An upper bound for permanents of nonnegative matrices

Alex Samorodnitsky*

Abstract

A recent conjecture of Caputo, Carlen, Lieb, and Loss, and, independently, of the author, states that the maximum of the permanent of a matrix whose rows are unit vectors in l_p is attained either for the identity matrix I or for a constant multiple of the all-1 matrix J.

The conjecture is known to be true for $p = 1$ (I) and for $p \geq 2$ (J).

We prove the conjecture for a subinterval of $(1, 2)$, and show the conjectured upper bound to be true within a subexponential factor (in the dimension) for all $1 < p < 2$. In fact, for p bounded away from 1, the conjectured upper bound is true within a constant factor.

This leads to a mild (subexponential) improvement in deterministic approximation factor for the permanent. We present an efficient deterministic algorithm that approximates the permanent of a nonnegative $n \times n$ matrix within $\exp\{n - O(n/\log n)\}$.

1 Introduction

Let $A = (a_{ij})$ be an $n \times n$ matrix. The *permanent* of A is defined as

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

Here S_n is the symmetric group on n elements.

This paper investigates upper bounds on the permanent of matrices with nonnegative entries. Bregman [3] resolved the Minc conjecture and proved a tight upper bound on the permanent of a zero-one matrix with given row sums. Here we are interested in upper bounds for matrices with general nonnegative entries. (For related work see also [17] and the references there.)

More specifically, given $1 \leq p \leq \infty$, we investigate the maximal possible value $U(n, p)$ of the permanent of a matrix whose rows are unit vectors in l_p^n. We give an upper bound on $U(n, p)$ which is tight up to a subexponential (in n) multiplicative factor. Since the permanent is a

*School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel.

multiliner function of its rows, this leads to an upper bound on the permanent of an arbitrary real matrix, given the l_p length of its rows.

Let us start with a conjecture claiming that there are only two possible matrices on which the maximum of the permanent can be attained. This conjecture is due to Caputo, Carlen, Lieb, and Loss [4], and, independently, to the author.

Conjecture 1.1: Let $1 \le p < \infty$. The maximum of the permanent of an $n \times n$ matrix whose rows are unit vectors in l_p is attained in one of two cases.

1. On the identity matrix. In this case the permanent is 1.
2. On a matrix all of whose entries are $n^{-1/p}$. In this case the permanent is $\frac{n!}{n^{n/p}}$.

In particular, the maximal possible value of the permanent is

$$U(n,p) = \max\left\{1, \frac{n!}{n^{n/p}}\right\} \tag{1}$$

$\blacksquare$

Here are some preliminary remarks. Let the dimension n be fixed. The function $f(p) = \frac{n!}{n^{n/p}}$ is increasing. Clearly $f(1) \le 1$ and $f(2) \ge 1$. It is easy to compute the unique value of p, lying in $[1,2]$ for which $f(p) = 1$, that is

$$p_c = \frac{n \log n}{\log n!} \tag{2}$$

Let I denote the identity matrix, and J the all-1 matrix. The conjecture claims that I is optimal for $p \in [1, p_c]$ and $n^{-1/p} \cdot J$ is optimal for $p \in [p_c, \infty]$.

In fact, it would suffice to prove the conjecture only for $p = p_c$.

Lemma 1.2:

- *Let $p_0 > 1$ be such that the matrix I is optimal for p_0. Then I is the only optimal matrix for all $1 \le p < p_0$.*
- *Let p_0 be such that the matrix $n^{1/p} \cdot J$ is optimal for p_0. Then $n^{-1/p} \cdot J$ is the only optimal matrix for all $p > p_0$.*

Let us now present the known results.

1. The case $p = 1$ is trivial. For any n only the identity matrix is optimal, and $U(n,1) = 1$.

2. The conjecture is also known to be true for $p \geq 2$. In this case the optimal matrix is $n^{-1/p} \cdot J$, and $U(n,p) = \frac{n!}{n^{n/p}}$. Different proofs of this fact were given in [12, 15, 10]. Later it was pointed out [9] that this case was, essentially, already dealt with in [16]. More specifically, the proof of [10] is a special case of an argument in [16] (Proposition 9.1.1, Appendix 1).

   To the best of our knowledge, the first published proof specifically treating this case appeared recently in [4]. Furthermore, this paper (independently) states Conjecture 1.1, attributing it also to P. Caputo.

   Let us also mention that results in [7] imply Conjecture 1.1 for $p \geq n$.

3. The case $1 < p < 2$. This case seems to be the most interesting.

   Clearly, one direction in (1) is trivially true: $U(n,p) \geq \max\left\{1, \frac{n!}{n^{n/p}}\right\}$.

   In the other direction, $U(n,p) \leq U(n,1/2) = \frac{n!}{n^{n/2}}$.

   This upper bound on $U(n,p)$ was improved in [4]. They show the function $U(n,p)$ to be logarithmically convex in $1/p$. This, together with the known values $U(n,1) = 1$ and $U(n,2) = \frac{n!}{n^{n/2}}$, lead to an upper bound

$$U(n,p) \leq \left(\frac{n!}{n^{n/2}}\right)^{2-2/p}$$

In this paper we show the conjecture to hold in the interval $[1, p_0]$ where

$$p_0 = \frac{n \log n - (n-1)\log(n-1)}{\log n}$$

For $n \geq 2$ holds $1 < p_0 < p_c \leq 2$.

