On teaching the approximation method for circuit lower bounds*

Oded Goldreich
Department of Computer Science
Weizmann Institute of Science, Rehovot, Israel.

March 15, 2023

Abstract

This text provides a basic presentation of the the approximation method of Razborov (*Matematicheskie Zametki*, 1987) and its application by Smolensky (*19th STOC*, 1987) for proving lower bounds on the size of $\mathcal{AC}^0[p]$-circuits that compute sums mod q (for primes $q \neq p$). The textbook presentations of the latter result concentrate on proving the special case of $q = 2$, and do not provide details on the proof of the general case. Furthermore, the presentations I have read tend to be too terse to my taste. The current text provides a detailed exposition of both the special case and the general case. Nevertheless, I agree with the common practice of covering only the case of $q = 2$ in class, and suggest leaving the general case (i.e, $q > 2$) for advanced reading.

Contents

*Partially supported by the Israel Science Foundation (grant No. 1041/18) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702).

1 Introduction (mainly for the teacher)

The approximation method of Razborov [5] and its application by Smolensky [4] for proving lower bounds on the size of $\mathcal{AC}^0[p]$-circuits that compute sums mod q (for primes $q \neq p$) are among the most celebrated results of circuit complexity. The textbook presentations of the latter result (see, e.g., [1, 3]) concentrate on proving the special case of $q = 2$, and do not provide details on the proof of the general case. Furthermore, the presentations I have read tend to be too terse to my taste. Aiming at making the proof of the result more accessible, I worked out a more detailed and friendly presentation of both the special case (i.e., $q = 2$) and the general case.

My presentation is aimed at graduate students interested in the theory of computation at large, and not necessarily at those focused on complexity theory, let alone circuit complexity. I assume that these students are familiar with the notion of Boolean circuits (including the notions of depth and fan-in). Such a familiarity is essential for the technical description. In addition, for the sake of perspective, I assume that the students are familiar with the P-vs-NP problem and the fact that our current state of knowledge regarding it is in its infancy. In particular, I assume that they know that even extremely modest (in comparison) separations (e.g., $\mathcal{AC}^0$ cannot compute Parity) are quite challenging to prove. Needless to say, if any of these assumptions does not hold, then one should start by correcting this state of affairs.

Turning to the actual contents, recall that proving lower bounds on the size of $\mathcal{AC}^0[p]$-circuits that compute sums mod q (for primes $q \neq p$) consists of two steps: (1) Showing that the computation of $\mathcal{AC}^0[p]$-circuits can be (well) approximated by low degree polynomials over $\mathrm{GF}(p)$; and (2) Showing that summation mod q cannot be approximated (well) by low degree polynomials over $\mathrm{GF}(p)$.

Step (1) is very intuitive and its presentation in standard texts is quite adequate; still, I will provide my own presentation of it (see Section 2.1 and Lemma 2 (in Section 2.2)). The problem is with the presentation of Step (2) and specifically with the case of $q > 2$. The special case of $q = 2$ is captured by Lemma 3, which is also presented in Section 2.2, whereas the general case is treated in Section 3. (Section 2.2 also presents a derivation of the lower bound for Majority, via a reduction of modular sums to it.)

I agree with the common practice of covering only Step (1) and the special case (i.e., $q = 2$) of Step (2) in class. I suggest using Section 2 as a basis for teaching this material, while leaving Section 3 (which covers the general case (i.e., $q \geq 2$)) for advanced reading. In contrast to these sections, Section 4 is not intended for the students: It presents a more complicated (alternative) proof of the general case, which may be of independent interest.

2 The basic material

The approximation method is pivoted at the discrepancy between the ability of *low degree* polynomials (over a small prime field) to approximate functions computed by certain classes of Boolean circuits and their inability to approximate modular sums modulo a different prime (i.e., different from the field size). In particular, we shall prove that, on the one hand, *low degree* polynomials over $\mathrm{GF}(p)$ can (well) approximate constant-depth (unbounded fan-in) circuits, but, on the other hand, they cannot (well) approximate the sum modulo q, for any prime $q \neq p$. The conclusion is that constant-depth (unbounded fan-in) circuits cannot compute such modular sums. (We mention that the foregoing holds also when the circuits are equipped with (unbounded fan-in) $\texttt{MOD}\,p$ gates.)

The foregoing emphasis on low degree is crucial: Polynomials of degree n over GF(2) can compute any n-variate Boolean function.[1] On the other hand, the number of low degree n-variate polynomials is much small than the number of n-variate Boolean function.[2]

The concept of approximation is also crucial here, because no low degree polynomial can perfectly agree (even) with extremely simple functions such as the n-wide AND (or OR).[3] Hence, we consider the class of functions that can be *well approximated by low degree polynomials*, showing that this class contains all functions computed by constant-depth (unbounded fan-in) circuits, but not some other simple functions (e.g., Majority and MOD q for q different from the field size). We start with an overview, which is focused on the first aspect.

2.1 Overview

The key observation that underlies the *approximation method* is that an unbounded fan-in OR-gate can be well approximated by a low degree polynomial (say, over GF(2)); that is, for every distribution $\mathcal{D}$ over $\{0,1\}^w$, there exists a degree ℓ polynomial $P : \mathrm{GF}(2)^w \to \mathrm{GF}(2)$ such that

$$\Pr_{(y_1,...,y_w)\sim\mathcal{D}}\Big[P(y_1,...,y_w) = \mathtt{OR}(y_1,...,y_w)\Big] \geq 1 - 2^{-\ell}. \tag{1}$$

We stress that the degree of the polynomial is logarithmic in the (reciprocal of the) desired error bound, and does not depend on the number of variables. Furthermore, the claim holds for any distribution $\mathcal{D}$ (rather than only for the uniform distribution over $\{0,1\}^w$).[4] The latter fact becomes crucial when we wish to replace an intermediate gate in a circuit (i.e., a gate that is fed by the outputs of other gates).

The foregoing claim is proved by observing that, for every $(y_1,...,y_w) \in \{0,1\}^w \setminus \{0^w\}$, a random linear function $L : \mathrm{GF}(2)^\ell \to \mathrm{GF}(2)$ satisfies $\Pr_L[L(y_1,...,y_w) = 1] = 1/2$, whereas $L(0,...,0) = 0$ for every linear function L. Hence, for every $(y_1,...,y_w) \in \{0,1\}^w$, it holds that

$$\Pr_{L_1,....,L_\ell}\Big[\mathtt{OR}(L_1(y_1,...,y_w),...,L_\ell(y_1,...,y_w)) = \mathtt{OR}(y_1,...,y_w)\Big] \geq 1 - 2^{-\ell},$$

where $L_1,...,L_\ell$ are random linear functions. It follows that there exists linear functions $L_1,...,L_\ell$ such that

$$\Pr_{(y_1,...,y_w)\sim\mathcal{D}}\Big[\mathtt{OR}(L_1(y_1,...,y_w),...,L_\ell(y_1,...,y_w)) = \mathtt{OR}(y_1,...,y_w)\Big] \geq 1 - 2^{-\ell}.$$

Hence, replacing $\mathtt{OR}(z_1,...,z_\ell)$ by the polynomial $1 - \prod_{j\in[\ell]}(1 - z_j)$, we derive Eq. (1); specifically, we define $P(y_1,...,y_w) \stackrel{\text{def}}{=} 1 - \prod_{j\in[\ell]}(1 - L_j(y_1,...,y_w))$, where the L_j's are the foregoing linear functions. A similar idea can be applied in GF(p), for any prime p, but in that case we raise the

[1] Hint: Write the function in DNF using terms of size n, and note that the conjunction of n literals can be computed by multiplying the corresponding linear function (e.g., $x_1 \wedge (\neg x_2) \wedge x_3$ can be computed by $x_1 \cdot (1 - x_2) \cdot x_3$).

[2] Hint: The unber of degree d mononimals over n varaiables is $\binom{n}{d}$.

[3] Hint: Use the fact that a non-zero degree d polynomial over GF(2) evaluates to 1 with probability at least 2^{-d}.

[4] Note that when seeking to approximate an OR-gate under the uniform distribution, we can use

$$\Pr_{(y_1,...,y_w)\in\{0,1\}^n}\Big[\mathtt{OR}(y_1,...,y_\ell) = \mathtt{OR}(y_1,...,y_w)\Big] \geq 1 - 2^{-\ell},$$

and then replace $\mathtt{OR}(y_1,...,y_\ell)$ by the polynomial $1 - \prod_{j\in[\ell]}(1 - y_j)$.

linear functions to power $p-1$ in order to guarantee an answer in $\{0,1\}$. Consequently, the degree of the polynomial is $(p-1)\cdot\ell$.

Applying the analogous replacement to all (OR and AND) gates of a depth d circuit of size s, we obtain a degree $d\cdot(p-1)\cdot\ell$ polynomial over GF(p) that approximates the circuit up to error of $s\cdot 2^{-\ell}$, even if the circuit has unbounded fan-in MOD p gates. The replacement process will be described and analyzed in detail in the proof of Lemma 2. In contrast, it can be shown that low degree polynomials cannot approximate the MOD q function with such a small error rate, where q is an arbitrary fixed prime different from p. Indeed, we said nothing about how the latter claim is proved: The special case of $q=2$ is stated in Lemma 3, whereas the general case is treated in Section 3.

2.2 The actual theorem and its proof

We start by recalling the general background and the relevant preliminaries.

General background. When discussing circuit (size and/or depth) lower bounds, the point is obtaining them for *explicit* functions; in contrast, it is trivial to get such lower bounds for non-explicit functions or even for functions of high (uniform) time complexity (i.e., just let the algorithm try all functions and all circuits). The question is what is "explicit" and the answer is undetermined; actually, one often wants things to be *as explicit as possible*. Still, the first choice would be that *explicit* means computable in polynomial-time. Often (see, e.g., [1, Sec. 6.1] and [2, Sec. 5.2.3]), one may require even more; e.g., computability in log-space. Here, we shall discuss size lower bounds for very explicit functions (e.g., Majority, Parity, etc).

Preliminaries. The class $\mathcal{AC}^0[m]$ consists of Boolean functions computable by families of circuits of polynomial size and constant depth having unbounded fan-in AND, OR, NOT, and MOD m gates, where $m>1$ is an arbitrary constant. The MOD m gates, denoted $\mathtt{MOD}_m$, are defined such that $\mathtt{MOD}_m(x_1,...,x_w)=0$ if $\sum_{i\in[w]}x_i\equiv 0 \pmod m$ and 1 otherwise (i.e., if $\sum_{i\in[w]}x_i \bmod m\in\{1,...,m-1\}$). Indeed, $\mathtt{MOD}_2$ coincides with XOR (equiv, the Parity function). We shall focus on the case that m is a prime number, while warning that the case of composite m's (even $m=6$) is wide open (with the exception of prime powers).[5]

The following result implies that, for any prime p, the class $\mathcal{AC}^0[p]$ cannot compute simple functions such as Majority or $\mathtt{MOD}_q$, where $q\neq p$. Indeed, this result does fit the intuition that functions of a "counting" falvor (other than $\mathtt{MOD}_p$ itself) cannot be computed by $\mathcal{AC}^0[p]$, but the

[5] Essentially, for every prime power p^e, it holds that $\mathcal{AC}^0[p^e]=\mathcal{AC}^0[p]$, because $\mathtt{MOD}_{p^e}$ can be implemented in $\mathcal{AC}^0[p]$ (and $\mathtt{MOD}_p$ can be implemented in $\mathcal{AC}^0[p^e]$ (e.g., by duplicating each input p^{e-1} times)). Specifically, we can compute $\mathtt{MOD}_{p^e}(x)$ using $\mathtt{MOD}_{p^{e-1}}$ and $\mathtt{MOD}_p$ gates as follows.

