Dimension-free $L_2$ Maximal Inequality for Spherical Means in the Hypercube

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

May 2, 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 $\varepsilon$ 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 $x$.

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 $\pi_x$ denote the evaluation map from $V \rightarrow \mathbb{R}$ defined by $\pi_x f = f(x)$, for $f \in V$. If $\mathcal{A} \subseteq \text{Hom}(V, V)$, (the vector space consisting of all linear mappings from $V$ into itself) the maximal operator $M_\mathcal{A} : V \rightarrow V$ is the sublinear operator in which $M_\mathcal{A}f$ is defined by

$$\pi_x M_\mathcal{A}f = \sup_{A \in \mathcal{A}} \pi_x Af$$

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

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

Applying $M$, we have the spherical maximal operator $M_S : V \rightarrow V$ defined by

$$\pi_x M_S f = \max_{0 \leq k \leq n} \pi_x S_k f$$

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

**Theorem 1.** There is a constant $A_1$ such that for all $n$, $\|M_S\|_{2 \rightarrow 2} < A_1$.

Equivalently, for all $n$ and $f$, $\|M_S f\|_2 \leq A_1 \|f\|_2$.

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 [26] for a good review. Most previous work does not explicitly consider finite

---

*Massachusetts Institute of Technology
†University of Illinois, Urbana-Champaign
‡California Institute of Technology

This notation choice is because our paper is replete with operators acting on functions, and the associative composition $\pi_x Af$ is preferable to the cumbersome $(Af)(x)$. 
or discrete metric spaces; however, see [27] 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 [9, 23], 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 $\{ x : \sum x_i = 0, \sum |x_i| \leq 2N \}$. Then $\|f\|_2^2 = 2N + 1$ while $\|M_S f\|_2^2 \in O(N^2)$. 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 $\mathbb{R}^n$, Theorem 1 has the following consequence: For nonnegative $\varepsilon$ less than some $\varepsilon_0$ and for all $n$, if $|F| < \varepsilon 2^n$ then there exists $x \in \mathbb{R}^n$ 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 has the following consequence: For nonnegative $\varepsilon$ less than some $\varepsilon_1$ 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 edges which lie in $F'$ is $O(\sqrt{\varepsilon})$.

(Define a function $f$ on vertices by $f(y) = \frac{|\{e \in F' : y \in e\}|}{|F'|}$ that lie in $F'$ is bounded for $k \leq n/2$ by $2\pi^2 S_k f$, and for $k > n/2$ by $2\pi^2 S_{k+1} f$.)

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 [13], VERTEX COVER [15], MAXIMUM CUT [14] and non-uniform SPARSEST CUT [6, 13], 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 in [4] and [28] 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 $(u, v) \in E$, 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 [13] 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 2. (UGC)** For any constants \(\epsilon, \delta > 0\), there is a \(k(\epsilon, \delta)\) such that for any \(k > k(\epsilon, \delta)\), it is NP-hard to distinguish between instances of Unique Games with alphabet size \(k\) where at least \(1 - \epsilon\) fraction of constraints are satisfiable and those where at most \(\delta\) fraction of constraints are satisfiable.

Numerous works have focused on designing approximation algorithms for Unique Games. While earlier papers [13, 32, 12, 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, 24], local expanders [2, 29], and more generally, graphs with a few large eigenvalues [17]. In [18], 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 \(1 - \epsilon\) satisfiable instance to think of it as originating from a completely satisfiable instance, where a malicious adversary picks an \(\epsilon\) 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 \(\epsilon\) 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 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 \(n\)th 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 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 $I^n$ begs the question whether there is a natural limit object $T$, such that for every $I^n$ there exists a single morphism $I^n \to T$ 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 $\text{Sen}$. 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., $\mathcal{S} = \{S_k\}_0^a$ as above.) Then the family $\text{Sen}(\mathcal{T}) = \{\text{Sen}(T)_k\}$, indexed by $k$ in the same range, consists of the operators

$$\text{Sen}(T)_k = \frac{1}{k + 1} \sum_{t = 0}^k T_t \quad \text{or} \quad \text{Sen}(T)_k = \frac{1}{k} \int_0^k T_t \, dt$$

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 3.** $\|M_{\mathcal{S}}\|_{2 \to 2} \in O(1 + \|M_{\text{Sen}(\mathcal{S})}\|_{2 \to 2})$.

