THEORY OF COMPUTING, Volume 10 (3), 2014, pp. 55–75
www.theoryofcomputing.org

SPECIAL ISSUE: ANALYSIS OF BOOLEAN FUNCTIONS

Dimension-Free L_2 Maximal Inequality for Spherical Means in the Hypercube

Aram W. Harrow* Alexandra Kolla† Leonard J. Schulman‡

Received February 21, 2013; Revised November 12, 2013; Published May 23, 2014

Abstract: We establish the maximal inequality claimed in the title. In combinatorial terms this has the implication that for sufficiently small $\varepsilon > 0$, for all n, any marking of an ε fraction of the vertices of the n-dimensional hypercube necessarily leaves a vertex x such that marked vertices are a minority of every sphere centered at x.

ACM Classification: G.3

AMS Classification: 42B25

Key words and phrases: maximal inequality, Fourier analysis, boolean hypercube

1 Introduction

Let $\mathbb{I}^n$ be the n-dimensional hypercube: the set $\{0,1\}^n$ equipped with Hamming metric

$$d(x,y) = |\{i : x_i \neq y_i\}| .$$

Let $V = \mathbb{R}^{\mathbb{I}^n}$ be the vector space of real-valued functions on the hypercube. For $x \in \mathbb{I}^n$, let π_x denote the *evaluation map*[1] from $V \to \mathbb{R}$ defined by $\pi_x f = f(x)$, for $f \in V$. If $\mathcal{A} \subseteq \mathrm{Hom}(V,V)$ (the vector space

*Supported by NSF grants CCF-0916400 and CCF-1111382, DARPA QuEST contract FA9550-09-1-0044 and ARO contract W911NF-12-1-0486. Part of this work was done while working at the University of Washington.

†Was at Microsoft Research at the beginning and during part of this work.

‡Supported by NSF grants CCF-0829909, CCF-1038578, CCF-1319745, and the NSF-supported Institute for Quantum Information and Matter. This work began during his visit in 2010 to the Theory Group at Microsoft Research, Redmond.

[1]This notation choice is because our paper is replete with operators acting on functions, and the associative composition $\pi_x A f$ is preferable to the cumbersome $(Af)(x)$.

 DOI: 10.4086/toc.2014.v010a003

consisting of all linear mappings from V into itself), the *maximal operator* $M_{\mathcal{A}} : V \to V$ is the sublinear operator in which $M_{\mathcal{A}} f$ is defined by

$$\pi_x M_{\mathcal{A}} f = \sup_{A \in \mathcal{A}} \pi_x A f \,. \tag{1.1}$$

Of interest is the family $\mathcal{S} = \{S_k\}_{k=0}^n$ of *spherical means*, the stochastic linear operators $S_k : V \to V$ given by

$$\pi_x S_k f = \frac{1}{\binom{n}{k}} \sum_{\{y : d(x,y)=k\}} \pi_y f \,.$$

Applying M, we have the *spherical maximal operator* $M_{\mathcal{S}} : V \to V$ defined by

$$\pi_x M_{\mathcal{S}} f = \max_{0 \le k \le n} \pi_x S_k f \,. \tag{1.2}$$

The result of this paper is the following dimension-free bound.

Theorem 1.1. *There is a constant $A_{\mathbb{I}}$ such that for all n, $\|M_{\mathcal{S}}\|_{2\to 2} < A_{\mathbb{I}}$.*

Equivalently, for all n and f, $\|M_{\mathcal{S}} f\|_2 \le A_{\mathbb{I}} \|f\|_2$.

Maximal inequalities for various function spaces such as $L_p(\mathbb{R}^n)$, and with stochastic linear operators defined by various distributions such as uniform on spheres (as above), uniform on balls, or according to the distribution of a random walk of specified length (ergodic averaging), have been extensively studied; see [27] for a good review. Most previous work does not explicitly consider finite or discrete metric spaces; however, see [28] for a maximal inequality on the free group on finitely many generators.

One may ask whether the hypercube bound should follow from known results for larger spaces. The hypercube of dimension greater than 1 does not embed isometrically in Euclidean space of any dimension [10, 25], so inequalities for Euclidean spaces do not seem to be a useful starting point. The hypercube does embed isometrically in $\mathbb{R}^n$ with the L_1 norm, but there is no maximal inequality for this metric. To see this, still in the context of discrete metric spaces, consider the space $\mathbb{Z}^2$ with the L_1 distance. Fixing any nonnegative integer N let f be the indicator function of

$$\left\{x : \sum x_i = 0, \sum |x_i| \le 2N\right\} .$$

Then $\|f\|_2^2 = 2N+1$ while $\|M_{\mathcal{S}} f\|_2^2 \in \Omega(N^2)$. A similar gap (between $O(N^{n-1})$ and $\Omega(N^n)$) occurs in any fixed dimension n, because there exists a set of size $O(N^{n-1})$ constituting a positive fraction of $\Omega(N^n)$ L_1-spheres, necessarily of many radii. It is therefore not the L_1 metric structure of the hypercube which makes a maximal inequality possible, but, essentially, its bounded side-length.

1.1 Combinatorial interpretation

In the special case that f is the indicator function of a set of vertices F in $\mathbb{I}^n$, Theorem 1.1 has the following consequence: For nonnegative ε less than some ε_0 and for all n, if $|F| < \varepsilon 2^n$ then there exists $x \in \mathbb{I}^n$ such that in every sphere about x, the fraction of points which lie in F is $O(\sqrt{\varepsilon})$.

The aspect of interest is that this holds for every sphere about x. The analogous claim for a fixed radius is a trivial application of the Markov inequality; by a union bound the same holds for any constant number of radii. Avoiding the union bound over all $n+1$ radii is the essence of the maximal inequality.

The combinatorial interpretation also has an edge version. Let F' be a set of edges in $\mathbb{I}^n$. Let the distance from a point to an edge be the distance to the closer point on that edge. Theorem 1.1 has the following consequence: For nonnegative ε less than some ε_1 and for all n, if $|F'| < \varepsilon n 2^n$ then there exists $x \in \mathbb{I}^n$ such that in every sphere about x, the fraction of edges which lie in F' is $O(\sqrt{\varepsilon})$. (Define a function f on vertices by $f(y) =$ the fraction of edges adjacent to y that lie in F'. Note that $\|f\|_2 \in O(\sqrt{\varepsilon})$. Apply Theorem 1.1 to f. For the desired conclusion observe that in the sphere of edges of distance k from x, the fraction of edges lying in F' is bounded for $k \le n/2$ by $2\pi_x S_k f$, and for $k > n/2$ by $2\pi_x S_{k+1} f$.)

1.2 Maximal inequalities and Unique Games on the hypercube

In this section we discuss one of the motivations for the present work, although our main result does not directly yield progress on the question.

Khot's *Unique Games Conjecture* (UGC) [14] has, for a decade, been the focus of great attention. Either resolution of the conjecture will have implications for the hardness of approximating NP-hard problems. Falsification, in particular, is likely to provide powerful new algorithmic techniques. On the other hand verification of UGC will imply that improving the currently best known approximation algorithms for many important computational problems such as MIN-2SAT-DELETION [14], VERTEX COVER [16], MAXIMUM CUT [15] and non-uniform SPARSEST CUT [6, 17], is NP-hard. In addition, in recent years, UGC has also proved to be intimately connected to the limitations of semidefinite programming (SDP). Making this connection precise, the authors of [4] and [29] show that if UGC is true, then for every constraint satisfaction problem (CSP) the best approximation ratio is given by a certain simple SDP. While the UGC question, in short, is arguably among the most important challenges in computational complexity, the current evidence in either direction is not strong, and there is no consensus regarding the likely answer.

A Unique Game instance is specified by an undirected constraint graph $G = (V, E)$, an integer k which is the alphabet size, a set of variables $\{x_u\}_{u \in V}$, one for each vertex u, and a set of permutations (constraints) $\pi_{uv} : [k] \to [k]$, one for each (u, v) s.t. $\{u, v\} \in E$, with $\pi_{uv} = (\pi_{vu})^{-1}$. An assignment of values in $[k]$ to the variables is said to satisfy the constraint on the edge $\{u, v\}$ if $\pi_{uv}(x_u) = x_v$. The optimization problem is to assign a value in $[k]$ to each variable x_u so as to maximize the number of satisfied constraints.