- For every $i\in[n]$, we compute $y_i=\mathtt{MOD}_{p^{e-1}}(x_1,...,x_i)$.
  Note that $(y_{i-1},y_i)=(1,0)$ holds if any only if $\sum_{j\in[i-1]}x_i\equiv -1 \pmod{p^{e-1}}$ and $x_i=1$.
  This implies that $|\{i\in\{2,...,n\}:(y_{i-1},y_i)=(1,0)\}|$ equals $\left\lfloor \sum_{i\in[n]}x_i/p^{e-1}\right\rfloor$.
- For every $i\in\{2,...,n\}$, we compute $z_i=\mathtt{AND}(y_{i-1},\neg y_i)$.
  Note that $\sum_{i\in[n]}x_i=p^{e-1}\cdot\left(\sum_{i\in\{2,...,n\}}z_i\right)+\left(\sum_{i\in[n]}x_i \bmod p\right)$.

Hence, $\mathtt{MOD}_{p^e}(x)=0$ if and only if both $\mathtt{MOD}_p(z)=0$ and $\mathtt{MOD}_p(x)=0$.

point is that we can actually prove this statement. In contrast, we do not know a proof of the equally intuitive conjecture that $\mathcal{AC}^0[6]$ cannot compute Majority.

Theorem 1 (size lower bounds for constant-depth Boolean circuits with AND, OR, NOT, and $\mathtt{MOD}_p$ gates, when p is a prime): *For any prime $p \geq 2$, the following holds.*

1. *Computing the majority of n bits by a depth d circuit with unbounded fan-in* AND, OR, NOT, *and* $\mathtt{MOD}_p$ *gates requires size* $\exp(\Omega(n^{1/2d}))$.
2. *For any prime q different from p, computing the* $\mathtt{MOD}_q$ *of n bits by a depth d circuit with unbounded fan-in* AND, OR, NOT, *and* $\mathtt{MOD}_p$ *gates requires size* $\exp(\Omega(n^{1/2d}))$.

In particular, Part 1 (with $p = 2$) implies that $\mathcal{AC}^0[2]$ cannot compute Majority, whereas Part 2 (with $p = 3$ and $q = 2$) implies that $\mathcal{AC}^0[3]$ (let alone $\mathcal{AC}^0$ itself) cannot compute Parity. In general, Part 2 implies that $\mathcal{AC}^0[p]$ cannot compute $\mathtt{MOD}_q$, for $q \neq p$.

We shall focus on proving Part 2, while noting that it implies a (weaker but sufficiently interesting) version of Part 1. In general, the proof of Theorem 1 combines two steps:

Step 1: Proving that the computation of $\mathcal{AC}^0[p]$ circuits can be well approximated by polynomials of low degree over the finite field of p elements, denoted $\mathrm{GF}(p)$.

This will be proved in Lemma 2 using the ideas presented in the overview (for the case of $p = 2$).

Step 2: Proving that the target functions (i.e., n-wise Majority and n-wise $\mathtt{MOD}_q$, for $q \neq p$) cannot be well approximated by low degree polynomials over $\mathrm{GF}(p)$.

This will be proved in Lemma 3 for $\mathtt{MOD}_2$ and any fixed prime $p \neq 2$, which is sufficiently interesting. The general case of $\mathtt{MOD}_q$ for any fixed primes $q \neq p$ is treated in Section 3.

We shall also show that the lower bound for computing $\mathtt{MOD}_q$, for any q (e.g., $q = 2$), yields a lower bound for Majority. Starting with Step 1, we prove the following.

Lemma 2 (approximating $\mathcal{AC}^0[p]$ by low degree polynomials over $\mathrm{GF}(p)$): *For any prime $p \geq 2$, let $C : \{0,1\}^n \to \{0,1\}$ be a depth d circuit of size s with unbounded fan-in* AND, OR, NOT, *and* $\mathtt{MOD}_p$ *gates. Then, there exists a degree D polynomial A over* $\mathrm{GF}(p)$ *such that*

$$\Pr_{x \in \{0,1\}^n} [A(x) = C(x)] > 1 - \frac{s}{\exp(\Omega(D^{1/d}))}$$

where the Omega-notation hides a $\frac{\log p}{p-1}$ *factor.*

Setting $D = \sqrt{n}$ and using a sufficiently small $s = \exp(\Omega(n^{1/2d}))$, we get an approximation error of $o(1)$. (In contrast, we shall later show that degree $\sqrt{n}$ polynomials over $\mathrm{GF}(p)$ have error rate $\Omega(1)$ with respect to $\mathtt{MOD}_q : \{0,1\}^n \to \{0,1\}$, for any fixed prime $q \neq p$.)

Proof: The proof is by induction on the structure of C, and all arithmetic expressions are in $\mathrm{GF}(p)$. Denoting the function computed by the top (output) gate by g, we consider the four possible cases.

1. *The top gate is a* NOT *gate*: If $g = \neg f$, then we approximate g by $\widetilde{g} \stackrel{\text{def}}{=} 1 - \widetilde{f}$, where $\widetilde{f}$ is the polynomial that approximates f.

   Note that $\widetilde{g}$ has the same degree as $\widetilde{f}$, and that $\widetilde{g}(x) \in \{0,1\}$ whenever $\widetilde{f}(x) \in \{0,1\}$. Furthermore, if $\widetilde{f}(x) = f(x)$, then $\widetilde{g}(x) = g(x)$. Hence, the current replacement adds no approximation error.

2. *The top gate is a* $\mathtt{MOD}_p$ *gate*: If $g = \mathtt{MOD}_p(f_1, ..., f_w)$, then we approximate g by the polynomial $\widetilde{g} \stackrel{\text{def}}{=} \left(\sum_{i\in[w]} \widetilde{f_i}\right)^{p-1}$, where $\widetilde{f_i}$ is the polynomial that approximates f_i.

   Note that $\widetilde{g}$ has degree $(p-1)\cdot\max_{i\in[w]}\{\deg(\widetilde{f_i})\}$, and that $\widetilde{g}(x) \in \{0,1\}$ for every $x \in \{0,1\}^n$. Furthermore, for every x, if $\widetilde{f_i}(x) = f_i(x)$ for every $i \in [w]$, then $\widetilde{g}(x) = g(x)$, because in that case

$$\widetilde{g}(x) \;=\; \left(\sum_{i\in[w]} f_i(x)\right)^{p-1} \;=\; \mathtt{MOD}_p(f_1(x), ..., f_w(x)),$$

   where the last equality uses the fact that $v^{p-1} = 1$ for every $v \in \mathrm{GF}(p) \setminus \{0\}$ (and $0^{p-1} = 0$). Again, the current replacement adds no approximation error.

3. *The top gate is an* OR *gate*: If $g = \mathtt{OR}(f_1, ..., f_w)$, then, using suitable (see next) linear functions $L_j : \mathrm{GF}(p)^w \to \mathrm{GF}(p)$, for $j = 1, ..., \ell$, where ℓ is currently a free parameter, we approximate g by the polynomial

$$\widetilde{g} \stackrel{\text{def}}{=} 1 - \prod_{j\in[\ell]} \left(1 - L_j(\widetilde{f_1}, ..., \widetilde{f_w})^{p-1}\right)$$

   where $\widetilde{f_i}$ is the polynomial that approximates f_i.

   Note that $\widetilde{g}$ has degree $\ell\cdot(p-1)\cdot\max_{i\in[w]}\{\deg(\widetilde{f_i})\}$, and that $1-\prod_{j\in[\ell]}\left(1 - L_j(v_1, ..., v_w)^{p-1}\right) \in \{0,1\}$ for every $v_1, ..., v_w \in \mathrm{GF}(p)$. Hence, $\widetilde{g}(x) \in \{0,1\}$ for every $x \in \{0,1\}^n$. The key issue is the selection of the L_j's, and the clue for it is provided by the following claim.

   Claim 2.1 (on a random linear combination of elements of a non-zero sequence): *Suppose that* $v_1, ..., v_w \in \{0,1\}$ *such that* $\mathtt{OR}(v_1, ..., v_w) = 1$. *Then, for a random linear function* $L : \mathrm{GF}(p)^w \to \mathrm{GF}(p)$, *it holds that* $\Pr_L[L(v_1, ..., v_w)=0] = 1/p$.

   Proof: The value of a random linear function $L : \mathrm{GF}(p)^w \to \mathrm{GF}(p)$ at a non-zero point $(v_1, ..., v_w) \in \mathrm{GF}(p)^w$ is uniformly distributed in $\mathrm{GF}(p)$, because $L(z_1, ..., z_w) = \sum_{i\in[w]} r_i z_i$, where the r_i's are uniformly and independently distributed. ■

   Selecting linear function $L_1, ..., L_\ell : \mathrm{GF}(p)^w \to \mathrm{GF}(p)$ uniformly at random, the following holds for every $x \in \{0,1\}^n$ such that $\widetilde{f_i}(x) \in \{0,1\}$ for every $i \in [w]$:

   - If $\widetilde{f_1}(x) = \cdots = \widetilde{f_w}(x) = 0$, then all L_j's evaluate to 0, which implies that $\prod_{j\in[\ell]}(1-L_j^{p-1})$ is identically 1.
   - If $\widetilde{f_i}(x) = 1$ for some $i \in [w]$, then each L_j evaluates to 0 with probability $1/p$ (equiv., each $1 - L_j^{p-1}$ evaluates to 1 with probability $1/p$), which implies that the product $\prod_{j\in[\ell]}(1 - L_j^{p-1})$ is 1 with probability $p^{-\ell}$.

Recalling that $1-\prod_{j\in[\ell]}(1-L_j^{p-1})\in\{0,1\}$ always holds, we get

$$\Pr_{L_1,...,L_w}\left[\mathtt{OR}(\widetilde{f}_1(x),...,\widetilde{f}_w(x))\neq 1-\prod_{j\in[\ell]}\left(1-L_j(\widetilde{f}_1(x),...,\widetilde{f}_w(x))^{p-1}\right)\right]\leq p^{-\ell} \qquad (2)$$

Using an averaging argument, it follows that there exists a choice of linear function $L_1,...,L_\ell:\mathrm{GF}(p)^w\to\mathrm{GF}(p)$ such that, for a uniformly distributed x, it holds that

$$\Pr_{x\in\{0,1\}^n}\left[\mathtt{OR}(\widetilde{f}_1(x),...,\widetilde{f}_w(x))\neq 1-\prod_{j\in[\ell]}\left(1-L_j(\widetilde{f}_1(x),...,\widetilde{f}_w(x))^{p-1}\right)\right]\leq p^{-\ell}. \qquad (3)$$

It follows that replacing $\mathtt{OR}(\widetilde{f}_1(x),...,\widetilde{f}_w(x))$ by $\widetilde{g}$ adds an approximation error of at most $p^{-\ell}$.

4. *The top gate is an* AND *gate*: Analogously, if $g=\mathtt{AND}(f_1,...,f_w)$, then we approximate g by $\widetilde{g}\stackrel{\text{def}}{=}\prod_{j\in[\ell]}\left(1-L_j(1-\widetilde{f}_1,...,1-\widetilde{f}_w\right)^{p-1})$, where $\widetilde{f}_i$ is the polynomial that approximates f_i and the L_j's are suitable linear functions.[6] Again, the current replacement adds an approximation error of $p^{-\ell}$.

