Bounds on the permanent and some applications

Leonid Gurvits* and Alex Samorodnitsky†

Abstract

We show that the permanent of a doubly stochastic $n \times n$ matrix $A = (a_{ij})$ is at least as large as $\prod_{i,j} (1 - a_{ij})^{1-a_{ij}}$ and at most as large as 2^n times this number. Combined with previous work, this improves on the deterministic approximation factor for the permanent, giving 2^n instead of e^n-approximation.
We also give a combinatorial application of the lower bound, proving S. Friedland's "Asymptotic Lower Matching Conjecture" for the monomer-dimer problem.

1 Introduction

The permanent of an $n \times n$ matrix $A = (a_{ij})$ is given by

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

Here S_n is the symmetric group on n elements.

The permanent is a classical mathematical notion, going back to Binet and Cauchy [20]. One part of its appeal is its strong, though seemingly spurious, similarity to the determinant. Another part is in its ability to count things. The permanent of a $0,1$ matrix A equals the number of perfect matchings in the bipartite graph it represents. The permanents are also useful in counting more complex subgraphs, such as Hamiltonian cycles ([7] and the references therein).

In fact, the permanent counts things in a very strong sense, since it is $\#P$ to compute [26], even for $0,1$ matrices. Hence, from the complexity point of view, the permanent is very different from the determinant. While the latter is efficiently computable, the permanent of nonnegative matrices is (probably) not. The natural question is, therefore, to try and approximate the permanent as efficiently as possible, and as well as possible.

We briefly discuss three different approaches to achieve this goal.

The Monte Carlo Markov Chain approach: As observed by Jerrum et al [14] an efficient procedure to sample uniformly from the set of all perfect matchings in a bipartite graph is

*Department of Computer Science, Grove School of Engineering, The City College of New York, New York, United States of America. Research supported by NSF grant 116143.

†School of Engineering and Computer Science, The Hebrew University of Jerusalem, Jerusalem, Israel. Research supported by BSF and ISF grants.

computationally equivalent to approximately counting the matchings. Broder [3] proposed to construct such a procedure by devising a random walk on an appropriate space, rapidly converging to its stationary distribution, which would be uniform on the set of perfect matchings (and assign a substantial weight to it). This was accomplished (and extended) in [14], giving an efficient randomized approximation algorithm for the permanent of a nonnegative matrix, up to any degree of precision, and providing a complete solution to the problem.

Exploiting the similarity to determinant: This is based on an observation of Godsil and Gutman [18], that, for a matrix $A = (a_{ij})$ with nonnegative entries, the random matrix $B = \left(\epsilon_{ij} \cdot \sqrt{a_{ij}}\right)$ where ϵ_{ij} are independent random variables with expectation 0 and variance 1, satisfies $Per(A) = \mathbb{E}\ Det^2(B)$. Hence, for an efficient randomized permanent approximation, it would suffice to show the random variable $Det^2(B)$ to be concentrated around its expectation. In [1] the random variables ϵ_{ij} were taken to be quaternionic Gaussians, leading to an efficient randomized approximation algorithm for the permanent, which achieves an approximation factor of about 1.3^n.

Using combinatorial bounds on the permanent: The permanent of a doubly stochastic matrix was shown to be at least $\frac{n!}{n^n} \approx e^{-n}$ in [5, 6], answering a question of van der Waerden. On the other hand, this permanent is (clearly) at most 1. Hence, we already know the permanent of a doubly stochastic matrix up to a factor of e^n. In [16] this fortuitous fact was exploited by showing an efficient reduction of the problem for general nonnegative matrices to that of doubly stochastic matrices. This was done via matrix (Sinkhorn's) scaling: for any matrix $A = (a_{ij})$ with nonnegative entries and positive permanent, one can efficiently find *scaling factors* $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ such that the matrix $B = (x_i \cdot a_{ij} \cdot y_j)$ is (almost) doubly stochastic. Since $Per(A) = \frac{1}{\prod_i x_i \cdot \prod_j y_j} \cdot Per(B)$ this constitutes a reduction, and in fact achieves e^n deterministic approximation for the permanent of a nonnegative matrix.

1.1 Our results

Our paper is a contribution to the third approach. One may say that, in a sense, it takes up where [16] left off. The algorithm of [16] reduces the problem to the case of doubly stochastic matrices, on which it "does nothing", that is returns 1 and quits. The natural next step would be to "actually look at the matrix", that is to come up with an efficiently computable function of the entries of the matrix, which would provide a non-trivial estimate of its permanent.

This is precisely what we do. This efficiently computable function of the doubly stochastic matrix $A = (a_{ij})$ is $F(A) = \prod_{i,j=1^n} (1 - a_{ij})^{1-a_{ij}}$.

We prove new lower and upper bounds for the permanent of a doubly stochastic matrix A, showing that for any such matrix it holds that

$$F(A) \leq Per(A) \leq 2^n \cdot F(A) \tag{1}$$

Combined with the preceding discussion, this gives our main algorithmic result.

Theorem 1.1: *There is a deterministic polynomial-time algorithm to approximate the permanent of a nonnegative matrix up to a multiplicative factor of 2^n.*

Let us now briefly describe the ideas leading to the bounds in (1).

We proceed via convex relaxation. That is, given a matrix A with nonnegative entries, we define a concave maximization problem, whose solution approximates $\log(Per(A))$.

Let us start with pointing out that approximating the permanent via matrix scaling may also be achieved by solving a convex optimization problem. In fact, what we need is to find the product of scaling factors $\prod_i x_i \cdot \prod_j y_j$ of A. This could be done in two different ways:

By solving a concave maximization problem:

$$\log\left(\frac{1}{\prod_i x_i \cdot \prod_j y_j}\right) = \max_{B\in\Omega_n} \sum_{1\le i,j\le i,j} b_{i,j} \log\left(\frac{a_{i,j}}{b_{i,j}}\right) \tag{2}$$

Here Ω_n is the set of all $n \times n$ doubly stochastic matrices.

And by solving a convex minimization problem:

$$\log\left(\frac{1}{\prod_i x_i \cdot \prod_j y_j}\right) = \inf_{x_1+...+x_n=0} \log\left(Prod_A\left(e^{x_1}, ..., e^{x_n}\right)\right), \tag{3}$$

where $Prod_A(x_1, ..., x_n)$ is the *product polynomial* of A,

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

Note that $Per(A)$ is the *mixed derivative* of $Prod_A$: $Per(A) = \frac{\partial^n}{\partial x_1 ... \partial x_n} Prod_A(0, ..., 0)$.

The relaxation (2) is very specifically tied to the permanent. On the other hand, (3) is much more general, in that it aims to approximate the mixed derivative of a homogeneous polynomial $p(x_1, ..., x_n)$ of degree n with non-negative coefficients, given via an evaluation oracle[1].

In [11], the relaxation (3) was shown to provide an e^n-approximation of the mixed derivative for a large class of homogeneous polynomials, containing the product polynomial. Moreover, it is the first step in a hierarchy of sharper relaxations given by considering

$$\gamma_i =: \inf_{x_1+...+x_i=0} \log(Q_i(e^{x_1}, ..., e^{x_n})),$$

where $Q_i(x_1, ..., x_i) = \frac{\partial^{n-i}}{\partial x_{i+1} ... \partial x_n} p(x_1, ..., x_i, 0, ..., 0)$.

If the (multivariate) polynomial p does not have roots with positive real parts (in this case it is known as *H-Stable*, or *hyperbolic*) then

$$G\left(deg_{Q_{i+1}}(i+1)\right) \cdot exp(\gamma_i) \le Per(A) \le exp(\gamma_i) \le exp(\gamma_{i+1}),$$

[1] Note that the product polynomial can be efficiently evaluated.

where $G(k) =: \left(\frac{k-1}{k}\right)^{k-1}$ and $deg_{Q_{i+1}}(i+1)$ is the degree of the variable x_{i+1} in the polynomial Q_{i+1}. In particular,

$$\frac{i!}{i^i} \cdot exp(\gamma_i) \leq Per(A) \leq exp(\gamma_i). \tag{4}$$

Considering this hierarchy turns out to be very useful, both from mathematical and from algorithmic points of view [11], [17]. Note that, when this approach is applied to the product polynomial $Prod_A$, the original matrix structure is essentially lost. But by giving up the matrix structure, we gain additional inductive abilities, leading, in particular, to a rather simple proof of (4).

