

A CERTIFIED UPPER BOUND FOR THE BEAT THE AVERAGE GAME

TEJASVI SINGH TOMAR

ABSTRACT. Let C be the supremum of $\mathbb{P}[X_1 + X_2 + X_3 < 2X_4]$ over all probability measures on $[0, \infty)$, where $X_1, \dots, X_4$ are independent with the given law. This is Problem 6.39 in the list of Georgiev, Gómez-Serrano, Tao and Wagner. Bellec and Fritz proved $0.4006953 \leq C \leq 673/1615 < 0.41672$ and conjectured that the lower bound is the exact value. We prove $C \leq 4006983/10000000 = 0.4006983$, which is within $3 \cdot 10^{-6}$ of the conjectured value. The proof applies the flag algebra method of Razborov to samples of eight points. After sorting, such a sample determines which of the 280 inequalities $2x_d > x_a + x_b + x_c$ among its coordinates hold, and the probability in question is the expected fraction of these inequalities that hold. Independence of the sample imposes positive semidefinite constraints on the distribution of these sign patterns, and since they hold for every measure, no limit argument is needed. We combine them in a semidefinite program, round its solution to an exact sum of squares, and evaluate the bound in exact rational arithmetic on all 45,319,744 patterns of eight points whose restrictions to seven points are realizable. Gordan certificates exclude the two of these patterns that exceed the bound, and trees of Gordan certificates show that the list of realizable patterns of seven points is complete. A checker that uses only the Python standard library verifies the certificate.

1. INTRODUCTION

For a probability measure μ on $[0, \infty)$ let $X_1, X_2, \dots$ be independent random variables with law μ , and put

$$P(\mu) = \mathbb{P}[X_1 + X_2 + X_3 < 2X_4], \quad C = \sup_{\mu} P(\mu).$$

Fritz asked for the value of C on MathOverflow [4]. Determining C is Problem 6.39 in the list of Georgiev, Gómez-Serrano, Tao and Wagner [5], who call it the beat the average game. The name comes from a game with three dice, in which the casino maximizes $\mathbb{P}[2X_1 < X_2 + X_3]$ [2, Section 1]. Bellec and Fritz [2] proved

$$(1.1) \quad 0.4006953 \dots = \max_{0 < p < 1} \frac{p(2-p)}{1+p+p^2+p^3} \leq C \leq \frac{673}{1615} < 0.41672.$$

Their lower bound comes from measures with an atom of mass p at 0 and equal masses at the points $1 - 2^{-i}$ for $1 \leq i \leq N$, and a scaling construction repeats this pattern at smaller and smaller scales [2, Propositions 3.3 and 4.1]. The maximum in (1.1) is attained at the root $p^* = 0.4743464 \dots$ of $p^4 - 4p^3 - 3p^2 - 2p + 2$ in $(0, 1)$ [2, Remark 4.2]. Yu [11] pointed out that this weight improves on $p = 1/2$. Bellec and Fritz conjectured that the lower bound in (1.1) is the exact value of C [2, Conjecture 5.1]. A computer search with the method of [8] found a measure with $P(\mu) \geq 0.389$ [5].

The upper bound in (1.1) rests on a lemma that Bellec and Fritz attribute to a MathOverflow answer of Lackner [2, Lemma 2.3]. For every $m \geq 4$, it bounds C by the largest fraction of the $4\binom{m}{4}$ inequalities $x_i + x_j + x_k < 2x_l$ among m nonnegative reals that can hold simultaneously. Bellec and Fritz computed this fraction for $m = 20$ by mixed-integer linear programming. For

2020 *Mathematics Subject Classification*. 60E15; 90C22, 52C35, 68V05.

Key words and phrases. Beat the average game, independent identically distributed random variables, flag algebras, semidefinite programming, sums of squares, computer-assisted proof.

$m = 25$ their computation gives $20898/50600 < 0.41301$, but they do not have a proof of this bound [2, Section 4].

Theorem 1.1. *For every probability measure μ on $[0, \infty)$ we have $P(\mu) \leq 4006983/10000000$. Hence $0.4006953 \leq C \leq 0.4006983$.*

The upper bound exceeds the conjectured value $0.4006953 \dots$ by less than $3 \cdot 10^{-6}$.

The method. We sample k points from μ , sort them, and record which of the $4\binom{k}{4}$ inequalities $2y_d > y_a + y_b + y_c$ among distinct sorted coordinates hold. Then $P(\mu)$ is the expected fraction of the inequalities that hold (Lemma 2.3), and Lacker's lemma bounds this expectation by its maximum over the sign patterns that occur. The distribution of the pattern, however, comes from an independent sample. Fix t sample points and two disjoint groups of $m - t$ further points. Given the fixed points, the two groups are independent and have the same law. Hence the matrix of the probabilities that the first group forms one configuration with the fixed points and the second group another is positive semidefinite. This is the positivity used in Razborov's flag algebras [9], and for independent samples it holds for every single measure (Lemma 3.1). We combine these constraints in a semidefinite program (Theorem 3.3 and Section 5).

