ldots, z_n) = \sum_{S \subseteq \{1, \ldots, n-1\}} (a_S + a_{S \cup \{n\}}) \prod_{i \in S} z_i $$

and

$$ g(z_1, \ldots, z_n) = \sum_{S \subseteq \{1, \ldots, n-1\}} (b_S + b_{S \cup \{n\}}) \prod_{i \in S} z_i. $$

Therefore, for any $w_1, w_2 \in D$, the $(n - 1)$-variate polynomials

$$ (z_1, \ldots, z_{n-1}) \mapsto \sum_{S \subseteq \{1, \ldots, n-1\}} (a_S + a_{S \cup \{n\}}) \prod_{i \in S} z_i $$

and

$$ (z_1, \ldots, z_{n-1}) \mapsto \sum_{S \subseteq \{1, \ldots, n-1\}} (a_S + a_{S \cup \{n\}}) \prod_{i \in S} z_i $$

are $D$-stable and hence by the induction hypothesis the $(n - 1)$-variate polynomial

$$ (z_1, \ldots, z_{n-1}) \mapsto \sum_{S \subseteq \{1, \ldots, n-1\}} (a_S + a_{S \cup \{n\}}) (b_S + b_{S \cup \{n\}}) \prod_{i \in S} z_i $$

is $D$-stable. Therefore, for any $z_1, \ldots, z_{n-1} \in D$, the bivariate polynomial

$$ (w_1, w_2) \mapsto \sum_{S \subseteq \{1, \ldots, n-1\}} (a_S + a_{S \cup \{n\}}) (b_S + b_{S \cup \{n\}}) \prod_{i \in S} z_i $$

is $D$-stable. Hence by Lemma 12.3, for any $z_1, \ldots, z_{n-1} \in D$, the univariate polynomial

$$ z \mapsto \sum_{S \subseteq \{1, \ldots, n-1\}} (a_S b_S + z a_{S \cup \{n\}} b_{S \cup \{n\}}) \prod_{i \in S} z_i $$

is $D$-stable. Therefore, $h$ is $D$-stable. \hfill \Box

(12.5) Corollary. Let $\alpha > 0$ and $\beta > 0$ be real and let

$$ f(z_1, \ldots, z_n) = \sum_{S \subseteq \{1, \ldots, n\}} a_S \prod_{i \in S} z_i $$

and

$$ g(z_1, \ldots, z_n) = \sum_{S \subseteq \{1, \ldots, n\}} b_S \prod_{i \in S} z_i $$

be polynomials such that

$$ f(z_1, \ldots, z_n) \neq 0 \quad \text{whenever} \quad |z_1| \leq \alpha, \ldots, |z_n| \leq \alpha $$

and

$$ g(z_1, \ldots, z_n) \neq 0 \quad \text{whenever} \quad |z_1| \leq \beta, \ldots, |z_n| \leq \beta. $$
Then for
\[ h(z_1, \ldots, z_n) = \sum_{S \subset \{1, \ldots, n\}} a_S b_S \prod_{i \in S} z_i \]
we have
\[ h(z_1, \ldots, z_n) \neq 0 \quad \text{whenever} \quad |z_1| \leq \alpha \beta, \ldots, |z_n| \leq \alpha \beta. \]

Sketch of Proof. The polynomials
\[ (z_1, \ldots, z_n) \mapsto f(\alpha z_1, \ldots, \alpha z_n) \quad \text{and} \quad (z_1, \ldots, z_n) \mapsto g(\beta z_1, \ldots, \beta z_n) \]
are \( \mathbb{D} \)-stable and hence by Theorem 12.2 the polynomial
\[ (z_1, \ldots, z_n) \mapsto h(\alpha \beta z_1, \ldots, \alpha \beta z_n) \]
is \( \mathbb{D} \)-stable. \( \square \)

(12.6) Ramifications. For a multi-index \( \alpha = (\alpha_1, \ldots, \alpha_n) \) of non-negative integers we denote \( z^\alpha = z_1^{\alpha_1} \cdots z_n^{\alpha_n} \), where we agree that \( z_i^{\alpha_i} = 1 \) if \( \alpha_i = 0 \). For \( \gamma = (\gamma_1, \ldots, \gamma_n) \) we say that \( \alpha \leq \gamma \) if \( \alpha_i \leq \gamma_i \) for \( i = 1, \ldots, n \), in which case we denote
\[ \binom{\gamma}{\alpha} = \prod_{i=1}^{n} \frac{\gamma_i!}{\alpha_i! (\gamma_i - \alpha_i)!} \]
Borcea and Brändén (2009) proved the following extension of Theorem 12.2. Suppose that the polynomials
\[ f(z_1, \ldots, z_n) = \sum_{\alpha \leq \gamma} a_\alpha \binom{\gamma}{\alpha} z^\alpha \quad \text{and} \quad g(z_1, \ldots, z_n) = \sum_{\alpha \leq \gamma} b_\alpha \binom{\gamma}{\alpha} z^\alpha \]
are \( \mathbb{D} \)-stable. Then the polynomial
\[ h(z_1, \ldots, z_n) = \sum_{\alpha \leq \gamma} a_\alpha b_\alpha \binom{\gamma}{\alpha} z^\alpha \]
is also \( \mathbb{D} \)-stable.

Ruelle (1971) proved the following extension of Lemma 12.3. Let \( A, B \subset \mathbb{C} \) be closed sets such that \( 0 \notin A \) and \( 0 \notin B \). Let
\[ p(z_1, z_2) = a + bz_1 + cz_2 + dz_1 z_2 \]
be a polynomial such that
\[ p(z_1, z_2) = 0 \quad \Rightarrow \quad z_1 \in A \quad \text{or} \quad z_2 \in B. \]
Then for the polynomial
\[ q(z) = a + dz \]
we have
\[ q(z) = 0 \quad \Rightarrow \quad z = -z_1 z_2 \quad \text{for some} \quad z_1 \in A \quad \text{and} \quad z_2 \in B. \]
13. The Lee-Yang Circle Theorem

The following remarkable result was obtained by Lee and Yang (1952).

(13.1) **Theorem.** Let \( A = (a_{ij}) \) be an \( n \times n \) Hermitian matrix such that \( |a_{ij}| \leq 1 \) for all \( i, j \). Then the roots of the univariate polynomial

\[
\text{Cut}_A(z) = \sum_{S \subset \{1, \ldots, n\}} z^{|S|} \prod_{i \in S, j \notin S} a_{ij}
\]

lie on the circle \( |z| = 1 \). As usual, we assume that the constant term (for \( S = \emptyset \)) and the coefficient of \( z^n \) (for \( S = \{1, \ldots, n\} \)) are equal to 1.

The proof is based on Theorem 12.2 and the following lemma.

(13.2) **Lemma.** Suppose that \( |a| \leq 1 \). Then

\[ 1 + az_1 + \overline{a}z_2 + z_1z_2 \neq 0 \quad \text{whenever} \quad |z_1| < 1 \quad \text{and} \quad |z_2| < 1. \]

**Sketch of Proof.** Solving the equation for \( z_2 \), say, we obtain

\[ z_2 = -\frac{1 + az_1}{\overline{a} + z_1}. \]

We claim that for the map

\[ \phi(z) = \frac{1 + az}{\overline{a} + z} \]

we have \( |\phi(z)| \geq 1 \) for all \( |z| < 1 \). Indeed

\[ |\phi(z)|^2 = \frac{(1 + az)(1 + \overline{a}z)}{(|\overline{a} + z|)(a + \overline{z})} \]

and

\[(1 + az)(1 + \overline{a}z) - (\overline{a} + z)(a + \overline{z}) = 1 + |a|^2|z|^2 - |a|^2 - |z|^2 = (1 - |a|^2)(1 - |z|^2) \geq 0.\]

\( \square \)

(13.3) **Sketch of Proof of Theorem 13.1.** Let us define an \( n \)-variate polynomial

\[ p_A(z_1, \ldots, z_n) = \sum_{S \subset \{1, \ldots, n\}} \prod_{i \in S, j \notin S} a_{ij} \prod_{i \in S} z_i. \]

