Non-abelian combinatorics and communication complexity

Emanuele Viola*

July 20, 2019

Finite groups provide an amazing wealth of problems of interest to complexity theory. And complexity theory also provides a useful viewpoint of group-theoretic notions, such as what it means for a group to be "far from abelian." The general problem that we consider in this survey is that of computing a *group product* $g = x_1 \cdot x_2 \cdot \cdots \cdot x_t$ over a finite group G. Several variants of this problem are considered in this survey and in the literature, including in [KMR66, Mix89, BC92, IL95, BGKL03, PRS97, Amb96, AL00, Raz00, MV13, Mil14, GV].

Some specific, natural computational problems related to g are, from hardest to easiest:

(1) Computing g,

(2) Deciding if $g = 1_G$, where 1_G is the identity element of G, and

(3) Deciding if $g = 1_G$ under the promise that either $g = 1_G$ or $g = h$ for a fixed $h \neq 1_G$.

Problem (3) is from [MV13]. The focus of this survey is on (2) and (3).

We work in the model of *communication complexity* [Yao79], with which we assume familiarity. For background see [KN97, RY19]. Briefly, the terms x_i in a product $x_1 \cdot x_2 \cdot \cdots \cdot x_t$ will be partitioned among collaborating parties – in several ways – and we shall bound the number of bits that the parties need to exchange to solve the problem.

Organization.

We begin in Section 1 with two-party communication complexity. In Section 2 we give a streamlined proof, except for a step that is only sketched, of a result of Gowers and the author [GV] about interleaved group products. In particular we present an alternative proof, communicated to us by Will Sawin, of a lemma from [GV]. We then consider two models of multi-party communication. In Sections 3 and 4 we consider number-in-hand protocols, and we relate the communication complexity to *mixing in quasirandom groups* [Gow08, BNP08]. In Section 5 we consider number-on-forehead protocols. We briefly discuss interleaved group products and the corresponding result of Gowers and the author [GV]. Then we consider the problem of separating deterministic and randomized communication. In Section 6 we give an exposition of a result by Austin [Aus16], and show that it implies a separation that matches the state-of-the-art [BDPW10] but applies to a different problem.

Some of the sections follow closely a set of lectures by the author [Vio17]; related material can also be found in the blog posts [Vio16a, Vio16b]. One of the goals of this survey is to present this material in a more organized matter, in addition to including new material. The text is interspersed with open problems; some are seemingly within reach, others are major and long-standing.

*Supported by NSF CCF award 1813930.

1 Two parties

Let G be a group and let us start by considering the following basic communication task. Alice gets an element $x \in G$ and Bob gets an element $y \in G$ and their goal is to check if $x \cdot y = 1_G$. How much communication do they need? Well, $x \cdot y = 1_G$ is equivalent to $x = y^{-1}$. Because Bob can compute y^{-1} without communication, this problem is just a rephrasing of the *equality* problem, which has a randomized protocol with constant communication. This holds for any group.

The same is true if Alice gets two elements x_1 and x_2 and they need to check if $x_1 \cdot y \cdot x_2 = 1_G$. Indeed, it is just checking equality of y and $x_1^{-1} \cdot x_2^{-1}$, and again Alice can compute the latter without communication.

Things get more interesting if both Alice and Bob get two elements and they need to check if the *interleaved product* of the elements of Alice and Bob equals 1_G, that is, if

$$x_1 \cdot y_1 \cdot x_2 \cdot y_2 = 1_G.$$

Now the previous transformations don't help anymore. In fact, the complexity depends on the group. If it is abelian then the elements can be reordered and the problem is equivalent to checking if $(x_1 \cdot x_2) \cdot (y_1 \cdot y_2) = 1_G$. Again, Alice can compute $x_1 \cdot x_2$ without communication, and Bob can compute $y_1 \cdot y_2$ without communication. So this is the same problem as before and it has a constant communication protocol.

For non-abelian groups this reordering cannot be done, and the problem seems hard. This can be formalized for certain groups that are "far from abelian" – or we can take this result as a definition of being far from abelian. One of the groups that works best in this sense is the following, first constructed by Galois in the 1830's.

Definition 1. The *special linear group* $SL(2, q)$ is the group of 2×2 invertible matrices over the field $\mathbb{F}_q$ with determinant 1.

The following result was asked in [MV13] and was proved in [GV].

Theorem 2. *Let $G = SL(2, q)$ and let $h \neq 1_G$. Suppose Alice receives $x_1, x_2 \in G$ and Bob receives $y_1, y_2 \in G$. They are promised that $x_1 \cdot y_1 \cdot x_2 \cdot y_2$ either equals 1_G or h. Deciding which case it is requires randomized communication $\Omega(\log |G|)$.*

This bound is tight as Alice can send her input, taking $O(\log |G|)$ bits. Omitting a step, we present the proof of this theorem in the next section.