Bounds on the Krawtchouk polynomials play a key role in this argument. We will introduce the polynomials and prove these bounds in Sec. 2, and then use them to prove Proposition 3 in Sec. 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 $p = (1 - e^{-t})/2$, we set

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

This has the following interpretation. $\pi_y N_t f$ 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 $i$th bit of $x$ as many times as there are events in the $i$th Poisson process. The $N_t$’s 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 Sec. 4) by a direct pointwise comparison that:

**Proposition 4.** $\|M_{\text{Sen}(\mathcal{S})}\|_{2 \to 2} \in O(\|M_{\text{Sen}(\mathcal{N})}\|_{2 \to 2})$.

Finally (see $[21]$) $\|M_{\text{Sen}(\mathcal{N})}\|_{2 \to 2} \leq 2\sqrt{2}$ (indeed, $\|M_{\text{Sen}(\mathcal{N})}\|_{p \to p} \leq (p/(p - 1))^{1/p}$ 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 by

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

**Remark 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 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 $[20]$. On the other hand, $\|M_S\|_{1 \to 1} = n + 1$, as can be seen by taking $f$ to be nonzero only on a single point.

---

$\text{Sen}$ is the terminology is to express that (as in the United States Senate) each block has equal weight without regard to size.
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 for nonnegative $f$; this simplifies some expressions and will be assumed throughout.

## 2 Fourier analysis and Krawtchouk polynomials

For $y \in \mathbb{I}^n$, define the character $\chi_y \in \mathbb{C}^n$ by $\pi_x \chi_y = (-1)^{x \cdot y} \sqrt{2^n}$. The normalization is chosen so that the $\chi_y$ form an orthonormal basis of $\mathbb{C}^n$. This basis simultaneously diagonalizes each $S_k$, as they commute with $\mathbb{I}^n$ as an abelian group. A direct calculation (see also [10]) shows that

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

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

$$\kappa_k^{(n)}(x) = \sum_{j=0}^{k} (-1)^j \binom{x}{j} \binom{n-x}{k-j} \binom{n}{k}^{-1}$$

We collect here some facts about Krawtchouk polynomials.

**Lemma 6.**

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:

$$\sum_{x=0}^{n} \kappa_k^{(n)}(x) \kappa_{\ell}^{(n)}(x) \binom{n}{x} \binom{n}{x} = \frac{2^n}{\binom{n}{k}} \delta_{k,\ell}.$$  

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 [22]), and the fourth claim is a weaker version of Theorem 8 of [19]. 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 Sec. 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 $\chi_y$ basis, and for $|y| = x$, their eigenvalue for $\chi_y$ is $(1-2t)^x$. Thus, our goal in this section will be to show that $\kappa_k(x)$ has similar behavior to $(1-2k/n)^x$. More precisely, we prove

**Lemma 7.** There is a constant $c > 0$ such that for all $n$ and integer $0 \leq x, k \leq n/2$,

$$|\kappa_k^{(n)}(x)| \leq e^{-c k x/n}.$$  

Due to Lemma 6 it suffices to bound $\kappa_k(x)$ only when $0 \leq k \leq x \leq n/2$.

---

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 [20] 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, [5] 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 6.3.

Setting $k = \ell$ in Lemma 6.3 and observing that all of the terms on the LHS are nonnegative, it follows that

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

(7)

Based on Stirling’s formula, Lemma 17.5.1 of [25] states that

$$
\left(\frac{n}{np}\right) \geq \frac{e^{nH_2(p)}}{\sqrt{8p(1-p)n}} \geq \frac{e^{nH_2(p)}}{\sqrt{2n}}
$$

(8)

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

Note that $H_2(0.14) > \ln(2)/2$, so, for $k, x \geq 0.14n$, (7) implies that $\kappa_k(x)^2 \leq 2n e^{(\ln(2) - 2H_2(0.14))n} \leq 2ne^{-c_1n}$ 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 $n \geq n_0$, $2ne^{-c_1n} \leq e^{-c_2n/2}$. So for all $n \geq n_0$ and all $0.14n \leq k, x \leq n/2$, $\kappa_k(x)^2 \leq e^{-2c_1(n/2)(n/2)/n} \leq e^{-2c_1kn/x}$.

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

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

It is immediate from Definition 4 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 WLOG that $k \leq 0.14n$.

It is convenient to make the change of variable

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

set

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

and expand $\mu_k(z) = \sum_{i=0}^k \alpha_{k,i}z^i$. $\mu_k$ 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 $y_1, \ldots, y_k$ be the roots of $\mu_k$. By Lemma 6.2 the multiset $\{y_1, \ldots, y_k\}$ is identical to the multiset $\{-y_1, \ldots, -y_k\}$. So we can write

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

(9)

It is immediate from Definition 4 that

$$
\mu_k^2(1) = 1
$$

(10)

Furthermore, by Lemma 6.3

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

(11)
We now obtain an upper bound on \( \mu_k^2(z) \) simply by maximizing (10) subject to the constraints (10) and (11). 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

\[
\frac{\partial}{\partial y_i^2} \frac{z^2 - y_i^2}{1 - y_i^2} = \frac{z^2 - 1}{(1 - y_i^2)^2} \leq 0.
\]

