NOTES ON COMBINATORIAL APPLICATIONS OF HYPERBOLIC POLYNOMIALS

ABSTRACT. These are notes on combinatorial applications of hyperbolic polynomials, one of the topics covered in my course "Topics in Convexity" in Winter 2013.

1. Hyperbolic polynomials and their hyperbolicity cones

(1.1) Definition. Let $p : \mathbb{R}^d \longrightarrow \mathbb{R}$ be a homogeneous polynomial of some degree $m > 0$, and let $u \neq 0$ be a vector. We say that p is *hyperbolic in the direction of* u if for every $x \in \mathbb{R}^d$ all the roots of the univariate polynomial

$$t \longmapsto p(x - tu)$$

are real.

(1.2) Example. Let

$$p(x) = x_1 \cdots x_d \quad \text{for} \quad x = (x_1, \ldots , x_d)$$

and let

$$u = (1, \ldots , 1).$$

Then

$$p(x - tu) = (x_1 - t) \cdots (x_n - t)$$

and p is hyperbolic in the direction of u (as well as in any other direction).

(1.3) Example. Let $\mathbb{R}^d = Sym_n$, the space of real symmetric $n \times n$ matrices, let

$$p(X) = \det X$$

and let

$$u = I, \quad \text{the identity matrix.}$$

Then

$$p(x - tu) = \det(X - tI)$$

and the roots are of the polynomial $t \longmapsto \det(X - tI)$ are the eigenvalues of X, which are all real. Hence $\det X$ is hyperbolic in the direction of the identity matrix.

Typeset by $\mathcal{AMS}$-TEX

(1.4) Differentiation. Let p be a homogeneous polynomial of degree $m > 1$, hyperbolic in the direction of $u = (u_1, \ldots, u_n)$. We define a polynomial q of degree $m-1$ by

$$q(x) = \frac{\partial p}{\partial u} = \sum_{i=1}^{d} u_i \frac{\partial p}{\partial x_i}.$$

It is then easy to see that q is hyperbolic in the direction of u. Indeed,

$$q(x - tu) = \sum_{i=1}^{d} u_i \frac{\partial p(x - tu)}{\partial x_i} = -\frac{d}{dt} p(x - tu)$$

and by Rolle's Theorem all the roots of the polynomial $t \longmapsto q(x - tu)$ are real and interlace the roots of p.

(1.5) Example. Differentiating $n - k$ times the polynomial of Example 1.2 we conclude that the *elementary symmetric polynomial*

$$p(x) = \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} x_{i_1} \cdots x_{i_k}$$

is hyperbolic in the direction of $u = (1, \ldots, 1)$ for any $k = 1, \ldots, n$ (Exercise).

(1.6) Example. Differentiating $n - k$ times the polynomial of Example 1.3 we conclude that the polynomial

$$p(X) = \sum_{\substack{J \subset \{1, \ldots, n\} \\ |J| = k}} \det X_J,$$

where the sum is taken over all k-subsets $J \subset \{1, \ldots, n\}$ and X_J is the $k \times k$ submatrix of X, consisting of the entries in the rows and columns indexed by the elements of J, is hyperbolic in the direction of I (Exercise).

(1.7) Definition. Let $p : \mathbb{R}^d \longrightarrow \mathbb{R}$ be a polynomial hyperbolic in the direction of u. We define the *hyperbolicity cone* by

$$K(p, u) = \Big\{ x \in \mathbb{R}^d : \text{ the roots of the polynomial } t \longrightarrow p(x - tu) \text{ are positive} \Big\}.$$

Strictly speaking, $K(p, u)$ is not a cone as we defined them, since $K(p, u)$ may not contain 0. It is not hard to show that the closure $\overline{K(p, u)}$ of $K(p, u)$ can be defined as

$$\overline{K(p, u)} = \Big\{ x \in \mathbb{R}^d : \text{ the roots of the polynomial } t \longrightarrow p(x - tu) \text{ are non-negative} \Big\}$$

(Exercise). We obtain some familiar cones as $\overline{K(p, u)}$.

(1.8) Example. Let $p = x_1 \cdots x_d$ and let $u = (1, \ldots, 1)$, as in Example 1.2. Then

$$K(p, u) = \operatorname{int} \mathbb{R}^d_+,$$

the set of all vectors in $\mathbb{R}^d$ with all coordinates positive and

$$\overline{K(p, u)} = \mathbb{R}^d_+$$

is the non-negative orthant in $\mathbb{R}^d$.

(1.9) Example. Let $p(X) = \det X$ and let $u = I$, as in Example 1.3. Then

$$K(p, u) = \operatorname{int} \mathcal{S}_+,$$

the set of all positive definite $n \times n$ symmetric matrices and

$$\overline{K(p, u)} = \mathcal{S}_+$$

is the cone of positive semidefinite matrices.

It does not look easy to describe the cones $K(p, u)$ in Example 1.5 (except when $k = 1$ or $k = 2$) and in Example 1.6 (except when $k = 1$). It is clear though that $K(p, u) \subset K(q, u)$ if q is obtained from p as in Section 1.4 (Exercise).

(1.10) Dependence of roots of a polynomial on its coefficients. We will often say that the roots of a univariate polynomial depend continuously on its coefficients. More precisely, let

$$p(z) = a_0 + a_1 z + \ldots + a_n z^n$$

be a complex polynomial, such that $p \not\equiv 0$, so $|a_0| + \ldots + |a_n| > 0$. Let

$$D = \Big\{ z \in \mathbb{C} : \quad |z - z_0| < \delta \Big\}$$

be an open disk in the complex plane centered at z_0 and of radius $\delta > 0$ and let $S = \partial D$ be the boundary circle of D. Suppose that p has exactly k roots, counting multiplicity, in D and no roots on S. Then there is an $\epsilon > 0$ such that if

$$q(z) = b_0 + b_1 z + \ldots + b_n z^n$$

is a polynomial satisfying

$$|a_j - b_j| < \epsilon \quad \text{for} \quad j = 0, \ldots, n$$

then q also has exactly k roots, counting multiplicity, in D. Indeed, by Cauchy's formula the number of roots in D of a polynomial f with no roots in S is expressed by the contour integral

$$\frac{1}{2\pi i} \oint_S \frac{f'(z)}{f(z)} \, dz,$$

and the integral depends on f continuously.

The following result was obtained by L. Gårding [Gå59]. We follow the exposition of J. Renegar [Re06].

(1.11) Theorem. *Let $p : \mathbb{R}^d \longrightarrow \mathbb{R}$ be a homogeneous polynomial of degree $m > 0$, hyperbolic in the direction of u. Suppose that $p(u) \neq 0$. Then*

(1) *The set $K(p, u)$ is the connected component of $\mathbb{R}^d \setminus \{x : p(x) = 0\}$ that contains u;*
(2) *For any $v \in K(p, u)$, the polynomial p is hyperbolic in the direction of v;*
(3) *For any $v \in K(p, u)$ we have $K(p, v) = K(p, u)$;*
(4) *The set $K(p, u)$ is convex.*

Proof. We prove Part 1 first. Since $p(u - u) = 0$ and $p(u) \neq 0$, the only root of the polynomial

$$t \longrightarrow p(u - tu) = (1 - t)^m p(u)$$

is $t = 1$. Hence $u \in K(p, u)$. Let C be the connected component of $\mathbb{R}^d \setminus \{x : p(x) = 0\}$ that contains u. Since the roots of a polynomial depend continuously on the polynomial (Section 1.10), for all $x \in C$ the roots of the polynomial $t \longmapsto p(x - tu)$ are positive, which proves that $C \subset K(p, u)$. It remains to show that the set $K(p, u)$ is path-connected.

