A NEW PROOF OF VAN DER WAERDEN'S INEQUALITY, AND BEYOND

SUVRIT SRA*

Abstract. We give a new, self-contained proof of van der Waerden's permanent inequality via a monotone-averaging argument. Specifically, we show that the minimum total correlation of a random permutation with prescribed doubly stochastic marginals never decreases when two rows or two columns of the marginal matrix are averaged. This quantity is largest at the uniform matrix, which yields van der Waerden's bound. The monotone-averaging rests on the quadratic Alexandrov–Fenchel inequality, which lets us extend the argument to many other problems. In this way we also prove the Friedland–Tverberg inequalities and Gurvits's coefficient bound for Lorentzian polynomials.

hexagon:2611.00118v1

1. Introduction

Van der Waerden's inequality states that every doubly stochastic $n \times n$ matrix A satisfies $\operatorname{per} A \geq n!/n^n$, with equality only at the matrix J_n/n whose entries all equal $1/n$. Falikman (1981) and Egorychev (1981) proved it through the structure of minimizers and the quadratic Alexandrov–Fenchel inequality, and Gurvits (2008) later gave a proof by capacity induction for real stable polynomials.

The permanent itself need not decrease as A moves to J_n/n along a straight segment (Cheon and Wanless, 2005, Problem 8), so a monotone proof must follow a different quantity. We look for one that also measures how far a matrix is from the extremal one, and whose monotonicity rests on an inequality available beyond permanents.

We give a short, self-contained proof of van der Waerden's inequality by a new monotonicity theorem. The quantity we follow is the minimum total correlation of a random permutation σ with prescribed marginal matrix, that is, the smallest value of $\sum_i \mathrm{H}(\sigma(i)) - \mathrm{H}(\sigma)$ over all such σ, where H is Shannon entropy. This difference measures how far the positions $\sigma(1), \ldots, \sigma(n)$ are from independent. We prove that the minimum never decreases when two rows or two columns of the marginal matrix are averaged (Theorem 2.1). Since finitely many averages reach J_n/n, it is largest there, and its maximum implies van der Waerden's inequality and its equality case.

We show that the derivative along an average is nonnegative because the permanent cofactors of the two averaged rows satisfy the quadratic Alexandrov–Fenchel inequality, the same input that Falikman and Egorychev used. Nothing else in the argument is specific to permanents, so we can apply it to every form with the same quadratic property, including the mixed derivatives of Lorentzian polynomials. At every step the argument keeps a nonnegative remainder.

Main contributions of the paper.

(1) A new monotonicity theorem (Theorem 2.1), which shows that the minimum total correlation of a random permutation with prescribed doubly stochastic marginals never decreases when two rows or two columns of the marginal matrix are averaged; it gives a short proof of van der Waerden's inequality with its equality case (Corollary 2.6).

2020 *Mathematics Subject Classification.* Primary 15A15; Secondary 52A39, 94A17, 05B35.

Key words and phrases. permanent, van der Waerden, entropy, Alexandrov–Fenchel, Lorentzian, capacity, volumes.

* TU Munich, Dept. of Mathematics, School of CIT; Garching, Germany. Email: s.sra@tum.de.

First public release: 31 October 2026, as v1 in the author's preprint repository, https://github.com/suvrit/preprints/tree/main/preprints/gurvits-lower; every version is kept there with its timestamp.

(2) Averaging proofs of the Friedland–Tverberg inequalities and their equality case (Theorem 3.1) and of Gurvits's coefficient bound for Lorentzian polynomials (Corollary 4.3). The latter gives lower bounds for weighted counts of matroid bases and for mixed volumes, and hence for the number of roots of generic polynomial systems (Section 5).
(3) Remainders for these bounds. The change along each average splits exactly into nonnegative entropy-production terms (Theorem 4.4), and the bounds at every order give explicit corrections; in particular, per A exceeds $n!/n^n$ by at least a multiple of $\|A-J_n/n\|_F^2$ whose constant depends only on n and the smallest entry of A (Corollary 5.3; Proposition 5.2 for mixed discriminants).

A second mechanism averages binary couplings toward independence and gives the stable-pairing bound of Anari and Oveis Gharan with an exact entropy remainder, which improves for two homogeneous polynomials of equal degree (Theorem 6.3 and Proposition 6.4). For permanents it needs no stability and corrects the Bethe lower bound, exactly on unions of even cycles; for regular bipartite graphs it improves Schrijver's bound by a factor exponential in the number of components (Corollary 6.6). Under a spectral gap, the averaging remainders are at least a multiple of the squared distance to equal rows (Corollary 4.5).

1.1. **Related work.** Table 1 lists the inequalities we prove, with the earlier bound or input for each. The endpoints are known, and we credit each where we use it; the averaging proofs and their remainders are new. The entropy remainders are exact but implicit, since evaluating them requires the maximum-entropy law, while those in item (3), Proposition 6.4, and Corollary 6.6 are explicit.

Table 1: Inequalities proved by averaging.

Setting	Inequality	Here	Earlier bound or input
Permanents	$\operatorname{per} A \geq n!/n^n$	Cor. 2.6	Egorychev (1981), Falikman (1981)
Partial matchings	Friedland–Tverberg with remainder	Thm. 3.1	Friedland (1982)
Lorentzian f	$f_\alpha \geq \binom{d}{\alpha}\alpha^\alpha d^{-d}\operatorname{Cap}_\alpha(f)$	Cor. 4.3	Gurvits (2009a)
	Product capacity with remainder	Cor. 4.7	Gurvits (2009a)
	Squarefree-coefficient sum	Cor. 5.7	Anari et al. (2017) (stable case)
Mixed volumes	Volume-capacity lower bound	(5.2)	Gurvits (2009b)
Mixed discriminants	Determinant-capacity lower bound	§5.1	Gurvits (2008)
	Explicit quadratic remainder	Prop. 5.2, Cor. 5.3	Gurvits (2006), Friedland and Gurvits (2006)
Matroid bases	Ordered entropy with remainder	Cor. 5.4	Coefficient bound and Anari et al. (2021)
	Weighted assignments; rainbow bases	Cor. 5.5, Cor. 5.6	Same inputs
Stable pairings	Capacity bound with entropy remainder	Thm. 6.3	Anari and Oveis Gharan (2017)
Permanents	Bethe bound with explicit correction	Cor. 6.6	Schrijver (1998), Gurvits (2012)

The closest antecedents of Theorem 2.1 concern different operations. Rothaus (1972) and Brégman (1973) studied maximum-entropy permutation laws with prescribed marginals and compared permanents through them. Cryan et al. (2021) derive entropy contraction from the same quadratic condition, and Lee and Seo (2026) compare exchangeable Lorentzian laws under coordinate deletion. Theorem 2.1 instead compares the minimum total correlation before and after each pair average. By Remark 2.8, its endpoint at J_n/n is equivalent to the capacity lower bound of Gurvits (2008), and the theorem adds monotonicity along the whole path.

Anari and Oveis Gharan (2017) proved the pairing bound for real stable polynomials and derived Schrijver's inequality from it (their Section 3.1). The variational form of the Bethe lower bound is

due to Gurvits (2012, Theorem 2.2), and Anari (2026) improves it by pairing rows. In Section 6.2 we relate these results to ours.

The companion paper gives a rowwise upper bound on the maximum entropy of a random permutation with prescribed marginals (Sra, 2026a, Theorem 5.4). The supplement (Sra, 2026b) transfers the Orlicz-norm bound of that paper to mixed derivatives of polynomials and collects further consequences of the averaging arguments.

2. Pair averaging and van der Waerden's inequality

The monotone quantity comes from the entropy of random permutations. For a doubly stochastic matrix $P \in \mathcal{B}_n$, let $\mathcal{M}(P)$ be the set of probability laws on the symmetric group $\mathfrak{S}_n$ whose marginal matrix is P, that is, under which $\sigma(i) = j$ with probability p_{ij}, and put

$$\mathsf{H}_{\mathrm{perm}}(P) = \max_{\mu \in \mathcal{M}(P)} \mathrm{H}(\mu), \qquad \mathcal{S}(P) = -\sum_{i,j} p_{ij} \log p_{ij}, \qquad \mathcal{I}(P) = \mathcal{S}(P) - \mathsf{H}_{\mathrm{perm}}(P),$$

where H denotes Shannon entropy, with natural logarithms and $0 \log 0 = 0$. Thus $\mathcal{I}(P)$ is the smallest total correlation that a random permutation with marginal matrix P can have. Averaging rows r and s replaces P by $T_{rs}(t)P$, where $T_{rs}(t) = (1-t)I_n + tS_{rs}$ and S_{rs} interchanges rows r and s.

Theorem 2.1 (Pair-averaging monotonicity)**.** *For every $P \in \mathcal{B}_n$, the function*

$$t \longmapsto \mathcal{I}(T_{rs}(t)P)$$

is nondecreasing on $[0, 1/2]$. The same assertion holds for right multiplication by a pair column-averaging matrix. Hence $\mathcal{I}$ is nondecreasing under every finite chain of pair row and column T-transforms.

Both $\mathcal{S}$ and $\mathsf{H}_{\mathrm{perm}}$ are nondecreasing along such averages, by concavity of entropy; for $\mathsf{H}_{\mathrm{perm}}$, mixing a maximizing law with its image under the transposition of rows r and s shows this. Theorem 2.1 says that $\mathcal{S}$ grows at least as fast.

The monotonicity proof differentiates $\mathcal{I}$ along an averaging path, so it needs $\mathcal{I}$ to be smooth with an explicit differential. Both come from the Gibbs form of the maximizing law. Maximum-entropy permutation laws with fixed marginals are classical (Brégman, 1973, Theorem 2 and equation (9)); see also Slater (1987); Verducci (1989); Agrawal et al. (2008). We give the details needed here.

The quantity $\mathcal{I}$ is well defined and nonnegative. Since P has unit row and column sums, Hall's condition holds on its positive support, and successively subtracting supported permutation matrices with positive weights gives a law in $\mathcal{M}(P)$. Compactness and strict concavity give a unique entropy maximizer, and entropy subadditivity gives $\mathcal{I}(P) \geq 0$. We write $U = J_n/n$ for the uniform matrix and $\mathcal{V} = \{X \in \mathbb{R}^{n \times n} : X\mathbf{1} = X^T\mathbf{1} = 0\}$ for the directions tangent to $\mathcal{B}_n$, and for $B > 0$ we let $\mu_B(\sigma) = \prod_i b_{i,\sigma(i)} / \operatorname{per} B$ be the Gibbs law of B.

Lemma 2.2 (Maximum-entropy parametrization)**.** *Every positive $P \in \mathcal{B}_n$ is the marginal matrix of a positive Gibbs law μ_B, which is its unique maximum-entropy law. The matrix B can be chosen smoothly in P with $B(U) = J_n$, and $\mathsf{H}_{\mathrm{perm}}$ and $\mathcal{I}$ are smooth in the relative interior of $\mathcal{B}_n$. For $X \in \mathcal{V}$,*

$$d\mathcal{I}(P)[X] = \sum_{i,j} X_{ij} (\log b_{ij} - \log p_{ij}). \tag{2.1}$$

Proof. For $P > 0$, write $P = \delta U + (1-\delta)R$ with $\delta > 0$ and $R \in \mathcal{B}_n$. Mixing a law for R with the uniform law gives a positive law in $\mathcal{M}(P)$. An entropy maximizer cannot have a zero mass, since moving toward this law would have infinite positive entropy derivative. Lagrange multipliers therefore give μ_B, with $b_{ij} = e^{\theta_{ij}}$. Conversely, relative entropy shows that every Gibbs law maximizes entropy among laws with its marginals.

Put $F(\theta) = \log \operatorname{per}(e^{\theta_{ij}})$, with entrywise exponentiation. Its gradient is the marginal matrix and its Hessian is the covariance of the permutation matrix. This covariance is positive definite on $\mathcal{V}$, because a constant statistic $\sum_i X_{i,\sigma(i)}$ has zero two-by-two alternating differences, so X is a sum of row and column potentials, and hence is zero if $X \in \mathcal{V}$. Restricting θ to $\mathcal{V}$, the inverse function theorem gives a smooth choice of B, with $B(U) = J_n$. Differentiating $\mathsf{H}_{\mathrm{perm}}(P) = F(\theta) - \langle P, \theta \rangle$ gives $d\mathsf{H}_{\mathrm{perm}}(P)[X] = -\langle \theta, X \rangle$, which proves (2.1). □

2.1. **The local entropy comparison.** Along a pair-averaging path, (2.1) reduces to an expression in a reversible stochastic matrix built from the cofactors of the two selected rows. Two Jensen inequalities control its sign.

Lemma 2.3 (Entropy and the quadratic spectral condition)**.** *Let T be a reversible stochastic $m \times m$ matrix with a strictly positive stationary probability vector π, and write $\langle u, v \rangle_\pi = \sum_j \pi_j u_j v_j$. Suppose*

$$\langle z, Tz \rangle_\pi \le \langle z, \mathbf{1} \rangle_\pi^2 \qquad (z \in \mathbb{R}^m). \tag{2.2}$$

Then, for every positive vector z,

$$\langle Tz - z, \log(Tz) \rangle_\pi \ge 0. \tag{2.3}$$

Equality holds exactly when Tz is constant.

Proof. Set $a = \langle z, \mathbf{1} \rangle_\pi$. Jensen's inequality with weights $\pi_j z_j / a$ and then Jensen's inequality for $s \log s$ give

$$\langle z, \log(Tz) \rangle_\pi \le a \log \frac{\langle z, Tz \rangle_\pi}{a} \le a \log a \le \langle Tz, \log(Tz) \rangle_\pi.$$

Stationarity gives $\langle Tz, \mathbf{1} \rangle_\pi = a$, so the last inequality is strict unless $Tz = a\mathbf{1}$. In that case all three are equalities. □

The cofactor matrices of the permanent satisfy the hypothesis of Lemma 2.3 because of the quadratic Alexandrov–Fenchel inequality (Falikman, 1981; Egorychev, 1981). We include a proof for the particular cofactor matrices in use, so that no permanent lower bound is assumed.

Lemma 2.4 (Cofactor Alexandrov–Fenchel inequality)**.** *Let $n \ge 2$ and $c_1, \dots, c_{n-2} \in (0, \infty)^n$. Let M be the symmetric matrix with diagonal zero and, for $j \ne k$,*

$$M_{jk} = \operatorname{per}\big((c_\ell)_{\ell=1}^{n-2} \text{ with columns } j, k \text{ deleted}\big).$$

For $n = 2$, the empty permanent is 1. Then

$$(x^T M y)^2 \ge (x^T M x)(y^T M y) \qquad (x \in \mathbb{R}^n,\ y > 0). \tag{2.4}$$

In particular M has at most one positive eigenvalue.