Khot [14] conjectured that it is NP-hard to distinguish between the cases when almost all, or very few, of the constraints of a Unique Game are satisfiable:

Conjecture 1.2 (UGC). *For any constants $\varepsilon, \delta > 0$, there is a $k(\varepsilon, \delta)$ such that for any $k > k(\varepsilon, \delta)$, it is NP-hard to distinguish between instances of Unique Games with alphabet size k where at least $1 - \varepsilon$ fraction of constraints are satisfiable and those where at most δ fraction of constraints are satisfiable.*

Numerous works have focused on designing approximation algorithms for Unique Games. While earlier papers [14, 32, 13, 5, 7] presented efficient algorithms for any instance that might have very poor approximation guarantees when run on the worst case input, recent works have focused on ruling out hardness for large classes of instances. Such instances include expanders [3, 26], local expanders [2, 30], and more generally, graphs with a few large eigenvalues [19]. In [20], hardness of random and semi-random distributions over instances is also ruled out. We note that in [1], a sub-exponential time algorithm for general instances is given.

This recent line of work can be seen as a strategy to disprove UGC by ruling out hardness of instances largely based on their spectral profile, and more specifically the number of “large enough" eigenvalues (small set expansion) of the instance’s constraint graph. Following this strategy, the question to ask next is what type of graphs are the ones on which all those techniques fail. The hypercube is typical of such graphs, with its spectrum lying in a “temperate zone" in terms of expansion. This property, along with the high symmetry of the hypercube, makes it a natural next frontier toward disproving UGC.

It is typical when dealing with a $1-\varepsilon$ satisfiable instance to think of it as originating from a completely satisfiable instance, where a malicious adversary picks an ε fraction of edges in the constraint graph and “spoils” them, by modifying their corresponding constraints. Some likely algorithms for Unique Games on the hypercube are seed-and-propagate (exhaustively range over assignments for a small subset of nodes and propagate the assignment from those nodes according to the constraints on the edges, using some fixed conflict resolution strategy), or “local SDP.” In either case the performance of these algorithms depends on whether the adversary can “rip up” the graph enough so that the algorithm has difficulty propagating the seeded values, or otherwise combining local solutions into larger regions—even though in a pure expansion sense the graph is not *so* easy to rip up that the algorithm could afford to work just in local patches of the graph and throw away all the edges between patches. A good strategy to show that the adversary can win is to show that the adversary can select an ε fraction of the edges of the graph in such a way that around every seed vertex x there is a sphere in which a majority of the of the edges have been selected by the adversary.

Theorem 1.1 shows that, to the contrary, no adversary has this power.

It is a challenging problem to analyze the performance of seed-and-propagate algorithms, but the maximal inequality is a favorable indication for the research program aimed at showing their effectiveness.

Regarding seed-and-propagate on the hypercube we only add that seeding at a single vertex is insufficient, as the graph is sufficiently weakly connected that a sub-linear number of edge removals suffices to partition it into components each of sub-linear size. However multiple seedings remains a very viable strategy.

1.3 Possible generalizations

Let $G=(V,E)$ be any finite connected graph, with shortest-path metric d_G. Let $G^{\square n}$ be the nth Cartesian power of G, the graph on V^n in which (v,w) is an edge if there is a unique i for which $(v_i,w_i)\in E$. The shortest path metric on $G^{\square n}$ is therefore the L_1 metric induced by d_G. Spherical operators and spherical maximal operators are now defined, and we conjecture that Theorem 1.1 holds for a suitable constant A_G. Our proof techniques would require modification even for the next simplest cases of G taken to be the path or the complete graph on three vertices.

In a different direction, the existence of a dimension-free bound for all $\mathbb{I}^n$ invites the question whether there is a natural limit object T, such that for every $\mathbb{I}^n$ there exists a single morphism $\mathbb{I}^n\to T$ and thus each n occurs in it as a special case.

1.4 Proof overview

Our proof is in two main steps, in each of which we obtain a maximal inequality for one class of operators based on comparison with another more tractable class. To introduce the first of these reductions we need to define the *senate operator*[2] Sen. Let $\mathcal{T} = \{T_k\}$ be any family of operators indexed by a parameter k which varies over an interval $[0,a]$ (a possibly infinite) of either nonnegative reals or nonnegative integers. (E. g., $\mathcal{S} = \{S_k\}_0^n$ as above.) Then the family $\mathrm{Sen}(\mathcal{T}) = \{\mathrm{Sen}(\mathcal{T})_k\}$, indexed by k in the same range, consists of the operators

$$\mathrm{Sen}(\mathcal{T})_k = \frac{1}{k+1}\sum_{\ell=0}^{k} T_\ell \quad \text{or} \quad \mathrm{Sen}(\mathcal{T})_k = \frac{1}{k}\int_0^k T_\ell \, d\ell$$

depending as k ranges over integers or reals, and taking the limit from above at $k=0$ in the continuous case.

In the first step of our argument we follow a comparison method due to Stein [31] to show that

Proposition 1.3. $\|M_{\mathcal{S}}\|_{2\to 2} \in O(1 + \|M_{\mathrm{Sen}(\mathcal{S})}\|_{2\to 2})$.

Bounds on the Krawtchouk polynomials play a key role in this argument. We will introduce the polynomials and prove these bounds in Section 2, and then use them to prove Proposition 1.3 in Section 3.

To introduce the second reduction we need to define the family of stochastic *noise operators* $\mathcal{N} = \{N_t\}_{t\geq 0}$ indexed by real t. Letting $p = (1-\mathrm{e}^{-t})/2$, we set

$$N_t = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} S_k \,.$$

This has the following interpretation. $\pi_x N_t f$ is the expectation of $\pi_y f$ where y is obtained by running n independent Poisson processes with parameter 1 from time 0 to time t, and re-randomizing the ith bit of x as many times as there are events in the ith Poisson process. The N_t form a semigroup: $N_{t_1}N_{t_2} = N_{t_1+t_2}$. The process is equivalent to a Poisson-clocked random walk on the hypercube.

We show (in Section 4) by a direct pointwise comparison that:

Proposition 1.4. $\|M_{\mathrm{Sen}(\mathcal{S})}\|_{2\to 2} \in O(\|M_{\mathrm{Sen}(\mathcal{N})}\|_{2\to 2})$.

Finally (see [23]) $\|M_{\mathrm{Sen}(\mathcal{N})}\|_{2\to 2} \leq 2\sqrt{2}$ (indeed, $\|M_{\mathrm{Sen}(\mathcal{N})}\|_{p\to p} \leq (p/(p-1))^{1/p}$ for $p > 1$) by previous results: the Hopf-Kakutani-Yosida maximal inequality and Marcinkiewicz's interpolation theorem. For the reader's convenience, we restate these results in the Appendix.

Combining these results, we have Theorem 1.1 by

$$\|M_{\mathcal{S}}f\|_2 \in O(\|f\|_2 + \|M_{\mathrm{Sen}(\mathcal{S})}f\|_2) \subseteq O(\|f\|_2 + \|M_{\mathrm{Sen}(\mathcal{N})}f\|_2) \subseteq O(\|f\|_2)\,.$$

Remark 1.5. While our main result is in terms of the $2 \to 2$ norm, many of our techniques generalize to other norms. Here we are limited by Proposition 1.3, which does not conveniently generalize to other norms. Very recently, this difficulty has been overcome by Krause; for a preliminary version of his work see [22]. On the other hand, $\|M_{\mathcal{S}}\|_{1\to 1} = n+1$, as can be seen by taking f to be nonzero only on a single point.

We are concerned in this paper solely with maximal operators for sets $\mathcal{A}$ of nonnegative matrices. For any such maximal operator $M_{\mathcal{A}}$, $|\pi_x M_{\mathcal{A}} f| \leq \pi_x M_{\mathcal{A}} |f|$. So it suffices to show Theorem 1.1 for nonnegative f; this simplifies some expressions and will be assumed throughout.

[2] The terminology is to express that (as in the United States Senate) each block has equal weight without regard to size.

2 Fourier analysis and Krawtchouk polynomials

For $y \in \mathbb{I}^n$, define the character $\chi_y \in \mathbb{R}^{\mathbb{I}^n}$ by $\pi_x \chi_y = (-1)^{x \cdot y}/\sqrt{2^n}$. The normalization is chosen so that the χ_y form an orthonormal basis of $\mathbb{R}^{\mathbb{I}^n}$. This basis simultaneously diagonalizes each S_k, as they commute with $\mathbb{I}^n$ as an abelian group. A direct calculation (see also [11]) shows that

$$S_k \chi_y = \kappa_k^{(n)}(|y|)\chi_y\,, \tag{2.1}$$

where $|y| = |\{i : y_i = 1\}|$, and $\kappa_k^{(n)}(|y|)$ is the normalized k^{th} Krawtchouk polynomial, defined by

$$\kappa_k^{(n)}(x) = \sum_{j=0}^{k} (-1)^j \frac{\binom{x}{j}\binom{n-x}{k-j}}{\binom{n}{k}}\,. \tag{2.2}$$

We collect here some facts about Krawtchouk polynomials.