Similar results are known for any *simple* group, see [GV] and [Sha16]. One such group that is "between" abelian groups and $SL(2, q)$ is the following.

Definition 3. The *alternating group* A_n is the group of even permutations of $1, 2, \ldots, n$.

If we work over A_n instead of $SL(2, q)$ in Theorem 2 then the communication complexity is $\Omega(\log \log |G|)$ [Sha16]. The latter bound is tight [MV13]: with knowledge of h, the parties can agree on an element $a \in \{1, 2, \ldots, n\}$ such that $h(a) \neq a$. Hence they only need to keep track of the image a. This takes communication $O(\log n) = O(\log \log |A_n|)$ because $|A_n| = n!/2$. In more detail, the protocol is as follows. First Bob sends $y_2(a)$. Then Alice sends $x_2 y_2(a)$. Then Bob sends $y_1 x_2 y_2(a)$ and finally Alice can check if $x_1 y_1 x_2 y_2(a) = a$.

Interestingly, to decide if $g = 1_G$ without the promise a stronger lower bound can be proved for many groups, including A_n, see Corollary 7 below.

Theorem 2 and the corresponding results for other groups also scale with the length of the product: for example deciding if $x_1 \cdot y_1 \cdot x_2 \cdot y_2 \cdots x_t \cdot y_t = 1_G$ over $G = SL(2, q)$ requires communication $\Omega(t \log |G|)$ which is tight. This is stated formally for the more general setting of multiparty communication in Theorem 18 below.

Problem 4. Understand for which groups Theorem 2 applies. In particular, is the communication large for every quasirandom group [Gow08]?

A strength of the above lower bounds is that they hold for any choice of h in the promise. Moreover, they in fact hold even if the parties only achieve a slight advantage over random guessing. This makes them equivalent to certain *mixing* results, discussed below in Section 4.1. Next we prove other lower bounds that do not hold for any choice of h and can be obtained by reduction from *disjointness*. First we show that for any non-abelian group G there exists an element h such that deciding if $g = 1_G$ or $g = h$ requires communication linear in the length of the product. Interestingly, the proof works for any non-abelian group. The choice of h is critical, as for some non-abelian G and h the problem is easy. For example: take any group G and consider $H := G \times \mathbb{Z}_2$ where $\mathbb{Z}_2$ is the group of integers with addition modulo 2. Distinguishing between $1_H = (1_G, 0)$ and $h = (1_G, 1)$ amounts to computing the parity of (the $\mathbb{Z}_2$ components of) the input, which takes constant communication.

Theorem 5. *Let G be a non-abelian group. There exists $h \in G$ such that the following holds. Suppose Alice receives $x_1, x_2, \ldots, x_t$ and receives $y_1, y_2, \ldots, y_t$. They are promised that $x_1 \cdot y_1 \cdot x_2 \cdot y_2 \cdot \cdots \cdot x_t \cdot y_t$ either equals 1_G or h. Deciding which case it is requires randomized communication $\Omega(t)$.*

Proof. We reduce from *unique set-disjointness*, defined below. For the reduction we encode the And of two bits $s, t \in \{0, 1\}$ as a group product. This encoding is similar to the famous puzzle that asks to hang a picture on a wall with two nails in such a way that the picture falls if either nail is removed. Since G is non-abelian, there exist $a, b \in G$ such that $a \cdot b \neq b \cdot a$, and in particular $a \cdot b \cdot a^{-1} \cdot b^{-1} = h$ with $h \neq 1$. We can use this fact to encode the And of s and t as

$$a^s \cdot b^t \cdot a^{-s} \cdot b^{-t} = \begin{cases} 1 & \text{if } \mathrm{And}(s, t) = 0 \\ h & \text{otherwise} \end{cases} .$$

In the disjointness problem Alice and Bob get inputs $x, y \in \{0, 1\}^t$ respectively, and they wish to check if there exists an $i \in [t]$ such that $x_i \wedge y_i = 1$. If you think of x, y as characteristic vectors of sets, this problem is asking if the sets have a common element or not. The communication of this problem is $\Omega(t)$ [KS92, Raz92]. Moreover, in the "unique" variant of this problem where the number of such i's is 0 or 1, the same lower bound $\Omega(t)$ still applies. This follows from [KS92, Raz92] – see also Proposition 3.3 in [AMS99]. For more on disjointness see the surveys [She14, CP10].