Proof. A real polynomial is real stable if it has no zero when all variables have positive imaginary parts. Positive directional differentiation preserves stability for homogeneous polynomials of positive degree with nonnegative coefficients. Indeed, if q is such a stable polynomial and $u > 0$, all roots α of $q(z + tu)$ have strictly negative imaginary parts when $\operatorname{Im} z > 0$. Its leading coefficient is $q(u) > 0$, and $D_u q(z)/q(z) = \sum_\alpha (-\alpha)^{-1}$ has negative imaginary part. Starting with $z_1 \cdots z_n$ and differentiating therefore gives the real stable quadratic

$$q(z) = D_{c_1} \cdots D_{c_{n-2}} (z_1 \cdots z_n) = \frac{1}{2} z^T M z.$$

For real x and positive y, stability and conjugacy make every root of $q(x + ty)$ real. Its nonnegative discriminant is $(x^T M y)^2 - (x^T M x)(y^T M y)$, proving (2.4). The final assertion follows by restricting the quadratic form to $\{x : x^T M y = 0\}$. □

Proof of Theorem 2.1. Assume $n \geq 2$; the case $n = 1$ is immediate. For positive P, Lemma 2.2 supplies Gibbs parameters, and we put the derivative along the averaging path into the form of Lemma 2.3. Let B be the Gibbs matrix of P, let x, y be its two selected rows, and let M be the cofactor matrix of the other rows from Lemma 2.4. Define

$$C = \operatorname{diag}(y) M \operatorname{diag}(y), \quad \varrho = C\mathbf{1}, \quad T = \operatorname{diag}(\varrho)^{-1} C, \quad z = x/y, \quad Z = x^T M y = \varrho^T z.$$

All divisions here are coordinatewise. The matrix T is stochastic and reversible for $\pi = \varrho/(\mathbf{1}^T\varrho)$. Congruence and (2.4) give $w^T C w \leq (w^T C\mathbf{1})^2/(\mathbf{1}^T C\mathbf{1})$ for real w, which normalizes to (2.2).

Since $\operatorname{per} B = x^T M y$, the corresponding rows u, v of P are

$$u_j = \frac{\varrho_j z_j}{Z}, \qquad v_j = \frac{\varrho_j (Tz)_j}{Z}, \qquad \log\frac{x_j}{u_j} - \log\frac{y_j}{v_j} = \log(Tz)_j.$$

In the direction changing only these two rows, (2.1) and Lemma 2.3 give

$$d\mathcal{I}(P)[(v-u, u-v)] = \frac{1}{Z}\sum_j \varrho_j\big((Tz)_j - z_j\big)\log(Tz)_j \geq 0. \tag{2.5}$$

At path parameter $t < 1/2$, the current row difference is $1-2t$ times the initial one, so the derivative along the original path is (2.5) divided by $1-2t$; continuity gives the endpoint $t = 1/2$. Since $T_{rs}(t) = S_{rs}T_{rs}(1-t)$ and $\mathcal{I}$ does not change when two rows are interchanged, $\mathcal{I}(T_{rs}(t)P) \geq \mathcal{I}(P)$ for every $t \in [0,1]$, which gives the chain assertion. Inversion of a random permutation shows $\mathsf{H}_{\mathrm{perm}}(P^T) = \mathsf{H}_{\mathrm{perm}}(P)$, while $\mathcal{S}(P^T) = \mathcal{S}(P)$, so the row assertion for P^T gives the column assertion.

For boundary P, set $P_\varepsilon = (1-\varepsilon)P + \varepsilon U$. Mixing feasible laws shows that $\mathsf{H}_{\mathrm{perm}}$ is concave, giving its lower radial limit. A convergent subsequence of maximizing laws gives the matching upper limit by continuity of entropy. Thus $\mathcal{I}(P_\varepsilon) \to \mathcal{I}(P)$. Since $T_{rs}(t)U = U$, the same radial argument applies after averaging, and passage to the limit proves the boundary statement. □

2.2. **The uniform endpoint.** Every doubly stochastic matrix reaches U through finitely many pair averages, so Theorem 2.1 makes U the maximizer of $\mathcal{I}$. The equality condition in Lemma 2.3 makes it the unique one.

Theorem 2.5 (Unique uniform maximizer)**.** *For $n \geq 2$,*

$$0 \leq \mathcal{I}(P) \leq \kappa_n := n\log n - \log(n!), \qquad P \in \mathcal{B}_n.$$

Equality in the upper bound holds precisely at $U = J_n/n$.

Proof. A finite chain of pair averages takes any list of rows to its mean. Inductively average the first $n-1$ rows to a common row u, and call the last row v. Successively mix the last row with the unused copies of u, using parameters $1/n, 1/(n-1), \ldots, 1/2$. Each step fixes another row at $((n-1)u+v)/n$, and double stochasticity makes this row uniform. Theorem 2.1 therefore gives $\mathcal{I}(P) \leq \mathcal{I}(U) = \kappa_n$. Here the uniform law gives $\mathsf{H}_{\mathrm{perm}}(U) = \log(n!)$, while $\mathcal{S}(U) = n\log n$.

For uniqueness, it suffices that the derivative (2.5) is positive near U whenever the two selected rows differ. At U, the smooth Gibbs choice in Lemma 2.2 gives $T = (J_n - I_n)/(n-1)$, which is invertible. It stays invertible near U. By Lemma 2.3, zero production in (2.5) there forces Tz constant, hence z constant and $u = v$. On the last nontrivial segment of a chain ending at U, the selected rows are unequal before the endpoint. Production is therefore strictly positive near that endpoint, giving $\mathcal{I}(P) < \mathcal{I}(U)$ whenever $P \neq U$, including boundary P. □

Corollary 2.6 (Van der Waerden's bound)**.** *For every doubly stochastic A,*

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

with equality precisely at J_n/n.

Proof. Let μ maximize entropy with marginals A, and put $\nu(\sigma) = \prod_i a_{i,\sigma(i)}/\operatorname{per} A$. Hall's argument gives $\operatorname{per} A > 0$, and ν is positive wherever μ is, even with zero entries. Jensen's inequality gives

$$0 \le \sum_\sigma \mu(\sigma) \log \frac{\mu(\sigma)}{\nu(\sigma)} = \log \operatorname{per} A + \mathcal{I}(A).$$

Theorem 2.5 gives $\log \operatorname{per} A \ge -\mathcal{I}(A) \ge -\kappa_n$, with equality only at U. Direct evaluation gives $\operatorname{per} U = n!/n^n$. The case $n = 1$ is immediate. □

Remark 2.7 (The spectral converse). For irreducible reversible T, the entropy inequality (2.3) for every $z > 0$ implies (2.2). Indeed, for $z = \mathbf{1} + \varepsilon h$ with $\langle h, \mathbf{1}\rangle_\pi = 0$, the left side is

$$\varepsilon^2 \langle Th, (T - I)h\rangle_\pi + O(\varepsilon^3).$$

On an eigenspace of eigenvalue λ, the quadratic coefficient is $\lambda(\lambda - 1)\|h\|_\pi^2$. Stochasticity gives $-1 \le \lambda \le 1$, and irreducibility makes 1 simple, so universal nonnegativity excludes $0 < \lambda < 1$.

Permanent kernels are irreducible because their off-diagonal entries are positive. If we fix the other Gibbs rows and $y > 0$ and vary $x = y \odot z$, then $z > 0$ is arbitrary in (2.5). Thus pair-averaging monotonicity for every interior marginal matrix recovers the cofactor Alexandrov–Fenchel condition. The quadratic input to the local comparison is therefore necessary as well as sufficient.

Remark 2.8. The endpoint $\mathcal{I} \le \kappa_n$ is equivalent to the capacity lower bound

$$\operatorname{per} B \ge \frac{n!}{n^n} \inf_{y>0} \frac{\prod_i \sum_j b_{ij} y_j}{\prod_j y_j} \qquad (B > 0)$$

of Gurvits (2008, Corollary 2.5 and Example 2.6); the permanent is the case $f = z_1 \cdots z_n$ of the equivalence proved for Lorentzian polynomials in Section 4.2. The endpoint alone thus recovers Gurvits's bound, while Theorem 2.1 compares $\mathcal{I}$ at successive points of every pair-averaging path.

3. Partial matchings by averaging within a face

Van der Waerden's bound counts perfect matchings, and the Friedland–Tverberg inequalities extend it to matchings of every size. We prove them, with their equality case and an entropy remainder, by averaging inside a face of a larger permutation polytope. For a nonnegative $n \times n$ matrix A, let

$$\sigma_k(A) = \sum_{\substack{R,C \subseteq [n] \\ |R| = |C| = k}} \operatorname{per} A[R, C].$$

For a bipartite adjacency matrix, this is the number of matchings of size k. Friedland (1982) proved that σ_k is uniquely minimized over doubly stochastic matrices at $U = \mathbf{1}\mathbf{1}^T/n$ for $2 \le k \le n$. Friedland and Gurvits subsequently obtained a generalization for positive hyperbolic polynomials (Friedland and Gurvits, 2006).

To bring Theorem 2.1 to bear, we embed matchings into permutations. Applied to prescribed marginal matrices, the completion construction of Anari et al. (2017, Appendix A, proof of Corollary 2.7) turns a law on matchings of size k into a permutation law on a fixed face of the permutation polytope, and pair averages of A preserve that face.

The averaging argument then needs a functional on matchings that plays the role of $\mathcal{I}$. Fix $2 \le k \le n$. For a doubly stochastic A, let $\mathsf{H}_k(A)$ be the maximum entropy of a law on size-k matchings whose edge marginals are $(k/n)A$, and put

$$\mathcal{F}_k(A) = \frac{k}{n}\mathcal{S}(A) - \mathsf{H}_k(A).$$

Selecting k edges uniformly from a random permutation with marginals A gives such a law. Like $\mathcal{I}$ for permutations, $\mathcal{F}_k$ compares the entropy of a random matching with the entropy of its prescribed edge marginals.

Theorem 3.1 (Averaging for partial matchings)**.** *Let $2 \le k \le n$. The functional $\mathcal{F}_k$ is nondecreasing under pair averaging of the rows or columns of a doubly stochastic matrix. Moreover,*

$$\mathcal{F}_k(A) \le \mathcal{F}_k(U) = -\log\left(\binom{n}{k}^2 \frac{k!}{n^k}\right),$$

with equality if and only if $A = U$. Consequently, writing $\Delta_{n,k}(A) = \mathcal{F}_k(U) - \mathcal{F}_k(A)$, one has

$$\sigma_k(A) \ge \binom{n}{k}^2 \frac{k!}{n^k} \exp(\Delta_{n,k}(A)), \qquad \Delta_{n,k}(A) \ge 0. \tag{3.1}$$

Equality in the Friedland–Tverberg bound $\sigma_k(A) \ge \binom{n}{k}^2 k!/n^k$ holds exactly at U.

Proof. ***Completion.*** To apply Theorem 2.1, we complete each size-k matching to a permutation by adding $h = n - k$ rows and columns. Let $N = n + h$, let $J_{a,b}$ be the $a \times b$ all-ones matrix, and set

$$P_k(A) = \begin{pmatrix} (k/n)A & J_{n,h}/n \\ J_{h,n}/n & 0_{h,h} \end{pmatrix}.$$

Blocks of size zero are omitted. This matrix is doubly stochastic. The zero lower-right block restricts permutations to a set Ω. Each allowed permutation uses exactly k upper-left entries, and each size-k matching there has $(h!)^2$ completions. Uniform conditional completion therefore adds $2\log(h!)$ to entropy. It also gives the prescribed marginals in the off-diagonal blocks, because each original vertex is unmatched with probability h/n and its added partner is uniform among h choices. The entropy chain rule now gives

$$\mathsf{H}_{\mathrm{perm}}(P_k(A)) = \mathsf{H}_k(A) + 2\log(h!), \qquad \mathcal{S}(P_k(A)) = \frac{k}{n}\mathcal{S}(A) - k\log(k/n) + 2h\log n.$$

Thus $\mathcal{I}(P_k(A)) - \mathcal{F}_k(A)$ is independent of A.

Monotonicity and the endpoint. A pair average of rows or columns of A acts on the corresponding first n rows or columns of $P_k(A)$. Applying Theorem 2.1 in dimension N proves monotonicity of $\mathcal{F}_k$, including at zero entries. A finite chain takes the first n rows to their common mean, with endpoint $P_k(U)$. At U, the uniform law on all $\binom{n}{k}^2 k!$ matchings has the required edge marginals, so

$$\mathsf{H}_k(U) = \log\left(\binom{n}{k}^2 k!\right), \qquad \frac{k}{n}\mathcal{S}(U) = k\log n.$$

These values give $\mathcal{F}_k(U)$, and the monotonicity gives $\Delta_{n,k}(A) \ge 0$.

The matching bound. The functional $\mathcal{F}_k$ controls σ_k through the Gibbs variational principle. Let μ maximize the matching entropy with edge marginals $(k/n)A$. Every matching in its support uses only positive entries of A. Jensen's inequality therefore gives

$$\log \sigma_k(A) \ge \mathsf{H}_k(A) + \frac{k}{n}\sum_{i,j} a_{ij}\log a_{ij} = -\mathcal{F}_k(A).$$

The same support argument applies at zero entries, and the feasible law ensures $\sigma_k(A) > 0$. Substituting the definition of $\Delta_{n,k}$ proves (3.1).

Strictness. As in the proof of Theorem 2.5, it suffices that the entropy derivative is positive near the endpoint whenever the selected rows differ, so we apply Lemma 2.3 on the relative interior of the face with support Ω. The uniform law on Ω has marginal matrix $P_k(U)$, so this endpoint lies in that relative interior. There the entropy maximizer has positive Gibbs parameters on all allowed entries and zero parameters in the lower-right block. These parameters can be chosen smoothly near $P_k(U)$, with value one on every allowed entry at the endpoint. Indeed, after removing the affine relations on the incidence vectors, the Hessian of the log partition function is a positive definite covariance matrix, so the inverse function theorem applies. For selected Gibbs rows x, y and cofactor matrix

M, the kernel in the zero-production criterion is $T = \operatorname{diag}(C\mathbf{1})^{-1}C$, with $C = \operatorname{diag}(y)M\operatorname{diag}(y)$ and $z = x/y$ coordinatewise. At the endpoint $x = y = \mathbf{1}$, this simplifies to $T = \operatorname{diag}(M\mathbf{1})^{-1}M$.

Select two of the first n rows. At the endpoint, their two-slot cofactor matrix M has zero diagonal and three off-diagonal values:

$$a = \frac{((n-2)!)^2}{(k-2)!}, \qquad b = \frac{(n-1)!(n-2)!}{(k-1)!}, \qquad c = \frac{n!(n-2)!}{k!}.$$

They correspond respectively to two original columns, one original and one added column, and two added columns; absent blocks are omitted. To count the first case, assign the last h rows injectively to the $n-2$ remaining original columns, then assign the remaining $n-2$ rows bijectively to the remaining columns. The count is $(n-2)!/(k-2)!$ times $(n-2)!$; the other two cases replace $n-2$ available original columns by $n-1$ or n. The two row sums of M are

$$r_{\mathrm{o}} = \frac{((n-1)!)^2}{(k-1)!}, \qquad r_{\mathrm{a}} = \frac{n!(n-1)!}{k!}.$$

Consequently $T = \operatorname{diag}(M\mathbf{1})^{-1}M$ has eigenvalue $-(k-1)/(n-1)^2$ on the zero-sum vectors supported on the n original columns. On the zero-sum vectors supported on the h added columns its eigenvalue is $-1/(n-1)$. For $h > 0$, the remaining two-dimensional space of vectors constant on each column class has quotient matrix

$$\frac{1}{n-1}\begin{pmatrix} k-1 & h \\ k & h-1 \end{pmatrix},$$

whose eigenvalues are 1 and $-1/(n-1)$. Thus the complete spectrum is

$$\left\{1, \quad -\frac{k-1}{(n-1)^2} \text{ with multiplicity } n-1, \quad -\frac{1}{n-1} \text{ with multiplicity } h\right\}.$$

For $h = 0$, the second group of nontrivial eigenvalues is absent. In particular, the kernel is nonsingular for $k \geq 2$.