Let us choose any $v \in K(p, u)$ and any real $s \geq 0$. Then $v + su \in K(p, u)$ since if t_0 is a root of the polynomial $t \longmapsto p(v + su - tu) = p(v - (t - s)u)$ then $t_0 - s$ is a root of the polynomial $t \longmapsto p(v - tu)$ and hence $t_0 - s > 0$. Then $t_0 = (t_0 - s) + s > 0$.

Let us fix a $\gamma > 0$ and let $v \in \mathbb{R}^d$ be a vector such that $\|v\| \leq \gamma$. For any $s > 0$ we can write

$$p\left(v + su - tu\right) = s^m p\left(s^{-1}v + u - (ts^{-1})u\right).$$

Since for any $s > 0$ the only root of the polynomial

$$t \longmapsto p\left(u - (ts^{-1})u\right)$$

is $t = s$ and

$$\|s^{-1}v\| \ \leq \ \gamma/s \longrightarrow 0 \quad \text{as} \quad s \longrightarrow +\infty,$$

by continuity (Section 1.10), we conclude that for all sufficiently large $s \geq s_0(\gamma)$ the roots of the polynomial

$$t \longrightarrow p(v + su - tu)$$

are all positive and hence $v + su \in K(p, u)$ for all sufficiently large $s \geq s_0(\gamma)$.

Now we are ready to present a path connecting any two points $v_1, v_2 \in K(p, u)$. Let us choose a $\gamma > 0$ such that $\|v_1\|, \|v_2\| < \gamma$. Then $\|v\| < \gamma$ for all $v \in [v_1, v_2]$ and let $s_0 > 0$ be a number such that $v + s_0 u \in K(p, u)$ as long as $\|v\| < \gamma$. The path consists of the three intervals:

$$[v_1, \ v_1 + s_0 u], \quad [v_2, \ v_2 + s_0 u] \quad \text{and} \quad [v_1 + s_0 u, \ v_2 + s_0 u],$$

which concludes the proof of Part 1.

We prove Part 2 now. Let us choose any $x \in \mathbb{R}^d$ and consider the polynomial $t \longmapsto p(x - tv)$. We must show that it has real roots only. Let $i = \sqrt{-1}$ and $\alpha > 0$. Fix a real $\beta > 0$ and consider the polynomial

$$t \longmapsto p(\beta x - tv + \alpha i u). \tag{1.11.1}$$

We claim that if $t \in \mathbb{C}$ is a root of the polynomial (1.11.1) then $\Im t > 0$ (the imaginary part of t is positive). If $\beta = 0$, we get the equation $p(\alpha i u - tv) = 0$. We note that $t = 0$ is not a root since $p(u) \neq 0$. By homogeneity, we can write the equation as $p\left(v - t^{-1}\alpha i u\right) = 0$ and since $v \in K(p, u)$, for every root t we must have $\alpha t^{-1} i$ real and positive, from which it follows that $t = \gamma i$ for some $\gamma > 0$. Now, if $\Im t \leq 0$ for some $\beta_0 > 0$, by continuity (see Section 1.10), for some $\beta_0 > \beta > 0$ the polynomial (1.11.1) will have a real root t. That would mean that $-\alpha i$ is a root of the polynomial $s \longmapsto p(\beta x - tv - su)$, which contradicts to the fact that p is hyperbolic in the direction of u.

Choosing $\beta = 1$ in (1.11.1), we conclude for all $\alpha > 0$ the roots of the polynomial

$$t \longrightarrow p(x - tv + \alpha i u)$$

satisfy $\Im t > 0$. Taking the limit as $\alpha \longrightarrow 0$, by continuity (Section 1.10), we conclude that $\Im t \geq 0$ for all roots t of the polynomial $t \longmapsto p(x - tv)$, which proves that p is hyperbolic in the direction of v, since complex roots of a real polynomial come in complex conjugate pairs $a \pm bi$.

Next, we prove Part 3. By Parts 1 and 2, both $K(p, u)$ and $K(p, v)$ are connected components of $\mathbb{R}^d \setminus \{x : p(x) = 0\}$. Since $v \in K(p, u)$ and $v \in K(p, v)$, we must have $K(p, u) = K(p, v)$.

Finally, we prove Part 4. Let us choose any $v_1, v_2 \in K(p, u)$ and let $v = \alpha v_1 + (1 - \alpha)v_2$ for some $0 \leq \alpha \leq 1$. We have to prove $v \in K(p, u)$, that is, that the roots of the polynomial

$$t \longrightarrow p\left(\alpha v_1 + (1 - \alpha)v_2 - tu\right) \tag{1.11.2}$$

are positive. Since $v_1 \in K(p, u)$, all roots of (1.11.2) are positive if $\alpha = 1$. Since $v_2 \in K(p, u)$, all roots of (1.11.2) are positive if $\alpha = 0$. Suppose that for some $0 < \alpha_0 < 1$ there is a non-positive root of (1.11.2). Since the roots of (1.11.2) are real for all real α, by continuity (Section 1.10), there will be an $0 < \alpha < 1$ such that $t = 0$ is a root of (1.11.2), that is,

$$p\left(\alpha v_1 + (1 - \alpha)v_2\right) = 0.$$

Then $s = (\alpha - 1)/\alpha$ is a negative root of the polynomial

$$s \longmapsto p(v_1 - s v_2) \tag{1.11.3}$$

However, by Part 2, the polynomial p is hyperbolic in the direction of v_2 and by Part 3, we have $K\left(p, v_2\right) = K(p, u)$, so $v_1 \in K\left(p, v_2\right)$ and the roots of (1.11.3) are all positive. □

2. Permanents and stable polynomials

We follow the exposition of L. Gurvits [Gu08].

(2.1) Definition. Let $A = (a_{ij})$ be an $n \times n$ matrix and let S_n be the symmetric group of all permutations of $\{1, \dots, n\}$. The *permanent* of A is defined by

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

Another way to define $\operatorname{per} A$ is as follows. Let $x_1, \dots, x_n$ be variables and let us define a polynomial

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

Then

$$\operatorname{per} A = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} p.$$

(2.2) Definition. Let $p : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a real polynomial. We say that p is *stable* if

$$p(z_1, \dots, z_n) \neq 0 \quad \text{provided} \quad \Im z_1, \dots, \Im z_n > 0$$

(recall that $\Im z = b$ for $z = a + bi$ and $i = \sqrt{-1}$).

Suppose that p is homogeneous. It is easy to see that p is stable if and only if for any vector $u = (u_1, \dots, u_n)$ where $u_1, \dots, u_n > 0$, the polynomial p is hyperbolic in the direction of u. Indeed, let us choose an $x \in \mathbb{R}^n$, and consider the univariate polynomial

$$t \longmapsto p(x - tu), \tag{2.2.1}$$

where $x = (x_1, \dots, x_n)$, $u = (u_1, \dots, u_n)$ and $u_j > 0$ for $j = 1, \dots, n$. If $p(z_1, \dots, z_n) = 0$ where $z_j = a_j + ib_j$ and $b_j > 0$ for $j = 1, \dots, n$ then $t = -i$ is a root of (2.2.1) for $x_j = a_j$ and $u_j = b_j$, so p is not hyperbolic in the direction of u. If the polynomial (2.2.1) has a root with $\Im t \neq 0$ then, since complex roots of real polynomials come in pairs of complex conjugates, $t = a + bi$ is root of (2.2.1) for some $a, b \in \mathbb{R}$ and $b < 0$. Then $p(z_1, \dots, z_n) = 0$, where $z_j = (x_j - au_j) - (bu_j)i$ for $j = 1, \dots, n$ and hence p is not stable.