We will reduce unique disjointness to group products. For $x, y \in \{0, 1\}^t$ we produce inputs for the group problem as follows:

$$x \to (a^{x_1}, a^{-x_1}, \ldots, a^{x_t}, a^{-x_t})$$
$$y \to (b^{y_1}, b^{-y_1}, \ldots, b^{y_t}, b^{-y_t}).$$

The group product becomes

$$\underbrace{a^{x_1} \cdot b^{y_1} \cdot a^{-x_1} \cdot b^{-y_1}}_{1 \text{ bit}} \cdot \cdots \cdot a^{x_t} \cdot b^{y_t} \cdot a^{-x_t} \cdot b^{-y_t}.$$

If there isn't an $i \in [t]$ such that $x_i \wedge y_i = 1$, then for each i the term $a^{x_i} \cdot b^{y_i} \cdot a^{-x_i} \cdot b^{-y_i}$ is 1_G, and thus the whole product is 1.

Otherwise, there exists a unique i such that $x_i \wedge y_i = 1$ and thus the product will be $1 \cdots 1 \cdot h \cdot 1 \cdots 1 = h$, with h being in the i-th position. If Alice and Bob can check if the above product is equal to 1, they can also solve the unique set disjointness problem, and thus the lower bound applies for the former. □

We required the uniqueness property, because otherwise we might get a product h^c that could be equal to 1 in some groups.

Next we prove a result for products of length just 4; it applies to non-abelian groups of the form $G = H^n$ and not with the promise.

Theorem 6. *Let H be a non-abelian group and consider $G = H^n$. Suppose Alice receives x_1, x_2 and Bob receives y_1, y_2. Deciding if $x_1 \cdot y_1 \cdot x_2 \cdot y_2 = 1_G$ requires randomized communication $\Omega(n)$.*

Proof. The proof is similar to the proof of Theorem 5. We use coordinate i of G to encode bit i of the disjointness instance. If there is no intersection in the latter, the product will be 1_G. Otherwise, at least some coordinate will be $\neq 1_G$. □

As a corollary we can prove a lower bound for A_n.

Corollary 7. *Theorem 6 holds for $G = A_n$.*

Proof. Note that A_n contains $(A_4)^{\lfloor n/4 \rfloor}$ and that A_4 is not abelian. □

Theorem 6 is tight for constant-size H.

Problem 8. Is Corollary 7 tight? The trivial upper bound is $O(\log |A_n|) = O(n \log n)$.

2 Proof of Theorem 2

Several related proofs of this theorem exist, see [GV15, GV, Sha16]. As in [GV], the proof that we present can be broken down in three steps. First we reduce the problem to a statement about conjugacy classes. Second we reduce this to a statement about trace maps. Third we prove the latter. We present the first step in a way that is similar but slightly different from the presentation in [GV]. The second step is only sketched, but relies on classical results about $SL(2, q)$ and can be found in [GV]. For the third we present a proof that was communicated to us by Will Sawin. We thank him for his permission to include it here.

2.1 Step 1

We would like to rule out randomized protocols, but it is hard to reason about them directly. Instead, we are going to rule out deterministic protocols on random inputs. First, for any group element $g \in G$ we define the distribution on quadruples $D_g := (x_1, y_1, x_2, (x_1 \cdot y_1 \cdot x_2)^{-1} g)$, where $x_1, y_1, x_2 \in G$ are uniformly random elements. Note the product of the elements in D_g is always g.

Towards a contradiction, suppose we have a randomized protocol P such that

$$\mathbb{P}[P(D_1) = 1] \geq \mathbb{P}[P(D_h) = 1] + \frac{1}{10}.$$

This implies a deterministic protocol with the same gap, by fixing the randomness.

We reach a contradiction by showing that for every deterministic protocol P using little communication, we have

$$|\Pr[P(D_1) = 1] - \Pr[P(D_h) = 1]| \leq \frac{1}{100}.$$

We start with the following standard lemma, which describes a protocol using product sets.

Lemma 9. *(The set of accepted inputs of) A deterministic c-bit protocol for a function $f : X \times Y \to Z$ can be written as a disjoint union of 2^c rectangles, where a rectangle is a product set of the form $A \times B$ with $A \subseteq X$ and $B \subseteq Y$ and where f is constant.*

Proof. (sketch) For every communication transcript t, let $S_t \subseteq G^2$ be the set of inputs giving transcript t. The sets S_t are disjoint since an input gives only one transcript, and their number is 2^c: one for each communication transcript of the protocol. The rectangle property can be proven by induction on the protocol tree. □

Next, we show that any rectangle $A \times B$ cannot distinguish D_1, D_h. The way we achieve this is by showing that the probability that $(A \times B)(D_g) = 1$ is roughly the same for every g, and is roughly the density of the rectangle. (We write $A \times B$ for the characteristic function of the set $A \times B$.) Without loss of generality we set $g = 1_G$. Let A have density α and B have density β. We aim to bound above

$$\left|\mathbb{E}_{a_1,b_1,a_2,b_2:a_1b_1a_2b_2=1} A(a_1, a_2)B(b_1, b_2) - \alpha\beta\right|,$$

where note the distribution of a_1, b_1, a_2, b_2 is the same as D_1.