The kernel remains nonsingular near $P_k(U)$. Approximating the zero Gibbs parameters by positive ones shows that the cofactor Alexandrov–Fenchel inequality remains valid on this face. Its off-diagonal cofactors are positive for $k \geq 2$, so the kernel and the entropy derivative are defined there. The zero-production criterion in Lemma 2.3 forces Tz constant, hence z constant and the two selected marginal rows equal. If $A \neq U$, the last nonconstant segment of a finite averaging chain has unequal selected rows before it reaches $P_k(U)$. Near that endpoint all allowed marginal entries are positive, even if the starting matrix is on the boundary. Production is strictly positive on this part of the segment, which proves $\Delta_{n,k}(A) > 0$. Direct evaluation at U completes the equality statement. □

The remainder $\Delta_{n,k}(A)$ is the integral of the permanent entropy production along any chain averaging the first n rows within this face, and Theorem 4.4 below decomposes that production into nonnegative local terms. Corollary 5.3 gives an explicit correction to $\sigma_k(A)$ in the entries of A.

4. Averaging for Lorentzian polynomials

Mixed derivatives of Lorentzian polynomials are symmetric multilinear forms whose two-slot contractions satisfy the quadratic Alexandrov–Fenchel inequality, as the permanent does in (2.4). We extend Theorem 2.1 to them; at the uniform endpoint the extension becomes a capacity lower bound for their coefficients.

To extend the argument, we replace random permutations by random words whose weights are the coefficients of a polynomial. Throughout this section, let

$$f(z) = \sum_{\alpha \in \mathcal{A}} f_\alpha z^\alpha, \qquad f_\alpha > 0, \quad \mathcal{A} \subset \{\alpha \in \mathbb{N}^m : |\alpha| = d\},$$

be a nonzero homogeneous polynomial of degree $d \geq 2$, and let $B \geq 0$ have d rows $b_1, \dots, b_d$ and m columns. Its mixed derivative is

$$\mathcal{D}_f(B) = [t_1 \cdots t_d] f(B^T t) = D_{b_1} \cdots D_{b_d} f.$$

For $f = z_1 \cdots z_n$ it is the permanent.

Expanding $\mathcal{D}_f(B)$ over words gives the analogue of the permanent's sum over permutations. A word $\omega \in [m]^d$ has a multiplicity vector $\alpha(\omega)$, and we write $\alpha! = \prod_j \alpha_j!$. Let $\Omega_f = \{\omega \in [m]^d : \alpha(\omega) \in \mathcal{A}\}$. For a word ω, its incidence matrix has a single one in row i, column ω_i. Expansion gives

$$\mathcal{D}_f(B) = \sum_{\omega \in \Omega_f} w_\omega \prod_i b_{i,\omega_i}, \qquad w_\omega = \alpha(\omega)! f_{\alpha(\omega)}.$$

The polytope $\mathcal{K}_f$ is the convex hull of these incidence matrices. A weighted entropy and its deficit then play the roles of $\mathsf{H}_{\mathrm{perm}}$ and $\mathcal{I}$. For $P \in \mathcal{K}_f$, with rows p_i, put

$$\mathsf{H}_f(P) = \max_{\mu:\, \mathrm{marg}\mu = P} \{\mathrm{H}(\mu) + \mathbb{E}_\mu \log w_\omega\}, \qquad \mathcal{I}_f(P) = \mathcal{S}(P) - \mathsf{H}_f(P),$$

where $\mathcal{S}(P) = \sum_i \mathrm{H}(p_i)$ as before. For the permanent, $\mathsf{H}_f = \mathsf{H}_{\mathrm{perm}}$ and $\mathcal{I}_f = \mathcal{I}$. In general $\mathcal{I}_f$ need not be nonnegative, because the coefficient weights need not be probabilities. Pair averages of rows preserve the column sums $q = P^T \mathbf{1}$, so averaging ends at the matrix $\bar{P}$ whose rows all equal q/d. The weighted entropy H_f is concave. Its maximizing law exists by compactness and is unique by strict concavity. For $B > 0$, Gibbs' variational formula is

$$\log \mathcal{D}_f(B) = \max_{P \in \mathcal{K}_f} \{\langle P, \log B \rangle + \mathsf{H}_f(P)\}. \tag{4.1}$$

The uniform endpoint $\bar{P}$ is evaluated through capacity. Let $\mathrm{Newt}(g)$ denote the Newton polytope of a polynomial g, the convex hull of its exponent vectors. For a polynomial g with nonnegative coefficients and an exponent vector a,

$$\mathrm{Cap}_a(g) = \inf_{z > 0} \frac{g(z)}{z^a}.$$

If $a \in \mathrm{Newt}(g)$, finite-dimensional convex duality gives

$$\log \mathrm{Cap}_a(g) = \max_{\nu:\, \mathbb{E}_\nu \alpha = a} \left\{ \mathrm{H}(\nu) + \sum_\alpha \nu_\alpha \log g_\alpha \right\}. \tag{4.2}$$

This entropy interpretation of capacity is standard (Straszak and Vishnoi, 2017a, Appendix A). The formula follows by applying the log-sum-exp variational formula to $g(e^u)$ and minimizing after subtracting $a \cdot u$. On a boundary face, the same argument applies after restricting to that face, by compactness of the finite probability simplex. In particular, the logarithm in (4.2) is finite on $\mathrm{Newt}(g)$.

Proposition 4.1 (The equal-row entropy)**.** *For every nonnegative coefficient polynomial f and $q \in \mathrm{Newt}(f)$,*

$$\mathsf{H}_f(\bar{P}) = \log(d!) + \log \mathrm{Cap}_q(f), \qquad \bar{p}_i = q/d. \tag{4.3}$$

Moreover $\mathsf{H}_f(P) \leq \mathsf{H}_f(\bar{P})$ for every $P \in \mathcal{K}_f$ with column sums q.

Proof. We relax the marginal constraint to the expected multiplicity vector q. Symmetrizing a law over slot permutations preserves that constraint and its expected log weight, does not decrease entropy, and gives every slot marginal q/d. Thus the relaxation is exact. For a fixed count α, uniformity on its $d!/\alpha!$ words maximizes conditional entropy. The resulting objective for a count law ν is

$$\mathrm{H}(\nu) + \sum_\alpha \nu_\alpha \log(d! f_\alpha), \qquad \mathbb{E}_\nu \alpha = q.$$

Formula (4.2) proves (4.3), including boundary exponents. The same symmetrization starting from a maximizing law for P proves the final assertion. □

Two established properties of homogeneous Lorentzian polynomials enter the averaging argument. Their support is M-convex, and every nonzero quadratic obtained from them by positive directional differentiation has at most one positive eigenvalue. Nonnegative linear substitutions, directional differentiation, and products preserve the class once the zero polynomial is admitted. Homogeneous stable polynomials, volume polynomials, and matroid basis generating polynomials belong to it (Brändén and Huh, 2020, Proposition 2.2, Theorems 2.10, 2.25, 2.30, 3.10, and 4.1, and Corollaries 2.11 and 2.32).

If we fix positive vectors in all but two slots of $\mathcal{D}_f$, the remaining symmetric bilinear form is given by a matrix M, the two-slot contraction. If f is Lorentzian, this matrix is nonnegative and has at most one positive eigenvalue. For positive y, it follows that

$$(x^T My)^2 \geq (x^T Mx)(y^T My) \qquad (x \in \mathbb{R}^m). \tag{4.4}$$

Indeed, the quadratic form is nonpositive on the M-orthogonal complement of y. The value $y^T My$ is positive because $f \neq 0$ and the contracting vectors are positive. Conversely this inequality implies the one-positive-eigenvalue condition. Thus the kernel hypothesis is exactly a quadratic Alexandrov–Fenchel hypothesis.

Theorem 4.2 (Averaging for Lorentzian polynomials)**.** *If f is Lorentzian and S_{rs} interchanges rows r, s, then*

$$t \longmapsto \mathcal{I}_f\big(((1-t)I + tS_{rs})P\big)$$

is nondecreasing for $0 \leq t \leq 1/2$. In particular,

$$\mathcal{I}_f(P) \leq \mathcal{I}_f(\bar{P}), \tag{4.5}$$

and consequently the weighted entropy satisfies

$$\mathsf{H}_f(P) \geq \log \mathrm{Cap}_q(f) + \log(d!) + \mathcal{S}(P) - d\,\mathrm{H}(q/d). \tag{4.6}$$

Proof. ***Gibbs parameters.*** As for permanents, we first parametrize interior marginal matrices by Gibbs laws. Let P_0 be the marginal matrix of the law proportional to w_ω, which is positive on Ω_f; by slot symmetry its rows are identical. Every P in the relative interior of $\mathcal{K}_f$ is $(1-\varepsilon)P' + \varepsilon P_0$ for some $P' \in \mathcal{K}_f$ and $\varepsilon > 0$, so mixing a law for P' with this law gives a law for P that is positive on Ω_f. The entropy maximizer must also be positive, since moving towards a positive law has one-sided entropy derivative $+\infty$ at any zero coordinate. Lagrange multipliers therefore give

$$\mu_B(\omega) = \frac{w_\omega \prod_i b_{i,\omega_i}}{\mathcal{D}_f(B)}, \qquad B > 0. \tag{4.7}$$

The Hessian of $\log \mathcal{D}_f(e^\theta)$ is the covariance of the incidence vector. It is positive definite on the span of differences of incidence vectors, because a zero-variance linear functional is constant on the support and hence annihilates this span. After these affine gauge directions are removed, the inverse function theorem proves smoothness of H_f in the relative interior and gives

$$d\mathcal{I}_f(P)[X] = \sum_{i,j} X_{ij}(\log b_{ij} - \log p_{ij}) \tag{4.8}$$

for tangent directions X. Every active marginal coordinate is positive there. The value is independent of the choice of Gibbs parameters, because their nonuniqueness consists of affine relations on the marginal polytope.

The local comparison. In the relative interior we use the Gibbs parameters (4.7) and bring the derivative into the form of Lemma 2.3. Let the two selected parameter rows be x, y, and let M

be the contraction of the other slots. Define

$$C = \operatorname{diag}(y) M \operatorname{diag}(y), \quad \varrho = C\mathbf{1}, \quad T = \operatorname{diag}(\varrho)^{-1} C, \quad \pi = \frac{\varrho}{\mathbf{1}^T \varrho}, \quad z = x/y. \tag{4.9}$$

Every ϱ_j is positive, because every retained coordinate occurs in the word support, slot symmetry places it in the selected slot, and all other parameter entries are positive. The matrix T is stochastic and reversible for π. Congruence and (4.4) give (2.2). Write $a = \langle z, \mathbf{1} \rangle_\pi$. The two marginal rows are

$$u_j = \frac{\pi_j z_j}{a}, \qquad v_j = \frac{\pi_j (Tz)_j}{a}. \tag{4.10}$$

Their gradient difference in (4.8) is $\log(Tz)_j$, so the swap-direction derivative is

$$d\mathcal{I}_f(P)[(v-u, u-v)] = \frac{1}{a} \langle Tz - z, \log(Tz) \rangle_\pi \geq 0. \tag{4.11}$$

At a point with path parameter $t < 1/2$, the original direction is the current swap direction divided by $1 - 2t$, which proves monotonicity on the open path; continuity gives its endpoint.

Boundary and endpoint. For boundary P, we approximate by $(1-\varepsilon)P + \varepsilon P_0$, which lies in the relative interior. Pair averages fix P_0, so they act on this approximation through P alone. Concavity of the maximum weighted entropy gives its lower radial limit. A convergent subsequence of maximizing laws gives the matching upper limit, by continuity of entropy on the finite support. Passing to the limit proves the boundary inequality.

The finite averaging construction in the proof of Theorem 2.5 takes any list of rows to their common mean. Applied here, it proves (4.5). Combining (4.5) with Proposition 4.1 gives (4.6). □

Only the two-slot quadratic condition (4.4) was used in the local comparison. For a nonnegative symmetric multilinear form, requiring this condition at every positive choice of the other slots is equivalent to its diagonal polynomial being Lorentzian (Brändén and Huh, 2020, Proposition 2.33 and Theorem 2.30).

The monotonicity need not be strict. For $f(z) = (a \cdot z)^d$ with $a > 0$, we have $\mathsf{H}_f(P) = \log(d!) + \mathcal{S}(P) + q \cdot \log a$, so $\mathcal{I}_f(P) = -\log(d!) - q \cdot \log a$ depends only on q, and every pair average leaves it unchanged.

Evaluating (4.6) at a deterministic word, we recover Gurvits's coefficient bound.

Corollary 4.3 (The Gurvits coefficient bound)**.** *For a degree-d Lorentzian polynomial and an integer $\alpha \in \mathbb{N}^m$ with $|\alpha| = d$,*

$$f_\alpha \geq \binom{d}{\alpha} \frac{\alpha^\alpha}{d^d} \operatorname{Cap}_\alpha(f), \qquad 0^0 = 1.$$

Proof. If $\alpha \in \mathcal{A}$, take a deterministic word of that count. Its marginal matrix has zero row entropy and weighted entropy $\log(\alpha! f_\alpha)$. The bound (4.6) gives the assertion. If $\alpha \notin \mathcal{A}$, M-convexity implies $\alpha \notin \operatorname{Newt}(f)$, so $\operatorname{Cap}_\alpha(f) = 0$ by a separating hyperplane and the assertion is trivial. □

Corollary 4.3 is an entropy derivation of a theorem of Gurvits (2009a, Corollary 3.6(2)); see also Gurvits and Leake (2026, Theorem 2.10). For stable polynomials there are stronger degree-sensitive factors (Gurvits, 2008). Gurvits and Leake (2026, Theorem 1.4) prove that the uniform polynomial is the unique minimizer of the all-ones coefficient among doubly stochastic Lorentzian polynomials, by a capacity induction that avoids the Alexandrov–Fenchel inequality.

4.1. **Entropy production.** The coefficient bound of Corollary 4.3 uses only the endpoint of the averaging path. The derivative along the path is nonnegative by Lemma 2.3, and its integral over a chain from P to $\bar{P}$ is the gain

$$\Delta_f(P) = \mathcal{I}_f(\bar{P}) - \mathcal{I}_f(P) \geq 0. \tag{4.12}$$

Retaining the Jensen gaps in Lemma 2.3 writes the derivative as two relative entropies and the logarithmic slack in the quadratic inequality. Their integral represents the gain as a nonnegative remainder in a capacity lower bound.

Theorem 4.4 (An exact production decomposition)**.** *Assume that f is Lorentzian, or that the selected two-slot contraction satisfies (4.4). With the positive Gibbs data of (4.9) and the two marginal rows of (4.10), set*

$$R = \sum_j \frac{u_j v_j}{\pi_j}, \qquad \eta_j = \frac{u_j v_j}{\pi_j R}.$$

Then $0 < R \le 1$, η is a probability vector, and

$$d\mathcal{I}_f(P)[(v-u, u-v)] = \mathrm{D}(v\|\pi) + \mathrm{D}(u\|\eta) - \log R \ge \tfrac{1}{2}\|v-\pi\|_1^2 - \log R \ge 0. \tag{4.13}$$

Here D *is relative entropy with natural logarithms.*

Proof. The quadratic bound gives $R = \langle z, Tz\rangle_\pi / a^2 \le 1$. Direct expansion yields

$$\mathrm{D}(v\|\pi) = \sum_j v_j \log(Tz)_j - \log a, \quad \mathrm{D}(u\|\eta) = \log(aR) - \sum_j u_j \log(Tz)_j.$$

Subtracting $\log R$ gives (4.11). Nonnegativity and Pinsker's inequality prove the estimates. □

As in the proof of Theorem 2.1, dividing (4.13) by $1-2t$ gives the derivative in the original path direction at parameter $t < 1/2$. Consequently $\Delta_f(P)$ is the sum of these nonnegative production integrals along any finite chain taking P to $\bar{P}$. Integrals ending at $1/2$ are improper integrals with finite value by entropy continuity. For boundary matrices the identity means the limit of the interior production integrals under the radial approximation in the proof of Theorem 4.2. The integrated value is independent of the chain, since it is a difference of endpoint values.

