Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 169 (2011)

Unleashing the power of Schrijver's permanental inequality with the help of the Bethe Approximation.

Leonid Gurvits *

February 29, 2012

Abstract

Let $A \in \Omega_n$ be doubly-stochastic $n \times n$ matrix. Alexander Schrijver proved in 1998 the following remarkable inequality

$$per(\widetilde{A}) \geq \prod_{1 \leq i,j \leq n} (1 - A(i,j)); \widetilde{A}(i,j) =: A(i,j)(1 - A(i,j)), 1 \leq i,j \leq n \tag{1}$$

We prove in this paper the following generalization (or just clever reformulation) of (1):
For all pairs of $n \times n$ matrices (P, Q), where P is nonnegative and Q is doubly-stochastic

$$\log(per(P)) \geq \sum_{1 \leq i,j \leq n} \log(1-Q(i,j))(1-Q(i,j)) - \sum_{1 \leq i,j \leq n} Q(i,j) \log\left(\frac{Q(i,j)}{P(i,j)}\right) \tag{2}$$

The main co rollary of (2) is the following inequality for doubly-stochastic matrices:

$$\frac{per(A)}{F(A)} \geq 1; F(A) =: \prod_{1 \leq i,j \leq n} (1 - A(i,j))^{1-A(i,j)}.$$

We use this inequality to prove Friedland's conjecture on monomer-dimer entropy, so called *Asymptotic Lower Matching Conjecture*
We present explicit doubly-stochastic $n \times n$ matrices A with the ratio $\frac{per(A)}{F(A)} = \sqrt{2}^n$ and conjecture that

$$\max_{A \in \Omega_n} \frac{per(A)}{F(A)} \approx \left(\sqrt{2}\right)^n.$$

If true, it would imply a deterministic poly-time algorithm to approximate the permanent of $n \times n$ nonnegative matrices within the relative factor $\left(\sqrt{2}\right)^n$.

*gurvits@lanl.gov. Los Alamos National Laboratory, Los Alamos, NM.

ISSN 1433-8092

1 The permanent

Recall that a $n \times n$ matrix A is called doubly stochastic if it is nonnegative entry-wise and its every column and row sum to one. The set of $n \times n$ doubly stochastic matrices is denoted by Ω_n. The set of $n \times n$ of row stochastic(i.e. when every row sum to one) is denoted by RS_n, the set of column stochastic(i.e. when every column sum to one) is denoted by CS_n.

Let $\Lambda(k, n)$ denote the set of $n \times n$ matrices with nonnegative integer entries and row and column sums all equal to k . We define the following subset of rational doubly stochastic matrices: $\Omega_{k,n} = \{k^{-1}A : A \in \Lambda(k, n)\}$.

Recall that the permanent of a square matrix A is defined by

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

The following inequality was conjectured by B.l. van der Waerden in 1926 and proved independently in 1981 by D.L. Falikman [13] and G.P. Egorychev [12]:

$$\min_{A \in \Omega_n} per(A) = \frac{n!}{n^n} =: vdw(n). \tag{3}$$

1.1 Schrijver-Valiant Conjecture and (main) Schrijver's permanental inequality

Define

$$\lambda(k, n) = \min\{per(A) : A \in \Omega_{k,n}\} = k^{-n} \min\{per(A) : A \in \Lambda(k, n)\};$$

$\theta(k) = \lim_{n\to\infty}(\lambda(k, n))^{\frac{1}{n}}$.

It was proved in [2] (also earlier in [1]) that, using our notations, $\theta(k) \leq G(k) =: (\frac{k-1}{k})^{k-1}$ and conjectured that $\theta(k) = G(k)$. Though the case of $k = 3$ was proved by M. Voorhoeve in 1979 [20] , this conjecture was settled only in 1998 [3] (17 years after the published proof of the Van der Waerden Conjecture). The main result of [3] (as many people, including myself, wrongly thought) is the remarkable **(Schrijver-bound)** :

$$\min\{per(A) : A \in \Omega_{k,n}\} \geq \left(\frac{k-1}{k}\right)^{(k-1)n} \tag{4}$$

The bound (4) is a corollary of another inequality for doubly-stochastic matrices:

$$per(\widetilde{A}) \geq \prod_{1\leq i,j\leq n} (1 - A(i, j)); A \in \Omega_n; \widetilde{A}(i, j) =: A(i, j)(1 - A(i, j)), 1 \leq i, j \leq n. \tag{5}$$

The proof of (5) in [3] is, in the words of its author, "highly complicated". Surprisingly, the only known to me application of (5) is the bound (4), which applies only to "very" rational doubly-stochastic matrices. The main goal of this paper is to show the amazing power of (5), which has been overlooked for 13 years.

2 A Generalization of Schrijver's permanental inequality

We prove in this section the following theorem, stated in [10] in a rather cryptic way.Fortunately, the paper cites [11] and M. Chertkov is my colleague in Los Alamos.
The statement in the current paper has been communicated to me by Misha Chertkov, to whom I am profoundly grateful.

Definition 2.1: Define for a pair (P,Q) of non-negative matrices the following functional:

$$CW(P,Q) =: \sum_{1 \leq i,j \leq n} \log(1-Q(i,j))(1-Q(i,j)) - \sum_{1 \leq i,j \leq n} Q(i,j) \log\left(\frac{Q(i,j)}{P(i,j)}\right). \quad (6)$$

(Note that for fixed P the functional $CW(P,Q) = \sum_{1 \leq i,j \leq n} F_{i,j}(Q(i,j))$ and $F_{i,j}(0) = 0$. If $Q \in \Omega_n$ is doubly-stochastic and $P = Diag(a_1,...,a_n)TDiag(b_1,...,b_n)$ then

$$CW(P,Q) = \sum_{1 \leq i \leq n} \log(a_i b_i) + CW(T,Q). \quad (7)$$

Therefore, WLOG we can consider only doubly-stochastic matrices P.
The functional $CW(P,Q)$ is concave in P and, rather surprisingly (see the 2011 arxiv version of [10]), concave in $Q \in \Omega_n$.
∎

Theorem 2.2: *Let P be non-negative $n \times n$ matrix. If $Per(P) > 0$ then $max_{Q \in \Omega_n} CW(P,Q)$ is attained and*

$$\log(Per(P)) \geq max_{Q \in \Omega_n} CW(P,Q) \quad (8)$$

(It is assumed that $0^0 = 1$.)
An equivalent statement of this theorem is

$$\log(Per(P)) \geq \sum_{1 \leq i,j \leq n} \log(1-Q(i,j))(1-Q(i,j)) - \sum_{1 \leq i,j \leq n} Q(i,j) \log\left(\frac{Q(i,j)}{P(i,j)}\right) : P \geq 0, Q \in \Omega_n \quad (9)$$

Proof: We will prove, to avoid trivial technicalities, just the positive case, i.e when $P(i,j) > 0, 1 \le i,j \le n$.
We compute first partial derivatives:

$$\frac{\partial}{\partial Q} CW(P,Q) = \{-2 - \log(1 - Q(i,j)) - \log(Q(i,j)) + \log(P(i,j)) : 1 \le i,j \le n\} \quad (10)$$

In the positive case, i.e. for the fixed positive P, the functional $CW(P,Q)$ is bounded and continuous on Ω_n. Therefore the maximum exists. Let $V \in \Omega_n$ be one of argmaximums, i.e.

$$CW(P,V) = max_{Q\in\Omega_n} CW(P,Q).$$

Then, after some column/row permutations

$$V = \begin{pmatrix} V_{1,1} & 0 & ... & 0 \\ 0 & V_{2,2} & 0 & ...0 \\ . & . & . & . \\ 0 & ... & 0 & V_{k,k} \end{pmatrix};$$

$$P = \begin{pmatrix} P_{1,1} & . & ... & . \\ . & P_{2,2} & . & .... \\ . & . & . & . \\ . & ... & . & P_{k,k} \end{pmatrix};$$