Unfortunately, we only know how to compute γ_i in $poly(n) \cdot 2^{n-i}$ oracle calls, which is polynomial-time only for $i = n - O(\log(n))$. In other words, this "hyperbolic polynomials" approach does not seem to break the e^n-barrier for the approximation of the permanent by a polynomial-time deterministic algorithm.

So, the challenge was to come up with a better convex relaxation. Such a relaxation was suggested in [4], and it is a generalization of (2). It is a special case of a well-known heuristics in Machine Learning, the so called *Bethe Approximation.* This heuristics is used to approximate *log partition functions* of the following type (appearing, in particular, in the analysis of *Belief Propagation* algorithms).

$$PF =: \log \left(\sum_{\substack{x_i \in S_i \\ i = 1 ... n}} \prod_i G_i\left(x_i\right) \cdot \prod_{(i,j) \in E} F_{i,j}\left(x_i, x_j\right) \right) \tag{5}$$

Here S_i are finite sets; $G_i(x_i)$ and $F_{i,j}(x_i, x_j)$ are given non-negative functions, and E is the set of edges of the associated undirected graph Γ.

If the graph Γ is a tree then PF can be efficiently evaluated, e.g. by dynamic programming. The Bethe Approximation is a heuristic to handle possible cycles. It turns out that $\log(Per(A)$ can be represented as in (5). This was first observed in [13]. In this paper we use a simplified version of this heuristic proposed in [4], which amounts to approximating the logarithm of the permanent of a nonnegative matrix A by

$$\max_{B \in \Omega_n} \sum_{i,j=1}^{n} (1 - b_{ij}) \log (1 - b_{ij}) + \sum_{i,j=1}^{n} b_{ij} \log \left(\frac{a_{ij}}{b_{ij}} \right). \tag{6}$$

We should mention that, according to [15], the physicists had already applied the Bethe Approximation to the closely related monomer-dimer problem as early as in late 1930s.

Lower bound

We prove that (6) is a lower bound on $\log(Per(A))$.

Theorem 1.2: *Let $A = (a_{ij})_{i,j=1}^{n}$ be a nonnegative matrix and let $B = (b_{ij})_{i,j=1}^{n}$ be a doubly stochastic matrix. Then*

$$Per(A) \geq \prod_{i,j=1}^{n} (1 - b_{ij})^{1-b_{ij}} \cdot \exp\left\{ - \sum_{i,j=1}^{n} b_{ij} \log \frac{b_{ij}}{a_{ij}} \right\} \tag{7}$$

Let us note that this claim was first stated (but not proved) in [27].

If A is doubly stochastic, setting $B = A$ in (7) gives the lower bound in (1).

Theorem 1.2 has an additional combinatorial application. We show it to imply S. Friedland's "Asymptotic Lower Matching Conjecture" for the monomer-dimer problem. We will go into details in Section 3.

Upper bound

We prove that 2^n times (6) is an upper bound on $\log(Per(A))$.

Theorem 1.3: *The permanent of a stochastic matrix $A = (a_{ij})$ satisfies*

$$Per(A) \leq C^n \cdot \prod_{ij} (1 - a_{ij})^{1-a_{ij}}$$

for some $C \leq 2$.

Note that this implies, in particular, that for a nonnegative matrix A, and its *doubly stochastic scaling* B, we have

$$Per(A) \leq 2^n \cdot \prod_{i,j=1}^{n} (1 - b_{ij})^{1-b_{ij}} \cdot \exp\left\{ - \sum_{i,j=1}^{n} b_{ij} \log \frac{b_{ij}}{a_{ij}} \right\}$$

Remark 1.4:

- Let

  $$CW(A, B) = \sum_{i,j=1}^{n} (1 - b_{ij}) \log (1 - b_{ij}) + \sum_{i,j=1}^{n} b_{ij} \log \left(\frac{a_{ij}}{b_{ij}} \right)$$

  The functional $CW(A, B)$ is clearly concave in A. Less obviously, it is concave in $B \in \Omega_n$ [27]. So, in principle, the concave maximization problem (6) can be solved in polynomial deterministic time by, say, the ellipsoid method.

We don't use the concavity in B in this paper. The algorithm we propose and analyze first scales the matrix A to a doubly-stochastic matrix D and outputs $\prod_{i,j=1}^{n}(1-d_{ij})^{1-d_{ij}}$ multiplied by the product of the scaling factors. So, when applied to a doubly-stochastic matrix, our algorithm has linear complexity.

There are several benefits in using this suboptimal algorithm. First: We can analyze it. Second: It is fast, and local (looking only at the entries) in the doubly-stochastic case. Third: it already improves on e^n-approximation. Fourth: it might allow (conjectural) generalizations to the hyperbolic polynomials setting, to be described in the journal version.

We also conjecture that our algorithm, might in fact turn out to be optimal. That is, that its worst case accuracy is the same as that of the Bethe Approximation (6).

- Let us remark that our results can be viewed as reasonably sharp bounds on a specific partition function in terms of its Bethe Approximation. To the best of our knowledge, this might be one of the first results of this type, and one of the first applications of the Bethe Approximation to theoretical computer science.

$\blacksquare$

Discussion. It would seem that the improvement of the approximation factor from one exponential to a smaller one leaves something to be desired. This is, of course, true. On the other hand, let us remark that any algorithm which considers only the distribution of the entries of the matrix cannot achieve better than $2^{n/2}$ approximation for the permanent. This was pointed out to us by [28]. In fact, consider the following two $0,1$ matrices, both having 2 ones in each row and column. The matrix A_1 is a block-diagonal matrix, with $n/2$ blocks of $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ on the diagonal (assume n is even). The matrix A_2 is the adjacency matrix of a $2n$-cycle, viewed as a bipartite graph with n vertices on each side. The permanent of A_1 is clearly $2^{n/2}$, while the permanent of A_2 is 2.

We conjecture that this optimal approximation factor of $2^{n/2}$ can be attained, by improving our upper bound.

Conjecture 1.5: The permanent of a doubly stochastic matrix $A = (a_{ij})$ satisfies

$$Per(A) \leq 2^{n/2} \cdot \prod_{ij} (1-a_{ij})^{1-a_{ij}}$$

$\blacksquare$

Note that this conjectured bound would be tight for the doubly stochastic matrix $\frac{1}{2} \cdot A_1$.

Organization: The organization of this paper is as follows: We discuss known combinatorial bounds for the permanent and their relation to our bounds in Section 2. We prove the lower bound in Section 3, and the upper bound in Sections 4 and 5.

2 Bounds for the permanent

2.1 Lower bounds

In general, the permanent of a nonnegative matrix may vanish. Hence, we need to impose additional constraints on the matrix to allow non-trivial lower bounds. Usually, the matrix is assumed to be *doubly stochastic*, that is to have row and column sums equal 1. In this case it is easy to see that the permanent has to be positive. The most famous bound for permanents is that of Egorychev [5] and Falikman [6], resolving the question of van der Waerden, and showing the permanent of a doubly stochastic matrix to be at least $\frac{n!}{n^n}$. This bound is tight and is attained on the matrix all of whose entries equal $1/n$.

If we impose additional constraints on the matrix, we may expect a stronger bound. The class $\Lambda(k,n)$ of integer matrices whose row and column sums equal k (adjacency matrices of k-regular bipartite graphs with multiple edges) was considered by Schrijver and Valiant [23]. Normalizing by k, one obtains a class of doubly stochastic matrices with entries of the form $\frac{m}{k}$ for integer m (and hence, with support of size at most k in each row and column). The authors conjectured the minimal permanent for this class to be at least $((k-1)/k)^{(k-1)n}$. This conjecture was proved in [24][2]. A more general bound from [24] will be of special interest to us: Let $B = (b_{ij})$ be a doubly stochastic matrix, and let $A = (b_{ij} \cdot (1 - b_{ij}))$. Then

$$Per(A) \geq \prod_{i,j=1}^{n} (1 - b_{ij}) \tag{8}$$

We observe, for future reference, that the matrix B is replaced by a new matrix A, obtained by applying a concave function $\phi(t) = t(1-t)$ entry-wise to A. For this new matrix, an explicit, efficiently computable, lower bound on the permanent is given.

All these bounds are very difficult technical results, some of them using advanced mathematical tools, such as the Alexandrov-Fenchel inequalities. Let us note that more general bounds (with easier proofs), implying all the results above, were given in [11], using the machinery of hyperbolic polynomials. The point we would like to make (for future comparison with the situation with upper bounds) is that the lower bounds for the permanent are hard to prove, but they are essentially optimal.