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 \( y_i = y_{\text{max}} \). In the former case, \( |z^2 - y_i^2|/(1 - y_i^2) = z^2 \). In the latter case,

\[
\frac{|z^2 - y_i^2|}{1 - y_i^2} \leq \frac{y_{\text{max}}^2 - z^2}{1 - y_{\text{max}}^2} \leq \frac{y_{\text{max}}^2}{1 - y_{\text{max}}^2} \leq 0.93.
\]

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

\[
\mu_k^2(z) \leq (\max(z^2, 0.93))^k \tag{13}
\]

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

If \( z^2 \leq 0.93 \) then (recalling \( x \leq n/2 \)) we have \( \kappa_k^2(x) \leq (0.93)^k \leq (0.93)^{2xk/n} = e^{-cxk/n} \) 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 3**

**Proposition 3** (restatement). \( \|M_S\|_{2 \rightarrow 2} \in O(1 + \|M_{\text{Sen}}(S)\|_{2 \rightarrow 2}) \).

We start by defining \( S = \{S_k\}_{0}^{\lfloor n/2 \rfloor} \). The operator \( M_S : V \rightarrow V \) is then defined by

\[
\pi_x M_S f = \max_{0 \leq k \leq \lfloor n/2 \rfloor} \pi_x S_k f.
\]

Define \( \iota := S_n \) to be the antipodal operator on \( \mathbb{R}^n \), i.e., the involution obtained by flipping all \( n \) bits of the argument of the function. Then \( \pi_x M_S f = \max\{\pi_x M_S f, \pi_x M_{\iota S} f\} \). So \( \|M_S\|_{2 \rightarrow 2} \leq \sqrt{2}\|M_{\iota S}\|_{2 \rightarrow 2} \).

**Claim 8.** There is a \( C < \infty \) such that for all \( n \) and \( f \), \( \|M_{\iota S} f\|_2 \leq C\|f\|_2 + \|M_{\text{Sen}}(S) f\|_2 \).

We prove the claim in the following two subsections.

3.1 **A method of Stein**

The bounds on \( \|M_{\iota S}\|_{2 \rightarrow 2} \) for even and odd radius \( \ell \) are technically distinct (but not in any interesting way). We present the arguments in parallel.
3.1.1 Even radius: $S_{2r}$ for $0 \leq r \leq r_{\text{max}} = \lfloor \lfloor n/2 \rfloor/2 \rfloor$

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

Abel’s lemma gives the following easily verified identity:

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

(14)

Hence we have the following pointwise (that is to say, valid at each point $x$) inequality for $0 \leq r \leq r_{\text{max}}$:

$$\pi_x(S_{2r} - \frac{1}{r+1} \sum_{k=0}^{r} S_{2k})f|^2 = \left[\sum_{k=1}^{r} \frac{\sqrt{k}}{r+1} \cdot \sqrt{\sum_{k=1}^{r} k[\pi_x(S_{2k} - S_{2(k-1)})f|^2} \leq \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

\leq \frac{1}{2} \sum_{k=1}^{r_{\text{max}}} k[\pi_x(S_{2k} - S_{2(k-1)})f|^2

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

$$\pi_x R_0 f = \left[\frac{1}{2} \sum_{k=1}^{r_{\text{max}}} k[\pi_x(S_{2k} - S_{2(k-1)})f|^2\right]^{1/2}$$

(15)

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

$$\|(S_{2r} - \text{Sen}(S)_{2r})f\|^2 \leq \|R_0 f\|^2$$

(16)

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

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

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

Abel’s lemma gives:

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

(17)

Hence we have the following pointwise inequality for $0 \leq r \leq r_{\text{max}} = \lfloor \lfloor n/2 \rfloor - 1 \rfloor/2 \rfloor$:
3.2 Bounding the error term

Define \( R_1 : V \to V \) by

\[
\pi_x R_1 f = \sqrt{\frac{1}{2} \sum_{k=1}^{\max} k[\pi_x(S_{2k+1} - S_{2k-1})f]^2}
\] (18)

so that for any \( r \leq r_{\text{max}} \),

\[
\|(S_{2r+1} - \text{Sen}(S)_{2r+1})f\|_2^2 \leq \|R_1 f\|_2^2
\] (19)

This suggests defining an “error term” operator \( R_1 : V \to V \) by

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

so that for any \( r \leq r_{\text{max}} \),

\[
\|(S_{2r+1} - \text{Sen}(S)_{2r+1})f\|_2^2 \leq \|R_1 f\|_2^2
\]