Because the distribution of (b_1, b_2) is uniform in G^2, the above can be rewritten as

$$\begin{aligned}
&\left|\mathbb{E}_{b_1,b_2} B(b_1, b_2)\mathbb{E}_{a_1,a_2:a_1b_1a_2b_2=1}(A(a_1, a_2) - \alpha)\right| \\
&\leq \sqrt{\mathbb{E}_{b_1,b_2} B(b_1, b_2)^2}\sqrt{\mathbb{E}_{b_1,b_2}\mathbb{E}^2_{a_1,a_2:a_1b_1a_2b_2=1}(A(a_1, a_2) - \alpha)}. \\
&= \sqrt{\beta}\sqrt{\mathbb{E}_{b_1,b_2,a_1,a_2,a_1',a_2':a_1b_1a_2b_2=a_1'b_1a_2'b_2=1} A(a_1, a_2)A(a_1', a_2') - \alpha^2}.
\end{aligned}$$

The inequality is Cauchy-Schwarz, and the step after that is obtained by expanding the square and noting that (a_1, a_2) is uniform in G^2, so that the expectation of the term $A(a_1, a_2)\alpha$ is α^2.

Now we do several transformations to rewrite the distribution in the last expectation in a convenient form. First, right-multiplying by b_2^{-1} we can rewrite the distribution as the uniform distribution on tuples such that

$$a_1b_1a_2 = a_1'b_1a_2'.$$

The last equation is equivalent to $b_1^{-1}(a_1')^{-1}a_1b_1a_2 = a_2'$.

We can now set a_1' to be a_1x^{-1} to rewrite the distribution of the four-tuple as

$$(a_1, a_2, a_1x^{-1}, C(x)a_2)$$

where we use $C(x)$ to denote a uniform element from the conjugacy class of x, that is $b^{-1}xb$ for a uniform $b \in G$.

Hence it is sufficient to bound

$$\left|\mathbb{E}A(a_1, a_2)A(a_1x^{-1}, C(x)a_2) - \alpha^2\right|,$$

where all the variables are uniform and independent.

With a similar derivation as above, this can be rewritten as

$$\begin{aligned}
&\left|\mathbb{E}A(a_1, a_2)\mathbb{E}(A(a_1x^{-1}, C(x)a_2) - \alpha)\right| \\
&\leq \sqrt{\mathbb{E}A(a_1, a_2)^2}\sqrt{\mathbb{E}_{a_1,a_2}\mathbb{E}_x^2(A(a_1x^{-1}, C(x)a_2) - \alpha)} \\
&= \sqrt{\alpha}\sqrt{\mathbb{E}A(a_1x^{-1}, C(x)a_2)A(a_1x'^{-1}, C(x')a_2) - \alpha^2}.
\end{aligned}$$

Here each occurrence of C denotes a uniform and independent conjugate. Hence it is sufficient to bound

$$\left|\mathbb{E}A(a_1x^{-1}, C(x)a_2)A(a_1x'^{-1}, C(x')a_2) - \alpha^2\right|.$$

We can now replace a_2 with $C(x)^{-1}a_2$. Because $C(x)^{-1}$ has the same distribution of $C(x^{-1})$, it is sufficient to bound

$$\left|\mathbb{E}A(a_1x^{-1}, a_2)A(a_1x'^{-1}, C(x')C(x^{-1})a_2) - \alpha^2\right|.$$

For this, it is enough to show that with high probability $1 - 1/|G|^{\Omega(1)}$ over x and x', the distribution of $C(x')C(x^{-1})$, over the choice of the two independent conjugates, has statistical distance $\leq 1/|G|^{\Omega(1)}$ from uniform.

2.2 Step 2

In this step we use information on the conjugacy classes of the group to reduce the latter task to one about the equidistribution of the trace map. Let Tr be the Trace map:

$$Tr\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} = a_1 + a_4.$$

We state the lemma that we want to show.