The diagonal blocks $V_{i,i}$ are indecomposable doubly-stochastic $d_i \times d_i$ matrices; $\sum_{1\le i\le k} d_i = n$ and $1 \le k \le n$. Clearly,

$$CW(P,V) = \sum_{1\le i\le k} CW(P_{i,i}, V_{i,i}).$$

As $\log(per(P)) \ge \sum_{1\le i\le k} \log(per(P_{i,i}))$ it is sufficient to prove that

$$\log(Per(P_{i,i})) \ge CW(P_{i,i}, V_{i,i}); 1 \le i \le k.$$

For blocks of size one, the inequality is trivial: $(1-1)^{1-1} - 1\log(\frac{1}{a}) = \log(a)$.
Consider a (indecomposable) block $V_{i,i}$ of size $d_i \ge 2$ and define its support

$$Supp(V_{i,i}) = \{(k,l) : V_{i,i}(k,l) > 0\}.$$

Note that $1 > V_{i,i}(k,l) > 0, (k,l) \in Supp(V_{i,i})$. Consider the following functional

$$L(W_{i,i}) =: \sum_{(k,l)\in Supp(W_{i,i}} \log(1-W_{i,i}(k,l))(1-W_{i,i}(k,l)) - \sum_{(k,l)\in Supp(V_{i,i}} W_{i,i}(k,l) \log\left(\frac{W_{i,i}(k,l)}{P(i,j)}\right)$$

defined on compact convex subset of doubly-stochastic matrices which are zero outside of $Supp(P_{i,i})$. We conclude that the functional $L(\dot{})$ is differentiable at $V_{i,i}$. Note that

$L(V_{i,i}) = CW(P_{i,i}, V_{i,i}))$.
We now can express the local extremality condition not on full Ω_{d_i} but rather on its compact convex subset of doubly-stochastic matrices which are zero outside of $Supp(P_{i,i})$. Using (10) and doing standard Lagrange multipliers respect to variables $V_{i,i}(k,l), (k,l) \in Supp(V_{i,i})$, we get that there exists real numbers $(\alpha_k; \beta_l)$ such that

$$-2 - \log(1 - V_{i,i}(k,l)) - \log(V_{i,i}(k,l)) + \log(P_{i,i}(k,l)) = \alpha_k + \beta_l : (k,l) \in Supp(V_{i,i}).$$

Which gives for some positive numbers a_k, b_l the following scaling:

$$P_{i,i}(k,l) = a_k b_l V_{i,i}(k,l)(1 - V_{i,i}(k,l)); (k,l) \in Supp(V_{i,i}). \tag{11}$$

It follows from the definition of the support that

1. $$P_{i,i} \geq Diag(a_k)\widetilde{V_{i,i}}Diag(b_l); \widetilde{V_{i,i}}(k,l) = V_{i,i}(k,l)(1 - V_{i,i}(k,l)). \tag{12}$$

2. Using the scalability (7) property, we get that

$$CW(P_{i,i}, V_{i,i}) = \sum \log(a_k) + \sum \log(b_l) + \sum_{(k,l) \in Supp(V_{i,i})} \log(1 - V_{i,i}(k,l)). \tag{13}$$

Finally it follows from (13) and Schriver's permanental inequality (5) that

$$\log(per(Diag(a_k)\widetilde{V_{i,i}}Diag(b_l)) \geq CW(P_{i,i}, V_{i,i});$$

and that

$$\log(per(P_{i,i})) \geq \log(per(Diag(a_k)\widetilde{V_{i,i}}Diag(b_l)) \geq CW(P_{i,i}, V_{i,i}).$$

∎

Remark 2.3: Note that the proof does not use concavity of $CW(P,V)$ in $V \in \Omega_n$. ∎

3 Corollaries

1. Schrijver's permanental inequality (5) is a particular case of (9). Indeed

$$CW(\tilde{V}, V) = \sum_{1 \leq i,j \leq n} \log(1 - V(i,j)) : V \in \Omega_n.$$

2. Let $P \in \Omega_n$ be doubly-stochastic $n \times n$ matrix. Then

$$log(per(P)) \geq CW(P,P) = \sum_{1 \leq i,j \leq n} \log(1 - P(i,j))(1 - P(i,j)).$$

We get the following important inequality, perhaps the main observation in this paper:

$$\frac{per(P)}{F(P)} \geq 1; F(P) =: \prod_{1 \leq i,j \leq n} (1 - P(i,j))^{1-P(i,j)}; P \in \Omega_n \tag{14}$$

The lower bound (14) suggests the importance of the following quantity:

$$UB(n) =: max_{P \in \Omega_n} \frac{per(P)}{F(P)}.$$

It is easy to show that the limit

$$UB =: \lim_{n \to \infty} (UB(n))^{\frac{1}{n}}$$

exists and $1 \leq UB \leq e$. There is obvious deterministic poly-time algorithm to approximate the permanent of nonnegative matrices within relative factor $UB(n)$. The current best rate is e^n. Therefore proving that $UB < e$ is of major algorithmic importance.

Remark 3.1: All previous lower bounds on the permanent of **doubly-stochastic** matrices $P \in \Omega_n$ depend only on the dimension n and the support of P. I.e. the previous bounds are structural. The beauty (and potential power) of our lower bound (14) is in its explicit dependence on the entries of P. We use (14) in Section 5 to settle important conjecture on the monomer-dimer entropy. ∎

Example 3.2: I. Let $P = aJ_n + bI_n, a = \frac{1}{2(n-1)}, b = \frac{n-2}{2(n-1)}$, i.e. the diagonal $P(i,i) = \frac{1}{2}, 1 \leq i \leq n$ and the off-diagonal entries are equal to $\frac{1}{2(n-1)}$.
It is easy to see that for these (a,b):

$$2^{-n+1} \leq per(aJ_n + bI_n) = n!a^n \sum_{0 \leq i \leq n} \frac{1}{i!} \left(\frac{b}{a}\right)^i \leq n!a^n exp\left(\frac{b}{a}\right).$$

Non-difficult calculations show that for this $P \in \Omega_n$

$$\frac{per(P)}{F(P)} \approx \left(\sqrt{\frac{e}{2}}\right)^n \tag{15}$$

II. Let $P \in \Omega_2 = \frac{1}{2}J_2$ be 2×2 "uniform" doubly-stochastic matrix. The direct inspection gives that

$$CW(P,Q) \equiv -2\log(2) = F(P), Q \in \Omega_n.$$

Consider now the direct sum $P_{2n} \in \Omega_{2n} = \frac{1}{2}J_2 \oplus ... \oplus \frac{1}{2}J_2$. Then

$$max_{Q\in\Omega_{2n}} CW(P_{2n}, Q) = \log(F(P_{2n})) = -2n\log(2). \tag{16}$$

Therefore in this case

$$\frac{per(P_{2n})}{F(P_{2n})} = 2^n. \tag{17}$$

Which gives the following lower bound on $UB(k)$ for even k:

$$UB(k) \geq (\sqrt{2})^k. \tag{18}$$

As $max_{Q\in\Omega_{2n}} CW(P_{2n}, Q) = \log(F(P_{2n}))$, this class of matrices also provides a counter-example to the non-trivial part of **Conjecture 15** in [10].
Is the bound (18) sharp? ∎

3. Recall the main function from [7]:

$$G(x) = \left(\frac{x-1}{x}\right)^{x-1}, x \geq 1.$$

Note that for $P \in \Omega_n$ the column product

$$CPR_j(P) =: \prod_{1\leq i\leq n} (1 - P(i,j))^{1-P(i,j)} \geq G(n). \tag{19}$$

Define C_j as the number of non-zero entries in the jth column then

$$CPR_j(P) =: \prod_{1\leq i\leq n} (1 - P(i,j))^{1-P(i,j)} \geq G(C_j). \tag{20}$$