It is interesting to compare p_0 with p_c. We have $p_c \leq \frac{\log n}{\log(n)-1} = 1 + \frac{1}{\log(n)-1}$. And $p_0 = \frac{\log n + (n-1)\log\frac{n}{n-1}}{\log n} \geq \frac{\log n + (n-1)/n}{\log n} = 1 + \frac{1}{\log n} - \frac{1}{n \log n}$. Thus p_c and p_0 are only about $\frac{1}{\log^2 n}$ apart.

The proximity of p_0 and p_c, together with log-convexity of $U(n,p)$, already suffice for giving an upper bound on $U(n,p)$ for all $p \in (1,2)$ which is tight up to a simply exponential factor (in n). The approach we take will lead to a somewhat tighter estimate, which has a subexponential error in the worst case.

Our main results are given in the following theorem.

Theorem 1.3: *Let n be fixed, and let $p_0 = \frac{n\log n-(n-1)\log(n-1)}{\log n}$.*

1. *The conjecture is true for $1 \leq p \leq p_0$. The identity matrix is optimal for for $1 \leq p \leq p_0$, and*

$$U(n,p) = 1$$

2. For $p_0 \le p \le 2$ holds

$$\max\left\{1, \frac{n!}{n^{n/p}}\right\} \le U(n,p) \le \exp\left\{(p-1)/p \cdot e^{1/(p-1)}\right\} \cdot \frac{n!}{n^{n/p}}$$

Observe, that this bound is $(n/\log n)$-tight in the worst case. For p bounded away from 1, this bound is tight within a constant factor.

1.1 Approximating the permanent

The original motivation for this study was computational. The goal is to construct an efficient deterministic algorithm that approximates the permanent of a given nonnegative matrix within a reasonable multiplicative factor. (A *randomized* algorithm to approximate the permanent with arbitrary precision was constructed in [11].)

In [13] this problem was reduced to the case in which the input matrix is doubly stochastic. This immeadiately gave an $\frac{n^n}{n!}$-approximation, since the permanent of a doubly stochastic matrix lies between $\frac{n!}{n^n}$ and 1. Here the upper bound is trivial, while the lower bound is a deep theorem of Egorychev [5] and Falikman [6], proving a conjecture by van der Waerden. In this light, it seems natural to look fore more informative upper bounds, which could lead to better approximation factors for the doubly-stochastic, and thus, for the general case.

Our results lead to an improvement of $\exp\{O(n/\log n)\}$ in the approximation factor. We note that a polynomial (in n) improvement in the approximation factor was recently obtained in [8].

The main tool is a permanental inequality which might be of independent interest. This inequality is an immediate consequence of Theorem 1.3.

Proposition 1.4: *Let $n \ge 2$ be an integer. Let $p_0 = \frac{n\log n-(n-1)\log(n-1)}{\log n}$. Then for any stochastic $n \times n$ matrix $A = (a_{ij})$ holds*

$$Per\left(\left(a_{ij}^{1/p_0}\right)\right) \le 1$$

Corollary 1.5: *There is a deterministic polynomial-time algorithm to approximate the permanent of a given nonnegative $n \times n$ matrix within a multiplicative factor of $\frac{n^n}{n!} \cdot e^{-\Omega\left(\frac{n}{logn}\right)}$.*

Proof: (Of the corollary) It is sufficient to present an algorithm approximating the permanent of a given doubly stochastic matrix within this factor.

Let $q_0 = 1/p_0$. Assume $n \ge 5$. Let A be a doubly stochastic matrix. Let $\sigma \in S_n$ be a permutation such that $\prod_{i=1}^n a_{i\sigma(i)}$ is maximal. [1] Then there are two cases.

[1]Finding σ amounts to finding a maximal weight perfect matching in a given bipartite graph with $2n$ vertices, and can be done efficiently.

- $\prod_{i=1}^{n} a_{i\sigma(i)} \geq 2^{-n}$. Then

$$2^{-n} \leq \prod_{i=1}^{n} a_{i\sigma(i)} \leq Per(A) \leq 1$$

- $\prod_{i=1}^{n} a_{i\sigma(i)} < 2^{-n}$. In this case, by the proposition,

$$\frac{n!}{n^n} \leq Per(A) \leq 2^{(q_0-1)n} \cdot Per\left(\left(a_{ij}^{q_0}\right)\right) \leq 2^{(q_0-1)n} \leq e^{-\Omega\left(\frac{n}{logn}\right)}$$

∎

1.2 A generalization of Minc's conjecture to general nonnegative matrices

The Minc conjecture, proved by Bregman, states that for a zero-one matrix A with r_i ones in row i, $1 \leq i \leq n$,

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

and equality holds if and only if A is a block-diagonal matrix, and all the blocks are all-1 matrices.[2]

Let $\phi :\ [0,1] \rightarrow [0,1]$ be a function taking $1/r$ to $1/\left(r!\right)^{1/r}$, for all integer r. Given a matrix A with entries in $[0,1]$, let $\phi(A)$ denote a matrix whose (ij)-th entry is $\phi\left(a_{ij}\right)$. Consider a stochastic matrix $A = (a_{ij})$ whose i-th row has entries with two possible values: r_i entries with value $1/r_i$ and $n - r_i$ entries valued 0. Then the Bregman bound implies

$$per(\phi(A)) \leq 1,$$

and equality holds iff A is a block-diagonal matrix with blocks which are constant multiples of all-1 matrices.

A natural way to extend ϕ to the whole interval $[0,1]$ is by taking $\phi(x) = \Gamma\left(1/x+1\right)^{-x}$, for all $0 < x \leq 1$, and setting $\phi(0) = 0$. The following conjecture generalizes the Minc conjecture.

