BRUNN-MINKOWSKI INEQUALITIES FOR CONTINGENCY TABLES AND INTEGER FLOWS

Alexander Barvinok

April 2006

Abstract. We establish approximate log-concavity for a wide family of combinatorially defined integer-valued functions. Examples include the number of non-negative integer matrices (contingency tables) with prescribed row and column sums (margins), as a function of the margins and the number of integer feasible flows in a network, as a function of the excesses at the vertices. As a corollary, we obtain approximate log-concavity for the Kostant partition function of type A. We also present an indirect evidence that at least some of the considered functions might be genuinely log-concave.

1. Introduction

(1.1) The Brunn-Minkowski inequality. The famous Brunn-Minkowski inequality states that for bounded Borel sets $A, B \subset \mathbb{R}^d$ and non-negative numbers α, β such that $\alpha + \beta = 1$ one has

$$\operatorname{vol}(\alpha A + \beta B) \geq \operatorname{vol}^{\alpha}(A) \operatorname{vol}^{\beta}(B),$$

where vol is the usual volume (Lebesgue measure) in Euclidean space $\mathbb{R}^d$ and

$$\alpha A + \beta B = \{\alpha x + \beta y : \quad x \in A, y \in B\}.$$

The inequality extends to finite families of sets in an obvious way: if $A_1, \ldots, A_p \subset \mathbb{R}^d$ are bounded Borel sets and $\alpha_1, \ldots, \alpha_p$ are non-negative numbers such that $\alpha_1 + \ldots + \alpha_p = 1$ then

$$\operatorname{vol}\left(\alpha_1 A_1 + \ldots + \alpha_p A_p\right) \geq \prod_{k=1}^{p} \operatorname{vol}^{\alpha_k}\left(A_k\right). \tag{1.1.1}$$

1991 *Mathematics Subject Classification.* 05A16, 52B12, 52B20, 52A41.

Key words and phrases. contingency tables, permanent, Brunn-Minkowski inequality, flow polytopes, integer points, log-concave functions, matrix scaling.

This research was partially supported by NSF Grant DMS 0400617.

Typeset by $\mathcal{AMS}$-TeX

The Brunn-Minkowski inequality plays an important role in almost all branches of mathematics, see [Ga02] for a survey. Inequality (1.1.1) was extended and generalized in numerous direction. In particular, we need its functional version, known as the Prékopa-Leindler inequality:

let $\alpha_1, \ldots, \alpha_p$ be non-negative numbers such that $\alpha_1 + \ldots + \alpha_p = 1$ and let $g, h_1, \ldots, h_p : \mathbb{R}^d \longrightarrow \mathbb{R}$ be Borel measurable non-negative functions such that

$$g\left(\alpha_1 x_1 + \ldots + \alpha_p x_p\right) \geq \prod_{k=1}^{p} h_k^{\alpha_k}(x_k) \quad \text{for all} \quad x_1, \ldots, x_k \in \mathbb{R}^d.$$

Then

$$\int_{\mathbb{R}^d} g(x)\ dx \geq \prod_{k=1}^{p} \left(\int_{\mathbb{R}^d} h_k(x)\ dx\right)^{\alpha_k}, \tag{1.1.2}$$

see for example, Section 6.1 of [Vi03] and Section 2.2 of [Le01]. We note that (1.1.1) is obtained from (1.1.2) by choosing h_k to be the indicator function of A_k, so that $h_k(x) = 1$ if $x \in A_k$ and $h_k(x) = 0$ if $x \notin A_k$ and g to be the indicator of $\alpha_1 A_1 + \ldots + \alpha_p A_p$. The inequality (1.1.2) remains valid if dx is replaced by a log-concave measure.

In this paper we obtain versions of inequality (1.1.1), respectively (1.1.2), for the number of integer points, respectively for the number of weighted integer points, in some special polytopes, known as *flow polytopes.*

(1.2) Contingency tables. Let $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ be positive integer vectors such that

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

An $m \times n$ non-negative integer matrix $D = (d_{ij})$ with the row sums $r_1, \ldots, r_m$ and the column sums $c_1, \ldots, c_n$ is called a *contingency table* with margins R and C. Geometrically, one can think of the set of contingency tables with prescribed margins as of the set of integer points in the *transportation polytope* $P(R, C)$ of $m \times n$ matrices $X = (x_{ij})$ satisfying the equations

$$\sum_{j=1}^{n} x_{ij} = r_i \quad \text{for} \quad i = 1, \ldots, m, \qquad \sum_{i=1}^{m} x_{ij} = c_j \quad \text{for} \quad j = 1, \ldots, n$$

and inequalities

$$x_{ij} \geq 0 \quad \text{for all} \quad i, j.$$

The numbers of contingency tables with prescribed margins are of interest because of their applications in statistics, combinatorics, and representation theory, see [DE85], [DG95], and [DG04].

We consider the number of *weighted* tables, defined as follows.

(1.3) Definition. Let $W = (w_{ij})$ be an $m \times n$ non-negative matrix. For $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$, we define

$$T(R, C; W) = \sum_D \prod_{ij} w_{ij}^{d_{ij}},$$

where the sum is taken over all $m \times n$ contingency tables $D = (d_{ij})$ with the margins (R, C). We agree that $0^0 = 1$.

Geometrically, $T(R, C; W)$ is the generating function over the set of integer points in a transportation polytope. We get the number of points if we choose $W = \mathbf{1}$, the matrix of all 1s.

(1.4) Integer flows. Let $G = (V, E)$ be a directed graph with the set V of vertices, the set E of edges, without multiple edges or loops. Suppose that an integer $a(v)$, called the *excess* v, is assigned to every vertex $v \in V$ so that

$$\sum_{v \in V} a(v) = 0.$$

A collection $x(e) : e \in E$ of non-negative integers is called an *integer feasible flow* in G if the balance condition is satisfied at every vertex

$$\sum_{e:\ \mathrm{head}(e)=v} x(e) - \sum_{e:\ \mathrm{tail}(e)=v} x(e) = a(v) \quad \text{for all} \quad v \in V.$$

If G does not contain directed cycles $v_1 \to v_2 \to \ldots \to v_k \to v_1$ then the set of feasible flows is compact, so the number of integer feasible flows is finite.