(2.3) Lemma. *Let $p(x_1, \dots, x_n)$ be a stable polynomial.*

(1) *Suppose that p contains a monomial αx_1^k for some $\alpha \neq 0$ and $k > 0$. Then the polynomial*

$$q = \frac{\partial}{\partial x_1} p$$

is stable.

(2) *Let $t \in \mathbb{R}$ be a number such that the polynomial*

$$r(x_2, \ldots, x_n) = p(t, x_2, \ldots, x_n)$$

is non-constant. Then r is stable.

Proof. To prove Part (1), let us fix any $z_2, \ldots, z_n$ such that $\Im z_j > 0$ for $j = 2, \ldots, n$ and consider a univariate polynomial

$$f(z) = p(z, z_2, \ldots, z_n).$$

Then f is non-constant, and since p is stable, all roots z of f satisfy the inequality $\Im z \leq 0$.

By the Gauss-Lucas Theorem, it follows that all roots z of $f' = q(z, z_2, \ldots, z_n)$ lie in the convex hull of the set of roots of f and hence also satisfy the inequality $\Im z \leq 0$. Therefore,

$$q(z_1, z_2, \ldots, z_n) \neq 0$$

if $\Im z_1, \ldots, \Im z_n > 0$, and hence q is stable.

To prove Part (2), suppose that $r(z_2, \ldots, z_n) = 0$ where $\Im z_2. \ldots, \Im z_n > 0$. Since r is non-constant, for some $(\alpha_2, \ldots, \alpha_n) \in \mathbb{C}^{n-1} \setminus \{0\}$, the univariate polynomial

$$f(z) = r(z_2 + \alpha_2 z, \ldots, z_n + \alpha_n z) = p(t, z_2 + \alpha_2 z, \ldots, z_n + \alpha_n z)$$

is non-constant and $z = 0$ is a root of f. By continuity (Section 1.10), for all sufficiently small $\epsilon > 0$, the polynomial

$$\tilde{f}(z) = p(t + i\epsilon, z_2 + \alpha_2 z, \ldots, z_n + \alpha_n z)$$

has a root w such that $\Im(z_2 + \alpha w), \ldots, \Im(z_n + \alpha_n w) > 0$, which contradicts the stability of p. □

(2.4) Lemma. *Suppose that a bivariate quadratic polynomial $p(x, y) = ax^2 + 2bxy + cy^2$ is stable. Then $b^2 \geq ac$.*

Proof. If $b^2 < ac$ then the univariate polynomial $ax^2 + 2bx + c$ has a pair of complex conjugate roots $\alpha \pm \beta i$ for some $\beta \neq 0$ (and hence we may assume that $\beta > 0$). By continuity (Section 1.10), for a sufficiently small $\epsilon > 0$, a point $y = 1 + \epsilon i$, $x = \tilde{\alpha} + \tilde{\beta} i$ with $\tilde{\beta} > 0$ is a root of the polynomial $p(x, y)$, which contradicts the stability of p. □

The following result is the consequence for permanents of the more general Alexandrov-Fenchel inequality for mixed volumes of convex bodies.

(2.5) Theorem. *Let A be an $n \times n$ non-negative matrix and let $a_1, \ldots, a_n$ be the columns of A. Then*

$$\operatorname{per}^2[a_1, \ldots, a_n] \;\geq\; \operatorname{per}[a_1, \ldots, a_{n-2}, a_{n-1}, a_{n-1}] \operatorname{per}[a_1, \ldots, a_{n-2}, a_n, a_n].$$

Proof. By continuity, we may assume that the entries a_{ij} of A are positive. Suppose that $\Im z_1, \ldots, \Im z_n > 0$ for some $z_1, \ldots, z_n \in \mathbb{C}$. Then

$$\Im\left(\sum_{j=1}^n a_{ij} z_j\right) > 0 \quad \text{and hence} \quad \sum_{j=1}^n a_{ij} z_j \neq 0.$$

Therefore,

$$p(z_1, \ldots, z_n) \neq 0 \quad \text{for} \quad p(x_1, \ldots, x_n) = \prod_{i=1}^n \left(\sum_{j=1}^n a_{ij} x_j\right)$$

and hence p is a stable polynomial. Note that p contains all monomials of degree n with positive coefficients. Repeatedly applying Part (1) of Lemma 2.3, we conclude that the polynomial

$$q = \frac{\partial^{n-2}}{\partial x_1 \cdots \partial x_{n-2}} p$$

is also stable. However, q is a quadratic polynomial in x_{n-1} and x_{n-2} and it is not hard to see that

$$q(x_{n-1}, x_n) = a x_{n-1}^2 + 2b x_{n-1} x_n + c x_n^2,$$

where

$$a = \frac{1}{2} \operatorname{per}[a_1, \ldots, a_{n-2}, a_{n-1}, a_{n-1}], \quad b = \frac{1}{2} \operatorname{per}[a_1, \ldots, a_n] \quad \text{and}$$
$$c = \frac{1}{2} \operatorname{per}[a_1, \ldots, a_{n-2}, a_n, a_n].$$

The proof now follows by Lemma 2.4. □

3. Stable polynomials and capacity

We follow L. Gurvits [Gu08].

(3.1) Definition. Let $p(x_1, \ldots, x_n)$ be a real polynomial with non-negative coefficients. The *capacity* of p is defined as

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

(3.2) Lemma. *Let $R(t)$ be a univariate polynomial of degree k with non-negative coefficients such that all roots of R are real. Then*

$$R'(0) \ \geq \ \left(\frac{k-1}{k}\right)^{k-1} \operatorname{cap}(R)$$

if $k > 1$ and

$$R'(0) = \operatorname{cap}(R)$$

if $k \leq 1$.

Proof. If $\deg R \leq 1$, so $R(t) = r_0 + r_1 t$ for some $r_0, r_1 \geq 0$ then clearly

$$\inf_{t>0} t^{-1} R(t) = r_1 = R'(0) \tag{3.2.1}$$

(the infimum is attained as $t \longrightarrow +\infty$) Suppose that $k \geq 2$. If $R(0) = 0$, so $R(t) = r_1 t + \ldots + r_k t^k$ for some non-negative $r_1, \ldots, r_k$, we still have (3.2.1), only that the infimum is attained as $t \longrightarrow 0$. Hence we can assume that $R(0) > 0$, and, scaling R if necessary, we assume that $R(0) = 1$.

Since the coefficients of R are non-negative, all roots are necessarily negative. Hence we can write

$$R(t) = \prod_{i=1}^{k} (1 + a_i t)$$

for some $a_1, \ldots, a_k > 0$. Then

$$R'(0) = \sum_{i=1}^{k} a_i.$$

Applying the inequality between the arithmetic and geometric means, we conclude that

$$R(t) \ \leq \ \left(1 + \frac{a_1 + \ldots + a_k}{k} t\right)^k = \left(1 + \frac{R'(0)}{k} t\right)^k.$$

Then

$$\operatorname{cap}(R) \ \leq \ \inf_{t>0} g(t) \quad \text{where} \quad g(t) = t^{-1} \left(1 + \frac{R'(0)}{k} t\right)^k.$$

Clearly $g(t) \longrightarrow +\infty$ if $t \longrightarrow +\infty$ or if $t \longrightarrow 0$, so the infimum of $g(t)$ is attained at a critical point. Solving the equation $g'(t) = 0$, we obtain

$$t = \frac{k}{(k-1)R'(0)} \quad \text{and} \quad g(t) = \left(\frac{k}{k-1}\right)^{k-1} R'(0),$$

which proves that

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

as desired. $\square$

(3.3) Remark. It is worth noting that

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

is a decreasing function of $k > 1$. Indeed, for

$$f(x) = (x-1)\ln(x-1) - (x-1)\ln x$$

we have

$$f'(x) = \ln\frac{x-1}{x} + \frac{1}{x} = \ln\left(1 - \frac{1}{x}\right) + \frac{1}{x} < 0 \quad \text{for} \quad x > 1.$$

Therefore, in Lemma 3.2 we can write

$$R'(0) \ \geq \ \left(\frac{k-1}{k}\right)^{k-1} \operatorname{cap}(R) \quad \text{provided} \quad \deg R \leq k.$$

(3.4) Theorem. *Let $p(x_1, \ldots, x_n)$ be stable polynomial of degree n with non-negative coefficients such that the coefficients of all monomials of degree n are positive. Then*

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} p \ \geq \ \frac{n!}{n^n} \operatorname{cap}(p).$$

Proof. We proceed by induction on n. For $n = 1$ we have $p(x_1) = ax_1 + b$ where $a > 0$, $b \geq 0$ and hence $p' = a$ and $\operatorname{cap}(p) = a$.