For an unordered pair \( \{i, j\} \) where \( 1 \leq i \neq j \leq n \), let us define

\[ p_{ij}(z_1, \ldots, z_n) = (1 + a_{ij}z_i + a_{ji}z_j + z_i z_j) \prod_{k \neq i, j} (1 + z_k). \]
We note that the coefficient of \( \prod_{m \in S} z_m \) in \( p_{ij} \) is equal 1 if \( i \notin S, j \notin S \) or if \( i \in S, j \in S \), is equal \( a_{ij} \) if \( i \in S, j \notin S \) and is equal \( a_{ji} \) if \( j \in S \) and \( i \notin S \). Therefore, \( p_A \) is the Hadamard (Schur) product of \( p_{ij} \) over all unordered pairs \( \{i, j\} \) of indices \( 1 \leq i \neq j \leq n \),

\[
p_A = \ast_{\{i,j\}} p_{ij}.
\]

For any \( 0 < \rho < 1 \), by Lemma 13.2 we have

\[
p_{ij}(z_1, \ldots, z_n) \neq 0 \quad \text{whenever} \quad |z_1| \leq \rho, \ldots, |z_n| \leq \rho.
\]

Applying Corollary 12.5, we conclude that

\[
p_A(z_1, \ldots, z_n) \neq 0 \quad \text{whenever} \quad |z_1| \leq \rho^{\binom{n}{2}}, \ldots, |z_n| \leq \rho^{\binom{n}{2}}.
\]

Since \( 0 < \rho < 1 \) was arbitrary, we conclude that

\[
p_A(z_1, \ldots, z_n) \neq 0 \quad \text{provided} \quad |z_1| < 1, \ldots, |z_n| < 1.
\]

In particular,

\[
\text{Cut}_A(z) \neq 0 \quad \text{provided} \quad |z| < 1.
\]

Since

\[
z^n \text{Cut}_A \left( \frac{1}{z} \right) = \overline{\text{Cut}_A (z)},
\]

it follows that

\[
\text{Cut}_A(z) \neq 0 \quad \text{provided} \quad |z| > 1.
\]

Hence the roots of \( \text{Cut}_A(z) \) lie on the circle \( |z| = 1 \).

\[\square\]

14. Complex zeros and computational complexity

(14.1) Lemma. Let \( g : \mathbb{C} \rightarrow \mathbb{C} \) be a polynomial of degree \( n \) and let \( \beta > 1 \) be real. Suppose that

\[
g(z) \neq 0 \quad \text{provided} \quad |z| \leq \beta.
\]

Let us choose a branch of

\[
f(z) = \ln g(z) \quad \text{for} \quad |z| \leq 1
\]

and let

\[
p_m(z) = f(0) + \sum_{k=1}^{m} \frac{f^{(k)}(0)}{k!} z^k
\]

35
be the degree \( m \) Taylor polynomial of \( f \). Then
\[
|f(z) - p_m(z)| \leq \frac{n}{(m + 1)\beta^m(\beta - 1)} \quad \text{provided} \quad |z| \leq 1.
\]

Sketch of Proof. Let \( \alpha_1, \ldots, \alpha_n \) be the roots of \( g(z) \), so that
\[
g(z) = g(0) \prod_{i=1}^{n} \left(1 - \frac{z}{\alpha_i}\right) \quad \text{and} \quad |\alpha_i| > \beta \quad \text{for} \quad i = 1, \ldots, n.
\]
Then
\[
f(z) = f(0) + \sum_{i=1}^{n} \ln \left(1 - \frac{z}{\alpha_i}\right).
\]
Assuming that \(|z| \leq 1\), we have
\[
\ln \left(1 - \frac{z}{\alpha_i}\right) = -\sum_{k=1}^{m} \frac{z^k}{k\alpha_i^k} + \xi_i,
\]
where
\[
|\xi_i| = \left| \sum_{k=m+1}^{\infty} \frac{z^k}{k\alpha_i^k} \right| \leq \frac{1}{m + 1} \sum_{k=m+1}^{\infty} \beta^{-k} = \frac{1}{(m + 1)\beta^m(\beta - 1)}
\]
and the proof follows. \( \square \)

As follows from Lemma 14.1, for any fixed \( \beta > 1 \), to approximate \( f(1) = \ln g(1) \) within an additive error \( 0 < \epsilon < 1 \), we can use the Taylor polynomial of \( f \) at \( z = 0 \) of degree \( m = O_\beta(\ln n - \ln \epsilon) \).

(14.2) Computing \( f^{(k)}(0) \). We have
\[
f'(z) = \frac{g'(z)}{g(z)} \quad \text{and hence} \quad g'(z) = f'(z)g(z).
\]
Differentiating the product \( k - 1 \) times, we get
\[
g^{(k)}(z) = \sum_{j=0}^{k-1} \binom{k-1}{j} f^{(k-j)}(z)g^{(j)}(z).
\]
Hence to compute
\[
f(0), f^{(1)}(0), \ldots, f^{(m)}(0)
\]
we need to know
\[
g(0), g^{(1)}(0), \ldots, g^{(m)}(0)
\]
and solve the non-degenerate triangular system (14.2.1) of linear equations. It follows from Lemma 14.1 that to approximate \( f(1) \) within an additive error of \( 0 < \epsilon < 1 \) (or, equivalently, to approximate \( g(1) \) with a relative error \( 0 < \epsilon < 1 \)), we need to know only \( O(\ln n - \ln \epsilon) \) lowest terms of \( g(z) \) (provided, of course, that there are no complex zeros in the vicinity).
Example: computing the partition function in the ferromagnetic Ising model. Let us $A = (a_{ij})$ be an $n \times n$ Hermitian matrix such that $|a_{ij}| \leq 1$ for all $i, j$ and let

$$\text{Cut}_A(z) = \sum_{S \subseteq \{1, \ldots, n\}} z^{|S|} \prod_{i \in S, j \notin S} a_{ij}. $$

As we discussed in class, up to a change of variables and some trivial factor, $\text{Cut}_A(z)$ is the partition function in the ferromagnetic Ising model with constant magnetic field,

$$\sum_{\sigma: \{1, \ldots, n\} \rightarrow \{-1, 1\}} \exp \left\{ \sum_{1 \leq i < j \leq n} b_{ij} \sigma(i) \sigma(j) + \sum_{i=1}^n c_i \sigma(i) \right\}. $$

Theorem 13.1 states that all roots of $\text{Cut}_A(z)$ lie on the circle $|z| = 1$. Let us fix a $0 < \gamma < 1$. Using Lemma 14.1 and Section 14.2, we conclude that to approximate $\ln \text{Cut}_A(z)$ with $|z| \leq \gamma$ within an additive error of $0 < \epsilon < 1$, we need only to compute

$$\prod_{i \in S, j \notin S} a_{ij}$$

for subsets $S \subseteq \{1, \ldots, n\}$ of logarithmic size $|S| = O_\gamma (\ln n - \ln \epsilon)$. There are only $n^{O(\ln n - \ln \epsilon)}$ of such subsets (as opposed to $2^n$ of all subsets $S$).

Using reciprocity

$$z^n \text{Cut}_A \left( \frac{1}{z} \right) = \overline{\text{Cut}_A(z)},$$

we conclude that the same is true if we want to approximate $\ln \text{Cut}_A(z)$ for $|z| > \gamma^{-1}$. Hence the only area where it appears to be hard to approximate $\ln \text{Cut}_A(z)$ is in the neighborhood of the circle $|z| = 1$. In other words, it appears to be hard to approximate the partition function in the ferromagnetic Ising model with constant external magnetic field only if the external magnetic field is close to 0.