Lemma 2.1.

1. k-x Symmetry: $\kappa_k^{(n)}(x) = \kappa_x^{(n)}(k)$.
2. Reflection symmetry: $\kappa_k^{(n)}(n-x) = (-1)^k \kappa_k^{(n)}(x)$.
3. Orthogonality:
$$\sum_{x=0}^{n} \kappa_k^{(n)}(x)\kappa_\ell^{(n)}(x)\binom{n}{x} = \frac{2^n}{\binom{n}{k}}\delta_{k,\ell}\,. \tag{2.3}$$
4. Roots: *The roots of $\kappa_k^{(n)}(x)$ are real, distinct, and lie in the range $n/2 \pm \sqrt{k(n-k)}$.*

The proofs of the first three claims are straightforward (see [24]), and the fourth claim is a weaker version of Theorem 8 of [21]. We sometimes abbreviate $\kappa_k(x) = \kappa_k^{(n)}(x)$.

Before going into further technical detail, we give an overview of our goals in this section. As we have noted in Section 1, maximal inequalities are easily proved for semigroups, such as the noise operators N_t. In some ways S_k resembles $N_{k/n}$, since N_t is approximately an average of S_k for $k = nt \pm \sqrt{nt(1-t)}$. While direct comparison is difficult (e. g., writing S_k as a linear combination of N_t necessarily entails large coefficients), we can argue that the spectra of these operators should be qualitatively similar.

Indeed, the N_t are also diagonal in the χ_y basis, and for $|y| = x$, their eigenvalue for χ_y is $(1-2t)^x$. Thus, our goal in this section will be to show that $\kappa_k(x)$ has similar behavior[3] to $(1-2k/n)^x$. More precisely, we prove

Lemma 2.2. *There is a constant $c > 0$ such that for all n and integer $0 \le x, k \le n/2$,*

$$|\kappa_k^{(n)}(x)| \le \mathrm{e}^{-ckx/n}. \tag{2.4}$$

Due to Lemma 2.1.1,2 it suffices to bound $\kappa_k(x)$ only when $0 \le k \le x \le n/2$.

[3] This qualitative similarity breaks down when k (or x) is close to $n/2$, although this case will not be the main contribution to our bounds. See the introduction to [18] for discussion of the properties of the $S_{n/2}$ operator.

Proof. The main complication in working with Krawtchouk polynomials is that they have several different forms of asymptotic behavior depending on whether x and k are in the lower, middle or upper part of their range; indeed, [9] breaks the asymptotic properties of $\kappa_k(x)$ into 12 different cases. However, for our purpose, we need only two different upper bounds on the Krawtchouk polynomials, based on whether k/n is greater than or less than 0.14; a somewhat arbitrary threshold that we will see the justification for below.

Case I: $k,x \geq 0.14n$**.** This is the simpler upper bound, which relies only on the orthogonality property Lemma 2.1.3. Setting $k = \ell$ in Lemma 2.1.3 and observing that all of the terms on the LHS are nonnegative, it follows that

$$\kappa_k^2(x) \leq \frac{2^n}{\binom{n}{k}\binom{n}{x}}. \tag{2.5}$$

Based on Stirling's formula, Lemma 17.5.1 of [8] states that

$$\binom{n}{np} \geq \frac{\mathrm{e}^{nH_2(p)}}{\sqrt{8p(1-p)n}} \geq \frac{\mathrm{e}^{nH_2(p)}}{\sqrt{2n}} \tag{2.6}$$

where $H_2(p) := -p\ln p - (1-p)\ln(1-p)$.

Note that $H_2(0.14) > \ln(2)/2$, so, for $k,x \geq 0.14n$, (2.5) implies that

$$\kappa_k^2(x) \leq 2n\mathrm{e}^{(\ln(2)-2H_2(0.14))n} \leq 2n\mathrm{e}^{-c_1 n}$$

for some $c_1 > 0.116$. Now let n_0 be sufficiently large ($n_0 = 100$ suffices) that $c_1 \geq 2\ln(2n_0)/n_0$; then for all $n \geq n_0$, $2n\mathrm{e}^{-c_1 n} \leq \mathrm{e}^{-c_1 n/2}$. So for all $n \geq n_0$ and all $0.14n \leq k,x \leq n/2$,

$$\kappa_k^2(x) \leq \mathrm{e}^{-2c_1(n/2)(n/2)/n} \leq \mathrm{e}^{-2c_1 kx/n}.$$

To handle the $n < n_0$ case, we define

$$c_2 = \min\left\{-(n/kx)\ln|\kappa_k^{(n)}(x)| : 1 \leq k \leq x \leq n/2, n < n_0\right\}.$$

It is immediate from Definition (2.2) that $|\kappa_k^{(n)}(x)| < 1$ if $1 \leq k,x \leq n-1$, so $c_2 > 0$.

Finally, the lemma follows with $c = \min\{2c_1, c_2\}$.

Case II: $k \leq 0.14n$ **or** $x \leq 0.14n$**.** By the symmetry between k and x, we can assume without loss of generality that $k \leq 0.14n$.

It is convenient to make the change of variable

$$x = (1-z)n/2,$$

set

$$\mu_k(z) = \kappa_k((1-z)n/2),$$

and expand μ_k as $\mu_k(z) = \sum_{i=0}^{k} \alpha_{k,i} z^i$. μ_k is either symmetric or anti-symmetric about 0, and we focus on bounding $|\mu_k|$ in the range $0 \leq z \leq 1$ corresponding to $0 \leq x \leq n/2$.

Let $y_1,\dots,y_k$ be the roots of μ_k. By Lemma 2.1.2 the multiset $\{y_1,\dots,y_k\}$ is identical to the multiset $\{-y_1,\dots,-y_k\}$. So we can write

$$\mu_k^2(z) = \alpha_{k,k}^2 \prod_{i=1}^{k}(z^2 - y_i^2)\,. \tag{2.7}$$

It is immediate from Definition (2.2) that

$$\mu_k^2(1) = 1\,. \tag{2.8}$$

Furthermore, by Lemma 2.1.4,

$$y_{\max} := \max_i y_i \leq \frac{2}{n}\sqrt{k(n-k)}\,. \tag{2.9}$$

We now obtain an upper bound on $\mu_k^2(z)$ simply by maximizing (2.7) subject to the constraints (2.8) and (2.9). Observe that

$$\mu_k^2(z) = \frac{\mu_k^2(z)}{\mu_k^2(1)} = \prod_{i=1}^{k}\frac{z^2-y_i^2}{1-y_i^2}\,.$$

Consider the problem of choosing y_i to maximize a single term $|z^2-y_i^2|/(1-y_i^2)$. Observe that

$$\frac{\partial}{\partial y_i^2}\frac{z^2-y_i^2}{1-y_i^2} = \frac{z^2-1}{(1-y_i^2)^2} \leq 0\,. \tag{2.10}$$

As a result, the maximum over $|y_i| < z$ is found at $y_i = 0$ and the maximum over $|y_i| > z$ is found at $y_i = y_{\max}$. In the former case, $|z^2-y_i^2|/(1-y_i^2) = z^2$. In the latter case,

$$\frac{|z^2-y_i^2|}{1-y_i^2} \leq \frac{y_{\max}^2 - z^2}{1-y_{\max}^2} \leq \frac{y_{\max}^2}{1-y_{\max}^2} \leq 0.93\,.$$

The last inequality uses the fact that $y_{\max} \leq 2\sqrt{0.14\cdot 0.86}$ (recalling that $k \leq 0.14n$). So $|z^2-y_i^2|/(1-y_i^2) \leq \max(z^2, 0.93)$, implying that

$$\mu_k^2(z) \leq (\max(z^2,0.93))^k\,. \tag{2.11}$$

If $z^2 \geq 0.93$ then we use $z = 1-2x/n \leq \mathrm{e}^{-2x/n}$ to obtain $\kappa_k^2(x) \leq \mathrm{e}^{-4xk/n}$.
If $z^2 \leq 0.93$ then (recalling $x \leq n/2$) we have

$$\kappa_k^2(x) \leq (0.93)^k \leq (0.93)^{2xk/n} = \mathrm{e}^{-cxk/n}$$

for $c = -2\ln(0.93)$. □

The threshold of 0.14 used here could be replaced by any p satisfying $H_2(p) > 1/2 > \sqrt{p(1-p)}$.