Suppose that $n > 1$. Let us fix any $x_2, \ldots, x_n > 0$ and consider the univariate polynomial $R(t) = p(t, x_2, \ldots, x_n)$. Then $\deg R = n$ and all roots of R are necessarily real, since if $R(z) = 0$ for some z with $\Im z \neq 0$ and complex roots come in pairs of complex conjugates, we may assume that $\Im z > 0$. Then, by continuity (Section 1.10), for a sufficiently small $\epsilon > 0$ the polynomial $\tilde{R}(t) = p(t, x_2 + i\epsilon, \ldots, x_n + i\epsilon)$ will have a root $\tilde{z}$ with $\Im \tilde{z} > 0$, which contradicts the stability of p. By Lemma 3.2, we have

$$\inf_{t>0} \frac{R(t)}{t} \ \leq \ \left(\frac{n}{n-1}\right)^{n-1} R'(0). \tag{3.4.1}$$

Let us define

$$q = \frac{\partial p}{\partial x_1} \quad \text{and} \quad r(x_2, \ldots, x_n) = q(0, x_2, \ldots, x_n).$$

Hence, by (3.4.1) we can write

$$
\begin{aligned}
\operatorname{cap}(p) &= \inf_{x_1,\dots,x_n>0} \frac{p\,(x_1,\dots,x_n)}{x_1\cdots x_n} \\
&\leq \left(\frac{n}{n-1}\right)^{n-1} \inf_{x_2,\dots,x_n>0} \frac{q\,(0,x_2,\dots,x_n)}{x_2\cdots x_n} \qquad (3.4.2)\\
&= \left(\frac{n}{n-1}\right)^{n-1} \operatorname{cap}(r).
\end{aligned}
$$

By Part (1) of Lemma 2.3, the polynomial q is stable and by Part (2) of Lemma 2.3, the polynomial r is stable of degree $n-1$ such that the coefficients of all monomial of degree $n-1$ are positive. By the induction hypothesis

$$
\operatorname{cap}(r) \ \leq\ \frac{(n-1)^{n-1}}{(n-1)!}\frac{\partial^{n-1}}{\partial x_2\cdots\partial x_n} r = \frac{(n-1)^{n-1}}{(n-1)!}\frac{\partial^{n}}{\partial x_1\cdots\partial x_n} p \qquad (3.4.3)
$$

Combining (3.4.3) and (3.4.2), we conclude

$$
\operatorname{cap}(p) \ \leq\ \left(\frac{n}{n-1}\right)^{n-1} \frac{(n-1)^{n-1}}{(n-1)!}\frac{\partial^{n}}{\partial x_1\cdots\partial x_n} p = \frac{n^n}{n!}\frac{\partial^{n}}{\partial x_1\cdots\partial x_n} p
$$

and the proof follows. □

(3.5) Remark. Suppose that $p\,(x_1,\dots,x_n)$ is a stable homogeneous polynomial of degree n with non-negative coefficients and that the degree of p in x_i is k_i for $i=1,\dots,n$. One can show that

$$
\frac{\partial^{n}}{\partial x_1\cdots\partial x_n} p \ \geq\ \operatorname{cap}(p)\prod_{i=1}^{n}\left(\frac{k_i-1}{k_i}\right)^{k_i-1}. \qquad (3.5.1)
$$

Indeed, p is hyperbolic in any direction $u=(u_1,\dots,u_n)$ where $u_1,\dots,u_n>0$ (see Definition 2.2) and hence by Theorem 1.11 so is its derivative $\partial p/\partial u$. To prove (3.5.1), in the proof of Theorem 3.4, instead of taking partial derivatives $\partial p/\partial x_i$, we take the derivative $\partial p/\partial u_i$, where the i-th coordinate of u_i is 1 and all other coordinates are ϵ for some small $\epsilon>0$ and notice that the coefficients of monomials of $R(t)$ of degree higher than k_i are $O(\epsilon)$, so taking the limit as $\epsilon \longrightarrow 0$, at the i-th step we can replace (3.4.1) by

$$
\inf_{t>0}\frac{R(t)}{t} \ \leq\ \left(\frac{k_i}{k_i-1}\right)^{k_i-1} R'(0).
$$

4. Capacity, permanents, and doubly stochastic matrices

We recall the definition of a convex function.

(4.1) Definition. A function $f:\mathbb{R}^d \longrightarrow \mathbb{R}$ is called *convex* if

$$f\big(\alpha x+(1-\alpha)y\big) \ \leq \ \alpha f(x)+(1-\alpha)f(y)$$
$$\text{for all} \quad x,y\in\mathbb{R}^d \quad \text{and all} \quad 0\leq\alpha\leq 1.$$

(4.2) Lemma. *Let $\lambda_1, \ldots, \lambda_n$ be reals and let $\alpha_1, \ldots, \alpha_n$ be positive reals. Then the function $f:\mathbb{R}\longrightarrow\mathbb{R}$,*

$$f(t)=\ln\left(\sum_{k=1}^{n}\alpha_k e^{\lambda_k t}\right)$$

is convex.

Proof. It suffices to check that $f''(t)\geq 0$ for all $t\in\mathbb{R}$. Writing

$$f(t)=\ln g(t) \quad \text{where} \quad g(t)=\sum_{k=1}^{n}\alpha_k e^{\lambda_k t},$$

we compute

$$f'(t)=\frac{g'(t)}{g(t)} \quad \text{and} \quad f''(t)=\frac{g''(t)g(t)-g'(t)g'(t)}{g^2(t)}.$$

Now,

$$\begin{aligned}
g''(t)g(t)-g'(t)g'(t) &= \sum_{i,j=1}^{n}\lambda_i^2\alpha_i\alpha_j e^{(\lambda_i+\lambda_j)t}-\sum_{i,j=1}^{n}\lambda_i\lambda_j\alpha_i\alpha_j e^{(\lambda_i+\lambda_j)t}\\
&=\sum_{\substack{\{i,j\}\\ i\neq j}}\left(\lambda_i^2+\lambda_j^2-2\lambda_i\lambda_j\right)\alpha_i\alpha_j e^{(\lambda_i+\lambda_j)t}\\
&=\sum_{\substack{\{i,j\}\\ i\neq j}}\left(\lambda_i-\lambda_j\right)^2\alpha_i\alpha_j e^{(\lambda_i+\lambda_j)t}\geq 0
\end{aligned}$$

and the proof follows. □

(4.3) Corollary. *Let $p(x_1, \ldots, x_n)$ be a real polynomial with non-negative coefficients. Then the function $f:\mathbb{R}^n\longrightarrow\mathbb{R}$,*

$$f(t_1, \ldots, t_n)=\ln p\left(e^{t_1}, \ldots, e^{t_n}\right)$$

is convex.