For the proof of Theorem 1.1 we take $k = 8$. Its certificate consists of trees of Gordan certificates, which generate the complete list of the 136,800 realizable sign patterns of 7 sorted points, Gordan certificates for two patterns of 8 points, and positive semidefinite matrices given as exact sums of squares (Sections 4 and 5). Our checker evaluates the bound in exact rational arithmetic on every pattern of 8 points whose restrictions to 7 points are realizable, and a second, independent implementation reproduces this computation. With seven points the same method gives $C \leq 0.40076$ from a certificate evaluated on all 136,800 realizable patterns, and with six points it gives $C \leq 0.402546$ (Table 1).

2. SIGN PATTERNS OF SAMPLES

We first express $P(\mu)$ through the distribution of a finite combinatorial object. For $k \geq 1$ write $[k] = \{1, \dots, k\}$ and

$$T_k = \{(d; a, b, c) : d \in [k], a < b < c, \{a, b, c\} \subseteq [k] \setminus \{d\}\}, \quad L_\tau(x) = 2x_d - x_a - x_b - x_c$$

for $\tau = (d; a, b, c) \in T_k$ and $x \in \mathbb{R}^k$. Then $|T_k| = 4\binom{k}{4}$; in particular, T_k is empty for $k \leq 3$. A point $x \in \mathbb{R}^k$ is *generic* if its coordinates are positive and pairwise distinct and $L_\tau(x) \neq 0$ for all $\tau \in T_k$. A k -*pattern* is a vector $M \in \{0, 1\}^{T_k}$. The *pattern* $\pi(x)$ of a generic point x has $\pi(x)_\tau = 1$ if $L_\tau(x) > 0$ and $\pi(x)_\tau = 0$ otherwise, where y is the increasing rearrangement of x . A pattern is *realizable* if it is the pattern of a generic point, and $\mathcal{R}_k$ denotes the set of realizable k -patterns. The *density* of a k -pattern with $k \geq 4$ is $d(M) = |T_k|^{-1} \sum_\tau M_\tau$. For a set $U = \{u_1 < \dots < u_m\} \subseteq [k]$ the *restriction* $M|_U$ is the m -pattern with $(M|_U)_{(d; a, b, c)} = M_{(u_d; u_a, u_b, u_c)}$. If x is generic with increasing rearrangement y , then $\pi(x)|_U$ is the pattern of $(y_{u_1}, \dots, y_{u_m})$.

Lemma 2.1. *C is the supremum of $P(\mu)$ over the absolutely continuous probability measures μ with $\mu((0, \infty)) = 1$.*

Proof. Let μ be a probability measure on $[0, \infty)$ and $\varepsilon > 0$. Let $U_1, \dots, U_4$ be uniform on $[0, 1]$, with $X_1, \dots, X_4, U_1, \dots, U_4$ independent, and put $Y_i = X_i + \varepsilon(1 + U_i)$. The Y_i are independent with a common law μ_ε , which is absolutely continuous and satisfies $\mu_\varepsilon([\varepsilon, \infty)) = 1$. We have

$$Y_1 + Y_2 + Y_3 - 2Y_4 = X_1 + X_2 + X_3 - 2X_4 + \varepsilon(1 + U_1 + U_2 + U_3 - 2U_4),$$

and the last term is at most 4ε . Hence $P(\mu_\varepsilon) \geq \mathbb{P}[X_1 + X_2 + X_3 - 2X_4 < -4\varepsilon]$, and the right side tends to $P(\mu)$ as $\varepsilon \rightarrow 0$. So the supremum in the lemma is at least $P(\mu)$. The reverse inequality holds because the supremum runs over a subset of the measures in the definition of C . $\square$

Lemma 2.2. *Let μ have no atoms and satisfy $\mu((0, \infty)) = 1$. Then $(X_1, \dots, X_k)$ is generic almost surely.*

Proof. For $i \neq j$ we have $\mathbb{P}[X_i = X_j] = \int \mu(\{x\}) d\mu(x) = 0$. For $\tau = (d; a, b, c)$ the variable X_d is independent of (X_a, X_b, X_c) , so $\mathbb{P}[L_\tau(X) = 0] = \mathbb{E} \mu(\{(X_a + X_b + X_c)/2\}) = 0$. The coordinates are positive almost surely because $\mu((0, \infty)) = 1$. $\square$

For such μ we write $p_\mu(M) = \mathbb{P}[\pi(X_1, \dots, X_k) = M]$. By Lemma 2.2, p_μ is a probability vector supported on $\mathcal{R}_k$.

Lemma 2.3. *Let $k \geq 4$, and let μ have no atoms and satisfy $\mu((0, \infty)) = 1$. Then $P(\mu) = \sum_M p_\mu(M) d(M)$.*

Proof. Let $x \in \mathbb{R}^k$ be generic and let σ be the permutation of $[k]$ with $x_{\sigma(1)} < \dots < x_{\sigma(k)}$, so that $y = (x_{\sigma(1)}, \dots, x_{\sigma(k)})$ is the increasing rearrangement of x . For $\tau = (d; a, b, c)$ let $\sigma(\tau) \in T_k$ be the tuple with first entry $\sigma(d)$ and remaining entries $\sigma(a), \sigma(b), \sigma(c)$ in increasing order. Then $\tau \mapsto \sigma(\tau)$ is a bijection of T_k and $L_{\sigma(\tau)}(y) = L_\tau(x)$. Hence $|T_k| d(\pi(x))$ is the number of $\tau \in T_k$ with $L_\tau(x) > 0$. We apply this to $X = (X_1, \dots, X_k)$ and take expectations:

$$|T_k| \sum_M p_\mu(M) d(M) = \sum_{\tau \in T_k} \mathbb{P}[L_\tau(X) > 0].$$

For $\tau = (d; a, b, c)$ the vector (X_a, X_b, X_c, X_d) has the law of (X_1, X_2, X_3, X_4) , so every term on the right equals $P(\mu)$. $\square$

Since p_μ is supported on $\mathcal{R}_k$, Lemmas 2.1 and 2.3 give $C \leq \max_{M \in \mathcal{R}_k} d(M)$. This is the bound of Lacker's lemma with $m = k$, because a point of $[0, \infty)^k$ can be perturbed to a generic point without violating the strict inequalities that it satisfies.