3 Senates dominate dictatorships: Proof of Proposition 1.3

Proposition 1.3 (restatement). $\|M_{\mathcal{S}}\|_{2\to 2} \in O(1+\|M_{\mathrm{Sen}(\mathcal{S})}\|_{2\to 2})$.

We start by defining $\overline{\mathcal{S}} = \{S_k\}_0^{\lfloor n/2 \rfloor}$. The operator $M_{\overline{\mathcal{S}}} : V \to V$ is then defined by

$$\pi_x M_{\overline{\mathcal{S}}} f = \max_{0\le k\le \lfloor n/2\rfloor} \pi_x S_k f\,.$$

Define $\iota := S_n$ to be the antipodal operator on $\mathbb{R}^{\mathbb{I}^n}$, i.e., the involution obtained by flipping all n bits of the argument of the function. Then $\pi_x M_{\mathcal{S}} f = \max\{\pi_x M_{\overline{\mathcal{S}}} f, \pi_x \iota M_{\overline{\mathcal{S}}} f\}$. So $\|M_{\mathcal{S}}\|_{2\to 2} \le \sqrt{2}\|M_{\overline{\mathcal{S}}}\|_{2\to 2}$. Proposition 1.3 therefore follows from:

Claim 3.1. *There is a $C<\infty$ such that for all n and f, $\|M_{\overline{\mathcal{S}}} f\|_2 \le C\|f\|_2 + \|M_{\mathrm{Sen}(\overline{\mathcal{S}})} f\|_2$.*

We prove the claim in the following two subsections.

3.1 A method of Stein

The bounds on $\|\overline{\mathcal{S}}_\ell\|_{2\to 2}$ for even and odd radius ℓ are technically distinct (but not in any interesting way). We present the arguments in parallel.

3.1.1 Even radius: S_{2r} for $0 \le r \le r_{\max} = \lfloor \lfloor n/2 \rfloor /2 \rfloor$

Note that for $n = 4m+a$, $0\le a\le 3$, this gives $r_{\max} = m$.

Abel's lemma gives the following easily verified identity:

$$S_{2r} - \frac{1}{r+1}\sum_{k=0}^{r} S_{2k} = \frac{1}{r+1}\sum_{k=1}^{r} k(S_{2k} - S_{2(k-1)})\,. \tag{3.1}$$

Hence we have the following pointwise (that is to say, valid at each point x) inequality for $0 \le r \le r_{\max}$, $r_{\max} = \lfloor \lfloor n/2 \rfloor /2 \rfloor$:

$$\begin{aligned}\left[\pi_x\Big(S_{2r} - \frac{1}{r+1}\sum_{k=0}^{r} S_{2k}\Big)f\right]^2 &= \left[\sum_{k=1}^{r} \frac{\sqrt{k}}{r+1}\cdot\sqrt{k}[\pi_x(S_{2k}-S_{2(k-1)})f]\right]^2 \\ &\le \sum_{k=1}^{r}\frac{k}{(r+1)^2}\sum_{k=1}^{r} k[\pi_x(S_{2k}-S_{2(k-1)})f]^2 \\ &= \frac{r}{2(r+1)}\sum_{k=1}^{r} k[\pi_x(S_{2k}-S_{2(k-1)})f]^2 \\ &\le \frac{1}{2}\sum_{k=1}^{r_{\max}} k[\pi_x(S_{2k}-S_{2(k-1)})f]^2\,.\end{aligned}$$

The first inequality is by Cauchy-Schwarz. This suggests defining an "error term" operator $R_0 : V \to V$ by

$$\pi_x R_0 f = \sqrt{\frac{1}{2} \sum_{k=1}^{r_{\max}} k[\pi_x (S_{2k} - S_{2(k-1)}) f]^2} \tag{3.2}$$

so that for any $r \le r_{\max}$,

$$\|(S_{2r} - \mathrm{Sen}(\mathcal{S})_{2r}) f\|_2^2 \le \|R_0 f\|_2^2 . \tag{3.3}$$

3.1.2 Odd radius: S_{2r+1} for $0 \le r \le r_{\max} = \lfloor (\lfloor n/2 \rfloor - 1)/2 \rfloor$

Note that for $n = 4m + a$, $0 \le a \le 3$, this gives

$$r_{\max} = \begin{cases} m-1 & \text{if } a \in \{0,1\}, \\ m & \text{if } a \in \{2,3\}. \end{cases}$$

Abel's lemma gives:

$$S_{2r+1} - \frac{1}{r+1} \sum_{k=0}^{r} S_{2k+1} = \frac{1}{r+1} \sum_{k=1}^{r} k(S_{2k+1} - S_{2k-1}) . \tag{3.4}$$

Hence we have the following pointwise inequality for $0 \le r \le r_{\max} = \lfloor (\lfloor n/2 \rfloor - 1)/2 \rfloor$:

$$\left[\pi_x \left(S_{2r+1} - \frac{1}{r+1} \sum_{k=0}^{r} S_{2k+1} \right) f \right]^2 =$$
$$\left[\sum_{k=1}^{r} \frac{\sqrt{k}}{r+1} \cdot \sqrt{k} [\pi_x (S_{2k+1} - S_{2k-1}) f] \right]^2$$
$$\dots \le \frac{1}{2} \sum_{k=1}^{r_{\max}} k[\pi_x (S_{2k+1} - S_{2k-1}) f]^2 .$$

This suggests defining an "error term" operator $R_1 : V \to V$ by

$$\pi_x R_1 f = \sqrt{\frac{1}{2} \sum_{k=1}^{r_{\max}} k[\pi_x (S_{2k+1} - S_{2k-1}) f]^2} \tag{3.5}$$

so that for any $r \le r_{\max}$,

$$\|(S_{2r+1} - \mathrm{Sen}(\mathcal{S})_{2r+1}) f\|_2^2 \le \|R_1 f\|_2^2 . \tag{3.6}$$

3.2 Bounding the error term

Define Rf by $\pi_x Rf := \max\{\pi_x R_0 f, \pi_x R_1 f\}$. Combining (3.3) and (3.6) we have for each $x \in \mathbb{I}^n$, each $f \in \mathbb{R}^{\mathbb{I}^n}$ and each $r \le n/2$,

$$|\pi_x(S_r - \mathrm{Sen}(\mathcal{S})_r)f| \le \pi_x Rf\,.$$

Maximizing the LHS over r, squaring and summing over x, we obtain that

$$\|M_{\overline{\mathcal{S}}} f - M_{\mathrm{Sen}(\overline{\mathcal{S}})} f\|_2 \le \|Rf\|_2\,.$$

Claim 3.1 (and Proposition 1.3) now follow from the following lemma.

Lemma 3.2. *There is a $C < \infty$ such that $\|R_0 f\|_2, \|R_1 f\|_2 \le C\|f\|_2$.*

Proof. As seen in the preliminaries, the operators S_k commute and share the eigenvectors $\{\chi_y\}_{y\in\mathbb{I}^n}$, with eigenvalues given by equation (2.1): $S_k\chi_y = \kappa_k^{(n)}(|y|)\chi_y$. Let E_x be the projection operator on the $\binom{n}{|y|}$-dimensional eigenspace spanned by $\{\chi_y\}_{|y|=x}$; so $S_k = \sum_{x=0}^n \kappa_k^{(n)}(x)E_x$. We calculate:

$$\begin{aligned}
\|R_0 f\|_2^2 &= \sum_{z\in\mathbb{I}^n} \frac{1}{2}\sum_{k=1}^{r_{\max}} k[((S_{2k} - S_{2(k-1)})f)(z)]^2 \\
&= \frac{1}{2}\sum_{k=1}^{r_{\max}} k \sum_{x=0}^{n} \left\|(\kappa_{2k}^{(n)}(x) - \kappa_{2(k-1)}^{(n)}(x))E_x f\right\|_2^2 \\
&= \frac{1}{2}\sum_{x=0}^{n} \|E_x f\|_2^2 \sum_{k=1}^{r_{\max}} k(\kappa_{2k}^{(n)}(x) - \kappa_{2(k-1)}^{(n)}(x))^2\,.
\end{aligned}$$