3.2 Bounding the error term

Define \( Rf \) by \( \pi_x Rf := \max\{\pi_x R_0 f, \pi_x R_1 f\} \). Combining (16) and (19) we have for each \( x \in \mathbb{F}^n \), each \( f \in \mathbb{R}^n \) and each \( r \leq n/2 \),

\[
|\pi_x(S_r - \text{Sen}(S)_r)f| \leq \pi_x Rf.
\]

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

\[
\|M_{\mathcal{F}} f - M_{\text{Sen}(\mathcal{F})} f\|_2 \leq \|Rf\|_2.
\]

Claim 8 (and Proposition 3) now follow from:

**Lemma 9.** There is a \( C < \infty \) such that \( \|R_0 f\|_2, \|R_1 f\|_2 \leq C\|f\|_2 \).

**Proof.** As seen in the preliminaries, the operators \( S_k \) commute and share the eigenvectors \( \{\chi_y\}_{y \in \mathbb{F}^n} \), with eigenvalues given by Eqn. (3): \( S_k \chi_y = \kappa^{(n)}_k (|y|) \chi_y \). Let \( E_x \) be the projection operator on the \( \binom{n}{|y|} \)-dimensional eigenspace spanned by \( \{\chi_y\}_{|y| = x} \); so \( S_k = \sum_{z=0}^n \kappa^{(n)}_k (x) E_z \). We calculate:

\[
\|R_0 f\|_2^2 = \sum_{z \in \mathbb{F}^n} \frac{1}{2} \sum_{k=1}^{\max} k[\|(S_{2k} - S_{2(k-1)})f(z)\|^2
\]

\[
= \frac{1}{2} \sum_{k=1}^{\max} \sum_{x=0}^n ||(\kappa^{(n)}_{2k}(x) - \kappa^{(n)}_{2(k-1)}(x))E_x f||_2^2
\]

\[
= \frac{1}{2} \sum_{x=0}^n ||E_x f||_2^2 \sum_{k=1}^{\max} k(\kappa^{(n)}_{2k}(x) - \kappa^{(n)}_{2(k-1)}(x))^2
\]

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

\[
\|R_1 f\|_2^2 = \sum_{z \in \mathbb{F}^n} \frac{1}{2} \sum_{k=1}^{\max} k[\|(S_{2k+1} - S_{2k-1})f(z)\|^2
\]

\[
= \frac{1}{2} \sum_{x=0}^n ||E_x f||_2^2 \sum_{k=1}^{\max} k(\kappa^{(n)}_{2k+1}(x) - \kappa^{(n)}_{2k-1}(x))^2
\]
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 \( 0 \leq x \leq n \),

\[
\left[ \sum_{k=1}^{\text{r}_{\max}} k\left( \kappa_{2k}^{(n)}(x) - \kappa_{2k-1}^{(n)}(x) \right)^2 \right] \leq C \tag{20a}
\]

\[
\left[ \sum_{k=1}^{\text{r}_{\max}} k\left( \kappa_{2k+1}^{(n)}(x) - \kappa_{2k}^{(n)}(x) \right)^2 \right] \leq C \tag{20b}
\]

Recall that it suffices by Lemma 6.2 to consider \( x \leq n/2 \). For \( x = 0 \), (20) is trivial as the LHS is 0.

For \( x > 0 \) we use Lemma 6.1 to rewrite the parenthesized term (with \( \ell = 2k \) or \( \ell = 2k + 1 \)) as follows:

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

To see this, recall that \( \binom{n}{\ell} \kappa_x^{(n)}(\ell) \) 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 \( \ell \).

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

\[
= \frac{4x(n-2)}{n(n-1)} \kappa_{x-1}^{(n-2)}(\ell - 2) = \frac{4x}{n} \left( \frac{n-x}{n-1} \right) \kappa_{x-1}^{(n-2)}(\ell - 2) \tag{21}
\]

The two terms on the LHS of (20) are now

\[
\sum_{k=1}^{\text{r}_{\max}} k\left( \kappa_{x}^{(n)}(2k) - \kappa_{x}^{(n)}(2k-2) \right)^2 = \sum_{k=1}^{\text{r}_{\max}} k\left( \frac{4x(n-x)}{n(n-1)} \kappa_{x-1}^{(n-2)}(2k-2) \right)^2 + \sum_{k=1}^{\text{r}_{\max}} k\left( \kappa_{x}^{(n)}(2k+1) - \kappa_{x}^{(n)}(2k-1) \right)^2
\]

\[
= \sum_{k=1}^{\text{r}_{\max}} k\left( \frac{4x(n-x)}{n(n-1)} \kappa_{x-1}^{(n-2)}(2k-1) \right)^2 \tag{22}
\]

For \( x = 1 \), quantities (22) come to \( \sum_{k=1}^{r_{\max}} 16k n^{-2} \) which is upper bounded by a constant.