Proof. It suffices to prove that the restriction of f onto every line in $\mathbb{R}^n$ is convex, that is,

$$f(\alpha_1+\beta_1 t, \ldots, \alpha_n+\beta_n t)=\ln p\left(e^{\alpha_1}e^{\beta_1 t}, \ldots, e^{\alpha_n}e^{\beta_n t}\right)$$

is a convex function of $t\in\mathbb{R}$. This follows now from Lemma 4.2. □

(4.4) Definition. An $n \times n$ matrix $A = (a_{ij})$ is called *doubly stochastic* if

$$\sum_{j=1}^{n} a_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, n, \quad \sum_{i=1}^{n} a_{ij} = 1 \quad \text{for} \quad j = 1, \ldots, n \quad \text{and}$$
$$a_{ij} \geq 0 \quad \text{for all} \quad i, j.$$

(4.5) Lemma. *Let $A = (a_{ij})$ be an $n \times n$ doubly stochastic matrix and let*

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

Then

$$\operatorname{cap}(p) = 1.$$

Proof. Since p is a homogeneous polynomial of degree n, we can write

$$\operatorname{cap}(p) = \inf_{\substack{x_1, \ldots, x_n > 0 \\ x_1 \cdots x_n = 1}} p(x_1, \ldots, x_n).$$

Substituting $x_i = e^{t_i}$, we conclude that

$$\operatorname{cap}(p) = \exp\left\{ \inf_{t_1 + \ldots + t_n = 0} f(t_1, \ldots, t_n) \right\} \quad \text{where}$$
$$f(t_1, \ldots, t_n) = \ln p\left(e^{t_1}, \ldots, e^{t_n}\right).$$

We claim that $t_1 = \ldots = t_n = 0$ is a critical point of f on the hyperplane $t_1 + \ldots + t_n = 0$. Computing the gradient of f at $t_1 = \ldots = t_n = 0$, we obtain

$$\frac{\partial f}{\partial t_j} = \sum_{i=1}^{n} \frac{a_{ij} e^{t_j}}{\sum_{j=1}^{n} a_{ij} e^{t_j}}$$

and hence

$$\frac{\partial f}{\partial t_j}\Big|_{t_1 = \ldots = t_n = 0} = \sum_{i=1}^{n} a_{ij} = 1,$$

where in the first equality we used that the column sums of A are 1's and in the second equality we used that the row sums of A are 1's.

Hence the gradient of f at $t_1 = \ldots = t_n = 0$ is orthogonal to the hyperplane $t_1 + \ldots + t_n = 0$ and so $t_1 = \ldots = t_n = 0$ is a critical point of $f(t)$ on the hyperplane. Since by Corollary 4.3 the function f is convex, we conclude that $t_1 = \ldots = t_n = 0$ is the minimum point of f on the hyperplane. Since $f(0, \ldots, 0) = 0$, the proof follows. □

Now we are ready to prove the famous van der Waerden inequality for permanents.

(4.6) Theorem. *Let A be an $n \times n$ doubly stochastic matrix. Then*

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

Proof. By continuity, without loss of generality we may assume that $a_{ij} > 0$ for all i, j. We define the polynomial $p(x_1, \ldots, x_n)$ as in Lemma 4.5. As in the proof of Theorem 2.5, we establish that p is stable. By Lemma 4.5, we have $\operatorname{cap}(p) = 1$, so by Theorem 3.4,

$$\operatorname{per} A = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} p \ \geq \ \frac{n!}{n^n} \operatorname{cap}(p) = \frac{n!}{n^n}.$$

□

(4.7) Remark. Suppose that A is doubly stochastic and contains not more than k non-zero entries in every column. Then the degree of p in every variable $x_1, \ldots, x_n$ does not exceed k. Replacing every zero entry a_{ij} by a small $\epsilon > 0$, and running the proof of Theorem 3.4, we observe that in (3.4.1), the coefficients of $R(t)$ of degree $k+1$ and higher are all $O(\epsilon)$. Therefore, as $\epsilon \longrightarrow 0$, we can replace $\left(\frac{n}{n-1}\right)^{n-1}$ in (3.4.1) by $\left(\frac{k}{k-1}\right)^{k-1}$. Hence we get the inequality

$$\operatorname{per} A \ \geq \ \left(\frac{k-1}{k}\right)^{(k-1)n}$$

(A. Schrijver's bound), see also Remark 3.5.

5. Matrix scaling and permanents

(5.1) Theorem. *Let $A = (a_{ij})$ be an $n \times n$ matrix such that $a_{ij} > 0$ for all i, j. Then there exists a doubly stochastic matrix $B = (b_{ij})$ and positive $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$ such that*

$$a_{ij} = b_{ij} \lambda_i \mu_j \quad \textit{for all} \quad i, j.$$

Proof. As in the proof of Lemma 4.5, we define a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$f(t_1, \ldots, t_n) = \sum_{i=1}^n \ln \left(\sum_{j=1}^n a_{ij} e^{t_j} \right)$$

and consider its minimum on the hyperplane $H \subset \mathbb{R}^n$ defined by the equation $t_1 + \ldots + t_n = 0$. First, we claim that the minimum of f on H is attained at some point. Let

$$M = \max_{i,j} \ln \frac{f(0, \ldots, 0)}{a_{ij}}.$$

If $t_j > M$ for some j, we have $f(t_1, \dots, t_n) > f(0, \dots, 0)$. On the other hand, since $t_1 + \dots + t_n = 0$, if $t_j < -nM$ for some j then $t_k > M$ for some $k \neq j$. Hence the minimum of f on the compact set

$$\{(t_1, \dots, t_n) : \quad |t_j| \leq nM \quad \text{for} \quad j = 1, \dots, n\} \cap H$$

is the minimum of f on H.

Let $t^* = (t_1^*, \dots, t_n^*)$ be the minimum point. Then the gradient of f at t^* should be proportional to the normal vector to H and hence for some α

$$\text{(5.1.1)} \qquad \frac{\partial f}{\partial t_j}\Big|_{t=t^*} = \sum_{i=1}^n \frac{a_{ij} e^{t_j^*}}{\sum_{k=1}^n a_{ik} e^{t_k^*}} = \alpha \quad \text{for} \quad j = 1, \dots, n.$$

Since

$$\sum_{j=1}^n \sum_{i=1}^n \frac{a_{ij} e^{t_j^*}}{\sum_{k=1}^n a_{ik} e^{t_k^*}} = n,$$

we conclude that $\alpha = 1$.

Let us define

$$\lambda_i = \sum_{k=1}^n a_{ik} e^{t_k^*} \quad \text{and} \quad \mu_j = e^{-t_j^*} \quad \text{for all} \quad i, j.$$

Then

$$a_{ij} = b_{ij} \lambda_i \mu_j \quad \text{where} \quad b_{ij} = \frac{a_{ij} e^{t_j^*}}{\sum_{k=1}^n a_{ik} e^{t_k^*}}.$$

Clearly, $B = (b_{ij})$ us a non-negative matrix and

$$\sum_{j=1}^n b_{ij} = 1 \quad \text{for} \quad i = 1, \dots, n.$$

From (5.1.1) with $\alpha = 1$ we get

$$\sum_{i=1}^n b_{ij} = 1 \quad \text{for} \quad j = 1, \dots, n.$$

□

(5.2) Scaling and permanents. Given a positive $n \times n$ matrix A, let us compute the numbers $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$ as in Theorem 5.1. Then

$$\operatorname{per} A = \left(\prod_{i=1}^n \lambda_i \right) \left(\prod_{j=1}^n \mu_j \right) \operatorname{per} B$$

and

$$\frac{n!}{n^n} \leq \operatorname{per} B \leq 1.$$

This allows us to estimate $\operatorname{per} A$ within a factor of $n!/n^n \approx e^{-n}$.