The inequality (19) gives a slightly weaker version of the celebrated Falikman-Egorychev-van der Waerden lower bound (3):

$$per(P) \geq \prod_{1\leq j\leq n} CPR_j(P) \geq \left(\frac{n-1}{n}\right)^{n(n-1)}$$

The inequality (20) gives a non-regular real-valued version of **(Schrijver-bound)**:

$$per(P) \geq \prod_{1\leq j\leq n} CPR_j(P) \geq \prod_{1\leq j\leq n} G(C_j) \tag{21}$$

In the worst case, the author's bound from [7] is better:

$$per(P) \geq \prod_{1\leq j\leq n} G\left(min(j, C_j)\right) \tag{22}$$

Perhaps, it is true that

Conjecture 3.3:

$$per(P) \geq \prod_{1\leq j\leq n} G\left(min(j, EC_j)\right)?$$

where the effective real-valued degree $EC_j = G^{-1}(CPR_j(P))$. ∎

4 Some historical remarks

The column products $CPR_j(P) =: \prod_{1 \leq i \leq n}(1 - P(i,j))^{1-P(i,j)} \geq G(C_j)$ have appeared in the permanent context before. Let $P = [a|b, .., |b] \in \Omega_n$ be doubly-stochastic matrix with 2 distinct columns. Then (Proposition 2.2 in [17])

$$Per(P) \geq CPR(1)vdw(n-1). \tag{23}$$

Let us recall a few notations from [7] and [5]:

1. The linear space of homogeneous polynomials with real (complex) coefficients of degree n and in m variables is denoted $Hom_R(m,n)$ ($Hom_C(m,n)$).
   We denote as $Hom_+(m,n)$ ($Hom_{++}(n,m)$) the closed convex cone of polynomials $p \in Hom_R(m,n)$ with nonnegative (positive) coefficients.

2. For a polynomial $p \in Hom_+(n,n)$ we define its **Capacity** as

$$Cap(p) = \inf_{x_i > 0, \prod_{1 \leq i \leq n} x_i = 1} p(x_1, \ldots, x_n) = \inf_{x_i > 0} \frac{p(x_1, \ldots, x_n)}{\prod_{1 \leq i \leq n} x_i}. \tag{24}$$

3. The following product polynomial is associated with a $n \times n$ matrix P:

$$Prod_P(x_1, \ldots, x_n) =: \prod_{1 \leq i \leq n} \sum_{1 \leq j \leq n} P(i,j)x_j. \tag{25}$$

   The permanent $per(P)$ is the mixed derivative of the polynomial $Prod_P$:

$$Per(P) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} Prod_P(0). \tag{26}$$

4. $$q_{(j)} =: \frac{\partial}{\partial x_j} Prod_P(x_1, \ldots, x_n) : x_j = 0.$$

   Note that the polynomials $q_{(j)} \in Hom_+(n-1, n-1)$
   For example, $q_{(n)} = \frac{\partial}{\partial x_n} Prod_P(x_1, \ldots, x_{n-1}, 0)$.

The following lower bound, which holds for all $P \in \Omega_n$, was proved in [5]:

$$Cap(q_{(j)}) \geq CPR_j(P), 1 \leq j \leq n. \tag{27}$$

Combining results from [7] (i.e. $Per(P) \geq vdw(n-1)Cap(q_{(j)}), 1 \leq j \leq n$) and (27) gives a different version of (14)

$$per(P) \geq \left(\prod_{1 \leq j \leq n} CPR_j(P) \right)^{\frac{1}{n}} vdw(n-1), P \in \Omega_n. \tag{28}$$

Or better

$$per(P) \geq (max_{1 \leq j \leq n} CPR_j(P))\, vdw(n-1), P \in \Omega_n. \tag{29}$$

Perhaps, it is even true that

Conjecture 4.1:

$$per(P) \geq \prod_{1 \leq j \leq n} Cap(q_{(j)}), P \in \Omega_n?$$

∎

A general, i.e. not doubly-stochastic and not just "permanental", version of Conjecture(4.1) is the following one:

Conjecture 4.2: Let $p \in Hom_+(n,n)$ be **H-Stable**, i.e. $p(z_1,...,z_n) \neq 0$ if the real parts $RE(z_i) > 0, 1 \leq i \leq n$. In other words, the homogeneous polynomial p does not have roots with positive real parts. Then the following inequality holds

$$\frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} p(0) \geq Cap(p) \prod_{1 \leq j \leq n} \frac{Cap(q_{(j)})}{Cap(p)} \tag{30}$$

∎

5 Some Partial Results Towards the Main Conjecture(s)

Let us formalize the main new question in the following Conjecture.

Conjecture 5.1: Let $P \in \Omega_n$ be doubly-stochastic matrix. Is it true that

1. **"Optimizational" Conjecture**

   $$per(P) \leq (\sqrt{2})^n exp(\max_{Q \in \Omega_n} CW(P,Q)).$$

   It will be explained below that **"Optimizational" Conjecture** gives provable deterministic polynomial(but not strongly) algorithm to approximate $per(P)$ with the factor $(\sqrt{2})^n$

2. **Strong Conjecture**

   $$per(P) \leq (\sqrt{2})^n F(P), F(P) =: \prod_{1 \leq i,j \leq n} (1-P(i,j))^{1-P(i,j)}.$$

   Strong Conjecture obviously gives deterministic strongly-polynomial algorithm to approximate $per(P)$ with the factor $(\sqrt{2})^n$

3. **Mild Conjecture**
$$per(P) \leq (\sqrt{2})^n (F(P))^c,$$
where $0 < c < 1$ is some universal constant. The case $c = \frac{1}{2}$ seems believable. As $per(P) \geq F(P)$ thus **Mild Conjecture** gives deterministic strongly-polynomial algorithm to approximate $per(P)$ with the factor
$(F(P))^{c-1} \leq\approx e^{n(1-c)} < e^n$.

∎

5.1 Some Basic Properties of $CW(P,Q)$

The "odd entropy" function $OE(p) = p\log(p) - (1-p)\log(1-p), 0 \leq p \leq 1$ is not convex on $[0,1]$. Yet, when lifted to the Simplex it becomes convex. This non-obvious result was proved in recent extended version of [10]. We present below a simpler and more general proof.

Definition 5.2: Call a function $f : [0,1] \rightarrow \mathbf{R}$ **simplex-convex** if the functional $f_{Sim(1)}(p_1, ..., p_n) =: f(p_1) + ... + f(p_n)$ is convex on the simplex $Sim_n(1) = \{(p_1, ..., p_n) : p_i \geq 0, 1 \leq i \leq n; \sum_{1 \leq i \leq n} p_i = 1$. ∎

Clearly, if f is convex then it is **simplex-convex** as well. We describe a much wider class of **simplex-convex** functions.

We need two simple facts.