Likewise: (here and below the value of $r_{\max}$ depends on whether R_0 or R_1 is being bounded)

$$\begin{aligned}
\|R_1 f\|_2^2 &= \sum_{z\in\mathbb{I}^n} \frac{1}{2}\sum_{k=1}^{r_{\max}} k[((S_{2k+1} - S_{2k-1})f)(z)]^2 \\
&= \frac{1}{2}\sum_{x=0}^{n} \|E_x f\|_2^2 \sum_{k=1}^{r_{\max}} k(\kappa_{2k+1}^{(n)}(x) - \kappa_{2k-1}^{(n)}(x))^2\,.
\end{aligned}$$

Since $\|f\|_2^2 = \sum_{x=0}^n \|E_x f\|_2^2$, it suffices to show that there is a $C < \infty$ such that for every $0 \le x \le n$,

$$\left[\sum_{k=1}^{r_{\max}} k\left(\kappa_{2k}^{(n)}(x) - \kappa_{2(k-1)}^{(n)}(x)\right)^2\right] \le C\,, \tag{3.7a}$$

$$\left[\sum_{k=1}^{r_{\max}} k\left(\kappa_{2k+1}^{(n)}(x) - \kappa_{2k-1}^{(n)}(x)\right)^2\right] \le C\,. \tag{3.7b}$$

Recall that it suffices by Lemma 2.1.2 to consider $x \le n/2$. For $x = 0$, (3.7) is trivial as the LHS is 0. For $x > 0$ we use Lemma 2.1.1 to rewrite the parenthesized term (with $\ell = 2k$ or $\ell = 2k+1$) as follows:

$$\kappa_x^{(n)}(\ell) - \kappa_x^{(n)}(\ell - 1) = -2\frac{\binom{n-1}{x-1}}{\binom{n}{x}}\kappa_{x-1}^{(n-1)}(\ell-1) = -\frac{2x}{n}\kappa_{x-1}^{(n-1)}(\ell-1)\,.$$

To see this, recall that $\binom{n}{x}\kappa_x^{(n)}(\ell)$ counts x-subsets of $\{1,\ldots,n\}$ according to the parity of their intersection with $\{1,\ldots,\ell\}$; now condition on whether the x-subset contains the element ℓ.

Consequently,

$$\kappa_x^{(n)}(\ell)-\kappa_x^{(n)}(\ell-2)=-\frac{2x}{n}\left(\kappa_{x-1}^{(n-1)}(\ell-1)+\kappa_{x-1}^{(n-1)}(\ell-2)\right)$$

which by a similar argument is

$$=-\frac{4x}{n}\frac{\binom{n-2}{x-1}}{\binom{n-1}{x-1}}\kappa_{x-1}^{(n-2)}(\ell-2)=-\frac{4x}{n}\frac{(n-x)}{(n-1)}\kappa_{x-1}^{(n-2)}(\ell-2)\,. \tag{3.8}$$

The two terms on the LHS of (3.7) are now

$$\sum_{k=1}^{r_{\max}}k\left(\kappa_x^{(n)}(2k)-\kappa_x^{(n)}(2k-2)\right)^2 \tag{3.9}$$

$$=\sum_{k=1}^{r_{\max}}k\left(\frac{4x(n-x)}{n(n-1)}\kappa_{x-1}^{(n-2)}(2k-2)\right)^2\sum_{k=1}^{r_{\max}}k\left(\kappa_x^{(n)}(2k+1)-\kappa_x^{(n)}(2k-1)\right)^2$$

$$=\sum_{k=1}^{r_{\max}}k\left(\frac{4x(n-x)}{n(n-1)}\kappa_{x-1}^{(n-2)}(2k-1)\right)^2. \tag{3.10}$$

For $x=1$, quantities (3.10) come to $\sum_{k=1}^{r_{\max}}16kn^{-2}$ which is upper bounded by a constant.

For $x>1$ we upper bound (3.10), first by

$$\frac{16x^2}{n^2}\sum_{k=1}^{r_{\max}}k\left(\kappa_{x-1}^{(n-2)}(2k-2)\right)^2\quad\text{and}\quad\frac{16x^2}{n^2}\sum_{k=1}^{r_{\max}}k\left(\kappa_{x-1}^{(n-2)}(2k-1)\right)^2$$

which in turn are upper bounded by (applying each value of $r_{\max}$):

$$\frac{16x^2}{n^2}\sum_{k=0}^{\lfloor n/2\rfloor}(k/2+1)\left(\kappa_{x-1}^{(n-2)}(k)\right)^2.$$

Now apply Lemma 2.2 to upper bound this by

$$\frac{16x^2}{n^2}\sum_{k=0}^{\infty}(k/2+1)\mathrm{e}^{-2c(x-1)k/(n-2)}=\frac{16x^2}{n^2}\sum_{k=0}^{\infty}(k/2+1)\mathrm{e}^{-\alpha k} \tag{3.11}$$

$$=\frac{x^2}{n^2}\frac{16(1-\mathrm{e}^{-\alpha})+8\mathrm{e}^{-\alpha}}{(1-\mathrm{e}^{-\alpha})^2} \tag{3.12}$$

$$\le 24\left(\frac{x}{n(1-\mathrm{e}^{-\alpha})}\right)^2$$

$$\le 24\left(\frac{x}{n\alpha}\right)^2$$

$$=24\left(\frac{(n-2)x}{2cn(x-1)}\right)^2$$

$$\le 24/c^2,$$

where in (3.11) we have defined $\alpha = 2c(x-1)/(n-2)$ and in (3.12) we have used the identity

$$\sum_{k=0}^{\infty} k\mathrm{e}^{-\alpha k} = \mathrm{e}^{-\alpha}(1-\mathrm{e}^{-\alpha})^{-2}.$$

This completes the requirement of equation (3.7). □

4 Comparing senate maximal functions: Proof of Proposition 1.4

Proposition 1.4 (restatement). $\|M_{\mathrm{Sen}(\mathcal{S})}\|_{2\to 2} \in O(\|M_{\mathrm{Sen}(\mathcal{N})}\|_{2\to 2})$.

Proof. The proof relies on pointwise comparison of maximal functions.

If A, B are matrices, write $A \le B$ if $B - A$ is a nonnegative matrix. If $\mathcal{A}, \mathcal{B}$ are sets of nonnegative matrices indexed by integers or reals, we write $\mathcal{A} \lesssim \mathcal{B}$ if for every $A \in \mathcal{A}$ there is a probability measure μ_A on $\mathcal{B}$ such that $A \le \int B \, d\mu_A(B)$. Observe that in this case for any nonnegative function f and any x, $\sup_{A\in\mathcal{A}} \pi_x A f \le \sup_{B\in\mathcal{B}} \pi_x B f$, and therefore for any norm, $\|M_{\mathcal{A}}\| \le \|M_{\mathcal{B}}\|$ (and in particular for $\|\cdot\|_{2\to 2}$).

For any $k > \lfloor n/2 \rfloor$,

$$\pi_x \, \mathrm{Sen}(\mathcal{S})_k f \le \pi_x \, \mathrm{Sen}(\mathcal{S})_{\lfloor n/2 \rfloor} (f + \iota f).$$

Therefore $\|M_{\mathrm{Sen}(\mathcal{S})}\|_{2\to 2} \le 2\|M_{\mathrm{Sen}(\overline{\mathcal{S}})}\|_{2\to 2}$. However, we will not compare $\mathrm{Sen}(\overline{\mathcal{S}})$ and $\mathrm{Sen}(\mathcal{N})$ directly. Instead, we will introduce a variant of $\mathcal{N}$ that more closely resembles $\overline{\mathcal{S}}$, but is no longer a semigroup.

Recall that N_t represents the average over independently flipping each bit with probability $p = (1-\mathrm{e}^{-t})/2$. Define $\tilde{N}_p$ to represent the same noise process but parameterized by p instead of t. Thus

$$N_t = \tilde{N}_{\frac{1-\mathrm{e}^{-t}}{2}} \qquad \text{and} \qquad \tilde{N}_p = N_{-\ln(1-2p)}.$$

While the sets $\{N_t\}_{t\ge 0}$ and $\{\tilde{N}_p\}_{p\in[0,1/2)}$ are of course the same, their Senate operators $\mathrm{Sen}(\mathcal{N})$ and $\mathrm{Sen}(\tilde{\mathcal{N}})$ are different:

$$\mathrm{Sen}(\mathcal{N})_T = \frac{1}{T}\int_0^T N_t \, \mathrm{d}t\,, \tag{4.1}$$

$$\mathrm{Sen}(\tilde{\mathcal{N}})_P = \frac{1}{P}\int_0^P \tilde{N}_p \, \mathrm{d}p = \frac{1}{2P}\int_0^{-\ln(1-2P)} \mathrm{e}^{-t} N_t \, \mathrm{d}t\,. \tag{4.2}$$

In (4.2), P is varies over $(0,1/2)$ and in (4.1), T can be any positive real number.