We now consider a more general notion than the permanent. For an $n \times n$ matrix A, and $1 \leq m \leq n$, let $Per_m(A)$ be the sum of permanents of all $m \times m$ submatrices of A. Note that if A is a $0,1$ matrix, the permanent counts the perfect matchings of the corresponding bipartite graph, while $Per_m(A)$ counts all the matchings with m edges. Friedland [9] stated a conjectured lower bound on Per_m for the class $\Lambda(k,n)$ of integer matrices[3]. This conjecture has significance in statistical physics and is a natural generalization of the Schrijver-Valiant conjecture. Partial results towards this conjecture were obtained in [10].

[2]Let us remark that the assumption on the rationality of the entries was removed in [11], making only the structure of the support matter.

[3]This lower bound is complicated, we will state it explicitly below.

Our results:

We restate our lower bound Theorem 1.2 here for the convenience of the reader:

Let $A = (a_{ij})_{i,j=1}^n$ be a nonnegative matrix and let $B = (b_{ij})_{i,j=1}^n$ be a doubly stochastic matrix. Then

$$Per(A) \geq \prod_{i,j=1}^{n} (1 - b_{ij})^{1-b_{ij}} \cdot \exp\left\{ - \sum_{i,j=1}^{n} b_{ij} \log \frac{b_{ij}}{a_{ij}} \right\}$$

We note that this lower bound is the first lower bound on the permanent which actually "looks at the matrix", that is depends explicitly on the entries of A, rather than on its support pattern.

Note that the bound (8) follows, by taking $A = (b_{ij} \cdot (1 - b_{ij}))$. Hence Theorem 1.2 is a generalization of (8). On the other hand, let us say that we view it as a corollary of (8), since it is proved by analysis of the first order optimality conditions on the RHS of the inequality above, viewed as a function on doubly stochastic matrices, and the key part of the analysis is applying (8).

The conjecture of Friedland. Let $\alpha(m, n, k) = \min_{A \in \Lambda(k,n)} Per_m(A)$. Think about m growing linearly in n and k being fixed[4]. Then $\alpha(m, n, k)$ is exponential in n, and we are interested in the exponent.

To be more precise, fix $p \in [0, 1]$ (this is the so called *limit dimer density*). Let $m(n) \leq n$ be an integer sequence with $\lim_{n \to \infty} \frac{m(n)}{n} = p$. Finally, let[5]

$$\beta(p, k) = \lim_{n \to \infty} \frac{1}{n} \log(\alpha(m(n), n, k))$$

The challenge is to find $\beta(p, k)$. S. Friedland had conjectured that, similarly to [24], one can replace the minimum in the definition of $\alpha(m, n, k)$ by an (explicitly computable) average over a natural distribution $\mu = \mu_{k,n}$ on $\Lambda(k, n)$ (see Section 3).

We show this conjecture to hold, deducing it from Theorem 1.2.

Theorem 2.1:

$$\beta(p, k) = \lim_{n \to \infty} \frac{1}{n} \log \left(\mathbb{E}_{\mu}(Per_{m(n)}(A)) \right)$$

Remark 2.2: Friedland's conjecture was proved, using the hyperbolic polynomials, in [10] for limit dimer densities of the form $p = \frac{k}{k+s}, s \in \mathbb{N}$. ▮

[4]The bounds below hold for any k, though.

[5]It follows from Theorem 2.1 that this definition is independent of the choice of the sequence $m(n)$ and that the limit exists.

2.2 Upper bounds

The notable upper bound for the permanents is due to Bregman [2], proving a conjecture of Minc. This is a bound for permanents of $0, 1$ matrices. For a $0, 1$ matrix A with r_i ones in the i^{th} row,

$$Per(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i} \tag{9}$$

To the best of our knowledge, there is no satisfying extension of this bound to general nonnegative matrices. We will now give a different view of (9), suggesting a natural way to extend it. Let $A = (a_{ij})$ be a stochastic matrix, whose values in the i^{th} row are either 0 or $1/r_i$. Let $B = (b_{ij})$ be a matrix with $b_{ij} = 0$ if $a_{ij} = 0$ and $b_{ij} = (1/r_i!)^{1/r_i}$ if $a_{ij} = 1/r_i$. Then: $Per(B) \leq 1$.

There is a natural construction of a function on the interval $[0, 1]$ taking $1/r$ to $(1/r!)^{1/r}$ for all integer r. This is the function $\phi_0(x) = \Gamma\left(\frac{1+x}{x}\right)^{-x}$.

Conjecture 2.3: ([22]) Let $A = (a_{ij})$ be a stochastic matrix, and let $B = (\phi_0(a_{ij}))$. Then $Per(B) \leq 1$. ∎

Unfortunately, we do not know how to prove this conjecture.

There is, however, a way to view it as a special (difficult) case in a general family of upper bounds for the permanent. The function $\phi_0(x) = \Gamma\left(\frac{1+x}{x}\right)^{-x}$ is a *concave* [25] increasing function taking $[0, 1]$ onto $[0, 1]$. We can ask for which concave functions ϕ of this form, Conjecture 2.3 holds. Note the similarity of this point of view with that of the bound (8). In both cases we apply a concave function entry-wise to the entries of a stochastic matrix and ask for an explicit efficiently computable upper (or lower) bound for the permanent of the obtained matrix.

Let ϕ be concave increasing function taking $[0, 1]$ onto $[0, 1]$. The function $\psi = \phi^{-1}$ is convex increasing taking $[0, 1]$ onto $[0, 1]$. It defines an *Orlicz norm* ([29]) $\|\cdot\|_\psi$ on $\mathbb{R}^n$ as follows: for $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$

$$\|v\|_\psi = s, \quad \text{where } s \text{ is such that } \sum_{i=1}^{n} \phi\left(\frac{|v_i|}{s}\right) = 1$$

Note that this is a generalization of the more familiar l_p norms. For $\psi(x) = x^p$, $\|\cdot\|_\psi = \|\cdot\|_p$.

If v is a stochastic vector, the vector $w = (\phi(v_1), \ldots, \phi(v_n))$ has $\|w\|_\psi = 1$. Thus, the question we are asking is: For which Orlicz norms $\|\cdot\|_\psi$, a matrix B whose rows are unit vectors in this norm has permanent at most 1. Using homogeneity of the norm and multilinearity of the permanent, we obtain an appealing form of the general family of upper bounds to consider: We want any nonnegative matrix B with rows $b_1, \ldots, b_n$ satisfy

$$Per(B) \leq \prod_{i=1}^{n} \|b_i\|_\psi \tag{10}$$

Our results: We prove (10) for a family of functions ψ. Theorem 1.3 follows as a corollary.

We note, that in strong contrast to the lower bounds case, our bounds are far from being optimal, and, in particular, are far from proving Conjecture 1.5 or Conjecture 2.3.

3 Proofs of the lower bounds

3.1 Proof of Theorem 1.2

Notation. We will denote by Ω_n the class of doubly stochastic $n \times n$ matrices. For a pair $P = (p_{ij}), Q = (q_{ij})$ of non-negative matrices, we let

$$CW(P,Q) = \sum_{i,j=1}^{n} (1-q_{ij}) \log(1-q_{ij}) - \sum_{i,j=1}^{n} q_{ij} \log\left(\frac{q_{ij}}{p_{ij}}\right)$$

Let P be a non-negative $n \times n$ matrix with positive permanent (which we may assume, without loss of generality). We will prove the theorem by showing

$$\log(Per(P)) \geq max_{Q \in \Omega_n} CW(P,Q)$$

Note that, by continuity, we may assume all the entries in P to be strictly positive. Then the functional $CW(P,Q)$ is bounded from above and continuous as function of Q on Ω_n. Therefore, the maximum is attained. Let $V \in \Omega_n$ be one of points at which it is attained.

We first isolate ones in the doubly-stochastic matrix V: up to rearrangement of the rows and columns, $V = \begin{pmatrix} I & 0 \\ 0 & T \end{pmatrix}$, where the doubly -stochastic matrix T does not have ones; and block-partition accordingly the matrix $P = \begin{pmatrix} P^{(1,1)} & P^{(1,2)} \\ P^{(2,1)} & P^{(2,2)} \end{pmatrix}$.

Note that $CW(P,V) = CW\left(P^{(2,2)}, T\right) + \sum_i \log\left(P_{i,i}^{(1,1)}\right)$.

Since $Per(P) \geq Per\left(P^{(1,1)}\right) \cdot Per\left(P^{(2,2)}\right) \geq \prod_i P_{i,i}^{(1,1)} \cdot Per\left(P^{(2,2)}\right)$, we only need to prove $\log\left(Per(P^{(2,2)})\right) \geq CW\left(P^{(2,2)}, T\right)$.

