# Who Will Win The US Presidential Election?

*Election and debate special from GLL*

Aram Harrow is both a computer scientist and a physicist—something that makes him unique and special. Few people can explain what a Turing Machine is and also what complementarity is. He has done some beautiful work in pure and applied mathematics, work that uses his dual abilities.

Today we wish to describe and interpret an intricate recent theorem of his in joint work with Alexandra Kolla and Leonard Schulman.

Aram is no stranger to these pages. Of course he has also been our honored participant in our year-long debate on the feasibility of building quantum computers, all during his current appointment in the Computer Science and Engineering department of the University of Washington in Seattle. To top it off he is also one of three mainstays on the prominent quantum computing blog created by Dave Bacon, and during our debate still managed to get a little bit of other work done.

Finally, in January he will join the MIT department of Physics as an assistant professor, and we congratulate him on that.

## The Paper

The paper's title, "*Dimension-free L2 maximal inequality for spherical means in the hypercube*," may not generate any 15-second sound bites. Nor may it help predict the outcome of the election, but it does contain an important new theorem.

Before we discuss it as a mathematical result, let's talk informally about it and why it is a strong theorem. The term "hypercube" is just the set $\{0,1\}^n$ of binary strings of length $n$—one of the favorite objects for computer scientists. In this world, the sphere of radius $r$ around a point $x$ is just the set of strings that are Hamming distance $r$ from $x$. Thus, if $x = 000$ and $r = 2$ the points in the sphere—sometimes we call them balls—are:

$$110, 101, 011.$$

Now imagine we select a small fraction of the points from the hypercube, say $\epsilon 2^n$ where as usual $\epsilon$ is a small positive number. The question their paper addresses and more is this: can one find a single point $x$ that is special in that every sphere around $x$ of any radius $r$ has less than half of its points in the selected set? Clearly, this seems quite reasonable, if $\epsilon$ is small enough.

This is a case where the intuition is correct, there always is such point. The difficulty is proving that it must exist. An obvious method of proving there is such a special point is to pick a random one, essentially use the probabilistic method. Alas this works for one radius or a few, but does not seem to be powerful enough to show that the point works for all spheres of all radii.

The probabilistic method is extremely powerful, but when it fails often the proof required can be quite difficult. This is the case with Aram new result. Actually in our opinion it is one of the most important aspects of the new paper. Since the probabilistic method does not always work, any new ideas that can solve problems where it fails are automatically of great importance. Here one idea that Aram tells me was important is a different kind of "complementarity," namely that $B_k(x) = B_{n-k}(x')$ where $x'$ is the vertex opposite $x$ in the hypercube.

Let's turn now to state the theorem in a mathematically precise manner and discuss its proof.

## The Result

The inequality of Aram's paper involves the $L_2$ norms of functions $f : \{0,1\}^n \longrightarrow \mathbb{R}$, defined by

$$||f||_2 = \text{ the square root of } \sum_x f(x)^2.$$

The other functions involved maximize the average of $f$ over the balls:

$$M_f(x) = \max_k \frac{1}{|B_k(x)|} \sum_{y \in B_k(x)} f(y).$$

The key clause is that the bound on the inequality does not depend on $n$. Here is their new result:

**Theorem 0.1.** *There is a constant $a > 0$ such that for all $n$ and all $k \leq n$,*

$$||M_f||_2 \leq a||f||_2$$

.

Pretty neat, no?

Well it is pretty neat, but perhaps needs some explanation, so here is a way to interpret it that may be helpful. Call $x$ "light" if $|f(x)|^2$ is below the average of the squares, i.e., $|f(x)|^2 < \frac{1}{2^n}||f||_2^2$. We might hope to "bulk $x$ up" by finding an average over one of the spheres enclosing $x$ that is greater. For many $f$'s there will be some $x$'s for which this makes a big difference, but the theorem shows that under $L_2$ norms the cumulative effect of this extra allowance will never be more than a fixed constant factor—one that is independent of the dimension.