Hence, replacing each of the gates by the corresponding polynomial adds an approximation error of at most $p^{-\ell}$, and so replacing all s gates yields an approximation error of at most $s\cdot p^{-\ell}$. The degree of the resulting polynomial, which approximates the circuit C, is at most $((p-1)\cdot\ell)^d$. Aiming at $((p-1)\cdot\ell)^d\leq D$, we use $\ell=\frac{1}{p-1}\cdot D^{1/d}$, and obtain an approximation error of $s\cdot p^{-D^{1/d}/(p-1)}$. ■

Digest. The key fact, captured by Claim 2.1, is that a random choice of a linear function is positively correlated with an OR-gate. (We note that a somewhat weaker result can be achieved by using a random 0-1 linear function (i.e., 0-1 coefficients only)).) In any case, Eq. (3) upper-bounds the approximation error of replacing a single OR-gate (by a suitable degree $(p-1)\cdot\ell$ polynomial), and the error bound of Lemma 2 follows by applying a union bound over all gates (while using an adequate setting of ℓ). Note that raising various $\mathrm{GF}(p)$-expressions to the power of $p-1$ guarantees that the resulting polynomial always yields a value in $\{0,1\}$.

Proving Part 1 of Theorem 1. As stated above, Part 2 of the theorem (i.e., a lower bound for $\mathtt{MOD}_q$) is established by combining Lemma 2 with a proof that a degree $\sqrt{n}$ polynomial over $\mathrm{GF}(p)$ cannot approximate the n-bit $\mathtt{MOD}_q$ function. While Part 1 (i.e., a lower bound for Majority) can be proven analogously (see Appendix), we establish a weaker version of Part 1 by observing that *any symmetric function* (e.g., Parity (i.e., $\mathtt{MOD}_2$)) *can be $\mathcal{AC}^0$-reduced to computing Majority.* This observation is proved next.

Let $\mathtt{TH}_k^n:\{0,1\}^n\to\{0,1\}$ denote the function that return 1 if and only if its input contains at least k ones, and note that $\mathtt{TH}_k^n(x)=\mathtt{TH}_{n+1}^{2n+1}(x1^{n+1-k}0^k)$, where $\mathtt{TH}_{n+1}^{2n+1}$ is the $(2n+1)$-bit Majority function. Now, letting $\mathrm{wt}(x_1,...,x_n)\stackrel{\text{def}}{=}|\{i\in[n]:x_i=1\}|$, suppose that for some $S\subseteq[n]$ the function

[6] Indeed, the expression for $\widetilde{g}$ can be obtained by observing that $g=\neg\mathtt{OR}(\neg f_1,...,\neg f_w)$.

$f : \{0,1\}^n \to \{0,1\}$ satisfies $f(x) = 1$ if and only if $\mathrm{wt}(x) \in S$. Then, for $S = \{s_1, ..., s_m\}$, it holds that

$$f(x) = \bigvee_{i \in [m]} \mathtt{AND}(\mathtt{TH}^n_{s_i}(x), \neg \mathtt{TH}^n_{s_i+1}(x)),$$

because $\mathtt{AND}(\mathtt{TH}^n_s(x), \neg \mathtt{TH}^n_{s+1}(x)) = 1$ if and only if $\mathrm{wt}(x) = s$. Hence, a lower bound on the size of $\mathcal{AC}^0[p]$-circuits of depth $d+3$ that compute f (e.g., $f = \mathtt{MOD}_2$) yields a lower bound on the size of $\mathcal{AC}^0[p]$-circuits of depth d that compute Majority. Actually, using specific features of the foregoing reduction, a lower bound on the size of $\mathcal{AC}^0[p]$-circuits of depth d that compute f yields a lower bound on the size of $\mathcal{AC}^0[p]$-circuits of depth d that compute Majority.[7]

Proving a special case of Part 2 of Theorem 1. We now turn to Part 2 of Theorem 1 (i.e., a lower bound for $\mathtt{MOD}_q$). Specifically, we prove a special case of Part 2 (i.e., the case of $q = 2$), by combining Lemma 2 with the following result –

Lemma 3 (on the error rate of low degree polynomials over $\mathrm{GF}(p)$ that approximate $\mathtt{MOD}_2$): *There exists a constant $\epsilon > 0$ such that, for any prime $p \geq 3$, any n-variate polynomial $Q : \mathrm{GF}(p)^n \to \mathrm{GF}(p)$ of degree at most $\sqrt{n}$ fails to compute the n-ary parity on at least $\epsilon \cdot 2^n$ of the n-bit inputs; that is,*

$$\Pr_{x \in \{0,1\}^n}[Q(x) \neq \mathtt{MOD}_2(x)] \geq \epsilon.$$

For any constant $\delta > 0$, the claim holds (with a different $\epsilon > 0$) also for polynomials of degree $\delta \cdot \sqrt{n}$: In particular, for $\delta < 1$ we get $\epsilon = 0.5 - O(\delta)$ whereas for $\delta > 1$ we get $\epsilon = \exp(-O(\delta^2))$; actually, the result holds also for non-constant δ (see Footnote 8 for details). The special case of Part 2 of Theorem 1 (i.e., the case of $q = 2$) follows by a suitable setting of parameters in Lemma 2. Specifically, in contrast to Lemma 3, this setting implies that *for a sufficiently small $c > 0$, it holds that $\mathcal{AC}^0[p]$-circuits of depth d and size $\exp(c \cdot n^{1/2d})$ can be approximated by degree $\sqrt{n}$ polynomials with an approximation error $o(1)$.*

Proof: Let $G \stackrel{\text{def}}{=} \{x \in \{0,1\}^n : Q(x) = \mathtt{MOD}_2(x)\}$ denote the set of inputs on which Q equals $\mathtt{MOD}_2$. We shall prove that G misses a constant fraction of $\{0,1\}^n$ by using Q to present a class of $p^{(1-\Omega(1))\cdot 2^n}$ polynomials that can compute $p^{|G|}$ different functions. Specifically, the low degree of Q will be used to upper-bound the degree of the polynomials in the foregoing class (which will contain only multi-linear polynomials).

The crucial step is a variable substitution that maps $x_i \in \{0,1\}$ to $(-1)^{x_i} \in \{\pm 1\} \equiv \{1, q-1\}$, where, for $x_i \in \{0,1\}$, it holds that $(-1)^{x_i} = (1 - x_i) \cdot (-1)^0 + x_i \cdot (-1)^1 = 1 - 2x_i$. The benefit of this technical step is that it related $\mathtt{MOD}_2(x_1, ..., x_n)$ to $\prod_{i \in [n]} (-1)^{x_i}$; that is, $(-1)^{\mathtt{MOD}_2(x_1,...,x_n)} = \prod_{i \in [n]} (-1)^{x_i}$. Accordingly, we consider the polynomial $R : \mathrm{GF}(p)^n \to \mathrm{GF}(p)$ defined as $R(y_1, ..., y_n) \stackrel{\text{def}}{=} 1 - 2 \cdot Q(x_1, ..., x_n)$, where $x_i = (1 - y_i)/2$ (equiv., $y_i = 1 - 2x_i$), noting that R has the same degree as Q. The salient feature of the polynomial R is that $R(y_1, ..., y_n) = \prod_{i \in [n]} y_i$ holds whenever the corresponding $(x_1, ..., x_n)$ is in G. This is the case because for every $(x_1, ..., x_n) \in \{0,1\}^n$ it holds

[7] Specifically, we use the fact that $f(x)$ equals $\sum_{i \in [m]} (\mathtt{TH}^n_{s_i}(x) \cdot (1 - \mathtt{TH}^n_{s_i+1}(x)))$, when the sum and the 2-argument multiplication are in $\mathrm{GF}(p)$, and the fact that a (depth d+3) circuit that has a corresponding form can be approximated by a polynomial that has degree twice the degree of the depth d circuit that computes $\mathtt{TH}^n_s$.

that

$$\begin{aligned}\prod_{i\in[n]}(1-2x_i) &= \prod_{i\in[n]}(-1)^{x_i}\\ &= (-1)^{\mathtt{MOD}_2(x_1,....,x_n)}\\ &= 1-2\cdot\mathtt{MOD}_2(x_1,....,x_n)\end{aligned}$$

whereas $(x_1,...,x_n)\in G$ implies that $\mathtt{MOD}_2(x_1,....,x_n)=Q(x_1,...,x_n)$. Hence, $\prod_{i\in[n]}(1-2x_i)=1-2\cdot Q(x_1,....,x_n)$ for every $(x_1,...,x_n)\in G$, which means that $\prod_{i\in[n]}y_i=R(y_1,....,y_n)$ for every $((1-y_1)/2,....,(1-y_n)/2)\in G$.

Consequently, letting $H\stackrel{\rm def}{=}\left\{y\in\{\pm1\}^n : R(y)=\prod_{i\in[n]}y_i\right\}$, and observing that $|H|=|G|$, we seek to upper-bound $|H|$. The key fact that we shall use is that R has a magical feature: *It is a degree $\sqrt{n}$ polynomial that, when restricted to H, equals a degree n polynomial* (specifically, $\prod_{i\in[n]}y_i$). (Indeed, the letter 'R' was chosen for evoking the degree-**r**eduction feature of R (when restricted to H).)

Towards upper-bounding $|H|$, we consider the class $\mathcal{F}$ of all function $f:H\to\mathrm{GF}(p)$, and note that $|\mathcal{F}|=p^{|H|}$. We first observe that each $f\in\mathcal{F}$ can be written as a linear combination of multi-linear monomials, because $\sigma^2=1$ for every $\sigma\in\{\pm1\}$. Hence, for every $y\in H$,

$$f(y)=\sum_{I\subseteq[n]}f_I\cdot\prod_{i\in I}y_i,$$

where the f_I's are in $\mathrm{GF}(p)$. Furthermore, for every $y\in H$, using $R(y)=\prod_{i\in[n]}y_i$ *we decrease the degree of any monomial to at most* $(n+\sqrt{n})/2$, because $\prod_{i\in I}y_i=\left(\prod_{i\in[n]}y_i\right)\cdot\prod_{i\in[n]\setminus I}y_i=R(y)\cdot\prod_{i\in[n]\setminus I}y_i$, which has degree $\sqrt{n}+(n-|I|)$, whereas either $|I|\leq(n+\sqrt{n})/2$ or $\sqrt{n}+(n-|I|)<(n+\sqrt{n})/2$. (We stress that $R(y)\cdot\prod_{i\in[n]\setminus I}y_i$ is a sum of *multi-linear* monomials of degree at most $(n+\sqrt{n})/2$, where here, and in the line above, we used $y_i^2=1$ for $y_i\in\{\pm1\}$.) Hence, letting $t\stackrel{\rm def}{=}(n+\sqrt{n})/2$, we get

$$\begin{aligned}f(y) &= \sum_{I\subseteq[n]:|I|\leq t}f_I\cdot\prod_{i\in I}y_i+\sum_{I\subseteq[n]:|I|>t}f_I\cdot\prod_{i\in I}y_i\\ &= \sum_{I\subseteq[n]:|I|\leq t}f_I\cdot\prod_{i\in I}y_i+\sum_{I\subseteq[n]:|I|>t}f_I\cdot R(y)\cdot\prod_{i\in[n]\setminus I}y_i\end{aligned}$$

which means that each $f\in\mathcal{F}$ can be represented as a linear combination of multi-linear monomials of degree at most $t=(n+\sqrt{n})/2$. Noting that the number of such monomials is $\sum_{i=0}^{t}\binom{n}{i}$, we conclude that $|\mathcal{F}|\leq p^{\sum_{i=0}^{t}\binom{n}{i}}$, which implies $|H|\leq\sum_{i=0}^{t}\binom{n}{i}$, and the claim follows (because $\sum_{i\leq0.5n+O(\sqrt{n})}\binom{n}{i}=(1-\Omega(1))\cdot2^n$).[8] ■

[8] More generally, denoting the degree of Q by D and letting $t=(n+D)/2$, we need to upper-bound $\sum_{i\leq t}\binom{n}{i}$. For $D\leq\sqrt{n}$, we use

$$\sum_{i\leq t}\binom{n}{i}\leq\sum_{i\leq(n-1)/2}\binom{n}{i}+(1+(D/2))\cdot O(2^n/\sqrt{n})\leq(0.5+O(D/\sqrt{n}))\cdot2^n$$

whereas for $D>\sqrt{n}$ we have $2^{-n}\cdot\sum_{i\leq t}\binom{n}{i}=1-\exp(-O(D^2/n))$.