Conjecture 1.6: For any stochastic matrix A holds

$$per(\phi(A)) \leq 1$$

and equality holds iff A is a block-diagonal matrix with blocks which are constant multiples of all-1 matrices. ∎

The function $\phi = \Gamma\left(1/x+1\right)^{-x}$ is strictly monotone and takes $[0,1]$ onto $[0,1]$. It is also concave [14].

Let $K = \{x \in \mathbb{R}^n;\ \sum_{i=1}^{n} \phi\left(x_i\right) \leq 1\}$. This is a convex ball in $\mathbb{R}^n$ defining a norm $\|\cdot\|_K$. Consider the following optimization problem: Choose n unit vectors $x^{(1)}...x^{(n)}$ in $\mathbb{R}^n$ endowed

[2] Up to a permutation of rows or columns.

with the norm $\|\cdot\|_K$ as rows of a matrix so that the permanent of this matrix is as large as possible. [3]

An alternative way to state Conjecture 1.6 is to say that all the optimal solutions to this optimization problem are obtained as follows: partition $\{1...n\}$ into disjoint subsets $S_1...S_k$. For each $j = 1...k$ choose all the vectors $x^{(i)}$, $i \in S_j$, to be equal to $\frac{1}{|S_j|} \cdot \mathbf{1}_{S_j}$, that is be $\frac{1}{|S_j|}$ on the coordinates in S_j, and 0 elsewhere.

The function ϕ and the norm it defines are somewhat compicated to deal with. A natural "easier" family of norms to consider as a test case are the l_p norms, $1 \leq p \leq \infty$. This, in fact, was the starting point of this study.

A word on our methods and an acknowledgement. Our proof of Theorem 1.3 proceeds along the lines of Bregman's proof of the Minc conjecture. A key inequality in that proof has to be replaced by a more general inequality of [1], quoted as Theorem 2.3 below. We are grateful to Leonid Gurvits for directing us to this inequality.

2 A recursive bound on U(n,p)

Let $1 \leq p \leq \infty$ be fixed. Let $q = 1/p$.

A vector $y = (y_1...y_n) \in \mathbb{R}^n$ is *stochastic* if its coordinates are nonnegative and sum to 1. Consider the following function defined on the set Δ of stochastic vectors:

$$P(y) = \sum_{i=1}^{n} y_i^q \prod_{j \neq i} (1 - y_j)^q$$

This is a continuous bounded function which attains its maximum on Δ.

Definition 2.1:

$$w(n,p) = \max_{y \in \Delta} P(y)$$

▮

The main claim of this section is:

Theorem 2.2:

$$U(n,p) \leq \prod_{k=1}^{n} w(k,p)$$

Proof: The proof is by induction on n. For $n = 1$, $U(1,p) = w(1,p) = 1$.

Consider an optimization problem

$$\text{Maximize} \quad Per\left(\lambda_{ij}^q\right)$$

[3] Replacing permanent with determinant one arrives to questions about the maximal volume subcube of K. These questions are of interest in convex geometry [2]. The two contexts seem to be very different, however.

Given

$$\lambda_{ij} \geq 0 \qquad \forall i \ \sum_{j=1}^{n} \lambda_{ij} = 1$$

Clearly the optimal value here is $U(n,p)$.

A key element of our proof is an inequality of [1], which we state next.

Theorem 2.3: *Let $p(x,\lambda)$ be a nonnegative function defined on a space $X \times \Lambda$ and let μ be a nonnegative weight function on X.*

Let $P(\lambda) = \sum_{x\in X} \mu(x)p(x,\lambda)$, and $Q(\lambda,\bar{\lambda}) = \sum_{x\in X} \mu(x)p(x,\lambda)\log p(x,\bar{\lambda})$.

Then $Q(\lambda,\bar{\lambda}) \geq Q(\lambda,\lambda)$ implies $P(\bar{\lambda}) > P(\lambda)$ unless $p(x,\lambda) = p(x,\bar{\lambda})$ for all x with $\mu(x) > 0$.

Now we apply Theorem 2.3 in our setting.

Let $X = S_n$ be the symmetric group on n elements, and Λ be the set of all stochastic matrices (λ_{ij}). Let $\mu(\sigma) = 1$ for all permutations $\sigma \in S_n$ and let $p(\sigma,\lambda) = \prod_{i=1}^{n} \lambda_{i,\sigma(i)}^{q}$, for $\sigma \in S_n$ and $\lambda \in \Lambda$. Then $P(\lambda) = \sum_{x\in X} \mu(x)p(x,\lambda) = Per\left(\lambda_{ij}^{q}\right)$.

Let $\lambda[i,j]$ be the $(n-1)\times(n-1)$ matrix obtained from λ by deleting i-th row and j-th column. Let $\lambda^q[i,j]$ be the matrix obtained from $\lambda[i,j]$ by raising each entry to q-th power. Let $\bar{\lambda} = (\bar{\lambda}_{ij})$ with

$$\bar{\lambda}_{ij} = \frac{\lambda_{ij}^{q} Per\left(\lambda^q[i,j]\right)}{Per\left(\lambda_{ij}^{q}\right)}$$

The following lemma is a direct consequence of Theorem 2.3.