Some interesting quantities can be defined as the number of integer feasible flows in an appropriate network. For example, we get the *Kostant partition function* (for the A_{n-1} root system) if $G = K_n$ is a complete graph with the set of vertices $V = \{1, \ldots, n\}$ and edges $E = \{i \to j :\ i > j\}$, cf. [B+04]. Given an integer vector $a = (a_1, \ldots, a_n)$ such that $a_1 + \ldots + a_n = 0$, the number $\phi(a)$ of integer feasible flows in K_n with the excess at i equal a_i is the value of the Kostant partition function at a.

Given a directed graph G on $|V| = n$ vertices and excesses $a(v)$ at its vertices, one can construct an $n \times n$ matrix $W = (w_{ij})$ with $w_{ij} \in \{0, 1\}$, a vector $R = (r_1, \ldots, r_n)$ of row sums and a vector $C = (c_1, \ldots, c_n)$ of column sums so that $T(R, C; W)$ is equal to the number of integer feasible flows in G. To that end, we identify $V = \{1, \ldots, n\}$. Given the excess a_i at the vertex i of G, we find an a priori upper bound $z_i \geq 0$ on the total incoming flow to i and let $r_i = z_i - a_i$ and $c_i = z_i$. Finally, we let $w_{ij} = 1$ if $i = j$ or $i \to j$ is an edge of G and let $w_{ij} = 0$ otherwise.

With a feasible flow $\{x_e :\ e \in E\}$ in G, we associate a contingency table $D = (d_{ij})$ as follows: we let $d_{ij} = x(e)$ provided $i = \mathrm{head}(e)$ and $j = \mathrm{tail}(e)$ and let

$$d_{ii} = r_i - \sum_{e:\ \mathrm{tail}(e)=i} x(e) = c_i - \sum_{e:\ \mathrm{head}(e)=i} x(e).$$

Further, we let $d_{ij} = 0$ if $w_{ij} = 0$. One can observe that this correspondence is a bijection between the integer feasible flows in G and the contingency tables enumerated by $T(R, C; W)$.

For example, for the Kostant partition function, we let $w_{ij} = 1$ if $i \geq j$ and $w_{ij} = 0$ otherwise and define $r_1 = 0$, $r_i = a_1 + \ldots + a_{i-1}$ for $i > 1$ and $c_j = a_1 + \ldots + a_j$ for $j \geq 1$. Noticing that $r_1 = c_n = 0$, we cross out the first row and the nth column and obtain the following description of the Kostant partition function.

Let us define the $(n-1) \times (n-1)$ matrix $W = (w_{ij})$ by

$$w_{ij} = \begin{cases} 1 & \text{if } i \geq j-1 \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$r_k = c_k = \sum_{i=1}^{k} a_k \quad \text{for} \quad k = 1, \ldots, n-1.$$

Then the Kostant partition function ϕ satisfies

$$\phi(a_1, \ldots, a_n) = T(R, C; W).$$

A version of the integer flow enumeration problem involves positive integer *capacities* $c(e)$ of edges and requires feasible flows to satisfy $x(e) \leq c(e)$. Given a directed graph G with capacities one can construct a directed graph G' without capacities so that the integer feasible flows in G' are in a bijection with the integer feasible flows in G. For that, an extra vertex is introduced for every edge of G with capacity, see [B+04].

2. Main results

Our main result is the following inequality relating numbers $T(R, C; W)$ of weighted contingency tables for different margins R and C.

(2.1) Theorem. *For a positive integer vector $B = (b_1, \ldots, b_l)$ we define*

$$|B| = \sum_{i=1}^{l} b_i \quad \textit{and} \quad \omega(B) = \prod_{i=1}^{l} \frac{b_i^{b_i}}{b_i!}.$$

Let $W = (w_{ij})$ be a non-negative $m \times n$ matrix, let $R_1, \ldots, R_p$ be positive integer m-vectors and let $C_1, \ldots, C_p$ be positive integer n-vectors such that

$$|R_1| = \ldots = |R_p| = |C_1| = \ldots = |C_p| = N.$$

Suppose further that $\alpha_1, \ldots, \alpha_p \geq 0$ are numbers such that $\alpha_1 + \ldots + \alpha_p = 1$. Let us define

$$R = \sum_{k=1}^{p} \alpha_k R_k \quad \text{and} \quad C = \sum_{k=1}^{p} \alpha_k C_k$$

and suppose that R and C are positive integer vectors.

Then

$$\frac{N^N}{N!} \frac{T(R,C;W)}{\omega(R)\omega(C)} \geq \prod_{k=1}^{p} \left(\frac{T\left(R_k, C_k; W\right)}{\min\{\omega(R_k),\ \omega(C_k)\}} \right)^{\alpha_k}.$$

Geometrically, for the transportation polytopes $P(R,C)$ and $P\left(R_k, C_k\right)$ we have

$$P(R,C) = \alpha_1 P\left(R_1, C_1\right) + \ldots + \alpha_p P\left(R_p, C_p\right),$$

cf. Section 1.2. On the other hand, the corresponding convex combination of integer points in $P\left(R_k, C_k\right)$ does not have to be an integer point in $P(R,C)$. Hence, the existence of an a priori relation between the numbers of integer points in $P\left(R_k, C_k\right)$ and $P\left(R, C\right)$ is not obvious (for a different approach to discrete Brunn-Minkowski inequalities, see [GG01]).

What follows is a chain of weaker inequalities which are easier to parse.