Exercises.

Prove that the numbers $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$ in Theorem 5.1 are unique up to an obvious rescaling:

$$\lambda_i := \lambda_i \tau, \quad \mu_j = \mu_j \tau^{-1} \quad \text{for all} \quad i, j.$$

This allows us to define a function F on positive $n \times n$ matrices by

$$F(A) = \left(\prod_{i=1}^{n} \lambda_i\right)\left(\prod_{j=1}^{n} \mu_j\right).$$

Prove that F is *log-concave*:

$$F\left(\frac{1}{2}A + \frac{1}{2}B\right) \ \geq \ \sqrt{F(A)F(B)}$$

for any two positive $n \times n$ matrices A and B.

6. Ramifications: mixed discriminants

We follow mostly L. Gurvits and A. Samorodnitsky [GS02] and L. Gurvits [Gu08].

(6.1) Definition. Let $Q_1, \ldots, Q_n$ be $n \times n$ real symmetric matrices. Then

$$p\left(x_1, \ldots, x_n\right) = \det\left(x_1 Q_1 + \ldots + x_n Q_n\right)$$

is a homogeneous polynomial of degree n and the mixed term

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} p = D\left(Q_1, \ldots, Q_n\right)$$

is called the *mixed discriminant* of $Q_1, \ldots, Q_n$.

(6.2) Lemma. *Suppose that the matrices $Q_1, \ldots, Q_n$ are positive semidefinite. Then*

$$D\left(Q_1, \ldots, Q_n\right) \geq 0.$$

Proof. Since $D\left(Q_1, \ldots, Q_n\right)$ is a continuous function of $Q_1, \ldots, Q_n$, without loss of generality we may assume that $Q_i \succ 0$ for $i = 1, \ldots, n$. We proceed by induction on n. Clearly, the statement is true for $n = 1$. Suppose that $n > 1$. Since $Q_1 \succ 0$, we can write $Q_1 = TT^*$ for some invertible $n \times n$ matrix T and then

$$D\left(Q_1, \ldots, Q_n\right) = (\det T)^2 D\left(I, T^{-1} Q_2 \left(T^*\right)^{-1}, \ldots, T^{-1} Q_n \left(T^*\right)^{-1}\right),$$

where I is an $n \times n$ identity matrix and the matrices $Q_i' = T^{-1} Q_i \left(T^*\right)^{-1}$ are positive semidefinite. Thus is suffices to prove that

$$D\left(I, Q_2, \ldots, Q_n\right) > 0 \quad \text{whenever} \quad Q_2, \ldots, Q_n \succ 0.$$

It is not hard to see that

$$D\left(I, Q_2, \ldots, Q_n\right) = \sum_{\substack{J \subset \{1, \ldots, n\} \\ |J| = n-1}} D\left(Q_2(J), \ldots, Q_n(J)\right),$$

where the sum is taken over all $(n-1)$-subsets of $\{1, \ldots, n\}$ and $Q_i(J)$ is the $(n-1) \times (n-1)$ submatrix of Q_i consisting of the entries with the row and column in J. Since $Q_i(J) \succ 0$ provided $Q_i \succ 0$, the proof follows. □

Exercises.

1. Let $u_1, \ldots, u_n$ be vectors from $\mathbb{R}^n$. Prove that

$$D\left(u_1 \otimes u_1, \ldots, u_n \otimes u_n\right) = \left(\det\left[u_1, \ldots, u_n\right]\right)^2,$$

where $[u_1, \ldots, u_n]$ is the $n \times n$ matrix with columns $u_1, \ldots, u_n$.

2. Let G be a connected graph with n vertices and m edges, colored with $n-1$ different colors. We introduce an arbitrary orientation on the edges of G and define the *incidence matrix* of G as an $n \times m$ matrix $A = (a_{ij})$ where

$$a_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ is the beginning of edge } j, \\ -1 & \text{if vertex } i \text{ is the end of edge } j, \\ 0 & \text{elsewhere.} \end{cases}$$

Let us remove an arbitrary row of A and let $a_1, \ldots, a_m$ be the columns of the resulting matrix, interpreted as vectors from $\mathbb{R}^{n-1}$. For $k = 1, \ldots, n-1$, let $J_k \subset \{1, \ldots, n\}$ be the set of edges of G colored with the k-th color and let

$$Q_k = \sum_{j \in J_k} a_j \otimes a_j.$$

Prove that $D\left(Q_1, \ldots, Q_{n-1}\right)$ is the number of spanning trees in G having exactly 1 edge of each color.

(6.3) Lemma. *Suppose that $Q_1, \ldots, Q_n \succ 0$. Then the polynomial*

$$p\left(x_1, \ldots, x_n\right) = \det\left(x_1 Q_1 + \ldots + x_n Q_n\right)$$

is a stable homogeneous polynomial of degree n and the coefficient of every monomial of p of degree n is positive.

Proof. Let us choose any $z_1, \ldots, z_n$ such that $\Im z_j > 0$ for $j = 1, \ldots, n$ and suppose that $p(z_1 Q_1 + \ldots + z_n Q_n) = 0$. Then the matrix

$$Q = \sum_{j=1}^{n} z_j Q_j$$

is not invertible and hence there is a vector $x \in \mathbb{C}^n \setminus \{0\}$ such that $Qx = 0$. Let us consider the standard inner product

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \overline{y_i}$$

in $\mathbb{C}^n$. Then

$$0 = \langle Qx, x \rangle = \sum_{i=1}^{n} z_i \langle Q_i x, x \rangle.$$

On the other hand, $\langle Q_i x, x \rangle$ are positive real numbers and we obtain a contradiction.

The coefficient of $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ in $p(x_1, \ldots, x_n)$ is

$$D\left(\underbrace{Q_1, \ldots, Q_1}_{\alpha_1 \text{ times}}, \ldots, \underbrace{Q_n, \ldots, Q_n}_{\alpha_n \text{ times}}\right)$$

and hence by Lemma 6.2 is positive. □

(6.4) Lemma. *Let $Q_1, \ldots, Q_n$ be $n \times n$ positive definite matrices such that*

$$\sum_{i=1}^{n} Q_i = I \quad \text{and} \quad \operatorname{tr} Q_i = 1 \quad \text{for} \quad i = 1, \ldots, n.$$

Then, for

$$p(x_1, \ldots, x_n) = \det(x_1 Q_1 + \ldots + x_n Q_n),$$

we have

$$\operatorname{cap}(p) = 1.$$

Proof. Let us define $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$f(t_1, \ldots, t_n) = \ln \det \left(\sum_{i=1}^{n} e^{t_i} Q_i \right)$$

and let $H \subset \mathbb{R}^n$ be the hyperplane defined by the equation $t_1 + \ldots + t_n = 0$. By Lemma 6.2 and Corollary 4.3, the function f is convex. It suffices to prove that the minimum of f on H is attained at $t_1 = \ldots, t_n = 0$, for which it suffices to prove that the gradient of f at $t_1 = \ldots = t_n = 0$ is proportional to the vector $(1, \ldots, 1)$. Since

$$\nabla (\ln \det X) = \left(X^T\right)^{-1},$$

denoting

$$S(t) = \sum_{i=1}^{n} e^{t_i} Q_i$$

we conclude that

$$\frac{\partial f}{\partial t_i} = \left\langle e^{t_i} Q_i,\ S^{-1}(t) \right\rangle = e^{t_i} \operatorname{tr}\left(Q_i S^{-1}(t)\right). \tag{6.4.1}$$

Hence

$$\left.\frac{\partial f}{\partial t_i}\right|_{t_1 = \ldots = t_n = 0} = 1$$

and the proof follows. □

The following result confirms a conjecture of Bapat.