Lemma 2.4:

$$Per\left(\bar{\lambda}_{ij}^{q}\right) \geq Per\left(\lambda_{ij}^{q}\right)$$

Proof: Consider the optimization problem of maximizing $Q\left(\lambda,\bar{\lambda}\right)$ given λ. We have

$$Q(\lambda,\bar{\lambda}) = \sum_{\sigma\in S_n} p(\sigma,\lambda)\log p(\sigma,\bar{\lambda}) = q\cdot \sum_{\sigma\in S_n} p(\sigma,\lambda)\sum_{i=1}^{n}\log\bar{\lambda}_{i\sigma(i)} = q\cdot\sum_{i,j=1}^{n}\lambda_{ij}^{q} Per\left(\lambda^q[i,j]\right)\log\bar{\lambda}_{ij}$$

The constraints on $\bar{\lambda}$ are that it is a stochastic matrix. Therefore we have n independent optimization problems of the form:

$$\text{Maximize} \ \sum w_j \log y_j \qquad \text{Given} \ \ y_j \geq 0, \ \ \sum y_j = 1,$$

where w_j are nonnegative constants. Assuming not all w_j are zero, which we may and will do in our case, the only solution of this problem is $y_j = \frac{w_j}{\sum_k w_k}$. This is a simple consequence of the concavity of the logarithm.

Fixing $1 \leq i \leq n$, and substituting $w_j = \lambda_{ij}^{q} Per\left(\lambda^q[i,j]\right)$ and $y_j = \bar{\lambda}_{ij}$, we see that optimal $\bar{\lambda}$ is given by $\bar{\lambda}_{ij} = \frac{\lambda_{ij}^{q} Per(\lambda^q[i,j])}{Per\left(\lambda_{ij}^{q}\right)}$. The claim of the lemma now follows from Theorem 2.3. ∎

Now, following [3], we write

$$Per\left(\lambda_{ij}^{q}\right) \le Per\left(\bar{\lambda}_{ij}^{q}\right) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} \bar{\lambda}_{i\sigma i}^{q} =$$

$$\sum_{\sigma \in S_n} \prod_{i=1}^{n} \left(\frac{\lambda_{i\sigma(i)}^{q} Per\left(\lambda^{q}[i,\sigma(i)]\right)}{Per\left(\lambda_{ij}^{q}\right)} \right)^{q} = \frac{1}{Per^{qn}\left(\lambda_{ij}^{q}\right)} \cdot \sum_{\sigma \in S_n} \prod_{i=1}^{n} \lambda_{i\sigma(i)}^{q^2} Per^{q}\left(\lambda^{q}[i,\sigma(i)]\right)$$

Let (λ_{ij}^{q}) be an optimal matrix, that is $Per\left((\lambda_{ij}^{q})\right) = U(n,p)$. Then

$$U(n,p)^{qn+1} \le \sum_{\sigma \in S_n} \prod_{i=1}^{n} \lambda_{i\sigma(i)}^{q^2} Per^{q}\left(\lambda^{q}[i,\sigma(i)]\right)$$

Consider the matrix $\lambda[i,j]$. This is an $(n-1)\times(n-1)$ matrix with row sums $r_k = 1-\lambda_{kj}$, for $k = 1...n$, $k \ne i$. Let R be the $(n-1)\times(n-1)$ diagonal matrix with $1/(1-r_k)$ on the diagonal. Then $(a_{ij}) = R\cdot\lambda[i,j]$ is a stochastic matrix, and therefore, by induction hypothesis, $Per(a_{ij}^{q}) \le U(n-1,p)$.

This means $Per\left(\lambda^{q}[i,j]\right) \le U(n-1,p)\cdot \prod_{k\ne i}(1-\lambda_{kj})^{q}$. Substituting this in the inequality above, we obtain

$$U(n,p)^{qn+1} \le U(n-1,p)^{qn} \cdot \sum_{\sigma \in S_n} \prod_{i=1}^{n} \left(\lambda_{i\sigma(i)}^{q^2} \prod_{k\ne i} (1-\lambda_{k\sigma(i)})^{q^2} \right) =$$

$$U(n-1,p)^{qn} \cdot \sum_{\sigma \in S_n} \left(\prod_{i=1}^{n} \lambda_{i\sigma(i)} \cdot \prod_{k,j:\ \sigma(k)\ne j} (1-\lambda_{kj}) \right)^{q^2} =$$

$$U(n-1,p)^{qn} \cdot \prod_{i,j} (1-\lambda_{ij})^{q^2} \cdot \sum_{\sigma \in S_n} \prod_{i=1}^{n} \left(\frac{\lambda_{i\sigma(i)}^{q}}{(1-\lambda_{i\sigma(i)})^{q}} \right)^{q}$$

The third term in this expression is the permanent of a matrix $\left(a_{ij}^{q}\right)$, where $a_{ij} = \frac{\lambda_{ij}^{q}}{(1-\lambda_{ij})^{q}}$.

Let $r_i = \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} \frac{\lambda_{ij}^{q}}{(1-\lambda_{ij})^{q}}$ be the row sums of this matrix. Then, $Per(a_{ij}^{q}) \le U(n,p)\cdot \prod_{i=1}^{n} r_i^{q}$. Substituting in the inequality above gives

$$U(n,p)^{qn} \le U(n-1,p)^{qn} \cdot \prod_{i,j} (1-\lambda_{ij})^{q^2} \cdot \left(\prod_{i=1}^{n} \sum_{j=1}^{n} \frac{\lambda_{ij}^{q}}{(1-\lambda_{ij})^{q}} \right)^{q}$$

Taking q-th roots of both sides this simplifies to

$$U(n,p)^{n} \le U(n-1,p)^{n} \cdot \prod_{i,j} (1-\lambda_{ij})^{q} \cdot \prod_{i=1}^{n} \sum_{j=1}^{n} \frac{\lambda_{ij}^{q}}{(1-\lambda_{ij})^{q}}$$