Remark 2.4. Two features of sorted patterns matter below. First, a permutation of the sample points permutes T_k , so the fraction of the inequalities that hold, and the flag counts of Section 3, depend on a sample only through its pattern (proofs of Lemmas 2.3 and 3.1). Second, realizability refers to points with positive coordinates, and this condition cannot be dropped. Indeed, L_τ is not invariant under translations. If $Z_1, \dots, Z_4$ are independent and identically distributed, then

$$\mathbb{P}[(Z_1 - c) + (Z_2 - c) + (Z_3 - c) < 2(Z_4 - c)] = \mathbb{P}[Z_1 + Z_2 + Z_3 - 2Z_4 < c],$$

which tends to 1 as $c \rightarrow \infty$. For this reason the inequality $x_1 > 0$ appears in every system of Section 4.

3. POSITIVITY FROM INDEPENDENCE

We now derive the constraints that independence imposes on p_μ . Fix integers $0 \leq t \leq m$. A *flag* of size m with t marked points is a pair $F = (N, J)$ of an m -pattern N and a set $J \subseteq [m]$ with $|J| = t$. Its *type* is the t -pattern $N|_J$. Let $k \geq 2m - t$ and let M be a k -pattern. For disjoint sets $J, A \subseteq [k]$ with $|J| = t$ and $|A| = m - t$ put

$$\mathfrak{fl}(M; J, A) = (M|_{J \cup A}, J'),$$

where $J' \subseteq [m]$ is the set of positions of the elements of J in the increasing enumeration of $J \cup A$. For flags F, G of size m with t marked points let

$$(3.1) \quad c(F, G; M) = \frac{1}{D} \# \{ (J, A, B) : \mathfrak{fl}(M; J, A) = F, \mathfrak{fl}(M; J, B) = G \},$$

where (J, A, B) runs over the triples of pairwise disjoint subsets of $[k]$ with $|J| = t$ and $|A| = |B| = m - t$, and $D = \binom{k}{t} \binom{k-t}{m-t} \binom{k-m}{m-t}$ is the number of these triples. Since $\mathfrak{fl}(M; J, A)$ and

$\text{fl}(M; J, B)$ both have type $M|_J$, we have $c(F, G; M) = 0$ unless F and G have the same type. Exchanging A and B shows that $c(F, G; M) = c(G, F; M)$.

Lemma 3.1. *Let μ have no atoms and satisfy $\mu((0, \infty)) = 1$, and let $k \geq 2m - t$. Let $\mathcal{F}$ be a finite set of flags of size m with t marked points, and let $Q = (Q_{FG})_{F, G \in \mathcal{F}}$ be a positive semidefinite matrix. Then*

$$\sum_M p_\mu(M) \sum_{F, G \in \mathcal{F}} Q_{FG} c(F, G; M) \geq 0.$$

Proof. For a point z with positive coordinates indexed by a finite set I , and for disjoint $J, A \subseteq I$ with $|J| = t$ and $|A| = m - t$, we define $g(z; J, A)$ as follows. If the values z_i with $i \in J \cup A$ are pairwise distinct and, listed in increasing order, form a generic point $w \in \mathbb{R}^m$, then $g(z; J, A) = (\pi(w), J')$, where $J' \subseteq [m]$ is the set of positions of the values z_j with $j \in J$ in this list. Otherwise $g(z; J, A) = *$, a symbol that is not a flag.

Let $x \in \mathbb{R}^k$ be generic and let σ be the permutation of $[k]$ with $x_{\sigma(1)} < \dots < x_{\sigma(k)}$. Then $\text{fl}(\pi(x); J, A) = g(x; \sigma(J), \sigma(A))$, since both sides are read off from the increasing list of the values $x_{\sigma(i)}$ with $i \in J \cup A$. As σ permutes the triples (J, A, B) ,

$$(3.2) \quad c(F, G; \pi(x)) = \frac{1}{D} \# \{ (J, A, B) : g(x; J, A) = F, g(x; J, B) = G \}.$$

Let $X = (X_1, \dots, X_k)$. By Lemma 2.2 and (3.2),

$$\sum_M p_\mu(M) c(F, G; M) = \frac{1}{D} \sum_{(J, A, B)} \mathbb{P}[g(X; J, A) = F, g(X; J, B) = G].$$

Put $J_0 = \{1, \dots, t\}$, $A_0 = \{t+1, \dots, m\}$ and $B_0 = \{m+1, \dots, 2m-t\}$. For each triple (J, A, B) there is a permutation ρ of $[k]$ with $\rho(J_0) = J$, $\rho(A_0) = A$ and $\rho(B_0) = B$. The vector $X \circ \rho = (X_{\rho(1)}, \dots, X_{\rho(k)})$ has the same law as X , and $g(X; J, A) = g(X \circ \rho; J_0, A_0)$ and $g(X; J, B) = g(X \circ \rho; J_0, B_0)$. Hence each term of the sum equals $\mathbb{P}[g(X; J_0, A_0) = F, g(X; J_0, B_0) = G]$.