Fact 5.3: Let $g : [0,1] \rightarrow \mathbf{R}$ be convex function; $g(0) = 0$. Then this function g is super-additive:

$$g(t_1 + ... + t_n) \geq g(t_1) + ... + g(t_n) : t_i \geq 0, 1 \leq i \leq n; t_1 + ... + t_n \leq 1.$$

Proof: Let $t_i \geq 0, 1 \leq i \leq n; t_1 + ... + t_n = s \leq 1$. Lift g to the simplex $sSim_n(1) =: Sim_n(s) = \{t_i \geq 0, 1 \leq i \leq n; t_1 + ... + t_n = s\}$:

$$\bar{g}(t_1, ..., t_n) = g(t_1) + ... + g(t_n).$$

The functional $\bar{g}$ is convex on the simplex $Sim_n(s)$. Therefore, its maximium is attained at the extreme points, i.e at the vectors $(s, 0, ..., 0), ..., (0, 0, ..., s)$. As $g(0) = 0$ we get that
$$\max_{Sim_n(s)} \bar{g}(t_1, ..., t_n) = g(s),$$
which finishes the proof. ∎

Fact 5.4: Consider the linear subspace $Sum_0 \subset \mathbf{R}^{n+1}$, $Sum_0 = \{(x_0, x_1, ..., x_n) : \sum_{0 \leq i \leq n} x_i = 0$.
Let $a_0 > 0; a_i > 0, 1 \leq i \leq n$. The diagonal matrix $D = Diag(-a_0, a_1, ..., a_n)$ is positive semidefinite on Sum_0, i.e $< DX, X > \geq 0, X \in Sum_0$ iff

$$\sum_{1 \leq i \leq n} \frac{1}{a_i} \leq \frac{1}{a_0}; a_i > 0, 1 \leq i \leq n. \tag{31}$$

Proof: The proof directly follows from the following easily checkable equality:

$$\min_{y_1+....+y_n=1} \sum_{1 \leq i \leq n} a_i y_i^2 = \left(\sum_{1 \leq i \leq n} a_i^{-1} \right)^{-1}.$$

∎

Theorem 5.5: *Let $f : [0,1] \to \mathbf{R}$ be continuous and twice differentiable on $(0,1)$ function. Define $g(t) = f(\frac{1}{2} + t), t \in [-\frac{1}{2}, \frac{1}{2}]$. Assume that the second derivative $g^{(2)}$ satisfies the following properties:*

1. $g^{(2)}(t) > 0, 0 > t > -\frac{1}{2}; g^{(2)}(0) \geq 0$.
2. $g^{(2)}(t) \geq -g^{(2)}(-t), 0 < t < \frac{1}{2}$.
3. $\lim_{t \to -\frac{1}{2}} \frac{1}{g^{(2)}(t)} = 0$.
4. *The function $\frac{1}{g^{(2)}(t)}$ is convex on $[-\frac{1}{2}, 0)$.*

Then the function f is **simplex-convex***.*

Remark 5.6: The "odd entropy" function $OE(p) = p\log(p) - (1-p)\log(1-p), 0 \leq p \leq 1$ satisfies the above properties:
Indeed, $\frac{1}{g^{(2)}(t)} = \frac{1}{OE^{(2)}(\frac{1}{2}+t)} = -\frac{1}{8t} + \frac{t}{2}$. ∎

Proof: As f is continuous it is sufficient to prove that $f(t_0) + ... + f(t_n)$ is convex in the interior of the simplex $Sim_{n+1}(1)$, i.e when $0 < t_i < 1$. Define $d_i =: f^{(2)}(t_i)$. We need to prove that $D =: Diag(d_0, d_1, ..., d_n)$ is positive semidefinite on Sum_0.
If $t_i \leq \frac{1}{2}$ then $d_i \geq 0$ and D is positive semidefinite. Otherwise, there is only one $t_i > \frac{1}{2}$, say

$$t_0 = \frac{1}{2} + s_0, \frac{1}{2} \geq s_0 > 0; t_i = \frac{1}{2} - s_i; \frac{1}{2} > s_i > 0, 0 \leq i \leq n.$$

Note that $d_i > 0, 1 \leq i \leq n$. If $f^{(2)}(t_0) = g^{(2)}(s_0) \geq 0$ we are done. Assume that $-\beta =: f^{(2)}(t_0) = g^{(2)}(s_0) < 0, \beta > 0$. Our goal, using Fact(5.4), is to prove that

$$(\beta)^{-1} \geq \sum_{1 \leq i \leq n} d_i^{-1}. \tag{32}$$

Note that

$$\frac{1}{2} - s_0 = \sum_{1 \le i \le n} (\frac{1}{2} - s_i).$$

Using the properties(2-4) above and Fact(5.3), applied to the convex function $\alpha(t) =: \frac{1}{g^{(2)}(t)}, t \in [0, \frac{1}{2}]$, we get that

$$(f^{(2)}(\frac{1}{2} - s_0))^{-1} \ge \sum_{1 \le i \le n} (d_i)^{-1}.$$

As $f^{(2)}(\frac{1}{2} + s_0) < 0$ we get from property(2) above that

$$\beta = -f^{(2)}(\frac{1}{2} + s_0) \le f^{(2)}(\frac{1}{2} - s_0).$$

Which gives the desired inequality (32). ∎

Remark 5.7: Recall the definition of the *Bregman Distance* associated with a convex functional f:

$$0 \le D_f(X||Y) = f(X) - f(Y) - < \bigtriangledown F_Y, X - Y > .$$

For instance, the **Kullback-Leibler Divergence** is the *Bregman Distance* associated with

$$f(p_1, ..., p_n) = \sum_{1 \le i \le n} p_i \log(p_i).$$

As we know that the "odd entropy" functional

$$OE(p_1, ..., p_n)) = \sum_{1 \le i \le m} p_i \log(p_i) - (1 - p_i) \log(1 - p_i)$$

is convex on the simplex $Sim_n(1)$, we can define a new divergence, which we call **Bethe Divergence**:

$$BD(X||Y) = \sum_{1 \le i \le n} \left(x_i \log(\frac{x_i}{y_i}) - (1 - x_i) \log \left(\frac{1 - x_i}{1 - y_i} \right) \right); X, Y \in Sim_n(1). \qquad (33)$$

It would be interesting to investigate statistical (or learning) applications of the **Bethe Divergence**. ∎

5.2 Some easy exact computations of $\max_{Q \in \Omega_n} CW(P, Q)$

The following fact is easy corollary of the **simplex-convexity** of the "odd entropy" function $OE(p) = p \log(p) - (1 - p) \log(1 - p), 0 \le p \le 1$.

Fact 5.8:

1. Let $p_i > 0, 1 \leq i \leq n$ be a positive vector, $n \geq 2$. Define

$$OD(q,p) = \sum_{1 \leq i \leq n} (1-q_i)\log(1-q_i) - q_i \log(\frac{q_i}{p_i}), q \in Sim_n(1).$$

   Then

$$\max_{(q_1,...,q_n) \in Sim_n(1)} OD(q,p) = \log(p_j)$$

   iff $p_j \geq \sum_{i \neq j} p_i$. We call such index j dominant.
   Note that if $n \geq 3$ then there exists at most one dominant index.
   If there is no dominant index then the maximum is attained in the interior of the simplex $Sim_n(1)$.

2. Let $p_i = const > 0, 1 \leq i \leq n$. Then

$$\max_{(q_1,...,q_n) \in Sim_n(1)} OD(q,p) = OD(\frac{e}{n}, p) = (n-1)\log(1-n^{-1}) + \log(n) + \log(const).$$

Remark 5.9: The first item of Fact (5.8) says that for $n \geq 3$ the extremum of $OD(q,p), p > 0$ is either an extreme point of the simplex(when the unique dominant index exists) or a point in the interior. This is in stark contrast with KLD-minimization, where the extremum has largest possible support. ∎

We will take advantage of the following corollary.

Corollary 5.10: *Let RS_n denote the set of $n \times n$ row-stochastic matrices.*
Let P be $n \times n$ diagonally dominant non-negative matrix. i.e. $P(i,i) \geq \sum_{j \neq i} P(i,j); 1 \leq i \leq n$. Then

$$\max_{Q \in \Omega_n} CW(P,Q) = \max_{Q \in RS_n} CW(P,Q) = \sum_{1 \leq i \leq n} \log(P(i,i)). \tag{34}$$

The following observation follows now from the scalability property (7).