Let λ_i be the i-th row vector of λ. Since λ is a stochastic matrix, λ_i is a stochastic vector. We have

$$\prod_{i,j} (1-\lambda_{ij})^q \cdot \prod_{i=1}^n \sum_{j=1}^n \frac{\lambda_{ij}^q}{(1-\lambda_{ij})^q} = \prod_{i=1}^n P(\lambda_i) \leq w(n,p)^n$$

Therefore $U(n,p) \leq U(n-1,p) \cdot w(n,p)$. The claim now follows from the induction hypothesis

$$U(n,p) \leq U(n-1,p) \cdot w(n,p) \leq w(n,p) \cdot \prod_{k=1}^{n-1} w(k,p) = \prod_{k=1}^{n} w(k,p)$$

∎

3 Proofs of the main results

Our first order of business is to determine $w(k,p)$, for $1 \leq k \leq n$. Let $1 < p < 2$ be fixed, and let $q = 1/p$.

Let $\theta(k) = k \cdot \left(\frac{(k-1)^{k-1}}{k^k}\right)^q$ for integer $k \geq 2$ and let $\theta(1) = 1$.

Theorem 3.1: *Fix $k \geq 1$. The maximum of $P(y) = \sum_{i=1}^k y_i^q \prod_{j \neq i}(1-y_j)^q$ is attained either at a standard basis vector and then $w(k,p) = \theta(1) = 1$, or at the all-$1/k$ vector, in which case $w(k,p) = \theta(k)$.*

The proof of Theorem 3.1 is technical and is relegated to Appendix.

We briefly discuss the claim of the theorem. Let I_k be the $k \times k$ identity matrix. Let J_k denote the matrix $k^{-1/p} \cdot J$, where J is the all-1 $k \times k$ matrix. Note that $\theta(k) = \frac{per(J_k)}{per(J_{k-1})}$. Therefore the theorem, combined with Theorem 2.2, says that for any $k \geq 2$

$$\frac{U(k,p)}{U(k-1,p)} \leq \max\left\{\frac{per\,(I_k)}{per\,(I_{k-1})}, \frac{per\,(J_k)}{per\,(J_{k-1})}\right\}$$

Let us observe that this inequality agrees well with Conjecture 1.1.

The last step before the proof of Theorem 1.3 is Lemma 1.2, which we prove now.

Proof: (Lemma 1.2)

The following notation will be convenient. For $1 \leq p \leq \infty$, let $\Omega(n,p)$ be the set of $n \times n$ matrices whose rows are unit vectors in l_p.

We need a following well-known fact. Let $1 \leq p < p' \leq \infty$. Let a be a vector in $\mathbb{R}^n$. Then

$$1 \leq \frac{\|a\|_p}{\|a\|_{p'}} \leq n^{\frac{1}{p}-\frac{1}{p'}} \tag{3}$$

Equality on the left is possible only for a multiple of a standard basis vector, and equality on the right is possible only for a multiple of the all-1 vector.

Let p_0 be such that the matrix I is optimal for p_0. Let $p < p_0$. Let $A \in \Omega(n,p)$ with rows $a_1 ... a_n$. Let $D = (d_{ij})$ be a diagonal matrix with $d_{ii} = \frac{\|a_i\|_p}{\|a_i\|_{p_0}}$. Then DA is in $\Omega(n,p_0)$ and therefore

$$per(A) = per\left(D^{-1} \cdot (DA)\right) = per\left(D^{-1}\right) \cdot per(DA) = \prod_{i=1}^{n} \frac{\|a_i\|_{p_0}}{\|a_i\|_p} \cdot per(DA)$$

$$\leq \prod_{i=1}^{n} \frac{\|a_i\|_{p_0}}{\|a_i\|_p} \cdot per(I) = \prod_{i=1}^{n} \frac{\|a_i\|_{p_0}}{\|a_i\|_p} \leq 1$$

By (3) equality is only possible if all the rows a_i are standard basis vectors, and A is the identity matrix, up to permuting coordinates.

This proves the first claim of the lemma. The proof of the second claim proceeds along similar lines, using second half of inequality (3). We omit the details. ∎

Proof: (Theorem 1.3)

Fix $p = p_0 = \frac{n \log n - (n-1)\log(n-1)}{\log n}$. Let $q = 1/p$. The value of p is chosen precisely so that $\theta(n) = n \cdot \left(\frac{(n-1)^{n-1}}{n^n}\right)^q = 1$.

By Theorem 2.2, Theorem 3.1, and Lemma 4.1

$$U(n,p) \leq \prod_{k=1}^{n} w(k,p) \leq \prod_{k=1}^{n} \max\{1, \theta(k)\} \leq (\max\{1, \theta(n)\})^n = 1$$

Therefore I is optimal for $p = p_0$. Lemma 1.2 completes the proof of the first claim of the theorem.

Now, to the second claim. Fix $p \in (1,2)$. Let $q = 1/p$. By Lemma 4.1 there is an integer k_0 such that $\theta(k) < 1$ for $k \leq k_0$ and $\theta(k) \geq 1$ for $k > k_0$. Since $\theta(k) = \frac{per(J_k)}{per(J_{k-1})}$, this means that $per\left(J_{k_0}\right) = \prod_{k=1}^{k_0} \theta(k) = \min_{k \geq 1} per\left(J_k\right)$.