The vectors $W = (X_1, \dots, X_t)$, $Y = (X_{t+1}, \dots, X_m)$ and $Y' = (X_{m+1}, \dots, X_{2m-t})$ are independent, and Y and Y' have the same law. For $w \in (0, \infty)^t$ let $\varphi_F(w) = \mathbb{P}[g((w, Y); J_0, A_0) = F]$. The flag $g(X; J_0, A_0)$ depends only on (W, Y) , and $g(X; J_0, B_0)$ is the same function of (W, Y') . By Fubini's theorem,

$$\mathbb{P}[g(X; J_0, A_0) = F, g(X; J_0, B_0) = G] = \mathbb{E}[\varphi_F(W) \varphi_G(W)].$$

Therefore the sum in the lemma equals $\mathbb{E}[\sum_{F, G \in \mathcal{F}} Q_{FG} \varphi_F(W) \varphi_G(W)]$, which is nonnegative because Q is positive semidefinite. $\square$

Remark 3.2. The flags in Lemma 3.1 need not be realizable or share a type. Up to normalization, $c(F, G; M)$ is the coefficient of M in the averaged product of F and G in Razborov's flag algebra [9], and Lemma 3.1 is the counterpart of the positivity of averaged squares. Razborov proves this positivity for limits of sequences of finite structures. For independent samples it holds for each measure separately, so we need neither limit objects nor de Finetti's theorem. Our structures are linearly ordered and have no nontrivial automorphisms, as in the flag algebras of permutations [1, 10]. This is why a flag can be written as a pattern together with a set of positions.

Theorem 3.3. *Let $k \geq 4$, and let $\mathcal{P}$ be a set of k -patterns with $\mathcal{R}_k \subseteq \mathcal{P}$. For $b = 1, \dots, r$ let $\mathcal{F}_b$ be a finite set of flags of size m_b with t_b marked points, where $2m_b - t_b \leq k$, and let Q_b be a positive semidefinite matrix indexed by $\mathcal{F}_b$. For a k -pattern M put*

$$(3.3) \quad \text{val}(M) = d(M) + \sum_{b=1}^r \sum_{F, G \in \mathcal{F}_b} (Q_b)_{FG} c(F, G; M).$$

Then $C \leq \max_{M \in \mathcal{P}} \text{val}(M)$.

Proof. Let μ be absolutely continuous with $\mu((0, \infty)) = 1$. By Lemma 2.3 and by Lemma 3.1 for each b ,

$$P(\mu) = \sum_M p_\mu(M) d(M) \leq \sum_M p_\mu(M) \text{val}(M) \leq \max_{M \in \mathcal{P}} \text{val}(M),$$

where the last step holds because p_μ is a probability vector supported on $\mathcal{R}_k \subseteq \mathcal{P}$. Lemma 2.1 completes the proof. $\square$

Patterns in $\mathcal{P} \setminus \mathcal{R}_k$ can only weaken the bound. Hence it suffices to know a superset of $\mathcal{R}_k$, and we construct one next.

4. COMPLETENESS OF THE LIST OF PATTERNS

We build sets $\mathcal{P}_k \supseteq \mathcal{R}_k$ by adding one point at a time. Restricting a generic point to a subset of its coordinates gives a generic point, so $M|_U \in \mathcal{R}_{|U|}$ for all $M \in \mathcal{R}_k$ and $U \subseteq [k]$. In particular each $M \in \mathcal{R}_k$ extends the pattern $M|_{[k-1]} \in \mathcal{R}_{k-1}$ by its values on the $4 \binom{k-1}{3}$ tuples of $T_k \setminus T_{k-1}$, the tuples that contain k .

Let P be a $(k-1)$ -pattern, let $n = |T_k \setminus T_{k-1}|$, and fix an enumeration $\tau_1, \dots, \tau_n$ of $T_k \setminus T_{k-1}$, which may depend on P . For $s \in \{0, 1\}^i$ with $0 \leq i \leq n$ let $S_P(s)$ be the system of strict inequalities

$$\begin{aligned} x_1 &> 0, & x_{j+1} - x_j &> 0 \quad (1 \leq j < k), \\ (2P_\tau - 1)L_\tau(x) &> 0 \quad (\tau \in T_{k-1}), & (2s_j - 1)L_{\tau_j}(x) &> 0 \quad (1 \leq j \leq i) \end{aligned}$$

in $x \in \mathbb{R}^k$. Each inequality has the form $r \cdot x > 0$ with $r \in \mathbb{Z}^k$. A *Gordan certificate* for a system of such inequalities consists of nonnegative integers y_r , one for each inequality and not all zero, with $\sum_r y_r r = 0$. A system with a Gordan certificate has no solution, since a solution x would give $0 = \sum_r y_r (r \cdot x) > 0$. Conversely, by Gordan's theorem [6], every system of strict homogeneous linear inequalities with rational coefficients and without a solution has a certificate with rational entries, and hence one with integer entries.