Digest. Lemma 3 starts with an algebraic manipulation that translates the existence of a low degree polynomial that approximates $\mathtt{MOD}_2$ (over $\{0,1\}^n$) to the existence of a low degree polynomial that allows to effectively (i.e., w.r.t $\{\pm 1\}^n$) decrease the degree of any monomial to $t \stackrel{\text{def}}{=} 0.5n + O(\sqrt{n})$, which in turn upper-bounds the quality of the initial approximation. Specifically, if the initial approximation is correct on N of the inputs (in $\{0,1\}^n$), then the degree reduction holds for N other inputs (in $\{\pm 1\}^n \subseteq \mathrm{GF}(p)^n$), which span a vector space of dimension N, whereas the space of functions (defined over these N inputs) that can be express as a linear combination of multi-linear monomials of degree at most t has dimension at most $\sum_{i \leq t} \binom{n}{i} = (1 - \Omega(1)) \cdot 2^n$. (Hence, $N \leq (1 - \Omega(1)) \cdot 2^n$.)

The observation that enables this miraculous degree reduction is that $(-1)^{\mathtt{MOD}_2(x)} = \prod_{i \in [n]} (-1)^{x_i}$. This observation is put into work by the mapping $x_i \mapsto (-1)^{x_i}$, which coincides (over $\{0,1\}$) with the linear function $L(\zeta) = 1 - 2\zeta = (1-\zeta) \cdot (-1)^0 + \zeta \cdot (-1)$.

Extension to arbitrary $q \neq p$. It is tempting to try to extend the foregoing argument to $\mathtt{MOD}_q$, when using a q^{th} root of unity (over the complex numbers) and the corresponding extension field of $\mathrm{GF}(p)$. Although this idea may not be work as stated, something along these lines does work (i.e., we consider a $(q-1)$-dimensional extension field and an element of multiplicative order q in it).[9] But the real difficulty is that, in general, unlike in the case of $q = 2$, it does not hold that $\mathtt{MOD}_q(x_1, ..., x_n)$ equals to $\mathtt{mod}_q(x_1, ..., x_n) \stackrel{\text{def}}{=} \sum_{i \in [n]} x_i \bmod q$, even when the x_i's are restricted to $\{0,1\}$. Loosely speaking, this difficulty is resolved by working with $\mathtt{mod}_q$ (rather than with $\mathtt{MOD}_q$).

Another difficulty is that the product of multi-linear monomials is not necessarily a multi-linear monomial, even when the variables are restricted to $\{1, \omega\}$, where ω denotes an element of order q in the extension field. (This is because, unlike in the case of $q = 2$, it does not holds that $\omega^e \in \{1, \omega\}$ for every $e \in \mathbb{Z}_q = \{0, 1, ..., q-1\}$.) This difficulty is resolved by observing that we can reduce the individual degrees of variables over $\{1, \omega\}$ by using adequate linear transformations; that is, for a variable $\zeta \in \{1, \omega\}$, for any $e \in \mathbb{Z}_q$, we can replace ζ^e with the linear (in ζ) function $\frac{\zeta - 1}{\omega - 1} \cdot \omega^e + \frac{\zeta - \omega}{1 - \omega}$.

The bottom-line is that, using the foregoing modifications, it is possible to prove an extension of Lemma 3 to arbitrary $q \neq p$, where this extension refers to approximating $\mathtt{mod}_q$ rather than to approximating $\mathtt{MOD}_q$. Although this extension does not fit Lemma 2, which refers to approximating $\mathtt{MOD}_q$, we shall also show how to bridge this gap. For details see Section 3.

3 Advanced reading

Recall that Lemma 3 relies on translating a low degree polynomial (over $\mathrm{GF}(p)$) that approximates $\mathtt{MOD}_2 \equiv \mathtt{mod}_2$ over $\{0,1\}^n$ to a low degree polynomial (also over $\mathrm{GF}(p)$) that approximates the product of n variables that assume values in $\{\pm 1\} = \{1, p-1\}$. As stated in the foregoing digest, it is tempting to try to extend this strategy to $\mathtt{mod}_q$ by using a q^{th} root of unity (over the complex numbers) and the corresponding extension field of $\mathrm{GF}(p)$.

One reason to be alarmed about this proposal is that it is unclear where we use the condition $p \neq q$. In fact, in case $p = q$, for every $e \in \mathbb{Z}$, it holds that $p^e - 1$ is not divisible by q, which means that the corresponding extension field has no element of multiplicative order q. In contrast, when $p \neq q$, it holds that $p^{q-1} \equiv 1 \pmod{q}$, which means that q divides $p^{q-1} - 1$, which implies the $(q-1)$-dimensional extension field of $\mathrm{GF}(p)$ has elements of multiplicative order q.

[9] See discussion at the beginning of Section 3.

A parenthetical question is whether the foregoing ($(q-1)$-dimensional) extension field (of $\mathrm{GF}(p)$) is spanned by the powers of the q^{th} root of unity (over the complex numbers). This is the case if and only if $(x^q-1)/(x-1)=\sum_{i=0}^{q-1} x^i$ is an irreducible polynomial over $\mathrm{GF}(p)$. Note that for some $p \neq q$ the answer is positive, whereas for others it is negative.

Notation: In light of the foregoing, fixing an arbitrary pair of primes, denoted $p \neq q$, we denote by $\mathcal{K}$ the $(q-1)$-dimensional extension field of $\mathrm{GF}(p)$, and denote by $\omega \in \mathcal{K}$ an arbitrary element of multiplicative order q. We shall also use the notation $\mathbb{Z}_q \stackrel{\text{def}}{=} \{0,1,...,q-1\}$.

Recalling that $\mathtt{mod}_q : \{0,1\}^n \to \mathbb{Z}_q$ is defined such that $\mathtt{mod}_q(x_1,...,x_n) \stackrel{\text{def}}{=} \sum_{i\in[n]} x_i \bmod q$, we face the fact that functions with range in $\mathrm{GF}(p)$ cannot possibly approximate $\mathtt{mod}_q$ well if $q > p$. This problem does not arise when $q < p$, because then we can embed $\mathbb{Z}_q$ in $\mathrm{GF}(p)$, let alone do so in a straightforward and transparent manner. Hence, we start with the case of $q < p$, and postpone the case of $q > p$ to later.

3.1 The case of $q < p$

With the foregoing preliminaries in place, we start by stating and proving the natural extension of Lemma 3. Note that this extension refers to $\mathtt{mod}_q$ rather than to $\mathtt{MOD}_q$, but $\mathtt{mod}_2 \equiv \mathtt{MOD}_2$ and indeed the extension coincides with Lemma 3 when $q = 2$.

Lemma 4 (on the error rate of low degree polynomials over $\mathrm{GF}(p)$ that approximate $\mathtt{mod}_q$): *There exists a constant $\epsilon > 0$ such that, for any prime $p > q$, any n-variate polynomial $Q : \mathrm{GF}(p)^n \to \mathrm{GF}(p)$ of degree at most $\sqrt{n}$ fails to compute $\mathtt{mod}_q$ on at least $\epsilon \cdot 2^n$ of the n-long inputs; that is,*

$$\Pr_{x\in\{0,1\}^n}[Q(x) \neq \mathtt{mod}_q(x)] \geq \epsilon.$$

As stated above, there is no point in considering the case of $p < q$, because in that case no polynomial over $\mathrm{GF}(p)$ can approximate $\mathtt{mod}_q$ with error rate smaller than $(q-p-o(1))/q$, since $\Pr_{x\in\{0,1\}^n}[\mathtt{mod}_q(x) \in \{0,1...,p-1\}] \leq \frac{p}{q} + \frac{q}{2^n}$. Note that, while Lemma 4 asserts that low degree polynomials over $\mathrm{GF}(p)$ cannot approximate $\mathtt{mod}_q$ well, the contrast with Lemma 2 requires obtaining such a result for $\mathtt{MOD}_q$. We shall address this gap later.

Proof: Following the proof strategy of Lemma 3, letting $G \stackrel{\text{def}}{=} \{x \in \{0,1\}^n : Q(x) = \mathtt{mod}_q(x)\}$, we shall prove that G misses a constant fraction of $\{0,1\}^n$ by using Q to present a class of $|\mathcal{K}|^{(1-\Omega(1))\cdot 2^n}$ polynomials that can compute $|\mathcal{K}|^{|G|}$ different functions. We prove this assertion by extending the proof of Lemma 3. The extension is conceptually straightforward, but involves more technicalities than the special case of $q = 2$. Details follow.

Again, the crucial step is a variable substitution. Here, we map $x_i \in \{0,1\}$ to $\omega^{x_i} \in \{1,\omega\} \subset \mathcal{K}$, where ω is an arbitrary fixed element of order q in the multiplicative group of the extension field $\mathcal{K}$. (Indeed, for $q = 2$, it holds that $\omega = -1$ and $\mathcal{K} = \mathrm{GF}(p)$.) The benefit of this substitution is that it allows to relate $\mathtt{mod}_q(x_1,...,x_n)$ to $\prod_{i\in[n]} \omega^{x_i}$; that is, $\omega^{\mathtt{mod}_q(x_1,...,x_n)} = \prod_{i\in[n]} \omega^{x_i}$. Noting that $\mathtt{mod}_q(x)$ is in $\mathbb{Z}_q$, it is useful to consider the mapping $\zeta \mapsto \omega^\zeta$ as defined over $\mathbb{Z}_q$, and actually to extend it to $\mathcal{K}$. We note that this mapping as well as its inverse can be computed by degree $q-1$ polynomials over $\mathcal{K}$, denoted M and M'; that is, there exists degree $q-1$ polynomials $M, M' : \mathcal{K} \to \mathcal{K}$ such that $M(\zeta) = \omega^\zeta$ and $M'(M(\zeta)) = \zeta$ for every $\zeta \in \mathbb{Z}_q$.