Therefore,

$$U(n,p_0) \leq \prod_{k=1}^{n} w(k,p_0) \leq \prod_{k=1}^{n} \max\{1, \theta(k)\} = \prod_{k=2}^{n} \max\left\{1, \frac{per\left(J_k\right)}{per\left(J_{k-1}\right)}\right\} =$$

$$\frac{per\left(J_n\right)}{per\left(J_{k_0}\right)} = \frac{per\left(J_n\right)}{\min_{k \geq 1} per\left(J_k\right)}$$

It remains to estimate the denominator on the right.

We have

$$\min_{k \geq 1} per\left(J_k\right) = \min_{k \geq 1} \frac{k!}{k^{qk}} \geq \min_{k \geq 1} \frac{k^{(1-q)k}}{e^k} \geq \min_{x \geq 1} \frac{x^{(1-q)x}}{e^x}$$

where in the last inequality an integer variable k is replaced with a real variable x. A simple analysis gives that the minumum on the right hand side is attained for $x = \exp\{q/(1-q)\} = \exp\{1/(p-1)\}$ and equals $\exp\left\{-(p-1)/p \cdot e^{1/(p-1)}\right\}$.

Therefore

$$U(n,p) \leq \exp\left\{(p-1)/p \cdot e^{1/(p-1)}\right\} \cdot per\,(J_n) = \exp\left\{(p-1)/p \cdot e^{1/(p-1)}\right\} \cdot \frac{n!}{n^{n/p}}$$

This completes the proof of the second claim and of the theorem. ∎

4 Appendix: A Proof of Theorem 3.1

We start with a useful property of the function θ. Let $1/2 < q < 1$ be a real number.

Lemma 4.1: *Let $k \geq 1$ and consider the continuous function $\theta(x) = x \cdot \left(\frac{(x-1)^{x-1}}{x^x}\right)^q$ of a real variable x on the interval $[1,k]$. If x_0 is a point of maximum of θ then $x_0 = 1$ or $x_0 = k$.*

Proof: It is convenient to deal with $f(x) = \ln(\theta(x)) = \ln x - q \cdot (x \ln x - (x-1)\ln(x-1))$. The derivative $f'(x) = \frac{1}{x} - q \ln \frac{x}{x-1} = \frac{q}{x} \cdot \left(\frac{1}{q} - x \ln \frac{x}{x-1}\right)$.

Consider the function $g(x) = x \ln \frac{x}{x-1}$ on $[1,\infty)$. The derivative $g'(x) = \ln \frac{x}{x-1} - \frac{1}{x-1} = \ln\left(1 + \frac{1}{x-1}\right) - \frac{1}{x-1}$ is strictly negative. At the endpoints, $g(1) = \infty$ and $g(\infty) = 1$. Therefore on $[1,\infty)$ the function g decreases from ∞ to 1. Since $\frac{1}{q} = p > 1$ this means that there exists a positive real number $x_q > 1$ depending only on q such that $f' < 0$ for $1 \leq x < x_q$, $f'(x_q) = 0$, and $f'(x) > 0$ for $x > x_q$.

Consequently, f is unimodal on $[1,\infty)$ with minimum in x_q. The claim of the lemma follows. ∎

The proof of the theorem proceeds by induction on k. For $k = 1$ the claim holds trivially. For $k = 2$ we have

$$P(y) = P(y_1, 1-y_1) = y_1^{2q} + (1-y_1)^{2q}$$

For $q > 1/2$, the function $f(x) = x^{2q} + (1-x)^{2q}$ attains its maximum on $[0,1]$ at 0 and at 1. This means that the points of maximum of P are standard basis vectors, and the claim holds.

Assume the theorem is true for $2 \leq l < k$.

Let $y^* \in \Delta$ be a point at which P attains maximum. If y^* has $1 < l < k$ non-zero coordinates, then the induction hypothesis implies y^* is the all-$1/l$ vector. This is to say $P(y^*) = \theta(l)$. However, Lemma 4.1 showed $\theta(l) < \max\{1, \theta(k)\}$, reaching a contradiction.

Therefore either y^* is a standard basis vector, in which case we are done, or y^* is an interior point of Δ. This is the remaining case. We will assume that y^* is not the all-$1/k$ vector and reach a contradiction.

Since y^* is an interior extremum point, we can use the first and the second order optimality conditions on the gradient and the Hessian of P at y^* to obtain information about y^*.

Let $s_i(y) = y^q \prod_{j \neq i} (1-y_j)^q$, for $i = 1...k$. Of course $P = \sum_{i=1}^k s_i$.

Lemma 4.2: *For all $i = 1...k$*

$$s_i\left(y^*\right) = y_i^* P\left(y^*\right)$$

Proof: We have $\frac{\partial s_i}{\partial y_i} = \frac{q s_i}{y_i}$ and, for $j \neq i$, $\frac{\partial s_i}{\partial y_j} = -\frac{q s_i}{1-y_j}$. Therefore

$$\frac{\partial P}{\partial y_j} = \sum_{i=1}^{n} \frac{\partial s_i}{\partial y_j} = \frac{\partial s_j}{\partial y_j} + \frac{\partial P - s_j}{\partial y_j} = q \cdot \left(\frac{s_j}{y_j} - \frac{P - s_j}{1 - y_j}\right) = q \cdot \frac{s_j - y_j P}{y_j(1 - y_j)}$$

The first order optimality conditions for y^* say that there is a constant λ such that for all $j = 1...k$ holds $\frac{\partial P}{\partial y_j}\left(y^*\right) = \lambda$. This means that for $j = 1...k$ holds $s_j\left(y^*\right) - y_j^* P\left(y^*\right) = \frac{\lambda}{q} y_j^* \left(1 - y_j^*\right)$.