Next, we want to approximate $f(z) = \ln g(z)$ at $z = 1$, assuming only that $g(z) \neq 0$ in a neighborhood $U$ of the interval $[0, 1] \subset \mathbb{C}$. The idea is to construct a polynomial $\phi: \mathbb{C} \rightarrow \mathbb{C}$ such that $\phi(0) = 1$, $\phi(1) = 1$ and $\phi$ maps the disc $\{z: |z| \leq \beta\}$ of some radius $\beta > 1$ into the neighborhood $U$. The composition $\tilde{g} = g(\phi(z))$ satisfies the conditions of Lemma 14.1. We have $\tilde{g}(1) = g(1)$, $\deg \tilde{g} = (\deg g)(\deg \phi)$ and to compute $\tilde{g}(0), \ldots, \tilde{g}(m)(0)$ we need to know only $g(0), \ldots, g^{(m)}(0)$, since $\phi(0) = 0$.

The following lemma provides a possible construction of $\phi$. 37
Lemma. For $0 < \rho < 1$, let us define

$$
\alpha = \alpha(\rho) = 1 - e^{-\frac{1}{\rho}}, \quad \beta = \beta(\rho) = \frac{1 - e^{-1 - \frac{1}{\rho}}}{1 - e^{-\frac{1}{\rho}}} > 1,
$$

$$
N = N(\rho) = \left\lfloor \left(1 + \frac{1}{\rho}\right) e^{1 + \frac{1}{\rho}} \right\rfloor, \quad \sigma = \sigma(\rho) = \sum_{m=1}^{N} \alpha^m \quad \text{and}
$$

$$
\phi(z) = \phi_\rho(z) = \frac{1}{\sigma} \sum_{m=1}^{N} \frac{(\alpha z)^m}{m}.
$$

Then $\phi(z)$ is a polynomial of degree $N$ such that $\phi(0) = 0$, $\phi(1) = 1$,

$$
-\rho \leq \Re \phi(z) \leq 1 + 2\rho \quad \text{and} \quad |\Im \phi(z)| \leq 2\rho \quad \text{provided} \quad |z| \leq \beta.
$$

Sketch of Proof. Consider the function

$$
F_\rho(z) = \rho \ln \frac{1}{1 - z} \quad |z| < 1,
$$

where we choose the branch of the logarithm so that $F_\rho(0) = 0$. Then

$$
|\Im F_\rho(z)| \leq \frac{\pi \rho}{2} \quad \text{and} \quad \Re F_\rho(z) \geq -\rho \ln 2.
$$

In addition,

$$
F_\rho(\alpha) = 1 \quad \text{and} \quad \Re F_\rho(z) \leq 1 + \rho \quad \text{provided} \quad |z| \leq 1 - e^{-1 - \frac{1}{\rho}}.
$$

We obtain $\phi(z)$ approximating $F_\rho(\alpha z)$ by its Taylor polynomial and scaling to $\phi(1) = 1$.

Note that we get

$$
\beta(\rho) = 1 + O \left( e^{-\frac{1}{\rho}} \right).
$$

One can prove that the dependence on $\rho$ is essentially optimal.

Example: computing the matching polynomial. Given a graph $G = (V, E; a)$ with non-negative weights on edges, we consider its matching polynomial

$$
\Mat_G(\lambda) = 1 + \sum_{\substack{e_1, \ldots, e_k \text{ matching}}} \lambda^k a(e_1) \cdots a(e_k),
$$

see Section 1. By Theorem 1.2, the roots of $\Mat_G(\lambda)$ are negative real. Suppose that we know a bound $\delta > 0$ such that

$$
\lambda \leq -\delta
$$
for the roots $\lambda$ of $\text{Mat}_G(\lambda)$ and that we want to approximate $\text{Mat}_G(\lambda)$ at some, possibly complex, $\lambda$. Suppose now that we have

$$|\lambda| \leq \gamma \delta \text{ and } |\pi - \arg \lambda| \geq \frac{1}{\gamma}$$

for some constant $\gamma > 1$, fixed in advance (so that $\lambda$ is away from the negative real axis and $|\lambda|$ is not too large). We define a polynomial

$$g(z) = \text{Mat}_G(\lambda z)$$

and our goal is to approximate $g(1)$. We note that $g(z) \neq 0$ for $z$ in some neighborhood, depending only on the constant $\gamma$, of the interval $[0, 1] \subset \mathbb{C}$ and that computing $g(0), \ldots, g^{(m)}(0)$ reduces to the enumeration of matchings of size at most $m$ in $G$. Using the above approach, for any $0 < \epsilon < 1$, we can compute $\ln \text{Mat}_G(\lambda)$ within an additive error $\epsilon$ in $|V|^O(\gamma \ln |V| - \ln \epsilon)$ time by enumerating the matchings in $G$ of size $O(\gamma \ln |V| - \ln \epsilon)$.

If we want to approximate $\text{Mat}_G(\lambda)$ for positive real $\lambda$ only, we can get a better complexity bound. Basically, we want to construct a polynomial $\phi : \mathbb{C} \rightarrow \mathbb{C}$ and a real $\beta > 1$ such that $\phi(0) = 0$, $\phi(1) = 1$ and the image of the disc $\{z : |z| \leq \beta\}$ under $\phi$ does not contain negative real smaller than $-\delta$.

**Lemma.** For $0 < \rho < 1$, let us define

$$\alpha = \alpha(\rho) = 1 - \sqrt{\rho \frac{1 + \rho}{1 + \rho}}, \quad \beta = \beta(\rho) = \frac{1 - 0.5\rho}{\alpha} > 1,$$

$$N = N(\rho) = \left\lceil \frac{7 + 4 \ln(1/\rho)}{\sqrt{\rho}} \right\rceil, \quad \sigma = \sigma(\rho) = \sum_{m=1}^{N} (m + 1)\alpha^m \quad \text{and}$$

$$\phi(z) = \phi_\rho(z) = \frac{1}{\sigma} \sum_{m=1}^{N} (m + 1)(\alpha z)^m.$$

Then $\phi$ is a polynomial of degree $N$ such that $\phi(0) = 0$, $\phi(1) = 1$ and the image of the disc $\{z : |z| \leq \beta\}$ under $\phi$ does not contain negative real smaller than $-2\rho$.

**Sketch of Proof.** Consider the function

$$G_\rho(z) = \frac{\rho}{(1 - z)^2} - \rho \quad \text{for} \quad |z| \leq 1 - 0.5\sqrt{\rho}.$$

Then $G_\rho(0) = 0$, $G_\rho(\alpha) = 1$ and the image of $G_\rho$ does not contain the cone

$$K = \{z : |\pi - \arg z| < \sqrt{\rho} - \rho\}.$$

We obtain $\phi(z)$ approximating $G_\rho(\alpha z)$ by its Taylor polynomial and scaling to $\phi(1) = 1$. \qed

We note that

$$\beta(\rho) = 1 + O(\sqrt{\rho})$$
15. Complex zeros of the permanent

(15.1) Theorem. Let $A = (a_{ij})$ be an $n \times n$ complex matrix such that

$$|1 - a_{ij}| \leq \frac{1}{2} \quad \text{for all } i, j.$$

Then $\text{per} A \neq 0$.

The proof is based on a simple geometric lemma.

(15.2) Lemma. Let $u_1, \ldots, u_n \in \mathbb{C} \setminus \{0\}$ be non-zero complex numbers such that the angle between any two does not exceed $\theta < 2\pi/3$.

