Vector Packing and Sharp Bounds for Matrices of Bounded Rank

Yair Lavi

14 November 2026

Abstract

We prove a packing theorem in real and complex normed spaces. Given a possibly infinite family of vectors (v_j) whose span has finite dimension $r \geq 1$, let M be the supremum of the norms of its finite linear combinations with coefficients of modulus at most one. If M is finite, the family splits into r nonempty sets, each with sum of norms at most M. The same partition works for every norm on the span, using that norm's value of M. We also determine exactly which finite lists of nonnegative numbers can be realized as the norms of a family of capacity one in some normed space of dimension at most k. We also obtain sharp rank-dependent bounds for products of the moduli of the entries on generalized diagonals and the permanent of the entrywise modulus, extending the known nonnegative-matrix bounds to real and complex matrices. The packing theorem also gives sharp support and entropy bounds for doubly stochastic matrices of bounded rank.

Keywords. Vector packing; matrix rank; majorization; stochastic matrix; permanent; entropy.
2020 Mathematics Subject Classification. Primary 46B07; Secondary 15A15, 15A45, 15B51, 52A21.

1 Introduction and main theorem

1.1 From a bound on the sum of norms to a partition

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(v_j)_{j \in J}$ be a labeled family in a normed space over $\mathbb{F}$, and put $E = \operatorname{span}_{\mathbb{F}}\{v_j : j \in J\}$. Repeated vectors may have different labels. A sum of nonnegative numbers over an arbitrary label set means the supremum of its finite subsums. All logarithms in this paper are natural. Define the *coefficient-sum capacity* by

$$M = M((v_j), \|\cdot\|) := \sup_{\substack{H \subseteq J \text{ finite} \\ |\alpha_j| \leq 1}} \left\| \sum_{j \in H} \alpha_j v_j \right\|. \tag{1.1}$$

This is the supremum of the norms obtainable by choosing signs or complex phases for finitely many vectors. Indeed, on each finite set H, convexity in each coefficient permits restriction to $\alpha_j \in \{-1, 1\}$ over $\mathbb{R}$, or to $|\alpha_j| = 1$ over $\mathbb{C}$. If E^* denotes the space of continuous linear functionals on E, with $\|\varphi\| = \sup_{\|v\| \leq 1} |\varphi(v)|$, duality gives

$$M = \sup_{\substack{\varphi \in E^* \\ \|\varphi\| \leq 1}} \sum_{j \in J} |\varphi(v_j)|. \tag{1.2}$$

To see this, express the norm in (1.1) as a supremum over φ, interchange the suprema, and align the signs or phases of the scalars $\varphi(v_j)$. The quantity M is the weak ℓ_1 norm of the family;

we refer to it here as its capacity. For finite families, (1.2) is the quantity μ_1 of [6, Section 2]. The equality of (1.1) and (1.2), and the restriction to signs or unit phases, are recorded in [6, 2.2].

For nonnegative numbers x_j, the *load* of a set F is $\sum_{j\in F} x_j$. A packing with load bound M is a partition of the labels into nonempty sets, each of load at most M. For a vector family, the load of a set F is $\sum_{j\in F} \|v_j\|$. For a finite family with $M > 0$, its *normalized norm profile* is $(\|v_j\|/M)_{j=1}^n$.

When $\dim E = r$, the classical dimension bound is $\sum_j \|v_j\| \le rM$; its relation to absolutely summing norms is recalled in Section 3.1. A bound on the total does not by itself say how the vectors can be partitioned. For example, three numbers equal to $3/5$ have total $9/5 < 2$, but cannot be placed in two sets of load at most one: every pair has load $6/5$. The result below shows that a family of vectors with capacity one whose span has dimension two cannot have this profile of norms. Our main theorem is the following.

1.2 One partition for every norm

Theorem 1.1 (vector packing)**.** *Let $(v_j)_{j\in J}$ be a family in a normed space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and put $E = \operatorname{span}_{\mathbb{F}}\{v_j : j \in J\}$. Suppose $1 \le r = \dim_{\mathbb{F}} E < \infty$ and the capacity M defined in* (1.1) *is finite. Then J has a partition into exactly r nonempty sets $F_1, \dots, F_r$ such that*

$$\sum_{j\in F_a} \|v_j\| \le M \qquad (1 \le a \le r). \tag{1.3}$$

Only countably many v_j are nonzero, and their norms are summable. The partition can be chosen independently of the norm on E: (1.3) *holds for every norm on E, with its own capacity* (1.1)*.*

For example, the columns of a nonzero matrix admit one partition for both $\|x\|_\infty = \max_i |x_i|$ and $\|x\|_1 = \sum_i |x_i|$, computing the capacity separately for each norm. The bound concerns the sum of the individual norms in each set, rather than the norm of their sum.

1.3 Sharpness

Both the number of sets and the load bound M are sharp. Let $e_1, \dots, e_r$ be the standard coordinate vectors in $\mathbb{F}^r$, with e_a equal to one in coordinate a and zero elsewhere. In the supremum norm, this family has capacity one, and each vector has norm one. No two can share a set of load at most one. The zero-span case has zero load; we exclude it from Theorem 1.1 because zero nonempty sets cannot partition a nonempty label set.

1.4 Consequences

The packing theorem has consequences for vector norms, matrix inequalities, and information bounds. A *row-stochastic* matrix has nonnegative entries and every row sum equal to one. A *column-stochastic* matrix is defined by the corresponding column condition; a matrix is *doubly stochastic* if both conditions hold.

- **Small norms and signed sums.** It bounds the smallest norm in a finite family, gives large signed sums or sums with unit phases (Corollary 3.1), and refines the classical bound on coordinate variations (Section 3.1).
- **Rank and eigenvalue location.** Column maxima and relative diagonals give packing bounds on rank (Corollary 3.2 and Theorem 3.3), recover classical nonsingularity criteria, and locate eigenvalues according to their geometric multiplicity (Corollary 3.4).

- **Attainable profiles and norm inequalities.** It characterizes the normalized norm profiles attainable as the family and space vary, and the possible stochastic column maxima (Theorem 4.1). It also gives sharp majorization, symmetric-function, and weighted-product bounds (Theorem 4.2, Corollary 4.3, and Theorem 4.4).
- **Diagonal products and permanents.** It recovers the known sharp stochastic bounds, extends them to real and complex matrices (Corollary 5.1 and Theorem 5.2), and controls powers and partial sums of absolute diagonal products (Corollary 5.3).
- **Prescribed diagonals.** It determines the minimum absolute entrywise mass for prescribed diagonal magnitudes (Theorem 5.4) and characterizes the attainable diagonals of row-stochastic matrices under a rank bound (Theorem 6.2).
- **Doubly stochastic matrices.** It gives sharp bounds on entropy, support size, and entrywise power sums in the stated exponent ranges (Theorem 6.3).
- **Mutual information.** It gives bounds depending on rank and the marginals, with sharpness when one marginal is prescribed and the other may vary (Theorem 6.4).

1.5 Relation to earlier work

In 2018, the author conjectured sharp rank-dependent maxima for diagonal products and permanents of stochastic matrices [8]. The sharp maxima and equality cases were established in [9] for diagonal products and in [10] for the permanent. Here we derive these maxima from vector packing and extend the inequalities to real and complex matrices, with rank imposed on the original matrix. The equality classifications remain separate results; we do not classify complex equality cases.

Theorem 1.1 refines the classical dimension estimate recalled in Section 1.1 [6, p. 33, item (4)], [7, (5)] by controlling the whole family through a partition. For n vectors in a real normed space, the small-norm bound in Corollary 3.1 is equivalent to the signed-sum theorem of Ambrus and Grundbacher [1, Theorems 2 and 7]. This consequence already supplies the small-column step in the real-matrix proof of the permanent bound. The full partition theorem holds over both fields, includes infinite families of finite capacity, and chooses one partition valid for every norm.

1.6 Organization

Section 2 proves Theorem 1.1. Sections 3 and 4 develop the partition consequences and the majorization bounds; Sections 5 and 6 apply them to matrix inequalities and stochastic matrices. Further comparisons with earlier inequalities are made at their point of use.

2 Proof of Theorem 1.1

The proof first constructs a partition and numerical weights using only linear algebra. For a finite family, we choose a partition and scalars of modulus one to maximize the absolute determinant of the resulting r set sums. Changing one phase forces the assigned coordinate weight to be nonnegative; moving one vector to another set forces that weight to dominate its other coordinates. The weights sum to one in each set. A recovery identity then converts these coordinate bounds into norm bounds, for every norm at once. Absolute summability and compactness extend the construction to infinite families.

2.1 A partition maximizing the determinant

Start with a finite family $v_1, \dots, v_n$ of nonzero vectors whose span has dimension r. Fix linear coordinates on E. For an assignment of the labels to r indexed sets $F_1, \dots, F_r$ (possibly empty) and a choice of scalars $\theta_j \in \mathbb{F}$ with $|\theta_j| = 1$, put

$$z_a = \sum_{j \in F_a} \theta_j v_j, \qquad D = \det[z_1 \ \cdots \ z_r]. \tag{2.1}$$

Over the reals, $|D|$ is the volume of the parallelepiped spanned by the set sums in the chosen coordinates. We seek an assignment whose volume cannot be increased by either of the two local changes described below. Over the complexes, the same determinant objective supplies the needed algebraic comparisons.

There is an assignment with $D \neq 0$. Choose linearly independent vectors $u_1, \dots, u_r$ from the family, let w be the sum of the remaining vectors, and define

$$D_0 = \det[u_1 \ u_2 \ \cdots \ u_r], \qquad D_1 = \det[w \ u_2 \ \cdots \ u_r].$$

Assign u_a to F_a with phase one, and all remaining vectors to F_1 with a common sign $\varepsilon \in \{-1, 1\}$. The resulting determinant is $D_0 + \varepsilon D_1$. Since $D_0 \neq 0$, at least one of these two choices is nonzero.

Allowing empty sets during the optimization, the assignments and unit phases form a compact space. Thus $|D|$ has a positive maximum. Fix a maximizing configuration. Its set sums form a basis, so its sets are in fact nonempty. Write $h_a : E \to \mathbb{F}$ for the coordinate functionals in this basis: $v = \sum_{a=1}^{r} h_a(v) z_a$.