Bounding an entropy gap by quadratic slack also appears in Cryan et al. (2021, proof of Lemma 10). Here the slack controls production along prescribed marginal averages. A uniform negative bound on the nontrivial kernel eigenvalues makes that control quantitative.

Corollary 4.5 (Strictness and a spectral remainder)**.** *Under the hypotheses of Theorem 4.4, put $a = \langle z, \mathbf{1}\rangle_\pi$. Then*

$$d\mathcal{I}_f(P)[(v-u, u-v)] = 0 \iff Tz = a\mathbf{1} \iff M(x - ay) = 0. \tag{4.14}$$

If, for some $0 < \lambda \le 1$,

$$\langle h, Th\rangle_\pi \le -\lambda \|h\|_\pi^2 \qquad \text{whenever } \langle h, \mathbf{1}\rangle_\pi = 0, \tag{4.15}$$

then, with $c_\lambda = \lambda/(1+\lambda)^2$,

$$d\mathcal{I}_f(P)[(v-u, u-v)] \ge c_\lambda \sum_j \frac{(u_j - v_j)^2}{\pi_j}. \tag{4.16}$$

If (4.15) holds with a common λ along a finite pair-averaging chain in the relative interior of $\mathcal{K}_f$ from P to $\bar{P}$, then

$$\Delta_f(P) \ge \frac{\lambda}{2(1+\lambda)^2} \|P - \bar{P}\|_F^2. \tag{4.17}$$

Proof. Lemma 2.3 gives the first equivalence in (4.14); the definition of T in (4.9) gives the second. Set $w = z/a - \mathbf{1}$. Theorem 4.4 gives

$$d\mathcal{I}_f(P)[(v-u, u-v)] \ge -\log R \ge 1 - R = -\langle w, Tw\rangle_\pi.$$

The nontrivial eigenvalues of the reversible stochastic matrix T are $-s$ with $\lambda \le s \le 1$. Since $s/(1+s)^2 \ge c_\lambda$ on this interval,

$$-\langle w, Tw\rangle_\pi \ge c_\lambda \|(I-T)w\|_\pi^2 = c_\lambda \sum_j \frac{(u_j - v_j)^2}{\pi_j},$$

proving (4.16).

On a segment with initial rows u, v and parameter $0 \le t \le \tau \le 1/2$, the current row difference is $(1-2t)(u-v)$. Dividing (4.16) by $1-2t$, using $\pi_j \le 1$, and integrating gives a gain of at least $c_\lambda \tau(1-\tau)\|u-v\|_2^2$. The squared Frobenius distance to $\bar{P}$ decreases by exactly $2\tau(1-\tau)\|u-v\|_2^2$. Summing over the chain proves (4.17); continuity handles $\tau = 1/2$. □

Without such a gap the deficit can stay constant. For the power of a linear form above, the nontrivial kernel eigenvalues are zero and $\mathcal{I}_f$ is constant at fixed column sums.

4.2. **Capacity bounds with remainders.** To obtain a capacity bound with the entropy remainder Δ_f, we evaluate H_f at the matrix that optimizes a transportation problem. For a $d \times m$ matrix $B > 0$, put $g_B(z) = \prod_i \langle b_i, z\rangle$. Rowwise Gibbs duality gives

$$\log \mathrm{Cap}_q(g_B) = \max_{\substack{P\ge 0,\ P\mathbf{1}=\mathbf{1} \\ P^T\mathbf{1}=q}} \{\langle P, \log B\rangle + \mathcal{S}(P)\}, \qquad |q| = d. \tag{4.18}$$

For $q = \mathbf{1}$ this is the formula of Straszak and Vishnoi (2017b, Theorem 2.1). The feasible transportation polytope is compact, including when some $q_j = 0$. The argument needs the optimizing matrix to be the marginal matrix of a law on allowed words, and for M-convex supports the next lemma shows that every matrix with unit row sums and column sums in $\mathrm{Newt}(f)$ is one.

Lemma 4.6 (The marginal polytope)**.** *If $\mathcal{A}$ is M-convex, then*

$$\mathcal{K}_f = \{P \ge 0 : P\mathbf{1} = \mathbf{1},\ P^T\mathbf{1} \in \mathrm{Newt}(f)\}. \tag{4.19}$$

Proof. We realize the right side as a matroid-intersection polytope. The integral-polymatroid description of an M-convex set gives the increasing integer submodular function $r(J) = \max_{\alpha\in\mathcal{A}} \sum_{j\in J} \alpha_j$, with $r([m]) = d$, and

$$\mathrm{Newt}(f) = \{q \ge 0 : q([m]) = d,\ q(J) \le r(J)\ \forall J\}, \qquad \mathcal{A} = \mathrm{Newt}(f) \cap \mathbb{Z}^m.$$

On $E = [d] \times [m]$, set $E_J = [d] \times J$. The standard matroid lift of an integer polymatroid (Helgason, 1974) has rank

$$\widetilde{r}(F) = \min_{J\subseteq[m]} \{r(J) + |F \setminus E_J|\}.$$

By a theorem of McDiarmid (1973), in the form of Bonin et al. (2023, Theorem 2.1), its independent sets are exactly the sets satisfying $|F \cap E_J| \le r(J)$ for every J. We intersect this matroid with the partition matroid allowing at most one element per row of E.

If P belongs to the right side of (4.19), then $p_{ij} \le 1$, and for every F, J,

$$P(F) \le P(E_J) + |F \setminus E_J| \le r(J) + |F \setminus E_J|.$$

Thus P lies in both independent-set polytopes and has total mass d. Edmonds' matroid-intersection polytope theorem (Edmonds, 1979) expresses it as a convex combination of common independent sets of size d. Each chooses one element per row and has its count vector in $\mathcal{A}$; it is therefore an allowed word incidence matrix. The reverse inclusion is immediate. □

The support hypothesis is needed. For $f = z_1z_2 + z_3z_4$, the rows $(1/2, 1/2, 0, 0)$ and $(0, 0, 1/2, 1/2)$ have column sum $(1/2, 1/2, 1/2, 1/2) \in \mathrm{Newt}(f)$, but they force the two letters of a word into different blocks, unlike every allowed word.

Corollary 4.7 (Product-factor capacity and its entropy remainder)**.** *Let f be Lorentzian, $q \in \mathrm{Newt}(f)$, and $B > 0$. If P_q maximizes (4.18), then*

$$\mathcal{D}_f(B) \geq \frac{d!}{d^d}\, q^q \,\mathrm{Cap}_q(f)\,\mathrm{Cap}_q(g_B)\exp(\Delta_f(P_q)). \tag{4.20}$$

In particular, dropping the remainder and taking the supremum over q gives

$$\mathcal{D}_f(B) \geq \frac{d!}{d^d} \sup_{q\in\mathrm{Newt}(f)} q^q \,\mathrm{Cap}_q(f)\,\mathrm{Cap}_q(g_B), \tag{4.21}$$

also for $B \geq 0$.

Proof. Lemma 4.6 makes every feasible P_q a word marginal. Equations (4.3) and (4.12) give

$$\mathsf{H}_f(P_q) = \mathcal{S}(P_q) - d\,\mathrm{H}(q/d) + \log(d!) + \log \mathrm{Cap}_q(f) + \Delta_f(P_q).$$

Inserting this law in (4.1) and using (4.18) and $\exp(-d\,\mathrm{H}(q/d)) = q^q/d^d$ proves (4.20). Dropping the nonnegative remainder and taking a supremum proves (4.21). Approximation by $B + \varepsilon \mathbf{1}\mathbf{1}^T$ proves the latter for nonnegative matrices, since $\mathrm{Cap}_q(g_{B+\varepsilon\mathbf{1}\mathbf{1}^T}) \downarrow \mathrm{Cap}_q(g_B)$ by monotonicity and evaluation at a fixed positive point. □

Expanding the mixed derivative gives $\mathcal{D}_f(B)/d! = \sum_{|\alpha|=d} \binom{d}{\alpha}^{-1} f_\alpha(g_B)_\alpha$, so (4.21) is a Lorentzian analogue of the coefficient-pairing conjecture of Gurvits (Gurvits and Leake, 2021, Conjecture 6.1) in the case of a product of linear forms; the conjecture is stated there for two real stable polynomials. In this case the unrefined bound follows from the classical coefficient bound (Gurvits, 2009a, Theorem 3.5(2)). Indeed, we apply it to $t \mapsto f(B^T t)$, use $f(z) \geq \mathrm{Cap}_q(f) z^q$, and apply the transposed form of the transportation duality (4.18):

$$\inf_{t>0} \frac{\prod_j (\sum_i b_{ij} t_i)^{q_j}}{\prod_i t_i} = q^q \,\mathrm{Cap}_q(g_B).$$

The log-sum-exp dual has objective $\langle P, \log B\rangle + \mathcal{S}(P) + \sum_j q_j \log q_j$ with the same constraints as (4.18); columns with $q_j = 0$ are omitted.

The full entropy endpoint (4.6) is also a consequence of this classical bound. Given P in the relative interior of $\mathcal{K}_f$ with column sums q, we choose its Gibbs parameter matrix B. Combining (4.21) with (4.18) gives

$$\mathsf{H}_f(P) = \log \mathcal{D}_f(B) - \langle P, \log B\rangle \geq \mathcal{S}(P) + \log(d!) + \log \mathrm{Cap}_q(f) - d\,\mathrm{H}(q/d).$$

Applying the radial continuity argument above to P and its equal-row matrix gives the boundary cases. Together with the proof of Corollary 4.7, this argument shows that the entropy endpoint and the unrefined product-capacity bound are equivalent.

The corrected right side of (4.20) equals

$$\exp\{\langle P_q, \log B\rangle + \mathsf{H}_f(P_q)\},$$

which is the Gibbs variational bound at the transportation optimizer. Pair averaging supplies an integral of nonnegative local terms for its improvement over the capacity baseline. Positive production at any interior point of a chain certifies a strict improvement.

For an explicit determinant-polynomial instance, take

$$f(z) = z_1^2 + 4z_1z_2 + z_2^2, \qquad B = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}, \qquad q = (1,1).$$

Then $\mathcal{D}_f(B) = 84$, $f(B^T t) = 33t_1^2 + 84t_1t_2 + 33t_2^2$, and its capacity is 150. Also $\mathrm{Cap}_q(f) = 6$, $\mathrm{Cap}_q(g_B) = 25$, and the transport optimizer has rows $(4/5, 1/5)$ and $(1/5, 4/5)$. Thus both unrefined lower bounds give 75.

Put $s = (3+\sqrt{73})/4$, so $2s^2 - 3s - 8 = 0$. The Gibbs matrix with rows $(s,1)$, $(1,s)$ has precisely those marginals and partition function $10(s+2)$. Consequently

$$\mathsf{H}_f(P_q) = \log(10(s+2)) - \frac{8}{5}\log s, \qquad \Delta_f(P_q) = \mathsf{H}_f(P_q) - 2\,\mathrm{H}(4/5,1/5) - \log 3.$$

The refined bound is

$$84 \geq 10(s+2)(4/s)^{8/5} > 75,$$

with middle value approximately 82.3712. The strict improvement also has an exact production certificate: the contraction is $M = \left(\begin{smallmatrix} 2 & 4 \\ 4 & 2 \end{smallmatrix}\right)$, and at the starting point

$$R = \left(\frac{2s^2+8s+2}{4s^2+4s+4}\right)^2 < 1,$$

since the denominator minus the numerator is $2(s-1)^2 > 0$. The initial production is therefore positive. Finally, $f(z) = \det(z_1\,\mathrm{diag}(2+\sqrt{3}, 2-\sqrt{3}) + z_2 I_2)$, so the example is also a mixed-discriminant instance of the correction, with commuting matrices.

5. Geometric and combinatorial inequalities

Theorem 4.2 and its corollaries hold for every Lorentzian polynomial. Geometry and combinatorics supply Lorentzian polynomials whose mixed derivatives are mixed volumes, mixed discriminants, intersection numbers, and weighted counts of matroid bases, and for these the corollaries become inequalities of van der Waerden type.

5.1. **Mixed volumes, mixed discriminants, and intersection numbers.** Let $K_1, \dots, K_m$ be compact convex bodies in $\mathbb{R}^d$, and suppose their Minkowski sum is full dimensional. We use the standard normalization $V(K,\dots,K) = \mathrm{Vol}(K)$. The volume polynomial is

$$f(z) = \mathrm{Vol}\Big(\sum_j z_j K_j\Big) = \sum_{|\alpha|=d} \frac{d!}{\alpha!} V(K_1[\alpha_1], \dots, K_m[\alpha_m]) z^\alpha. \tag{5.1}$$

For $L_i = \sum_j b_{ij} K_j$, multilinearity gives

$$\mathcal{D}_f(B) = d! V(L_1, \dots, L_d).$$

The Lorentzian property follows from classical mixed-volume Alexandrov–Fenchel theory (Brändén and Huh, 2020, Theorem 4.1). All entropy comparisons and production identities therefore apply to the coefficient weights in (5.1). In particular, applying Corollary 4.3 to $t \mapsto \mathrm{Vol}(\sum_i t_i L_i)$ gives

$$V(L_1, \dots, L_d) \geq \frac{1}{d^d}\,\mathrm{Cap}_{\mathbf{1}}\left(\mathrm{Vol}\Big(\sum_i t_i L_i\Big)\right). \tag{5.2}$$

Alternatively, for every $q \in \mathrm{Newt}(f)$, Corollary 4.7 gives

$$V(L_1, \dots, L_d) \geq \frac{q^q}{d^d}\,\mathrm{Cap}_q(f)\,\mathrm{Cap}_q(g_B), \tag{5.3}$$

with the entropy correction of (4.20) when $B > 0$. Gurvits proved (5.2), together with stronger factors involving the affine dimensions of the bodies (Gurvits, 2009b, Theorems 2.1 and 2.4), and (5.3) follows from the classical coefficient bound as in Section 4.2. When $L_1, \dots, L_d$ are the Newton polytopes of a system of Laurent polynomials, $d!V(L_1,\dots,L_d)$ is the number of isolated roots in $(\mathbb{C}^*)^d$ of a generic system with these polytopes (Bernshtein, 1975), so (5.2) bounds that generic root count from below.

Let $A_1, \dots, A_m$ be Hermitian positive semidefinite matrices of size d, with $\sum_j A_j$ positive definite, and put

$$f(z) = \det\Big(\sum_j z_j A_j\Big).$$

This polynomial has nonnegative coefficients and is real stable (Borcea et al., 2009, Proposition 3.2), hence Lorentzian (Brändén and Huh, 2020, Proposition 2.2). We normalize the mixed discriminant by $\mathrm{MD}(A,\dots,A)=\det A$. For $Q_i=\sum_j b_{ij}A_j$,

$$\mathcal{D}_f(B)=d!\,\mathrm{MD}(Q_1,\dots,Q_d),\qquad f_\alpha=\frac{d!}{\alpha!}\,\mathrm{MD}(A_1[\alpha_1],\dots,A_m[\alpha_m]).$$

Equations (5.2) and (5.3) hold with V,Vol replaced by $\mathrm{MD},\det$, with the same entropy correction. The same applies to intersection numbers of nef divisor classes on an irreducible projective variety, whose intersection polynomial is Lorentzian (Brändén and Huh, 2020, Theorem 4.6).