Hence Proposition 1.4 is established in two subsidiary claims:

Lemma 4.1. $\mathrm{Sen}(\tilde{\mathcal{N}}) \lesssim \mathrm{Sen}(\mathcal{N})$.

Lemma 4.2. $\mathrm{Sen}(\overline{\mathcal{S}}) \lesssim C \cdot \mathrm{Sen}(\tilde{\mathcal{N}})$ *for some constant* $C > 0$.

□

Proof of Lemma 4.1. For each P we will write $\mathrm{Sen}(\tilde{\mathcal{N}})_P$ as a convex combination of $\mathrm{Sen}(\mathcal{N})_T$ for different values of T. By considering the action of each side on the constant function we will see that it suffices to write $\mathrm{Sen}(\tilde{\mathcal{N}})_P$ as a bounded nonnegative combination of $\mathrm{Sen}(\mathcal{N})_T$ for different values of T (i. e., a linear combination with coefficients that are nonnegative and whose sum is bounded).

Let $f(t) := \mathrm{e}^{-t}/2P$ and $\tau := -\ln(1-2P)$ so that (4.2) becomes

$$\begin{aligned}
\mathrm{Sen}(\tilde{\mathcal{N}})_P &= \int_0^\tau \mathrm{d}t\, f(t) N_t && (4.3)\\
&= \int_0^\tau \mathrm{d}t\, f(\tau) N_t + \int_0^\tau \mathrm{d}t (f(t) - f(\tau)) N_t && (4.4)\\
&= \tau f(\tau)\,\mathrm{Sen}(\mathcal{N})_\tau + \int_0^\tau \mathrm{d}t \left(\int_t^\tau \mathrm{d}T (-f'(T)) \right) N_t && (4.5)\\
&= \tau f(\tau)\,\mathrm{Sen}(\mathcal{N})_\tau + \int_0^\tau \mathrm{d}T \int_0^T \mathrm{d}t (-f'(T)) N_t && (4.6)\\
&= \tau f(\tau)\,\mathrm{Sen}(\mathcal{N})_\tau + \int_0^\tau \mathrm{d}T (-T f'(T))\,\mathrm{Sen}(\mathcal{N})_T\,. && (4.7)
\end{aligned}$$

Since $f'(T) < 0$, (4.7) is the desired nonnegative combination. We have written the proof in this way to stress that the only features of (4.1) and (4.2) used are that $\mathrm{Sen}(\mathcal{N})_T$ is a "flat" distribution over N_t and $\mathrm{Sen}(\tilde{\mathcal{N}})_P$ has N_t weighted by a weakly decreasing function. □

Proof of Lemma 4.2. To compare $\mathrm{Sen}(\bar{\mathcal{S}})$ and $\mathrm{Sen}(\tilde{\mathcal{N}})$, we need to show that for any $K \le n/2$, we can find a distribution over P such that $\mathrm{Sen}(\bar{\mathcal{S}})_K$ is pointwise $\le$ the appropriate average over $\mathrm{Sen}(\tilde{\mathcal{N}})_P$ times a constant. In fact, it will suffice to consider a distribution that is concentrated on a single value of P. Define

$$P_K := \min\left(\frac{K+\sqrt{K}}{n}, \frac{1}{2}\right).$$

In Lemma 4.3, we will show that $\mathrm{Sen}(S)_K \le C \cdot \mathrm{Sen}(\tilde{\mathcal{N}})_{P_K}$, thus implying that $\mathrm{Sen}(\bar{\mathcal{S}}) \lesssim C \cdot \mathrm{Sen}(\tilde{\mathcal{N}})$. The idea behind Lemma 4.3 is that for each $k \le K$, there are significant contributions to the S_k coefficient of $\mathrm{Sen}(\tilde{\mathcal{N}})_{P_K}$ for p throughout the range $[k/n, (k+\sqrt{k})/n]$. This window has width $\sqrt{k}/n$, contributes $\Omega(1/\sqrt{k})$ weight to S_k at each point and is normalized by $1/P_K \approx n/K$. The total contribution is thus $\Omega(1/K)$.

Lemma 4.3. *Let $n \ge 9$, $K \le n/2$ and $P_K = \min\left(\frac{K+\sqrt{K}}{n}, \frac{1}{2}\right)$. Then* $\mathrm{Sen}(S)_K \le 3\mathrm{e}^{20} \cdot \mathrm{Sen}(\tilde{\mathcal{N}})_{P_K}$,

The true constant is certainly much better, and perhaps closer to $1/2$. We remark that $n \ge 9$ can be assumed without loss of generality, since $\|M_{\mathcal{S}}\|_{2\to 2} \le n+1$ by the triangle inequality.

Proof. Observe that $P := P_K \le 2K/n$.

For $0 \le k \le K$, we now compare the coefficient of S_k in $\mathrm{Sen}(\mathcal{S})_K$ (where it is $1/(K+1)$) to its value in $\mathrm{Sen}(\tilde{\mathcal{N}})_P$, where it is

$$\frac{1}{P}\int_0^P \mathrm{d}p\, B(n,p,k)\,.$$

Denote this latter quantity by a_k.

Consider first $k=0$. If $K=0$ and $P=0$, then S_0 has weight 1 in both cases. Otherwise $P \geq 1/n$, and

$$a_0 \geq \frac{1}{P}\int_0^{1/n}(1-p)^n \mathrm{d}p \geq \frac{n}{2K}\cdot\frac{1}{n}\cdot\left(1-\frac{1}{n}\right)^n \geq \frac{1}{8K}\,.$$

This last inequality uses the fact that $(1-1/n)^n$ is increasing with n and thus is $\geq 1/4$ for $n \geq 2$.

Next, suppose $k+\sqrt{k} \leq n/2$. Then $P \geq (k+\sqrt{k})/n$. We will consider only the contribution to a_k resulting from $0 \leq p-k/n \leq \sqrt{k}/n$. First, use Stirling's formula to bound

$$\binom{n}{k} \geq \exp(nH_2(k/n))\sqrt{\frac{n}{9k(n-k)}} \geq \exp(nH_2(k/n))\frac{1}{3\sqrt{k}}\,,$$

implying that

$$B\big(n,n/k,k\big) \geq \frac{1}{3\sqrt{k}}\,. \tag{4.8}$$

Next, observe that

$$\frac{\mathrm{d}}{\mathrm{d}q}\ln B(n,q,k) = \frac{k-nq}{q(1-q)}\,. \tag{4.9}$$

When $k/n \leq q \leq p \leq (k+\sqrt{k})/n$, we have

$$k-nq \geq k-np \geq -\sqrt{k} \qquad \text{and} \qquad q(1-q) \geq \frac{k}{n}\left(1-\frac{k}{n}\right) \geq \frac{k}{2n}\,.$$

Thus, $0 \leq p-k/n \leq \sqrt{k}/n$ implies that

$$\begin{aligned}\ln\frac{B(n,p,k)}{B(n,k/n,k)} &= \int_{k/n}^{p}\mathrm{d}q\frac{k-nq}{q(1-q)} && (4.10\text{a})\\ &\geq \left(p-\frac{k}{n}\right)\frac{k-np}{k/2n} && (4.10\text{b})\\ &\geq -2\,. && (4.10\text{c})\end{aligned}$$

Combining (4.8) and (4.10), we find that $B(n,p,k) \geq 1/6\mathrm{e}^2\sqrt{k}$ for $k/n \leq p \leq (k+\sqrt{k})/n$. Thus

$$a_k \geq \frac{1}{P}\int_{k/n}^{(k+\sqrt{k})/n}\mathrm{d}p\,B(n,p,k) \geq \frac{n}{2K}\cdot\frac{\sqrt{k}}{n}\cdot\frac{1}{6\mathrm{e}^2\sqrt{k}} \geq \frac{1}{12\mathrm{e}^2K}\,.$$

Finally, we consider the case when $k+\sqrt{k} \geq n/2$. In this regime, we have $P=1/2$. Also, $k \leq n/2$, so $K \geq k \geq n/2-\sqrt{n/2} \geq n/4$ (assuming in the last step that $n \geq 8$).