For \( x > 1 \) we upper bound (22), 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 \( r_{\max} \)):

\[
\frac{16x^2}{n^2} \sum_{k=0}^{n/2} (k/2 + 1) \left( \kappa_{x-1}^{(n-2)}(k) \right)^2.
\]
Now apply Lemma 7 to upper bound this by
\[
\frac{16x^2}{n^2} \sum_{k=0}^{\infty} \frac{(k/2 + 1)e^{-2c(x-1)k/(n-2)}}{(k/2 + 1)e^{-\alpha k}}
\]
\[
= \frac{16x^2}{n^2} \sum_{k=0}^{\infty} \frac{(k/2 + 1)e^{-\alpha k}}{(1 - e^{-\alpha})^2}
\]
\[
\leq 24 \left( \frac{x}{n(1 - e^{-\alpha})} \right)^2
\]
\[
\leq 24 \left( \frac{(n-2)x}{2c(n(x-1))} \right)^2
\]
\[
\leq 24/e^2,
\]
where in (23) we have defined \( \alpha = \frac{2c(x-1)}{n-2} \) and in (24) we have used the identity \( \sum_{k=0}^{\infty} ke^{-\alpha k} = e^{-\alpha} \left( 1 - e^{-\alpha} \right)^{-2} \).

This completes the requirement of Eqn. 20.

4 Comparing senate maximal functions: proof of Proposition 4

Proposition 4 (restatement).
\[
\|M_{\text{Sen} \left( S \right)}\|_2 \to_2 \in O(\|M_{\text{Sen} \left( N \right)}\|_2 \to_2).
\]

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

If \( A, B \) are matrices, write \( A \leq B \) if \( B - A \) is a nonnegative matrix. If \( A, B \) are sets of nonnegative matrices indexed by integers or reals, we write \( A \lesssim B \) if for every \( A \in A \) there is a probability measure \( \mu_A \) on \( B \) such that \( A \leq \int B \, d\mu_A(B) \). Observe that in this case for any nonnegative function \( f \) and any \( x \), \( \sup_{A \in A} \pi_x Af \leq \sup_{B \in B} \pi_x Bf \), and therefore for any norm, \( \|M_A\| \leq \|M_B\| \) (and in particular for \( \| \cdot \|_2 \to_2 \)).

For any \( k > \lceil n/2 \rceil \),
\[
\pi_x \text{Sen}(S)_{k} f \leq \pi_x \text{Sen}(S)_{\lfloor n/2 \rfloor} (f + \epsilon f)
\]
Therefore \( \|M_{\text{Sen}(S)}\|_2 \to_2 \leq 2\|M_{\text{Sen}(\bar{S})}\|_2 \to_2 \). However, we will not compare \( \text{Sen}(\bar{S}) \) and \( \text{Sen}(\bar{N}) \) directly. Instead, we will introduce a variant of \( \bar{N} \) that more closely resembles \( \bar{S} \), but is no longer a semigroup.

Recall that \( N_t \) represents the average over independently flipping each bit with probability \( p = (1 - e^{-t})/2 \). Define \( \bar{N}_p \) to represent the same noise process but parameterized by \( p \) instead of \( t \). Thus
\[
N_t = \bar{N}_{1-e^{-t}} \quad \text{and} \quad \bar{N}_p = N_{-\ln(1-2p)}.
\]

While the sets \( \{N_t\}_{t \geq 0} \) and \( \{\bar{N}_p\}_{p \in [0,1/2]} \) are of course the same, their Senate operators \( \text{Sen}(\bar{N}) \) and \( \text{Sen}(\bar{N}) \) are different:
\[
\text{Sen}(\bar{N})_T = \frac{1}{T} \int_0^T N_t \, dt \tag{25}
\]
\[
\text{Sen}(\bar{N})_p = \frac{1}{P} \int_0^P \bar{N}_p \, dp = \frac{1}{2P} \int_0^{-\ln(1-2P)} e^{-t} N_t \, dt \tag{26}
\]
In (26), $P$ is varies over $(0, 1/2)$ and in (28), $T$ can be any positive real number.

Hence Proposition 3 is established in two subsidiary claims:

**Lemma 10.** $\text{Sen}(\tilde{N}) \leq \text{Sen}(N)$.

**Lemma 11.** $\text{Sen}(\tilde{S}) \leq C \cdot \text{Sen}(N)$ for some constant $C > 0$.

**Proof of Lemma 11.** For each $P$ we will write $\text{Sen}(\tilde{N})_P$ as a convex combination of $\text{Sen}(N)_T$ for different values of $T$. By considering the action of each side on the constant function we will see that it suffices to write $\text{Sen}(\tilde{N})_P$ as a bounded nonnegative combination of $\text{Sen}(N)_T$ for different values of $T$ (i.e. a linear combination with coefficients that are nonnegative and whose sum is bounded).