Doubly stochastic tuples. Mixed discriminants of doubly stochastic tuples of matrices generalize permanents of doubly stochastic matrices. For them, averaging gives the partial bounds at every order, and keeping one slack term turns these bounds into an explicit remainder in the matrices. For matrices of order n, we write $\mathsf{D}=n!\,\mathrm{MD}$, so $\mathsf{D}(I,\dots,I)=n!$, and we assume throughout that

$$A_i\succeq 0,\qquad \operatorname{tr}A_i=1,\qquad \sum_{i=1}^{n}A_i=I.$$

For $0\le k\le n$, put

$$\mathsf{S}_k(A)=\sum_{\substack{S\subseteq[n]\\|S|=k}}\mathsf{D}\big((A_i)_{i\in S},I^{[n-k]}\big),\qquad \mathsf{S}_0(A)=n!.$$

Here $I^{[r]}$ means r copies of I. The following lower bound on S_k is a special case of the generalized Friedland–Tverberg inequality of Friedland and Gurvits (2006, Theorem 3.1), applied to the positive hyperbolic polynomial q_k below; for $k=n$ it is Gurvits's solution of Bapat's conjecture (Gurvits, 2006). We include their padding proof because Corollary 4.3 now supplies its coefficient bound by averaging. We write $D_{\mathbf{1}}=\sum_i\partial_{x_i}$.

Proposition 5.1 (Partial mixed-discriminant bound)**.** *For every $0\le k\le n$,*

$$\mathsf{S}_k(A)\ge\binom{n}{k}\frac{n!}{n^k}. \tag{5.4}$$

Proof. The cases $k=0,1$ are identities. For $k\ge 2$, write

$$f(x)=\det\Big(\sum_i x_iA_i\Big),\qquad q_k(x)=\frac{k!}{n!}D_{\mathbf{1}}^{n-k}f(x),\qquad \ell(x)=\frac{1}{n}\sum_i x_i.$$

The polynomial q_k is k-homogeneous. Since $f(\mathbf{1})=1$ and $\partial_i f(\mathbf{1})=\operatorname{tr}A_i=1$, homogeneity gives

$$q_k(\mathbf{1})=1,\qquad \partial_i q_k(\mathbf{1})=\frac{k}{n}.$$

The determinant polynomial is stable, because if every z_i has positive imaginary part, then the imaginary part of $\sum_i z_iA_i$ is positive definite, so this matrix is nonsingular. It is therefore Lorentzian; nonnegative directional differentiation and multiplication by a nonnegative linear form preserve this class (Brändén and Huh, 2020, Proposition 2.2, Corollaries 2.11 and 2.32, and Theorem 2.30). Thus $p_k=q_k\ell^{n-k}$ is Lorentzian, with $p_k(\mathbf{1})=1$ and $\partial_i p_k(\mathbf{1})=1$. Since its coefficients form a probability law on exponent vectors with mean $\mathbf{1}$, Jensen's inequality gives $p_k(x)\ge\prod_i x_i$ for $x>0$. Consequently $\mathrm{Cap}_{\mathbf{1}}(p_k)=1$.

Corollary 4.3 therefore gives

$$[x_1\cdots x_n]p_k\ge\frac{n!}{n^n}.$$

Only squarefree monomials of q_k can contribute on the left, so

$$[x_1\cdots x_n]p_k=\frac{k!}{n!}\,\frac{(n-k)!}{n^{n-k}}\mathsf{S}_k(A).$$

Rearranging proves (5.4). □

To express the remainder in the input matrices, define

$$\eta = n \min_i \lambda_{\min}(A_i), \qquad V(A) = \sum_i \|A_i - I/n\|_F^2 = \sum_i \operatorname{tr}(A_i^2) - 1.$$

The assumptions imply $0 \le \eta \le 1$.

Proposition 5.2 (Explicit quadratic remainder)**.** *For $n \ge 2$ and $2 \le k \le n$,*

$$\mathsf{S}_k(A) \ge \binom{n}{k}\frac{n!}{n^k} + \binom{n-2}{k-2}\frac{(n-2)!}{2n^{k-2}}\eta^{k-2}V(A). \tag{5.5}$$

The convention $\eta^0 = 1$ applies also at $\eta = 0$. In particular,

$$\mathsf{D}(A_1, \dots, A_n) \ge \frac{n!}{n^n} + \frac{(n-2)!}{2n^{n-2}}\eta^{n-2}\sum_i \|A_i - I/n\|_F^2. \tag{5.6}$$

For $k = 2$, (5.5) is an identity. For every $k \ge 2$, the coefficient of $V(A)$ is sharp to second order at $A_i = I/n$.

Proof. If $\eta = 1$, all $A_i = I/n$, so the assertion is immediate. Otherwise we split off the uniform part $(\eta/n)I$ of each A_i; what remains is again a doubly stochastic tuple, and $\mathsf{S}_k(A)$ expands into its slacks. We set

$$B_i = \frac{A_i - \eta I/n}{1 - \eta}.$$

Multilinearity and counting the sets that contain a fixed j-element subset give

$$\mathsf{S}_k(A) = \sum_{j=0}^{k} \binom{n-j}{k-j}\left(\frac{\eta}{n}\right)^{k-j}(1-\eta)^j \mathsf{S}_j(B). \tag{5.7}$$

Set $\delta_j(B) = \mathsf{S}_j(B) - \binom{n}{j}n!/n^j$. Proposition 5.1 says $\delta_j(B) \ge 0$, and the binomial identity $\binom{n-j}{k-j}\binom{n}{j} = \binom{n}{k}\binom{k}{j}$ shows that (5.7) is exactly

$$\mathsf{S}_k(A) - \binom{n}{k}\frac{n!}{n^k} = \sum_{j=2}^{k} \binom{n-j}{k-j}\left(\frac{\eta}{n}\right)^{k-j}(1-\eta)^j \delta_j(B). \tag{5.8}$$

The order-two term is explicit. Differentiating the determinant twice at the identity gives

$$\mathsf{D}(B_i, B_j, I^{[n-2]}) = (n-2)!(1 - \operatorname{tr}(B_iB_j)).$$

Since $\sum_i B_i = I$, summing over $i < j$ yields

$$\delta_2(B) = \frac{(n-2)!}{2}\left(\sum_i \operatorname{tr}(B_i^2) - 1\right) = \frac{(n-2)!}{2}V(B).$$

Keeping only $j = 2$ in (5.8) and using $(1-\eta)^2 V(B) = V(A)$ proves the bound. At $k = 2$ no terms were discarded.

For the second-order assertion, let $A_i(\varepsilon) = I/n + \varepsilon H_i$, where the H_i are Hermitian, $\operatorname{tr} H_i = 0$, and $\sum_i H_i = 0$. Multilinearity gives

$$\mathsf{S}_k(A(\varepsilon)) = \binom{n}{k}\frac{n!}{n^k} + \binom{n-2}{k-2}\frac{(n-2)!}{2n^{k-2}}\varepsilon^2 \sum_i \operatorname{tr}(H_i^2) + O(\varepsilon^3).$$

Indeed, the linear term vanishes; the quadratic term uses $\mathsf{D}(H_i, H_j, I^{[n-2]}) = -(n-2)!\operatorname{tr}(H_iH_j)$ and $\sum_{i<j}\operatorname{tr}(H_iH_j) = -\frac{1}{2}\sum_i \operatorname{tr}(H_i^2)$. Also $\eta(A(\varepsilon)) = 1 + O(|\varepsilon|)$, proving the asserted local sharpness. □

For positive definite tuples, $\eta > 0$, so (5.6) is strict unless every $A_i = I/n$. For singular tuples and $n > 2$, its displayed correction vanishes; it does not by itself settle boundary equality. At $n = 2$, it is the exact identity $\mathsf{D}(A_1, A_2) = \frac{1}{2} + \frac{1}{2}V(A)$.

If $A_i = \operatorname{diag}(a_{i1}, \dots, a_{in})$, the tuple conditions say that the matrix with entries a_{ij} is doubly stochastic, and $\mathsf{S}_k(A)$ is $(n-k)!$ times its σ_k from Section 3. Proposition 5.1 then recovers the Friedland–Tverberg constant $\binom{n}{k}^2 k!/n^k$, and Proposition 5.2 corrects it in the entries of the matrix.

Corollary 5.3 (Explicit partial-permanent remainder)**.** *Let A be an $n \times n$ doubly stochastic matrix, $a = \min_{i,j} a_{ij}$, and $2 \le k \le n$. Then, with $b_{n,k} = \binom{n}{k}^2 k!/n^k$,*

$$\sigma_k(A) \ge b_{n,k} + \frac{1}{2}\binom{n-2}{k-2}^2 (k-2)!\, a^{k-2}\|A - U\|_F^2. \tag{5.9}$$

Here $a^0 = 1$, including when $a = 0$. The bound is an identity for $k = 2$, and its quadratic coefficient is asymptotically exact as $A \to U$.

Proof. We apply Proposition 5.2 to the diagonal matrices whose diagonals are the rows of A. Then $\eta = na$, $V = \|A - U\|_F^2$, and $\mathsf{S}_k = (n-k)!\,\sigma_k(A)$. Dividing by $(n-k)!$ and using $(n-2)!/(n-k)! = \binom{n-2}{k-2}(k-2)!$ gives (5.9). The identity at $k = 2$ and the local sharpness are those of Proposition 5.2, since diagonal perturbations are allowed there. □

For example, for the matrix with diagonal entries $1/2$ and off-diagonal entries $1/4$ in dimension three, the permanent is $1/4$. The lower bound in (5.9) is $2/9 + 1/64 = 137/576$, compared with the unrefined value $2/9$. For other quantitative permanent estimates, see Lavi (2026, Theorem 1.2).

5.2. **Matroids and strongly Rayleigh laws.** For a multiaffine polynomial, the mixed derivative is a weighted count of assignments of labelled slots to bases. Let $f(z) = \sum_{S \in \mathcal{B}} f_S z^S$ be homogeneous and multiaffine of degree d. Then

$$\mathcal{D}_f(B) = \sum_{S \in \mathcal{B}} f_S \operatorname{per}(B_S),$$

the weighted partition function for assigning d labelled slots bijectively to a chosen set S; if $f(\mathbf{1}) = 1$, it is $\mathbb{E}_S \operatorname{per}(B_S)$. The averaging theorems need f to be Lorentzian, which arbitrary positive basis weights need not be. A fixed-size strongly Rayleigh distribution has a homogeneous stable generating polynomial and is therefore covered. A general strongly Rayleigh law may first be conditioned on a cardinality with positive probability, which preserves the strongly Rayleigh property (Borcea et al., 2009, Corollary 4.18). Every unweighted matroid basis polynomial is Lorentzian, including those of matroids without a stable basis polynomial (Brändén and Huh, 2020, Theorem 3.10). Positive external fields preserve this property.

For such f, (4.6) bounds the maximum weighted entropy of ordered bases with prescribed slot marginals. For an unweighted matroid, the entropy estimate of Anari et al. (2021, Theorem 5.2, Corollary 5.6, and Lemma 5.8) applies to both a basis law and its complement. Keeping both entropy estimates strengthens the bound for ordered bases; to state it, we write

$$L(q) = -\sum_j q_j \log q_j, \qquad \chi(q) = \max\{0, L(\mathbf{1} - q) - L(q)\}, \qquad \kappa_d = \log \frac{d^d}{d!}.$$

Corollary 5.4 (Ordered matroid bases)**.** *Let f be the unweighted basis generating polynomial of a matroid of rank d on $[m]$, and let $P \in \mathcal{K}_f$ have column sums q. Then*

$$\mathsf{H}_f(P) \ge \mathcal{S}(P) - \kappa_d + \chi(q) + \Delta_f(P). \tag{5.10}$$

Proof. The prescribed-marginal construction of Anari et al. (2021, Lemma 5.8) gives a basis law ν with marginals q such that ν and its complement ν^* are log-concave. Their entropy estimate

gives $\mathrm{H}(\nu) \geq L(q)$ and $\mathrm{H}(\nu) = \mathrm{H}(\nu^*) \geq L(\mathbf{1}-q)$. The construction includes boundary marginals by limits of external fields. By (4.2),

$$\log \mathrm{Cap}_q(f) \geq \max\{L(q), L(\mathbf{1}-q)\} = L(q) + \chi(q). \tag{5.11}$$

Equations (4.3) and (4.12) give

$$\mathsf{H}_f(P) = \mathcal{S}(P) - \kappa_d - L(q) + \log \mathrm{Cap}_q(f) + \Delta_f(P),$$

which proves the claim. □

For the free matroid of rank n, the polynomial is $z_1 \cdots z_n$, the column sums are $q = \mathbf{1}$, and $\chi(q) = 0$. Dropping $\Delta_f(P)$ recovers the bound $\mathcal{I} \leq \kappa_n$ of Theorem 2.5.

External fields carry these bounds to weighted bases. For positive external fields λ, put $f_\lambda(z) = f(\lambda_1 z_1, \ldots, \lambda_m z_m)$. Then

$$\mathsf{H}_{f_\lambda}(P) = \mathsf{H}_f(P) + q \cdot \log \lambda, \qquad \mathrm{Cap}_q(f_\lambda) = \lambda^q \, \mathrm{Cap}_q(f), \qquad \Delta_{f_\lambda}(P) = \Delta_f(P).$$

For the normalized polynomial $f_\lambda / f(\lambda)$, the normalization subtracts $\log f(\lambda)$ from both the entropy and the logarithmic capacity and leaves the averaging remainder unchanged. These identities hold on the boundary as well, since the log weight added to every admissible law has expectation $q \cdot \log \lambda$.

The ordered entropy bound also compares a weighted assignment count with a relaxation involving only the slot marginals. For a $d \times m$ matrix $B \geq 0$, let

$$\mathcal{P}_M(B) = \{P \in \mathcal{K}_f : p_{ij} = 0 \text{ whenever } b_{ij} = 0\}, \qquad F_B(P) = \langle P, \log B \rangle + \mathcal{S}(P),$$

where zero terms are omitted in the inner product. If $\mathcal{P}_M(B) \neq \varnothing$, set

$$Z_M(B) = \sum_{S \text{ basis of } M} \mathrm{per}(B_S), \qquad \rho_M(B) = \max_{P \in \mathcal{P}_M(B)} F_B(P).$$

Corollary 5.5 (Weighted assignments)**.** *Let $P \in \mathcal{P}_M(B)$ have column sums q. Then*

$$Z_M(B) \geq \frac{d!}{d^d} \exp(F_B(P) + \chi(q) + \Delta_f(P)). \tag{5.12}$$

Consequently, the relaxation value $\rho_M(B)$ satisfies

$$\frac{d!}{d^d} \, e^{\rho_M(B)} \leq Z_M(B) \leq e^{\rho_M(B)}. \tag{5.13}$$

Moreover, with $g_B(z) = \prod_i \langle b_i, z \rangle$ as in Section 4.2,

$$e^{\rho_M(B)} = \sup_{q \in \mathrm{Newt}(f)} \mathrm{Cap}_q(g_B).$$

Proof. The Gibbs formula (4.1) gives

$$\log Z_M(B) = \max_{P \in \mathcal{P}_M(B)} \{\langle P, \log B \rangle + \mathsf{H}_f(P)\}.$$

For each feasible P, its maximum-entropy law is supported on allowed assignments, because a zero marginal excludes that coordinate from every word of positive probability. Applying (5.10) proves (5.12); applying entropy subadditivity $\mathsf{H}_f(P) \leq \mathcal{S}(P)$ proves the upper bound in (5.13). The lower bound follows by choosing a maximizer of F_B and dropping the two nonnegative corrections. Finally, Lemma 4.6 lets us optimize first at fixed column sums, and (4.18) then gives the last identity. The same variational formula holds for zero entries of B after restricting the support. □

The factor $d!/d^d$ itself also follows from the classical coefficient bound and (5.11), through Corollary 4.7; the averaging adds the entropy remainder at the chosen marginals. Corollary 5.5 thus specializes the entropy approach to matroid counting of Anari et al. (2021) to assignments with prescribed slot marginals.