2.2 Two local changes: changing a phase and moving a vector

Lemma 2.1 (coordinate weights)**.** *For the maximizing configuration chosen above, $D \neq 0$, and its basis coordinates satisfy*

$$d_j := \theta_j h_a(v_j) = \max_{1 \le b \le r} |h_b(v_j)| \in [0, 1] \quad (j \in F_a), \qquad \sum_{j \in F_a} d_j = 1. \tag{2.2}$$

Proof. By Cramer's rule,

$$\det[z_1 \ \cdots \ z_{a-1} \ v \ z_{a+1} \ \cdots \ z_r] = h_a(v) D \qquad (v \in E, \ 1 \le a \le r).$$

Fix $j \in F_a$ and write $d = \theta_j h_a(v_j)$. First change only its phase, from θ_j to $u\theta_j$ with $|u| = 1$. Column a in (2.1) becomes $z_a + (u-1)\theta_j v_j$. By multilinearity, the changed determinant D' satisfies

$$\frac{D'}{D} = 1 + (u-1)\theta_j h_a(v_j) = 1 - d + ud. \tag{2.3}$$

Over $\mathbb{R}$, the sign flip $u = -1$ and maximality give $|1 - 2d| \le 1$, hence $d \in [0, 1]$. Over $\mathbb{C}$, the largest modulus of (2.3) over the unit circle is $|1 - d| + |d|$, so this sum is at most one. Its reverse inequality follows from $(1 - d) + d = 1$. Equality in the triangle inequality forces d to be real and in $[0, 1]$, including the endpoints.

Next move j from F_a to F_b, $b \neq a$, with a new unit phase η. Columns a, b become $z_a - \theta_j v_j$ and $z_b + \eta v_j$. Expanding both columns, the term containing v_j twice vanishes. The remaining terms give

$$\frac{D'}{D} = 1 - \theta_j h_a(v_j) + \eta h_b(v_j) = 1 - d + \eta h_b(v_j). \tag{2.4}$$

The move is allowed even if it empties F_a: the resulting determinant is zero, as (2.4) also predicts. This covers singleton sets.

Since $1-d \geq 0$, choose η so that $\eta h_b(v_j) = |h_b(v_j)|$. Maximality gives $1-d+|h_b(v_j)| \leq 1$. Thus $|h_b(v_j)| \leq d$ for $b \neq a$, while $|h_a(v_j)| = d$. Finally, $\sum_{j \in F_a} d_j = h_a(z_a) = 1$. □

Remark. A global maximum is sufficient but not necessary. A configuration with $D \neq 0$ has the same coordinate properties if $|D|$ cannot be increased by changing one phase or moving one label with any new unit phase. The hypothesis $D \neq 0$ is essential to this assertion.

2.3 From coordinate weights to bounds for every norm

The purpose of the weights is to express each vector as a coefficient sum of the whole family, with all coefficients controlled by one number.

Lemma 2.2 (finite packing)**.** *For a finite family satisfying the hypotheses of Theorem 1.1, one partition satisfies* (1.3) *for every norm on E, using that norm's own capacity.*

Proof. Remove the zero vectors, which does not change the capacity, and apply the preceding construction to the label set J_+ of the nonzero vectors. For $\ell \in J_+$, let $a(\ell)$ be its assigned set index. Expansion in the basis z_b and substitution of the set sums give the recovery identity

$$v_j = \sum_b h_b(v_j) z_b = \sum_{\ell \in J_+} \theta_\ell h_{a(\ell)}(v_j) v_\ell. \tag{2.5}$$

Every coefficient in the last sum has modulus at most d_j. Since $v_j \neq 0$, (2.2) gives $d_j > 0$. Divide (2.5) by d_j; the coefficients now have modulus at most one. Definition (1.1) gives $\|v_j\| \leq M d_j$. Summing over F_a and using $\sum_{j \in F_a} d_j = 1$ proves (1.3).

The assignment, phases, and weights depend only on the vectors and the fixed coordinates. The same identity therefore proves the inequality for every norm on E. Assign the zero labels to any of the r nonempty sets, with weight zero. □

The recovery identity also gives a useful statement involving no norm: for every $\varphi \in E^*$,

$$|\varphi(v_j)| \leq d_j \sum_\ell |\varphi(v_\ell)|. \tag{2.6}$$

This follows by applying φ to (2.5) and using the triangle inequality. It explains why the weights can serve all norms at once.

Remark (Lewis's variational method)**.** In the real case this determinant maximization is an instance of Lewis's variational method [11]; see also Pisier [16, Theorem 3.1] for the formulation using an arbitrary norm. For any norm α on the real $r \times r$ matrices, a maximizer Z of $|\det U|$ over $\alpha(U) \leq 1$ is invertible and satisfies $\alpha(Z) = 1$ and

$$|\operatorname{tr}(Z^{-1}U)| \leq r\alpha(U) \qquad \text{for every } U,$$

with equality at $U = Z$. This follows by differentiating the determinant along line segments in the unit ball, then using symmetry and homogeneity. For our family, let $V = [v_1 \ \cdots \ v_n]$ in the fixed coordinates and give $r \times r$ matrices U the quotient norm $\alpha(U) = \min_{VX=U} \max_j \|\operatorname{row}_j X\|_1$, where X is $n \times r$. The determinant of VX is affine in each row of X: terms involving that row twice have two proportional columns and vanish. A maximum of its absolute value on $\max_j \|\operatorname{row}_j X\|_1 \leq 1$ therefore occurs with every row a signed coordinate vector. Those rows give precisely the signed partitions in (2.1). The direct proof above also covers complex phases.

2.4 Infinite families

To repeat the determinant argument, we first need the set sums to converge uniformly over all assignments and phases. Finite dimension supplies this through absolute summability.

Lemma 2.3 (absolute summability)**.** *Under the hypotheses of Theorem 1.1, only countably many v_j are nonzero and $\sum_j \|v_j\| < \infty$.*

Proof. Choose a basis $u_1, \dots, u_r$ of E and its coordinate functionals $g_1, \dots, g_r$. These are continuous in finite dimension. For each b and finite $H \subseteq J$, choose coefficients in (1.1) aligning the scalars $g_b(v_j)$. It follows that

$$\sum_{j \in H} |g_b(v_j)| \le M\|g_b\|, \qquad \sum_{j \in J} \|v_j\| \le M \sum_{b=1}^{r} \|u_b\|\|g_b\| < \infty. \tag{2.7}$$

The second inequality uses $v_j = \sum_b g_b(v_j) u_b$ and the triangle inequality, then takes finite subsums. For each positive integer N, only finitely many vectors have norm at least $1/N$. The union of these finite sets contains every nonzero label. □

Proof of Theorem 1.1. Lemma 2.2 proves the finite case. In general, $J_+ = \{j : v_j \neq 0\}$ is countable by Lemma 2.3. Define z_a, D by (2.1) for every assignment of J_+ to r indexed sets and every choice of unit phases. The sums converge absolutely. After enumerating J_+, their tails are bounded uniformly over all assignments and phases by the tail of $\sum_j \|v_j\|$. Thus the set sums, and hence the determinant, are continuous on the compact product of the finite assignment spaces and the phase circles (sign choices over $\mathbb{R}$).

The construction in Section 2.1 applies with w equal to the absolutely convergent sum of the remaining vectors. It again gives $D_0 \pm D_1$ with $D_0 \neq 0$. Consequently $|D|$ has a positive maximum. At a maximizer, the one-label phase and move arguments give (2.2), including $\sum_{j \in F_a} d_j = 1$; continuity of h_a justifies applying it to the convergent set sum.

The recovery identity (2.5) converges absolutely. For fixed j, every finite partial sum on its right has norm at most Md_j by (1.1). Passing to the limit yields $\|v_j\| \le Md_j$, and summing over each F_a proves (1.3). Zero labels can again be assigned arbitrarily with weight zero. The sets remain nonempty because $D \neq 0$.

For any other norm $\|\cdot\|'$ on E, norm equivalence gives $\|x\|' \le C\|x\|$ for all $x \in E$ and some finite constant C. Its capacity satisfies $M' \le CM < \infty$, and the same sums converge in that norm. The partition and weights therefore give (1.3) with each norm's own capacity. Equation (2.6) also extends by absolute convergence. □

3 Classical inequalities and their partition refinements

3.1 Dimension bounds, signed sums, and coordinate variations

Summing (1.3) gives $\sum_j \|v_j\| \le rM$. For a finite-dimensional normed space E, define

$$\pi_1(\mathrm{id}_E) = \sup_{\substack{\text{finite families } (v_j) \subset E \\ M((v_j)) > 0}} \frac{\sum_j \|v_j\|}{M((v_j))}. \tag{3.1}$$

This is the absolutely summing norm of the identity. Since the span of each family in (3.1) has dimension at most $\dim E$, the preceding bound yields $\pi_1(\mathrm{id}_E) \le \dim E$. This estimate is classical: Kadets–Snobar [7, (5)] give an equivalent formulation over the reals, and Jameson [6, p. 33, item (4)] gives the estimate over both fields. Theorem 1.1 adds a partition controlling the individual loads.

Corollary 3.1 (small norms and large signed sums)**.** *Let $v_1, \dots, v_n$ be vectors in a normed space over $\mathbb{F}$, whose span has dimension $r \geq 1$, and let M be their capacity. Then*

$$\min_{1 \leq j \leq n} \|v_j\| \leq \frac{M}{\lceil n/r \rceil}.$$

In particular, for n unit vectors, some signed sum of all n vectors has norm at least $\lceil n/r \rceil$ over $\mathbb{R}$; over $\mathbb{C}$, the same conclusion holds with unit phases in place of signs.

Proof. By Theorem 1.1, one packing set contains at least $\lceil n/r \rceil$ labels and has load at most M. Its smallest norm is therefore at most $M/\lceil n/r \rceil$. For unit vectors this gives $M \geq \lceil n/r \rceil$, and the capacity of a finite family is attained by signs or unit phases. □

The real bound is the sharp bound of Ambrus and Grundbacher [1, Theorem 2], applied to the unit ball in the real span of the vectors. Repeating each coordinate vector as evenly as possible in the supremum norm shows sharpness over both fields.

For n vectors in a real normed space, the small-norm bound in Corollary 3.1 also follows by scaling from [1, Theorem 7], applied in their span with the dual unit ball of the restricted norm.