Lemma 10. *Let* $a := \begin{pmatrix} 0 & 1 \\ 1 & w \end{pmatrix}$ *and* $b := \begin{pmatrix} v & 1 \\ 1 & 0 \end{pmatrix}$. *For all but* $O(1)$ *values of* $w \in \mathbb{F}_q$ *and* $v \in \mathbb{F}_q$, *the distribution of*

$$Tr\left(au^{-1}bu\right)$$

is $O(1/q)$ *close to uniform over* $\mathbb{F}_q$ *in statistical distance.*

To give some context, in $SL(2,q)$ the conjugacy class of an element is essentially determined by the trace. Moreover, we can think of a and b as generic elements in G. So the lemma can be interpreted as saying that for typical $a, b \in G$, taking a uniform element from the conjugacy class of b and multiplying it by a yields an element whose conjugacy class is uniform among the classes of G. Using that essentially all conjugacy classes are equal, and some of the properties of the trace map, one can show that the above lemma implies that for typical x, x' the distribution of $C(x')C(x^{-1})$ is close to uniform. For more on how this fits we refer the reader to [GV].

2.3 Step 3

We now present a proof of Lemma 10. The high-level argument of the proof is the same as in [GV] (Lemma 5.5), but the details may be more accessible and in particular the use of the Lang-Weil theorem [LW54] from algebraic geometry is replaced by a more elementary argument. For simplicity we shall only cover the case where q is prime. We will show that for all but $O(1)$ values of $v, w, c \in \mathbb{F}_q$, the probability over u that $Tr(au^{-1}bu) = c$ is within $O(1/q^2)$ of $1/q$, and for the others it is at most $O(1/q)$. Summing over c gives the result.

We shall consider elements b whose trace is unique to the conjugacy class of b. (This holds for all but $O(1)$ conjugacy classes – see for example [GV] for details.) This means that the distribution of $u^{-1}bu$ is that of a uniform element in G conditioned on having trace b. Hence, we can write the probability that $Tr(au^{-1}bu) = c$ as the number of solutions in x to the following three equations (divided by the size of the group, which is $q^3 - q$):

$$\begin{aligned} x_3 + x_2 + wx_4 &= c && (Tr(ax) = c), \\ x_1 + x_4 &= v && (Tr(x) = Tr(b)), \\ x_1x_4 - x_3x_3 &= 1 && (Det(x) = 1). \end{aligned}$$

We use the second one to remove x_1 and the first one to remove x_2 from the last equation. This gives

$$(v - x_4)x_4 - (c - x_3 - wx_4)x_3 = 1.$$

This is an equation in two variables. Write $x = x_3$ and $y = x_4$ and use distributivity to rewrite the equation as

$$-y^2 + vy - cx + x^2 + wxy = 1.$$

At least since Lagrange it has been known how to reduce this to a Pell equation $x^2 + dy^2 = e$. This is done by applying an invertible affine transformation, which does not change the number of solutions. First set $x = x - wy/2$. Then the equation becomes

$$-y^2 + vy - c(x - wy/2) + (x - wy/2)^2 + w(x - wy/2)y = 1.$$

Equivalently, the cross-term has disappeared and we have

$$y^2(-1 - w^2/4) + y(v + cw/2) + x^2 - cx = 1.$$

Now one can add constants to x and y to remove the linear terms, changing the constant term. Specifically, let $h := (v + cw/2)/2$ and set $y = y - h$ and $x = x + c/2$. The equation becomes

$$(y - h)^2(-1 - w^2/4) + (y - h)2h + (x + c/2)^2 - c(x + c/2) = 1.$$

The linear terms disappear, the coefficients of x^2 and y^2 do not change and the equation can be rewritten as

$$y^2(-1-w^2/4)+h^2(-1-w^2/4)-2h^2+x^2+(c/2)^2-c^2/2=1.$$

So this is now a Pell equation

$$x^2+dy^2=e$$

where $d:=(-1-w^2/4)$ and

$$e:=1+h^2(3+w^2/4)+(c/2)^2=1+(v^2+(cw/2)^2+cvw)(1/4)(3+w^2/4)+(c/2)^2.$$

For all but $O(1)$ values of w we have that d is non-zero. Moreover, for all but $O(1)$ values of v,w the term e is a non-zero polynomial in c. (Specifically, for any $v\neq 0$ and any w such that $3+w^2/4\neq 0$.) So we only consider the values of c that make it non-zero. Those where $e=0$ give $O(q)$ solutions, which is fine. We conclude with the following lemma.

Lemma 11. *For d and e non-zero, and prime q, the number of solutions over $\mathbb{F}_q$ to the Pell equation*

$$x^2+dy^2=e$$

is within $O(1)$ of q.

This is a basic result from algebraic geometry that can be proved from first principles.

Proof. If $d=-f^2$ for some $f\in\mathbb{F}_q$, then we can replace y with fy and we can count instead the solutions to the equation

$$x^2-y^2=e.$$

Because $x^2-y^2=(x-y)(x+y)$ we can set $x':=x-y$ and $y':=x+y$, which preserves the number of solutions, and rewrite the equation as

$$x'y'=e.$$

Because $e\neq 0$, this has $q-1$ solutions: for every non-zero y' we have $x'=e/y'$.