(2.2) Corollary. *Under the conditions of Theorem 2.1, let*

$$R = (r_1, \ldots, r_m), \quad C = (c_1, \ldots, c_n), \quad a = \min\{m, n\}, \quad \text{and}$$

$$s = N/a \quad \text{where} \quad N = \sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j.$$

Then we have

(1)

$$\frac{N^N}{N!} \min\left\{ \prod_{i=1}^{m} \frac{r_i!}{r_i^{r_i}}, \quad \prod_{j=1}^{n} \frac{c_j!}{c_j^{c_j}} \right\} T(R,C;W) \geq \prod_{k=1}^{p} T^{\alpha_k}\left(R_k, C_k; W\right).$$

(2)

$$\frac{N^N}{N!} \frac{\Gamma^a(s+1)}{s^N} T(R,C;W) \geq \prod_{k=1}^{p} T^{\alpha_k}\left(R_k, C_k; W\right).$$

(3) *There is an absolute constant $\kappa > 0$ such that*

$$(\kappa s)^{\frac{1}{2}(a-1)} T(R,C;W) \geq \prod_{k=1}^{p} T^{\alpha_k}\left(R_k, C_k; W\right).$$

Suppose that $w_{ij} \in \{0,1\}$ for all i,j. Then $T(R,C;W)$ is the number of integer points in the flow polytope $P(R,C;W)$ defined in the space of $m \times n$ matrices (x_{ij}) by the equations

$$\begin{aligned} &\sum_{j=1}^{n} x_{ij} = r_i \quad \text{for} \quad i=1,\dots,m \\ &\sum_{i=1}^{m} x_{ij} = c_j \quad \text{for} \quad j=1,\dots,n, \quad \text{and} \\ &x_{ij} = 0 \quad \text{whenever} \quad w_{ij} = 0 \end{aligned}$$

and inequalities

$$x_{ij} \geq 0 \quad \text{provided} \quad w_{ij} = 1.$$

Generally speaking, the correction term $(\kappa s)^{(a-1)/2}$ is small compared to the value of $T(R,C;W)$. If we scale $R \longmapsto tR$, $C \longmapsto tC$ for a positive integer t, the number of integer points in $P(tR,tC;W)$ grows as a polynomial of t of degree $d = \dim P(R,C;W)$, see, for example, Section 4.6 of [St97], which can be as high as $d=(m-1)(n-1)$ in the case of the transportation polytope (see Section 1.2) with $w_{ij} \equiv 1$. On the other hand, the correction term $(\kappa s)^{(a-1)/2}$ is a polynomial in t of degree only $(\min\{m,n\}-1)/2$.

As another extreme case, let us consider the situation when the numbers r_i, c_j are uniformly bounded, while m and n grow. In this case, $T(R,C;W)$ grows roughly as $(\kappa_1 N)^N$, as long as the number of zeros in each row and column of the 0-1 matrix W is uniformly bounded, cf. [Be74]. The correction term is about κ_2^N for some absolute constants $\kappa_1, \kappa_2 > 0$.

Let us choose an $m \times n$ matrix c_{ij} and let us define matrix $W(t) = (w_{ij}(t))$ by $w_{ij}(t) = \exp\{tc_{ij}\}$. One can observe that

$$\lim_{t \longrightarrow +\infty} t^{-1} \ln T\big(R,C;W(t)\big) = \max\left\{\sum_{ij} c_{ij}x_{ij} : \ (x_{ij}) \in P(R,C) \cap \mathbb{Z}^{m \times n}\right\}.$$

In words: the limit is equal to the maximum value of the linear function defined by matrix (c_{ij}) on the set of integer points in the transportation polytope $P(R,C)$, see Section 1.2. Thus any estimate of the type

$$\alpha(R,C)T(R,C;W) \geq \prod_{k=1}^{p} T^{\alpha_k}\left(R_k, C_k; W\right),$$

where $\alpha(R,C)$ is a factor depending on R and C alone, implies that if $x_k \in P(R_k, C_k)$ are integer points then the point $\alpha_1 x_1 + \ldots + \alpha_p x_p$ lies inside the convex hull of the set of integer points of $P(R,C)$. This, of course, also follows from the fact that the vertices of $P(R,C)$ are integer.

One can ask, naturally, whether the bound in Theorem 2.1 can be strengthened. In particular, the following question is of interest:

- Is it true that under conditions of Theorem 2.1, one has

$$T(R,C;W) \geq \prod_{k=1}^{p} T^{\alpha_k}\left(R_k, C_k; W\right)? \tag{2.3}$$

Or, perhaps, does the above inequality hold in some interesting special cases, for example, when $W = \mathbf{1}$, the $m \times n$ matrix of all 1s, so that $T(R,C;W)$ is the number of contingency tables with the row sums R and column sums C?

There is some circumstantial evidence that the $T(R,C;\mathbf{1})$ might indeed satisfy (2.3). We note that the value of $T(R,C;\mathbf{1})$ does not change if the entries of R and and C are arbitrarily permuted. Let $a = (\alpha_1, \ldots, \alpha_n)$ and $b = (\beta_1, \ldots, \beta_n)$ be integer vectors such that

$$\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \quad \text{and} \quad \beta_1 \geq \beta_2 \geq \ldots \geq \beta_n.$$

We say that a *dominates* b (denoted $a \trianglerighteq b$) if

$$\sum_{i=1}^{k} \alpha_i \geq \sum_{i=1}^{k} \beta_i \quad \text{and} \quad k = 1, \ldots, n-1 \quad \text{and} \quad \sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \beta_i.$$

Equivalently, $a \trianglerighteq b$ if b is a convex combination of vectors obtained from a by permutations of coordinates.