For a nonzero finite-dimensional subspace $L \subseteq \ell_1(J)$, where $\ell_1(J)$ consists of scalar families with finite sum of absolute values, define its *coordinate variations* by

$$\kappa_j(L) = \sup_{\substack{x \in L \\ x \neq 0}} \frac{|x_j|}{\sum_\ell |x_\ell|}. \tag{3.2}$$

For $L \subseteq \mathbb{R}^n$ or $L \subseteq \mathbb{C}^n$, Hoffman [5, (1.1), (1.2), (4.1)] obtained the sharp bound $\sum_j \kappa_j(L) \leq \dim L$ from the Fan–Hoffman rank inequality. Theorem 1.1 refines this bound by partitioning the coordinate labels into $\dim L$ sets, each satisfying $\sum_{j \in F_a} \kappa_j(L) \leq 1$. This refinement holds over both fields and for arbitrary J. Indeed, regard the coordinate functionals $v_j(x) = x_j$ as vectors in the dual of $(L, \|\cdot\|_1)$. They span L^*, since their common annihilator in L is zero. Their norms are $\kappa_j(L)$, and their capacity is one by phase alignment and the identity $\|x\|_1 = \sum_j |x_j|$. Theorem 1.1 therefore partitions the coordinate labels into $\dim L$ nonempty sets, each satisfying

$$\sum_{j \in F_a} \kappa_j(L) \leq 1.$$

3.2 Absolute column maxima and rank from relative diagonals

For a nonzero $m \times n$ matrix A over $\mathbb{F}$, define its *absolute column maxima* $p_j(A)$ and its largest absolute row sum $R(A)$ by

$$p_j(A) = \max_i |a_{ij}|, \qquad R(A) = \max_i \sum_j |a_{ij}|. \tag{3.3}$$

Its columns Ae_j, in the supremum norm on $\mathbb{F}^m$, have norms $p_j(A)$, and their span has dimension $\operatorname{rank}_{\mathbb{F}} A$. Their capacity is $R(A)$: the triangle inequality gives the upper bound, and aligning the entries of a row attaining $R(A)$ gives equality.

Corollary 3.2 (absolute column maxima)**.** *If $r = \operatorname{rank} A > 0$, the column labels partition into r nonempty sets with*

$$\sum_{j \in F_a} p_j(A) \leq R(A) \qquad (1 \leq a \leq r). \tag{3.4}$$

Here r is the rank of A itself, even when the entrywise modulus matrix $|A| = (|a_{ij}|)$ has greater rank.

Proof. Apply Theorem 1.1 to the columns of A in the supremum norm. □

To state a rank consequence, define the *packing number* $P(x)$ of $x \in [0,1]^n$ as the least number of sets into which its positive coordinates can be partitioned, each set having sum at most one. Set $P(0) = 0$. Equivalently, $P(x)$ is the minimum number of bins of capacity one needed to pack items whose sizes are the positive coordinates of x, as in the classical bin packing problem. Zero coordinates can be adjoined to any existing set; for $k \geq 1$, $P(x) \leq k$ is equivalent to a partition of all n labels into at most k nonempty sets of load at most one. The convention at zero lets us include zero matrices in the following statement.

For an $n \times n$ matrix $A = (a_{ij})$, a *generalized diagonal* consists of the positions $(i, \sigma(i))$, where σ is a permutation of $[n] = \{1, \ldots, n\}$. Its product is $\prod_i a_{i,\sigma(i)}$. Write S_n for the set of all permutations of $[n]$.

For a square matrix, write $\rho_i = \sum_j |a_{ij}|$ and $\gamma_j = \sum_i |a_{ij}|$ for its absolute row and column sums. For a permutation σ, define the two *relative diagonals* by

$$t_i = \frac{|a_{i,\sigma(i)}|}{\rho_i}, \qquad t'_i = \frac{|a_{i,\sigma(i)}|}{\gamma_{\sigma(i)}}, \tag{3.5}$$

with either ratio defined to be zero when its denominator is zero.

Theorem 3.3 (packing rank bound)**.** *For every real or complex $n \times n$ matrix A and every permutation σ of $\{1, \ldots, n\}$, the relative diagonals t, t' defined in* (3.5) *satisfy*

$$\operatorname{rank} A \geq \max\{P(t), P(t')\}. \tag{3.6}$$

Proof. The zero matrix is immediate. Otherwise divide each nonzero row by its positive absolute sum, leaving zero rows unchanged. This is invertible diagonal row scaling, so rank is preserved, and the resulting matrix has $R = 1$. Its absolute column maxima dominate t, after relabeling column $\sigma(i)$ by i. Corollary 3.2 packs these maxima, hence also the relative diagonal, into $\operatorname{rank} A$ sets. Apply the same reasoning to the transpose, with permutation σ^{-1}, for t'. □

3.3 Fan–Hoffman, Lévy–Desplanques, and Ostrowski

Since a partition into r sets of load at most one has total load at most r, (3.6) contains the Fan–Hoffman inequality [4, Theorem 1.3]

$$\operatorname{rank} A \geq \sum_{i:\rho_i > 0} \frac{|a_{ii}|}{\rho_i}. \tag{3.7}$$

The bound $\sum_j p_j(A) \leq rR(A)$ on the sum of absolute column maxima follows already from this classical inequality. Choose a row attaining each absolute column maximum, and form an $n \times n$ matrix B whose jth row is the row chosen for column j, allowing repetitions. Then $|b_{jj}| = p_j(A)$, $\operatorname{rank} B \leq r$, and its absolute row sums $s_j \leq R(A)$. Fan–Hoffman gives $\sum_{s_j > 0} p_j(A)/s_j \leq r$; a zero s_j has $p_j(A) = 0$. This proves the bound on the sum of absolute column maxima. Its stochastic form also appears, after transposition, in Riazanov and Vyalyi [17, Theorem 4.3].

For nonsingularity, assume $n \geq 2$ and write $R_i = \sum_{j \neq i} |a_{ij}|$ for the off-diagonal row sum. If $|a_{ii}| > R_i$ for every i, then every relative diagonal entry is greater than $1/2$. No two can share a set of load at most one, so (3.6) gives rank n. This is the Lévy–Desplanques strict diagonal-dominance criterion.

Ostrowski's pairwise criterion [14, Zusatz zu Satz VI, p. 77], [15, Satz II with $\alpha = 1$] is recovered in the same way. When $\rho_i, \rho_j > 0$, direct cross-multiplication gives

$$t_i + t_j \le 1 \quad \Longleftrightarrow \quad |a_{ii}|\,|a_{jj}| \le R_i R_j. \tag{3.8}$$

Thus $|a_{ii}|\,|a_{jj}| > R_i R_j$ for every $i \neq j$ rules out every shared set and forces nonsingularity. The strict hypothesis itself rules out zero rows. For a row-stochastic matrix, $R_i = 1 - a_{ii}$, so Ostrowski's condition is precisely $a_{ii} + a_{jj} > 1$ for all pairs. Every singular row-stochastic matrix therefore has two diagonal entries whose sum is at most one.

Ostrowski's criterion implies that among any $r+1$ columns of a rank-r row-stochastic matrix, two column maxima have sum at most one. Choose a row attaining the maximum for each. If two chosen rows coincide, the two maxima sum to at most one. Otherwise their square submatrix is singular. Its off-diagonal row sums are at most $1 - p_j$, so the contrapositive of Ostrowski's criterion gives $p_i p_j \le (1-p_i)(1-p_j)$, hence $p_i + p_j \le 1$, for some pair. Corollary 3.2 retains more information: it gives one partition of all column labels, with a load bound for every set.

Write $c^{[m]}$ for m repetitions of c. For an example of the additional information in the partition, take five relative diagonal entries all equal to $2/5$. Their sum is two, and every pair has sum at most one. A set can nevertheless hold at most two of them, so $P((2/5)^{[5]}) = 3$ and (3.6) forces rank at least three. This bound is attained by the row-stochastic matrix

$$\begin{pmatrix} 2/5 & 2/5 & 1/5 & 0 & 0 \\ 2/5 & 2/5 & 1/5 & 0 & 0 \\ 1/5 & 0 & 2/5 & 2/5 & 0 \\ 1/5 & 0 & 2/5 & 2/5 & 0 \\ 3/5 & 0 & 0 & 0 & 2/5 \end{pmatrix}. \tag{3.9}$$

It has only three distinct rows, so its rank is at most three; the relative-diagonal packing bound gives the reverse inequality.

3.4 Eigenvalue location and geometric multiplicity

For a complex $n \times n$ matrix $A = (a_{ij})$ and $z \in \mathbb{C}$, set

$$t_i(z) = \frac{|a_{ii} - z|}{|a_{ii} - z| + R_i}, \tag{3.10}$$

with value zero when the denominator vanishes. The *geometric multiplicity* of an eigenvalue λ is $\mu = \dim\ker(A - \lambda I)$, where I is the identity matrix.

Corollary 3.4 (eigenvalue localization)**.** *Let $A = (a_{ij}) \in \mathbb{C}^{n\times n}$, and let λ be an eigenvalue of A with geometric multiplicity $\mu = \dim\ker(A - \lambda I)$. With $t(\lambda)$ defined by* (3.10)*, we have*

$$P(t(\lambda)) \le n - \mu. \tag{3.11}$$

The same conclusion holds with off-diagonal column sums in place of R_i. If $\mu < n$, put $h = \lceil n/(n-\mu) \rceil \ge 2$. There is an h-element set $S \subseteq [n]$ such that

$$\sum_{i\in S} t_i(\lambda) \le 1, \qquad |\lambda - a_{ii}| \le \frac{R_i}{h-1} \quad \text{for some } i \in S. \tag{3.12}$$

Proof. Apply Theorem 3.3 to $A - \lambda I$, whose rank is $n - \mu$. If this rank is positive, (3.11) permits a partition of all n labels into at most $n - \mu$ nonempty sets, including when all relative entries are zero. Some set contains at least h labels. Any h of them give the first inequality in (3.12), and one has $t_i(\lambda) \le 1/h$. Cross-multiplication gives the second inequality; it also holds if the denominator in (3.10) is zero. If $\mu = n$, then $A = \lambda I$ and (3.11) reads $P(0) = 0$; no value of h is needed. □

In particular, at most $n - \mu$ of the numbers $t_i(\lambda)$ exceed $1/2$, since no two such numbers can lie in one packing set. Thus λ lies in at least μ Gershgorin discs. This recovers the theorem of Taussky and Stein [4, p. 119].