Let d be the dimension of matrices $P^{(2,2)}, T$. We express the local extremality conditions for T not on the full Ω_d but rather in the interior of the compact convex subset of doubly-stochastic $d \times d$ matrices supported on the support of $T = (t_{kl})$.

We first compute the partial derivatives (writing them out for general d-dimensional P, Q)

$$\frac{\partial}{\partial q_{ij}} CW(P,Q) = -2 - \log(1-q_{ij}) - \log(q_{ij}) + \log(p_{ij}) \quad 1 \leq i,j \leq d$$

By the first order optimality conditions for T, we get that there exists real numbers $\{\alpha_k\}, \{\beta_l\}$ such that

$$-2 - \log(1-t_{kl}) - \log(t_{kl}) + \log\left(P_{kl}^{(2,2)}\right) = \alpha_k + \beta_l;\ (k,l) \in Supp(T)$$

Which gives, for some positive numbers $\{a_k\}, \{b_l\}$ the following scaling:

$$P_{kl}^{(2,2)} = a_k b_l \cdot t_{kl} (1 - t_{kl})\,; \;\; (k,l) \in Supp(T)$$

Now, we can conclude the proof.

1. It follows from the definition of the support that (applying the inequality below entry-wise)

$$P^{(2,2)} \geq Diag\,(a_k) \cdot \widetilde{T} \cdot Diag(b_l); \text{ where } \; \widetilde{T}_{kl} = t_{kl} (1 - t_{kl})$$

2. It follows from doubly-stochasticity of T that

$$CW(P^{(2,2)}, T) = \sum \log (a_k) + \sum \log (b_l) + \sum_{(k,l) \in Supp(T)} \log (1 - t_{kl}) \tag{11}$$

Finally it follows from (11) and (8) that

$$\log \left(Per \left(Diag\,(a_k) \cdot \widetilde{T} \cdot Diag\,(b_l) \right) \right) \geq CW \left(P^{(2,2)}, T \right)$$

and therefore

$$\log \left(Per \left(P^{(2,2)} \right) \right) \geq \log \left(Per \left(Diag\,(a_k) \cdot \widetilde{T} \cdot Diag\,(b_l) \right) \right) \geq CW \left(P^{(2,2)}, T \right)$$

∎

3.2 Proof of Theorem 2.1

Let us first recall the following well known identity (see, for instance, [8]), expressing $Per_m(A)$ as a single permanent:

$$Per_m(A) = ((n-m)!)^{-2} \cdot Per(L), \quad L = \begin{pmatrix} A & J_{n,n-m} \\ J_{n,n-m}^T & 0 \end{pmatrix}$$

where $J_{n,n-m}$ is $n \times (n-m)$ matrix of all ones. If the matrix $A \in c \cdot \Omega_n$ (i.e. proportional to a doubly-stochastic matrix) then it is easy to scale the matrix L. In particular, if $A \in \Lambda(k,n)$ then

$$Per_m(A) = \frac{Per(K)}{a^m b^{2(n-m)} ((n-m)!)^2} \tag{12}$$

where $K \in \Omega_{2n-m}$ is defined as follows

$$K = \begin{pmatrix} a \cdot A & b \cdot J_{n,n-m} \\ (b \cdot J_{n,n-m})^T & 0 \end{pmatrix}$$

with $p = \frac{m}{n}$, $a = \frac{p}{k} = \frac{m}{kn}$, $b = \frac{1}{n}$.

We note that the identity (12) follows from the diagonal scaling:

$$K = \left(\sqrt{a} I_n \oplus \frac{b}{\sqrt{a}} I_{n-m}\right) \cdot L \cdot \left(\sqrt{a} I_n \oplus \frac{b}{\sqrt{a}} I_{n-m}\right)$$

To proceed with the proof, we will need the following simple claim, following from the convexity of $(1-x)\log(1-x)$.

Proposition 3.1: *Let $p_1, ..., p_k$ be non-negative numbers, with $0 \le p_i \le 1$ and $\sum_{i=1}^k p_i = s$. Then, setting $b = \frac{s}{k}$,*

$$\prod_{i=1}^{k} (1-p_i)^{1-p_i} \ge (1-b)^{k(1-b)}$$

Our main claim is:

Theorem 3.2: *Let $A \in \Lambda(k,n)$, Let $1 \le m \le n$ and let $p = \frac{m}{n}$. Then the following inequality holds*[6]*:*

$$Per_m(A) \ge \frac{(\frac{k-p}{k})^{n(k-p)} \cdot (1-n^{-1})^{(1-n^{-1})2n^2(1-p)}}{(\frac{p}{k})^{np} \cdot n^{-2n(1-p)} \cdot ((n(1-p))!)^2} \tag{13}$$

Proof: Apply the lower bound in (1) to the doubly-stochastic matrix K and use (12). If A is boolean then this already gives the inequality we need. In the non-boolean case an immediate application of Proposition 3.1 finishes the proof. ▮

Proof of Theorem 2.1.

First, we define the distribution μ on $\Lambda(k,n)$. Consider the following construction of a matrix $A \in \Lambda(k,n)$. For a permutation $\pi \in S_{kn}$, let $M = M_\pi$ be the standard representation of π as a $kn \times kn$ matrix of zeroes and ones. Now, view M in the natural way as a $k \times k$ block matrix $M = (M_{ij})$, where each block M_{ij} is an $n \times n$ matrix. Finally, set $A = A(\pi) = \sum_{i,j=1}^k M_{ij}$. The distribution μ is the one induced on $\Lambda(k,n)$ by the uniform distribution on S_{kn}.

We point out that the expectation $\mathbb{E}_\mu(Per_m(A))$ is known (see for instance [9], [10]). In particular, if $\lim_{n\to\infty} \frac{m(n)}{n} = p \in [0,1]$ then the following equality holds:

$$\lim_{n\to\infty} \frac{\log\left(\mathbb{E}_\mu\left(Per_{m(n)}(A)\right)\right)}{n} = p\log\left(\frac{k}{p}\right) - 2(1-p)\log(1-p) + (k-p)\log\left(1-\frac{p}{k}\right) \tag{14}$$

The claim of the theorem follows directly from (13), (14), and Stirling's formula. ▮

[6] Assuming, for typographic simplicity, all the relevant values on LHS to be integer.

4 Proofs of the upper bounds

Recall that we are interested in upper bounds of the form given in (10). We prove the following general claim.

Theorem 4.1: *Let ψ be a convex increasing thrice differentiable function taking $[0,1]$ onto $[0,1]$. Assume ψ has the following properties*

1. *The function $x \cdot \frac{\psi'(x)}{\psi(x)}$ is increasing.*
2. *The function $x \cdot \frac{\psi''(x)}{\psi'(x)}$ is increasing.*
3. $$\psi\left(e^{-r/e}\right) + \psi\left(r \cdot e^{-r/e}\right) \geq 1 \quad \text{for } 0 \leq r \leq 1$$

Then, for any nonnegative matrix B with rows $b_1, \ldots, b_n$ it holds that

$$Per(B) \leq \prod_{i=1}^{n} \|b_i\|_\psi$$

For this theorem to be useful, we need to provide examples of functions it applies to. We now give an example of a function ψ satisfying the conditions of the theorem. Let $a \approx 1.54$ be the unique root of the equation $\frac{1-\ln a}{a} = \frac{1}{e}$.

Lemma 4.2: *The function*

$$\psi_a(x) = 1 - (1-x) \cdot a^x$$

satisfies the conditions of Theorem 4.1.

We now show how to deduce Theorem 1.3 from Theorem 4.1, using the function ψ_a. We start with a technical lemma.

Lemma 4.3:

- *For any stochastic vector $x = (x_1, \ldots, x_n)$, the maximum of the entries of the vector $\left(\frac{x_j}{\prod_{k=1}^n (1-x_k)^{1-x_k}}\right)_{j=1}^n$ is at most $e^{1/e} \approx 1.44$.*
- *Let ψ_a be the function in Lemma 4.2. Then for any stochastic vector $x = (x_1, \ldots, x_n)$ holds*[7]

  $$\sum_{j=1}^{n} \psi_a\left(\frac{x_j}{C \cdot \prod_{k=1}^n (1-x_k)^{1-x_k}}\right) \leq 1$$

  for some $e^{1/e} \leq C \leq 2$.

[7]Note that by the first claim of the lemma, all the arguments of ψ in LHS are in the allowed range $[0,1]$.