1. Let $u = u_1 + \ldots + u_n$. Then

$$|u| \geq \left( \cos \frac{\theta}{2} \right) \sum_{i=1}^{n} |u_i|.$$

2. Let

$$v = \sum_{i=1}^{n} \alpha_i u_i \quad \text{and} \quad w = \sum_{i=1}^{n} \beta_i u_i,$$

where $\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_n$ are complex numbers such that

$$|1 - \alpha_i| \leq \delta \quad \text{and} \quad |1 - \beta_i| \leq \delta$$

for some

$$0 \leq \delta < \cos \frac{\theta}{2}$$

and $i = 1, \ldots, n$. Then $v \neq 0$, $w \neq 0$ and the angle between $v$ and $w$ does not exceed

$$2 \arcsin \frac{\delta}{\cos \frac{\theta}{2}}.$$

Sketch of Proof. Since $\theta < 2\pi/3$, the origin is not a convex combination of the vectors $u_1, \ldots, u_n$ and hence $u_1, \ldots, u_n$ lie in an angle measuring at most $\theta$. To prove Part (1), consider the orthogonal projection of $u_i$ onto the bisector of the angle. The length of the orthogonal projection is at least $|u_i| \cos \frac{\theta}{2}$. Hence the length of the orthogonal projection of $u$ onto the bisector is at least

$$\left( \cos \frac{\theta}{2} \right) \sum_{i=1}^{n} |u_i|,$$

and the proof of Part (1) follows.
To prove Part (2), we write \( v = u + (v - u) \) and note that we have

\[
|u| \geq \left( \cos \frac{\theta}{2} \right) \sum_{i=1}^{n} |u_i|
\]

by Part (1) while

\[
|v - u| \leq \delta \sum_{i=1}^{n} |u_i| < |u|.
\]

Hence \( v \neq 0 \) and the angle between \( v \) and \( u \) does not exceed

\[
\arcsin \frac{|v - u|}{|u|} \leq \arcsin \frac{\delta}{\cos \frac{\theta}{2}}.
\]

Similarly, \( w \neq 0 \) and the angle between \( w \) and \( u \) does not exceed

\[
\arcsin \frac{\delta}{\cos \frac{\theta}{2}}.
\]

The proof of Part (2) now follows. \( \square \)

**15.3 Sketch of Proof of Theorem 15.1.** We prove by induction on \( n \) that if \( A \) and \( B \) are two \( n \times n \) matrices as in Theorem 15.1 that differ in at most one row or in at most one column then the angle between \( \text{per} A \neq 0 \) and \( \text{per} B \neq 0 \) does not exceed \( \pi/2 \).

If \( n = 1 \) then the angle between \( \text{per} A \neq 0 \) and \( \text{per} B \neq 0 \) does not exceed

\[
2 \arcsin \frac{1}{2} = \frac{\pi}{3} < \frac{\pi}{2}.
\]

Suppose that \( n > 1 \). Without loss of generality, we assume that \( B \) is obtained from \( A \) by replacing \( a_{1j} \) with \( b_{1j} \) for \( j = 1, \ldots, n \). Then

\[
\text{per} A = \sum_{j=1}^{n} a_{1j} \text{per} A_j \quad \text{and} \quad \text{per} B = \sum_{j=1}^{n} b_{1j} \text{per} A_j,
\]

where \( A_j \) is the \((n - 1) \times (n - 1)\) matrix obtained from \( A \) by crossing out the first row and the \( j \)-th column. Since any two \( A_{j_1} \) and \( A_{j_2} \) differ in at most one column, by the induction hypothesis the angle between any two \( \text{per} A_{j_1} \neq 0 \) and \( \text{per} A_{j_2} \neq 0 \) does not exceed \( \pi/2 \). Applying Part (2) Lemma 15.2 with \( u_j = \text{per} A_j \), \( \theta = \pi/2 \) and \( \delta = 0.5 \), we conclude that the angle between \( \text{per} A \neq 0 \) and \( \text{per} B \neq 0 \) does not exceed

\[
2 \arcsin \frac{\delta}{\cos \frac{\theta}{2}} = 2 \arcsin \frac{0.5}{\cos \frac{\pi}{4}} = 2 \arcsin \frac{\sqrt{2}}{2} = \frac{\pi}{2},
\]

which concludes the induction. \( \square \)
(15.4) **Higher-dimensional permanents.** We can define the \( k \)-dimensional permanent that enumerates perfect matchings in \( k \)-partite hypergraph. We consider the case of \( k = 3 \). Here \( A = (a_{ijk}) \) is an \( n \times n \times n \) array (tensor) and we define its permanent as

\[
\text{PER} A = \sum_{\sigma,\tau \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)\tau(i)}.
\]

(15.5) **Theorem.** Let \( A = (a_{ijk}) \) be an \( n \times n \times n \) complex tensor such that

\[
|1 - a_{ijk}| \leq \frac{\sqrt{6}}{9} \approx 0.272 \quad \text{for all} \quad 1 \leq i, j, k \leq n.
\]

Then

\[
\text{PER} A \neq 0.
\]

**Sketch of Proof.** Let us define

\[
\delta = \frac{\sqrt{6}}{9} \quad \text{and} \quad \theta = 2 \arcsin \frac{1}{\sqrt{6}} < \frac{\pi}{3}.
\]

We prove by induction on \( n \) that if \( A \) and \( B \) are two tensors as in the theorem that differ in at most one “slice” (obtained by fixing one coordinate and letting the other two vary) then the angle between \( \text{PER} A \neq 0 \) and \( \text{PER} B \neq 0 \) does not exceed \( \theta \).

If \( n = 1 \), the angle between \( \text{PER} A \neq 0 \) and \( \text{PER} B \neq 0 \) does not exceed

\[
2 \arcsin \delta < \theta.
\]

If \( n > 1 \), without loss of generality, we assume that \( B \) is obtained from \( A \) by replacing \( a_{1jk} \) by \( b_{1jk} \) for \( j, k = 1, \ldots, n \). Then

\[
\text{PER} A = \sum_{j,k=1}^{n} a_{1jk} \text{PER} A_{jk} \quad \text{and} \quad \text{PER} B = \sum_{j,k=1}^{n} b_{1jk} \text{PER} A_{jk},
\]

where \( A_{jk} \) is the \((n-1) \times (n-1) \times (n-1)\) tensor obtained from \( A \) by crossing out the slices containing \( a_{1jk} \). Since any two \( A_{j_1k_1} \) and \( A_{j_2k_2} \) differ in at most two slices, by the induction hypothesis the angle between \( \text{PER} A_{j_1k_1} \neq 0 \) and \( \text{PER} A_{j_2k_2} \neq 0 \) does not exceed \( 2\theta < 2\pi/3 \). Applying Part (2) of Lemma 15.2, we conclude that the angle between \( \text{PER} A \neq 0 \) and \( \text{PER} B \neq 0 \) does not exceed

\[
2 \arcsin \frac{\delta}{\cos \theta} = 2 \arcsin \frac{\delta}{1 - 2 \sin^2 \frac{\theta}{2}} = 2 \arcsin \frac{\sqrt{6}/9}{2/3} = 2 \arcsin \frac{1}{\sqrt{6}} = \theta,
\]

which completes the induction. \( \square \)

Theorems 15.1 and 15.5 imply that logarithms of (multi-dimensional) permanents are sometimes well-approximated by low degree polynomials.