Consider $p \in [1/2-1/\sqrt{n}, 1/2]$. Assume that $n \geq 9$ so that $1/2-1/\sqrt{n} \geq 1/6$. We can then use (4.9) to bound

$$\begin{aligned}\ln\frac{B(n,\frac{1}{2}-\frac{1}{\sqrt{n}},k)}{B(n,k/n,k)} &\geq -\left(\frac{k}{n}-\frac{1}{2}+\frac{1}{\sqrt{n}}\right)\frac{k-\frac{n}{2}+\sqrt{n}}{\frac{1}{6}\cdot\frac{1}{2}}\\ &\geq -12\left(1-\frac{1}{\sqrt{2}}\right)^2 \geq -20\,.\end{aligned}$$

Similarly

$$\ln \frac{B(n,1/2,k)}{B(n,k/n,k)} \geq \left(\frac{1}{2} - \frac{k}{n}\right) \frac{k-n/2}{1/12} \geq -6\,.$$

We conclude that

$$a_k \geq \frac{1}{P} \int_{\frac{1}{2}-\frac{1}{\sqrt{n}}}^{\frac{1}{2}} \mathrm{d}p\, B(n,p,k) \geq 2 \cdot \frac{1}{\sqrt{n}} \cdot \frac{1}{3\mathrm{e}^{20}\sqrt{k}} \geq \frac{1}{3\mathrm{e}^{20} K}\,.$$

□

This completes the proof of Lemma 4.2. □

A Maximal inequalities on semigroups

In this appendix, we review the maximal ergodic inequality that we use to show $\|M_{\mathrm{Sen}(\mathcal{N})}\|_{1\to 1,w} \leq 1$.

Theorem A.1 (maximal ergodic inequality)**.** *Let A be a positive sublinear operator with $\|A\|_{1\to 1} \leq 1$ and $\|A\|_{\infty\to\infty} \leq 1$. Let $\mathcal{A}$ denote the (discrete) semigroup generated by A. Then*

$$\|M_{\mathrm{Sen}(\mathcal{A})}\|_{1\to 1,w} \leq 1\,.$$

This inequality was originally proved by Kakutani and Yosida (in the case when A is linear) and Hopf (for general semigroups). However since the proof of [12] is simple and self-contained, we restate it here.

Proof. For f an arbitrary function and T a nonnegative integer, define $E_f^T = I[M_{\mathrm{Sen}(\mathcal{A})_{\leq T}} f \geq 0]$. We will first prove that

$$\langle E_f^T, f\rangle \geq 0\,, \tag{A.1}$$

for all T. To see that the theorem follows from this claim, apply (A.1) to $f-\lambda$ and we find that $\lambda \|E_{f-\lambda}^T\|_1 \leq \langle E_{f-\lambda}^T, f\rangle \leq \|f\|_1$. Thus $\|E_{f-\lambda}^T\|_1 \leq \lambda^{-1}\|f\|_1$. Since this inequality holds for all T, it implies that $\|I[M_{\mathrm{Sen}(\mathcal{A})} f \geq \lambda]\|_1$ is also $\leq \lambda^{-1}\|f\|_1$.

We now return to the proof of (A.1). We define for this purpose an unweighted Senate operator:

$$\overline{\mathrm{Sen}}(\mathcal{A})_T = \sum_{t=0}^{T} A_t\,.$$

We abbreviate $M_T := M_{\mathrm{Sen}(\mathcal{A})_{\leq T}}$ and $\bar{M}_T := M_{\overline{\mathrm{Sen}}(\mathcal{A})_{\leq T}}$. Observe that $I[M_T f \geq 0] = I[\bar{M}_T f \geq 0]$. Also define, for any function g, the function $(g)^+$ to be the nonnegative part of g.

Thus, if $0 \leq t \leq T$, then we have $\overline{\mathrm{Sen}}(\mathcal{A})_t f \leq \bar{M}_T f \leq (\bar{M}_T f)^+$. This implies that

$$f + A(\bar{M}_T f)^+ \geq f + A\overline{\mathrm{Sen}}(\mathcal{A})_t f = \overline{\mathrm{Sen}}(\mathcal{A})_{t+1} f\,.$$

Thus, $f \geq \overline{\mathrm{Sen}}(\mathcal{A})_t f - A((\bar{M}_T f)^+)$, (including an $t=0$ case that can be checked separately) and maximizing over $0 \leq t \leq T$, we have

$$f \geq \bar{M}_T f - A((\bar{M}_T f)^+)\,.$$

Now we take the inner product of both sides with E_f^T and find

$$\begin{aligned}
\langle E_f^T, f\rangle &\geq \langle E_f^T, \bar{M}_T f - A((\bar{M}_T f)^+)\rangle && \text{(A.2)}\\
&= \langle E_f^T, (\bar{M}_T f)^+ - A((\bar{M}_T f)^+)\rangle && \text{(A.3)}\\
&= \|(\bar{M}_T f)^+\|_1 - \langle E_f^T, A((\bar{M}_T f)^+)\rangle && \text{(A.4)}\\
&\geq \|(\bar{M}_T f)^+\|_1 - \|A((\bar{M}_T f)^+)\|_1 && \text{(A.5)}\\
&\geq 0\,. && \text{(A.6)}
\end{aligned}$$

The final inequality uses the fact that $\|A\|_{1\to 1} \leq 1$. □

For our purposes, we will want to convert the bound on the $1 \to 1, w$ norm into a bound on $p \to p$ norms, especially for $p = 2$. This is achieved by the Marcinkiewicz interpolation theorem [33].

Theorem A.2. *Let A be a sublinear operator with $\|A\|_{p\to p,w} \leq N_p$ and $\|A\|_{q\to q,w} \leq N_q$. Then for any $1 \leq p < r < q \leq \infty$, we have*

$$\|A\|_{r\to r} \leq 2N_p^{\delta} N_q^{1-\delta} \left(\frac{r(q-p)}{(r-p)(q-r)}\right)^{1/r},$$

with $\delta = p(q-r)/r(q-p)$.

In our case, a maximal operator $M_{\mathcal{A}}$ has $\|M_{\mathcal{A}}\|_{\infty\to\infty,w} = 1$ and if $\mathcal{A}$ is a positive contractive semigroup, then Theorem A.1 implies that $\|M_{\mathcal{A}}\|_{1\to 1,w} \leq 1$ as well. This implies that for any $1 < p$, we have

$$\|M_{\mathcal{A}}\|_{p\to p} \leq 2\left(\frac{p}{p-1}\right)^{1/p}, \tag{A.7}$$

which is $2\sqrt{2}$ when $p = 2$.

Acknowledgments

Thanks to Gil Kalai for a stimulating conversation which pointed us in the direction of maximal inequalities, to Konstantin Makarychev and Yury Makarychev for discussions about UGC on the hypercube, and to Yuval Peres and Terence Tao for helpful comments.

References

[1] SANJEEV ARORA, BOAZ BARAK, AND DAVID STEURER: Subexponential algorithms for Unique Games and related problems. In *Proc. 51st FOCS*, pp. 563–572. IEEE Comp. Soc. Press, 2010. [doi:10.1109/FOCS.2010.59] 57

[2] SANJEEV ARORA, RUSSELL IMPAGLIAZZO, WILLIAM MATTHEWS, AND DAVID STEURER: Improved algorithms for Unique Games via divide and conquer. *Electron. Colloq. on Comput. Complexity (ECCC)*, 17:41, 2010. ECCC. 57

[3] SANJEEV ARORA, SUBHASH KHOT, ALEXANDRA KOLLA, DAVID STEURER, MADHUR TULSIANI, AND NISHEETH K. VISHNOI: Unique Games on expanding constraint graphs are easy. In *Proc. 40th STOC*, pp. 21–28. ACM Press, 2008. [doi:10.1145/1374376.1374380] 57

[4] PER AUSTRIN: Towards sharp inapproximability for any 2-CSP. *SIAM J. Comput.*, 39(6):2430–2463, 2010. Preliminary version in FOCS'07. [doi:10.1137/070711670] 57

[5] MOSES CHARIKAR, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: Near-optimal algorithms for Unique Games. In *Proc. 38th STOC*, pp. 205–214. ACM Press, 2006. [doi:10.1145/1132516.1132547] 57