Let $f(t) := e^{-t}/2P$ and $\tau := -\ln(1 - 2P)$ so that (26) becomes

$$\text{Sen}(\tilde{N})_P = \int_0^\tau dt f(t)N_t = \int_0^\tau dt f(\tau)N_t + \int_0^\tau dt f(t) - f(\tau))N_t$$

(27)

$$= \tau f(\tau) \text{Sen}(N)_\tau + \int_0^\tau dt \left( \int_t^\tau dT(-f'(T)) \right) N_t$$

(28)

$$= \tau f(\tau) \text{Sen}(N)_\tau + \int_0^\tau dT \int_0^T dt (-f'(T))N_t$$

(29)

$$= \tau f(\tau) \text{Sen}(N)_\tau + \int_0^\tau dT (-Tf'(T)) \text{Sen}(N)_T$$

(30)

Since $f'(T) < 0$, (31) is the desired nonnegative combination. We have written the proof in this way to stress that the only features of (26) and (28) used are that $\text{Sen}(N)_T$ is a “flat” distribution over $N_t$ and $\text{Sen}(\tilde{N})_P$ has $N_t$ weighted by a weakly decreasing function.

**Proof of Lemma 12** To compare $\text{Sen}(\tilde{S})$ and $\text{Sen}(\tilde{N})$, we need to show that for any $K \leq n/2$, we can find a distribution over $P$ such that $\text{Sen}(\tilde{S})_K$ is pointwise $\leq$ the appropriate average over $\text{Sen}(\tilde{N})_P$ times a constant. In fact, it will suffice to consider a distribution that is concentrated on a single value of $P$. Define $P_K := \min(\frac{K+\sqrt{K}}{n}, \frac{1}{2})$. In Lemma 12 we will show that $\text{Sen}(S)_K \leq C \cdot \text{Sen}(\tilde{N})_P$, thus implying that $\text{Sen}(\tilde{S}) \leq C \cdot \text{Sen}(N)$. The idea behind Lemma 12 is that for each $k \leq K$, there are significant contributions to the $S_k$ coefficient of $\text{Sen}(\tilde{N})_P$ 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 $1/K$. The total contribution is thus $\Omega(1/K)$.

**Lemma 12.** Let $n \geq 9$, $K \leq n/2$ and $P_K = \min(\frac{K+\sqrt{K}}{n}, \frac{1}{2})$. Then $\text{Sen}(S)_K \leq 3e^{20} \cdot \text{Sen}(\tilde{N})_P$.

The true constant is certainly much better, and perhaps closer to $1/2$. We remark that $n \geq 9$ can be assumed WLOG, since $\|M_S\|_{2 \to 2} \leq n + 1$ by the triangle inequality.

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

For $0 \leq k \leq K$, we now compare the coefficient of $S_k$ in $\text{Sen}(\tilde{S})_K$ (where it is $1/(K + 1)$) to its value in $\text{Sen}(\tilde{N})_P$, where it is $\frac{1}{P} \int_0^p dp B(n, p, k)$. 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

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

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

12
Next, suppose \( k + \sqrt{k} \leq \frac{n}{2} \). Then \( P \geq \frac{k + \sqrt{k}}{n} \). 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

\[
B(n, \frac{n}{k}, k) \geq \frac{1}{3\sqrt{k}}. \tag{32}
\]

Next, observe that

\[
\frac{d}{dq} \ln B(n, q, k) = \frac{k - nq}{q(1-q)}. \tag{33}
\]

When \( k/n \leq q \leq (k + \sqrt{k})/n \), we have \( k - nq \geq k - np \geq -\sqrt{k} \) and \( q(1-q) \geq \frac{k}{n} \). Thus, \( 0 \leq p - k/n \leq \sqrt{k}/n \) implies that

\[
\ln \frac{B(n, p, k)}{B(n, \frac{n}{k}, k)} = \int_{k/n}^{p} \frac{k - nq}{q(1-q)} \geq \left( p - \frac{k}{n} \right) \frac{k - np}{k/2n} \geq -2. \tag{34a}
\]

Combining (32) and (34), we find that \( B(n, p, k) \geq 1/6e^2\sqrt{k} \) for \( k/n \leq p \leq (k + \sqrt{k})/n \). Thus

\[
a_k \geq \frac{1}{P} \int_{k/n}^{(k+\sqrt{k})/n} dp \ B(n, p, k) \geq \frac{n}{2K} \frac{\sqrt{k}}{n} \frac{1}{6e^2\sqrt{k}} \geq \frac{1}{12e^2K}. \tag{32}
\]