An *extension tree* for P is a set $\mathcal{T}$ of sign vectors of lengths $0, \dots, n$ that contains the empty vector and all prefixes of its elements, such that for each $s \in \mathcal{T}$ of length $i < n$ and each $e \in \{0, 1\}$ either $(s, e) \in \mathcal{T}$ or $S_P(s, e)$ has a Gordan certificate. A vector $s \in \mathcal{T}$ of length n is a *leaf*, and it defines the k -pattern that agrees with P on T_{k-1} and takes the value s_j at τ_j .

Proposition 4.1. *Let $\mathcal{P}_{k-1}$ be a set of $(k-1)$ -patterns with $\mathcal{R}_{k-1} \subseteq \mathcal{P}_{k-1}$, and fix an extension tree for each $P \in \mathcal{P}_{k-1}$. Then the set $\mathcal{P}_k$ of the patterns defined by the leaves of these trees satisfies $\mathcal{R}_k \subseteq \mathcal{P}_k$.*

Proof. Let $M \in \mathcal{R}_k$. Then $M = \pi(x)$ for a generic x with $0 < x_1 < \dots < x_k$, and $P = M|_{[k-1]} \in \mathcal{R}_{k-1} \subseteq \mathcal{P}_{k-1}$. Put $s_j = M_{\tau_j}$ in the enumeration of the tree for P . For each i , the point x solves $S_P(s_1, \dots, s_i)$, so this system has no Gordan certificate. By induction on i , the vector $(s_1, \dots, s_i)$ belongs to the tree for all $i \leq n$. The leaf $(s_1, \dots, s_n)$ defines M . $\square$

Starting from the set $\mathcal{P}_3 = \mathcal{R}_3$ of the single empty pattern, Proposition 4.1 produces sets $\mathcal{P}_4, \dots, \mathcal{P}_7$ that contain all realizable patterns. We computed the trees with the linear programming solver HiGHS [7] in floating point. We drop a branch only with a Gordan certificate, which we compute exactly from the support of a floating-point solution of the alternative system $y \geq 0$, $\sum_r y_r r = 0$, $\sum_r y_r = 1$. We keep a leaf only together with an integer point that realizes its pattern, and we check this point in integer arithmetic.

A second enumeration walks through the cells of the arrangement of the hyperplanes $L_\tau = 0$ inside the cone $0 < x_1 < \dots < x_k$ by changing one sign at a time. Since these hyperplanes are

pairwise distinct and distinct from the walls of the cone, a generic segment between two cells crosses them one at a time; hence the walk reaches every cell. It decides each step by an integer point, by a Gordan certificate, or, for $k = 7$, by a restriction to 6 points that is not in the list that the walk found for $k = 6$. It finds the same sets for $k \leq 7$, with an integer point for each pattern. We check these points against the patterns rebuilt from the trees. Hence $\mathcal{P}_k = \mathcal{R}_k$ for $k \leq 7$, although the proof uses only $\mathcal{R}_k \subseteq \mathcal{P}_k$.

The maximal density is $1/2$ on $\mathcal{P}_4$ and on $\mathcal{P}_5$, and Table 1 gives it for larger k . As a check, the value $7/15$ for $k = 6$ agrees with the computation of Bellec and Fritz for $m = 6$ [2, Section 4].

Eight points. For $k = 8$ we do not build extension trees. Let $\mathcal{J}_8$ be the set of 8-patterns M with $M|_U \in \mathcal{P}_7$ for every 7-element set $U \subseteq [8]$. A tuple of T_8 involves four points and hence lies in a 7-element set. Therefore M is determined by its restrictions to 7-element sets. For an 8-pattern M let $S(M)$ be the system

$$x_1 > 0, \quad x_{j+1} - x_j > 0 \quad (1 \leq j < 8), \quad (2M_\tau - 1)L_\tau(x) > 0 \quad (\tau \in T_8).$$

Lemma 4.2. *Let $\mathcal{R}_7 \subseteq \mathcal{P}_7$, and let E be a set of 8-patterns M such that $S(M)$ has a Gordan certificate. Then $\mathcal{P}_8 = \mathcal{J}_8 \setminus E$ satisfies $\mathcal{R}_8 \subseteq \mathcal{P}_8$.*

Proof. Let $M \in \mathcal{R}_8$, realized by a generic point x with $0 < x_1 < \dots < x_8$. Then $M|_U \in \mathcal{R}_7 \subseteq \mathcal{P}_7$ for every 7-element set U , so $M \in \mathcal{J}_8$. The point x solves $S(M)$, so $M \notin E$. $\square$

We obtain $\mathcal{J}_8$ by joining $\mathcal{P}_7$ with itself. The restriction $P = M|_{[7]}$ runs over $\mathcal{P}_7$. For $i \in [7]$ the restriction $M|_{[8] \setminus \{i\}}$ lies in the set of patterns $N \in \mathcal{P}_7$ whose restriction to $[6]$, which corresponds to the points $[7] \setminus \{i\}$, equals $P|_{[7] \setminus \{i\}}$. For $1 \leq i < j \leq 7$ the restrictions of $M|_{[8] \setminus \{i\}}$ and $M|_{[8] \setminus \{j\}}$ to the points $[8] \setminus \{i, j\}$ must agree. Conversely, a family of restrictions with these properties comes from exactly one 8-pattern. Indeed, a tuple $\tau \in T_8$ lies in the restrictions to $[8] \setminus \{u\}$ for the four points $u \notin \tau$, and the conditions make the four values of τ agree. There are 45,319,744 such patterns.