## An Election Interpretation

It is election season in the United States again. One hears often about how demographic and personal characteristics channel one's political preferences, and that there are "typical Democratic voters" and "typical Republican voters." The categories are often binary or close to it: male/female, married/single, urban/suburban-or-rural, over age 50/under age 50, dependent/independent, college degree/not, vegetarian/carnivore, and so on.

We can define them by positions on single issues: pro-life/pro-choice, gun-rights/gun-control, for/against gay marriage, immigration amnesty, anything like that. We can also add categories that seem irrelevant: righthanded/lefthanded, dark/light eyes or hair, tall/short, nearsighted/not, Coke/Pepsi. We can even include divisions that are not roughly balanced in the population, such as thin/over-weight, and so can break multi-way categories into binary pairs, such as X/not-X for every ethnic group X. Then the possible combinations of values of $n$ characteristics correspond to the vertices of the $n$-dimensional hypercube.

What justifies saying that a combination $x$ of characteristics is "typical" for a party is that if you vary some of them—any number $k$ of them—you still find mostly people who vote for that party. Importantly this should be true for changing *any* set of $k$ characteristics, not just specific ones. Given the strength of this requirement, do typical voters exist? The fact of categories that see much "single-issue voting" may make this seem unlikely, but the following statement holds at least in some cases, and constitutes a way to approach the issues in their paper:

**Theorem 0.2.** *If a party wins by a large enough landslide, then it has typical voters.*

That is, there are voters $x$ such that not only did $x$ vote for the winner, but for any $k > 0$, if you look at all voters $y$ who differ from

$x$ in $k$ characteristics, the winner got a majority of votes from those $y$ as well. For large values of $k$ this is unsurprising—the winner won by a lot, so Markov's inequality or Chernoff bounds applied to large balls, plus an application of the union bound, should yield the existence of such $x$. The point is that it holds for *all* values of $k$, with the threshold victory margin being independent of $n$. Since smaller values of $k$ are the issue, the voters involved really do share most characteristics with $x$.

Well this comes with some caveats so we need to look closer.

## The One-Vertex, One-Vote Case

Let us first make the simplifying assumption that the population and characteristics are such that every combination has **exactly one** voter. Then we define an indicator function $f$ by:

$$f(x) = 1 \text{ if } x \text{ voted for the } \textit{loser}\text{; } f(x) = 0 \text{ otherwise.}$$

Later we will ask what happens if some points $x$ are omitted from the sample, and if multiple points are allowed, boundedly many or not. Since $f(x)^2 = f(x)$, the $L_2$-norm is just the square root of the loss margin. That is, if the loser won an $\epsilon$ portion of the vote, then

$$||f||_2 = \sqrt{\sum_x f(x)^2} = \sqrt{\epsilon} \cdot \sqrt{2^n}.$$

By the theorem applied to $f$, and squaring both sides,

$$\sum_x M_f(x)^2 \leq a^2 \epsilon \cdot 2^n.$$

Dividing by $2^n$ makes the left side an average. There must exist some $x$ giving at most the average value, so there exists some $x$ such that

$$M_f(x) = \max\{\frac{1}{|B_k(x)|} \sum_{y \in B_k(x)} f(y) \leq a\sqrt{\epsilon}.$$

Taking $\epsilon$ small enough that $a\epsilon < 1/4$ makes $\frac{1}{|B_k(x)|} \sum_{y \in B_k(x)} f(y) < \frac{1}{2}$ for *all* $k$. Thus the loser achieves no more than a minority over all of the balls centered on $x$.

This makes $x$ a typical voter. By extending the averaging argument, one can show there are many typical voters. However, things become interesting when we vary the requirements on the space. Moreover, their theorem actually fails when the $L_2$-norm is replaced by the $L_1$-norm, and all this hints at the considerable complexity involved in their proof. Let us explore some of these other possibilities.