One can show that

$$T\left(R_1, C_1; \mathbf{1}\right) \geq T\left(R_2, C_2; \mathbf{1}\right) \quad \text{provided} \quad R_2 \trianglerighteq R_1 \quad \text{and} \quad C_2 \trianglerighteq C_1. \tag{2.4}$$

The proof consists of two steps. First, assuming that $R = (r_1 \geq r_2 \geq \ldots \geq r_m)$ and $C = (c_1 \geq c_2 \geq \ldots \geq c_n)$ we express $T(R,C;\mathbf{1})$ in terms of *Kostka numbers*,

$$T(R,C;\mathbf{1}) = \sum_{A} K_{AR} K_{AC},$$

where the sum is taken over all $A = (a_1 \geq a_2 \geq \ldots \geq a_s)$, see Section 6.I of [Ma95]. Then we apply the inequality

$$K_{AB_2} \leq K_{AB_1} \quad \text{provided} \quad B_2 \trianglerighteq B_1,$$

see Section 7.I of [Ma95]. Inequality (2.4) is consistent with the hypothesis (2.3).

To prove Theorem 2.1, we represent $T(R,C;W)$ as the expectation of the permanent of a random $N \times N$ matrix A with exponentially distributed entries using a result from [Ba05]. Then using the theory of matrix scaling [MO68], [RS89], [L+00], we represent $\operatorname{per} A$ as the product of a "large and tame" and a "small and wild" factors. The "tame" factor contributes the bulk to the expectation and it satisfies the conditions of the Prékopa-Leindler inequality (1.1.2), the fact that ultimately results in the inequality of Theorem 2.1. The "wild" factor is harder to analyze, but it does not vary much since it lies within the low bound provided by the van der Waerden estimate [Eg81], [Fa81] and the upper bound provided by the Bregman-Minc estimate [Br73]. It contributes to the correction term in Theorem 2.1 and Corollary 2.2.

We discuss preliminaries in Section 3 and present the proofs of Theorem 2.1 and Corollary 2.2 in Section 4.

3. A permanental representation of $T(R, C; W)$

Recall that the *permanent* of an $N \times N$ matrix $A = (a_{ij})$ is defined by

$$\operatorname{per} A = \sum_{\sigma \in S_N} \prod_{i=1}^{N} a_{i\sigma(i)},$$

where S_N is the symmetric group of all permutations of $\{1, \ldots, N\}$. We say that a random variable γ has the *standard exponential distribution* if

$$\mathbf{P}(\gamma > t) = \begin{cases} e^{-t} & \text{if } t > 0 \\ 1 & \text{otherwise.} \end{cases}$$

The following result expressing $T(R, C; W)$ as the expectation of the permanent of a random matrix was proved in [Ba05].

(3.1) Theorem. *Given a positive integer m-vector $R = (r_1, \ldots, r_m)$ and a positive integer n-vector $C = (c_1, \ldots, c_n)$ such that*

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

and an $m \times n$ matrix $W = (w_{ij})$, we construct an $N \times N$ random matrix A as follows: the set of rows of A is represented as a disjoint union of m subsets of cardinalities $r_1, \ldots, r_m$ whereas the set of columns of A is represented as a disjoint union of n subsets of cardinalities $c_1, \ldots, c_n$, so that A is represented as a block matrix of mn blocks $r_i \times c_j$. Let let $G = (g_{ij})$ be the $m \times n$ matrix with $g_{ij} = w_{ij}\gamma_{ij}$, where γ_{ij} are independent standard exponential random variables. We fill the (i, j)th block $r_i \times c_j$ of $A = A(G)$ by the copies of g_{ij}. Then

$$T(R, C; W) = \frac{\mathbf{E} \operatorname{per} A}{r_1! \cdots r_m! c_1! \cdots c_n!}.$$

For the sake of completeness, we present a proof of Theorem 3.1 in Section 5.

Next, we need some results on matrix scaling, in particular as described in [MO68] and [RS89].

(3.2) Matrix scaling. Let $G=(g_{ij})$ be a positive $m \times n$ matrix and let $r_1, \ldots, r_m$ and $c_1, \ldots, c_n$ be positive numbers such that

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

Then there exist a unique positive $m \times n$ matrix $L=(l_{ij})$ and positive numbers $\mu_1, \ldots, \mu_m$ and $\lambda_1, \ldots, \lambda_n$ such that

$$\begin{aligned} \sum_{j=1}^n l_{ij} &= r_i \quad \text{for} \quad i=1, \ldots, m, \\ \sum_{i=1}^m l_{ij} &= c_j \quad \text{for} \quad j=1, \ldots, n \end{aligned}$$

and such that

$$g_{ij} = l_{ij}\mu_i\lambda_j \quad \text{for all} \quad i,j.$$

Moreover, the numbers λ_i, μ_j are unique up to a scaling

$$\mu_i \longmapsto \mu_i\tau, \lambda_j \longmapsto \lambda_j\tau^{-1} \quad \text{for some} \quad \tau>0 \quad \text{and all} \quad i,j$$

and can be obtained as follows.

Let

$$F(G;x,y) = \sum_{i=1}^m \sum_{j=1}^n g_{ij}\xi_i\eta_j \quad \text{for}$$
$$x=(\xi_1, \ldots, \xi_m) \quad \text{and} \quad y=(\eta_1, \ldots, \eta_n).$$

Then $F(G;x,y)$ attains a unique minimum on the set of pairs (x,y) of vectors defined by the equations

$$\prod_{i=1}^m \xi_i^{r_i} = 1 \quad \text{and} \quad \prod_{j=1}^n \eta_j^{c_j} = 1$$

and inequalities

$$\xi_i > 0 \quad \text{for} \quad i=1, \ldots, m \quad \text{and} \quad \eta_j > 0 \quad \text{for} \quad j=1, \ldots, n.$$

Assuming that $x^*=(\xi_1^*, \ldots, \xi_m^*)$ and $y^*=(\eta_1^*, \ldots, \eta_n^*)$ is the minimum point, we may let

$$\mu_i = \frac{F(G;x^*,y^*)}{N\xi_i^*} \quad \text{and} \quad \lambda_j = \frac{1}{\eta_j^*} \quad \text{for all} \quad i,j,$$

see [RS89] and [MO68].