For disjoint color classes $E_1, \dots, E_d$ partitioning $[m]$, an assignment is a rainbow basis when slot i must choose from E_i. A vector $q \in \mathrm{Newt}(f)$ with $q(E_i) = 1$ determines the feasible matrix $P(q)_{ij} = q_j \mathbf{1}_{\{j \in E_i\}}$.

Corollary 5.6 (Rainbow bases)**.** *Let M be a rank-d matroid, with the color classes above and positive element weights λ_j. For every $q \in \mathrm{Newt}(f)$ satisfying $q(E_i) = 1$,*

$$\sum_{\substack{S \text{ basis of } M \\ |S \cap E_i| = 1 \ (1 \le i \le d)}} \prod_{j \in S} \lambda_j \ge \frac{d!}{d^d} \exp(q \cdot \log \lambda + \max\{L(q), L(\mathbf{1} - q)\} + \Delta_f(P(q))).$$

Proof. We apply (5.12) to $b_{ij} = \lambda_j \mathbf{1}_{\{j \in E_i\}}$. Each rainbow basis has a unique allowed assignment, and $F_B(P(q)) = q \cdot \log \lambda + L(q)$. □

The supplement specializes these bounds to weighted rainbow spanning trees and to the homogenized multivariate Tutte polynomial (Sra, 2026b, Section S5.2).

5.3. **Sums of squarefree coefficients.** Anari et al. (2017, Theorem 1.2) bound a squarefree coefficient sum in their analysis of Nash social welfare, and this bound also admits an averaging proof. Their padding construction (Anari et al., 2017, Lemma 4.1) reduces the sum to a single coefficient. Applying Corollary 4.3 to the padded polynomial gives their constant throughout the Lorentzian class.

For a homogeneous polynomial f of degree $d \le m$, put

$$C_{\mathrm{sf}}(f) = \sum_{\substack{S \subseteq [m] \\ |S| = d}} f_{\mathbf{1}_S}, \qquad \mathcal{R}(f) = \inf_{\substack{z > 0 \\ z^S \ge 1 \ (|S| = d)}} f(z), \qquad \mathcal{Q}_{m,d} = \{q \in [0,1]^m : |q| = d\}.$$

As usual, a product with exponent zero is interpreted as one.

Corollary 5.7 (Squarefree coefficient sums)**.** *Let f be a nonzero Lorentzian polynomial of degree $1 \le d \le m$ with $C_{\mathrm{sf}}(f) > 0$. Set $h = m - d$ and $c_{m,d} = m! h^h / (m^m h!)$. Then*

$$C_{\mathrm{sf}}(f) \ge c_{m,d} \sup_{q \in \mathcal{Q}_{m,d}} \frac{\mathrm{Cap}_q(f)}{(\mathbf{1} - q)^{\mathbf{1} - q}} \ge c_{m,d} \mathcal{R}(f) \ge e^{-d} \mathcal{R}(f). \tag{5.14}$$

Proof. The case $m = 1$ is an identity, so assume $m \ge 2$. The product $F(z) = f(z)(z_1 + \cdots + z_m)^h$ is Lorentzian by Brändén and Huh (2020, Corollary 2.32 and Theorem 2.30), and

$$[z_1 \cdots z_m] F = h! C_{\mathrm{sf}}(f).$$

For $q \in \mathcal{Q}_{m,d}$, weighted arithmetic–geometric mean gives

$$(z_1 + \cdots + z_m)^h \ge \frac{h^h z^{\mathbf{1} - q}}{(\mathbf{1} - q)^{\mathbf{1} - q}}.$$

For $h = 0$ this is the identity $1 = 1$. Multiplying by $f(z) \ge \mathrm{Cap}_q(f) z^q$ and applying Corollary 4.3 to F proves the first inequality.

The capacity identity

$$\mathcal{R}(f) = \sup_{q \in \mathcal{Q}_{m,d}} \mathrm{Cap}_q(f) \tag{5.15}$$

follows from Anari et al. (2017, Lemma 4.3) and uses only nonnegative coefficients. To see it directly, we set $z = e^u$ and minimize $\log f(e^u)$ subject to $\langle \mathbf{1}_S, u \rangle \ge 0$ for every d-set S. This convex program has a strictly feasible point. If $\mathcal{R}(f) = 0$, weak duality makes both sides of (5.15) zero; otherwise its value is finite. Strong duality identifies its value with the supremum of $\log \mathrm{Cap}_q(f)$,

where $q = \sum_S \lambda_S \mathbf{1}_S$ and $\lambda_S \geq 0$. Homogeneity forces $\sum_S \lambda_S = 1$, so these vectors fill $\mathcal{Q}_{m,d}$. Since $(\mathbf{1} - q)^{\mathbf{1}-q} \leq 1$, this proves the second inequality. Finally,

$$c_{m,d} = \prod_{r=h+1}^{m} \left(\frac{r-1}{r}\right)^{r-1} \geq e^{-d},$$

where the factor for $r = 1$ is one. □

The averaging proof also records the slack in the coefficient step. We view the identity matrix I_m as the deterministic word $(1, \dots, m)$ for F. Since $\mathsf{H}_F(I_m) = \log(h! C_{\mathrm{sf}}(f))$, the equal-row identity (4.3) and the definition (4.12) give

$$C_{\mathrm{sf}}(f) = \frac{m!}{m^m h!} \operatorname{Cap}_{\mathbf{1}}(F) \exp(\Delta_F(I_m)).$$

Thus the first two lower bounds in (5.14) can be multiplied by $\exp(\Delta_F(I_m))$. Theorem 4.4 expresses this remainder through nonnegative production along row averaging, with the boundary interpreted by continuity. The numerical constant and the padding argument are those of Anari et al. (2017).

For a product of linear forms, the matroid bounds of Section 5.2 improve the constant $c_{m,d}$ to $d!/d^d$; the supplement derives this improvement and its consequence for the allocation model of Anari et al. (2017) (Sra, 2026b, Section S5.3).

6. Binary averaging and stable pairings

Row averaging compares optimized entropies at different marginals. For a pairing of two polynomials, a second comparison keeps the marginals fixed and averages a binary coupling toward independence. Applied within the diagonal contractions of Anari and Oveis Gharan (2017), it proves their capacity bound with an exact entropy remainder. Stability guarantees that successive contractions satisfy the required two-variable inequality; for permanents, matching switching supplies that inequality.

The coefficient pairing $\sum_S f_S g_S$ arises from $f(y)g(z)$ by contracting the variables y_i and z_i in pairs, and the argument tracks capacity through these contractions. The capacity definition and the variational formula (4.2) apply also to nonhomogeneous polynomials, and we write $h(t) = -t \log t - (1-t)\log(1-t)$ for the binary entropy. A real polynomial $\Psi(y, z)$ in variables $y, z \in \mathbb{R}^m$ is *bistable* if $\Psi(y, -z)$ is real stable; in this section Ψ has nonnegative coefficients and is multiaffine in its $2m$ variables. The diagonal contraction

$$\mathcal{C}_i \Psi = (1 + \partial_{y_i} \partial_{z_i}) \Psi|_{y_i = z_i = 0}$$

keeps the terms in which both selected variables occur or neither occurs, and then deletes them.

Lemma 6.1 (Diagonal contraction)**.** *If Ψ is bistable, then $\mathcal{C}_i \Psi$ is bistable or zero. For positive fixed values of the other variables, write the selected slice as*

$$p(y_i, z_i) = a y_i z_i + b y_i + c z_i + d.$$

Then $ad \geq bc$.

Proof. For the preservation claim, we apply the sign change and write $R(u, v, w) = A(w)uv + B(w)u + C(w)v + D(w)$, which is stable. Since $-1/u$ lies in the upper half-plane whenever u does,

$$uR(u, -1/u, w) = B(w)u^2 + (D(w) - A(w))u - C(w)$$

is stable. Differentiating in u and specializing $u = 0$ shows that $D - A$ is stable or zero. Undoing the sign change identifies this polynomial with $\mathcal{C}_i \Psi$. The preservation statement is Anari and Oveis Gharan (2017, Lemma 4.4); the argument here uses only differentiation, real specialization, and a fractional linear substitution.

The bivariate polynomial $p(y,-z) = -ayz + by - cz + d$ is stable or zero by real specialization of the other variables. If $ad < bc$, then $b, c > 0$, and for y in the upper half-plane the root $z = (by+d)/(ay+c)$ also has positive imaginary part, since

$$\operatorname{Im} \frac{by+d}{ay+c} = \frac{bc-ad}{|ay+c|^2} \operatorname{Im} y > 0.$$

Such a root contradicts stability. □

Lemma 6.2 (A binary entropy comparison)**.** *Let $p(y,z) = ayz + by + cz + d$, where $a, b, c, d \geq 0$ and $ad \geq bc$. For $0 \leq q \leq 1$,*

$$a + d \geq e^{-h(q)} \operatorname{Cap}_{(q,q)}(p). \tag{6.1}$$

If all four coefficients are positive and $0 < q < 1$, let μ maximize weighted entropy with weights $(w_{00}, w_{10}, w_{01}, w_{11}) = (d, b, c, a)$, subject to $\mathbb{E}X = \mathbb{E}Y = q$. Put $\lambda = \log(ad/(bc))$ and $r = a/(a+d)$, and write $\mathrm{I}_\mu(X;Y) = \mathrm{D}(\mu \| \mu_X \otimes \mu_Y)$. Then

$$\log(a+d) + h(q) - \log \operatorname{Cap}_{(q,q)}(p) = \mathrm{I}_\mu(X;Y) + \lambda \mu(0,1) + \mathrm{D}(\operatorname{Ber}(q) \| \operatorname{Ber}(r)). \tag{6.2}$$

Moreover, for

$$p_t(y,z) = ayz + e^{t\lambda/2}(by + cz) + d, \qquad 0 \leq t \leq 1,$$

the capacity $\operatorname{Cap}_{(q,q)}(p_t)$ is nondecreasing and

$$\log(a+d) + h(q) - \log \operatorname{Cap}_{(q,q)}(p) = \mathrm{D}(\operatorname{Ber}(q) \| \operatorname{Ber}(r)) + \int_0^1 \lambda \mu_t(0,1)\, dt, \tag{6.3}$$

where μ_t is the corresponding weighted-entropy maximizer.

Proof. First suppose the four coefficients are positive and $0 < q < 1$. The constrained entropy maximizer has full support, and $\mu(0,1) = \mu(1,0) = s$. By (4.2),

$$\begin{aligned} \log \operatorname{Cap}_{(q,q)}(p) &= \mathrm{H}(\mu) + (1-q)\log d + q \log a - s\lambda \\ &= 2h(q) - \mathrm{I}_\mu(X;Y) + (1-q)\log d + q \log a - s\lambda. \end{aligned}$$

Subtracting from $\log(a+d) + h(q)$ gives (6.2), since the remaining Bernoulli term is relative entropy. All three terms are nonnegative.

For the interpolation, the coefficient ratio is $ad/(b_t c_t) = e^{(1-t)\lambda} \geq 1$, so the slices remain bistable. Strict concavity and the implicit function theorem give a unique positive maximizer depending smoothly on t. Differentiating the variational optimum gives

$$\frac{d}{dt} \log \operatorname{Cap}_{(q,q)}(p_t) = \frac{\lambda}{2}(\mu_t(1,0) + \mu_t(0,1)) = \lambda \mu_t(0,1) \geq 0.$$

At $t = 1$, the coefficient array has rank one, so X and Y are independent under its Gibbs law and

$$\log \operatorname{Cap}_{(q,q)}(p_1) = 2h(q) + (1-q)\log d + q \log a.$$

Integration proves (6.3).

For the remaining cases, $q = 0, 1$ gives capacities d, a, respectively. If $ad > 0$ but $bc = 0$, any supported law with equal marginals is diagonal, so $\log \operatorname{Cap}_{(q,q)}(p) = h(q) + (1-q)\log d + q \log a$; weighted arithmetic–geometric mean gives even $\operatorname{Cap}_{(q,q)}(p) \leq a + d$. If $ad = 0$, then $bc = 0$ and no supported law has equal marginals in $(0,1)$, so that capacity is zero. These cases prove (6.1) in full generality. □

The interpolation averages the binary coupling toward independence. Indeed, if $s_t = \mu_t(0,1)$, its Gibbs odds ratio is

$$\frac{(q - s_t)(1 - q - s_t)}{s_t^2} = e^{(1-t)\lambda}.$$

The left side strictly decreases with s_t, so s_t increases from s_0 to $q(1-q)$. For $\lambda > 0$, this gives

$$\mu_t = (1-\theta_t)\mu_0 + \theta_t(\mathrm{Ber}(q) \otimes \mathrm{Ber}(q)), \qquad \theta_t = \frac{s_t - s_0}{q(1-q) - s_0} \in [0,1].$$

When $\lambda = 0$, the coupling is already independent and the path is constant. The averaging comparison is local to each bivariate slice; it does not require a simultaneous interpolation of all coefficients of Ψ. Iteration requires the condition $ad \geq bc$ to persist under contraction.

Iterating Lemma 6.2 along the diagonal contractions gives the multiaffine pairing bound of Anari and Oveis Gharan (2017, Theorem 1.1), now with its exact loss. Lemma 6.1 supplies the closure that the iteration needs.

Theorem 6.3 (Multiaffine stable pairings)**.** *Let f, g be multiaffine real stable polynomials in m variables with nonnegative coefficients, and let $q \in [0,1]^m$. Then*

$$\sum_{S \subseteq [m]} f_S g_S \geq q^q (\mathbf{1}-q)^{\mathbf{1}-q} \operatorname{Cap}_q(f) \operatorname{Cap}_q(g). \tag{6.4}$$

Suppose both capacities are positive. Put $\Phi(y,z) = f(y)g(z)$ and $\Phi_i = \mathcal{C}_i \cdots \mathcal{C}_1 \Phi$, and write $\operatorname{Cap}(\Phi_i)$ for the capacity of Φ_i at the exponents q_j of y_j and z_j, $j > i$. Then the slacks

$$\delta_i = \log \operatorname{Cap}(\Phi_i) - \log \operatorname{Cap}(\Phi_{i-1}) + h(q_i)$$

are nonnegative, and

$$\sum_{S \subseteq [m]} f_S g_S = q^q (\mathbf{1}-q)^{\mathbf{1}-q} \operatorname{Cap}_q(f) \operatorname{Cap}_q(g) \exp\Big(\sum_{i=1}^{m} \delta_i\Big). \tag{6.5}$$

Proof. The polynomial Φ is bistable, because real coefficients imply that a stable polynomial is also nonvanishing when all variables lie in the lower half-plane. For a bistable Ψ, we fix all variables other than y_i, z_i at positive values. By Lemma 6.1 the resulting slice satisfies $ad \geq bc$, so (6.1) applies to it. Taking the infimum over the fixed variables gives

$$\operatorname{Cap}_{(q_{-i}, q_{-i})}(\mathcal{C}_i \Psi) \geq e^{-h(q_i)} \operatorname{Cap}_{(q,q)}(\Psi), \tag{6.6}$$

where q_{-i} omits the ith coordinate. Every Φ_i is bistable or zero by Lemma 6.1, so (6.6) applies at each step. If both capacities of f and g are positive, then inductively every $\operatorname{Cap}(\Phi_i)$ is positive and $\delta_i \geq 0$. The full contraction is

$$\Phi_m = \sum_{S \subseteq [m]} f_S g_S, \qquad \operatorname{Cap}(\Phi_0) = \operatorname{Cap}_q(f) \operatorname{Cap}_q(g),$$

so the slacks telescope to (6.5). Since $e^{-\sum_i h(q_i)} = q^q(\mathbf{1}-q)^{\mathbf{1}-q}$, nonnegativity of the slacks gives (6.4), which is immediate when one capacity is zero. □

Each slack separates into a local binary loss and a minimization loss. Let $v^{(k)}$ be a minimizing sequence for $\operatorname{Cap}(\Phi_i)$, and put $K_i(v) = \operatorname{Cap}_{(q_i,q_i)}(\Phi_{i-1}(v,\cdot,\cdot))$. Writing $v = (y^{\mathrm{rem}}, z^{\mathrm{rem}})$ for the variables with index $j > i$ and $v^{q^{\mathrm{rem}}} = (y^{\mathrm{rem}} z^{\mathrm{rem}})^{q_{i+1:m}}$ for the remaining denominator,

$$\delta_i = \lim_{k \to \infty} \left[\log \frac{\Phi_i(v^{(k)}) e^{h(q_i)}}{K_i(v^{(k)})} + \log \frac{K_i(v^{(k)})}{(v^{(k)})^{q^{\mathrm{rem}}} \operatorname{Cap}(\Phi_{i-1})} \right].$$

Both terms are nonnegative. For positive local coefficients the first has the representations (6.2)–(6.3). Thus no attainment assumption is needed for the capacity minima. The total loss is independent of the elimination order, although its individual terms need not be.