[6] SHUCHI CHAWLA, ROBERT KRAUTHGAMER, RAVI KUMAR, YUVAL RABANI, AND D. SIVAKUMAR: On the hardness of approximating MULTICUT and SPARSEST-CUT. *Comput. Complexity*, 15(2):94–114, 2006. Preliminary version in CCC'05. [doi:10.1007/s00037-006-0210-9] 57

[7] EDEN CHLAMTÁČ, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: How to play Unique Games using embeddings. In *Proc. 47th FOCS*, pp. 687–696. IEEE Comp. Soc. Press, 2006. [doi:10.1109/FOCS.2006.36] 57

[8] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory*. Series in Telecommunication. John Wiley and Sons, New York, 1991. 61

[9] DIEGO DOMINICI: Asymptotic analysis of the Krawtchouk polynomials by the WKB method. *The Ramanujan Journal*, 15(3):303–338, 2008. See also at arXiv. [doi:10.1007/s11139-007-9078-9] 61

[10] PER ENFLO: On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8(2):103–105, 1970. [doi:10.1007/BF02589549] 56

[11] PHILIP FEINSILVER AND JERZY KOCIK: Krawtchouk matrices from classical and quantum walks. In *Algebraic Methods in Statistics and Probability*, volume 287 of *Contemporary Mathematics*, pp. 83–96. AMS, 2001. See also at arXiv. [doi:10.1090/conm/287] 60

[12] ADRIANO GARSIA: A simple proof of E. Hopf's maximal ergodic theorem. *Indiana Univ. Math. J. (J. Math. Mech.)*, 14(3):381–382, 1965. [doi:10.1512/iumj.1965.14.14027] 70

[13] ANUPAM GUPTA AND KUNAL TALWAR: Approximating Unique Games. In *Proc. 17th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'06)*, pp. 99–106. ACM Press, 2006. [ACM:1109569] 57

[14] SUBHASH KHOT: On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pp. 767–775. ACM Press, 2002. [doi:10.1145/509907.510017] 57

[15] SUBHASH KHOT, GUY KINDLER, ELCHANAN MOSSEL, AND RYAN O'DONNELL: Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007. Preliminary version in FOCS'04. See also at ECCC. [doi:10.1137/S0097539705447372] 57

[16] SUBHASH KHOT AND ODED REGEV: Vertex cover might be hard to approximate to within $2-\varepsilon$. *J. Comput. System Sci.*, 74(3):335–349, 2008. Preliminary version in CCC'03. [doi:10.1016/j.jcss.2007.06.019] 57

[17] SUBHASH KHOT AND NISHEETH K. VISHNOI: The Unique Games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proc. 46th FOCS*, pp. 53–62. IEEE Comp. Soc. Press, 2005. [doi:10.1109/SFCS.2005.74] 57

[18] BO'AZ KLARTAG AND ODED REGEV: Quantum one-way communication can be exponentially stronger than classical communication. In *Proc. 43rd STOC*, pp. 31–40. ACM Press, 2011. See also at ECCC and arXiv. [doi:10.1145/1993636.1993642] 60

[19] ALEXANDRA KOLLA: Spectral algorithms for Unique Games. *Comput. Complexity*, 20(2):177–206, 2011. Preliminary version in CCC'10. See also at ECCC. [doi:10.1007/s00037-011-0011-7] 57

[20] ALEXANDRA KOLLA, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: How to play Unique Games against a semi-random adversary: Study of semi-random models of Unique Games. In *Proc. 52nd FOCS*, pp. 443–452. IEEE Comp. Soc. Press, 2011. See also at arXiv. [doi:10.1109/FOCS.2011.78] 57

[21] ILIA KRASIKOV: Nonnegative quadratic forms and bounds on orthogonal polynomials. *Journal of Approximation Theory*, 111(1):31–49, 2001. [doi:10.1006/jath.2001.3570] 60

[22] BEN KRAUSE: Dimension-free maximal inequalities for spherical means in the hypercube. Technical report, 2013. [arXiv:1309.4466] 59

[23] ULRICH KRENGEL: *Ergodic Theorems*. Walter de Gruyter, 1985. 59

[24] VLADIMIR I. LEVENSHTEIN: Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces. *IEEE Trans. Inform. Theory*, 41(5):1303–1321, 1995. [doi:10.1109/18.412678] 60

[25] NATHAN LINIAL AND AVNER MAGEN: Least-distortion Euclidean embeddings of graphs: Products of cycles and expanders. *J. Combin. Theory Ser. B*, 79(2):157–171, 2000. [doi:10.1006/jctb.2000.1953] 56

[26] KONSTANTIN MAKARYCHEV AND YURY MAKARYCHEV: How to play Unique Games on expanders. In *Proc. 8th Internat. Workshop on Approximation and Online Algorithms (WAOA'10)*, pp. 190–200, 2010. See also at ECCC. [doi:10.1007/978-3-642-18318-8_17] 57

[27] ASSAF NAOR AND TERENCE TAO: Random martingales and localization of maximal inequalities. *Journal of Functional Analysis*, 259(3):731–779, 2010. See also at arXiv. [doi:10.1016/j.jfa.2009.12.009] 56

[28] AMOS NEVO AND ELIAS M. STEIN: A generalization of Birkhoff's pointwise ergodic theorem. *Acta Mathematica*, 173(1):135–154, 1994. [doi:10.1007/BF02392571] 56

[29] PRASAD RAGHAVENDRA: Optimal algorithms and inapproximability results for every CSP? In *Proc. 40th STOC*, pp. 245–254. ACM Press, 2008. [doi:10.1145/1374376.1374414] 57

[30] PRASAD RAGHAVENDRA AND DAVID STEURER: Graph expansion and the Unique Games conjecture. In *Proc. 42nd STOC*, pp. 755–764. ACM Press, 2010. [doi:10.1145/1806689.1806792] 57

[31] ELIAS M. STEIN: On the maximal ergodic theorem. *Proc. Natl. Acad. Sci. USA*, 47(12):1894–1897, 1961. PNAS. 59

[32] LUCA TREVISAN: Approximation algorithms for Unique Games. *Theory of Computing*, 4(5):111–128, 2008. Preliminary version in FOCS'05. See also at ECCC. [doi:10.4086/toc.2008.v004a005] 57

[33] ANTONI ZYGMUND: On a theorem of Marcinkiewicz concerning interpolation of operations. *J. Math. Pures Appl.*, 9(35):223–248, 1956. 71

AUTHORS

Aram W. Harrow
Assistant Professor
Massachusetts Institute of Technology
aram@mit.edu
http://web.mit.edu/aram/

Alexandra Kolla
Assistant Professor
University of Illinois at Urbana-Champaign
akolla@illinois.edu
http://www.cs.illinois.edu/homes/akolla/

Leonard J. Schulman
Professor
California Institute of Technology
schulman@caltech.edu
http://users.cms.caltech.edu/~schulman/

ABOUT THE AUTHORS

ARAM W. HARROW graduated from M.I.T. in 2005 with a Ph. D. in physics; his advisor was Isaac Chuang. He has since worked at the University of Bristol for five years as a lecturer in computer science and math, and at the University of Washington as a research professor in computer science, before returning to MIT as an assistant professor of physics. His research focuses mainly on quantum computing, quantum information and connections between these fields and other areas of math, physics and computer science.

ALEXANDRA KOLLA received her Ph. D. in Computer Science from U.C. Berkeley under the supervision of Umesh Vazirani. She was a postdoc at The Institute for Advanced Study and Microsoft Research in Redmond. Since 2012, she has been on the faculty of the University of Illinois at Urbana Champaign. She is a fellow at the Simons Institute in Berkeley for the academic year 2013-2014. Her research focuses mostly on spectral graph theory, approximation algorithms, complexity theory, convex programming, and quantum computing.

LEONARD J. SCHULMAN received the B. S. in Mathematics in 1988 and the Ph. D. in Applied Mathematics in 1992, both from the Massachusetts Institute of Technology. Since 2000 he has been on the faculty of the California Institute of Technology. He has also held appointments at UC Berkeley, the Weizmann Institute of Science, the Georgia Institute of Technology, and the Mathematical Sciences Research Institute. He has received the MIT Bucsela prize in mathematics, an NSF mathematical sciences postdoctoral fellowship, an NSF CAREER award, and the IEEE Schelkunoff prize. He is the director of the Caltech Center for the Mathematics of Information, is on the faculty of the Institute for Quantum Information and Matter, and serves as Editor-in-Chief of the SIAM Journal on Computing. His research is in several overlapping areas: algorithms and communication protocols; combinatorics and probability; coding and information theory; quantum computation.