More generally, for every integer $m \ge 2$, each packing set contains at most $m - 1$ entries exceeding $1/m$. Consequently,

$$\bigl|\{i : |\lambda - a_{ii}| \le R_i/(m-1)\}\bigr| \ge n - (m-1)(n-\mu).$$

For $\mu < n$, taking $m = h$ gives at least $n - (h-1)(n-\mu) \ge 1$ indices satisfying the disc bound in (3.12).

For $n \ge 2$, this recovers Brauer's ovals of Cassini [2, p. 21], [15, Satz V with $\alpha = 1$]: every eigenvalue satisfies $|\lambda - a_{ii}|\,|\lambda - a_{jj}| \le R_i R_j$ for some $i \ne j$. If $A - \lambda I \ne 0$, two labels share a packing set, and (3.8) applies. A zero denominator gives the same inequality directly; if $A = \lambda I$, it is immediate.

Summing all loads in (3.11) gives $\sum_i t_i(\lambda) \le n - \mu$, the Fan–Hoffman rank bound [4, Theorem 1.3] applied to $A - \lambda I$.

For $1 \le \mu < n$, Fan–Hoffman [4, Theorem 1.7] gives $|\lambda - a_{ii}| \le \beta_i R_i$ for some i, whenever the positive parameters $\beta_1, \dots, \beta_n$ satisfy $\sum_{i=1}^{n} (1 + \beta_i)^{-1} \le \mu$. Taking $\beta_i = (n-\mu)/\mu$ for every i makes this sum equal to μ and gives the common radius factor $(n-\mu)/\mu$. The disc bound in (3.12) improves this equal-parameter specialization, since

$$\frac{1}{h-1} \le \frac{n-\mu}{\mu}, \qquad h = \left\lceil \frac{n}{n-\mu} \right\rceil.$$

Equality holds exactly when $n - \mu$ divides n. When $\mu \le n/2$, we have $h = 2$, and the disc bound reduces to the usual Gershgorin inclusion.

Geometric multiplicity cannot be replaced by algebraic multiplicity, the multiplicity as a root of the characteristic polynomial. For

$$A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix},$$

zero has algebraic multiplicity two but geometric multiplicity one, and $t(0) = (1/2, 1/2)$ has packing number one. Replacing geometric multiplicity by algebraic multiplicity in (3.11) would give $P(t(0)) \le 2 - 2 = 0$, whereas $P(t(0)) = 1$.

4 Attainable profiles and sharp majorization bounds

The partition is the exact constraint on finite normalized norm profiles when the normed space is allowed to vary. A simpler majorization bound then gives sharp answers for a broad class of symmetric functions. We separate these two statements because the majorization inequalities alone do not characterize which profiles are possible.

4.1 Exact norm and column-maximum profiles

Theorem 4.1 (attainable profiles)**.** *Let n, k be integers with $1 \le k \le n$, and let $x \in [0,1]^n$. The following are equivalent:*

1. *$x_j = \|v_j\|/M$ for a family $v_1, \dots, v_n$ of capacity $M > 0$ whose span has dimension at most k in some normed space over $\mathbb{F}$.*
2. *x is the vector of column maxima of an $n \times n$ row-stochastic matrix of rank at most k.*
3. *$\sum_j x_j \ge 1$ and $P(x) \le k$.*

The stochastic realization can be chosen of rank exactly $P(x)$. A vector realization exists in $\mathbb{F}^k$ with the supremum norm.

Proof. For 1 implies 3, Theorem 1.1 gives $P(x) \le k$, while the triangle inequality gives $M \le \sum_j \|v_j\|$. For 2 implies 1, use the matrix columns in the supremum norm; their capacity is one.

For 3 implies 2, take a packing into $r = P(x)$ nonempty sets $F_1, \dots, F_r$, assigning zero labels to these sets as necessary. Here $r \ge 1$ because $\sum_j x_j \ge 1$. For each a, construct a probability row $w^{(a)}$ by setting $w_j^{(a)} = x_j$ on F_a and distributing its remaining mass outside F_a, with $0 \le w_j^{(a)} \le x_j$. This is possible because

$$\sum_{j \notin F_a} x_j \ge 1 - \sum_{j \in F_a} x_j. \tag{4.1}$$

For completeness, the deficit can be assigned to the outside coordinates one at a time up to their upper bounds x_j; (4.1) ensures that it is exhausted. Set row i of A equal to $w^{(a)}$ whenever $i \in F_a$. There are at most r distinct rows, so $\operatorname{rank} A \le r$. Every entry in column j is at most x_j, and the rows of its own set attain x_j. Thus the column maxima are exactly x. Corollary 3.2 gives the reverse rank bound $\operatorname{rank} A \ge P(x) = r$. Retaining one representative row per set realizes the same column norms and capacity in $\mathbb{F}^r$; append zero coordinates to embed in $\mathbb{F}^k$. □

4.2 The balanced reciprocal profile

Write $n = qk + s$, where $q \ge 1$ and $0 \le s < k$. The *balanced reciprocal profile* is

$$b_{n,k} = \big((1/q)^{[q(k-s)]}, (1/(q+1))^{[(q+1)s]}\big). \tag{4.2}$$

It has n coordinates and total k. To obtain it, choose k set sizes differing by at most one, and give each coordinate in a set of size t the value $1/t$.

We use the following conventions for majorization. Write $x^{\downarrow}$ and $x^{\uparrow}$ for the coordinates of x in decreasing and increasing order. For vectors of the same length, $u \prec y$ means $\sum_{j=1}^{\ell} u_j^{\downarrow} \le \sum_{j=1}^{\ell} y_j^{\downarrow}$ for every ℓ, with equality for the full sum. A function F is *Schur-concave* if $u \prec y$ implies $F(u) \ge F(y)$, and *Schur-convex* if the inequality is reversed. These conditions in particular make the function invariant under coordinate permutations. Coordinatewise monotonicity is a separate condition.

Theorem 4.2 (sharp majorization bounds)**.** *Let n, k be integers with $1 \le k \le n$, and write $n = qk + s$, where $0 \le s < k$. If $x \in [0,1]^n$ and $P(x) \le k$, then, for every integer $1 \le \ell \le n$,*

$$\sum_{j=1}^{\ell} x_j^{\uparrow} \le \sum_{j=1}^{\ell} (b_{n,k})_j^{\uparrow} = \begin{cases} \ell/(q+1), & \ell \le (q+1)s, \\ s + (\ell - (q+1)s)/q, & \ell \ge (q+1)s. \end{cases} \tag{4.3}$$

Consequently, for every coordinatewise nondecreasing Schur-concave function F on $[0,\infty)^n$,

$$F(x) \le F(b_{n,k}). \tag{4.4}$$

For every coordinatewise nonincreasing Schur-convex function G on the same domain, $G(x) \ge G(b_{n,k})$. If every coordinate of x is positive, the conclusions for F and G also hold for functions defined only on $(0,\infty)^n$, under the same respective monotonicity and Schur-concavity or Schur-convexity assumptions. All these bounds are attained at $x = b_{n,k}$. This profile is realizable both as a normalized norm profile and as the vector of column maxima of a doubly stochastic matrix of rank k.

Proof. There are three steps: increase each set's load to one, average its coordinates, and balance the set sizes. Start with a packing of all labels into at most k nonempty sets and split sets until there are exactly k; this is possible because $k \le n$. The same construction applies when $x = 0$. Let these sets be $F_1, \dots, F_k$, put $n_a = |F_a|$, and write $\mathbf{n} = (n_1, \dots, n_k)$. Increase coordinates within each set until its sum is one, giving $y \ge x$. Averaging within each set gives

$$u(\mathbf{n}) = ((1/n_1)^{[n_1]}, \dots, (1/n_k)^{[n_k]}) \prec y. \tag{4.5}$$

This majorization follows, for example, by averaging all permutations within each set: for every ℓ, the sum of the ℓ largest coordinates is a convex function and cannot increase under this averaging.

Next we prove $b_{n,k} \prec u(\mathbf{n})$. Suppose two set sizes are p, t, with $t \ge p+2$, and replace them by $p+1, t-1$. In the affected $p+t$ coordinates, the old decreasing partial sum exceeds the new one by $\ell/p - \ell/(p+1) \ge 0$ for $\ell \le p$. For $p+1 \le \ell \le p+t$, the difference is

$$\left(1 + \frac{\ell - p}{t}\right) - \left(1 + \frac{\ell - p - 1}{t-1}\right) = \frac{p + t - \ell}{t(t-1)} \ge 0. \tag{4.6}$$

The total sums agree. Adjoining unchanged coordinates preserves majorization: the largest-ℓ sum of a concatenation is the maximum over all ways of taking a prefix from each part. Repeating this size balancing terminates, since it decreases $\sum_a n_a^2$, and yields $b_{n,k} \prec u(\mathbf{n})$.

Monotonicity and Schur-concavity now give

$$F(x) \le F(y) \le F(u(\mathbf{n})) \le F(b_{n,k}). \tag{4.7}$$

This proves (4.4). Taking F to be the sum of the ℓ smallest coordinates gives (4.3), with the displayed value computed from (4.2). Conversely, (4.3) implies (4.4) by the standard weak-majorization characterization [13, 3.A.8]. Taking $F = -G$ proves the dual statement. When x is strictly positive, every vector in the fill, average, and balance argument is strictly positive, proving this extension. Finally, take n_a copies of e_a/n_a for each balanced size in the supremum norm. Their capacity is one and their profile is $b_{n,k}$. The stochastic realization is the matrix $B_{n,k}$ defined in (4.8). □

Define J_t to be the $t \times t$ matrix all of whose entries are $1/t$, and write $\oplus$ for block diagonal direct sum. The matrix

$$B_{n,k} = J_q^{\oplus(k-s)} \oplus J_{q+1}^{\oplus s} \tag{4.8}$$

is doubly stochastic and has rank k. Its column maxima and its entries on the main diagonal both form the vector $b_{n,k}$. This matrix attains equality in many of the inequalities below.

The converse of Theorem 4.2 fails: the inequalities (4.3) do not imply $P(x) \le k$. For $n = 6, k = 3$, the vector $x = ((3/5)^{[5]}, 0)$ satisfies all of (4.3) against $b_{6,3} = ((1/2)^{[6]})$: its smallest-ℓ sum is

$3(\ell-1)/5 \le \ell/2$ for $1 \le \ell \le 6$. But its five positive coordinates require five sets. Thus Theorem 4.1 contains information that (4.3) alone does not retain.