Finally, we need some estimates for permanents.

(3.3) Estimates for permanents. Recall that an $N \times N$ matrix $B = (b_{ij})$ is called *doubly stochastic* if it is non-negative

$$b_{ij} \geq 0 \quad \text{for} \quad i, j = 1, \ldots, N$$

and all row and column sums are equal to 1:

$$\sum_{j=1}^N b_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, N \quad \text{and}$$
$$\sum_{i=1}^N b_{ij} = 1 \quad \text{for} \quad j = 1, \ldots, N.$$

The van der Waerden conjecture proved by G.P. Egorychev [Eg81] and D.I. Falikman [Fa81] asserts that

$$\operatorname{per} B \geq \frac{N!}{N^N} \tag{3.3.1}$$

if B is a doubly stochastic $N \times N$ matrix, see also Chapter 12 of [LW01].

The following upper bound was conjectured by H. Minc and proved by L.M. Bregman [Br73], see also Chapter 11 of [LW01].

Let $B = (b_{ij})$ be an $N \times N$ matrix such that $b_{ij} \in \{0, 1\}$ for all i, j and let

$$\sum_{j=1}^N b_{ij} = s_i \quad \text{for} \quad i = 1, \ldots, N.$$

Then

$$\operatorname{per} B \leq \prod_{i=1}^N (s_i!)^{1/s_i}.$$

We will need the following corollary of the Bregman-Minc inequality, see [So03].

Let $B = (b_{ij})$ be an $N \times N$ matrix such that

$$\sum_{j=1}^N b_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, N \quad \text{and}$$
$$0 \leq b_{ij} \leq \frac{1}{s_i} \quad \text{for} \quad j = 1, \ldots, N$$

and positive integers $s_1, \ldots, s_N$. Then

$$\operatorname{per} B \leq \prod_{i=1}^N \frac{(s_i!)^{1/s_i}}{s_i}. \tag{3.3.2}$$

Of course, similar estimates hold if we interchange rows and columns.

4. Proofs

In this section, we prove Theorem 2.1 and Corollary 2.2.

(4.1) Notation. Given an $m \times n$ positive matrix $G = (g_{ij})$, let us define

$$F(G; x, y) = \sum_{i=1}^{m} \sum_{j=1}^{n} g_{ij} \xi_i \eta_j \quad \text{for}$$
$$x = (\xi_1, \ldots, \xi_m) \quad \text{and} \quad y = (\eta_1, \ldots \eta_n).$$

For positive vectors $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$, we define

$$\begin{aligned} f(G; R, C) = &\min F(G; x, y) \\ &\text{for} \quad x = (\xi_1, \ldots, \xi_m) \quad \text{and} \quad y = (\eta_1, \ldots, \eta_n) \\ &\text{subject to} \quad \prod_{i=1}^{m} \xi_i^{r_i} = \prod_{j=1}^{n} \eta_j^{c_j} = 1 \quad \text{and} \\ &\xi_i, \eta_j > 0 \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n, \end{aligned}$$

see Section 3.2. We recall notation

$$|R| = \sum_{i=1}^{m} r_i \quad \text{and} \quad |C| = \sum_{j=1}^{n} c_j.$$

First, we establish a certain convexity property of $f(G; R, C)$.

(4.2) Lemma. *Let $G_1, \ldots, G_p$ be positive $m \times n$ matrices, let $R_1, \ldots, R_p$ be positive m-vectors, and let $C_1, \ldots, C_p$ be positive n-vectors such that*

$$|R_1| = \ldots = |R_p| = |C_1| = \ldots = |C_p|.$$

Suppose further that $\alpha_1, \ldots, \alpha_p \geq 0$ are numbers such that $\alpha_1 + \ldots + \alpha_p = 1$. Let us define

$$G = \sum_{k=1}^{p} \alpha_k G_k, \quad R = \sum_{k=1}^{p} \alpha_k R_k, \quad \text{and} \quad C = \sum_{k=1}^{p} \alpha_k C_k.$$

Then

$$f(G; R, C) \geq \prod_{k=1}^{p} f^{\alpha_k} (G_k; R_k, C_k).$$

Proof. Suppose that

$$R_k = (r_{1k}, \ldots, r_{mk}), \quad C_k = (c_{1k}, \ldots, c_{nk}), \quad R = (r_1, \ldots, r_m), \quad \text{and}$$
$$C = (c_1, \ldots, c_n).$$

In particular,

$$
(4.2.1) \qquad
\begin{aligned}
r_i &= \sum_{k=1}^{p} \alpha_k r_{ik} \quad \text{for} \quad i=1,\ldots,m \quad \text{and} \\
c_j &= \sum_{k=1}^{p} \alpha_k c_{jk} \quad \text{for} \quad j=1,\ldots,n.
\end{aligned}
$$

Let $x=(\xi_1,\ldots,\xi_m)$ and $y=(\eta_1,\ldots,\eta_n)$ be positive vectors such that

$$
(4.2.2) \qquad \prod_{i=1}^{m} \xi_i^{r_i} = \prod_{j=1}^{n} \eta_j^{c_j} = 1.
$$

Then

$$
F(G;x,y) = \sum_{k=1}^{p} \alpha_k F(G_k;x,y) \geq \prod_{k=1}^{p} F^{\alpha_k}(G_k;x,y).
$$

Let

$$
t_k = \left(\prod_{i=1}^{m} \xi_i^{r_{ik}}\right)^{1/|R|} \quad \text{and} \quad s_k = \left(\prod_{j=1}^{n} \eta_j^{c_{jk}}\right)^{1/|C|} \quad \text{for} \quad k=1,\ldots,p.
$$