42
(15.6) **Approximating permanents.** Let \( A = (a_{ij}) \) be an \( n \times n \) matrix with complex entries such that
\[
|1 - a_{ij}| \leq 0.49 \quad \text{for all} \quad i, j
\]
and let \( J \) be the \( n \times n \) matrix filled with 1s. We define a univariate polynomial
\[
g(z) = \text{per}(J + z(A - J)),
\]
so that \( g(0) = n! \) and \( g(1) = \text{per} \ A \). It follows from Theorem 15.1 that
\[
g(z) \neq 0 \quad \text{provided} \quad |z| \leq \beta \quad \text{for} \quad \beta = \frac{0.5}{0.49} > 1.
\]
We define
\[
f(z) = \ln g(z) \quad \text{for} \quad |z| \leq 1
\]
(we choose a branch of the logarithm so that \( f(0) = \ln n! \) is real). Let
\[
(15.6.2) \quad p_m(z) = f(0) + \sum_{k=1}^{m} \frac{f^{(k)}(0)}{k!} z^k
\]
be the Taylor polynomial of \( f(z) \) at \( z = 0 \). By Lemma 14.1,
\[
|f(1) - p_m(1)| \leq \frac{n}{\beta^m(\beta - 1)(m + 1)}.
\]
In particular, to approximate \( f(1) = \ln \text{per} \ A \) within an additive error of \( 0 < \epsilon < 1 \), it suffices to choose \( m = O(\ln n - \ln \epsilon) \). Let us show that \( p_m(1) \) is a degree \( m \) polynomial in the entries of \( A \). Since
\[
g(z) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} (1 + z (a_{i\sigma(i)} - 1)) ,
\]
we have
\[
g^{(k)}(0) = \sum_{\sigma \in S_n} \sum_{(i_1, \ldots, i_k)} (a_{i_1\sigma(i_1)} - 1) \cdots (a_{i_k\sigma(i_k)} - 1)
\]
\[
= (n - k)! \sum_{(i_1, \ldots, i_k)} \sum_{(j_1, \ldots, j_k)} (a_{i_1j_1} - 1) \cdots (a_{i_kj_k} - 1) ,
\]
where \((i_1, \ldots, i_k)\) and \((j_1, \ldots, j_k)\) independently run over ordered \( k \)-tuples of distinct numbers from \( \{1, \ldots, n\} \). Hence \( g^{(k)}(0) \) is a polynomial of degree \( k \) in the entries of \( A \). From Section 14.2, \( f^{(k)}(0) \) is a polynomial of degree \( k \) in the entries of \( A \) and hence by (15.6.2), the value of \( p_m(1) \) is a polynomial of degree \( m \) in the entries of \( A \). Hence we conclude that for any \( 0 < \epsilon < 1 \), there is a polynomial \( p = p_\epsilon \) of degree \( O(\ln n - \ln \epsilon) \) in the entries of \( A \) such that
\[
|\ln \text{per} \ A - p(A)| \leq \epsilon
\]
for any \( A \) satisfying (15.6.1).
16. The hypergraph matching polynomial

(16.1) Definitions. Let \( V \) be a finite set and let \( E \) be a collection of \( d \)-subsets of \( V \). The pair \( H = (V, E) \) is called a \( d \)-uniform hypergraph (we assume that \( d \geq 2 \)). Elements of \( V \) are called vertices of \( H \) whereas elements of \( E \) are called edges of \( H \). The degree of a vertex \( v \) is the number of edges that contain \( v \). A collection, possibly empty, of pairwise disjoint edges is called a matching of \( H \).

Let \( w : E \rightarrow \mathbb{C} \) be a map assigning complex weights to the edges of \( H \). We define the matching polynomial by

\[
\text{Mat}_H(w) = \sum_{e_1, \ldots, e_k \text{ is a matching}} w(e_1) \cdots w(e_k)
\]

(the empty matching contributes constant term 1 to \( \text{Mat}_H(w) \)). In particular, if \( w(e) = \lambda \) for all \( e \in E \), then

\[
\text{Mat}_H(\lambda) = \sum_{k=0}^{\infty} \text{(the number of } k \text{-matchings in } H \text{)} \lambda^k
\]

enumerates matchings in \( H \) according to the number of edges in \( H \).

Given a set \( S \subset V \), we denote by \( H - S \) the hypergraph with set \( V' = V \setminus S \) of vertices and set \( E' = \{ e \in E : e \cap S = \emptyset \} \) of edges. We note the following identity:

\[
(16.1.1) \quad \text{Mat}_H(w) = \text{Mat}_{H - v}(w) + \sum_{e \in E : v \in e} w(e) \text{Mat}_{H - e}(w) \quad \text{for all } v \in V,
\]

where the \( \text{Mat}_{H - v}(w) \) accounts for the matchings not containing \( v \) and the sum accounts for the matching containing \( v \) (we use \( H - v \) as a shorthand for \( H - \{v\} \)).

(16.2) Theorem. Suppose that

\[
\sum_{e \in E : v \in e} |w(e)| \leq \frac{(d-1)^{d-1}}{d^d} \quad \text{for all } v \in V.
\]

Then

\[
\text{Mat}_H(w) \neq 0.
\]

Sketch of Proof. We prove by induction on \( |V| \) that

\[
(16.2.1) \quad \text{Mat}_H(w) \neq 0
\]

and, moreover,

\[
(16.2.2) \quad \left| 1 - \frac{\text{Mat}_{H - v}(w)}{\text{Mat}_H(w)} \right| \leq \frac{1}{d - 1} \quad \text{for all } v \in V.
\]
If \(|V| < d\) then (16.2.1) and (16.2.2) trivially hold.

Suppose that \(|V| \geq d\) and let \(v \in V\) be a vertex. By the induction hypothesis, we have \(\text{Mat}_{H-v}(w) \neq 0\) and from (16.1.1), we get

\[
\text{Mat}_{H}(w) = 1 + \sum_{e \in E: v \in E} w(e) \frac{\text{Mat}_{H-e}(w)}{\text{Mat}_{H-v}(w)}.
\]

Let \(e = \{v, v_2, \ldots, v_d\}\) be an edge containing \(v = v_1\). Telescoping, we get

\[
\frac{\text{Mat}_{H-e}(w)}{\text{Mat}_{H-v}(w)} = \frac{\text{Mat}_{H-e}(w)}{\text{Mat}_{H-\{v_1,v_2,\ldots,v_{d-1}\}}(w)} \cdots \frac{\text{Mat}_{H-\{v_1,v_2\}}(w)}{\text{Mat}_{H-v}(w)}.
\]

Applying the induction hypothesis to each of the \((d-1)\) factors, we get

\[
\left| \frac{\text{Mat}_{H-\{v_1,v_2,\ldots,v_k\}}(w)}{\text{Mat}_{H-\{v_1,v_2,\ldots,v_{k-1}\}}(w)} \right| \leq \frac{d}{d-1} \quad \text{for} \quad k = 2, \ldots, d
\]

and hence

\[
\left| \frac{\text{Mat}_{H-e}(w)}{\text{Mat}_{H-v}(w)} \right| \leq \left( \frac{d}{d-1} \right)^{d-1}.
\]

Then, from (16.2.3) we conclude

\[
\left| 1 - \frac{\text{Mat}_{H}(w)}{\text{Mat}_{H-v}(w)} \right| \leq \frac{(d-1)^{d-1}}{d^d} \cdot \left( \frac{d}{d-1} \right)^{d-1} = \frac{1}{d},
\]

from which (16.2.1)–(16.2.2) follow. \[\square\]

(16.3) Corollary. Let \(H = (V, E)\) be a \(d\)-uniform hypergraph such that the degree of every vertex of \(H\) does not exceed \(\Delta \geq 1\). Suppose that \(w(e) = \lambda\) for some \(\lambda \in \mathbb{C}\) and all \(e \in E\). Then

\[
\text{Mat}_{H}(\lambda) \neq 0 \quad \text{provided} \quad |\lambda| \leq \frac{1}{\Delta} \frac{(d-1)^{d-1}}{d^d}.
\]

As follows from Section 14, \(\ln \text{Mat}_{H}(\lambda)\) is approximated within an additive error of \(\epsilon\) by a low (linear in \(\ln |V| - \ln \epsilon\)) degree polynomial provided

\[
|\lambda| \leq \frac{0.99 (d-1)^{d-1}}{\Delta \cdot d^d}.
\]

In other words, the value of \(\text{Mat}_{H}(\lambda)\) for such \(\lambda\) to a large extent depends on the number of matchings in \(H\) of size \(O(\ln |V| - \ln \epsilon)\).
17. The independence polynomial and its complex zeros

(17.1) Definitions. Let \( G = (V, E) \) be a graph a set \( S \subset V \) is called independent if no two vertices of \( S \) span an edge of \( G \). The empty set \( S = \emptyset \) is always independent. Suppose that a complex number \( z_v \), called the activity at \( v \in V \), is attached to every vertex \( v \in V \). We define the independence polynomial of \( G \) by

\[
\text{ind}_G(z) = \sum_{S \subset V} \prod_{v \in S} z_v,
\]

where for \( S = \emptyset \) the corresponding product in (17.1.1) is 1. For \( v \in V \) we define the neighborhood \( N_v \) of \( v \) in \( G \) by

\[
N_v = \{ u \in V : \{ u, v \} \in E \}.
\]

We note that

\[
(17.1.2) \quad \text{ind}_G(z) = \text{ind}_{G-v}(z) + z_v \text{ind}_{G-v-N_v}(z),
\]

where \( G - v \) is the graph obtained from \( G \) by deleting \( v \) with all adjacent edges and \( G - v - N_v \) is the graph obtained from \( G \) by deleting \( v \) and the vertices from \( N_v \), together with all adjacent edges.