So now we can assume that $d\neq -f^2$ for any $f\in\mathbb{F}_q$. Because the number of squares is $(q+1)/2$, the range of x^2 has size $(q+1)/2$. Similarly, the range of $e-dy^2$ also has size $(q+1)/2$. Hence these two ranges intersect, and there is a solution (a,b).

We take a line passing through (a,b): for parameters $s,t\in\mathbb{F}$ we consider pairs $(a+t,b+st)$. There is a bijection between such pairs with $t\neq 0$ and the points (x,y) with $x\neq a$. Because the number of solutions with $x=a$ is $O(1)$, using that $d\neq 0$, it suffices to count the solutions with $t\neq 0$.

The intuition is that this line has two intersections with the curve $x^2+dy^2=e$. Because one of them, (a,b), lies in $\mathbb{F}_q$, the other has to lie as well there. Algebraically, we can plug the pair in the expression to obtain the equivalent equation

$$a^2+t^2+2at+d(b^2+s^2t^2+2bst)=e.$$

Using that (a,b) is a solution this becomes

$$t^2+2at+ds^2t^2+2dbst=0$$

We can divide by $t \neq 0$. Obtaining

$$t(1 + ds^2) + 2a + 2dbs = 0.$$

We can now divide by $1 + ds^2$ which is non-zero by the assumption $d \neq -f^2$. This yields

$$t = (-2a - 2dbs)/(1 + ds^2).$$

Hence for every value of s there is a unique t giving a solution. This gives q solutions. □

3 Number-in-hand

In this section we consider the following three-party number-in-hand problem: Alice gets $x \in G$, Bob gets $y \in G$, Charlie gets $z \in G$, and they want to know if $x \cdot y \cdot z = 1_G$. The communication depends on the group G. We present next two efficient protocols for abelian groups, and then a communication lower bound for other groups.

3.1 A randomized protocol for the hypercube

We begin with the simplest setting. Let $G = (\mathbb{Z}_2)^n$, that is n-bit strings with bit-wise addition modulo 2. The parties want to check if $x + y + z = 0^n$. They can do so as follows. First, they pick a hash function h that is linear: $h(x + y) = h(x) + h(y)$. Specifically, for a uniformly random $a \in \{0, 1\}^n$ define $h_a(x) := \sum a_i x_i \mod 2$. Then, the protocol is as follows.

- Alice sends $h_a(x)$,
- Bob send $h_a(y)$,
- Charlie accepts if and only if $h_a(x) + h_a(y) + h_a(z) = 0$.

The hash function outputs 1 bit, so the communication is constant. By linearity, the protocol accepts iff $h_a(x + y + z) = 0$. If $x + y + z = 0$ this is always the case, otherwise it happens with probability $1/2$.

3.2 A randomized protocol for $\mathbb{Z}_N$

This protocol is from [Vio14]. For simplicity we only consider the case $N = 2^n$ here – the protocol for general N is in [Vio14]. Again, the parties want to check if $x+y+z = 0 \mod N$. For this group, there is no 100% linear hash function but there are almost linear hash functions $h : \mathbb{Z}_N \to \mathbb{Z}_{2^\ell}$ that satisfy the following properties. Note that the inputs to h are interpreted modulo N and the outputs modulo 2^ℓ.

1. for all a, x, y there is $c \in \{0, 1\}$ such that $h_a(x + y) = h_a(x) + h_a(y) + c$,
2. for all $x \neq 0$ we have $\mathbb{P}_a[h_a(x) \in \{-2, -1, 0, 1, 2\}] \leq O(1/2^\ell)$,
3. $h_a(0) = 0$.

Assuming some random hash function h that satisfies the above properties the protocol works similarly to the previous one:

- Alice sends $h_a(x)$,
- Bob sends $h_a(y)$,
- Charlie accepts if and only if $h_a(x) + h_a(y) + h_a(z) \in \{-2, -1, 0\}$.

We set $\ell = O(1)$ to achieve constant communication. To prove correctness of the protocol, first note that $h_a(x) + h_a(y) + h_a(z) = h_a(x + y + z) - c$ for some $c \in \{0, 1, 2\}$. Then consider the following two cases:

- if $x + y + z = 0$ then $h_a(x + y + z) - c = h_a(0) - c = -c$, and the protocol is always correct.
- if $x+y+z \neq 0$ then the probability that $h_a(x+y+z)-c \in \{-2, -1, 0\}$ for some $c \in \{0, 1, 2\}$ is at most the probability that $h_a(x + y + z) \in \{-2, -1, 0, 1, 2\}$, which is $\leq 2^{-\Omega(\ell)}$; so the protocol is correct with high probability.

The hash function..