Corollary 5.11: *Assume that there exist two diagonal matrices D_1, D_2 such that the matrix $P = D_1 A D_2$ is diagonally dominant, i.e. $B(i,i) \geq \sum_{j \neq i} B(i,j)$. Then*

$$\max_{Q \in \Omega_n} CW(P,Q) = \sum_{1 \leq i \leq n} \log(P(i,i)).$$

5.3 Regular Bipartite Graphs

Let $RB(r,n)$ denote the set of $n \times n$ boolean matrices with row and column sums all equal to r. Note that if $A \in RB(r,n)$ then $\frac{1}{r}A$ is doubly-stochastic and

$$F(\frac{1}{r}A) = \left(\frac{r-1}{r}\right)^{n(r-1)} = G(r)^n.$$

The celebrated Bregman's upper bound [14] gives that

$$per(\frac{1}{r}A) \leq \left(\frac{r!^{\frac{1}{r}}}{r}\right)^n =: B_r^n.$$

Therefore

$$\frac{per(\frac{1}{r}A)}{F(\frac{1}{r}A)} \leq \left(\frac{B_r}{G(r)}\right)^n \leq \left(\frac{B_2}{G(2)}\right)^n = (\sqrt{2})^n.$$

Therefore, **Strong Conjecture** holds on the sets $RB(r,n)$.
Let $CO(\frac{1}{r}RB(r,n))$ be the convex hull. It follows from linearity of the permanent in individual rows that

$$per(\frac{1}{r}A) \leq \left(\frac{r!^{\frac{1}{r}}}{r}\right)^n = B_r^n, A \in CO(\frac{1}{r}RB(r,n)).$$

The following observation(most likely known) follows fairly directly from the classical J.Edmonds' result that the intersection of two matroid polytopes is the polytope of the intersection of the corresponding two matroids with the same ground set.

Proposition 5.12: *The convex hull*

$$CO(\frac{1}{r}RB(r,n)) = \{A \in \Omega_n : A(i,j) \leq \frac{1}{r}; 1 \leq i,j \leq n\}.$$

Corollary 5.13:

$$CO(\frac{1}{r+1}RB(r+1,n)) \subset CO(\frac{1}{r}RB(r,n)), 1 \leq r \leq n-1.$$

We only can state (rather trivial) upper bound

$$\frac{per(P)}{F(P)} \leq \frac{B_r^n}{G(n)^n} \leq \left(\frac{r!^{\frac{1}{r}}e}{r}\right)^n : P \in CO(\frac{1}{r}RB(r,n)). \tag{35}$$

It follows from (35) that **Strong Conjecture** holds on $CO(\frac{1}{r}RB(r,n)), r \geq 6$.

5.4 Diagonally Dominant Matrices

Lemma 5.14: *Let A be $n \times n$ non-negative matrix. Then*

$$Per(A) \leq \prod_{1 \leq i \leq n} (A(i,i)^2 + (\sum_{j \neq i} A(i,j)^2))^{\frac{1}{2}}. \tag{36}$$

Proof: Follows from linearity of the permanent in individual rows and the following generalized Holder's inequality

$$|\prod_{1 \leq i \leq n} a_i + \prod_{1 \leq i \leq n} b_i| \leq \prod_{1 \leq i \leq n} (|a_i|^n + |b_i|^n)^{\frac{1}{n}}. \tag{37}$$

∎

Corollary 5.15: *If A is Diagonally Dominant then the* **"Optimizational" Conjecture** *holds, i.e.*

$$per(A) \leq (\sqrt{2})^n exp(\max_{Q \in \Omega_n} CW(A,Q)).$$

6 A proof of Friedland's *Asymptotic Lower Matching Conjecture*

6.1 Two models for random regular bipartite graphs with multiple edges

We denote as $RI(r,n)$ the set of $n \times n$ non-negative integer matrices with row and column sums all equal r:

$$RI(r,n) = \{\{A(i,j); 1 \leq i,j \leq n\} : A(i,j) \in Z_+; Ae = A^T e = re\}.$$

1. **The Pairing Model**: Consider a random, respect to uniform distribution, permutation $\pi \in S_{rn}$ of length rn and its standard matrix representation, pictured as a block matrix:

$$M_\pi = \begin{pmatrix} M_\pi(1,1) & M_\pi(1,2) & ... & M_\pi(1,r) \\ ... & ... & ... & ... \\ M_\pi(r,1) & M_\pi(r,2) & ... & M_\pi(r,r) \end{pmatrix},$$

   where each block is a (boolean) $n \times n$ matrix. **The Pairing Model** for a random matrix in $RI(r,n)$ corresponds to a random matrix $BM(r,n) =: \sum_{1 \leq i,j \leq r} M_\pi(i,j)$. This model was used in the context of the permanent in [2].

2. **The sum of r independent permutation matrices**: Another model is just the sum of r independent permutation matrices:

$$HW(r,n) =: \sum_{1 \le i \le r} M_{\sigma_i},$$

where $\sigma_i \in S_n, 1 \le i \le r$ are independent uniformly disributed permutations of length n. This model was used by Herbert Wilf [1]. As in [2], the main goal and result of [1] was the asymptotics of the expected value of the permanent:

$$\lim_{n\to\infty} (E(per(HW(r,n)))^{\frac{1}{n}} = \lim_{n\to\infty} (E(per(BM(r,n)))^{\frac{1}{n}} = rG(r). \qquad (38)$$

It is worth noticing that the proof in [1] is much more involved than in [2]. One of the corollaries of (38) is the following inequality

$$\lim_{n\to\infty} (\min_{A \in RI(r,n)} (per(A))^{\frac{1}{n}} \le rG(r), \qquad (39)$$

which was proved much later to be equality.
Let $prob_1(r,n)$ be the probability of the event $BM(r,n) \in RB(r,n)$, where $RB(r,n)$ is the set of $n \times n$ boolean matrices with r ones in each row and column; $prob_2(r,n)$ be the probability of the event $HW(r,n) \in RB(r,n)$. Brian McKay conjectured in [23] that for fixed r(we present here a simplified expression)

$$prob_1(r,n) = exp\left(-\frac{(r-1)^2}{2} + O(n^{-1})\right). \qquad (40)$$

This conjecture was proved almost 20 years after in [22], moreover it holds for $r = o(\sqrt{n})$. The proof in [23] is rather involved and has nothing to do with the permanent.
On the other hand, it is easy to see that

$$\frac{1}{(n!)^{r-1}} \prod_{1 \le i \le r-1} \min_{A \in RB(n-i,n)} per(A) \le prob_2(r,n) \le \frac{1}{(n!)^{r-1}} \prod_{2 \le i \le r} \max_{A \in RB(n-i,n)} per(A). \quad (41)$$

We can use now various lower bounds on $\min_{A \in RB(n-i,n)} per(A)$ and the Bregman's upper bound $((n-i)!)^{\frac{n}{n-i}}$ on $\max_{A \in RB(n-i,n)} per(A)$.
Using just the Van Der Waerden-falikman-Egorychev (or even Bang-Friedland) bound we get that

$$\prod_{1 \le i \le r-1} (\frac{n-i}{n})^n \le prob_2(r,n) \le \prod_{1 \le i \le r-1} \frac{((n-i)!)^{\frac{i}{n-i}}}{(n-i+1)...n} \qquad (42)$$

The best current lower bound (22) gives

$$\prod_{1 \le i \le r-1} G(n-i)^i \frac{(n-i)^i}{(n-i+1)...n} \le prob_2(r,n) \le \prod_{1 \le i \le r-1} \frac{((n-i)!)^{\frac{i}{n-i}}}{(n-i+1)...n} \qquad (43)$$

For a fixed r, as 42 as well 43 give the following asymptotic for $prob_2(r,n)$

$$prob_2(r,n) \approx exp\left(-\frac{r(r-1)}{2}\right),$$

which is less than (40).
Ian Wanless noticed in [21] that

$$\min_{A \in RB(r,n)} per(A) \leq (prob_1(r,n))^{-1} E(per(BM(r,n)).$$

Together with (40) it implies that

$$\lim_{n\to\infty} \left(\min_{A \in RB(r,n)} (per(A))^{\frac{1}{n}} \right) \leq rG(r), \tag{44}$$

which is the main conclusion of [21]. We sketched above an alternative, simpler way to get the same result by combining Herbert Wilf's 1966 paper and Van Der Waerden-Falikman-Egorychev Inequality and their recent refinements.