(6.5) Theorem. *Let $Q_1, \ldots, Q_n$ be $n \times n$ positive semidefinite matrices such that*

$$\sum_{i=1}^{n} Q_i = I \quad \textit{and} \quad \operatorname{tr} Q_i = 1 \quad \textit{for} \quad i = 1, \ldots, n.$$

Then

$$D(Q_1, \ldots, Q_n) \ \geq \ \frac{n!}{n^n}.$$

Proof. Without loss of generality we assume that $Q_1, \ldots, Q_n \succ 0$. The proof follows by Lemma 6.3, Theorem 3.4 and Lemma 6.4. □

Here is a version of scaling for mixed discriminants.

(6.6) Theorem. *Let $Q_1, \ldots, Q_n$ be $n \times n$ positive definite matrices. Then there are $n \times n$ positive definite matrices $B_1, \ldots, B_n$, an invertible $n \times n$ matrix T and positive reals $\lambda_1, \ldots, \lambda_n$ such that*

$$\sum_{i=1}^{n} B_i = I, \quad \operatorname{tr} B_i = 1 \quad \textit{and} \quad Q_i = \lambda_i T B_i T^* \quad \textit{for} \quad i = 1, \ldots, n.$$

Proof. As in the proof of Lemma 6.4, we define a convex function

$$f(t_1, \ldots, t_n) = \ln \det \left(\sum_{i=1}^{n} e^{t_i} Q_i \right)$$

and the hyperplane H defined by the equation $t_1 + \ldots + t_n = 0$. It is not hard to see (cf. the proof of Theorem 5.1) that f attains its minimum on H at some point $t_1^*, \ldots, t_n^*$, at which point the gradient of f is proportional to the vector $(1, \ldots, 1)$. By (6.4.1), we obtain that for some α and

$$S = \sum_{i=1}^n e^{t_i^*} Q_i,$$

we have

(6.6.1) $$e^{t_i^*} \operatorname{tr}\left(Q_i S^{-1}\right) = \alpha \quad \text{for} \quad i = 1, \ldots, n.$$

Since

$$n\alpha = \sum_{i=1}^n e^{t_i^*} \operatorname{tr}\left(Q_i S^{-1}\right) = \operatorname{tr}\left(\sum_{i=1}^n e^{t_i^*} Q_i S^{-1}\right) = \operatorname{tr}\left(S S^{-1}\right) = n,$$

we conclude that

(6.6.2) $$\alpha = 1$$

Since $S \succ 0$, we can write $S = TT^*$ for an invertible $n \times n$ matrix T. Then

(6.6.3) $$\operatorname{tr}\left(Q_i S^{-1}\right) = \operatorname{tr}\left(Q_i \left(T^{-1}\right)^* T^{-1}\right) = \left(T^{-1} Q_i \left(T^{-1}\right)^*\right)$$

and we define

$$B_i = e^{t_i^*} T^{-1} Q_i \left(T^{-1}\right)^* \quad \text{and} \quad \lambda_i = e^{-t_i^*} \quad \text{for} \quad i = 1, \ldots, n.$$

Clearly, $B_1, \ldots, B_n$ are positive definite matrices and

$$Q_i = \lambda_i T B_i T^* \quad \text{for} \quad i = 1, \ldots, n.$$

By (6.6.1)–(6.6.3) we have

$$\operatorname{tr} B_i = 1 \quad \text{for} \quad i = 1, \ldots, n.$$

Finally,

$$\sum_{i=1}^n B_i = \sum_{i=1}^n e^{t_i^*} T^{-1} Q_i \left(T^{-1}\right)^* = T^{-1} \left(\sum_{i=1}^n e^{t_i^*} Q_i\right) \left(T^{-1}\right)^* = T^{-1} S \left(T^*\right)^{-1} = I.$$

□

We note that

$$D\left(Q_1, \ldots, Q_n\right) = (\det T)^2 \left(\prod_{i=1}^n \lambda_i\right) D\left(Q_1, \ldots, Q_n\right).$$

Exercise.

Prove that $D\left(Q_1, \ldots, Q_n\right) \leq 1$, where $Q_1, \ldots, Q_n$ are positive semidefinite matrices such that $Q_1 + \ldots + Q_n = I$.

7. Upper bounds for permanents

Our goal is to prove the following inequality conjectured by Minc and proved by Bregman.

(7.1) Theorem. *Let $A = (a_{ij})$ be an $n \times n$ matrix such that $a_{ij} \in \{0, 1\}$ for all i and j and let*

$$r_i = \sum_{j=1}^{n} a_{ij} \quad \textit{for} \quad i = 1, \ldots, n.$$

Then

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

If all r_i are equal, the inequality is sharp, as the example of a block-diagonal matrix with n/r diagonal $r \times r$ blocks filled by 1's demonstrates.

The following corollary is due to A. Samorodnitsky.

(7.2) Corollary. *Suppose that $A = (a_{ij})$ is a stochastic $n \times n$ matrix, that is $a_{ij} \geq 0$ for all i, j and*

$$\sum_{j=1}^{n} a_{ij} = 1 \quad \textit{for all} \quad i = 1, \ldots n. \tag{7.2.1}$$

Suppose further that

$$a_{ij} \ \leq \ \frac{1}{b_i} \quad \textit{for} \quad j = 1, \ldots, n \tag{7.2.2}$$

and some positive integer $b_1, \ldots, b_n$. Then

$$\operatorname{per} A \ \leq \ \prod_{i=1}^{n} \frac{(b_i!)^{1/b_i}}{b_i}.$$

Proof. Let us fix all but the i-th row of an $n \times n$ matrix A. Then $\operatorname{per} A$ is a linear function in $a_i = (a_{i1}, \ldots, a_{in})$. Let us consider the polytope P_i of all n-vectors $a_i = (a_{i1}, \ldots, a_{in})$ such that (7.2.1) and (7.2.2) hold. Then the maximum of $\operatorname{per} A$ on P_i is attained at an extreme point of P_i, which necessarily has $a_{ij} \in \{0, 1/b_{ij}\}$ for all j. Indeed, if $0 < a_{ij_1} < 1/b_i$ for some j_1 then there will be another $j_2 \neq j_1$ such that $0 < a_{ij_2} < 1/b_i$ (we use that b_i is an integer) and the perturbation $a_{ij_1} := a_{ij_1} \pm \epsilon$, $a_{ij_2} := a_{ij_2} \mp \epsilon$ shows that a_i is not an extreme point of P_i. Hence the maximum point of $\operatorname{per} A$ on the matrices satisfying (7.2.1) and (7.2.2) is attained at a matrix A where $a_{ij} \in \{0, 1/b_{ij}\}$ for all i and j. Let B be the matrix obtained from A by multiplying the i-th row by b_i. Then

$$\operatorname{per} A = \left(\prod_{i=1}^{n} \frac{1}{b_i} \right) \operatorname{per} B \quad \text{and} \quad \operatorname{per} B \ \leq \ \prod_{i=1}^{n} (b_i!)^{1/b_i}$$

by Theorem 7.1. □

(7.3) Permanents of doubly stochastic matrices with small entries.