Given a graph \( G \), we define the line graph \( L(G) \) of \( G \) as follows: vertices of \( L(G) \) are the edges of \( G \) and two vertices of \( L(G) \) span an edge provided the corresponding edges of \( G \) share a vertex. It is easy to see that \( \text{Mat}_G = \text{ind}_{L(G)} \), so the independence polynomial generalizes the matching polynomial of a graph. It also generalizes the matching polynomial of a hypergraph.

The following theorem was proved by Dobrushin (1996); the proof below is due to Csikvári and Frenkel (2016) who follow Borgs (2006).

(17.2) Theorem. Let \( G = (V, E) \) be a graph and let \( 0 < r_v < 1 \) be real numbers attached to the vertices \( v \) of \( G \). Suppose that

\[
|z_v| \leq (1 - r_v) \prod_{u \in N_v} r_u \quad \text{for all} \quad v \in V.
\]

Then

\[
\text{ind}_G(z) \neq 0.
\]

Sketch of Proof. For a subset \( A \subset V \), let \( G(A) \) be the graph induced on \( A \). We prove by induction on \( |A| \) that \( \text{ind}_{G(A)}(z) \neq 0 \) and that for any \( B \subset A \) we have

\[
(17.2.1) \quad \frac{|\text{ind}_{G(B)}(z)|}{|\text{ind}_{G(A)}(z)|} \leq \prod_{u \in A \setminus B} r_u^{-1}.
\]

46
If \( A = \emptyset \), both sides of (17.2.1) are 1. Suppose that \( A \neq \emptyset \) and let \( v \in A \). We first check (17.2.1) for \( B = A \setminus \{v\} \). Applying (17.1.2), we have
\[
\text{ind}_{G(A)}(z) = \text{ind}_{G(A-v)}(z) + z_v \text{ind}_{G(A-v-N_v)}(z).
\]
By the induction hypothesis,
\[
\text{ind}_{G(A)}(z) - v(z) = \text{ind}_{G(A-v)}(z) \neq 0.
\]
Hence we get
\[
(17.2.2) \quad \frac{\text{ind}_{G(A)}(z)}{\text{ind}_{G(A-v)}(z)} = 1 + z_v \frac{\text{ind}_{G(A-v-N_v)}(z)}{\text{ind}_{G(A-v)}(z)}.
\]
Again, by the induction hypothesis,
\[
\left| \frac{\text{ind}_{G(A-v-N_v)}(z)}{\text{ind}_{G(A-v)}(z)} \right| = \left| \frac{\text{ind}_{G(A-v)}(z)}{\text{ind}_{G(A-v)}(z)} \right| \leq \prod_{u \in A, \{u,v\} \in E} r_u^{-1}.
\]
Therefore,
\[
\left| z_v \frac{\text{ind}_{G(A-v-N_v)}(z)}{\text{ind}_{G(A-v)}(z)} \right| \leq 1 - r_v
\]
and from (17.2.2) we get
\[
(17.2.3) \quad \left| \frac{\text{ind}_{G(A)}(z)}{\text{ind}_{G(A-v)}(z)} \right| \geq r_v, \quad \text{so that} \quad \left| \frac{\text{ind}_{G(A-v)}(z)}{\text{ind}_{G(A)}(z)} \right| \leq r_v^{-1}
\]
and (17.2.1) follows for \( B = A \setminus \{v\} \).

Next, we establish (17.2.1) for an arbitrary \( B \subset A \). If \( B = A \), then both sides of (17.2.1) are 1. Suppose that \( B \neq A \) and let us choose \( v \in A \setminus B \). Then
\[
\left| \frac{\text{ind}_{G(B)}(z)}{\text{ind}_{G(A)}(z)} \right| = \left| \frac{\text{ind}_{G(B)}(z)}{\text{ind}_{G(A-v)}(z)} \right| \left| \frac{\text{ind}_{G(A-v)}(z)}{\text{ind}_{G(A)}(z)} \right|
\]
and (17.2.1) follows by the induction hypothesis and (17.2.3). \( \square \)

Let \( \Delta = \Delta(G) \) be the largest degree of a vertex in \( G \). Choosing
\[
r_v = \frac{\Delta}{\Delta + 1} \quad \text{for all} \quad v \in V,
\]
we get that
\[
\text{ind}_G(z) \neq 0 \quad \text{provided} \quad |z_v| \leq \frac{\Delta^\Delta}{(\Delta + 1)^{\Delta + 1}} \quad \text{for all} \quad v \in V.
\]
Scott and Sokal (2005) proved a somewhat stronger bound, stating that
\[
\text{ind}_G(z) \neq 0 \quad \text{provided} \quad |z_v| \leq \frac{(\Delta - 1)^{\Delta-1}}{\Delta^\Delta} \quad \text{for all} \quad v \in V.
\]
One can obtain the proof along the lines of the proof of Theorem 16.2.
18. The chromatic polynomial of a graph

(18.1) Definition. Let \( G = (V, E) \) be a graph. A proper \( n \)-coloring of \( G \) is a map \( \phi : V \rightarrow \{1, \ldots, n\} \) such that

\[
\phi(u) \neq \phi(v) \quad \text{whenever} \quad \{u, v\} \in E.
\]

For a positive integer \( n \), we denote by \( \chi_G(n) \) the number of proper \( n \)-colorings of \( G \). The following result is due to Tutte.

(18.2) Theorem. Let \( G = (V, E) \) be a graph. Then

1. For \( k = 1, \ldots, |V| \) there exist integer \( a_k = a_k(G) \) such that

\[
(-1)^{|V| - k} a_k \geq 0 \quad \text{for} \quad k = 1, \ldots, |V|
\]

and

\[
\chi_G(n) = \sum_{k=1}^{|V|} a_k n^k \quad \text{for all positive integer} \quad n.
\]

2. For \( k = 1, \ldots, |V| \) there exist integer \( b_k = b_k(G) \) such that

\[
b_k \geq 0 \quad \text{for} \quad k = 1, \ldots, |V|
\]

and

\[
\chi_G(n) = \sum_{k=1}^{|V|} b_k \binom{n}{k} \quad \text{for all positive integer} \quad n.
\]

Sketch of Proof. We prove Part (1) by the induction on the number \(|E|\) of edges. If \( E = \emptyset \) then \( \chi_G(n) = n^{|V|} \) and the result follows. Suppose that \( E \neq \emptyset \) and let \( e \in E \) be an edge. Let \( G - e \) be the graph obtained from \( G \) by deleting \( e \). That is, \( G - e = (V, E') \) where \( E' = E \setminus \{e\} \). Let \( G/e \) be the graph obtained from \( G \) by contracting \( e \). That is, assuming that \( e = \{u, v\} \), we let \( G/e = (V', E') \) where

\[
V' = (V \setminus \{u, v\}) \cup \{w\},
\]

where \( w \) is a new vertex and \( E' \) is obtained from \( E \) by deleting \( e \) and replacing every edge \( \{a, u\} \) for \( a \neq v \) by \( \{a, w\} \) and every edge \( \{b, v\} \) for \( b \neq u \) by \( \{b, w\} \). It is easy to see that

\[
\chi_G(n) = \chi_{G-e}(n) - \chi_{G/e}(n).
\]

Applying the induction hypothesis to \( G - e \) and \( G/e \), we conclude the proof of Part 1.