6.2 Monomer-Dimer Problem

.

Let $per_m(A)$ denote the sum of permanents of all $m \times m$ submatrices of A:

$$per_m(A) =: \sum_{|S|=|T|=m} per(A_{S,T}).$$

Define the following two quantities

$$EMD_1(r,n;m) = E(per_m(BM(r,n))), EMD_2(r,n;m) = E(per_m(HW(r,n))).$$

A rather direct generalization of derivations in [2] and [1] gives the following asymptotics

$$\lim_{n\to\infty, \frac{m}{n} \to t \in [0,1]} \frac{\log(EMD_1(r,n;m))}{n} = g_r(t) =: t\log(\frac{r}{t}) - 2(1-t)\log(1-t) + (r-t)\log(1-\frac{t}{r}),$$

and

$$\lim_{n\to\infty, \frac{m}{n} \to t \in [0,1]} \frac{\log(EMD_2(r,n;m))}{n} = g_r(t).$$

It follows that

$$\lim_{n\to\infty, \frac{m}{n} \to t \in [0,1]} \frac{\min_{A \in RI(r,n)} \log(per_m(A))}{n} \leq g_r(t). \tag{45}$$

The Wanless argument gives the same inequality for the boolean case

$$\lim_{n\to\infty, \frac{m}{n} \to t \in [0,1]} \frac{\min_{AinRB(r,n)} \log(per_m(A)}{n} \leq g_r(t). \tag{46}$$

The Friedland's *Asymptotic Lower Matching Conjecture* asserts (after [2], [3]) that

$$\lim_{n\to\infty, \frac{m}{n}\to t\in[0,1]} \frac{\min_{A\in RB(r,n)} \log(per_m(A))}{n} \geq g_r(t). \tag{47}$$

We prove in this paper a slightly stronger result:

$$\lim_{n\to\infty, \frac{m}{n}\to t\in[0,1]} \frac{\min_{AinRI(r,n)} \log(per_m(A)}{n} \geq g_r(t). \tag{48}$$

Of course, as we explained above using Wanless argument, the inequalities ($\geq$) in (47, 48) imply equalities.
The *Lower Matching Conjecture* asserts that

$$per_m(A) \geq D(r;m,n) =: \binom{n}{m}^2 (\frac{r-t}{r})^{n(r-t)}(tr)^t; A \in RB(r,n), t =: \frac{m}{n}. \tag{49}$$

We prove in this paper the wollowing weeker inequality but for more general class of matrices, i.e for $A \in RI(r,n)$:

$$per_m(A) \geq SF(r,n,m) =: \frac{(\frac{r-t}{r})^{n(r-t)}(1-n^{-1})^{(1-n^{-1})2n^2(1-t)}}{(\frac{t}{r})^{nt} n^{-2n(1-t)}((n(1-t))!)^2} \tag{50}$$

We note that

$$\frac{D(r;m,n)}{SF(r,n,m)} = \left(\frac{G(n)^{n-m}}{G(m+1)...G(n)}\right)^2 > 1, m < n, \tag{51}$$

where $G(x) = (\frac{x-1}{x})^{x-1}, x \geq 1$.

The following simple Fact will be used below.

Fact 6.1:

1. Define the following function $G(x,t) = (\frac{x-t}{x})^{x-t}, x \geq t \geq 0$. For a fixed $t > 0$ the function $G(x,t)$ is decreasing in x.
2. Let $(a_1, ..., a_k$ be positive numbers, $\sum_{1\leq i\leq k} a_i = 1$. Then
$$\prod_{1\leq i\leq k} (1-ta_i)^{1-ta_i} \geq G(k,t); 0 \leq t \leq 1.$$

Theorem 6.2: *Let $A \in RI(r,n)$. For a positive integer $m \leq n$ define*

$$t = \frac{m}{n}, \alpha = \frac{t}{r}.$$

Then the following lower bound holds:

$$per_m(A) \geq SF(r,n,m) =: \frac{(1-\alpha)^{(1-\alpha)nr}(1-n^{-1})^{(1-n^{-1})2n^2(1-t)}}{\alpha^{nt} n^{-2n(1-t)}((n(1-t))!)^2} \tag{52}$$

(Notice that $(1-\alpha)^{(1-\alpha)nr} = G(r,t)^n$.)

Proof:

1. Step 1.
   Consider the following $2n-m \times 2n-m$ matrix

$$K = \begin{pmatrix} aA & bJ_{n,n-m} \\ (bJ_{n,n-m})^T & 0 \end{pmatrix}, \tag{53}$$

   where $a = \alpha = \frac{t}{r}$, $b = \frac{1}{n}$, and $J_{n,n-m}$ is $n \times n-m$ matrix of all ones.
   It is easy to check that this matrix K is doubly-stochastic. Importantly, the following identity holds:

$$per_m(A) = \frac{per(K)}{a^m b^{2(n-m)}((n-m)!)^2}. \tag{54}$$

2. Step 2.
   We apply the inequality (14) to the doubly-stochastic matrix K

$$per(K) \geq \left(\prod_{1 \leq i,j \leq n} (1 - \frac{t}{r} A(i,j))^{(1-\frac{t}{r})A(i,j)} \right) \left(1 - \frac{1}{n}\right)^{(1-\frac{1}{n})2n^2(1-t)}. \tag{55}$$

3. Step 3.
   Let d_j be the number of non-zero entries in the jth column of A. Notice that $d_j \leq r$ and $\sum_{A(i,j)\neq 0} \frac{A(i,j)}{r} = 1$. It follows from Fact (6.1) that

$$\prod_{1 \leq i,j \leq n} \left(1 - \frac{t}{r} A(i,j)\right)^{(1-\frac{t}{r})A(i,j)} \geq \prod_{1 \leq j \leq n} G(d_j, t) \geq G(r,t)^n \tag{56}$$

   Which gives the following lower bound on the permanent of K:

$$per(K) \geq G(r,t)^n (1 - \frac{1}{n})^{(1-\frac{1}{n})2n^2(1-t)} \tag{57}$$

4. Step 3.
   Finally, we get (52) by combining the (nontrivial, new) inequality (57) with the (trivial, well known) identity (54).

∎

Remark 6.3: We can express $SF(r,n,m)$ in terms of the function $G, G(x) = \left(\frac{x-1}{x}\right)^{x-1}, x \geq 1$:

$$SF(r,n,m) = \frac{(1-\alpha)^{(1-\alpha)nr}}{\alpha^{nt}} \frac{G(n)^{2n(1-t)}}{(G(1)...G(n(1-t)))^2 (1-t)^{2n(1-t)}} \tag{58}$$

∎

The following more general result is proved in the very same way.

Theorem 6.4: *Let $P \in \Omega_n$. Then*

$$per_m(P) \geq \frac{\left(\prod_{1 \leq i \leq n} (1 - \frac{m}{n} P(i,j))^{1 - \frac{m}{n} P(i,j)}\right) G(n)^{2(n-m)}}{(\frac{m}{n})^m n^{-2n(1-t)}((n(1-t))!)^2}. \quad (59)$$

Using Fact(6.1) one can get various corollaries of Theorem(6.4) expressed in terms of the support of doubly-stochastic matrix P.