Summing over j we obtain

$$\frac{\lambda}{q} \cdot \sum_{j=1}^{k} y_j^* \left(1 - y_j^*\right) = 0,$$

implying $\lambda = 0$. That is, for all $j = 1...k$ holds $s_j\left(y^*\right) = y_j^* P\left(y^*\right)$. ∎

Corollary 4.3: *The coordinates of y^* have two distinct values a and b with $a < 1 - q < b$.*

Proof: Let $i \neq j$ be two distinct indices. By the lemma at y^* we have $s_i = y_i^* P$ and $s_j = y_j^* P$. This implies

$$\frac{y_i^*}{y_j^*} = \frac{s_i}{s_j} = \frac{\left(y_i^*\right)^q \left(1 - y_j^*\right)^q}{\left(y_j^*\right)^q \left(1 - y_i^*\right)^q}$$

This means $\left(y_i^*\right)^{1-q} \left(1 - y_i^*\right)^q = \left(y_j^*\right)^{1-q} \left(1 - y_j^*\right)^q$. Let $f(x) = x^{1-q}(1-x)^q$. We have shown that $f\left(y_i^*\right) = f\left(y_j^*\right)$. Since the argument does not depend on the choice of i and j, this implies f has the same value on all y_i^*, $i = 1...k$.

The function f is a concave function on $[0, 1]$ vanishing at the endpoints, with maximum at $1 - q$. Therefore f takes each value at most twice, at two points lying on different sides of $1 - q$. Bearing in mind that y^* is not a constant vector, the claim of the corollary follows. ∎

Next, we compute the Hessian of P. We have, for $i \neq j \neq t$

$$\frac{\partial^2 s_i}{\partial y_i^2} = -\frac{q(1-q)s_i}{y_i^2}; \qquad \frac{\partial^2 s_i}{\partial y_j^2} = -\frac{q(1-q)s_i}{\left(1 - y_j\right)^2};$$

$$\frac{\partial^2 s_i}{\partial y_i \partial y_j} = -\frac{q^2 s_i}{y_i\left(1 - y_j\right)}; \qquad \frac{\partial^2 s_i}{\partial y_j \partial y_t} = \frac{q^2 s_i}{\left(1 - y_j\right)\left(1 - y_t\right)}$$

Let $H = H(y)$ be the Hessian of P at y. Then

$$H(j,j) = \frac{\partial^2 P}{\partial y_j^2} = \sum_{i=1}^{k} \frac{\partial^2 s_i}{\partial y_j^2} = -q(1-q) \cdot \left(\frac{s_j}{y_j^2} + \frac{P - s_j}{(1 - y_j)^2}\right)$$

Similarly

$$H(j,t) = \frac{\partial^2 P}{\partial y_j \partial y_t} = \sum_{i=1}^{k} \frac{\partial^2 s_i}{\partial y_j \partial y_t} = \frac{\partial^2 s_j}{\partial y_j \partial y_t} + \frac{\partial^2 s_t}{\partial y_j \partial y_t} + \frac{\partial^2 (P - s_j - s_t)}{\partial y_j \partial y_t} =$$

$$-q^2 \cdot \left(\frac{s_j}{y_j(1-y_t)} + \frac{s_t}{y_t(1-y_j)} \right) + q^2 \cdot \frac{P - s_j - s_t}{(1-y_j)(1-y_t)}$$

At y^* we have $s_i = y_i^* P$ for all $i = 1...k$. Therefore for $H = H(y^*)$ we have

$$H(j,j) = -q(1-q) \cdot \frac{P}{y_j^*(1-y_j^*)}$$

and

$$H(j,t) = q^2 P \cdot \left(\frac{1}{(1-y_j^*)(1-y_t^*)} - \frac{1}{(1-y_j^*)} - \frac{1}{(1-y_t^*)} \right) = -q^2 \cdot \frac{P}{(1-y_j^*)(1-y_t^*)}$$

Lemma 4.4: *y^* has only one coordinate with value b. (And therefore $k-1$ coordinates with value a.)*

Proof: We can write the Hessian at y^* as $H = -qP \cdot (A + D)$, where A is a rank-1 matrix with $a_{ij} = \frac{q}{(1-y_i^*)(1-y_j^*)}$, and D is a diagonal matrix with $d_{ii} = \frac{1-q-y_i^*}{y_i^*(1-y_i^*)^2}$.

The second order optimality conditions for y^* say that H is negative semidefinite on the subspace V of the vectors in $\mathbb{R}^k$ orthogonal to the all-1 vector. This means that the matrix $B = A + D$ is positive semidefinite on V.

Assume for the moment that y^* has two b-valued coordinates. Let these be the first two coordinates. This means that $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1/(1-b)^2 & 1/(1-b)^2 \\ 1/(1-b)^2 & 1/(1-b)^2 \end{bmatrix}$, and $\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} = \begin{bmatrix} \frac{1-q-b}{b(1-b)^2} & 0 \\ 0 & \frac{1-q-b}{b(1-b)^2} \end{bmatrix}$. Note, that since $b > 1-q$, the diagonal values of the second matrix are negative.