Accordingly, we consider the polynomial $R : \mathcal{K}^n \to \mathcal{K}$ defined as $R(y_1, ..., y_n) \stackrel{\text{def}}{=} M(Q(x_1, ..., x_n))$, where $x_i = M'(y_i)$ (equiv. (for $x_i \in \{0,1\}$), $y_i = M(x_i)$), while noting that R has degree $(q-1)^2 \cdot \sqrt{n}$. Actually, defining R (over $\mathcal{K}$) requires viewing Q as a polynomial over $\mathcal{K}$, which means replacing the field operations of GF(p) by field operations of $\mathcal{K}$. The salient feature of the polynomial R is that $R(y_1, ..., y_n) = \prod_{i\in[n]} y_i$ holds whenever the corresponding $(x_1, ..., x_n) = (M'(y_1), ..., M'(y_n))$ is in G. This is the case because for every $(x_1, ..., x_n) \in \{0,1\}^n$ it holds that

$$\begin{aligned}\prod_{i\in[n]} M(x_i) &= \prod_{i\in[n]} \omega^{x_i} \\ &= \omega^{\mathtt{mod}_q(x_1,....,x_n)} \\ &= M(\mathtt{mod}_q(x_1, ..., x_n))\end{aligned}$$

whereas $(x_1, ..., x_n) \in G$ implies that $\mathtt{mod}_q(x_1, ..., x_n) = Q(x_1, ..., x_n)$. Hence, $\prod_{i\in[n]} M(x_i) = M(Q(x_1, ..., x_n))$ for every $(x_1, ..., x_n) \in G$, which means that $\prod_{i\in[n]} y_i = R(y_1, ..., y_n)$ for every $(y_1, ..., y_n) \in \{1, \omega\}^n$ such that $(M'(y_1), ..., M'(y_n)) \in G$.

Consequently, letting $H \stackrel{\text{def}}{=} \left\{y \in \{1,\omega\}^n : R(y) = \prod_{i\in[n]} y_i\right\}$, and observing that $|H| \geq |G|$, we seek to upper-bound $|H|$. The key fact that we shall use is that R has a magical feature: *It is a degree $(q-1)^2 \cdot \sqrt{n}$ polynomial that, when restricted to H, equals a degree n polynomial* (specifically, $\prod_{i\in[n]} y_i$).

Towards upper-bounding $|H|$, we consider the class $\mathcal{F}$ of all function $f : H \to \mathcal{K}$, and note that $|\mathcal{F}| = |\mathcal{K}|^{|H|}$. We first show that each $f \in \mathcal{F}$ can be written as a linear combination of multi-linear monomials. This is the case because, for any distinct $\alpha, \beta \in \mathcal{K}$, any function $g : \{\alpha, \beta\} \to \mathcal{K}$ can be written as a linear function; that is,

$$L_{\alpha,\beta,g}(\zeta) \stackrel{\text{def}}{=} \frac{\zeta - \beta}{\alpha - \beta} \cdot g(\alpha) + \frac{\zeta - \alpha}{\beta - \alpha} \cdot g(\beta) \tag{4}$$

satisfies $L_{\alpha,\beta,g}(\alpha) = g(\alpha)$ and $L_{\alpha,\beta,g}(\beta) = g(\beta)$. In particular, for any $e_1, ..., e_n \in \mathbb{Z}_q$, we can replace $\prod_{i\in[e]} y_i^{e_i}$ with $\prod_{i\in[e]} L_{1,\omega,g_{e_i}}(y_i)$, where $g_e(\zeta) = \zeta^e$, because we wish to preserve the value only over $(y_1, ..., y_n) \in \{1, \omega\}^n$. Hence, for every $y \in H$, it holds that

$$f(y) = \sum_{I\subseteq[n]} f_I \cdot \prod_{i\in I} y_i,$$

where the f_I's are in $\mathcal{K}$. Furthermore, for every $y \in H$, using $R(y) = \prod_{i\in[n]} y_i$ *we decrease the number of variables that appear in any monomial to at most* $(n + (q-1)^2 \cdot \sqrt{n})/2$. This is the case because

$$\prod_{i\in I} y_i = \left(\prod_{i\in[n]} y_i\right) \cdot \prod_{i\in[n]\setminus I} y_i^{q-1} = R(y) \cdot \prod_{i\in[n]\setminus I} y_i^{q-1},$$

where we use $y_i^q = 1$ for each $i \in I$ (while relying on $y_i \in \{1, \omega\}$), whereas $R(y) \cdot \prod_{i\in[n]\setminus I} y_i^{q-1}$ is a linear combination of monomials such that each contain at most $\sqrt{n} + (n - |I|)$ variables (and so, either $|I| \leq (n + (q-1)^2 \cdot \sqrt{n})/2$ or $\sqrt{n} + (n - |I|) < (n + (q-1)^2 \cdot \sqrt{n})/2$). Thus, letting $t \stackrel{\text{def}}{=} (n + (q-1)^2 \cdot \sqrt{n})/2$, we get

$$f(y) = \sum_{I\subseteq[n]:|I|\leq t} f_I \cdot \prod_{i\in I} y_i + \sum_{I\subseteq[n]:|I|>t} f_I \cdot \prod_{i\in I} y_i$$

$$= \sum_{I\subseteq[n]:|I|\leq t} f_I \cdot \prod_{i\in I} y_i + \sum_{I\subseteq[n]:|I|>t} f_I \cdot R(y) \cdot \prod_{i\in[n]\setminus I} y_i^{q-1}$$

which means that each $f \in \mathcal{F}$ can be represented as a linear combination of monomials such that each monomial contains at most $t = (n+(q-1)^2\cdot\sqrt{n})/2$ variables. Replacing powers of the various variables by the corresponding linear function (i.e., y_i^e is replaced by L_{1,ω,g_e}, where $g_e(\zeta) = \zeta^e$), it follows that each $f \in \mathcal{F}$ can be represented as a linear combination of multi-linear monomials of degree at most t. Noting that the number of such monomials is $N \stackrel{\text{def}}{=} \sum_{i=0}^{t}\binom{n}{i}$, we conclude that $|\mathcal{F}| \leq |\mathcal{K}|^N$, and the claim follows (because $N = (1-\exp(-\Omega(q^4)))\cdot 2^n = (1-\Omega(1))\cdot 2^n$). ■

Digest. The proof of Lemma 4 mimics the proof of Lemma 3, while replacing the field $\mathrm{GF}(p)$ by its $q-1$ dimensional extension $\mathcal{K}$ and replacing $-1 \in \mathrm{GF}(p)$ by $\omega \in \mathcal{K}$ (of multiplicative order q). We highlight two additional modifications:

1. Although the inputs to $\mathtt{mod}_q$ are bits, the mapping $\zeta \mapsto \omega^\zeta$ is defined over $\mathbb{Z}_q$ (in order to accommodate also the output value). This mapping and its inverse are performed by degree $(q-1)$ polynomials over $\mathcal{K}$, and consequently the degree of R is larger by a factor of $(q-1)^2$ than the degree of Q.

2. In two places, arbitrary powers of $y_i \in \{1,\omega\}$ are replaced by linear functions of y_i (presented in Eq. (4)). This replacement allows the continued pivoting of the argument on multi-linear (low degree) polynomials that compute functions in $\mathcal{F}$.

We stress that Lemma 4 refers to approximating $\mathtt{mod}_q : \{0,1\}^n \to \mathbb{Z}_n$, whereas the contrast with Lemma 2 (which allows establishing Part 2 of Theorem 1) refers to approximating $\mathtt{MOD}_q : \{0,1\}^n \to \{0,1\}$. This gap is addressed next.

Bridging the gap (between approximating $\mathtt{MOD}_q : \{0,1\}^n \to \{0,1\}$ and approximating $\mathtt{mod}_q : \{0,1\}^n \to \mathbb{Z}_q$). Recall that Lemma 4 asserts that low degree polynomials over $\mathrm{GF}(p)$ cannot approximate $\mathtt{mod}_q$ well; specifically, every degree $\sqrt{n}$ polynomial over $\mathrm{GF}(p)$ approximates $\mathtt{mod}_q$ with error rate $\Omega(1)$. However, the contrast with Lemma 2 requires obtaining such a result for $\mathtt{MOD}_q$ (i.e., showing that low degree polynomials over $\mathrm{GF}(p)$ have error rate $\Omega(1)$ with respect to $\mathtt{MOD}_q$). We prove the contrapositive: Starting with a polynomial (over $\mathrm{GF}(p)$) that approximates $\mathtt{MOD}_q$ with error rate $o(1)$, we present a polynomial of the same degree (over the same field) that approximates $\mathtt{mod}_q$ with error rate $o(1)$.

We first observe that computing $\mathtt{mod}_q : \{0,1\}^n \to \mathbb{Z}_q$ can be reduced to computing $\mathtt{MOD}_q : \{0,1\}^{n+q} \to \{0,1\}$. This is done by observing that

$$\mathtt{mod}_q(x) = \sum_{i\in[q-1]} (1 - \mathtt{MOD}_q(x1^{q-i}0^i))\cdot i,$$

which holds because $\mathtt{MOD}_q(x1^{q-i}0^i)) = 0$ if and only if $\mathtt{mod}_q(x1^{q-i}0^i) = 0$ (which holds if and only if $\mathtt{mod}_q(x) = i$). Analogously, an approximating polynomial Q for $\mathtt{MOD}_q$ can be converted to an approximating polynomial Q' for $\mathtt{mod}_q$; that is, $Q'(x) = \sum_{i\in[q-1]}(1-Q(x1^{q-i}0^i))\cdot i$. Note that although the resulting polynomial Q' preserves the degree of Q, it does does not preserve the error

rate of Q; yet, the error rate of Q' is at most a $(q-1)\cdot 2^q$ factor larger than the error rate of Q; that is,

$$\begin{aligned}\Pr_{x\in\{0,1\}^n}[Q'(x)\neq \mathtt{mod}_q(x)] &\leq \sum_{i\in[q-1]} \Pr_{x\in\{0,1\}^n}[Q(x1^{q-i}0^i)\neq \mathtt{MOD}_q(x1^{q-i}0^i)] \\ &\leq (q-1)\cdot \frac{\Pr_{z\in\{0,1\}^{n+q}}[Q(z)\neq \mathtt{MOD}_q(z)]}{2^{-q}}\end{aligned}$$

which we can tolerate (because q is a constant whereas the error rate of Q is $o(1)$).

3.2 The case of $q > p$

The hypothesis $q < p$ was (only) used in Section 3.1 in order to allow for an embedding of $\mathbb{Z}_q$ in $\mathrm{GF}(p)$. In addition, the association of $\mathbb{Z}_q$ with $\{0,1,...,q-1\}$ and of $\mathrm{GF}(p)$ with $\{0,1,...,p-1\}$ allowed for a straightforward embedding that was not even stated explicitly. Essentially, all that is needed when turning to the case of $q > p$ is to pick an integer $e > 1$ such that $q < p^e$ (e.g., $e = \lceil \log_p q \rceil$), and consider an embedding of $\mathbb{Z}_q$ in $\mathrm{GF}(p)^e$. Denoting the embedding by $\psi : \mathbb{Z}_q \to \mathrm{GF}(p)^e$, specific modifications to Section 3.1 include:

- In Lemma 4, we consider $Q : \mathrm{GF}(p)^n \to \mathrm{GF}(p)^e$, and state the hypothesis as

  $$\Pr_{x\in\{0,1\}^n}[Q(x)\neq \psi(\mathtt{mod}_q(x))] \geq \epsilon.$$

  Note that there is no need to apply ψ to the (binary) input; we only apply it to the desired output (which is in $\mathbb{Z}_q$).

  Similarly, we start the proof by defining $G \stackrel{\text{def}}{=} \{x\in\{0,1\}^n : Q(x)=\psi(\mathtt{mod}_q(x))\}$.