For the hash function we can use one analyzed in [DHKP97]. Let a be a random odd number modulo 2^n. Define

$$h_a(x) := (a \cdot x \gg n - \ell) \mod 2^\ell$$

where the product $a \cdot x$ is integer multiplication, and $\gg$ is bit-shift. In other words we output the bits $n - \ell + 1, n - \ell + 2, \ldots, n$ of the integer product $a \cdot x$.

We now verify that the above hash function family satisfies the three properties we required above.

Property (3) is trivially satisfied.

For property (1), notice that if in the addition $a \cdot x + a \cdot y$ the carry into the $n - \ell + 1$ bit is 0 then

$$(a \cdot x \gg n - \ell) + (a \cdot y \gg n - \ell) = (a \cdot x + a \cdot y) \gg n - \ell;$$

otherwise

$$(a \cdot x \gg n - \ell) + (a \cdot y \gg n - \ell) + 1 = (a \cdot x + a \cdot y) \gg n - \ell.$$

This concludes the proof for property (1).

Finally, we prove property (2). We start by writing $x = s \cdot 2^c$ where s is odd. So the binary representation of x looks like

$$(\cdots\cdots 1 \underbrace{0 \cdots 0}_{c \text{ bits}}).$$

The binary representation of the product $a \cdot x$ for a uniformly random a looks like

$$(\textit{uniform}\ 1 \underbrace{0 \cdots 0}_{c \text{ bits}}).$$

We consider the two following cases for the product $a \cdot x$:

1. If $a \cdot x = (\underbrace{\textit{uniform}\ 1}_{\ell \textit{ bits}} \overbrace{00}^{2 \textit{ bits}} \cdots 0)$, or equivalently $c \geq n - \ell + 2$, the output never lands in the bad set $\{-2, -1, 0, 1, 2\}$;
2. Otherwise, the hash function output has $\ell - O(1)$ uniform bits. For any set B, the probability that the output lands in B is at most $|B| \cdot 2^{-\ell+O(1)}$.

3.3 Quasirandom groups

What happens in other groups? The hash function used in the previous result was fairly non-trivial. Do we have an almost linear hash function for 2×2 matrices? The answer is negative. For $SL_2(q)$ and A_n the problem is hard, even under the promise. For a group G the complexity can be expressed in terms of a parameter d which comes from representation theory. We will not formally define this parameter here, but several qualitatively equivalent formulations can be found in [Gow08]. Instead the following table shows the d's for the groups we've introduced.

G	:	abelian	A_n	$SL_2(q)$
d	:	1	$\Omega(\frac{\log \lvert G\rvert}{\log\log \lvert G\rvert})$	$\lvert G\rvert^{\Omega(1)}$

Theorem 12. *Let G be a group, and let $h \neq 1_G$. Let d be the minimum dimension of any irreducible representation of G. Suppose Alice, Bob, and Charlie receive x, y, and z respectively. They are promised that $x \cdot y \cdot z$ either equals 1_G or h. Deciding which case it is requires randomized communication complexity $\Omega(\log d)$.*

This result is tight for the groups we have discussed so far. The arguments are the same as before. Specifically, for $SL_2(q)$ the communication is $\Omega(\log |G|)$. This is tight up to constants, because Alice and Bob can send their elements. For A_n the communication is $\Omega(\log\log |G|)$. This is tight as well, as the parties can again just communicate the images of an element a such that $h(a) \neq a$, as discussed in Section 2. This also gives a computational proof that d cannot be too large for A_n, i.e., it is at most $(\log |G|)^{O(1)}$. For abelian groups we get nothing, matching the efficient protocols given above.

Problem 13. Is Theorem 12 tight for every group? For the upper bounds, can we always find a protocol that works even without the promise?

4 Proof of Theorem 12

First we discuss several "mixing" lemmas for groups, then we come back to protocols and see how to apply one of them there.

4.1 mixing

We want to consider "high entropy" distributions over G, and state a fact showing that the multiplication of two such distributions "mixes" or in other words increases the entropy. To define entropy we use the norms $\|A\|_c = (\sum_x A(x)^c)^{\frac{1}{c}}$. Our notion of (non-)entropy will be $\|A\|_2$. Note that $\|A\|_2^2$ is exactly the *collision probability* $\mathbb{P}[A = A']$ where A' is independent and identically distributed to A. The smaller this quantity, the higher the entropy of A. For the uniform distribution U we have $\|U\|_2^2 = 1/|G|$ and so we can think of $1/|G|$ as maximum entropy. If A is uniform over $\Omega(|G|)$ elements, we have $\|A\|_2^2 = O(1/|G|)$ and we think of A as having "high" entropy.