Now, let $v \in V$, $v = (1, -1, 0, \ldots, 0)$. Then clearly

$$vBv^t = 2\frac{1-q-b}{b(1-b)^2} < 0,$$

contradicting positive semidefinitness of B. This means that y^* has only one coordinate valued b. ∎

Consider the set $\Delta_1 \subset \Delta$ of stochastic vectors y with $y_2 = ... = y_k = \frac{1-y_1}{k-1}$. The preceding lemma implies that there is a maximum point y^* of P in Δ_1. Moreover $b = y_1^* > 1/k$.

P, restricted to Δ_1, is a function of one variable $x = y_1$ and is given by

$$P(x) = x^q \left(1 - \frac{1-x}{k-1}\right)^{(k-1)q} + (k-1)\left(\frac{1-x}{k-1}\right)^q (1-x)^q \left(1 - \frac{1-x}{k-1}\right)^{(k-2)q} =$$

$$\frac{1}{(k-1)^{(k-1)q}} \cdot \left(x^q(k-2+x)^{(k-1)q} + (k-1)(1-x)^{2q}(k-2+x)^{(k-2)q}\right)$$

We will show that on the interval $[1/k, 1]$ this function attains its maximum either at $1/k$ or at 1. This means, recalling $y_1^* > 1/k$, that y^* is a standard basis vector. This is a contradiction to previous assumptions, and will complete the proof of the theorem.

Lemma 4.5: *Let $k \geq 3$ be an integer, let $1/2 < q < 1$ be a real number, and let f be a function on $[1/k, 1]$ given by*

$$f(x) = x^q(k-2+x)^{(k-1)q} + (k-1)(1-x)^{2q}(k-2+x)^{(k-2)q}$$

Then f attains its maximum either at $1/k$ or at 1.

Proof: We compute the derivative of f.

$$f'(x) = qx^{q-1}(k-2+x)^{(k-1)q} \ + \ (k-1)qx^q(k-2+x)^{(k-1)q-1} \ -$$

$$2(k-1)q(1-x)^{2q-1}(k-2+x)^{(k-2)q} \ + \ (k-1)(k-2)q(1-x)^{2q}(k-2+x)^{(k-2)q-1} =$$

$$q(k-2+kx)(k-2+x)^{(k-2)q-1}x^{q-1} \ \cdot \ \left((k-2+x)^q \ - \ (k-1)x^{1-q}(1-x)^{2q-1}\right)$$

This means that the sign of f is determined by the sign of $(k-2+x)^q - (k-1)x^{1-q}(1-x)^{2q-1}$.

Since $t \mapsto t^q$ is monotone increasing, we can, as well, check the sign of

$$h(x) = (k-2+x) \ - \ (k-1)^{1/q}x^{1/q-1}(1-x)^{2-1/q}$$

The function $h(x)$ is strictly convex on $[1/k, 1]$, with $h\left(1/k\right) = 0$ and $h(1) = k-1 > 0$.

Therefore, there are two possible options.

- $h > 0$ on $(1/k, 1]$. This means that f attains its maximum at 1.
- There is a point $x \in (1/k, 1)$ such that $h < 0$ on $(1/k, x)$ and $h > 0$ on $(x, 1)$. This means that f attains its maximum at one of the endpoints $1/k$ or 1, and we are done.

∎

This completes the proof of Theorem 3.1.

5 Acknowledgements

We are grateful to Leonid Gurvits for several very helpful discussions. We also thank Shmuel Friedland, Nati Linial, and Michael Navon for valuable conversations.

References

[1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, Ann. of Math. Stat., vol. 41, 1, 1970 pp. 164-171.

[2] K. Ball, *An elementary introduction to modern convex geometry*, Flavors of Geometry, MSRI Publications, Vol.31, 1997.

[3] L. M. Bregman, *Some properties of nonnegative matrices and their permanents*, Soviet Math. Dokl., vol. 14, 4, 1973, pp. 945-949.

[4] E. Carlen, M. Loss, and E. H. Lieb, *An inequality of Hadamard type for permanents*, Mathematics ArXiv NT/0508096, 2005.

[5] G.P. Egorychev, *The solution of van der Waerden's problem for permanents*, Advances in Math., 42, 299-305, 1981.

[6] D. I. Falikman, *Proof of the van der Waerden's conjecture on the permanent of a doubly stochastic matrix*, Mat. Zametki 29, 6: 931-938, 957, 1981, (in Russian).

[7] S. Friedland, *A study of the van der Waerden conjecture and its generalizations*, Linear and Multilinear Algebra, 6, 1978, pp. 123-143.

[8] L. Gurvits, *Hyperbolic polynomials approach to van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications*, to appear in STOC 2006.

[9] L. Gurvits, personal communication.

[10] L. Gurvits, A. Samorodnitsky, an unpublished manuscript, 2001.

[11] M. Jerrum, A. Sinclair and E. Vigoda, *A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries*, Proc. 33 ACM Symp. on Theory of Computing, ACM, 2001.

[12] N. Linial, A. Samorodnitsky, an unpublished manuscript, 1998.

[13] N. Linial, A. Samorodnitsky and A. Wigderson, *A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents*, Combinatorica, vol. 20, 4, 2000.

[14] D. Moews, $\Gamma(x+1)^{1/x}$ *is concave*, personal communication to the author of [17].

[15] M. Navon, *Some notes on the permanent*, a project report, submitted as a part of MSc thesis requirements, Hebrew University, 2004.

[16] Y. Nesterov, A. Nemirovski, **Interior Point Polynomial Methods in Convex Programming**, SIAM, Philadelphia, 1994.

[17] G. W. Soules, *New permanental upper bounds for nonnegative matrices*, Linear and Multilinear Algebra 51, 2003, pp. 319-337.