Then

$$
F(G_k;x,y) = t_k s_k F\left(G_k; t_k^{-1}x, s_k^{-1}y\right) \geq t_k s_k f\left(G_k;R_k,C_k\right),
$$

since vectors $t_k^{-1}x$ and $s_k^{-1}y$ satisfy (4.2.2) with r_i and c_j replaced by r_{ik} and c_{jk} respectively. Therefore,

$$
F(G;x,y) \geq \prod_{k=1}^{p} t_k^{\alpha_k} s_k^{\alpha_k} f^{\alpha_k}\left(G_k;R_k,C_k\right).
$$

On the other hand, by (4.2.1) and (4.2.2), we have

$$
\prod_{k=1}^{p} t_k^{\alpha_k} = \left(\prod_{i=1}^{m} \xi_i^{\sum_{k=1}^{p} \alpha_k r_{ik}}\right)^{1/|R|} = \left(\prod_{i=1}^{m} \xi_i^{r_i}\right)^{1/|R|} = 1,
$$

and, similarly,

$$
\prod_{k=1}^{p} s_k^{\alpha_k} = \left(\prod_{j=1}^{n} \eta_j^{\sum_{k=1}^{p} \alpha_k c_{jk}}\right)^{1/|C|} = \left(\prod_{j=1}^{n} \eta_j^{c_j}\right)^{1/|C|} = 1.
$$

Since the inequality

$$
F(G;x,y) \geq \prod_{k=1}^{p} f^{\alpha_k}\left(G_k;R_k,C_k\right)
$$

holds for any positive x and y satisfying (4.2.2), the proof follows. □

Next, we consider block matrices A as in Theorem 3.1.

(4.3) Lemma. *Let $G=(g_{ij})$ be an $m \times n$ positive matrix. Let $R=(r_1, \ldots, r_m)$ and $C=(c_1, \ldots, c_n)$ be positive integer vectors such that $|R|=|C|=N$. Let us consider the $N \times N$ block matrix A, where the (i,j)th block of size $r_i \times c_j$ is filled by copies of g_{ij}. Then there exists an $N \times N$ block matrix B with the same block structure as A and such that*

(1) *Matrix B is doubly stochastic;*
(2) *The entries in the (i,j)th block of B do not exceed $\min\{1/r_i,\ 1/c_j\}$;*
(3) *We have*

$$\operatorname{per} A = N^{-N} f^N(G; R, C) \left(\prod_{i=1}^m r_i^{r_i}\right) \left(\prod_{j=1}^n c_j^{c_j}\right) \operatorname{per} B;$$

(4)

$$\frac{N!}{N^N} \leq \operatorname{per} B \leq \min\left\{\prod_{i=1}^m \frac{r_i!}{r_i^{r_i}}, \quad \prod_{j=1}^n \frac{c_j!}{c_j^{c_j}}\right\}.$$

Proof. Let $L=(l_{ij})$ be the $m \times n$ positive matrix and let μ_i, $i=1, \ldots, m$, and λ_j, $j=1, \ldots, n$, be positive numbers such that

$$g_{ij} = l_{ij}\mu_i\lambda_j \quad \text{for} \quad i=1, \ldots, m \quad \text{and} \quad j=1, \ldots, n$$

and such that

$$\sum_{j=1}^n l_{ij} = r_i \quad \text{for} \quad i=1, \ldots, m \quad \text{and}$$
$$\sum_{i=1}^m l_{ij} = c_j \quad \text{for} \quad j=1, \ldots, n,$$

see Section 3.2.

Let us divide the entries in the (i,j)th block of A by the product $\mu_i r_i \lambda_j c_j$. We get the matrix B with the entries in the (i,j)th block equal to $l_{ij}/r_i c_j$. It is seen now that B is doubly stochastic and that the entries in the (i,j)th block of B do not exceed $\min\{1/r_i,\ 1/c_j\}$. Furthermore,

$$\operatorname{per} A = \left(\prod_{i=1}^m (\mu_i r_i)^{r_i}\right) \left(\prod_{j=1}^n (\lambda_j c_j)^{c_j}\right) \operatorname{per} B.$$

On the other hand, if one computes μ_i and λ_j by optimizing $F(G; x, y)$ as in Section 3.2, one gets

$$\prod_{i=1}^m \mu_i^{r_i} = \frac{f^N(G; R, C)}{N^N} \quad \text{and} \quad \prod_{j=1}^n \lambda_j^{c_j} = 1,$$

which completes the proof of Part (3).

Part (4) follows by Parts (1) and (2) and estimates (3.3.1) and (3.3.2). $\square$

Now we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. Without loss of generality, we assume that $w_{ij} > 0$ for all i, j.

In the space $\mathrm{Mat}(m, n)$ of $m \times n$ real matrices $G = (g_{ij})$, we consider the exponential measure dG with the density

$$\prod_{ij} w_{ij}^{-1} \exp\{-g_{ij}/w_{ij}\} \quad \text{if} \quad g_{ij} > 0 \quad \text{for all} \quad i, j$$

and 0 elsewhere.

We note that dG is a log-concave measure.