Coordinatewise monotonicity in Theorem 4.2 is essential when totals may differ. The Schur-concave function $-\sum_j x_j \log x_j$, with $0 \log 0 = 0$, has value $\log 3$ at the rank-one profile $(1/3, 1/3, 1/3)$, exceeding its value $\log 2$ at $b_{3,2}$.

4.3 Products and symmetric functions of vector norms

For $x \in [0,\infty)^n$, the mth *elementary symmetric function* is $e_m(x) = \sum_{|S|=m} \prod_{j\in S} x_j$, for $1 \le m \le n$; set $e_0(x) = 1$. Let $[z^m]f(z)$ denote the coefficient of z^m in a polynomial. Define

$$\begin{aligned} D_{n,k} &= q^{-q(k-s)}(q+1)^{-(q+1)s}, \\ E_m(n,k) &= [z^m](1+z/q)^{q(k-s)}(1+z/(q+1))^{(q+1)s}, \\ N_{n,k} &= (k-s)q^2 + s(q+1)^2. \end{aligned} \tag{4.9}$$

For $b = b_{n,k}$, these constants are the product of its coordinates, its mth elementary symmetric function, and the sum of its reciprocals:

$$D_{n,k} = \prod_{j=1}^{n} b_j, \qquad E_m(n,k) = e_m(b), \qquad N_{n,k} = \sum_{j=1}^{n} \frac{1}{b_j}.$$

Corollary 4.3 (products and symmetric functions)**.** *Let n, k be integers with $1 \le k \le n$, and let $v_1, \dots, v_n$ be vectors in a real or complex normed space. Suppose $1 \le \dim \operatorname{span}\{v_1, \dots, v_n\} \le k$, and let M be their capacity. With $x_j = \|v_j\|/M$, we have*

$$\begin{aligned} \prod_j \|v_j\| &\le M^n D_{n,k}, \\ e_m(\|v_1\|, \dots, \|v_n\|) &\le M^m E_m(n,k) \quad (1 \le m \le n), \\ \frac{n}{\sum_j 1/x_j} &\le \frac{n}{N_{n,k}}. \end{aligned} \tag{4.10}$$

In the last line the left side is defined to be zero if any $x_j = 0$. The constants are attained by the repeated-coordinate construction in the proof of Theorem 4.2, and by the column maxima of $B_{n,k}$.

Proof. The elementary symmetric functions are increasing and Schur-concave on the nonnegative orthant. Hold all but two coordinates fixed. Then $e_m(x) = c_0 + c_1(x_i + x_j) + c_2 x_i x_j$, where c_0, c_1, c_2 depend only on the remaining coordinates and $c_2 \ge 0$. Averaging the pair at fixed sum increases $x_i x_j$, and hence e_m. The same is true of partial averaging, which makes the pair closer while preserving its sum. Majorization is characterized by finite sequences of such averaging steps and coordinate permutations [13, 4.A.1]. Apply Theorem 4.2 and the identity $\prod_j (1 + z x_j) = \sum_m e_m(x) z^m$. The product is the case $m = n$. The function $G_\varepsilon(x) = \sum_j 1/(x_j + \varepsilon)$ is decreasing and Schur-convex for every $\varepsilon > 0$. Apply the dual bound and let $\varepsilon \downarrow 0$ to get $\sum_j 1/x_j \ge N_{n,k}$ for positive x. If a coordinate is zero, the asserted harmonic-mean inequality holds by its stated convention. □

The product constant also has a direct explanation. Starting with the partition from Theorem 1.1, split sets until there are exactly k nonempty sets. This is possible because $k \le n$, and splitting preserves the load bound M. Within a set of size n_a, the arithmetic–geometric mean inequality gives a product at most $(M/n_a)^{n_a}$. Thus the total product is bounded by $M^n \prod_a n_a^{-n_a}$.

Strict convexity of $t \log t$ shows that $\sum_a n_a \log n_a$ is minimized at balanced positive integer sizes: transferring one unit from c to a decreases it whenever $c \geq a+2$. This recovers $D_{n,k}$.

For $n = 3, k = 2$, we have $b_{3,2} = (1, 1/2, 1/2)$, so

$$\prod_j x_j \leq \frac{1}{4} < \left(\frac{2}{3}\right)^3, \qquad e_2(x) \leq \frac{5}{4} < \frac{4}{3}, \qquad \frac{3}{\sum_j 1/x_j} \leq \frac{3}{5}. \tag{4.11}$$

The sum bound $\sum_j x_j \leq 2$ alone gives only $8/27$ for the product and $4/3$ for e_2. More generally, when $k \nmid n$, the nonconstant balanced reciprocal profile strictly improves the Maclaurin bounds $\binom{n}{m}(k/n)^m$ based only on $\sum_j x_j \leq k$, for $2 \leq m \leq n$.

4.4 Weighted products and entropy of partitions

Equal weights lead to balanced cardinalities. For unequal weights, the sharp constant is instead a finite optimization over the masses of the partition sets. For a probability vector $w = (w_1, \ldots, w_n)$, define its *Shannon entropy* by $H(w) = -\sum_j w_j \log w_j$, with $0 \log 0 = 0$. For strictly positive w and $1 \leq k \leq n$, put

$$H_k^*(w) = \max_{\mathcal{F}} H(W_1, \ldots, W_k), \qquad W_a = \sum_{j \in F_a} w_j, \tag{4.12}$$

where $\mathcal{F} = (F_1, \ldots, F_k)$ runs over partitions of $[n]$ into k nonempty sets. Equivalently, if X has distribution w, this is the maximum of $H(f(X))$ over deterministic functions f taking exactly k values on the support of X. This optimization was studied in [3].

Theorem 4.4 (weighted products)**.** *Let n, k be integers with $1 \leq k \leq n$, and let $v_1, \ldots, v_n$ be vectors in a real or complex normed space with $1 \leq \dim \operatorname{span}\{v_1, \ldots, v_n\} \leq k$. Let M be their coefficient-sum capacity, and let $w = (w_1, \ldots, w_n)$ satisfy $w_j > 0$ and $\sum_j w_j = 1$. Then*

$$\prod_j \|v_j\|^{w_j} \leq M \exp\big(H_k^*(w) - H(w)\big). \tag{4.13}$$

The constant is attained among vector families of capacity M whose span has dimension at most k. For every fixed permutation σ of $[n]$, the exact maximum of $\prod_{i=1}^n a_{i,\sigma(i)}^{w_i}$ over row-stochastic $n \times n$ matrices A with $\operatorname{rank} A \leq k$ is $\exp\big(H_k^*(w) - H(w)\big)$.

Proof. Put $x_j = \|v_j\|/M$. Apply Theorem 1.1 and split sets as necessary to obtain k nonempty sets, each with load at most one. If some $x_j = 0$, the product is zero. Otherwise, in a set F_a of mass W_a, the weighted arithmetic–geometric mean inequality gives

$$\sum_{j \in F_a} w_j \log \frac{x_j}{w_j/W_a} \leq W_a \log\left(\sum_{j \in F_a} x_j\right) \leq 0. \tag{4.14}$$

Summing yields $\sum_j w_j \log x_j \leq H(W) - H(w)$, which proves (4.13), since $H(W) \leq H_k^*(w)$.

For an entropy-maximizing partition, take $v_j = M(w_j/W_a)e_a$ when $j \in F_a$ in the supremum norm. Each coordinate sum is M, so the capacity is exactly M, and equality holds. For a row-stochastic matrix A of rank at most k, apply (4.13) to the reordered columns $v_i = Ae_{\sigma(i)}$ in the supremum norm. Their capacity is one, their span has dimension at most k, and $a_{i,\sigma(i)} \leq \|v_i\|_\infty$. Hence $\prod_{i=1}^n a_{i,\sigma(i)}^{w_i} \leq \prod_{i=1}^n \|v_i\|_\infty^{w_i} \leq \exp\big(H_k^*(w) - H(w)\big)$.

For the matrix realization, give every row indexed by F_a the probability row with entries w_j/W_a on F_a and zero elsewhere. The k distinct rows have disjoint supports and are independent. The resulting matrix has rank k, and its diagonal entries attain equality in (4.13). A column permutation gives any prescribed generalized diagonal. □

For $w_j = 1/n$, maximizing $H(W)$ is the same as minimizing $\sum_a n_a \log n_a$, so (4.13) becomes the nth root of the product bound. With arbitrary weights, equal cardinalities need not be optimal. For $w = (3/5, 1/10, 1/10, 1/10, 1/10)$ and $k = 2$, one set must have mass at least $3/5$. Entropy is maximized by putting the largest weight alone, giving masses $3/5, 2/5$ and set sizes one and four.

5 Sharp matrix inequalities

Throughout this section A is a real or complex $n \times n$ matrix, $1 \le k \le n$, and $\operatorname{rank}_{\mathbb{F}} A \le k$. The absolute row and column sums are ρ_i and γ_j, as defined in Section 3.2. The constants use $n = qk+s$, $0 \le s < k$.

5.1 Products along generalized diagonals

The sharp stochastic diagonal-product bound was conjectured in [8, Conjecture 4] and proved, with its equality cases, in [9]. The product bound in (4.10) gives the following inequality for real and complex matrices.

Corollary 5.1 (sharp bound for generalized diagonal products)**.** *Let n, k be integers with $1 \le k \le n$, and let A be a real or complex $n \times n$ matrix with* $\operatorname{rank} A \le k$*. Then, for every $\sigma \in S_n$,*

$$\prod_{i=1}^{n} |a_{i,\sigma(i)}| \le D_{n,k} \min\left\{ \prod_i \rho_i, \prod_j \gamma_j \right\}. \tag{5.1}$$

Proof. A zero row or column makes the left side zero. Otherwise divide each row by its positive absolute sum, obtaining a matrix $\widehat{A}$ of the same rank whose column family has capacity one in the supremum norm. Its absolute column maxima $p_j = p_j(\widehat{A})$ satisfy

$$\prod_i |\widehat{a}_{i,\sigma(i)}| \le \prod_j p_j \le D_{n,k} \tag{5.2}$$

by Corollary 4.3. Restore the row factors. Apply the same argument to A^{T} to obtain the column factors. □

The balanced stochastic matrix $B_{n,k}$ in (4.8) attains (5.1) on the main diagonal, and both products of row and column sums are one. Thus the constant is sharp even among doubly stochastic matrices.