- In the proof of Lemma 4, we replace the mapping $M : \mathcal{K} \to \mathcal{K}$ and $M' : \mathcal{K} \to \mathcal{K}$ with the mappings $M : \mathcal{K}^e \to \mathcal{K}$ and $M' : \mathcal{K} \to \mathcal{K}$ such that $M(\psi(\zeta)) = \omega^\zeta$ for every $\zeta \in \mathbb{Z}_q$ and $M'(\omega^\zeta) = \zeta$ for every $\zeta \in \{0,1\}$. We stress that now M is computed by an e-variate polynomial of individual degree $p-1$ and M' is computed by a linear function.[10] We then define $R : \mathcal{K}^n \to \mathcal{K}$ such that $R(y_1,...,y_n) \stackrel{\text{def}}{=} M(Q(M'(y_1),...,M'(y_n)))$, while noting that R has degree $e\cdot(p-1)\cdot\sqrt{n}$, and observe that for $(x_1,...,x_n)\in G$ it holds that

  $$\begin{aligned}R(\omega^{x_1},...,\omega^{x_n}) &= M(Q(M'(\omega^{x_1}),...,M'(\omega^{x_n}))) \\ &= M(Q(x_1,...,x_n)) \\ &= M(\psi(\mathtt{mod}_q(x_1,...,x_n))) \\ &= \omega^{\mathtt{mod}_q(x_1,...,x_n)} \\ &= \prod_{i\in[n]} \omega^{x_i}.\end{aligned}$$

[10] Indeed, the definition of M' differs from the one used in the proof of Lemma 4, but we could have used the current definition also there (because we only refer to the value of M' on $\{1,\omega\}$). In contrast, an extension of the definition of M' that was used in the proof of Lemma 4 would let $M' : \mathcal{K} \to \mathcal{K}^e$ be computed by an e-long sequence of univariate polynomials of degree $q-1$ such that $M'(\omega^\zeta) = \psi(\zeta)$ for every $\zeta \in \mathbb{Z}_q$. (In that case, $R(y_1,...,y_n) \stackrel{\text{def}}{=} M(Q(M'(y_1),...,M'(y_n)))$, where Q operates on embeddings of elements of $\mathbb{Z}_q$ in $\mathcal{K}^e$, would have had degree $e\cdot(p-1)\cdot(q-1)\cdot\sqrt{n}$.

Letting $H \stackrel{\text{def}}{=} \left\{y \in \{1,\omega\}^n : R(y) = \prod_{i\in[n]} y_i\right\}$, and observing that $|H| \geq |G|$ (as before), we proceed exactly as in the second part of the proof (i.e., the part in which $|H|$ is upper-bounded).

- When bridging the gap between $\mathtt{MOD}_q : \{0,1\}^n \to \{0,1\}$ and $\mathtt{mod}_q : \{0,1\}^n \to \mathbb{Z}_q$, we define $Q'(x) = \sum_{i\in[q-1]}(1 - Q(x1^{q-i}0^i)) \cdot \psi(i) \in \mathrm{GF}(p)^e$.

We stress that the foregoing refers to a restating of Lemma 4 in which the approximation of $\mathtt{mod}_q$ is provided by an e-long sequence of (n-variant) polynomials over $\mathrm{GF}(p)$, where $e = \lceil \log_p q \rceil$. An alternative presentation, which avoids the distinction between the case of $q < p$ and the case of $q > p$, can be obtained by considering $\omega^{\mathtt{mod}_q(x)}$ as a representation of $\mathtt{mod}_q(x)$; this means starting with a (low degree) polynomial $Q : \mathcal{K}^n \to \mathcal{K}$ and lower-bounding $\Pr_{x\in\{0,1\}^n}[Q(x) \neq \omega^{\mathtt{mod}_q(x)}]$.

4 Beyond the recommended reading

Recall that $\mathtt{mod}_q$ was defined over $\{0,1\}^n$ and Lemma 4 refers to the error rate (w.r.t $\mathtt{mod}_q$) of any low-degree polynomial $Q : \mathrm{GF}(p)^n \to \mathrm{GF}(p)$; that is, the error rate is $\Pr_{x\in\{0,1\}^n}[Q(x) \neq \mathtt{mod}_q(x)]$. It is natural to extend $\mathtt{mod}_q$ over $\mathbb{Z}_q^n$ (i.e., let $\mathtt{mod}'_q : \mathbb{Z}_q^n \to \mathbb{Z}_q$ be defined as $\mathtt{mod}'_q(x) = \sum_{i\in[n]} x_i \bmod q$) and consider the error rate (w.r.t $\mathtt{mod}'_q$) of any low-degree polynomial $Q : \mathrm{GF}(p)^n \to \mathrm{GF}(p)$; that is, the error rate is $\Pr_{x\in\mathbb{Z}_q^n}[Q(x) \neq \mathtt{mod}'_q(x)]$.

As shown at the end of Section 4.1, approximating $\mathtt{mod}_q$ (by low degree polynomials over $\mathrm{GF}(p)$) is reducible to approximating $\mathtt{mod}'_q$ (by such polynomials). The converse holds too, and combining the converse reduction with a lower bound on the error rate of low degree polynomials wrt $\mathtt{mod}'_q$ yields an alternative proof of Lemma 4. The resulting proof is more complicated than the one presented in Section 3, but it has its own merits. Indeed, in this section, we shall prove that the error rate of any degree $\sqrt{n}$ polynomial over $\mathrm{GF}(p)$ wrt $\mathtt{mod}'_q$ is lower-bounded by a positive constant. The proof will extend and modify the proof that was presented in Section 3.

When trying to extend Lemma 4 to $\mathtt{mod}'_q$, we face (again) the fact the product of multi-linear monomials is not necessarily a multi-linear monomial. This difficulty was resolved in the proof of Lemma 4 by reducing all individual degrees to 1, while capitalizing on the fact that we only cared about the value of the corresponding univariate functions at two points (i.e., 1 and ω). This is no longer the case in the current context, because we care about the values of these univariate functions at all powers of ω. However, this is not a problem because we can consider all monomials of total degree at most t and individual degree at most $q-1$. Their number will be contrasted with the number of functions over the subset of $\mathbb{Z}_q^n$ on which Q agrees with $\mathtt{mod}'_q$.

Notation and organization: As in Section 3, fixing an arbitrary pair of primes, denoted $p \neq q$, we denote by $\mathcal{K}$ the $q-1$ dimensional extension field of $\mathrm{GF}(p)$, and denote by $\omega \in \mathcal{K}$ an arbitrary element of multiplicative order q. We distinguish again between the case of $q < p$, where we can embed $\mathbb{Z}_q$ in $\mathrm{GF}(p)$, and the case of $q > p$, where we embed $\mathbb{Z}_q$ in $\mathrm{GF}(p)^e$ for $e = \lceil \log_p q \rceil$.

4.1 The case of $q < p$

Recall that $\mathtt{mod}'_q : \mathbb{Z}_q^n \to \mathbb{Z}_q$ is defined by $\mathtt{mod}'_q(x_1, ..., x_n) \stackrel{\text{def}}{=} \sum_{i\in[n]} x_i \bmod q$.

Lemma 5 (on the error rate of low degree polynomials over GF(p) that approximate $\mathtt{mod}'_q$): *There exists a constant $\epsilon > 0$ such that, for any prime $p > q$, any n-variate polynomial $Q : \mathrm{GF}(p)^n \to \mathrm{GF}(p)$ of degree at most $\sqrt{n}$ fails to compute $\mathtt{mod}'_q$ on at least $\epsilon \cdot q^n$ of the n-long inputs; that is,*

$$\Pr_{x \in \{0,1,...,q-1\}^n}[Q(x) \neq \mathtt{mod}'_q(x)] \geq \epsilon.$$

Proof: Let $G \stackrel{\text{def}}{=} \{x \in \mathbb{Z}_q^n : Q(x) = \mathtt{mod}'_q(x)\}$ denote the set of inputs on which Q equals $\mathtt{mod}'_q$. We shall prove that G misses a constant fraction of $\mathbb{Z}_q^n$ by using Q to present a class of $|\mathcal{K}|^{(1-\Omega(1))\cdot q^n}$ polynomials that can compute $|\mathcal{K}|^{|G|}$ different functions. We prove this assertion by extending the proof of Lemma 4. The extension is conceptually straightforward, although it differs in its technical details.

Again, the crucial step is a variable substitution. Here, we map $x_i \in \mathbb{Z}_q$ to ω^{x_i}, noting that $\omega^{\mathtt{mod}'_q(x_1,...,x_n)} = \prod_{i\in[n]} \omega^{x_i}$. Note that the mapping $\zeta \mapsto \omega^\zeta$ as well as its inverse can be computed by degree $q-1$ polynomials over $\mathcal{K}$, denoted M and M'; that is, there exists degree $q-1$ polynomials $M, M' : \mathcal{K} \to \mathcal{K}$ such that $M(\zeta) = \omega^\zeta$ and $M'(M(\zeta)) = \zeta$ for every $\zeta \in \mathbb{Z}_q$. Accordingly, we consider the polynomial $R : \mathcal{K}^n \to \mathcal{K}$ defined as $R(y_1, ..., y_n) \stackrel{\text{def}}{=} M(Q(x_1, ..., x_n))$, where $x_i = M'(y_i)$ (resp., $y_i = M(x_i)$ when $x_i \in \mathbb{Z}_q \subset \mathrm{GF}(p)$). (Again, defining R (over $\mathcal{K}$) requires viewing Q as a polynomial over $\mathcal{K}$, which means replacing the field operations of GF(p) by field operations of $\mathcal{K}$.) Note that although G is defined differently than in the proof of Lemma 4, the definitions of M, M' and R (and the following feature of R) remain intact.

The salient feature of the degree $(q-1)^2 \cdot \sqrt{n}$ polynomial R is that $R(y_1, ..., y_n) = \prod_{i\in[n]} y_i$ holds whenever the corresponding $(x_1, ..., x_n)$ is in G (i.e., $(M'(y_1), ..., M'(y_n)) \in G$).[11] Hence, $\prod_{i\in[n]} M(x_i)$ equals $M(Q(x_1, ..., x_n))$ for every $(x_1, ..., x_n) \in G$, which means that $\prod_{i\in[n]} y_i = R(y_1, ..., y_n)$ for every $(y_1, ..., y_n) \in \{\omega^e : e \in \mathbb{Z}_q\}^n$ such that $(M'(y_1), ..., M'(y_n)) \in G$. Thus, letting $H \stackrel{\text{def}}{=} \left\{y \in \{\omega^e : e \in \mathbb{Z}_q\}^n : R(y) = \prod_{i\in[n]} y_i\right\}$, and observing that $|H| \geq |G|$, we seek to upper-bound $|H|$. The key fact that we shall use is that R has a magical feature: *It is a polynomial of degree at most $(q-1)^2 \cdot \sqrt{n}$ that, when restricted to H, equals a degree n polynomial* (specifically, $\prod_{i\in[n]} y_i$). Indeed, the foregoing mimics the proof of Lemma 4, and the deviation comes in the next paragraph.

Towards upper-bounding $|H|$, we consider the class $\mathcal{F}$ of all function $f : H \to \mathcal{K}$, and note that $|\mathcal{F}| = |\mathcal{K}|^{|H|}$. We first observe that each $f \in \mathcal{F}$ can be written as a linear combination of monomials of *individual degree at most* $q-1$, because $\sigma^q = 1$ for every $\sigma \in \{\omega^e : e \in \mathbb{Z}_q\}$. Furthermore, for every $y \in H$, using $R(y) = \prod_{i\in[n]} y_i$ *we shall decrease the total degree of any monomial to at most* $t \stackrel{\text{def}}{=} ((q-1)\cdot n + 2(q-1)^3 \cdot \sqrt{n})/2$, by multiplication with a small power of R. Specifically, consider an arbitrary monomial $\prod_{i\in[n]} y_i^{e_i}$, where all e_i's are in $\mathbb{Z}_q$. Then, there exists $j \in \mathbb{Z}_q$ such that $\sum_{i\in[n]}(e_i + j \bmod q) \leq (q-1)\cdot n/2$, where the latter sum is over the integers, because

[11] As in the proof of Lemma 4, this is the case because for every $(x_1, ..., x_n) \in \mathbb{Z}_q^n$ it holds that

$$\begin{aligned}\prod_{i\in[n]} M(x_i) &= \prod_{i\in[n]} \omega^{x_i} \\ &= \omega^{\mathtt{mod}'_q(x_1,....,x_n)} \\ &= M(\mathtt{mod}'_q(x_1, ..., x_n))\end{aligned}$$

whereas $(x_1, ..., x_n) \in G$ implies $\mathtt{mod}'_q(x_1, ..., x_n) = Q(x_1, ..., x_n)$.