Corollary 6.5: *Fix a positive integer r and consider a sequence of pairs (n, m) such that*

$$n \to \infty, \frac{m}{n} \to t \in (0,1).$$

Then

$$\frac{\log(SF(r,n,m))}{n} \to g_r(t) = t \log(\frac{r}{t}) - 2(1-t)\log(1-t) + (r-t)\log(1-\frac{t}{r}) \quad (60)$$

Together with inequalities (45, 46) this solves Asymptotic Lower Matching Conjecture

$$\lim_{n \to \infty} \frac{\log(\min_{A \in RB(r,n)} per_m(A))}{n} = \lim_{n \to \infty} \frac{\log(\min_{A \in RI(r,n)} per_m(A))}{n} = g_r(t). \quad (61)$$

Proof: We only need to prove (60). The proof follows either from the Stirling approximation of the factorial or from the representation (58), using the well known fact that $\lim_{n \to \infty} G(n) = e^{-1}$.
∎

Remark 6.6:

1. The representation $\frac{n!}{n^n} = \prod_{1 \leq i \leq n} G(i)$ provides very simple derivation of the Stirling formula.

2. The first published statement of *Asymptotic Lower Matching Conjecture* appeared in [15].The author learned about the statement of (61) from Shmuel Friedland in 2005.
   The main result of [6] (and of 2006 arxiv version) was the limit equality (61) for $t = \frac{r}{r+s}, s = 0, 1, 2, \ldots$. The fairly self-contained and simple proof in [6] was based on the "hyperbolic polynomials approach" introduced first in [17]. The actual result in [6] was stated in terms of **sums of mixed derivatives** of general positive hyperbolic polynomials (the same as **H-Stable** in [7]), albeit for a restricted range of the parameter t. The proof in the present paper is not general at all, it works

only for the m-permanent, i.e. for the class of polynomials $Sym_m(y_1, ..., y_n)$, where y_i are linear forms with non-negative coefficients. But in this case the full range of densities $t \in [0, 1]$ is covered.
Whether it can be generalized to general **H-Stable** polynomials remains open.
Our proof of *Asymptotic Lower Matching Conjecture* illustrates once more how badly had the "Bethe Restatement" of Schrijver's inequality (5) been overlooked.
The author did some search on *Google Scholar* and found, to his amazement, that the Bethe approximation is one the oldest heuristics for the monomer-dimer problem, goes back to 1930s. So, the recent Bethe Approximation approach(as a heuristic) to the permanent is, in a way, a rediscovery. Apparently, the first recent publication in this direction was [16].
How cool is it that this classical statistical physics stuff was one of the main keys to rigorously settle the *Asymptotic Lower Matching Conjecture*! Of course, it would have been rather useless without the amazing Schrijver's inequality (5). Note that the validity of Conjecture 4.1 also implies *Asymptotic Lower Matching Conjecture*.
It would be great to prove Conjecture 4.1 using **H-Stable** polynomials.

3. The following equality holds for the doubly-stochastic matrices K as in (53):

$$F(K) = \max_{Q \in \Omega_n} CW(K, Q)$$

∎

7 A disproof of a positive correlation conjecture due to [Lu,Mohr,Szekely]

Let A be $n \times n$ stochastic matrix, i.e. the rows of A are probabilistic distributions on $\{1, ..., n\}$; $(e_1, ..., e_n)$ is the standard basis in R^n.
Let $\mathbf{V} =: (V_1, ..., V_n)$ be a n-tuple of independent random vectors:

$$Prob(V_i = e_k) = A(i, k); 1 \leq i, k \leq n.$$

The distribution of the sum $V_1 + ... + V_n$ coincides with the vector of the coefficients of the product polynomial

$$Prod_A, Prod_A(x_1, ..., x_n) = \prod_{1 \leq i \leq n} \sum_{1 \leq j \leq n} A(i, j) x_j,$$

i.e. the probability $Prob(V_1 + ... + V_n = (\omega_1, ..., \omega_n))$ is the coefficient $a_{\omega_1,...,\omega_n}$ of the monomial $\prod_{1 \leq i \leq n} x_i^{\omega_i}$ in the polynomial $Prod_A$. In particular,

$$per(A) = Prob(V_1 + ... + V_n = e), \tag{62}$$

where $e = (1, 1, ..., 1)$ is the vector of all ones.
Notice that the expected value $E(V_1 + ... + V_n) = (c_1, ..., c_n)$, where c_j is the sum of the jth column of A. Thus in the doubly-stochastic case

$$per(A) = Prob(V_1+...+V_n = E(V_1+...+V_n)) = Prob(||V_1+...+V_n-E(V_1+...+V_n)||_2^2 < 2), \quad (63)$$

and the lower bounds on the permanent of doubly-stochastic matrices can be viewed as concentration inequalities for sums of independent random vectors. This interpretation raises a number of natural questions:

1. What are the lower bounds on $Prob(||V_1+...+V_n-E(V_1+...+V_n)||_2^2 \leq R \leq n(n-1)$ in the doubly-stochastic case? Van Der Waerden-Falikman-Egorychev gives the lower bound $\frac{n!}{n^n} \approx exp(-n)$ for $R < 2$.
   This question, albeit for distributions associated with **H-Stable** polynomials, was asked by the author in [9].

2. Is it possible to use this probabilistic interpretation to get new lower bounds, like (14) in this paper?

3. Is there a lower bound, similar to Van Der Waerden-Falikman-Egorychev, on $Prob(||V_1 + ... + V_n - E(V_1 + ... + V_n)||_2^2 \leq 2)$ for stochastic matrices,perhaps with some different, yet small, radius?

4. The coefficients of the products polynomials $Prod_A, A \geq 0$, and of more general **H-Stable** and **Strongly Log-Concave** polynomials [**?**], satisfy a lot of log-concave like inequalities. Perhaps one use them to prove new concentration inequalies of type we listed above?

5. We invite the reader to raise more questions.

Remark 7.1: We presented above very simple and effective "classical" generator to sample the distribution $Dist = \{a_{\omega_1,...,\omega_n} : (\omega_1, ..., \omega_n) \in Z_+^n, \omega_1 + ... + \omega_n = n\}$. The similar problem for the doubly-stochastic polynomial

$$Per_U(x_1, ..., x_n) = per(UDiag(x_1, ..., x_n)U^*),$$

where U is $n \times n$ complex unitary matrix, is of major importance in **Quantum Computing**. The generator in this paper can be viewed as a classical approximation.
If $p \in Hom_+(n, n)$ is doubly-stochastic and log-concave on R_+^n then its coefficients satisfy the inequality

$$p_{\omega_1,...,\omega_n} \leq \prod_{1 \leq i \leq n} \omega_i^{-\omega_i}.$$

The permanental polynomials $Per_U(x_1, ..., x_n)$ have much veaker upper bounds:

$$q_{\omega_1,...,\omega_n} \leq \prod_{1 \leq i \leq n} \frac{(\omega_i)!}{\omega_i^{\omega_i}}.$$

$\blacksquare$

Define the following n events:

$$NE_i = \{(V_1, ..., V_n) : V_i \notin \{V_j, j \neq i\}\}; 1 \leq i \leq n.$$

Equivalently

$$per(A) = Prob(\cap_{1 \leq i \leq n} NE_i). \tag{64}$$

The authors of [19] noticed that $Prob(NE_i) = \sum_{1 \leq j \leq n} A(i,j) \prod_{k \neq i}(1 - A(k,j)$ and conjectured the following beautiful positive correlation inequality for doubly-stochastic matrices $A \in Omega_n$:

$$per(A) \geq G(A) =: \prod_{1 \leq i \leq n} Prob(EV_i) = \prod_{1 \leq i \leq n} \sum_{1 \leq j \leq n} A(i,j) \prod_{k \neq i}(1 - A(k,j). \tag{65}$$

It is easy to see that $G(A) \geq F(A), A \in \Omega_n$ and $G(A) = F(A)$ in the regular case, i.e. when $A \in r^{-1}RB(r,n); 1 \leq r \leq n$. Therefore in this regular case the inequality (65) holds and is equivalent to the **(Schrijver-bound)** (4).
Apparently the authors of [19] did a substantial numerical validation of the conjecture on random matrices of modest size.
Surprisingly, the Monomer-Dimer Problem provides a probabilistic counter-example.
We will present finite families $F_n \subset \Omega_n$ such that $G(A) = Const, A \in F(n)$ but the average with some weigths of the permanent over F_n is exponentially smaller than $Const$.