5. THE CERTIFICATE

5.1. The semidefinite program. For fixed k , $\mathcal{P}$ and flag sets $\mathcal{F}_b$, the best bound that Theorem 3.3 gives is the value of the semidefinite program

$$\text{minimize } \lambda \quad \text{subject to} \quad \lambda \geq \text{val}(M) \quad (M \in \mathcal{P}), \quad Q_b \succeq 0 \quad (b = 1, \dots, r).$$

Since $c(F, G; M)$ vanishes for flags of different types, we lose nothing by giving each type its own block. For $k = 8$ we start from all flags (N, J) with N realizable and with (t, m) equal to $(1, 4)$, $(2, 4)$, $(4, 5)$, $(5, 6)$ or $(6, 7)$. These pairs satisfy $2m - t \leq 8$. There is one t -pattern for $t \leq 3$, and there are 3, 30 and 1,290 realizable patterns of 4, 5 and 6 points, which give the blocks. The other pairs with $m \geq 4$, $t < m$ and $2m - t \leq 8$ are $(0, 4)$, $(3, 4)$, $(2, 5)$, $(3, 5)$ and $(4, 6)$. We do not use them; the last three would add 300, 300 and 19,350 flags.

The solver and the cutting planes below need not be trusted, because only the rounded matrices of Section 5.2 enter the proof. We solve the program numerically with the interior-point solver CSDP [3] and add the constraints as cutting planes. We solve it for a subset of $\mathcal{J}_8$, add patterns that violate their constraints by more than 10^{-7} , and repeat while some of the most violated patterns are realizable. A pattern enters the subset only after a linear program finds a point that realizes it. We round this point to an integer point and check it in integer arithmetic, so the subsets lie in $\mathcal{R}_8$. Each round drops the flags that occur in no pattern of the current subset (Table 1 counts the remaining ones). This leaves 596 of the 1,290 blocks of the pair $(6, 7)$, so the certificate has 631 blocks. A bound $\sum_b \text{tr } Q_b \leq 200$ keeps the intermediate programs bounded.

The certificates for $k = 6$ and $k = 7$ come from the same program over $\mathcal{P}_6$ and $\mathcal{P}_7$, with the pairs of Table 1. For $m \leq 3$ all m -patterns are empty. Hence $c(F, G; M)$ does not depend on M , and the pair $(1, 3)$ in the certificate for $k = 6$ only adds a nonnegative constant to every value.

5.2. Rounding. Let $\tilde{Q}$ be a numerical block. We compute an eigendecomposition $\tilde{Q} = \sum_i \alpha_i u_i u_i^\top$ in floating point and discard the eigenvalues at most $10^{-13} \max(1, \max_i \alpha_i)$. For each remaining i we round $2^{40} u_i$ to an integer vector v_i and α_i to a positive multiple w_i of 2^{-50} . For $k = 6$ the vectors use 2^{30} in place of 2^{40} . The block of the certificate is the rational matrix

$$Q = \sum_i w_i (2^{-40} v_i) (2^{-40} v_i)^\top,$$

which is positive semidefinite by construction. Rounding changes the values $\text{val}(M)$ slightly. We do not estimate this change but evaluate $\text{val}(M)$ exactly for the rounded matrices. For these matrices exactly two patterns of $\mathcal{J}_8$ have $\text{val}(M) > 4006983/10000000$. Both are unrealizable. The certificate contains a Gordan certificate for each of them, and we apply Lemma 4.2 with E consisting of these two patterns. The claimed bound $\lambda = 4006983/10000000$ is the exact maximum of val over $\mathcal{P}_8 = \mathcal{J}_8 \setminus E$, rounded up to seven decimal places.

5.3. The exact check. For $k = 8$ the checker evaluates val through patterns of seven points, by the following identities.

Lemma 5.1. *Let $k \geq 5$ and let M be a k -pattern. Let U run over the $(k-1)$ -element subsets of $[k]$, and read $M|_U$ as a $(k-1)$ -pattern.*

- (a) $d(M) = \frac{1}{k} \sum_U d(M|_U)$.
- (b) If $2m - t \leq k - 1$, then $c(F, G; M) = \frac{1}{k} \sum_U c(F, G; M|_U)$ for flags F, G of size m with t marked points.
- (c) If $(t, m) = (k-2, k-1)$, then $c(F, G; M)$ is $\frac{1}{k(k-1)}$ times the number of ordered pairs (a, b) of distinct elements of $[k]$ with $F = \text{fl}(M; [k] \setminus \{a, b\}, \{a\})$ and $G = \text{fl}(M; [k] \setminus \{a, b\}, \{b\})$.