For nonnegative A, inequality (5.1) and its full equality classification are proved in [9, Theorem 1.1(2)]. For complex A, the case $k \mid n$ already follows from Fan–Hoffman [4, Theorem 1.3] and the arithmetic–geometric mean inequality, with $D_{n,k} = (k/n)^n$; the case $k = n-1$ follows from Ostrowski's criterion [14, Zusatz zu Satz VI, p. 77], [15, Satz II with $\alpha = 1$], with $D_{n,n-1} = 1/4$. Corollary 5.1 gives the sharp bound in the full real and complex setting, without classifying its equality cases.

5.2 The permanent of the entrywise modulus

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

$$\operatorname{per} A = \sum_{\sigma \in S_n} \prod_i a_{i,\sigma(i)}.$$

The sharp stochastic permanent bound was conjectured in [8, Conjecture 1] and proved in [10]. We obtain it here from the packing theorem, together with its real and complex extension. Put

$$C_{n,k} = \left(\frac{q!}{q^q}\right)^{k-s} \left(\frac{(q+1)!}{(q+1)^{q+1}}\right)^{s}. \tag{5.3}$$

Theorem 5.2 (sharp permanent bound)**.** *Let n, k be integers with $1 \le k \le n$, and let A be a real or complex $n \times n$ matrix with* $\operatorname{rank} A \le k$*. Then*

$$|\operatorname{per} A| \le \operatorname{per}|A| \le C_{n,k} \min\left\{\prod_i \rho_i, \prod_j \gamma_j\right\}. \tag{5.4}$$

The rank hypothesis is on A; $|A|$ may have greater rank. The balanced stochastic matrix $B_{n,k}$ attains the constant.

The proof uses induction on the size. Packing supplies a column whose entries are all small. Expansion along that column reduces the problem to smaller matrices, while a scalar optimization bounds the expansion coefficients. The resulting factor is exactly the ratio of two successive balanced-block constants.

Proof. The first inequality is the triangle inequality. A zero row or column makes $\operatorname{per}|A| = 0$, so assume all absolute row sums are positive and divide each row by its sum. Rank is preserved. It remains to prove $\operatorname{per}|A| \le C_{n,k}$ when every absolute row sum equals one. Fix k and induct on $n \ge k$. For $n = k$, the permanent is a nonnegative subsum of the expansion of $\prod_i \sum_j |a_{ij}| = 1$, and $C_{k,k} = 1$.

Suppose $n > k$ and set $h = \lceil n/k \rceil \ge 2$. Corollary 3.1, applied to the columns in the supremum norm, gives a column j_* whose absolute column maximum is at most $1/\lceil n/r \rceil \le 1/h$, where $r = \operatorname{rank} A \le k$. Put $\beta_i = |a_{i,j_*}|$; every β_i is at most $1/h$.

Let A_i be the submatrix obtained by deleting row i and column j_* from A *itself, not* $|A|$. The surviving row $\ell \ne i$ has absolute sum $1 - \beta_\ell > 0$. Dividing that row by $1 - \beta_\ell$ gives a matrix $\widetilde{A}_i$ with absolute row sums one. Deletion cannot increase rank and invertible row scaling preserves rank, so $\operatorname{rank} \widetilde{A}_i \le k$. Since $n - 1 \ge k$, induction gives

$$\operatorname{per}|A_i| \le C_{n-1,k} \prod_{\ell \ne i} (1 - \beta_\ell). \tag{5.5}$$

No rank assumption has been applied to $|A_i|$. Expanding the nonnegative permanent $\operatorname{per}|A|$ along column j_* and using (5.5) yields

$$\operatorname{per}|A| \le C_{n-1,k} \Psi(\beta), \qquad \Psi(\beta) = \sum_i \beta_i \prod_{\ell \ne i} (1 - \beta_\ell). \tag{5.6}$$

The function Ψ is affine in each coordinate separately. Replacing coordinates one at a time by endpoints without decreasing Ψ shows that its maximum on $[0, 1/h]^n$ occurs at a vertex. A vertex with $t \ge 1$ nonzero coordinates has value $g_t = (t/h)(1 - 1/h)^{t-1}$; the all-zero vertex has value zero. For positive integers t,

$$\frac{g_{t+1}}{g_t} = \frac{t+1}{t} \frac{h-1}{h}. \tag{5.7}$$

This ratio exceeds one for $t < h - 1$, equals one for $t = h - 1$, and is less than one for $t > h - 1$. Since $2 \le h \le n$, both maximizing values $t = h - 1$ and $t = h$ are available. Hence

$$\Psi(\beta) \le g_{h-1} = g_h = (1 - 1/h)^{h-1}. \tag{5.8}$$

Let $f(t) = t!/t^t$. The balanced composition of n into k positive integer parts has largest part h. Reducing one largest part by one gives the balanced composition of $n-1$. Therefore

$$\frac{C_{n,k}}{C_{n-1,k}} = \frac{f(h)}{f(h-1)} = \left(1 - \frac{1}{h}\right)^{h-1}. \tag{5.9}$$

Combining (5.6), (5.8), and (5.9) gives

$$\operatorname{per}|A| \le C_{n-1,k}(1-1/h)^{h-1} = C_{n,k}, \tag{5.10}$$

which closes the induction. Undo the row scaling, then apply the same result to A^{T}, whose permanent and rank are unchanged, to obtain both products in (5.4). Each J_t has rank one and permanent $t!/t^t$; their balanced direct sum has permanent $C_{n,k}$ and both products of row and column sums equal to one. □

This recovers the sharp stochastic permanent bound proved in [10], where its equality cases are classified. Theorem 5.2 establishes the sharp bound for arbitrary real or complex matrices and does not classify their equality cases. Marcus and Minc [12, Corollary 3.2] proved $|\operatorname{per} A|^2 \le \frac{1}{n}\sum_{i=1}^{n} \alpha_i^{2n}$, where $\alpha_1, \dots, \alpha_n$ are the singular values of A. For doubly stochastic A, they obtained $\operatorname{per} A \le \sqrt{\operatorname{rank} A/n}$ [12, Corollary 3.3]. For rank at most k, the sharp bound in (5.4) is $C_{n,k}$. The improvement can be substantial: when $n = 2k$ and $\operatorname{rank} A = k$, (5.4) gives $\operatorname{per} A \le 2^{-k}$, compared with $1/\sqrt{2}$ from the Marcus–Minc rank bound. More generally, (5.4) bounds $\operatorname{per}|A|$ in terms of the absolute row and column sums, with the rank condition imposed on A.

5.3 Powers and sums of diagonal products

The preceding two bounds jointly control the entire vector of absolute diagonal products. Write $D = D_{n,k}$, $C = C_{n,k}$, and

$$\begin{aligned} Q &= \min\left\{\prod_i \rho_i, \prod_j \gamma_j\right\}, \qquad z_\sigma = \prod_i |a_{i,\sigma(i)}|, \\ m_0 &= C/D = (q!)^{k-s}((q+1)!)^s. \end{aligned} \tag{5.11}$$

Thus m_0 is an integer. Set $z(A) = (z_\sigma)_{\sigma \in S_n} \in [0,\infty)^{n!}$. The vector $z(B_{n,k})$ has exactly m_0 nonzero coordinates, all equal to D: a permutation contributes if and only if it preserves every block. For $t \ge 1$, let $|A|^{\circ t}$ denote the matrix obtained by taking the tth power of each entry of $|A|$.

Corollary 5.3 (majorization of absolute diagonal products)**.** *Let n, k be integers with $1 \le k \le n$, and let A be a real or complex $n \times n$ matrix with $\operatorname{rank} A \le k$. Then, for every coordinatewise nondecreasing Schur-convex function Φ on $[0,\infty)^{n!}$,*

$$\Phi(z(A)) \le \Phi\big(Qz(B_{n,k})\big). \tag{5.12}$$

In particular, for $t \ge 1$ and $1 \le m \le n!$,

$$\begin{aligned} &\operatorname{per}(|A|^{\circ t}) \le CD^{t-1}Q^t, \\ &\sum_{j=1}^{m} z_j^{\downarrow}(A) \le Q\min\{mD, C\}. \end{aligned} \tag{5.13}$$

All bounds are attained by $A = B_{n,k}$, for which $Q = 1$.

Proof. If $Q = 0$, all diagonal products vanish. Otherwise normalize $y = z(A)/Q$. Corollary 5.1 and Theorem 5.2 give $0 \le y_\sigma \le D$ and $\sum_\sigma y_\sigma \le C = m_0 D$. Increase coordinates, without exceeding D, until the total is C. This is possible because $m_0 \le n!$. The resulting vector is majorized by $(D^{[m_0]}, 0^{[n!-m_0]})$: each decreasing prefix is bounded both by its length times D and by the total C. Monotonicity and Schur-convexity prove (5.12). Take $\Phi(z) = \sum z_\sigma^t$ or the sum of the m largest coordinates to get (5.13). □

No rank restriction on $|A|^{\circ t}$ is used. The range $t \ge 1$ is essential to this conclusion: at $n = 3, k = 2, t = 1/2$, $\operatorname{per}(J_3^{\circ 1/2}) = 2/\sqrt{3} > 1$, whereas $\operatorname{per}((J_1 \oplus J_2)^{\circ 1/2}) = 1$.

5.4 Prescribed diagonals and minimum entrywise mass

We next minimize the *absolute entrywise mass* $\sum_{i,j} |a_{ij}|$ when the diagonal magnitudes are prescribed. The sharp answer is another partition optimization; in this case each set contributes the square of a sum of square roots.

For prescribed diagonal magnitudes $\delta_i > 0$ and rank at most k, the Fan–Hoffman inequality (3.7), together with Cauchy–Schwarz, gives

$$\sum_{i,j} |a_{ij}| \ge \frac{1}{k} \left(\sum_i \sqrt{\delta_i} \right)^2 .$$

For unit diagonal magnitudes, this is the bound n^2/k. The following theorem determines the exact minimum by retaining the partition constraint.

Theorem 5.4 (prescribed-diagonal mass)**.** *For $\delta_1, \dots, \delta_n > 0$ and $1 \le k \le n$, over either $\mathbb{R}$ or $\mathbb{C}$,*

$$\min_{\substack{\operatorname{rank} A \le k \\ |a_{ii}| = \delta_i}} \sum_{i,j} |a_{ij}| = \min_{\mathcal{F}} \sum_{F \in \mathcal{F}} \left(\sum_{i \in F} \sqrt{\delta_i} \right)^2 , \tag{5.14}$$

where the right minimum is over partitions of $[n]$ into at most k nonempty sets. It can equivalently be taken over exactly k sets.