7.1 The Construction

Consider either of two random models in $RI(r,n)$, say a random matrix $BM(r,n) \in RI(r,n)$. In induces a conditional distribution on $RB(r,n)$, i.e. a random matrix $CBM(r,n) \in RB(r,n)$ with the distribution

$$Prob(CBM(r,n) = A \in RB(r,n)) = \frac{prob(CBM(r,n) = A \in RB(r,n))}{prob\{BM(r,n) \in RB(r,n)\}}.$$

The Wanless argument gives that

$$\lim_{n \to \infty, \frac{m}{n} \to t \in [0,1]} \frac{\log(E(per_m(CBM(r,n))))}{n} \leq g_r(t).$$

Let $K \in \Omega_{2n-m}$ be the following random doubly-stochastic matrix

$$K = \begin{pmatrix} aCBM(r,n) & bJ_{n,n-m} \\ (bJ_{n,n-m})^T & 0 \end{pmatrix}, \tag{66}$$

$a = \frac{t}{r}, t = \frac{m}{n}; b = \frac{1}{n}$. By the direct inspection, we get that

$$G(K) = \left(t(1-\frac{t}{r})^{r-1}(1-\frac{1}{n})^{n(1-t)}) + (1-t)(1-\frac{1}{n})^{n-1}\right)^n \left((1-\frac{1}{n})^{n(1-t)-1}(1-\frac{t}{r})^{r(1-t)}\right)^{n(1-t)},$$

and

$$F(K) = \left(1-\frac{t}{r}\right)^{(r-t)n} \left(1-\frac{1}{n}\right)^{(n-1)2n(1-t)}.$$

Recall that

$$per_m(CBM(r,n)) = \frac{per(K)}{a^m b^{2(n-m)}((n-m)!)^2}. \quad (67)$$

The conjecture (65) would imply, if true, that

$$per_m(CBM(r,n))a^m b^{2(n-m)}((n-m)!)^2 \geq G(K). \quad (68)$$

Which would give

$$f(r,n,m) =: E(per_m(CBM(r,n))a^m b^{2(n-m)}((n-m)!)^2) \geq G(K). \quad (69)$$

But

$$\lim_{n\to\infty, \frac{m}{n}\to t\in[0,1]} \frac{\log(f(r,n,m))}{n} = (r-t)\log(1-\frac{t}{r}) - 2(1-t) =: M(t)$$

And

$$\lim_{n\to\infty, \frac{m}{n}\to t\in[0,1]} \frac{\log(G(K))}{n} = \log\left(t(1-\frac{t}{r})^{r-1}e^{-(1-t)} + (1-t)e^{-1}\right) - (1-t)^2 + r(1-t)\log\left(1-\frac{t}{r}\right) =: S(t).$$

The final observation is the following strict inequality

$$S(t) > M(t), 0 < t < 1,$$

which follows from the strict concavity of the logarithm and the inequality

$$(1-\frac{t}{r})^{r-1}e^{-(1-t)} > e^{-1}, 0 < t \leq 1.$$

8 Credits and Conclusion

The Definition (2.1) apparently has rich and important stat-physics meaning centered around so called *Bethe Approximation.Bethe Approximation* is also one of the main Heuristics in modern practice of Machine Learning, especially in inference on graphical models (it is quite rare for a Heuristic from Machine Learning to have such amazing proof power).

Although this stat-physics background was not used in the current paper, it and its developers(to be named in the final version) deserve a lot of praise: don't forget that many very good mathematicians have completely overlooked seemingly simple Theorem 2.2. It would be fantastic to have a rigorous and readable proof of Theorem 2.2 based on new(age) methods. The author is a bit skeptical at this point: any such proof would essentially reprove very hard Schrijver's permanental bound. The other avenue is to better understand and possibly to simplify the original Schrijver's proof, perhaps it has some deep stat-physics meaning.
It is possible that one can use higher order approximation(the Bethe Approximation being of order two, it involves marginals of subsets of cardinality two). Luckily, this order two case is covered by Schrijver's lower bound (5). The higher order cases will probably need new lower bounds (involving subpermanents?). It looks like a beginning of a beautiful(and hard) new line of research.
Our proof of Friedland's **monomer-dimer entropy** conjecture illustrates the power of Theorem 2.2. Interestingly, **monomer-dimer entropy** is the classical topic in stat-physics. The author is not a physicist,passionately so, even after 11 years at Los Alamos. Yet, there is a certain justice in the coincidence that some roots of this paper can be traced back to Hans Bethe...what a great group of creative people worked in New Mexico back then!

9 Acknowledgements

Without *Google Scholar* the author would have not stumbled on [10].
It is my great pleasure to thank Pascal Vontobel for many things, including catching quite a few typos in the previous version, but especially for stating Theorem(2.2).

References

[1] H.S. Wilf, On the permanent of a doubly stochastic matrix, Canad. J. Math. 18 (1966) 758-761.

[2] A. Schrijver and W.G.Valiant, On lower bounds for permanents, Indagationes Mathematicae 42 (1980) 425-427

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

[4] A. Schrijver, Matching, edge-colouring, dimers, in: "Graph-Theoretic Concepts in Computer Science" (H.L. Bodlaender, ed.), Lecture Notes in Computer Science 2880, Springer-Verlag, Berlin, 2003, pp. 13–22.

[5] M. Laurent and A. Schrijver, On Leonid Gurvits' proof for permanents, Amer. Math. Monthly 117 (2010), no. 10, 903-911.

[6] 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.

[7] 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).

[8] 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.

[9] L. Gurvits, Combinatorial and algorithmic aspects of hyperbolic polynomials, 2004; available at http://xxx.lanl.gov/abs/math.CO/0404474.

[10] P.O. Vontobel, The Bethe permanent of a non-negative matrix, in Proc. of Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on, 29 2010-Oct. 1 2010; available at Pascal Vontobel home page.

[11] Yusuke Watanabe and Michael Chertkov, Belief propagation and loop calculus for the permanent of a non-negative matrix, Journal of Physics A: Mathematical and Theoretical Volume 43 Number 24, 2010.

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

[13] 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).

[14] L.M. Bregman, Some properties of nonnegative matrices and their permanents, Soviet Math. Dokl, 1973.

[15] S. Friedland, E. Krop, Per H. Lundow, K. Markstrm; Validations of the Asymptotic Matching Conjectures, arXiv:math/0603001, 2006.

[16] B. Huang, T. Jebara; Approximating the Permanent with Belief Propagation, arXiv:0908.1769, 2009.

[17] L. Gurvits, A proof of hyperbolic van der Waerden conjecture : the right generalization is the ultimate simplification, Electronic Colloquium on Computational Complexity (ECCC)(103): (2005) and arXiv:math/0504397.

[18] L. Gurvits, Hyperbolic polynomials approach to Van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications, *Proc. 38 ACM Symp. on Theory of Computing (StOC-2006)*,417-426, ACM, New York, 2006.

[19] L. Lu, A. Mohr, L. Szekely, Quest for Negative Dependency Graphs, preprint, 2012.

[20] M. Voorhoeve, A lower bound for the permanents of certain (0,1) matrices, Indagationes Mathematicae 41 (1979) 83-86.

[21] I.M. Wanless, Addendum to Schrijver's work on minimum permanents, Combinatorica 26 (6) (2006) 743-745.

[22] B.D. McKay, X. Wang, Asymptotic enumeration of 0-1 matrices with equal row sums and equal column sums, Linear Algebra and its Applications; Volume 373, 2003, Pages 273-287.

[23] B.D. McKay, Asymptotics for 0-1 matrices with prescribed line sums, Enumeration and Design, Academic Press, Canada (1984), pp. 225-238

ECCC ISSN 1433-8092

http://eccc.hpi-web.de