## Weighted Measures

Now suppose we have a weighted measure $\mu(x)$ on the hypercube. With regard to this measure,

$$||f||_2^2 = \sum_x f(x)^2 \mu(x).$$

It also now makes sense to define weighted versions of the averages over the Hamming balls, leading to the operator

$$M_f^*(x) = \max_k \frac{\sum_{y \in B_k(x)} f(y)\mu(y)}{\sum_{y \in B_k(x)} \mu(y)}.$$

Does their theorem now hold in the form that there exists a universal constant $a_* > 0$ such that for all $f$, $||M_f^*||_2 \leq ||f||_2$? Expanded, this now means:

**Theorem?**

$$\sum_x M_f^*(x)^2 \mu(x) \leq a_* \sum_x f(x)^2 \mu(x).$$

The paper does not address this directly, and we understand from private communication with Aram that he and his co-authors have not considered it yet. If it is true, then we can obtain the existence of "typical voters" in a more-general setting.

We assume that the $\mu(x)$-many voters who share the same characteristics $x$ all vote the same way. (If this is not true, we must be able to find some characteristic that divides them, and we are free to add it as a co-ordinate to every vector because the count $\mu(x)$ is allowed to be anything including zero.) Let $D$ be the set of $x$ that vote for the *loser*. Then $wt(D) = \sum_{x \in D} \mu(x)$ is the total number of votes for $D$, out of a total electorate of $E = \sum_x \mu(x)$. Define $f$ to be the 0-1 valued indicator function of $D$ as before. Then

$$||f||_2^2 = \sum_{x \in D} \mu(x) = wt(D),$$

while

$$||M_f^*||_2^2 = \sum_x \max_k \left[ \frac{wt(B_k(x) \cap D)}{wt(B_k(x))} \right]^2 \mu(x).$$

The "Theorem?" and dividing both sides by $E$ yields

$$\sum_x \max_k \left[\frac{wt(B_k(x) \cap D)}{wt(B_k(x))}\right]^2 \mu(x) \leq a'^2 \frac{wt(D)}{E}.$$

Again an averaging argument with respect to $E = \sum_x \mu(x)$ yields the existence of $x$ such that for all $k$,

$$\frac{wt(B_k(x) \cap D)}{wt(D)} < a_* \sqrt{\frac{wt(D)}{E}}.$$

Here the right-hand side has $\sqrt{\ell}$ where $\ell$ is the loser's percentage of the total vote, and a large enough victory makes $a'\sqrt{\ell} < 1/2$. This implies that the loser still has a minority in every Hamming ball centered on the voter $x$.

## Can we use the present inequality?

We can try to apply the given result to the function

$$f_*(x) = f(x)\sqrt{\mu(x)}.$$

Then we obtain

$$||f_*||_2^2 = \sum_x f(x)^2 \mu(x),$$

while (using the original $M_f$ again)

$$||M_{f_*}||_2^2 = \sum_x \max_k \left[\frac{1}{|B_k(x)|} \sum_{y \in B_k(x)} f(y)\sqrt{\mu(y)}\right]^2.$$

In general this does not look very tractable, nor does the theorem $||M_f||_2 \leq a||f||_2$ seem either to imply or follow from the desired weighted-case inequality above. In our specific election case, we might try

$f_*(x) = \sqrt{\mu(x)}$ if voters with $x$ chose the loser; $f(x) = 0$ otherwise,

Then $||f_*||_2 = wt(D)$, and the theorem proved gives

$$\sum_x \max_k \left[\frac{1}{|B_k(x)|} \sum_{y \in B_k(x)} \sqrt{\mu(y)}\right]^2 \leq a^2 wt(D).$$

The averaging argument, however, is still in terms of $2^n$ not $E$ as just above, so we get the existence of an $x$ such that for all $k$,

$$\frac{\sum_{y \in B_k(x)} \sqrt{\mu(y)}}{\binom{n}{k}} \leq a\sqrt{\frac{wt(D)}{2^n}}.$$

Neither side of this inequality yields the desired interpretation; the Cauchy-Schwarz inequality can relate the numerator to $wt(B_k(x) \cap D)$, but going in the wrong inequality direction. This hints how tricky the considerations in the paper are.

## Some Civics Notes on the Proof

## Open Problems