Proof. A triple (J, A, B) covers $s = 2m - t$ points and lies in exactly $k - s$ of the sets U , and $\text{fl}(M; J, A)$ is the same flag whether read in M or in $M|_U$, with the elements of U numbered increasingly. Since $\binom{k}{t} \binom{k-t}{m-t} \binom{k-m}{m-t} = \frac{k}{k-s} \binom{k-1}{t} \binom{k-1-t}{m-t} \binom{k-1-m}{m-t}$, this gives (b). Part (a) follows in the same way, since a tuple of T_k covers four points and $|T_k| = \frac{k}{k-4} |T_{k-1}|$. Part (c) is (3.1) with $J = [k] \setminus \{a, b\}$, $A = \{a\}$, $B = \{b\}$ and $D = k(k-1)$. $\square$

The checker `check639_k8.py` has 278 lines and uses only the Python standard library and the checker `check639.py` of 286 lines for $k \leq 7$. They share with the search programs only the conventions for indexing T_k and the file formats. They perform four steps.

- (i) They rebuild $\mathcal{P}_4, \dots, \mathcal{P}_7$ from the trees. For every absent branch they build the rows of $S_P(s, e)$ from their definitions and check that the stored row indices are valid, that the stored multipliers are positive integers, and that the combination of the rows with these multipliers is zero. They also check that the trees are given for all patterns of the previous level, that each path decides every tuple of $T_k \setminus T_{k-1}$ exactly once, and that every certificate uses only rows of $S_P(s, e)$.
- (ii) They form each Q_b as an exact rational matrix from the terms (w_i, v_i) . They check that all stored numbers are integers, that $w_i \geq 0$, and that $2m_b - t_b \leq 8$, with $(t_b, m_b) = (6, 7)$ whenever $2m_b - t_b = 8$.
- (iii) For each $N \in \mathcal{P}_7$ they compute $f_7(N)$, the sum of $d(N)$ and of the contributions of the blocks with $2m_b - t_b \leq 7$, from (3.1) with $k = 7$, enumerating all triples (J, A, B) . They

then form $\mathcal{J}_8$ by the join of Section 4 and compute, for each $M \in \mathcal{J}_8$,

$$\text{val}(M) = \frac{1}{8} \sum_{|U|=7} f_7(M|_U) + \frac{1}{28} \sum_{1 \leq i < j \leq 8} q(F_{ij}, F_{ji}), \quad F_{ij} = \text{fl}(M; [8] \setminus \{i, j\}, \{i\}),$$

where $q(F, G)$ is the sum of $(Q_b)_{FG}$ over the blocks b with $(t_b, m_b) = (6, 7)$ that contain both F and G . This is Lemma 5.1 together with the symmetry of the matrices Q_b .

- (iv) They check that every $M \in \mathcal{J}_8$ with $\text{val}(M) > \lambda$ comes with a Gordan certificate for $S(M)$, built from its definition, and that $\text{val}(M) \leq \lambda$ for all other $M \in \mathcal{J}_8$.

Proof of Theorem 1.1. The checker accepts the certificate. By step (i), Proposition 4.1 and step (iv), Lemma 4.2 applies to the set E of the two excluded patterns, so $\mathcal{R}_8 \subseteq \mathcal{P}_8 = \mathcal{J}_8 \setminus E$. By step (ii) the matrices Q_b are positive semidefinite and satisfy the hypothesis $2m_b - t_b \leq 8$ of Theorem 3.3. Steps (iii) and (iv) give $\max_{M \in \mathcal{P}_8} \text{val}(M) = 0.4006982904 \dots \leq 4006983/10000000$, and Theorem 3.3 gives $C \leq 4006983/10000000$. The lower bound is (1.1). $\square$

For $k = 6$ and $k = 7$, `check639.py` alone performs step (i) for the levels 4 to k and the checks of step (ii) with the condition $2m_b - t_b \leq k$, and it evaluates $\text{val}(M)$ from the definitions for every $M \in \mathcal{P}_k$. A second implementation, written independently of `check639.py` and `check639.k8.py`, reads only the flags, the matrices and the claimed bound from a certificate, and for $k = 8$ also the excluded patterns. It takes the sets of patterns from the cell walk of Section 4, computes densities and flags from the integer points that realize the patterns instead of from their bits, and enumerates pairs of m -sets $U_1, U_2 \subseteq [k]$ with $|U_1 \cap U_2| = t$ instead of triples (J, A, B) . It obtains the same value $\text{val}(M)$ for every pattern, for the certificates with $k = 6$ and $k = 7$ in Table 1. For $k = 8$ it evaluates the blocks with $2m_b - t_b \leq 7$ in the same way on its own list of 7-patterns, forms the join of Section 4 from the bits of these patterns, and evaluates the candidates by Lemma 5.1. It finds 45,319,744 candidates, the same maximum and the same two patterns above λ , and it verifies their Gordan certificates with rows that it builds itself. It also evaluates $\text{val}(M)$ directly from the definition (3.1), without Lemma 5.1, for the 200 largest values below λ , for a sample of 30,021 further candidates and for the two patterns of E , and it obtains the same values.

TABLE 1. The certificates of Theorem 1.1 ($k = 8$) and of the bounds $C \leq 0.402546$ ($k = 6$) and $C \leq 0.40076$ ($k = 7$). Sizes are in bytes. The trees cover the levels 4 to $\min(k, 7)$. The check times are wall-clock times on a shared server, followed by the number of processes. For $k = 8$ the set $\mathcal{P}_8$ is $\mathcal{J}_8$ without the two patterns of E .