Given the lemma, Theorem 1.3 follows immediately: In fact, by the definition of $\|\cdot\|_{\psi}$, we have for any stochastic vector x,

$$\|x\|_{\psi_a} \leq C \cdot \prod_{k=1}^{n} (1 - x_k)^{1-x_k}$$

Hence, by Theorem 4.1, for any stochastic matrix B, whose rows are stochastic vectors $b_1, \ldots, b_n$,

$$Per(B) \leq \prod_{i=1}^{n} \|b_i\|_{\psi_a} \leq C^n \cdot \prod_{i,j=1}^{n} (1 - b_{ij})^{1-b_{ij}}$$

giving Theorem 1.3.

The full proofs of the claims in this section are given in the next section.

5 Full proofs of the claims for the upper bound

5.1 Proof of Theorem 4.1

A word on notation. We denote by $\|x\|_{\psi}$ the norm of a vector x in $\mathbb{R}^k$, without stating k explicitly. Thus, we may and will compare $\|\cdot\|_{\psi}$-norms of vectors of different dimensions.

We denote by A_{ij} the submatrix of a matrix A obtained by removing the i^{th} row and the j^{th} column of A.

The proof is by induction on the dimension n. For $n = 1$ the claim holds since for a scalar $a \in \mathbb{R}$,

$$Per(a) = a = \|a\|_{\psi}$$

The second equality is due to the fact that $\psi(1) = 1$.

Assume the theorem holds for $n - 1$. The induction step from $n - 1$ to n is incorporated in the following lemma.

Lemma 5.1: *Let $\phi_* : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a scalar function defined by*

$$\phi_*(r) = \min_{y \in \mathbb{R}_+^{n-1}:\ \|y\|_{\psi} = 1} \|(y, r)\|_{\psi}$$

Assume ϕ_* satisfies the following functional inequality: For any $r_1, \ldots, r_n \in \mathbb{R}_+$

$$\prod_{k=1}^{n} \phi_* (r_k) \geq \sum_{k=1}^{n} r_k \tag{15}$$

Then, if the theorem holds for $n - 1$, it holds also for n.

Proof: of Lemma 5.1

Write the rows of the $n \times n$ matrix A as $a_k = (x_k, b_k)$, with $x_k \in \mathbb{R}^{n-1}$ and $b_k = a_{kn} \in \mathbb{R}$.

Clearly, if any of a_k is 0 the claim of the theorem holds. The other boundary case we need to treat separately is the case in which one of the vectors x_k is 0. Without loss of generality, assume $x_1 = 0$. Expanding the permanent with respect to the first row, and using the induction hypothesis for A_{1n}, we have

$$Per(A) = a_{1n} \cdot Per\left(A_{1n}\right) \leq a_{1n} \cdot \prod_{k=2}^{n} \|x_k\|_\psi \leq \prod_{k=1}^{n} \|a_k\|_\psi$$

establishing the theorem in this case.

Assume none of x_k is 0. Expanding the permanent of A with respect to the last column, and using the induction hypothesis, we have

$$Per(A) = \sum_{i=1}^{n} b_i \cdot Per\left(A_{in}\right) \leq \sum_{i=1}^{n} b_i \cdot \prod_{j \neq i} \|x_j\|_\psi = \prod_{j=1}^{n} \|x_j\|_\psi \cdot \sum_{i=1}^{n} \frac{b_i}{\|x_i\|_\psi}$$

Hence, to prove the theorem for A, we need to show

$$\sum_{i=1}^{n} \frac{b_i}{\|x_i\|_\psi} \leq \prod_{k=1}^{n} \frac{\|\left(x_k, b_k\right)\|_\psi}{\|x_k\|_\psi}$$

Let $r_k = b_k / \|x_k\|_\psi$, $y_k = x_k / \|x_k\|_\psi$. Then the inequality translates to

$$\prod_{k=1}^{n} \|\left(y_k, r_k\right)\|_\psi \geq \sum_{i=1}^{n} r_i$$

which follows from (15), since $\|y_k\|_\psi = 1$, and hence $\|\left(y_k, r_k\right)\|_\psi \geq \phi_*\left(r_k\right)$. ∎

It remains to prove (15).

First, we observe that the function ϕ_* has an explicit form.

Lemma 5.2:

$$\phi_*(r) = \|(1, r)\|_\psi$$

Proof: (**of Lemma 5.2**)

We may assume $r > 0$, otherwise the claim of the lemma holds trivially.

Consider the optimization problem of minimizing $\|(y, r)\|_\psi$ for y in the unit sphere of the norm in $\mathbb{R}^{n-1}$. Note that the minimum is attained, since we are looking for the minimum of a continuous function in a compact set.

Let y_* be a point of minimum. We will show y_* to be a unit vector, implying the claim of the lemma.

First step: We show y_* to be constant on its support.

Since $\|(y_*, r)\|_\psi = \phi_*(r)$, we have

$$\left\| \left(\frac{y_*}{\phi_*(r)}, \frac{r}{\phi_*(r)} \right) \right\|_\psi = 1 \leq \left\| \left(\frac{y}{\phi_*(r)}, \frac{r}{\phi_*(r)} \right) \right\|_\psi$$

for any y of norm 1. Therefore $z_* = \frac{y_*}{\phi_*(r)}$ is a point of minimum of $\sum_{i=1}^{n-1} \psi(z_i)$ in the domain $D = \{z : \ \|z\|_\psi = 1/\phi_*(r)\}$.

Consider this new optimization problem. Set $a = \phi_*(r)$ for typographic convenience. Note $a > 1$, since, by assumption, $r > 0$. Then

$$D = \left\{ z \in R_+^{n-1} : \ \sum_{i=1}^{n-1} \psi(az_i) = 1 \right\}$$

We know that z_* is a point of minimum of the target function $\sum_{i=1}^{n-1} \psi(z_i)$ on D.

Let $S = S(z_*)$ be the support of z_*. The first order optimality conditions for z_* imply that there exists a constant $\lambda \in \mathbb{R}$ such that for any $i \in S$,

$$\frac{\psi'(z_i)}{\psi'(az_i)} = \lambda \cdot a \tag{16}$$

We would like to deduce from this that z_* (and hence also y_*) is constant on its support S.

Let $\eta(x) = \ln \psi'(e^x)$. We claim that η is strictly convex on $(-\infty, 0]$. In fact, $\eta'(x) = \frac{e^x \psi''(e^x)}{\psi'(e^x)}$, which is strictly increasing in x, by the second assumption of the theorem.

Note that $\psi'(x) = \exp\{\eta(\ln x)\}$. Therefore (16) is equivalent to

$$\eta(\ln(z_i)) - \eta(\ln(z_i) + \ln(a)) = \ln(\lambda \cdot a)$$

And this can't hold for different values of z_i if η is strictly convex. This shows z_* is constant on S, completing the first step.

Second step: $|S| = 1$.

Let $|S| = k$, for some $1 \leq k \leq n-1$.

Since $\sum_{i \in S} \psi(a \cdot (z_*)_i) = 1$ and z_* is constant on S, we have for all $i \in S$, $(z_*)_i = (1/a) \cdot \psi^{-1}(1/k)$. Therefore

$$\sum_{i=1}^{n-1} \psi((z_*)_i) = k \cdot \psi\left(\frac{\psi^{-1}(1/k)}{a}\right) \tag{17}$$

Consider the function $f(x) = (1/x) \cdot \psi\left(\frac{\psi^{-1}(x)}{a}\right)$. We will show this function to decrease on the interval $[0,1]$. This would imply the minimum over k of LHS of (17) is attained at $k=1$, completing this step.

Taking the first derivative, and denoting $\alpha = \psi^{-1}$, we need to verify for $x \in (0,1)$

$$0 > f'(x) = -\frac{1}{x^2} \cdot \psi\left(\frac{\alpha(x)}{a}\right) + \frac{1}{x} \cdot \psi'\left(\frac{\alpha(x)}{a}\right) \cdot \frac{\alpha'(x)}{a}$$

That is,

$$\psi\left(\frac{\alpha(x)}{a}\right) > \frac{x}{a} \cdot \psi'\left(\frac{\alpha(x)}{a}\right) \cdot \alpha'(x)$$

$$\psi\left(\frac{\alpha(x)}{a}\right) \cdot \psi'(\alpha(x)) > \frac{x}{a} \cdot \psi'\left(\frac{\alpha(x)}{a}\right)$$

Since $x = \psi(\alpha(x))$, we want to show

$$\frac{\psi'(\alpha(x))}{\psi(\alpha(x))} > \frac{1}{a} \cdot \frac{\psi'\left(\frac{\alpha(x)}{a}\right)}{\psi\left(\frac{\alpha(x)}{a}\right)} \iff \alpha(x) \cdot \frac{\psi'(\alpha(x))}{\psi(\alpha(x))} > \frac{\alpha(x)}{a} \cdot \frac{\psi'\left(\frac{\alpha(x)}{a}\right)}{\psi\left(\frac{\alpha(x)}{a}\right)}$$

That is, it suffices to show that $y \cdot \frac{\psi'(y)}{\psi(y)}$ increases in y, and this is true by the first assumption of the theorem.