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 (32) to bound

\[
\ln \frac{B(n, \frac{1}{2} - \frac{1}{\sqrt{n}}, k)}{B(n, \frac{n}{k}, k)} \geq - \left( \frac{k}{n} - \frac{1}{2} \right) \frac{k - \sqrt{n}}{1/2} \geq -12 \left( 1 - \frac{1}{\sqrt{2}} \right)^2 \geq -20.
\]

Similarly

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

We conclude that

\[
a_k \geq \frac{1}{P} \int_{\frac{1}{2} - \frac{1}{\sqrt{n}}}^{\frac{1}{2}} dp \ B(n, p, k) \geq 2 \cdot \frac{1}{\sqrt{n}} \frac{1}{3e^{20} \sqrt{k}} \geq \frac{1}{3e^{20} K}. \tag{32}
\]

This completes the proof of Lemma 11.
A Maximal inequalities on semigroups

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

**Theorem 13** (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 \( A \) denote the (discrete) semigroup generated by \( A \). Then

\[
\| M_{\text{Sen}(A)} \|_{1 \to 1,w} \leq 1.
\]

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 \([11]\) is simple and self-contained, we restate it here.

**Proof.** For \( f \) an arbitrary function and \( T \) a nonnegative integer, define \( E_T^f = I[M_{\text{Sen}(A)}]f \geq 0 \). We will first prove that

\[
\langle E_T^f, f \rangle \geq 0,
\]

for all \( T \). To see that the theorem follows from this claim, apply \([35]\) to \( f - \lambda \) and we find that

\[
\lambda \| E_T^{f-\lambda} \|_1 \leq \langle E_T^{f-\lambda}, f \rangle \leq \| f \|_1.
\]

Thus \( \| E_T^{f-\lambda} \|_1 \leq \lambda^{-1} \| f \|_1 \). Since this inequality holds for all \( T \), it implies that \( \| I[M_{\text{Sen}(A)}]f \geq \lambda \|_1 \) is also \( \leq \lambda^{-1} \| f \|_1 \).

We now return to the proof of \((35)\). We define for this purpose an unweighted Senate operator:

\[
\text{Sen}(A)_T = \sum_{t=0}^T A_t.
\]

We abbreviate \( M_T := M_{\text{Sen}(A)_{\leq T}} \) and \( \bar{M}_T := M_{\text{Sen}(A)_{\leq T}} \). 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 \( \text{Sen}(A)_t f \leq M_T f \leq (\bar{M}_T f)^+ \). This implies that

\[
f + A(\bar{M}_T f)^+ \geq f + A\text{Sen}(A)_t f = \text{Sen}(A)_{t+1} f.
\]

Thus, \( f \geq \text{Sen}(A)_t f - A((\bar{M}_T f)^+) \), (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 \( E_T^f \) and find

\[
\langle E_T^f, f \rangle \geq \langle E_T^f, \bar{M}_T f - A((\bar{M}_T f)^+) \rangle = \langle E_T^f, (\bar{M}_T f)^+ - A((\bar{M}_T f)^+) \rangle = \| (\bar{M}_T f)^+ \|_1 - \langle E_T^f, A((\bar{M}_T f)^+) \rangle \geq \| (\bar{M}_T f)^+ \|_1 - \| A((\bar{M}_T f)^+) \|_1 \geq 0.
\]

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 14.** 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 \( 1 \leq p < r < q < \infty \), 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 $M_A$ has $\|M_A\|_{\infty \to \infty, w} = 1$ and if $A$ is a positive contractive semigroup, then Theorem 13 implies that $\|M_A\|_{1 \to 1, w} \leq 1$ as well. This implies that for any $1 < p$, we have

$$\|M_A\|_{p \to p} \leq 2 \left(\frac{p}{p-1}\right)^{1/p},$$

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 Kostantin Makarychev and Yury Makarychev for discussions about UGC on the hypercube, to Yuval Peres for helpful comments, and to Terence Tao for suggestions along the lines of Sec. 3. AWH was funded by NSF grants 0916400, 0829937 and 0803478 and DARPA QuEST contract FA9550-09-1-0044. AK was at Microsoft Research at the beginning and during part of this work. LJS was funded in part 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.

References

[1] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. In Proceedings of the 51st IEEE Symposium on Foundations of Computer Science, 2010.

[2] S. Arora, R. Impagliazzo, W. Matthews, and D. Steurer. Improved algorithms for unique games via divide and conquer. In In Electronic Colloquium on Computational Complexity, pages TR10–041, 2010.