To prove Part (2), we define \( b_k \) as the number of proper \( k \)-colorings \( \phi \) with \(|\phi(V)| = k\). \( \square \)
Theorem 18.2 allows us to define the chromatic polynomial of a graph \( G = (V, E) \) by
\[
\chi_G(z) = \sum_{k=1}^{\lvert V \rvert} a_k(G) z^k = \sum_{k=1}^{\lvert V \rvert} b_k(G) \binom{z}{k}
\]
where \( \binom{z}{k} = \frac{z(z-1) \cdots (z-k+1)}{k!} \) for all \( z \in \mathbb{C} \).

19. More on the roots of the independence polynomial

Our goal is to prove the following result of Scott and Sokal (2005).

(19.1) Theorem. Let \( G = (V, E) \) be a graph and let \( x_v \geq 0 \) be non-negative real activities at the vertices of \( G \). For \( \zeta \in \mathbb{C} \), let us define the activities \( z_v = \zeta x_v \) and let
\[
g(\zeta) = \text{ind}_G(z).
\]
Then among the roots of \( g(\zeta) \) nearest to 0, there is necessarily a negative real root.

The proof uses a certain cluster expansion of \( \ln \text{ind}_G(z) \). Given a multiset \( S \) of vertices of \( G \) (that is, some vertices may be repeated several times and some may be missing), we define a graph \( G(S) \) with set \( S \) of vertices as follows: clones of \( v, u \in V \) in \( S \) span an edge of \( G(S) \) if \( u = v \) or if \( \{u, v\} \) is an edge in \( G \). We are interested in the chromatic polynomial \( \chi_{G(S)} \) and, in particular, in its first coefficient \( a_1(G(S)) \), see Theorem 18.2.

(19.2) Theorem. We have
\[
\ln \text{ind}_G(z) = \sum_S \frac{1}{\mu_1! \cdots \mu_n!} a_1(G(S)) \prod_{v \in S} z_v,
\]
where the sum is taken over all non-empty multisets \( S \) of vertices of \( G \), \( \mu_i \) is the multiplicity of \( v_i \) in \( G \) and the series converges absolutely and uniformly on compact sets for all \( z_v \) in a sufficiently small neighborhood of 0 in \( \mathbb{C}^V \).

Sketch of Proof. Applying the Binomial Theorem, we write
\[
\exp \{ x \ln \text{ind}_G(z) \} = (\text{ind}_G(z))^x = \left( 1 + \sum_{\substack{S \subset V \text{ independent} \\ S \neq \emptyset}} \prod_{v \in S} z_v \right)^x = 1 + \sum_{k=1}^{\infty} \binom{x}{k} \left( \sum_{\substack{S \subset V \text{ independent} \\ S \neq \emptyset}} \prod_{v \in S} z_v \right)^k.
\]
where
\[
\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}
\]
and the series converges absolutely and uniformly on compact sets in a neighborhood of \(z_v = x = 0\).

Now,
\[
\left( \sum_{\substack{S \subseteq V \\ S \neq \emptyset}} \prod_{v \in S} z_v \right)^k = \sum_{S_1,\ldots,S_k \text{ non-empty independent}} \prod_{v \in S_1 \sqcup \ldots \sqcup S_k} z_v,
\]
where the last sum is taken over all ordered \(k\)-tuples of non-empty independent sets in \(G\) and \(S_1 \sqcup \ldots \sqcup S_k\) is the multiset that is the disjoint union of \(S_1,\ldots,S_k\).

Next, we observe that for any non-empty multiset \(S\) of vertices of \(V\), a proper coloring of \(G(S)\) using exactly \(k\) colors corresponds to a representation \(S = S_1 \sqcup \ldots \sqcup S_k\) where \(S_i\) is an independent set in \(G\) consisting of the vertices whose clones in \(G(S)\) are colored with the \(i\)-th color. Moreover, we obtain the same decomposition \(S = S_1 \sqcup \ldots \sqcup S_k\) into the disjoint ordered union of non-empty independent sets from exactly \(\mu_1! \cdots \mu_n!\) different proper \(k\)-colorings of \(G(S)\), since an arbitrary permutation of the colors of the clones results in the same decomposition.

Therefore,
\[
\exp \{ x \ln \text{ind}_G(z) \} = 1 + \sum_{S} \frac{1}{\mu_1! \cdots \mu_n!} \left( \sum_{k=1}^\infty \binom{x}{k} b_k(G(S)) \right) \prod_{v \in S} z_v
\]
\[
= 1 + \sum_{S} \frac{\chi_G(S)(x)}{\mu_1! \cdots \mu_n!} \prod_{v \in S} z_v,
\]
where the sum is taken over all non-empty multisets \(S\), \(\mu_i\) is the multiplicity of \(v_i\) in \(S\) and \(b_k(G(S))\) is the number of proper \(k\)-colorings of \(G(S)\) using exactly \(k\) colors, see Section 18.

Since
\[
\ln \text{ind}_G(z) = \left. \frac{d}{dx} \exp \{ x \ln \text{ind}_G(z) \} \right|_{x=0} \quad \text{and} \quad \left. \frac{d}{dx} \chi_G(x) \right|_{x=0} = a_1(G(S)),
\]
the proof follows. \(\square\)

(19.3) Sketch of proof of Theorem 19.1. From Theorem 19.2, we can write
\[
\ln g(\zeta) = \sum_{S} \frac{a_1(G(S))}{\mu_1! \cdots \mu_n!} \zeta^{|S|} \prod_{v \in S} x_v.
\]
By Part (1) of Theorem 18.2, we have
\[
(-1)^{|S|-1} a_1(G(S)) \geq 0
\]
and hence we have

\[(19.3.1) \quad \ln g(\zeta) = \sum_{m=1}^{\infty} c_m \zeta^m,\]

where \(c_m\) are real coefficients such that

\[(19.3.2) \quad (-1)^m c_m \leq 0.\]

The distance to the nearest root of \(g(\zeta)\) is the radius of convergence of (19.3.1). Because of (19.3.2), if (19.3.1) diverges for some \(\zeta\), it diverges to \(-\infty\) at \(-|\zeta|\). Consequently, if \(\rho\) is the radius of convergence of (19.3.1), then

\[
\lim_{\zeta \to -\rho} \ln g(\zeta) = -\infty
\]

(we consider the limit from the left) and \(\zeta = -\rho\) is the root of \(g\), nearest to the origin. \(\square\)

20. A DYNAMICAL SYSTEM AND ITS APPLICATIONS

For a fixed \(\lambda > 0\) and an integer \(d \geq 2\), we consider the transformation

\[T(x) = T_{\lambda,d}(x) = \frac{1}{1 + \lambda x^{d-1}} \quad \text{for} \quad 0 \leq x \leq 1.\]

Clearly, \(T\) maps the unit interval into itself, so we consider the \(n\)-th iteration \(T^n\) of the map \(T\). Our immediate goal is to prove the following result.

(20.1) **Theorem.** There exists a unique fixed point \(x_0 = x_0(\lambda, d)\) such that \(0 < x_0 < 1\) and \(T(x_0) = x_0\). Let

\[\lambda_c = \lambda_c(d) = \frac{(d-1)^{d-1}}{(d-2)^d}.\]

If \(\lambda < \lambda_c\) then

\[\lim_{n \to \infty} T^n(x) = x_0 \quad \text{for all} \quad 0 \leq x \leq 1.\]

If \(\lambda > \lambda_c\) then there exist \(0 < x_- < x_0\) and \(1 > x_+ > x_0\) such that \(T(x_-) = x_+\), \(T(x_+) = x_-\) and

\[
\lim_{n \to \infty} T^{2n}(x) = \begin{cases} 
  x_- & \text{provided} \quad 0 \leq x < x_0 \\
  x_+ & \text{provided} \quad 1 \geq x > x_0.
\end{cases}
\]

**Sketch of Proof.** It is convenient to make a change of variables

\[x = e^{-s}, \quad 0 \leq s \leq +\infty,\]
just as we did in Section 2, so that $T$ is written as

$$T(s) = \ln \left( 1 + \lambda e^{-s(d-1)} \right).$$

Then $T(s)$ decreases from $\ln(1+\lambda) > 0$ to 0 and hence there is a unique $a = a(\lambda,d)$ such that $T(a) = a$. We have $x_0 = e^{-a}$. Since for $s < +\infty$ we have $T_{\lambda,d}(s)$ increasing with $\lambda$ from 0 at $\lambda = 0$ to $+\infty$, we conclude that $a$ increases with $\lambda$ from 0 to $+\infty$. It is convenient to parameterize everything in terms of $a$ (as opposed to $\lambda$).