Proof. The positive diagonal guarantees $\rho_i > 0$. By Theorem 3.3, the relative diagonal $t_i = \delta_i/\rho_i$ has a packing $F_1, \dots, F_r$ with $r \le k$. Cauchy–Schwarz within each set gives

$$\left(\sum_{i \in F_a} \sqrt{\delta_i} \right)^2 = \left(\sum_{i \in F_a} \sqrt{\rho_i} \sqrt{t_i} \right)^2 \le \left(\sum_{i \in F_a} \rho_i \right) \left(\sum_{i \in F_a} t_i \right) \le \sum_{i \in F_a} \rho_i . \tag{5.15}$$

Summing proves the lower bound by the right side of (5.14). Conversely, for any partition $\mathcal{F}$, put on each $F \times F$ the block $u_F u_F^{\mathsf{T}}$, where $u_F = (\sqrt{\delta_i})_{i \in F}$, and put zero outside the blocks. This nonnegative matrix has diagonal δ, rank $|\mathcal{F}|$, and exactly the prescribed mass on the right side of (5.14). Splitting a set with positive square-root sums a, b replaces $(a+b)^2$ by $a^2 + b^2$, decreasing the objective. Thus a minimum uses exactly k sets. □

For $\delta_i = 1$, the minimum is

$$\min_{n_1 + \cdots + n_k = n} \sum_a n_a^2 = N_{n,k} = \frac{n^2}{k} + \frac{s(k-s)}{k}, \tag{5.16}$$

where the sizes are positive integers. The minimum is attained by the direct sum of $k-s$ all-ones blocks of order q and s all-ones blocks of order $q+1$. For unit diagonal magnitudes, this refines the lower bound $\sum_{i,j}|a_{ij}| \geq n^2/k$ by the exact correction $s(k-s)/k$, which is positive precisely when $k \nmid n$. The analogous lower bound $\|A\|_F^2 \geq N_{n,k}$ for the squared *Frobenius norm* $\|A\|_F^2 = \sum_{i,j}|a_{ij}|^2$ is false. For example,

$$A = \begin{pmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{pmatrix} \tag{5.17}$$

has rank two (its eigenvalues are $0, 3/2, 3/2$), unit diagonal, and $\|A\|_F^2 = 9/2 < 5 = N_{3,2}$. Its absolute entrywise mass is six, consistent with (5.16).

6 Stochastic and doubly stochastic matrices

6.1 Stochastic column maxima and attainable diagonals

For row-stochastic matrices the column family has capacity one in the supremum norm. We record the partition explicitly, since it is the input to every application in this section.

Corollary 6.1 (stochastic column maxima)**.** *If a row-stochastic $m \times n$ matrix has rank r and column maxima $p_j = \max_i a_{ij}$, then its column labels partition into r nonempty sets with*

$$\sum_{j \in F_a} p_j \leq 1 \qquad (1 \leq a \leq r). \tag{6.1}$$

The same partition controls any choice of one entry from each column: $\sum_{j \in F_a} a_{i_j j} \leq 1$ for arbitrary row indices i_j. Transposition gives the corresponding row-maximum partition for a column-stochastic matrix.

Proof. This is Corollary 3.2 with $R(A) = 1$. The chosen entries are bounded by their column maxima. □

Corollary 3.1 gives $\min_j p_j \leq 1/\lceil n/r \rceil$. For square matrices this contains the diagonal-entry bound of [10, Theorem 3.1]. Conversely, that theorem yields the column-maximum bound, even for rectangular matrices, by the row selection used after (3.7). By transposition it extends [10, Corollary 3.3] from doubly stochastic to all square column-stochastic matrices: some row has every entry at most $1/\lceil n/r \rceil$.

Theorem 4.1 characterizes which column-maximum profiles are attainable. For prescribed diagonals, the condition $\sum_j x_j \geq 1$ is unnecessary except at rank one.

Theorem 6.2 (attainable stochastic diagonals)**.** *Let $n \geq 2$, $1 \leq k \leq n$, and $d \in [0,1]^n$. If $k \geq 2$, then d is the diagonal of a row-stochastic $n \times n$ matrix of rank at most k if and only if $P(d) \leq k$. If $k = 1$, the attainable diagonals are exactly the probability vectors, namely those with $\sum_i d_i = 1$. The same statements hold for any fixed generalized diagonal.*

Proof. A diagonal is dominated, coordinate by coordinate, by the column maxima, so (6.1) proves necessity for $k \geq 2$. For sufficiency, take a packing of the labels into at most k sets. If it has fewer than two sets, split it into two nonempty sets; when $d = 0$, start with any such split. This is possible because $n, k \geq 2$.

For each resulting set F_a, form a probability row whose entries on F_a are d_j, placing its deficit $1 - \sum_{j \in F_a} d_j$ in one column belonging to another set, and zero elsewhere. Give this row to every

row index in F_a. The matrix has at most k distinct rows, hence rank at most k, and its diagonal is exactly d. The deficit cannot change a diagonal entry because it is placed outside F_a.

A rank-one row-stochastic matrix has identical rows: its rows are proportional, and their sums are all one. Its diagonal therefore lists one probability row. Conversely, identical rows equal to d realize every probability vector. Column permutation gives the last assertion. □

For square row- or column-stochastic matrices of rank at most k, Corollary 5.1 and Theorem 5.2 give the sharp maxima $D_{n,k}$ and $C_{n,k}$ for a fixed diagonal product and the permanent. Their full stochastic equality classifications remain those of [9, 10]. The diagonal characterization above concerns row-stochastic matrices; it does not assert a characterization for doubly stochastic matrices.

6.2 Doubly stochastic entropy, support, and power sums

Column sums provide an additional constraint: each column now sums to one. This converts bounds on the column maxima into bounds on all entries. For $\alpha > 0$, consider the entrywise power sum $\sum_{i,j} a_{ij}^{\alpha}$. Write

$$\begin{aligned} T_{n,k}(\alpha) &= (k-s)q^{2-\alpha} + s(q+1)^{2-\alpha}, \\ L_{n,k} &= (k-s)q\log q + s(q+1)\log(q+1), \end{aligned} \tag{6.2}$$

with $n = qk + s$, $0 \le s < k$, and recall $N_{n,k}$ from (4.9). The *support size* is the number of positive entries.

Theorem 6.3 (entropy, support, and power sums)**.** *If A is an $n \times n$ doubly stochastic matrix and $1 \le k \le n$ is an integer with* $\operatorname{rank} A \le k$*, then*

$$\begin{aligned} |\{(i,j) : a_{ij} > 0\}| &\ge N_{n,k}, \\ \sum_{i,j} a_{ij}^{\alpha} &\ge T_{n,k}(\alpha) \quad (0 < \alpha < 1), \\ -\sum_{i,j} a_{ij}\log a_{ij} &\ge L_{n,k}, \\ \sum_{i,j} a_{ij}^{\alpha} &\le T_{n,k}(\alpha) \quad (1 < \alpha \le 2). \end{aligned} \tag{6.3}$$

Here $0\log 0 = 0$. At $\alpha = 1$, the power sum equals n. All bounds are attained by $B_{n,k}$. They also hold with $A = |Z|$ when Z is real or complex, $\operatorname{rank} Z \le k$*, and $|Z|$ is doubly stochastic, regardless of the rank of $|Z|$.*

The support and entropy bounds are limiting forms of the power-sum inequalities. As α decreases to zero, the power sum tends to the number of positive entries, while $T_{n,k}(\alpha)$ tends to $N_{n,k}$. At $\alpha = 1$, both sides equal n. Subtracting n, dividing by $\alpha - 1$, and taking the limit gives the entropy bound, since

$$\left.\frac{d}{d\alpha}\sum_{i,j} a_{ij}^{\alpha}\right|_{\alpha=1} = \sum_{i,j} a_{ij}\log a_{ij}, \qquad T'_{n,k}(1) = -L_{n,k}.$$

Proof. Let $p_j = \max_i a_{ij}$. Corollary 6.1 gives $P(p) \le \operatorname{rank} A \le k$, and every p_j is positive because its column sum is one. For $1 < \alpha \le 2$,

$$\sum_{i,j} a_{ij}^{\alpha} \le \sum_j p_j^{\alpha-1} \sum_i a_{ij} = \sum_j p_j^{\alpha-1} \le \sum_j (b_{n,k})_j^{\alpha-1} = T_{n,k}(\alpha). \tag{6.4}$$

For $1 < \alpha \le 2$, the power $t^{\alpha-1}$ is increasing and concave, so $x \mapsto \sum_j x_j^{\alpha-1}$ is coordinatewise nondecreasing and Schur-concave. Here we use that a sum of the same concave function over the coordinates is Schur-concave. The final inequality in (6.4) therefore follows from Theorem 4.2. For $0 < \alpha < 1$, the pointwise inequality reverses on positive entries. The power $t^{\alpha-1}$ is decreasing and convex, so the corresponding sum is coordinatewise nonincreasing and Schur-convex on the positive orthant. The dual part of Theorem 4.2, in its form for positive vectors, reverses the final inequality as well, giving the bound for $0 < \alpha < 1$ in (6.3). Zero entries contribute zero to the power sum and are omitted in this pointwise comparison.

For entropy, $a_{ij} \le p_j$ on positive entries gives

$$-\sum_{i,j} a_{ij} \log a_{ij} \ge -\sum_j \log p_j \ge -\log D_{n,k} = L_{n,k}, \tag{6.5}$$

using the column sums and Corollary 4.3. A column of total one and maximum p_j needs at least $1/p_j$ positive entries. Summing and using the reciprocal bound in Corollary 4.3 gives the support inequality.

In $B_{n,k}$, a block of size t has t^2 entries equal to $1/t$. Its contributions to a power sum, entropy, and support are respectively $t^{2-\alpha}$, $t \log t$, and t^2. These give equality throughout. For $A = |Z|$, Corollary 3.2 applied to Z gives the same packing of the column maxima into sets of load at most one. The remainder of the proof is unchanged. □

At $\alpha = 2$, (6.3) reads $\|A\|_F^2 \le k$. This case also has a short proof using singular values, which applies to the modulus extension. Indeed, if $A = |Z|$ is doubly stochastic and $\operatorname{rank} Z \le k$, the unit absolute row and column sums of Z, together with Cauchy–Schwarz, imply that every singular value of Z is at most one. Consequently,

$$\|A\|_F^2 = \|Z\|_F^2 \le \operatorname{rank} Z \le k.$$

The entropy and support bounds retain the integer sizes that the coarser estimates $L_{n,k} \ge n \log(n/k)$ and $N_{n,k} \ge n^2/k$ discard.