Given positive integer vectors $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$ and a positive $m \times n$ matrix G, let $A(G; R, C)$ be the $N \times N$ block matrix constructed as in Theorem 3.1. Then, by Theorem 3.1,

$$T(R, C; W) = \left(\prod_{i=1}^{m} \frac{1}{r_i!}\right)\left(\prod_{j=1}^{n} \frac{1}{c_j!}\right) \int_{\mathrm{Mat}(m,n)} \operatorname{per} A(G; R, C)\ dG.$$

From Lemma 4.3,

$$\begin{aligned} T(R, C; W) \geq & \frac{N!}{N^N} N^{-N} \left(\prod_{i=1}^{m} \frac{r_i^{r_i}}{r_i!}\right)\left(\prod_{j=1}^{n} \frac{c_j^{c_j}}{c_j!}\right) \\ & \times \int_{\mathrm{Mat}(m,n)} f^N(G; R, C)\ dG \\ = & \frac{N!}{N^N} N^{-N} \omega(R)\omega(C) \int_{\mathrm{Mat}(m,n)} f^N(G; R, C)\ dG. \end{aligned}$$

Similarly, letting $R_k = (r_{1k}, \ldots, r_{mk})$ and $C_k = (c_{1k}, \ldots, c_{nk})$, by Theorem 3.1 we obtain

$$T(R_k, C_k; W_k) = \left(\prod_{i=1}^{m} \frac{1}{r_{ik}!}\right)\left(\prod_{j=1}^{n} \frac{1}{c_{jk}!}\right) \int_{\mathrm{Mat}(m,n)} \operatorname{per} A(G; R_k, C_k)\ dG,$$

and from Lemma 4.3

$$\begin{aligned} T\left(R_k, C_k; W\right) \leq & N^{-N} \left(\prod_{i=1}^{m} \frac{r_{ik}^{r_{ik}}}{r_{ik}!}\right)\left(\prod_{j=1}^{n} \frac{c_{jk}^{c_{jk}}}{c_{jk}!}\right) \\ & \times \min\left\{\prod_{i=1}^{m} \frac{r_{ik}!}{r_{ik}^{r_{ik}}}, \quad \prod_{j=1}^{n} \frac{c_{jk}!}{c_{jk}^{c_{jk}}}\right\} \int_{\mathrm{Mat}(m,n)} f^N(G; R_k, C_k)\ dG \\ = & N^{-N} \min\bigl\{\omega(R_k),\ \omega(C_k)\bigr\} \int_{\mathrm{Mat}(m,n)} f^N(G; R_k, C_k)\ dG. \end{aligned}$$

By Lemma 4.2, for any positive matrices $G_1, \ldots, G_k$ we have

$$f(G; R, C) \geq \prod_{k=1}^{p} f^{\alpha_k}\left(G_k; R_k, C_k\right), \quad \text{where} \quad G = \sum_{k=1}^{p} \alpha_k G_k.$$

Applying the Prékopa-Leindler inequality (1.1.2), we obtain

$$\int_{\mathrm{Mat}(m,n)} f^N(G; R, C)\, dG \geq \prod_{k=1}^{p} \left(\int_{\mathrm{Mat}(m,n)} f^N\left(G; R_k, C_k\right) dG \right)^{\alpha_k}.$$

Therefore,

$$\frac{N^N}{N!} \frac{T(R, C; W)}{\omega(R)\omega(C)} \geq \prod_{k=1}^{p} \left(\frac{T\left(R_k, C_k; W\right)}{\min\{\omega(R_k),\ \omega(C_k)\}} \right)^{\alpha_k}$$

and the proof follows. □

Proof of Corollary 2.2. We use that the function

$$s \longmapsto \frac{b^b}{\Gamma(b+1)}, \quad b > 0$$

is log-convex. Therefore, the function

$$\omega\left(b_1, \ldots, b_l\right) = \prod_{i=1}^{l} \frac{b_i^{b_i}}{\Gamma(b_i + 1)}$$

is log-convex on the positive orthant $b_1 > 0, \ldots, b_l > 0$.

Thus we have

$$\omega(R) \geq \prod_{k=1}^{p} \omega^{\alpha_k}\left(R_k\right) \quad \text{and} \quad \omega(C) \geq \prod_{k=1}^{p} \omega^{\alpha_k}\left(C_k\right).$$

Hence

$$\frac{N^N}{N!} \frac{T(R, C; W)}{\omega(C)} \geq \prod_{k=1}^{p} T^{\alpha_k}\left(R_k, C_k; W\right) \quad \text{and}$$

$$\frac{N^N}{N!} \frac{T(R, C; W)}{\omega(R)} \geq \prod_{k=1}^{p} T^{\alpha_k}\left(R_k, C_k; W\right),$$

from which Part (1) follows.

Similarly, since ω is log-convex,

$$\omega(R) \geq \omega\left(|R|/m, \ldots, |R|/m\right) \quad \text{and} \quad \omega(C) \geq \omega\left(|C|/n, \ldots, |C|/n\right),$$

from which Part (2) follows.

Finally, by Stirling's formula

$$(2\pi s)^{1/2} s^s e^{-s} e^{\frac{1}{12s+1}} < \Gamma(s+1) < (2\pi s)^{1/2} s^s e^{-s} e^{\frac{1}{12s}}$$

and Part(3) follows. □

5. Appendix: Proof of Theorem 3.1

Let

$$\{1, \ldots, N\} = \bigcup_{i=1}^{m} \mathrm{Row}_i \quad \text{and} \quad \{1, \ldots, N\} = \bigcup_{j=1}^{n} \mathrm{Col}_j$$

be the partition of the set of rows, respectively the set of columns of A, into non-overlapping blocks such that

$$\left| \mathrm{Row}_i \right| = r_i, \quad \left| \mathrm{Col}_j \right| = c_j \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n$$

and the $\mathrm{Row}_i \times \mathrm{Col}_j$ block of A is filled by the copies of $g_{ij} = w_{ij}\gamma_{ij}$.

Let $\phi : \{1, \ldots, N\} \longrightarrow \{1, \ldots, N\}$ be a bijection. We denote by

$$t_\phi = \prod_{k=1}^{N} a_{k\phi(k)}$$

the corresponding term of $\operatorname{per} A$ and introduce the $m \times n$ contingency table $D = D(\phi)$, called the *pattern* of ϕ as follows: we have $D = (d_{ij})$ where d_{ij} is the number of indices $k \in \mathrm{Row}_i$ such that $\phi(k) \in \mathrm{Col}_j$. Clearly, the row sums of D are $r_1, \ldots, r_m$ while the column sums of D are $c_1, \ldots, c_n$.