$\mathrm{E}_{j\in\mathbb{Z}_q}[e_i + j \bmod q] = \mathrm{E}_{j\in\mathbb{Z}_q}[j] = (q-1)/2$. Using this $j \in \mathbb{Z}_p$ and writing

$$\left(\prod_{i\in[n]} y_i\right)^j \cdot \left(\prod_{i\in[n]} y_i^{e_i}\right) \;=\; \prod_{i\in[n]} y_i^{e_i+j \bmod q} \tag{5}$$

it follows that the r.h.s of Eq. (5) has total degree at most $(q-1)\cdot n/2$. Hence, using suitable $j_{\overline{e}}$'s in $\mathbb{Z}_q$ (i.e., for every $\overline{e} = (e_1, ..., e_n)$, we use $j_{\overline{e}} \in \mathbb{Z}_q$ such that $\sum_{i\in[n]}(e_i + j_{\overline{e}} \bmod q) \leq (q-1)\cdot n/2$, where the latter sum is over the integers), we can write any $f \in \mathcal{F}$ as

$$\begin{aligned} f(y) &= \sum_{\overline{e}=(e_1,...,e_n)\in\mathbb{Z}_q^n} f_{\overline{e}} \cdot \prod_{i\in[n]} y_i^{e_i} \\ &= \sum_{\overline{e}=(e_1,...,e_n)\in\mathbb{Z}_q^n} f_{\overline{e}} \cdot \left(\prod_{i\in[n]} y_i\right)^{q-j_{\overline{e}} \bmod q} \cdot \prod_{i\in[n]} y_i^{e_i+j_{\overline{e}} \bmod q} \\ &= \sum_{\overline{e}=(e_1,...,e_n)\in\mathbb{Z}_q^n} f_{\overline{e}} \cdot R(y)^{q-j_{\overline{e}} \bmod q} \cdot \prod_{i\in[n]} y_i^{e_i+j_{\overline{e}} \bmod q} \end{aligned}$$

where the $f_{\overline{e}}$'s are in $\mathcal{K}$. We note that $R(y)^{q-j_{\overline{e}} \bmod q} \cdot \prod_{i\in[n]} y_i^{e_i+j_{\overline{e}} \bmod q}$ is a linear combination of monomials of total degree at most $(q-1)\cdot\deg(R) + ((q-1)\cdot n/2) \leq t$ and individual degree at most $q-1$, where the later fact uses $y_i^q = 1$ for $y_i \in \{\omega^e : e \in \mathbb{Z}_q\}$. This means that each $f \in \mathcal{F}$ can be represented as a linear combination of monomials of total degree at most t and individual degree at most $q-1$. The number of such monomials is $(1 - \exp(-O(q^4)))\cdot q^n$, because the probability that the sum of n independent random variables that are uniformly and independently distributed in $\mathbb{Z}_q$ exceeds $t = 0.5(q-1)n + (q-1)^3\sqrt{n}$ is $\exp(-O(q^4))$. It follows that $|\mathcal{F}| \leq |\mathcal{K}|^{(1-\Omega(1))\cdot q^n}$. ■

Reducing the approximation of $\mathtt{mod}'_q : \mathbb{Z}_q^n \to \mathbb{Z}_q$ to approximating $\mathtt{mod}_q : \{0,1\}^n \to \mathbb{Z}_q$. The reduction is quite straightforward: Essentially, we may $\mathbb{Z}_q$ to $(q-1)$-bit long strings such that $w \in \mathbb{Z}_q$ is mapped to a string of Hamming weight w. That is, we use the unary encoding $U : \mathbb{Z}_q \to \{0,1\}^{q-1}$ such that $U(\sigma) = 1^\sigma 0^{q-1-\sigma}$, while noting that this encoding can be computed by an $(q-1)$-long sequence of degree $q-1$ univariate polynomials over $\mathrm{GF}(p)$, where $\mathbb{Z}_q \subseteq \mathrm{GF}(p)$ relies on $q < p$. Using this encoding, we can reduce computing $\mathtt{mod}'_q : \mathbb{Z}_q^n \to \mathbb{Z}_q$ to computing $\mathtt{mod}_q : \{0,1\}^{(q-1)\cdot n} \to \mathbb{Z}_q$ by observing that

$$\mathtt{mod}'_q(x_1, ..., x_n) = \mathtt{mod}_q(U(x_1), ..., U(x_n)).$$

Analogously, an approximating polynomial Q for $\mathtt{mod}_q$ can be converted to an approximating polynomial Q' for $\mathtt{mod}'_q$; that is, $Q'(x) = Q(U(x_1), ..., U(x_n))$. Indeed, the degree of Q' is $q-1$ times larger than the degree of Q, but the real problem is that the error rate is not preserved; in fact, the error rate may increase by a factor of $(2^{q-1}/q)^n$, which we cannot afford.

Thus, an additional idea is needed. Specifically, letting $n' = (q-1)\cdot n$, we use a random bijection $\Pi : \{0,1\}^{n'} \to \{0,1\}^{n'}$ that preserves the Hamming weight of the n'-bit long string; that is $\Pi(z_1, ..., z_{n'}) = (z_{\pi(1)}, ..., z_{\pi(n')})$, where π is a random permutation of $[n']$. Note that if $X = (X_1, ..., X_n)$ is uniformly distributed in $\mathbb{Z}_q^n$, then $\Pi(U(X_1), ..., U(X_n))$ is $o(1)$-close to be uniformly distributed in $\{0,1\}^{n'}$, because the total variation distance between the number of 1's

in $(U(X_1),...,U(X_n))$ (which equals $\sum_{i\in[n]} X_i$) and the number of 1's in a uniformly distributed n'-bit long string vanishes with n. Hence,

$$\begin{aligned}
&\Pr_{(x_1,...,x_n)\in\mathbb{Z}_q^n,\pi\in\mathrm{Sym}_{n'}}[Q(\Pi(U(x_1),...,U(x_n)))\neq \mathtt{mod}'_q(x_1,....,x_n)]\\
&= \Pr_{(x_1,...,x_n)\in\mathbb{Z}_q^n,\pi\in\mathrm{Sym}_{n'}}[Q(\Pi(U(x_1),...,U(x_n)))\neq \mathtt{mod}_q(U(x_1),....,U(x_n))]\\
&\leq \Pr_{(z_1,...,z_{n'})\in\{0,1\}^{n'}}[Q(z_1,...,z_{n'})\neq \mathtt{mod}_q(z_1,....,z_{n'})]+o(1)
\end{aligned}$$

and it follows that there exists a permutation π such that the corresponding (routing permutation) Π_π (i.e., $\Pi_\pi(z_1,....,z_{n'})=(z_{\pi(1)},....,z_{\pi(n')})$) satisfies

$$\begin{aligned}
&\Pr_{(x_1,...,x_n)\in\mathbb{Z}_q^n}[Q(\Pi_\pi(U(x_1),...,U(x_n)))\neq \mathtt{mod}'_q(x_1,....,x_n)]\\
&\leq \Pr_{(z_1,...,z_{n'})\in\{0,1\}^{n'}}[Q(z_1,...,z_{n'})\neq \mathtt{mod}_q(z_1,....,z_{n'})]+o(1).
\end{aligned}$$

Using this π, we redefine $Q'(x)\stackrel{\mathrm{def}}{=}Q(\Pi_\pi(U(x_1),...,U(x_n)))$, and obtain the desired approximating polynomial (while observing that Π_π only permutes variables).

Reducing the approximation of $\mathtt{mod}_q:\{0,1\}^n\to\mathbb{Z}_q$ to the approximating $\mathtt{mod}'_q:\mathbb{Z}_q^n\to\mathbb{Z}_q$. We first present a randomized reduction of computing $\mathtt{mod}_q$ to approximating $\mathtt{mod}'_q$. Essentially, on input $x=(x_1,...,x_n)\in\{0,1\}^n$, the reduction selects a random $r=(r_1,...,r_n)\in\mathbb{Z}_q^n$, and outputs the sum (mod q) of $q-\mathtt{mod}'_q(r)$ and $\mathtt{mod}'_q(r+x \bmod q)$, where $(r+x \bmod q)=((r_1+x_1 \bmod q),...,(r_n+x_n \bmod q))$. Analogously, an approximating polynomial Q for $\mathtt{mod}'_q$ can be converted to a "randomized polynomial" Q_r that computes $\mathtt{mod}_q$ correctly (w.h.p.) on each input. Specifically, observing that the (2-argument) addition (modulo q) can be computed by a bivariate polynomial of individual degree $q-1$, denoted A, and letting $Q_r(x)\stackrel{\mathrm{def}}{=}A(q-\mathtt{mod}'_q(r),Q(A(r_1,x_1),...,A(r_n,x_n)))$, *for every* $x\in\{0,1\}^n$, we have

$$\begin{aligned}
\Pr_{r\in\mathbb{Z}_q^n}[Q_r(x)=\mathtt{mod}_q(x)] &= \Pr_{r\in\mathbb{Z}_q^n}[Q(r+x \bmod q)=\mathtt{mod}_q(r+x \bmod q)]\\
&= \Pr_{r\in\mathbb{Z}_q^n}\left[Q(r)=\mathtt{mod}'_q(r)\right],
\end{aligned}$$

which means that (for every x) the randomized polynomial Q_r (in which $q-\mathtt{mod}'_q(r)$ is a constant) computes $\mathtt{mod}_q(x)$ with error probability that equals the error rate of Q w.r.t $\mathtt{mod}'_q:\mathbb{Z}_q^n\to\mathbb{Z}_q$. Hence,

$$\begin{aligned}
\mathrm{E}_{r\in\mathbb{Z}_q^n}\left[\Pr_{x\in\{0,1\}^n}[Q_r(x)=\mathtt{mod}_q(x)]\right] &= \mathrm{E}_{x\in\{0,1\}^n}\left[\Pr_{r\in\mathbb{Z}_q^n}[Q_r(x)=\mathtt{mod}_q(x)]\right]\\
&= \Pr_{r\in\mathbb{Z}_q^n}\left[Q(r)=\mathtt{mod}'_q(r)\right].
\end{aligned}$$

It follows that there exists $r\in\mathbb{Z}_q^n$ such that the error rate of Q_r w.r.t $\mathtt{mod}_q:\{0,1\}^n\to\mathbb{Z}_q$ (i.e., $\Pr_{x\in\{0,1\}^n}[Q_r(x)\neq\mathtt{mod}_q(x)]$) is upper-bounded by the error rate of Q w.r.t $\mathtt{mod}'_q:\mathbb{Z}_q^n\to\mathbb{Z}_q$. Observing that the degree of Q_r is $(q-1)^2$ times the degree of Q, the claim follows.

4.2 The case of $q>p$

As in Section 3, the hypothesis $q<p$ was (only) used in Section 4.1 in order to allow for an embedding of $\mathbb{Z}_q$ in GF(p). In addition, the association of $\mathbb{Z}_q$ with $\{0,1,...,q-1\}$ and of GF(p) with $\{0,1,...,p-1\}$ allowed for a straightforward embedding that was not even stated explicitly.