By algebraic polarization, Anari and Oveis Gharan (2017, Theorem 4.1) extend (6.4) to all real stable polynomials with nonnegative coefficients, with the factor $e^{-|q|} q^q$ for the pairing $\sum_\alpha \alpha! f_\alpha g_\alpha$; the supplement derives this extension from Theorem 6.3 (Sra, 2026b, Section S5.1).

6.1. **Retaining common linear constraints.** A contraction need not lose one binary-entropy factor at every coordinate. A common linear constraint can force the last variables to agree, so they can be summed without that loss. Retaining such coordinates gives an explicit correction, with no intermediate capacities to compute.

Proposition 6.4 (Common linear constraints)**.** *Under the assumptions of Theorem 6.3, suppose that a real matrix W and a vector c satisfy*

$$W\mathbf{1}_A = c \quad \text{whenever } f_A > 0 \text{ or } g_A > 0.$$

If the columns of W indexed by $S \subseteq [m]$ are linearly independent, then, for every $q \in [0,1]^m$,

$$\sum_A f_A g_A \geq \exp\left(-\sum_{j \notin S} h(q_j)\right) \operatorname{Cap}_q(f) \operatorname{Cap}_q(g). \tag{6.7}$$

In particular, when f, g are homogeneous of the same degree,

$$\sum_A f_A g_A \geq q^q (\mathbf{1}-q)^{\mathbf{1}-q} \operatorname{Cap}_q(f) \operatorname{Cap}_q(g) \exp\left(\max_j h(q_j)\right). \tag{6.8}$$

Proof. If either capacity is zero, the claim is immediate. Otherwise we start with $\Phi(y,z) = f(y)g(z)$ and contract every coordinate outside S. A surviving term originated from a pair A, B that agrees outside S. The common constraints give $W_S(\mathbf{1}_{A\cap S} - \mathbf{1}_{B\cap S}) = 0$, so independence of the selected columns forces $A \cap S = B \cap S$. Thus the remaining polynomial Q contains only terms $y^T z^T$, and its full contraction is $Q(\mathbf{1},\mathbf{1})$. Evaluation at $(\mathbf{1},\mathbf{1})$ gives $Q(\mathbf{1},\mathbf{1}) \geq \operatorname{Cap}_{(q_S,q_S)}(Q)$. Applying (6.6) to the coordinates outside S gives (6.7). For homogeneous polynomials we take $W = (1,\dots,1)$ and leave uncontracted one coordinate that maximizes $h(q_j)$. □

If both polynomials have the same degree in each block of a partition, we may retain one coordinate from each block. The correction in (6.8) then becomes the sum of the largest binary entropy in each block.

6.2. **The Bethe bound and Schrijver's inequality.** We represent a permanent as a pairing by giving each edge one variable in a row selector and one in a column selector. The common monomials are then exactly the perfect matchings, and Anari and Oveis Gharan (2017, Section 3.1) derive Schrijver's inequality from their pairing bound in this way. The Bethe lower bound is stated through the following functional: for $B > 0$ and a doubly stochastic P, we write

$$H_{\mathrm{B}}(P) = \sum_{i,j} [-p_{ij}\log p_{ij} + (1-p_{ij})\log(1-p_{ij})], \qquad \beta_B(P) = \langle P, \log B\rangle + H_{\mathrm{B}}(P).$$

The entropy conventions extend these expressions to zero entries whenever $\operatorname{supp} P \subseteq \operatorname{supp} B$.

For the permanent, a switching argument supplies every local inequality needed by the binary-entropy comparison, without stability. We record this elementary form of the monomer–dimer correlation identity of Heilmann and Lieb (Heilmann and Lieb, 1972).

Lemma 6.5 (Matching switching)**.** *Let $G = (R \sqcup C, F)$ be bipartite, with nonnegative edge weights w_e and vertex weights ξ_v, and put*

$$Z_G = \sum_{M \text{ matching in } G} \prod_{e \in M} w_e \prod_{v \notin V(M)} \xi_v.$$

If $r \in R$ and $c \in C$, then

$$Z_G Z_{G-r-c} \geq Z_{G-r} Z_{G-c}. \tag{6.9}$$

All deleted graphs inherit the remaining vertex and edge weights, and $Z_\varnothing = 1$. More precisely, the defect is a sum over the simple paths γ from r to c,

$$Z_G Z_{G-r-c} - Z_{G-r} Z_{G-c} = \sum_{\gamma: r \rightsquigarrow c} \left(\prod_{e \in \gamma} w_e \right) Z^2_{G - V(\gamma)}. \tag{6.10}$$

Proof. Color a matching of $G - r$ red and a matching of $G - c$ blue. Their union consists of alternating paths and cycles, with a doubled edge regarded as a cycle. The component containing r cannot contain c, because a path joining them would start blue and end red, so would have even length, whereas every such path in G has odd length. Interchange the colors on this component. The result is a red matching of G and a blue matching of $G - r - c$. This map preserves the product of the edge and vertex weights, and it is injective because the same interchange recovers its input. When r is isolated in the union, its unmatched-vertex weight simply moves from the blue matching to the red matching. Summing the preserved weights proves (6.9).

The pairs outside the image have an alternating path from r to c, starting and ending red. This path is unique and has odd length. For a fixed path γ, its colors are forced; deleting it leaves two arbitrary matchings of $G - V(\gamma)$. Their total weight is the corresponding summand in (6.10), which proves the identity. □

Corollary 6.6 (A Bethe bound with a binary-entropy correction)**.** *Let $B \geq 0$ and let P be doubly stochastic with* $\operatorname{supp} P \subseteq \operatorname{supp} B$*. If $E_1, \dots, E_s$ are the edge sets of the connected components of the bipartite support graph of B, then*

$$\log \operatorname{per} B \geq \beta_B(P) + \sum_{a=1}^{s} \max_{e \in E_a} h(p_e). \tag{6.11}$$

In particular, dropping the correction recovers the Bethe lower bound. For every doubly stochastic A,

$$\operatorname{per}(A \odot (\mathbf{1}\mathbf{1}^T - A)) \geq \prod_{i,j} (1 - a_{ij}). \tag{6.12}$$

Proof. We follow the contraction argument of Theorem 6.3, with matching switching in place of stability. We use variables indexed by the positive entries of B, and put

$$f(y) = \prod_i \left(\sum_{j: b_{ij} > 0} b_{ij} y_{ij} \right), \qquad g(z) = \prod_j \left(\sum_{i: b_{ij} > 0} z_{ij} \right).$$

Their coefficient pairing equals $\operatorname{per} B$. To apply the binary-entropy comparison without a stability assumption, we contract an arbitrary set F of edge-variable pairs and write $G_F = (R \sqcup C, F)$. Expansion of the row and column selectors gives the exact invariant

$$\left(\prod_{e \in F} \mathcal{C}_e \right) f(y) g(z) = Z_{G_F}(\xi), \qquad \xi_{r_i} = \sum_{j: (i,j) \notin F} b_{ij} y_{ij}, \quad \xi_{c_j} = \sum_{i: (i,j) \notin F} z_{ij}.$$

Here Z_{G_F} carries the edge weights b_e and the vertex weights ξ, and the sums use only the support of B. Indeed, a contracted edge must be selected at both ends or at neither end. The edges selected at both ends form a matching, and each unmatched vertex selects one of its remaining half-edges.

For an uncontracted edge $e = (r, c)$, we fix all other variables positively and write $\xi_r = u + b_e y_e$, $\xi_c = v + z_e$. The resulting slice is $a y_e z_e + b y_e + c z_e + d$, with

$$a = b_e Z_{G_F - r - c}, \qquad b = b_e Z_{G_F - r}, \qquad c = Z_{G_F - c}, \qquad d = Z_{G_F},$$

where the vertex weights at r, c in these four partition functions are u, v. Lemma 6.5 therefore gives $ad \geq bc$. We apply Lemma 6.2 at each contraction and take infima over the remaining variables, exactly as in (6.6). The terminal common-constraint argument of Proposition 6.4 uses only the

surviving monomials, so it also applies here. Thus the component correction follows from binary entropy and matching switching alone.

Since each variable occurs in only one factor of either polynomial, weighted arithmetic–geometric mean gives

$$\mathrm{Cap}_P(f) = \prod_{i,j} \left(\frac{b_{ij}}{p_{ij}}\right)^{p_{ij}}, \qquad \mathrm{Cap}_P(g) = \prod_{i,j} p_{ij}^{-p_{ij}}.$$

Equality is attained or approached at the corresponding row or column proportions. Products are over the support of B, and zero exponents are interpreted by continuity. Every support component has equally many row and column vertices, because P has unit row and column sums. Thus f, g have the same degree in each component's variables. In the preceding contraction argument, we retain an edge of largest binary entropy in each component. The logarithm of the uncorrected right side is

$$\sum_{i,j} [p_{ij} \log b_{ij} - p_{ij} \log p_{ij} + (1 - p_{ij}) \log(1 - p_{ij})] = \beta_B(P),$$

which proves (6.11).

For $0 < a_{ij} < 1$, we take $P = A$ and $b_{ij} = a_{ij}(1 - a_{ij})$, and drop the nonnegative correction. The exponent simplifies to $\beta_B(A) = \sum_{i,j} \log(1-a_{ij})$, proving (6.12). Approximation by $(1-\varepsilon)A+\varepsilon \mathbf{1}\mathbf{1}^T/n$ and continuity handle the boundary; the case $n = 1$ is immediate. □

Equation (6.12) is Schrijver's inequality (Schrijver, 1998). Its variational form without the correction is due to Gurvits (Gurvits, 2012, Theorem 2.2), and Anari and Oveis Gharan derive it from their pairing bound. Here the switching lemma replaces their stability input, and Lemma 6.2 retains the entropy loss that gives the component correction. Through the same multiaffine inequality of Anari and Oveis Gharan, Anari (2026, Theorem 6) multiplies the Bethe lower bound by gains from pairs of rows and obtains a deterministic approximation of the permanent within a factor c^n for some $c < \sqrt{2}$ (Anari, 2026, Theorem 1). Row averaging does not make $\mathsf{H}_{\mathrm{perm}} - H_{\mathrm{B}}$ monotone (Appendix A.3), so the correction comes from the binary comparison rather than from Theorem 2.1. The switching inequality also enters the entropy interpolation of Dong and Jain (2026, Claim 2.5 and Theorem 2.1), whose parameter is the probability that a vertex is covered by a matching.

The path identity locates the local interaction, since it gives

$$ad - bc = b_e \sum_{\gamma : r \rightsquigarrow c \text{ in } G_F} \left(\prod_{f \in \gamma} b_f\right) Z^2_{G_F - V(\gamma)}.$$

The local determinant is therefore positive only if the contracted edge closes a cycle with previously contracted edges, and with positive vertex weights every such path contributes strictly. These cycle terms enter the entropy-production formula (6.3) through $\lambda = \log(ad/(bc))$.

The same correction applies directly to the maximum permutation entropy $\mathsf{H}_{\mathrm{perm}}$. Let $E_1, \dots, E_s$ now be the support components of a doubly stochastic P. Finite Gibbs duality on this face gives

$$\mathsf{H}_{\mathrm{perm}}(P) = \inf_{\substack{b_{ij} > 0 \text{ on } \operatorname{supp} P \\ b_{ij} = 0 \text{ elsewhere}}} \{\log \operatorname{per} B - \langle P, \log B \rangle\}.$$

Indeed, the same maximum-entropy argument as in Lemma 2.2 applies after restricting to the allowed permutations and removing their affine relations. Taking this infimum in (6.11) yields

$$\mathsf{H}_{\mathrm{perm}}(P) \geq H_{\mathrm{B}}(P) + \sum_{a=1}^{s} \max_{e \in E_a} h(p_e). \tag{6.13}$$

This bound is exact whenever every support component is a single edge or an even cycle. A cycle has two perfect matchings; its marginals alternate between t and $1 - t$, so its maximum entropy is $h(t)$,

its Bethe entropy is zero, and its correction is $h(t)$. Distinct components can be made independent, which adds their maximum entropies. A single edge contributes zero.

For example, let B be the adjacency matrix of an r-regular bipartite graph on $n+n$ vertices, with s connected components and $r \geq 2$. Taking $P = B/r$ gives

$$\operatorname{per} B \geq e^{sh(1/r)} \left(\frac{(r-1)^{r-1}}{r^{r-2}} \right)^n .$$

When $r = 2$, every component is an even cycle, and the bound 2^s is exact.

7. Further directions

Effective remainders. The explicit quadratic corrections of Section 5.1 keep the slack of the partial bounds at order two. A different route is to bound the negative kernel eigenvalues in Corollary 4.5 directly from the input data, which could give effective entropy remainders even when the minimum entry or eigenvalue vanishes.

Degree-sensitive coefficient bounds. Let p be a stable homogeneous polynomial of degree n in n variables with nonnegative coefficients and positive capacity. Gurvits's degree-sensitive bound is

$$[x_1 \cdots x_n]p \geq \operatorname{Cap}_{\mathbf{1}}(p) \prod_{i=1}^{n} g(\min\{i, \deg_{x_i} p\}), \qquad g(r) = \left(\frac{r-1}{r} \right)^{r-1}, \quad g(1) = 1,$$

with an arbitrary ordering of the variables (Gurvits, 2008). This refinement does not hold for all Lorentzian polynomials.

To see this, we set $s = y + z$ and consider

$$f(x, y, z) = s^3 + 3xs^2 + 3x^2 s.$$

Its support is $\{\alpha \in \mathbb{Z}^3_{\geq 0} : |\alpha| = 3,\ \alpha_x \leq 2\}$, an M-convex set. Moreover,

$$\partial_x f = 3s^2 + 6xs, \qquad \partial_y f = \partial_z f = 3(x+s)^2.$$

The first Hessian has one positive, one negative, and one zero eigenvalue, and the other two have rank one and are positive semidefinite. Thus f is Lorentzian. Its mixed coefficient is $[xyz]f = 6$, whereas $yz \leq s^2/4$ gives

$$\operatorname{Cap}_{(1,1,1)}(f) = 4 \inf_{t>0} (t^{-1} + 3 + 3t) = 4(3 + 2\sqrt{3}).$$

Equality in the first reduction is attained at $y = z$, and the one-variable minimum is attained at $t = x/s = 1/\sqrt{3}$. Ordering the variables as y, z, x gives degree factors $g(1)g(2)g(2) = 1/4$, which would require $6 \geq 3 + 2\sqrt{3}$. An averaging proof of the degree-sensitive bound must therefore use additional structure of stable polynomials.