Thus

$$\mathbf{E} \operatorname{per} A = \sum_{\phi} \mathbf{E}\, t_\phi \quad \text{and} \quad \mathbf{E}\, t_\phi = \prod_{ij} w_{ij}^{d_{ij}} d_{ij}!,$$

where $D = (d_{ij})$ is the pattern of ϕ and the sum is taken over all bijections $\phi : \{1, \ldots, N\} \longrightarrow \{1, \ldots, N\}$. Here we use that $\mathbf{E}\,\gamma^d = d!$ for the standard exponential random variable γ.

It remains to count the bijections ϕ having a given pattern $D = (d_{ij})$. Let $D = (d_{ij})$ be a contingency table with the row sums $r_1, \ldots, r_m$ and the column sums $c_1, \ldots, c_n$. To choose a bijection ϕ with the pattern D, for each $i = 1, \ldots, m$ we choose a partition

$$\mathrm{Row}_i = \bigcup_{j=1}^{n} \mathrm{Row}_{ij} \quad \text{where} \quad \left| \mathrm{Row}_{ij} \right| = d_{ij}$$

and for each $j = 1, \ldots, n$ we choose a partition

$$\mathrm{Col}_j = \bigcup_{i=1}^{m} \mathrm{Col}_{ij} \quad \text{where} \quad \left| \mathrm{Col}_{ij} \right| = d_{ij},$$

altogether in

$$\left(\prod_{i=1}^{m} \frac{r_i!}{d_{i1}! \cdots d_{in}!} \right) \left(\prod_{j=1}^{n} \frac{c_j!}{d_{1j}! \cdots d_{mj}!} \right)$$

ways. Finally, we identify ϕ by choosing bijections $\phi_{ij} : \mathrm{Row}_{ij} \longrightarrow \mathrm{Col}_{ij}$ altogether in

$$\prod_{ij} d_{ij}!$$

ways. Therefore, there are

$$\left(\prod_{i=1}^{m} r_i!\right)\left(\prod_{j=1}^{n} c_j!\right)\left(\prod_{ij} d_{ij}!\right)^{-1}$$

bijections $\phi : \{1, \ldots, N\} \longrightarrow \{1, \ldots, N\}$ with the pattern $D = (d_{ij})$. Summarizing,

$$\mathbf{E} \operatorname{per} A = \left(\prod_{i=1}^{m} r_i!\right)\left(\prod_{j=1}^{n} c_j!\right) \sum_{D} \prod_{ij} w_{ij}^{d_{ij}},$$

where the sum is taken over all contingency tables $D = (d_{ij})$ with the row sums $r_1, \ldots, r_m$ and the column sums $c_1, \ldots, c_n$, which completes the proof.

Acknowledgment

I am grateful to Alexander Yong for many useful conversations.

References

[Ba05] A. Barvinok, *Enumerating contingency tables via random permanents*, preprint arXiv math.CO/0511596 (2005).

[Be74] E.A. Bender, *The asymptotic number of non-negative integer matrices with given row and column sums*, Discrete Math. **10** (1974), 217–223.

[Br73] L.M. Bregman, *Certain properties of nonnegative matrices and their permanents*, Dokl. Akad. Nauk SSSR **211** (1973), 27–30.

[B+04] W. Baldoni-Silva, J.A. De Loera, and M. Vergne, *Counting integer flows in networks*, Found. Comput. Math. **4** (2004), 277–314.

[DE85] P. Diaconis and B. Efron, *Testing for independence in a two-way table: new interpretations of the chi-square statistic. With discussions and with a reply by the authors*, Ann. Statist. **13** (1985), 845–913.

[DG95] P. Diaconis and A. Gangolli, *Rectangular arrays with fixed margins*, Discrete probability and algorithms (Minneapolis, MN, 1993), IMA Vol. Math. Appl., vol. 72, Springer, New York, 1995, pp. 15–41.

[DG04] P. Diaconis and A. Gamburd, *Random matrices, magic squares and matching polynomials*, Research Paper 2, Electron. J. Combin. **11** (2004), 26 pp.

[Eg81] G.P. Egorychev, *The solution of van der Waerden's problem for permanents*, Adv. in Math. **42** (1981), 299–305.

[Fa81] D.I. Falikman, *Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix (Russian)*, Mat. Zametki **29** (1981), 931–938.

[Ga02] R.J. Gardner, *The Brunn-Minkowski inequality*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), 355–405.

[GG01] R.J. Gardner and P. Gronchi, *A Brunn-Minkowski inequality for the integer lattice*, Trans. Amer. Math. Soc. **353** (2001), 3995–4024.

[Le01] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.

[LW01] J.H. van Lint and R.M. Wilson, *A Course in Combinatorics. Second edition*, Cambridge University Press, Cambridge, 2001.

[L+00] N. Linial, A. Samorodnitsky, and A. Wigderson, *A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents*, Combinatorica **20** (2000), 545–568.

[Ma95] I.G. Macdonald, *Symmetric Functions and Hall Polynomials. Second edition. With contributions by A. Zelevinsky*, Oxford Mathematical Monographs. Oxford Science Publications, The Clarendon Press, Oxford University Press, New York, 1995.

[MO68] A. Marshall and I. Olkin, *Scaling of matrices to achieve specified row and column sums*, Numer. Math. **12** (1968), 83–90.

[RS89] U. Rothblum and H. Schneider, *Scalings of matrices which have prespecified row sums and column sums via optimization*, Linear Algebra Appl. **114/115** (1989), 737–764.

[St97] R.P. Stanley, *Enumerative Combinatorics. Vol. 1*, Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, Cambridge, 1997.

[So03] G.W. Soules, *New permanental upper bounds for nonnegative matrices*, Linear Multilinear Algebra **51** (2003), 319–337.

[Vi03] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043

E-mail address: `barvinok@umich.edu`