This completes the second step and the proof of Lemma 5.2.

∎

As the next step towards the proof of (15), we give a sufficient condition for a function $g : \mathbb{R}_+ \to \mathbb{R}_+$ to satisfy the functional inequality stated in (15) for ϕ_*.

Lemma 5.3: *If*

$$g(x) \geq \begin{cases} e^{x/e} & for \quad 0 \leq x \leq e \\ x & otherwise \end{cases}$$

then $\prod_{k=1}^{n} g(r_k) \geq \sum_{k=1}^{n} r_k$.

Proof: Let $0 \leq r_1 \leq r_2 \leq \ldots \leq r_n$ be given, and assume $r_k < e$, $r_{k+1} \geq e$.

First assume $k < n$. Write $y = \sum_{i=1}^{k} r_i$, $z = \sum_{j=k+1}^{n} r_j$. Clearly, $z \geq e$. Note that, by assumption,

$$\prod_{j=k+1}^{n} g(r_j) \geq \prod_{j=k+1}^{n} r_j \geq \sum_{j=k+1}^{n} r_j = z$$

We have

$$\prod_{i=1}^{n} g(r_i) = \prod_{i=1}^{k} g(r_i) \cdot \prod_{j=k+1}^{n} g(r_j) \geq e^{1/e \cdot \sum_{i=1}^{k} r_i} \cdot z = e^{y/e} \cdot z$$

It remains to show $e^{y/e} \cdot z \geq y + z$ for $z \geq e$. Since $e^x \geq x + 1$, we have

$$e^{y/e} \geq y/e + 1 \geq \frac{y+z}{z}$$

and we are done in this case.

The other case to consider is $k = n$. Write $y = \sum_{i=1}^{k} r_i$. In this case we need to show $e^{y/e} \geq y$ for all $y \geq 0$. This again follows from the inequality $e^x \geq x + 1$, substituting $x = y/e - 1$. ▮

To prove (15) and complete the proof of the theorem, it remains to verify $\phi_*(r) = \|(1,r)\|_\psi$ satisfies the assumptions of Lemma 5.3. First, clearly,

$$\phi_*(r) \geq \|r\|_\psi = r$$

Next, $\phi_*(r) \geq e^{r/e}$ iff

$$\psi\left(e^{-r/e}\right) + \psi\left(r \cdot e^{-r/e}\right) \geq 1 \tag{18}$$

So we need to verify this for $0 \leq r \leq e$.

We now claim that we may reduce the problem to a subinterval.

Lemma 5.4: *Let ψ be an increasing differentiable convex function, taking $[0,1]$ to itself. If $\psi\left(e^{-r/e}\right) + \psi\left(r \cdot e^{-r/e}\right) \geq 1$ on $[0,1]$, then this also holds for $[0,e]$.*

Observe that the third assumption of the theorem is that (18) holds for $r \in [0,1]$. Thus, proving the lemma will complete the proof of the theorem.

Proof: Set

$$h(r) = \psi\left(e^{-r/e}\right) + \psi\left(r \cdot e^{-r/e}\right)$$

Then

$$h'(r) = \left(e^{-r/e} - \frac{1}{e} r e^{-r/e}\right) \cdot \psi'\left(re^{-r/e}\right) - \frac{1}{e} e^{-r/e} \cdot \psi'\left(e^{-r/e}\right)$$

First, we claim that h' is nonnegative on $[1, e-1]$. In fact, on this interval $re^{-r/e} \geq e^{-r/e}$. Consequently, by convexity of ψ, $\psi'\left(re^{-r/e}\right) \geq \psi'\left(e^{-r/e}\right)$. Hence

$$h'(r) \geq \psi'\left(e^{-r/e}\right) \cdot e^{-r/e} \cdot \left(1 - \frac{r+1}{e}\right) \geq 0$$

Next, we claim that $h(e-r) \geq h(r)$ for $0 \leq r \leq 1$. We need to show that

$$\psi\left((e-r)\cdot e^{-(e-r)/e}\right) + \psi\left(e^{-(e-r)/e}\right) \geq \psi\left(e^{-r/e}\right) + \psi\left(r \cdot e^{-r/e}\right)$$

Let a, b be the arguments on LHS, and c, d on RHS. Note $a \geq b$ and $c \geq d$. Since ψ is convex and increasing, it will suffice to show $a+b \geq c+d$ and $a \geq c$ (this would imply (a,b) majorizes (c,d)).

- We argue $a+b \geq c+d$. Let $f(x) = (x+1)e^{-x/e}$, and let $g(x) = f(e-x)$. We want to show $g(x) \geq f(x)$ for $0 \leq x \leq 1$. Note that f is increasing on $[0, e-1]$ and decreasing on $[e-1, e]$, so both f and g are increasing on $[0,1]$. First, we argue $f' \geq g'$. In fact, we have

  $$f'(x) = \frac{1}{e}\cdot((e-1)-x)\cdot e^{-x/e} \geq g'(x) = \frac{1}{e}\cdot(1-x)\cdot e^{-(e-x)/e}$$

  So, it would suffice to check $g(1) \geq f(1)$ which, after simplification, is the same as $e^{1/e} \geq 2^{1/2}$. And this is true.

- We argue $a \geq c$, that is $(e-r)\cdot e^{-(e-r)/e} \geq e^{-r/e}$ on $[0,1]$. Let $g(x)$ be the first function, and $f(x)$ the second. Note that $f(0) = g(0) = 1$. Hence, it suffices to prove $f' \leq g'$. We have $f'(x) = -1/e \cdot e^{-x/e}$ and $g'(x) = -e^{-(e-x)/e} + \frac{e-x}{e}\cdot e^{-(e-x)/e}$. Therefore

  $$g'(x) - f'(x) = \frac{1}{e}\cdot\left((e-x)\cdot e^{-(e-x)/e} + e^{-x/e} - e\cdot e^{-(e-x)/e}\right) =$$

  $$\frac{1}{e}\cdot\left(e^{-x/e} - x\cdot e^{-(e-x)/e}\right) \geq 0$$

∎

5.2 Proof of Lemma 4.2

We will prove the lemma in greater generality, that is for all functions $\psi = \psi_a$, with $\frac{1}{e} \leq \frac{1-\ln a}{a} < 1$.

First, we compute the first three derivatives of ψ.

$$\psi'(x) = (1-(1-x)\cdot\ln a)\cdot a^x$$

$$\psi''(x) = \ln a\cdot(2-(1-x)\cdot\ln a)\cdot a^x$$

$$\psi'''(x) = \ln^2 a\cdot(3-(1-x)\cdot\ln a)\cdot a^x$$

We now prove the required properties of ψ.

1. For $1 < a < e$, the function ψ is increasing strictly convex taking $[0,1]$ to $[0,1]$. In fact, by observation, $\psi' > 0$ for $0 \le x \le 1$ and $\psi'' > 0$ for $0 \le x \le 1$.

2. The function $x \cdot \frac{\psi'(x)}{\psi(x)}$ is strictly increasing for $1 < a < \sqrt{e}$.[8]
   It suffices to show for $0 < x < 1$

   $$\left(\psi' + x\psi''\right) \cdot \psi > x \left(\psi'\right)^2$$

   For typographic convenience, write $b = \ln a$. Substituting the expressions for ψ and its derivatives, and introducing notation

   $$P(x) = b^2x^2 + \left(2b - 2b^2\right)x + (1-b)^2, \qquad Q(x) = b^2x^2 + \left(3b - b^2\right)x + (1-b),$$

   we need to verify

   $$Q(x) \cdot \Big(1 - (1-x) \cdot e^{bx}\Big) > xP(x) \cdot e^{bx}$$

   Observe that Q is strictly positive on $(0,1)$. Rearranging, we need to show

   $$e^{-bx} > x \cdot \frac{P(x)}{Q(x)} + (1-x) = 1 - x \cdot \frac{Q(x) - P(x)}{Q(x)}$$

   Since $e^{-bx} > 1 - bx$ on $(0,1)$, it suffices to show $(Q-P)/Q \ge b$, that is $(1-b) \cdot Q \ge P$. And this is directly verifiable, for $x \in (0,1)$ and $b \in (0,1/2)$.