Sample points k	6	7	8
Bound λ	201273/500000	10019/25000	4006983/10000000
$\max_{M \in \mathcal{P}_k} \text{val}(M)$	0.4025451283...	0.4007589336...	0.4006982904...
$ T_k $; $\max_{M \in \mathcal{P}_k} d(M)$	60; 7/15	140; 16/35	280; 25/56
Patterns evaluated	$ \mathcal{R}_6 = 1,290$	$ \mathcal{R}_7 = 136,800$	$ \mathcal{J}_8 = 45,319,744$
Gordan certificates	4,597	967,177	967,177 + 2
Pairs (t, m)	(1, 3), (2, 4), (3, 4), (4, 5)	(1, 4), (2, 4), (4, 5), (5, 6)	as for $k = 7$, and (6, 7)
Flags used	3, 18, 12, 150	12, 18, 150, 3,924	12, 18, 150, 3,576, 17,206
Realizable flags	3, 18, 12, 150	12, 18, 150, 7,740	12, 18, 150, 7,740, 957,600
Blocks; flags (= terms)	6; 183	35; 4,104	631; 20,962
Denominators of v_i ; of w_i	2^{30} ; 2^{50}	2^{40} ; 2^{50}	2^{40} ; 2^{50}
Certificate file	93,872	8,949,389	40,068,357
Trees (JSON)	197,892	54,847,476	54,847,476
Check	1 s, 4	51 s, 24	4.6 min, 24
Second implementation	agrees	agrees	agrees

6. THE DUAL SOLUTION

The dual of the semidefinite program of Section 5 asks for a probability vector y on $\mathcal{P}$ that maximizes $\sum_M y(M) d(M)$ subject to $\sum_M y(M) c_b(M) \succeq 0$ for each b , where $c_b(M) = (c(F, G; M))_{F, G \in \mathcal{F}_b}$. By Lemmas 2.3 and 3.1, the vector p_μ is feasible for every measure μ without atoms that satisfies $\mu((0, \infty)) = 1$, with objective value $P(\mu)$. So the dual is a relaxation of the problem of maximizing $P(\mu)$. Numerically, the dual solution of the last program for $k = 7$, on 4,146 patterns and with the trace bound ignored (its multiplier is $5 \cdot 10^{-11}$), gives weight above 10^{-6} to 27 patterns. More than 99.97% of its mass lies on the 162 patterns observed in 300,000 samples of 7 points from the construction of Bellec and Fritz [2, Propositions 3.3 and 4.1] with $p = 0.474346$ and $N = 50$ in [2, (4.4)], truncated after nine scales, with a uniform variable at the finest scale and a ratio between consecutive scales small enough not to affect the patterns. Conversely, more than 92% of the samples fall on the 27 patterns with weight above 10^{-6} . The relaxation and the lower bound construction thus favor the same patterns, which is consistent with Conjecture 5.1 of [2].

DATA AVAILABILITY

The certificates, the checkers, the second implementation and the search programs are archived at doi:10.5281/zenodo.23506055. The programs and the certificates for $k = 6$ and $k = 7$ are also available at <https://github.com/tejasvi/beat-the-average-bound>.

USE OF AI TOOLS

Generative AI tools were used to find and prove the results, to write the programs, and to draft the text. The author reviewed the arguments and the programs and takes responsibility for the content.

REFERENCES

- [1] J. Balogh, P. Hu, B. Lidický, O. Pikhurko, B. Udvari and J. Volec, *Minimum number of monotone subsequences of length 4 in permutations*, Combin. Probab. Comput. **24** (2015), 658–679.
- [2] P. C. Bellec and T. Fritz, *On maximizing the probability of a linear inequality between iid random variables*, Amer. Math. Monthly (2026), doi:10.1080/00029890.2026.2702871; arXiv:2412.15179v5 (titled *Optimizing over iid distributions and the Beat the Average game* on arXiv), whose numbering we cite.
- [3] B. Borchers, *CSDP, a C library for semidefinite programming*, Optim. Methods Softw. **11** (1999), 613–623.
- [4] T. Fritz, *How large can $\mathbf{P}[X_1 + X_2 + X_3 < 2X_4]$ get?*, MathOverflow question 474916, 2024, <https://mathoverflow.net/q/474916>.
- [5] B. Georgiev, J. Gómez-Serrano, T. Tao and A. Z. Wagner, *Mathematical exploration and discovery at scale*, arXiv:2511.02864, 2025.
- [6] P. Gordan, *Ueber die Auflösung linearer Gleichungen mit reellen Coefficienten*, Math. Ann. **6** (1873), 23–28.
- [7] Q. Huangfu and J. A. J. Hall, *Parallelizing the dual revised simplex method*, Math. Program. Comput. **10** (2018), 119–142.
- [8] A. Novikov et al., *AlphaEvolve: A coding agent for scientific and algorithmic discovery*, arXiv:2506.13131, 2025.
- [9] A. A. Razborov, *Flag algebras*, J. Symbolic Logic **72** (2007), 1239–1282.
- [10] J. Šliachan and W. Stromquist, *Improving bounds on packing densities of 4-point permutations*, Discrete Math. Theor. Comput. Sci. **19** (2018), no. 2, paper 3.
- [11] V. Yu, answer to [4], MathOverflow, 2024, <https://mathoverflow.net/a/484853>.

ORCID: <https://orcid.org/0000-0003-4668-3639>