[3] S. Arora, S. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. Vishnoi. Unique Games on expanding constraint graphs are easy. In Proceedings of the 40th ACM Symposium On The Theory Of Computing, volume 40, pages 21–28, 2008.

[4] P. Austrin. Towards sharp inapproximability for any 2-csp. In Proceedings of the 48th IEEE Symposium on Foundations of Computer Science, volume 48, pages 307–317, 2007.

[5] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. In Proceedings of the ACM Symposium on the Theory of Computing, volume 38, pages 205–214, 2006.

[6] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In Computational Complexity, volume 15(2), pages 94–114, 2006.

[7] E. Chlamtac, K. Makarychev, and Y. Makarychev. How to play unique games using embeddings. In Proceedings of the IEEE Symposium on Foundations of Computer Science, volume 48, pages 687–696, 2006.

[8] D. Dominici. Asymptotic analysis of the Krawtchouk polynomials by the WKB method. The Ramanujan Journal, 15:303–338, 2008. arXiv:math/0501042.

[9] P. Enflo. On the nonexistence of uniform homeomorphisms between $L_p$-spaces. Ark. Mat., 8:103–105, 1969.
[10] P. Feinsilver and J. Kocik. Krawtchouk matrices from classical and quantum walks. *Contemporary Mathematics*, 287:83–96, 2002. arXiv:quant-ph/0702173.

[11] A. Garsia. A simple proof of Hopf’s maximal ergodic theorem. *J. Math. Mech.*, 14:381–382, 1965.

[12] A. Gupta and K. Talwar. Approximating unique games. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, volume 24, pages 99–106, 2006.

[13] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the ACM Symposium on the Theory of Computing*, volume 34, pages 767–775, 2002.

[14] S. Khot, G. Kindler, E. Mossel, and R. O’Donnell. Optimal inapproximability results for max-cut and other 2-variable CSPs? In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, volume 45, pages 146–154, 2004.

[15] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2 - \varepsilon$. In *Proceedings of the IEEE Conference on Computational Complexity*, volume 18, pages 379–386, 2003.

[16] S. Khot and N. K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into $\ell_1$. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, volume 46, pages 53–62, 2005.

[17] A. Kolla. Spectral algorithms for Unique Games. In *Proceedings of the 25th ACM Conference on Computational Complexity*, volume 25, pages 122–130, 2010.

[18] A. Kolla, K. Makarychev, and Y. Makarychev. How to play Unique Games against a semi-random adversary. In *Proceedings of the 43rd ACM Symposium on Foundations of Computer Science*, 2011.

[19] I. Krasikov. Nonnegative quadratic forms and bounds on orthogonal polynomials. *Journal of Approximation Theory*, 111(1):31–49, 2001.

[20] B. Krause. Dimension-free maximal inequalities for spherical means in the hypercube, 2013. arXiv:1309.4466.

[21] U. Krengel. *Ergodic Theorems*. Walter de Gruyter, 1985.

[22] V. I. Levenshtein. Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces. *IEEE Trans. Inf. Th.*, 41(5):1303–1321, Sep 1995.

[23] N. Linial and A. Magen. Least-distortion Euclidean embeddings of graphs: products of cycles and expanders. *J. Combin. Theory Ser. B*, 79(2):157–171, 2000.

[24] K. Makarychev and Y. Makarychev. How to play unique games on expanders. In *Proceedings of the eighth Workshop on Approximation and Online Algorithms*, 2010.

[25] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Series in Telecommunication. John Wiley and Sons, New York, 1991.

[26] A. Naor and T. Tao. Random martingales and localization of maximal inequalities. *Journal of Functional Analysis*, 259(3):731–779, 2010. arXiv:0912.1140.

[27] A. Nevo and E. M. Stein. A generalization of Birkhoff’s pointwise ergodic theorem. *Acta Mathematica*, 173:135–154, 1994.

[28] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th ACM Symposium On The Theory Of Computing*, volume 40, pages 245–254, 2008.
[29] P. Raghavendra and D. Steurer. Graph expansion and the Unique Games conjecture. In Proceedings of the 42nd ACM Symposium On The Theory Of Computing, volume 42, pages 755–764, 2010.

[30] O. Regev and B. Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In Proceedings of the 43rd annual ACM symposium on Theory of computing, STOC ’11, pages 31–40, 2011. arXiv:1009.3640.

[31] E. M. Stein. On the maximal ergodic theorem. Proc Natl Acad Sci U S A, 47(12):1894–1897, December 1961.

[32] L. Trevisan. Approximation algorithms for unique games. In Proceedings of the IEEE Symposium on Foundations of Computer Science, volume 46, pages 197–205, 2005.

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