3. The function $x \cdot \frac{\psi''(x)}{\psi'(x)}$ is strictly increasing for $1 < a < \sqrt{e}$.
   This is true iff

   $$\left(\psi''(x) + x\psi'''(x)\right) \cdot \psi'(x) > x \cdot \left(\psi''(x)\right)^2$$

   Since $\psi''' > 0$, it suffices to prove

   $$\psi''(x) \cdot \psi'(x) \ge x \cdot \left(\psi''(x)\right)^2 \quad \iff \quad x \cdot \psi''(x) \le \psi'(x)$$

   Substituting the expressions for the derivatives of ψ and simplifying, we need to verify

   $$bx(2 - (1-x)b) \le 1 - (1-x)b$$

   This is a quadratic inequality in x. For $0 < b < 1/2$, the interval between the roots of this quadratic is easily seen to contain $[0,1]$, and we are done.

[8] It is easy to check that all a for which $\frac{1}{e} \le \frac{1 - \ln a}{a} < 1$ lie in this interval.

4. $$\psi\left(e^{-r/e}\right)+\psi\left(r\cdot e^{-r/e}\right)\geq 1 \quad \text{for } 0\leq r\leq 1$$

As in the proof of Lemma 5.4, we set

$$h(r)=\psi\left(e^{-r/e}\right)+\psi\left(r\cdot e^{-r/e}\right)$$

Hence

$$h'(r)=\left(e^{-r/e}-\frac{1}{e}re^{-r/e}\right)\cdot\psi'\left(re^{-r/e}\right)-\frac{1}{e}e^{-r/e}\cdot\psi'\left(e^{-r/e}\right)$$

Observe $h(0)=1$. Hence, it suffices to prove $h'\geq 0$ on $[0,1]$. Equivalently, for $0\leq r\leq 1$,

$$\frac{\psi'\left(r\cdot e^{-r/e}\right)}{\psi'\left(e^{-r/e}\right)}\geq\frac{1}{e-r}$$

Set $y=e^{-r/e}$. Clearly $e^{-1/e}\leq y\leq 1$. We will show a stronger statement

$$\frac{\psi'\left(ry\right)}{\psi'\left(y\right)}\geq\frac{1}{e-r}$$

for all y in the range. Similarly to the argument in the first step in the proof of Lemma 5.2, $\ln\left(\psi'\left(e^x\right)\right)$ is convex in x, which implies the LHS is decreasing in y, so it suffices to prove the inequality for $y=1$. Substituting the expression for ψ' and again writing b for $\ln a$, we need to verify

$$(e-r)\cdot(1-(1-r)b)\geq e^{b(1-r)},$$

for $0\leq r\leq 1$. At $r=0$, we need to check $e\geq e^b/(1-b)=a/(1-\ln a)$, which is satisfied with equality, by the assumption. Clearly, RHS decreases in r. By a direct calculation, the derivative of LHS is positive, that is LHS is increasing, completing the proof.

5.3 Proof of Lemma 4.3

For the first claim, we need a technical lemma.

Lemma 5.5: *Let $x=(x_1,\ldots,x_n)$ be a stochastic vector. Let $y=x_1$. Then*

$$\prod_{k=1}^{n}\left(1-x_k\right)^{1-x_k}\geq\frac{(1-y)^{1-y}}{e^{1-y}}$$

Proof: We need to show

$$\prod_{k=2}^{n}\left(1-x_k\right)^{1-x_k}\geq e^{y-1}\quad\iff\quad\sum_{k=2}^{n}\left(1-x_k\right)\ln\left(1-x_k\right)\geq y-1$$

for nonnegative $x_2, \ldots, x_n$ summing to $a := 1-y$. Let x_* be minimizer of $f(x_2, \ldots, x_n) = \sum_{k=2}^n (1-x_k)\ln(1-x_k)$ on this domain. Let S be the support of x_*. The first order regularity conditions state the existence of a constant λ such that

$$\ln(1-(x_*)_k) = \lambda$$

for all $k \in S$. This means that $(x_*)_k$ are constant on S.

Let $s = |S|$. Then $f(x_*) = (s-a)\ln\left(\frac{s-a}{s}\right)$. It remains to argue

$$(s-a)\ln\left(\frac{s-a}{s}\right) \geq -a,$$

for all integer $s \geq 1$. In fact, the function $g(s) = (s-a)\ln\left(\frac{s-a}{s}\right)$ of the real variable s is non-increasing on $[1,\infty)$, since $g'(s) = \ln(1-a/s) + a/s \leq 0$. And it is easy to see that $g(s)$ tends to $-a$ as $s \to \infty$.

▮

This means that

$$\frac{x_1}{\prod_{k=1}^n (1-x_k)^{1-x_k}} \leq \frac{ye^{1-y}}{(1-y)^{1-y}}$$

The following lemma concludes the proof of the first claim of Lemma 4.3.

Lemma 5.6: *The function* $f(y) = \frac{ye^{1-y}}{(1-y)^{1-y}}$ *on* $[0,1]$ *is upperbounded by* $e^{1/e}$.

Proof: The maximum ye^{1-y} on $[0,1]$ is 1 and the minimum of $(1-y)^{1-y}$ on $[0,1]$ is $e^{-1/e}$.

▮

We move to the second claim of Lemma 4.3, repeating its claim for convenience. Let ψ be the function in Lemma 4.2. Then for any stochastic vector $x = (x_1, \ldots, x_n)$ holds

$$\sum_{j=1}^n \psi\left(\frac{x_j}{2 \cdot \prod_{k=1}^n (1-x_k)^{1-x_k}}\right) \leq 1$$

The proof contains two steps, given in the following lemmas.

Lemma 5.7: *Let a stochastic vector* $x = (x_1, \ldots, x_n)$ *be given, and let* $y = \max_i x_i$ *be its maximal coordinate. Then, for any convex increasing function* ψ *taking* $[0,1]$ *to itself, and for any constant* $C \geq e^{1/e}$ *it holds that*

$$\sum_{j=1}^n \psi\left(\frac{x_j}{C \cdot \prod_{k=1}^n (1-x_k)^{1-x_k}}\right) \leq \frac{1}{y} \cdot \psi\left(\frac{ye^{1-y}}{C \cdot (1-y)^{1-y}}\right) \tag{19}$$

Lemma 5.8: *Let ψ be the function in Lemma 4.2. Then*

$$\frac{1}{y} \cdot \psi \left(\frac{ye^{1-y}}{2 \cdot (1-y)^{1-y}} \right) \leq 1$$

for $0 < y \leq 1$.

It remains to prove the lemmas.

Proof of Lemma 5.7

Let $x = (x_1, \ldots, x_n)$ be a stochastic vector with maximal entry y. By Lemmas 5.5 and 5.6, all the arguments of ψ in the LHS of (19) are upperbounded by $m = \frac{ye^{1-y}}{C \cdot (1-y)^{1-y}} \leq 1$ and their sum is at most $\frac{e^{1-y}}{C \cdot (1-y)^{1-y}}$. Since ψ is convex and increasing, the maximum of LHS under these constraints is attained when $s = \lfloor 1/y \rfloor$ of these arguments equal m and the remaining non-zero one equals $\frac{e^{1-y}}{C \cdot (1-y)^{1-y}} - s \cdot m$, which gives

$$\sum_{j=1}^{n} \psi \left(\frac{x_j}{C \cdot \prod_{k=1}^{n} (1-x_k)^{1-x_k}} \right) \leq s \cdot \psi \left(\frac{ye^{1-y}}{C \cdot (1-y)^{1-y}} \right) + \psi \left(\frac{(1-sy)e^{1-y}}{C \cdot (1-y)^{1-y}} \right) \tag{20}$$

Let $t = (1-sy)/y$. Since $s = \lfloor 1/y \rfloor$, we have $t \leq 1$. Since ψ is convex, increasing, and $\psi(0) = 0$, we have that $\psi(t \cdot x) \leq t \cdot \psi(x)$ for any $t \leq 1$, $0 \leq x \leq 1$. Therefore

$$\psi \left(\frac{(1-sy)e^{1-y}}{C \cdot (1-y)^{1-y}} \right) \leq \frac{1-sy}{y} \cdot \psi \left(\frac{ye^{1-y}}{C \cdot (1-y)^{1-y}} \right)$$

and the RHS of (20) is at most

$$\frac{1}{y} \cdot \psi \left(\frac{ye^{1-y}}{C \cdot (1-y)^{1-y}} \right),$$

completing the proof of the lemma. ∎

Proof of Lemma 5.7

We have $\psi(x) = 1 - (1-x) \cdot a^x$, where $a \approx 1.54$ is determined by the identity $(1 - \ln a)/a = \frac{1}{e}$.
Set $f(y) = \frac{1}{y} \cdot \psi \left(\frac{ye^{1-y}}{2 \cdot (1-y)^{1-y}} \right)$.