Let A be a 5×5 doubly stochastic matrix of rank at most two, and let $H(A_{i\cdot})$ denote the entropy of row i, viewed as a probability vector. Theorem 6.3 gives the following bounds, with entropy expressed as an average over rows:

$$\frac{1}{5}\sum_i H(A_{i\cdot}) \ge \frac{2\log 2 + 3\log 3}{5}, \qquad \sum_{i,j} a_{ij}^{3/2} \le \sqrt{2} + \sqrt{3}. \tag{6.6}$$

Both inequalities are attained by $J_2 \oplus J_3$, which has 13 nonzero entries.

The column-sum assumption is essential: identical rows concentrated in one column give a rank-one row-stochastic matrix with row entropy zero and only n positive entries.

6.3 Information bounds with prescribed marginals

Let $\Pi = (\pi_{ij})$ be the $m \times n$ joint probability matrix of finite random variables X, Y, with m input labels and n output labels. Delete labels of probability zero, so the marginals $\mu_i = \sum_j \pi_{ij}$ and $\nu_j = \sum_i \pi_{ij}$ are positive. The *mutual information* of X and Y is

$$I(X;Y) = \sum_{i,j:\pi_{ij}>0} \pi_{ij} \log \frac{\pi_{ij}}{\mu_i \nu_j}. \tag{6.7}$$

Riazanov and Vyalyi [17, Theorem 4.1] proved

$$I(X;Y) \le \log \operatorname{rank} \Pi.$$

Theorem 6.4 refines this bound by incorporating the marginals.

The entropy optimization H_k^* was defined in (4.12). It measures the largest information that could be retained by recording only which set of a partition into k sets contains the label.

Theorem 6.4 (marginal-sensitive information bound)**.** *If* $\operatorname{rank} \Pi \le k \le \min\{m, n\}$*, then*

$$I(X;Y) \le \min\{H_k^*(\mu), H_k^*(\nu)\} \le \log k. \tag{6.8}$$

Each one-marginal bound is sharp when the other marginal and its alphabet may be chosen. With both marginals prescribed, (6.8) *is a bound; attainment is not asserted in general.*

Proof. The row-stochastic matrix $A = (a_{ij})$, with $a_{ij} = \pi_{ij}/\mu_i$, records the conditional distribution of Y given X. It is called the channel from X to Y and has the same rank as Π. Its column maxima p_j are positive and pack into at most k sets; split to exactly k nonempty sets. Since $\nu_j = \sum_i \mu_i a_{ij}$, (6.7) gives

$$\begin{aligned} I(X;Y) &= H(\nu) + \sum_{i,j:a_{ij}>0} \mu_i a_{ij} \log a_{ij} \\ &\le H(\nu) + \sum_j \nu_j \log p_j \\ &\le H_k^*(\nu). \end{aligned} \tag{6.9}$$

The last step is (4.14) with weights ν_j, applied within the actual packing sets: their masses $W_a = \sum_{j \in F_a} \nu_j$ give $\sum_j \nu_j \log p_j \le H(W) - H(\nu)$. Apply the same argument to Π^{T} for the other marginal. The entropy of a probability vector with k entries is at most $\log k$, by concavity of $-x \log x$.

For sharpness with ν fixed, choose an entropy-maximizing partition and let X be its set index, with probabilities W_a. Conditional on $X = a$, choose $Y = j \in F_a$ with probability ν_j/W_a. The output marginal is ν, the channel has rank k because its rows have disjoint nonempty supports, and X is determined by Y. Direct substitution into (6.7) gives $I(X;Y) = H(W) = H_k^*(\nu)$. Transposition proves the analogous assertion for μ. □

For uniform input on m labels, balanced set sizes maximize (4.12). Consequently, writing $m = qk + s$ with $0 \le s < k$,

$$I(X;Y) \le \log m - \frac{L_{m,k}}{m} = H\big((q/m)^{[k-s]}, ((q+1)/m)^{[s]}\big). \tag{6.10}$$

The right side is strictly below $\log k$ when $0 < s < k$. This refinement requires only a uniform input marginal; the output need not be uniform. A deterministic map onto the balanced partition sets attains it. If both marginals are uniform on m labels, $\Pi = B_{m,k}/m$ attains it. At $m = 5, k = 2$ the value is $H(2/5, 3/5) < \log 2$, continuing the example in (6.6).

Remark (sketch of an alternative proof by row elimination)**.** The row-elimination argument of Riazanov and Vyalyi [17] also yields the marginal refinement in (6.8). Whenever more than k nonzero rows remain, choose a nontrivial real dependence $\sum_i c_i \pi_{i\cdot} = 0$, and replace row i by $(1 + \varepsilon c_i)\pi_{i\cdot}$, restricting ε so that every factor is nonnegative. This preserves the column marginal and the conditional distribution in each surviving row. Mutual information is therefore affine in

ε. The coefficients c_i have both signs, so either direction reaches a value at which a row vanishes. Choose a direction in which information does not decrease. Row scaling and deletion cannot increase rank, so repetition leaves at most k nonzero rows.

Viewed as a channel from the fixed output marginal to these row labels, mutual information is convex in the channel: the joint law and the product of its marginals depend linearly on that channel when its input marginal is fixed, and the function $(a, b) \mapsto a \log(a/b)$ is jointly convex. Every channel is a convex combination of deterministic channels, so one of them has at least as much information.

A deterministic channel partitions the output labels into at most k nonempty sets. Refining such a partition to exactly k sets cannot decrease its entropy. This gives $I(X;Y) \le H_k^*(\nu)$; transposition gives $I(X;Y) \le H_k^*(\mu)$.

The packing proof also gives a modulus extension: (6.8) holds if $\Pi = |Z|$ for a real or complex matrix Z with rank at most k. Normalize the rows of Z by μ_i and apply Corollary 3.2 to its absolute column maxima; then (6.9) applies verbatim. No bound on rank $|Z|$ is required.

6.4 Limits of balanced-block extremality

The balanced stochastic matrix $B_{n,k}$ attains all the bounds in Theorem 6.3, but this does not make it an extremizer for every stochastic objective. Already the cubic power sum distinguishes two rank-two doubly stochastic matrices of order four:

$$\sum_{i,j} (J_2 \oplus J_2)_{ij}^3 = 1, \qquad \sum_{i,j} (J_1 \oplus J_3)_{ij}^3 = 1 + \frac{1}{3} = \frac{4}{3}. \tag{6.11}$$

Thus the balanced stochastic matrix does not maximize the cubic power sum. The packing proof explains its own range: in (6.4), the function of a column maximum is $p^{\alpha-1}$, which is concave precisely for $1 < \alpha \le 2$ and becomes convex above two. Optimizing general convex entrywise functions and treating all exponents above two calls for an additional reduction. Those questions, together with general divergences, are reserved for a separate treatment.

Acknowledgments

The proofs in this paper were developed by OpenAI's GPT models (Sol and Astra) and Anthropic's Claude Fable models under the author's guidance. The author has reviewed the resulting proof arguments. The author takes full responsibility for the contents of this paper.

References

[1] G. Ambrus and F. Grundbacher, "Large Signed Sums and the Polarization Constant of Convex Bodies," arXiv:2502.13752v2, 2026.

[2] A. Brauer, "Limits for the characteristic roots of a matrix. II," *Duke Mathematical Journal* **14** (1947), 21–26. doi:10.1215/S0012-7094-47-01403-8.

[3] F. Cicalese, L. Gargano and U. Vaccaro, "Bounds on the Entropy of a Function of a Random Variable and Their Applications," *IEEE Transactions on Information Theory* **64** (2018), 2220–2230. doi:10.1109/TIT.2017.2787181. Preprint: arXiv:1712.07906.

[4] K. Fan and A. J. Hoffman, "Lower bounds for the rank and location of the eigenvalues of a matrix," in O. Taussky (ed.), *Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*, National Bureau of Standards Applied Mathematics Series **39**, U.S. Government Printing Office, Washington, D.C., 1954, 117–130.

[5] A. J. Hoffman, "On the variation of co-ordinates in subspaces," *Annali di Matematica Pura ed Applicata* **86** (1970), 53–59. doi:10.1007/BF02415709.

[6] G. J. O. Jameson, *Summing and Nuclear Norms in Banach Space Theory*, London Mathematical Society Student Texts **8**, Cambridge University Press, 1987.

[7] M. I. Kadets and M. G. Snobar, "Some functionals over a compact Minkovskii space," *Mathematical Notes* **10** (1971), 694–696. doi:10.1007/BF01106467.

[8] Y. Lavi, "The permanent and diagonal products on the set of nonnegative matrices with bounded rank," arXiv:1808.00016v1, 2018.

[9] Y. Lavi, "The Maximum Diagonal Product of a Stochastic Matrix of Bounded Rank," arXiv:2611.01763v1, 2026.

[10] Y. Lavi, "The Maximum Permanent of a Stochastic Matrix of Bounded Rank," arXiv:2609.27911v1, 2026.

[11] D. R. Lewis, "Ellipsoids defined by Banach ideal norms," *Mathematika* **26** (1979), 18–29. doi:10.1112/S0025579300009566.

[12] M. Marcus and H. Minc, "Generalized matrix functions," *Transactions of the American Mathematical Society* **116** (1965), 316–329. doi:10.1090/S0002-9947-1965-0194445-9.

[13] A. W. Marshall, I. Olkin and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed., Springer, 2011. doi:10.1007/978-0-387-68276-1.

[14] A. Ostrowski, "Über die Determinanten mit überwiegender Hauptdiagonale," *Commentarii Mathematici Helvetici* **10** (1937), 69–96. doi:10.1007/BF01214284.

[15] A. Ostrowski, "Ueber das Nichtverschwinden einer Klasse von Determinanten und die Lokalisierung der charakteristischen Wurzeln von Matrizen," *Compositio Mathematica* **9** (1951), 209–226. Numdam: CM_1951___9___209_0.

[16] G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge Tracts in Mathematics **94**, Cambridge University Press, 1989.

[17] A. Riazanov and M. Vyalyi, "Exploring the bounds on the positive semidefinite rank," arXiv:1704.06507v1, 2017.