Essentially, all that is needed when turning to the case of $q > p$ is to pick an integer $e > 1$ such that $q < p^e$, and consider an embedding of $\mathbb{Z}_q$ in $\mathrm{GF}(p)^e$. As in Section 3.2, denoting the embedding by $\psi : \mathbb{Z}_q \to \mathrm{GF}(p)^e$, specific modifications to Section 4.1 include:

- In Lemma 5, we consider $Q : \mathrm{GF}(p)^{en} \to \mathrm{GF}(p)^e$, and state the hypothesis as

  $$\Pr_{x \in \mathbb{Z}_q^n}[Q(\psi(x)) \neq \psi(\mathtt{mod}_q(x))] \geq \epsilon$$

  where $\psi(x_1, ..., x_n) = (\psi(x_1), ..., \psi(x_n))$ and $\mathtt{mod}_q : \mathbb{Z}_q^n \to \mathbb{Z}_q$ remains intact.

  Similarly, we start the proof by defining $G \stackrel{\mathrm{def}}{=} \{x \in \mathbb{Z}_q^n : Q(\psi(x)) = \psi(\mathtt{mod}_q(x))\}$

- In the proof of Lemma 5, we use the mappings $M : \mathcal{K}^e \to \mathcal{K}$ and $M' : \mathcal{K} \to \mathcal{K}^e$ such that $M(\psi(\zeta)) = \omega^\zeta$ and $M'(\omega^\zeta) = \psi(\zeta)$ for every $\zeta \in \mathbb{Z}_q$, while noting that now M is computed by an e-variate polynomial of individual degree $p - 1$ and M' is computed by an e-long sequence of univariate polynomials of degree $q - 1$. We then define $R : \mathcal{K}^n \to \mathcal{K}$ such that $R(y_1, ..., y_n) \stackrel{\mathrm{def}}{=} M(Q(M'(y_1), ..., M'(y_n)))$, while noting that R has degree $e \cdot (p-1) \cdot (q-1) \cdot \sqrt{n}$, and observe that for $(x_1, ..., x_n) \in G$ it holds that

  $$\begin{aligned} R(\omega^{x_1}, ..., \omega^{x_n}) &= M(Q(M'(\omega^{x_1}), ..., M'(\omega^{x_n}))) \\ &= M(Q(\psi(x_1), ..., \psi(x_n))) \\ &= M(\psi(\mathtt{mod}'_q(x_1, ..., x_n))) \\ &= \omega^{\mathtt{mod}'_q(x_1, ..., x_n)} \\ &= \prod_{i \in [n]} \omega^{x_i}. \end{aligned}$$

  Letting $H \stackrel{\mathrm{def}}{=} \left\{ y \in \{\omega^e : e \in \mathbb{Z}_q\}^n : R(y) = \prod_{i \in [n]} y_i \right\}$, and observing that $|H| \geq |G|$ (as before), we proceed exactly as in the second part of the proof.

- When reducing $\mathtt{mod}'_q$ to $\mathtt{mod}_q$ we use the same unary encoding $U : \mathbb{Z}_q \to \{0,1\}^{q-1}$, but compute it as a function over $\mathrm{GF}(p)^e$; that is, when defining Q', we use $U'(\psi(\zeta)) = U(\zeta)$, for $\zeta \in \mathbb{Z}_q$. Specifically, we compute the individual bits of $U(\zeta)$ by using e-variate polynomials that act on $\psi(\zeta) \in \mathrm{GF}(p)^e$, where $\zeta \in \mathbb{Z}_q$. In other words, for $x \in \mathbb{Z}_q^n$, we define the polynomial $Q'(\psi(x)) \stackrel{\mathrm{def}}{=} Q(\Pi_\pi(U'(\psi(x_1)), ..., U'(\psi(x_n))))$, and note that its value equals $Q(\Pi_\pi(U(x_1), ..., U(x_n)))$.

- When reducing $\mathtt{mod}_q$ to $\mathtt{mod}'_q$, we are given $Q : \mathrm{GF}(p)^{en} \to \mathrm{GF}(p)^e$ that approximates the value of $\mathtt{mod}'_q : \mathbb{Z}_q^n \to \mathbb{Z}_q$ (embedded in $\mathrm{GF}(p)^e$ using ψ), and derive $Q_r : \mathrm{GF}(p)^n \to \mathrm{GF}(p)^e$ that approximates the value of $\mathtt{mod}_q : \{0,1\}^n \to \mathbb{Z}_q$. Here, for $r \in \mathbb{Z}_q^n$, we define $Q_r(x) \stackrel{\mathrm{def}}{=} A(\psi(q - \mathtt{mod}'_q(r)), Q(A(\psi(r_1), \psi(x_1)), ..., A(\psi(r_n), \psi(x_n))))$, where $A : \mathrm{GF}(p)^e \times \mathrm{GF}(p)^e \to \mathrm{GF}(p)^e$ is an e-long sequence of e-variate polynomials that compute (a representation of) addition mod q.

Appendix: Low degree polynomials and approximating Majority

Recall that $\mathtt{TH}^n_k : \{0,1\}^n \to \{0,1\}$ denotes the function that return 1 if and only if its input contains at least k ones (i.e., $\mathtt{TH}^n_k(x) = 1$ iff $\mathrm{wt}(x) \geq k$), and that $\mathtt{TH}^n_k(x) = \mathtt{TH}^{2n+1}_{n+1}(x1^{n+1-k}0^k)$, where $\mathtt{TH}^{2n+1}_{n+1}$ is the $(2n+1)$-bit Majority function. Hence, for any $t \in [n]$, a lower bound for $(2n+1)$-bit Majority follows from a lower bound for $\mathtt{TH}^n_t$. Letting $t(n) \stackrel{\mathrm{def}}{=} \lceil (n+\sqrt{n})/2 \rceil$, we prove that low degree polynomials over GF(2) cannot approximate $\mathtt{TH}^n_{t(n)}$ well.

Lemma 6 (on the error rate of low degree polynomials over GF(2) that approximate $\mathtt{TH}^n_{t(n)}$): *Any n-variate polynomial $Q : \mathrm{GF}(2)^n \to \mathrm{GF}(2)$ of degree smaller than $\sqrt{n}$ fails to compute $\mathtt{TH}^n_{t(n)}$ on $\Omega(2^n/\sqrt{n})$ of the n-bit inputs; that is,*

$$\Pr_{x \in \{0,1\}^n}[Q(x) \neq \mathtt{TH}^n_t(x)] = \Omega(1/\sqrt{n}).$$

Combining (or rather contrasting) Lemma 6 with Lemma 2 establishes Part 1 of Theorem 1 for the special case of $p = 2$.

Proof: Writing $x \leq s$ if for every $i \in [n]$ it holds that $x_i \leq s_i$, we first observe that *for every $s \in \{0,1\}^n$ such that $\mathrm{wt}(s) \geq \sqrt{n}$ it holds that* $\sum_{x \in \{0,1\}^n : x \leq s} Q(x) \equiv 0 \pmod 2$. The claim holds by considering each monomial of Q, and observing that for $I \subseteq [n]$ of size smaller than $\sqrt{n}$ it holds that

$$\begin{aligned} \sum_{x \in \{0,1\}^n : x \leq s} \prod_{i \in I} x_i &= |\{x \in \{0,1\}^n : (x \leq s) \wedge (\forall i \in I)\, x_i = 1\}| \\ &= \begin{cases} 2^{\mathrm{wt}(s)-|I|} & \text{if } I \subseteq \{i \in [n] : s_i = 1\} \\ 0 & \text{otherwise (i.e., } I \cap \{i \in [n] : s_i = 0\} \neq \emptyset) \end{cases} \end{aligned}$$

whereas $\mathrm{wt}(s) - |I| \geq 1$ (because $\mathrm{wt}(s) \geq \sqrt{n}$ and $|I| < \sqrt{n}$).

Letting $W_t \stackrel{\mathrm{def}}{=} \{x \in \{0,1\}^n : \mathrm{wt}(x) = t(n)\}$ and $D \stackrel{\mathrm{def}}{=} \{x \in \{0,1\}^n : Q(x) \neq \mathtt{TH}^n_{t(n)}\}$, we consider a Boolean matrix with rows corresponding to W_t and columns corresponding to D such that the $(r,c)^{\mathrm{th}}$ entry equals $\chi(r \geq c)$, where χ is a 0-1 indicator that equals 1 if and only if the condition holds. We consider the rank of this matrix over GF(2). Specifically, for every $r \in W_t$, we shall show that there exists a linear combination of the columns that yields a unit vector with 1 in row r. In particular, consider $D_r \stackrel{\mathrm{def}}{=} \{c \in D : c \leq r\}$. Then, for every $r' \in W_t$, it holds that

$$\begin{aligned} \sum_{c \in D_r} \chi(r' \geq c) &= \sum_{c \in D} \chi((c \leq r) \wedge (c \leq r')) \\ &= \sum_{x \leq r \wedge r'} \chi(x \in D) \end{aligned}$$

where $(r_1, ..., r_n) \wedge (r'_1, ..., r'_n) = (r_1 \wedge r'_1, ..., r_n \wedge r'_n)$. Observing that $x \in D$ (i.e., $Q(x) \neq \mathtt{TH}^n_{t(n)}(x)$) holds if and only if $Q(x) + \mathtt{TH}^n_{t(n)}(x) \equiv 1 \pmod 2$, we have

$$\begin{aligned} \sum_{x \leq r \wedge r'} \chi(x \in D) &\equiv \sum_{x \leq r \wedge r'} (Q(x) + \mathtt{TH}^n_{t(n)}(x)) \pmod 2 \\ &\equiv \sum_{x \leq r \wedge r'} Q(x) + \sum_{x \leq r \wedge r'} \mathtt{TH}^n_{t(n)}(x) \pmod 2. \end{aligned}$$

Noting that $\mathrm{wt}(r \wedge r') \geq \mathrm{wt}(r) + \mathrm{wt}(r') - n = 2 \cdot t(n) - n \geq \sqrt{n}$ and using the initial claim (with $s = r \wedge r'$), it follows that (mod 2) the first sum equals 0. On the other hand, the second sum equals 0 if $r' \neq r$ (because then $\mathrm{wt}(r \wedge r') < t(n)$) and equals 1 otherwise (due to the contribution of the term associated with $x = r$).

Hence, we showed that, for every $r \in W_t$, there exists a linear combination of the columns that yields a unit vector with 1 in row r. It follows that the matrix has rank at least $|W_t|$, which in turn implies that $|D| \geq |W_t| = \binom{n}{t(n)} = \Theta(2^n/\sqrt{n})$. The lemma follows. ■

Acknowledgements

I am grateful to Avishay Tal for extremely helpful comments.

References

[1] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[2] O. Goldreich. *Computational Complexity: A Conceptual Perspective*. Cambridge University Press, 2008.

[3] S. Jukna. *Boolean Function Complexity: Advances and Frontiers*. Algorithms and Combinatorics, Vol. 27, Springer, 2012.

[4] R. Smolensky. Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity. In *19th ACM Symposium on the Theory of Computing*, pages 77–82, 1987.

[5] A. Razborov. Lower bounds on the size of bounded-depth networks over a complete basis with logical addition. In *Matematicheskie Zametki*, Vol. 41, No. 4, pages 598–607, 1987 (in Russian). English translation in *Mathematical Notes of the Academy of Sci. of the USSR*, Vol. 41 (4), pages 333–338, 1987.