First, we claim that the maximum of f on $[0, 1]$ is attained for $y \leq y_0 = 0.51$.

In fact, setting $A(y) = \frac{ye^{1-y}}{C \cdot (1-y)^{1-y}}$, we have $f(y) = \psi(A(y))/y$ and $f'(y) \leq 0$ iff

$$\frac{A(y) \cdot \psi'(A(y))}{\psi(A(y))} \leq \frac{1}{1 + y \ln(1-y)}$$

Recall that the function $(x \cdot \psi'(x))/\psi(x)$ is increasing on $[0, 1]$. Therefore the maximum of LHS is at most $a = \psi'(1)$. We claim that RHS is at least that, for $0.51 \leq y \leq 1$. In fact, this is easy to see that RHS is an increasing function of y. Computing this function at $y_0 = 0.51$, we see that it is greater than a, and we are done. Hence $f'(0) < 0$ for $y_0 \le y \le 1$, and the maximum of f is attained outside this interval.

It remains to show $f(y) \le 1$ on $I = [0, y_0]$. Equivalently,

$$1 - (1 - A(y)) \cdot a^{A(y)} \le y \quad \Longleftrightarrow \quad \ln(a) \cdot A(y) + \ln(1 - A(y)) \ge \ln(1 - y) \tag{21}$$

The function $\ln(a) \cdot x + \ln(1-x)$ is decreasing in x, and therefore we decrease LHS by substituting a larger value for $A(y)$. We prove (21) for $y \in I$ by covering I with several intervals, and, in each interval, replacing $A(y)$ by a different linear function which majorizes it in this interval.

First, we need a technical lemma.

Lemma 5.9: *The function* $g(y) = (1/y) \cdot \ln\left(\frac{1-ry}{1-y}\right)$ *decreases on* $(0, 1/r)$ *for any* $r > 1$.

Proof: We will show $g' \le 0$. Computing the derivative and simplifying, we need to show

$$(1 - ry) \cdot \ln\left(\frac{1-y}{1-ry}\right) \le (r-1) \cdot \frac{y}{1-y}$$

Since $\ln(1+x) \le x$, we may replace the logarithm on LHS with $\frac{(r-1)y}{1-ry}$, leading to a trivially true inequality. ∎

We now prove (21) in several steps. Observe, for future use, that the function $h(y) = e^{1-y}/(1-y)^{1-y}$ decreases on $[0, 1]$.

- The maximum of $h(y)$ on $[0, 1]$ is $e = h(0)$. Set $r = e/2$. Then $A(y) \le ry$ for $y \in [0, 1]$. (Note $ry < 1$ for $y \in I$.) Hence, if we show

  $$\ln(a) \cdot ry + \ln(1 - ry) \ge \ln(1 - y)$$

  for y in some interval I_1, it would imply (21) in this interval. Rearranging, we need to show

  $$\frac{1}{y} \cdot \ln\left(\frac{1-ry}{1-y}\right) \ge \ln(a) \cdot r$$

  By the lemma, LHS is a decreasing function of y, hence it suffices to check this inequality at the right endpoint of I_1. It holds at $y = 0.3$, and therefore we may take $I_1 = [0, 0.3]$ and (21) holds in this interval.

- It remains to check (21) in $[0.3, 0.51]$. In this interval, the maximum of h equals $h(0.3)$ and therefore $A(y) \le ry$ for $r = h(0.3)/2$. Repeating the same argument, with the new value of r, we extend the validity of (21) to $I_2 = [0, 0.4]$. Reiterating, with new values of r, we get, in two more steps, to progressively larger intervals $[0, 0.48]$, and, finally, to $[0, 0.51]$.

The lemma is proved. ∎

References

[1] A. I. Barvinok, *Polynomial Time Algorithms to Approximate Permanents and Mixed Discriminants Within a Simply Exponential Factor.* Random Struct. Algorithms 14(1): 29-61 (1999)

[2] L. M. Bregman, *Certain properties of nonnegative matrices and their permanents*, Soviet Math. Dokl. 14, 945-949, 1973.

[3] A. Z. Broder, *How hard is it to marry at random? (On the approximation of the permanent)*, in Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC), ACM, New York, 1986, pp. 50-58. (Erratum in Proceedings of the 20th Annual ACM Symposium on Theory of Computing, 1988, pp. 551.)

[4] Michael Chertkov, Lukas Kroc, Massimo Vergassola, *Belief Propagation and Beyond for Particle Tracking*, http://arxiv.org/abs/0806.1199, 2008.

[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] A. Ferber, M. Krivelevich and B. Sudakov, *Counting and packing Hamilton cycles in dense graphs and oriented graphs*, preprint.

[8] S. Friedland, *A proof of a generalized van der Waerden conjecture on permanents*, Linear and Multilinear Algebra 11 (1982), no. 2, 107-120.

[9] S. Friedland, E. Krop, P. H. Lundow, K. Markstrm, *Validations of the Asymptotic Matching Conjectures*, arxiv preprint arXiv:math/0603001, 2006.

[10] S. Friedland and L. Gurvits, *Lower Bounds for Partial Matchings in Regular Bipartite Graphs and Applications to the Monomer-Dimer Entropy*, Combinatorics, Probability and Computing, 2008.

[11] L. Gurvits, *Van der Waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all*, Electronic Journal of Combinatorics 15 (2008).

[12] L. Gurvits, *A polynomial-time algorithm to approximate the mixed volume within a simply exponential factor.*, Discrete Comput. Geom. 41 (2009), no. 4, 533-555.

[13] B. Huang and T. Jebara, *Approximating the Permanent with Belief Propagation.*, New York Academy of Sciences Machine Learning Symposium 2007. Poster and abstract.

[14] M. Jerrum, A. Sinclair, and E. Vigoda, *A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries.*, J. ACM 51(4): 671-697 (2004)

[15] Heilmann, Ole J.; Lieb, Elliott H. *Theory of monomer-dimer systems.*, Comm. Math. Phys. 25 (1972), 190-232.

[16] N. Linial, A. Samorodnitsky, A.Wigderson, *A Deterministic Strongly Polynomial Algorithm for Matrix Scaling and Approximate Permanents.*, Combinatorica 20(4): 545-568 (2000)

[17] M. Laurent, A. Schrijver, *On Leonid Gurvits's proof for permanents*, American Mathematical Monthly; 117(10):903-911.

[18] L. Lovasz and M. D. Plummer, **Matching Theory**, North Holland, Amsterdam 1986.

[19] Adam Marcus, Daniel A. Spielman, Nikhil Srivastava, *Interlacing Families I: Bipartite Ramanujan Graphs of All Degrees*, arXiv:1304.4132 [math.CO], 2013.

[20] H. Minc, **Permanents**, Encyclopeadia of Mathematics and its Applications, vol. 6, Addison-Wesley, Reading, Mass., 1978.

[21] M. Rudelson, O. Zeitouni, *Singular values of Gaussian matrices and permanent estimators*, arXiv:1301.6268, 2013.

[22] A. Samorodnitsky, *An upper bound for permanents of nonnegative matrices.*, J. Comb. Theory, Ser. A 115(2): 279-292 (2008)

[23] A. Schrijver and W.G.Valiant, *On lower bounds for permanents*, Indagationes Mathematicae 42, pp. 425-427, 1980.

[24] A. Schrijver, *Counting 1-factors in regular bipartite graphs*, Journal of Combinatorial Theory, Series B 72 (1998) 122-135.

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

[26] L. G. Valiant, *The complexity of computing the permanent*, Theoretical Computer Science, 8(2), 189-201, 1979.

[27] P.O. Vontobel, *The Bethe permanent of a non-negative matrix*, in Proc. of Commu- nication, Control, and Computing (Allerton), 2010.

[28] A. Wigderson, personal communication.

[29] A. Zygmund, **Trigonometric series**, Volume 1 and 2 combined (3rd ed.), Cambridge University Press, 2002.