Pairings beyond bistability. The diagonal contractions in Section 6 supply the additional stable-polynomial input needed for the Anari–Oveis Gharan pairing inequalities. Corollary 4.7 applies to a different class, a Lorentzian factor paired with a product of linear forms. Extending its prescribed-marginal comparison to two general Lorentzian factors requires a replacement for bistable contraction; the coefficient-pairing problem is discussed by Gurvits and Leake (2021, Conjecture 6.1).

The direct coefficientwise product does not preserve the quadratic condition required by row averaging. Indeed, both

$$f = (x_1 + x_2)(x_3 + x_4), \qquad g = (x_1 + x_3)(x_2 + x_4)$$

are stable, but the product of their corresponding coefficients is

$$\sum_{\alpha} [x^\alpha] f \, [x^\alpha] g \, x^\alpha = x_1 x_4 + x_2 x_3.$$

Its Hessian has two positive eigenvalues. Retaining separate variables for the two factors avoids this product in the stable case. In the matching specialization, switching supplies the required local

inequality without stability, which suggests seeking correlation inequalities for other supports that allow binary averaging to be iterated.

Real linear polarization. The factorial constant of Corollary 4.3 does not give the real polarization inequality for products of linear functionals on a sphere, because signs enter as well. The supplement isolates this obstruction and proves a variational special case (Sra, 2026b, Section S6.3).

AI USAGE

The results in this paper were developed through extensive use of an AI assistant (OpenAI's GPT Pro, in a session that identified itself as GPT-6 Astra Pro). None of the claims were obtained zero-shot. The author's proposal of a Schur-type averaging statement for permanents led to Theorem 2.1; at the author's request the argument was compared with the proofs of Falikman, Egorychev, and Gurvits, and tested on mixed volumes and Lorentzian polynomials, leading to Sections 4–5. The AI assistant carried out the derivations, the literature search, and checks of the arguments. The partial-matching and binary-averaging results (Sections 3 and 6) were developed later in the same way, at the author's request, with AI assistants carrying out the derivations and independent checks. The author also used AI assistants (GPT and Claude) to polish the writing.

ACKNOWLEDGEMENTS

The author (SS) thanks the Alexander von Humboldt (AvH) Foundation for their generous support via a Humboldt Professorship in AI.

REFERENCES

Shipra Agrawal, Zizhuo Wang, and Yinyu Ye. Parimutuel betting on permutations. In *Internet and Network Economics (WINE 2008)*, volume 5385 of *Lecture Notes in Computer Science*, pages 126–137. Springer, 2008. doi: 10.1007/978-3-540-92185-1_21.

Nima Anari. Beyond the Bethe approximation of the permanent. Preprint, arXiv:2608.28031v2, 2026.

Nima Anari and Shayan Oveis Gharan. A generalization of permanent inequalities and applications in counting and optimization. Preprint, arXiv:1702.02937, 2017.

Nima Anari, Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. Nash social welfare, matrix permanent, and stable polynomials. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:12, 2017. doi: 10.4230/LIPIcs.ITCS.2017.36. Theorem numbering refers to arXiv:1609.07056v2.

Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials I: entropy and a deterministic approximation algorithm for counting bases of matroids. *Duke Math. J.*, 170:3459–3504, 2021. doi: 10.1215/00127094-2020-0091. Theorem numbering refers to arXiv:1807.00929.

D. N. Bernshtein. The number of roots of a system of equations. *Funct. Anal. Appl.*, 9:183–185, 1975. doi: 10.1007/BF01075595.

Joseph E. Bonin, Carolyn Chun, and Tara Fife. The natural matroid of an integer polymatroid. *SIAM J. Discrete Math.*, 37:1751–1770, 2023. doi: 10.1137/22M1521122. Preprint: arXiv:2209.03786.

Julius Borcea, Petter Brändén, and Thomas M. Liggett. Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.*, 22:521–567, 2009. doi: 10.1090/S0894-0347-08-00618-8.

Petter Brändén and June Huh. Lorentzian polynomials. *Ann. of Math. (2)*, 192:821–891, 2020. doi: 10.4007/annals.2020.192.3.4.

L. M. Brégman. Some properties of nonnegative matrices and their permanents. *Dokl. Akad. Nauk SSSR*, 211(1):27–30, 1973. URL https://www.mathnet.ru/eng/dan37755. In Russian; English translation in Soviet Math. Dokl. **14** (1973), 945–949.

Gi-Sang Cheon and Ian M. Wanless. An update on Minc's survey of open problems involving permanents. *Linear Algebra Appl.*, 403:314–342, 2005. doi: 10.1016/j.laa.2005.02.030.

Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. *Ann. Probab.*, 49(1):506–525, 2021. doi: 10.1214/20-AOP1453. Preprint: arXiv:1903.06081.

Dingding Dong and Vishesh Jain. A deterministic $(1+\varepsilon)^n$ approximation for the permanent of a nonnegative matrix. Preprint, arXiv:2609.11049v1, 2026.

Jack Edmonds. Matroid intersection. *Ann. Discrete Math.*, 4:39–49, 1979. doi: 10.1016/S0167-5060(08)70817-3.

G. P. Egorychev. Proof of the van der Waerden conjecture for permanents. *Siberian Math. J.*, 22 (6):854–859, 1981. doi: 10.1007/BF00968054. Translated from Sibirsk. Mat. Zh. **22** (1981), no. 6, 65–71.

D. I. Falikman. Proof of the van der Waerden conjecture regarding the permanent of a doubly stochastic matrix. *Math. Notes*, 29(6):475–479, 1981. doi: 10.1007/BF01163285. Translated from Mat. Zametki **29** (1981), no. 6, 931–938.

Shmuel Friedland. A proof of a generalized van der Waerden conjecture on permanents. *Linear and Multilinear Algebra*, 11(2):107–120, 1982. doi: 10.1080/03081088208817436.

Shmuel Friedland and Leonid Gurvits. Generalized Friedland–Tverberg inequality: applications and extensions, 2006. URL https://arxiv.org/abs/math/0603410. arXiv:math/0603410, version 2; published in revised form as "Lower bounds for partial matchings in regular bipartite graphs and applications to the monomer–dimer entropy", Combin. Probab. Comput. 17 (2008) 347–361.

Leonid Gurvits. The van der Waerden conjecture for mixed discriminants. *Advances in Mathematics*, 200(2):435–454, 2006. doi: 10.1016/j.aim.2004.12.002. Preprint: arXiv:math/0406420.

Leonid Gurvits. Van der Waerden/Schrijver–Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. *Electron. J. Combin.*, 15(1):Research Paper 66, 2008. doi: 10.37236/790. Preprint: arXiv:0711.3496.

Leonid Gurvits. On multivariate Newton-like inequalities. In *Advances in Combinatorial Mathematics*, pages 61–78. Springer, 2009a. doi: 10.1007/978-3-642-03562-3_4. Theorem numbering refers to arXiv:0812.3687.

Leonid Gurvits. A polynomial-time algorithm to approximate the mixed volume within a simply exponential factor. *Discrete Comput. Geom.*, 41:533–555, 2009b. doi: 10.1007/s00454-009-9147-5. Theorem numbering refers to arXiv:cs/0702013v4.

Leonid Gurvits. Unleashing the power of Schrijver's permanental inequality with the help of the Bethe approximation. Preprint, arXiv:1106.2844v11, 2012.

Leonid Gurvits and Jonathan Leake. Counting matchings via capacity-preserving operators. *Combin. Probab. Comput.*, 30:956–981, 2021. doi: 10.1017/S0963548321000122.

Leonid Gurvits and Jonathan Leake. Unique minimizers for permanents, mixed discriminants, and log-concave polynomials. Preprint, arXiv:2609.21083, 2026. Revision 2, 15 October 2026.

Ole J. Heilmann and Elliott H. Lieb. Theory of monomer-dimer systems. *Communications in Mathematical Physics*, 25(3):190–232, 1972. doi: 10.1007/BF01877590.

Thorkell Helgason. Aspects of the theory of hypermatroids. In *Hypergraph Seminar*, volume 411 of *Lecture Notes in Math.*, pages 191–213. Springer, Berlin, 1974.

Yair Lavi. The Marcus–Minc transform inequality, 2026. URL https://arxiv.org/abs/2609.29262. Preprint, arXiv:2609.29262, version 1.

Joonkyung Lee and Jaehyeon Seo. Antiferromagnetic models are clique-minimizing. Preprint, arXiv:2608.17920, 2026.

C. J. H. McDiarmid. Independence structures and submodular functions. *Bull. London Math. Soc.*, 5:18–20, 1973.

O. S. Rothaus. Study of the permanent conjecture and some generalizations. *Bull. Amer. Math. Soc.*, 78(5):749–752, 1972. doi: 10.1090/S0002-9904-1972-13017-9.

Alexander Schrijver. Counting 1-factors in regular bipartite graphs. *Journal of Combinatorial Theory, Series B*, 72(1):122–135, 1998. doi: 10.1006/jctb.1997.1798.

Paul B. Slater. Maximum entropy convex decompositions of doubly stochastic and nonnegative matrices. *Environment and Planning A*, 19(3):403–407, 1987. doi: 10.1068/a190403.

Suvrit Sra. A sharp Brégman–Minc bound for permanents. Companion preprint, 2026a.

Suvrit Sra. Supplement: Gamma–Orlicz upper bounds for mixed derivatives. Supplement to the companion preprints, 2026b.

Damian Straszak and Nisheeth K. Vishnoi. Real stable polynomials and matroids: optimization and counting. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC 2017)*, pages 370–383, 2017a. doi: 10.1145/3055399.3055457. Appendix numbering refers to arXiv:1611.04548.

Damian Straszak and Nisheeth K. Vishnoi. On convex programming relaxations for the permanent. Preprint, arXiv:1701.01419, 2017b.

Joseph S. Verducci. Minimum majorization decomposition. In *Contributions to Probability and Statistics*, pages 160–173. Springer, New York, 1989. doi: 10.1007/978-1-4612-3678-8_12.

Appendix A. Counterexamples

Theorem 2.1 is stated for chains of pair averages of entire rows or columns. The examples below show that this restriction is needed, and that the Bethe entropy gap of Section 6.2 is not monotone even under such averages.

A.1. **Majorization of the entries.** The order generated by pair averages is not majorization of the flattened list of entries. For example, let S be a cyclic permutation matrix of order four. The doubly stochastic matrices

$$A = \operatorname{diag}(J_2/2, J_2/2), \qquad B = (I_4 + S)/2$$

both have eight entries equal to $1/2$ and eight zero entries, while their maximum-entropy permutation laws are uniform on four and two supported permutations, respectively. Consequently

$$\mathcal{I}(A) = 2\log 2, \qquad \mathcal{I}(B) = 3\log 2,$$

so $\mathcal{I}$ is not a symmetric function of the flattened entries, as ordinary Schur-concavity would require.

A.2. **Concavity and arbitrary stochastic mixing fail.** Nor can the pair-averaging theorem be replaced by a concavity assertion. Let $B(t)$ be the 3×3 matrix with diagonal t and off-diagonal entries 1. Its Gibbs marginal matrix $P(x)$ has diagonal

$$x = \frac{t(t^2+1)}{t^3+3t+2}$$

and off-diagonal entries $(1-x)/2$. If N is the number of fixed points under this Gibbs law, differentiation of the exponential family gives

$$\frac{d^2}{dx^2}\mathcal{I}(P(x)) = \frac{9}{\operatorname{Var} N} - \frac{3}{x(1-x)},$$

because $\mathbb{E}N = 3x$, $d(3x)/d\log t = \operatorname{Var} N$, and $d\mathsf{H}_{\mathrm{perm}}(P(x))/dx = -3\log t$. At $t = 1/3$,

$$x = \frac{5}{41}, \qquad \operatorname{Var} N = \frac{513}{1681}, \qquad \frac{d^2}{dx^2}\mathcal{I}(P(x)) = \frac{1681}{1140} > 0.$$

Arbitrary doubly stochastic left multiplication cannot replace a chain of pair T-transforms either. Fix $r = 3/4$, and put

$$B_\varepsilon = \begin{pmatrix} \varepsilon & 1 & 1 \\ 1 & \varepsilon & 1 \\ r & 1 & \varepsilon \end{pmatrix}, \qquad Z_\varepsilon = 1 + r + (2+r)\varepsilon + \varepsilon^3.$$

Its cofactor matrix is

$$C_\varepsilon = \begin{pmatrix} 1+\varepsilon^2 & r+\varepsilon & 1+r\varepsilon \\ 1+\varepsilon & r+\varepsilon^2 & r+\varepsilon \\ 1+\varepsilon & 1+\varepsilon & 1+\varepsilon^2 \end{pmatrix},$$

so $P_\varepsilon = (B_\varepsilon \odot C_\varepsilon)/Z_\varepsilon$ is an interior doubly stochastic matrix. Let Π cyclically reorder its rows as $(2,3,1)$. Formula (2.1) gives

$$d\mathcal{I}(P_\varepsilon)[\Pi P_\varepsilon - P_\varepsilon] = -\sum_{i,j} (\Pi P_\varepsilon - P_\varepsilon)_{ij} \log (C_\varepsilon)_{ij}.$$

As $\varepsilon \downarrow 0$,

$$P_\varepsilon \longrightarrow \frac{1}{1+r} \begin{pmatrix} 0 & r & 1 \\ 1 & 0 & r \\ r & 1 & 0 \end{pmatrix},$$

and the limiting matrix $-\log C_\varepsilon$ is zero except at $(1,2),(2,2),(2,3)$, where it equals $-\log r$. Therefore

$$\lim_{\varepsilon \downarrow 0} d\mathcal{I}(P_\varepsilon)[\Pi P_\varepsilon - P_\varepsilon] = (-\log r)\frac{1-2r}{1+r} = -\frac{2}{7}\log\frac{4}{3} < 0.$$

Hence the doubly stochastic matrix $(1-t)I + t\Pi$ decreases $\mathcal{I}$ for small $\varepsilon > 0$ and then small $t > 0$, so the pair-averaging order in Theorem 2.1 is a genuine restriction.

A.3. **The Bethe entropy gap is not monotone.** Let $\mathcal{J}(P) = \mathsf{H}_{\mathrm{perm}}(P) - H_{\mathrm{B}}(P)$, with H_{B} as in Section 6.2. Equation (6.13) bounds this gap by diagonal contractions. Row averaging need not increase this gap, even in the interior of the doubly stochastic polytope.

Consider the Gibbs matrix B, with $\operatorname{per} B = 377$, and its marginals P:

$$B = \begin{pmatrix} 3 & 3 & 1 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 1 & 3 & 3 \\ 2 & 3 & 1 & 1 \end{pmatrix}, \qquad P = \frac{1}{377}\begin{pmatrix} 168 & 123 & 32 & 54 \\ 48 & 39 & 122 & 168 \\ 39 & 32 & 183 & 123 \\ 122 & 183 & 40 & 32 \end{pmatrix}.$$

Let u, v be its first two rows and x, y the first two rows of B. Along the path replacing u, v by $(1-t)u + tv,\ tu + (1-t)v$, the entropy differential gives

$$\frac{d}{dt}\mathcal{J}(P(t))\Big|_{t=0} = \sum_j (v_j - u_j) \log \frac{(1-u_j)y_j u_j}{(1-v_j)x_j v_j}.$$

Substitution yields the exact expression

$$377\ \frac{d}{dt}\mathcal{J}(P(t))\Big|_{t=0} = 6 \log R, \qquad R = \left(\frac{282}{209}\right)^{20}\left(\frac{6591}{5207}\right)^{14}\left(\frac{736}{1037}\right)^{15}\left(\frac{459}{616}\right)^{19} < \frac{1}{4}.$$

The final inequality follows by clearing the positive denominators. Thus $\mathcal{J}$ decreases under a sufficiently small pair average. This counterexample concerns the proposed monotonicity; the Bethe lower bound remains valid.