Together with the van der Waerden bound (Theorem 4.6), the Bregman-Minc bound (Theorem 7.1) implies that per A does not vary much if A is a doubly stochastic matrix with small entries. Indeed, suppose that A is an $n \times n$ doubly stochastic matrix. Then, by Theorem 4.6, we have

$$\ln \operatorname{per} A \ \geq \ \ln \frac{n!}{n} = -n + O(\ln n) \quad \text{as} \quad n \longrightarrow +\infty$$

by Stirling's formula. Suppose additionally that

$$a_{ij} \ \leq \ \frac{1}{b} \quad \text{for all} \quad i, j$$

and some positive integer b. Then, by Corollary 7.2,

$$\ln \operatorname{per} A \ \leq \ \frac{n}{b} \ln b! - n \ln b = -n + O\left(\frac{n \ln b}{b}\right) \quad \text{as} \quad b \longrightarrow +\infty.$$

In other words, the permanent of an $n \times n$ doubly stochastic matrix with uniformly small entries is close to e^{-n}.

We present A. Schriver's proof of Theorem 7.1 [Sc78].

(7.4) Lemma. *For positive $t_1, \ldots, t_r$ we have*

$$\left(\sum_{i=1}^r t_r\right)^{\sum_{i=1}^r t_r} \ \leq \ \left(r^{\sum_{i=1}^r t_r}\right) \prod_{i=1}^r t_i^{t_i}.$$

Proof. We observe that $f(x) = x \ln x$ is convex for $x > 0$. Indeed, $f'(x) = \ln x + 1$ and $f''(x) = 1/x > 0$. Therefore,

$$f\left(\frac{1}{r} \sum_{i=1}^r t_i\right) \ \leq \ \frac{1}{r} \sum_{i=1}^r f(t_i).$$

Exponentiating both sides of the inequality, we get the desired result. □

We will also use the following obvious row-expansion formula for the permanent.

(7.5) Lemma. *Let $A = (a_{ij})$ be an $n \times n$ matrix. For $1 \leq i, k \leq n$, let A_{ik} be the $(n-1) \times (n-1)$ matrix obtained from A by crossing out the i-th row and k-th column. Then, for any $1 \leq i \leq n$, we have*

$$\operatorname{per} A = \sum_{k=1}^n a_{ik} \operatorname{per} A_{ik}.$$

□

(7.6) Proof of Theorem 7.1. We proceed by induction on n. The case of $n=1$ is clear. Suppose that $n>1$. Without loss of generality, we may assume that $\operatorname{per} A>0$. We bound the expression

$$(7.6.1)\qquad (\operatorname{per} A)^{n \operatorname{per} A} = \prod_{i=1}^{n} (\operatorname{per} A)^{\operatorname{per} A}.$$

To bound the i-th factor in the product, we use the i-th row expansion together with Lemma 7.4. Since $a_{ik} \in \{0,1\}$, from Lemma 7.5, we can write

$$\operatorname{per} A = \sum_{k:\ a_{ik}=1} \operatorname{per} A_{ik}.$$

Letting $t_k = \operatorname{per} A_{ik}$, from Lemma 7.4 we obtain

$$(7.6.2)\qquad (\operatorname{per} A)^{\operatorname{per} A} \ \leq \ r_i^{\operatorname{per} A} \prod_{k:\ a_{ik}=1} (\operatorname{per} A_{ik})^{\operatorname{per} A_{ik}}.$$

Let S_n be the symmetric group of all permutations of the set $\{1, \ldots, n\}$ and let

$$S = \bigl\{\sigma \in S_n : \quad a_{i\sigma(i)} = 1 \quad \text{for} \quad i=1, \ldots, n\bigr\}$$

be the set of all permutations contributing to $\operatorname{per} A$. Then

$$(7.6.3)\qquad |S| = \operatorname{per} A \quad \text{and} \quad |\{\sigma \in S : \quad \sigma(i)=k\}| = \begin{cases} \operatorname{per} A_{ik} & \text{if } a_{ik}=1 \\ 0 & \text{if } a_{ik}=0. \end{cases}$$

It follows from (7.6.3) that

$$(7.6.4)\qquad \prod_{\sigma \in S} \left(\prod_{i=1}^{n} r_i \operatorname{per} A_{i\sigma(i)} \right) = r_i^{\operatorname{per} A} \prod_{k:\ a_{ik}=1} (\operatorname{per} A_{ik})^{\operatorname{per} A_{ik}}.$$

Now we apply the induction hypothesis to each of the $(n-1)\times(n-1)$ matrix $A_{i\sigma(i)}$. The rows of $A_{i\sigma(i)}$ are obtained from the rows of A by crossing out the $(j, \sigma(i))$-th entry of A for $j \neq i$ and crossing out the i-th row entirely. Hence, applying the induction hypothesis, we obtain

$$\operatorname{per} A_{i\sigma(i)} \ \leq \prod_{\substack{j:\ j\neq i \\ a_{j\sigma(i)}=0}} (r_j!)^{1/r_j} \prod_{\substack{j:\ j\neq i \\ a_{j\sigma(i)}=1}} (r_j-1)!^{1/(r_j-1)}.$$

Let us fix any permutation $\sigma \in S$. Then

$$(7.6.5)\qquad \left(\prod_{i=1}^{n} r_i \operatorname{per} A_{i\sigma(i)} \right) \ \leq \ \prod_{i=1}^{n} \left(r_i \prod_{\substack{j:\ j\neq i \\ a_{j\sigma(i)}=0}} (r_j!)^{1/r_j} \prod_{\substack{j:\ j\neq i \\ a_{j\sigma(i)}=1}} (r_j-1)!^{1/(r_j-1)} \right).$$

Now, for any $j = 1, \ldots, n$ the number of indices $i \neq j$ such that $a_{j\sigma(i)} = 0$ is precisely $n - r_j$ whereas the number of indices $i \neq j$ such that $a_{j\sigma(i)} = 1$ is precisely $r_j - 1$. Hence, for any $\sigma \in S$ we have

$$\prod_{i=1}^{n} \left(r_i \prod_{\substack{j:\ j \neq i \\ a_{j\sigma(i)}=0}} (r_j!)^{1/r_j} \prod_{\substack{j:\ j \neq i \\ a_{j\sigma(i)}=1}} (r_j - 1)!^{1/(r_j-1)} \right) = \prod_{j=1}^{n} \Big(r_j \, (r_j!)^{(n-r_j)/r_j} \, (r_j - 1)! \Big) = \prod_{j=1}^{n} (r_j!)^{n/r_j} . \tag{7.6.6}$$

Combining (7.6.1)–(7.6.6), we obtain

$$(\operatorname{per} A)^{n \operatorname{per} A} \leq \prod_{\sigma \in S} \left(\prod_{j=1}^{n} (r_j!)^{n/r_j} \right) = \left(\prod_{j=1}^{n} (r_j!)^{1/r_j} \right)^{n \operatorname{per} A},$$

and the proof follows. □

References

[Gå59] L. Gårding, *An inequality for hyperbolic polynomials*, J. Math. Mech. **8** (1959), 957–965.

[Gu08] L. Gurvits, *Van der Waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. With a corrigendum*, Research Paper 66, 26 pp, Electron. J. Combin. **15** (2008).

[GS02] L. Gurvits and A. Samorodnitsky, *A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary*, Discrete Comput. Geom. **27** (2002), 531–550.

[Re06] J. Renegar, *Hyperbolic programs, and their derivative relaxations*, Found. Comput. Math. **6** (2006), 59–79.

[Sc78] A. Schrijver, *A short proof of Minc's conjecture*, J. Combinatorial Theory Ser. A **25** (1978), 80–83.