We have

$$\ln \left( 1 + \lambda e^{-a(d-1)} \right) = a \quad \text{and hence} \quad \lambda = e^{ad} - e^{a(d-1)} = e^d \left( 1 - e^{-a} \right).$$

Now,

$$T'(a) = \frac{\lambda(d-1)e^{-a(d-1)}}{1 + \lambda e^{-a(d-1)}} = (d-1)(e^{-a} - 1).$$

We observe that

- If $a > \ln \frac{d-1}{d-2}$ then $T'(a) < -1$ and $\lambda > \lambda_c$
- and
- If $a < \ln \frac{d-1}{d-2}$ then $0 > T'(a) > -1$ and $\lambda < \lambda_c$.

In particular, $a$ is locally attracting for $\lambda < \lambda_c$ and repelling if $\lambda > \lambda_c$.

Next, we consider the second iteration $T^2(s)$. Clearly, the function $s \mapsto T^2(s)$ is increasing and $T^2(a) = a$. It is not hard to show that the function $s \mapsto T^2(s)$ has at most one inflection point: it is either concave or changes from convex to concave as $s$ grows. To see that, notice that

$$\frac{1}{(T^2(s))^7} = \frac{(1 + \lambda e^{-s(d-1)})^d + \lambda (1 + \lambda e^{-s(d-1)})}{\lambda^2(d-1)^2e^{-s(d-1)}}$$

is a positive linear combination of $y^k$ for $k = -1, 0, \ldots, d-1$ and $y = e^{-s(d-1)}$.

Now, if $b < a$ is a fixed point of $T^2$ then $c = T(b) > a$ is also a fixed point of $T^2$. We observe that $T^2$ cannot have more than three fixed points, since then the function $s \mapsto T^2(s)$ would have too many inflection points. If $T^2$ has a unique fixed point, necessarily $a$, then

$$(T^2(s))' \bigg|_{s=a} < 1,$$

(since $T^2(s) > s$ for $s < a$ and $T^2(s) < s$ for $s > a$) from which $T'(a) > -1$ and $\lambda < \lambda_c$. It follows then that $a$ is the global attraction point of $T^2$ and hence of $T$.

If $T^2$ has three fixed points $b < a < c$ then

$$(T^2(s))' \bigg|_{s=a} > 1,$$

(since $T^2(s) < s$ for $b < s < a$ and $T^2(s) > s$ for $c > s > a$) from which $T'(a) < -1$ and $\lambda > \lambda_c$. It follows that $b$ is the attraction point of $T^2$ for $0 \leq s < a$ while $c$ is the attraction point of $T^2$ for $s > a$. We let $x_- = e^{-c}$ and $x_+ = e^{-b}$. \qed
The probability space of independent sets. Given a graph $G = (V, E)$ and positive real activities $z_v = \lambda$ for $v \in V$, we consider the set of independent sets $S$ in $G$ as a probability space with

$$\Pr (S) = \frac{\lambda^{\mid S\mid}}{\text{ind}_G(\lambda)}.$$ 

We write the basic relation (17.1.2) as

$$\text{ind}_G(\lambda) = \text{ind}_{G-v}(\lambda) + \lambda \text{ind}_{G-v-N_v}(\lambda),$$

from which

$$(20.2.1) \quad \frac{\text{ind}_{G-v}(\lambda)}{\text{ind}_G(\lambda)} = \left( 1 + \lambda \frac{\text{ind}_{G-v-N_v}(\lambda)}{\text{ind}_{G-v}(\lambda)} \right)^{-1}.$$ 

The left hand side of (20.2.1) is the probability that a random $S$ does not contain vertex $v$ (we say that $v$ is unoccupied if $v \notin S$ and occupied if $v \in S$). The ratio

$$\frac{\text{ind}_{G-v-N_v}(\lambda)}{\text{ind}_{G-v}(\lambda)}$$

in the right hand side of (20.2.1) is the probability that $S$ does not contain any of the neighbors of $v$, given that it does not contain $v$.

Formula (20.2.1) simplifies if $G$ is the (semi) $d$-regular three $T^d_n$ with $n+1$ levels. The vertex $v$ (root) at the zeroth level is connected to $d-1$ vertices at the first level, every vertex at the $k$-th level for $1 \leq k \leq n-1$ is connected to one vertex at the $(k-1)$st level and $d-1$ vertices at the $(k+1)$st level, and the vertices at the $n$-th level are leaves. Let $p_n$ be the probability that the root of $T^d_n$ is unoccupied. Then (20.2.1) can be written as

$$p_n = \frac{1}{1 + \lambda p_{n-1}^{d-1}}$$

with the initial condition

$$p_0 = \frac{1}{1 + \lambda}.$$ 

It follows from Theorem 20.1 that for

$$\lambda < \lambda_c = \frac{(d - 1)^{d-1}}{(d - 2)^d}$$

there is a limit

$$(20.2.2) \quad \lim_{n \to \infty} p_n.$$
Let $p_u^n$ be the conditional probability that the root of $T_d^n$ is unoccupied, given that every leaf is unoccupied. Then $p_u^n$ satisfies the same recurrence relation

$$p_u^n = \frac{1}{1 + \lambda (p_u^{n-1})^{d-1}}$$

with the initial condition

$$p_u^0 = 1.$$

Again, Theorem 20.1 implies that if $\lambda < \lambda_c$ then there is a limit

$$\lim_{n \to \infty} p_u^n$$

and the limits (20.2.2) and (20.2.3) are equal.

Let $p_o^n$ be the conditional probability that the root of $T_d^n$ is unoccupied, given that every leaf is occupied. Then $p_o^n$ satisfies the same recurrence relation

$$p_o^n = \frac{1}{1 + \lambda (p_o^{n-1})^{d-1}}$$

with the initial condition

$$p_o^0 = 0.$$

Again, Theorem 20.1 implies that if $\lambda < \lambda_c$ then there is a limit

$$\lim_{n \to \infty} p_o^n$$

and the limits (20.2.2), (20.2.3) and (20.2.4) are equal. In other words, for subcritical activity $\lambda < \lambda_c$ the probability that the root of $T_d^n$ is occupied stabilizes as $n$ grows and does not depend on whether the leaves are occupied.

Weitz (2006) proved that for any graph with maximum degree $d$ of a vertex, the probability that a given vertex $v$ is occupied asymptotically does not depend on whether some vertices remote from $v$ are occupied provided $\lambda < \lambda_c = (d - 1)^{d-1}/(d - 2)^d$.

What happens for supercritical activities $\lambda > \lambda_c$? The asymptotic probability $p_o$ of the root of $T_d^n$ to be unoccupied depends on the parity of $n$. As follows by Theorem 20.1, the limits

$$\lim_{n \to \infty} p_{2n} \quad \text{and} \quad \lim_{n \to \infty} p_{2n+1}$$

exist and not equal.

If we assume that the leaves of $T_d^n$ at the level $n$ are occupied, it makes the vertices at the level $n - 1$ unoccupied with probability 1, which makes the vertices at the level $n - 2$ more likely to be occupied (compared to the tree with no conditions on the leaves), which makes the vertices at the level $n - 3$ less likely to be occupied, etc. For subcritical activities $\lambda < \lambda_c$, the differences disappear as we are getting closer to the root, but for supercritical activities $\lambda > \lambda_c$, they persist.