Because the entropy of U is small, we can think of the distance between A and U in the 2-norm

as being essentially the entropy of A:

$$\begin{aligned}\|A-U\|_2^2 &= \sum_{x\in G}\left(A(x)-\frac{1}{|G|}\right)^2 \\ &= \sum_{x\in G} A(x)^2 - 2A(x)\frac{1}{|G|}+\frac{1}{|G|^2} \\ &= \|A\|_2^2 - \frac{1}{|G|} \\ &= \|A\|_2^2 - \|U\|_2^2 \\ &\approx \|A\|_2^2.\end{aligned}$$

Lemma 14. *[Gow08, BNP08] If X, Y are independent over G, then*

$$\|X\cdot Y - U\|_2 \le \|X\|_2\|Y\|_2\sqrt{\frac{|G|}{d}},$$

where d is the minimum dimension of an irreducible representation of G.

By this lemma, for high entropy distributions X and Y, we get $\|X\cdot Y-U\|_2 \le \frac{O(1)}{\sqrt{|G|d}}$. The factor $1/\sqrt{|G|}$ allows us to pass to *statistical distance* $\|.\|_1$ using Cauchy-Schwarz:

$$\|X\cdot Y-U\|_1 \le \sqrt{|G|}\|X\cdot Y-U\|_2 \le \frac{O(1)}{\sqrt{d}}. \tag{1}$$

This is the way in which we will use the lemma.

Another useful consequence of this lemma, which however we will not use directly, is this. Suppose now you have *three* independent, high-entropy variables X, Y, Z. Then for every $g \in G$ we have

$$|\mathbb{P}[X\cdot Y\cdot Z = g] - 1/|G|| \le \|X\|_2\|Y\|_2\|Z\|_2\sqrt{\frac{|G|}{d}}. \tag{2}$$

To show this, set $g = 1_G$ without loss of generality and rewrite the left-hand-side as

$$|\sum_{h\in G}\mathbb{P}[X=h](\mathbb{P}[YZ=h^{-1}]-1/|G|)|.$$

By Cauchy-Schwarz this is at most

$$\sqrt{\sum_h \mathbb{P}^2[X=h]}\sqrt{\sum_h(\mathbb{P}[YZ=h^{-1}]-1/|G|)^2} = \|X\|_2\|YZ-U\|_2$$

and we can conclude by Lemma 14. Hence the product of three high-entropy distributions is close to uniform in a point-wise sense: each group element is obtained with probability roughly $1/|G|$.

At least over $SL(2,q)$, there exists an alternative proof of this fact that does not mention representation theory (see [GV] and [Vio16a, Vio16b]).

With this notation in hand, we conclude by stating a "mixing" version of Theorem 1. For more on this perspective we refer the reader to [GV].

Theorem 15. *Let $G = SL(2, q)$. Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ be two distributions over G^2. Suppose X is independent from Y. Let $g \in G$. We have*

$$|\mathbb{P}[X_1 Y_1 X_2 Y_2 = g] - 1/|G|| \leq |G|^{1-\Omega(1)} \|X\|_2 \|Y\|_2.$$

For example, when X and Y have high entropy over G^2 (that is, are uniform over $\Omega(|G|^2)$ pairs), we have $\|X\|_2 \leq \sqrt{O(1)/|G|^2}$, and so $|G|^{1-\Omega(1)}\|X\|_2\|Y\|_2 \leq 1/|G|^{1+\Omega(1)}$. In particular, $X_1Y_1X_2Y_2$ is $1/|G|^{\Omega(1)}$ close to uniform over G in statistical distance.

4.2 Back to protocols

As in the beginning of Section 2, for any group element $g \in G$ we define the distribution on triples $D_g := (x, y, (x \cdot y)^{-1} g)$, where $x, y \in G$ are uniform and independent. Note that the product of the elements in D_g is always g. Again as in Section 2, it suffices to show that for every *deterministic* protocols P using little communication we have

$$|\Pr[P(D_1) = 1] - \Pr[P(D_h) = 1]| \leq \frac{1}{100}.$$

Analogously to Lemma 9, the following lemma describes a protocol using rectangles. The proof is nearly identical and is omitted.

Lemma 16. *(The set of accepted inputs of) A deterministic c-bit number-in-hand protocol with three parties can be written as a disjoint union of 2^c rectangles, that is product sets of the form $A \times B \times C$.*

Next we show that these product sets cannot distinguish these two distributions D_1, D_h, via a straightforward application of lemma 14.

Lemma 17. *For all $A, B, C \subseteq G$ we have $|\mathbb{P}(A \times B \times C)(D_1) = 1] - \mathbb{P}[(A \times B \times C)(D_h) = 1]| \leq 1/d^{\Omega(1)}$.*

Proof. For any $h \in G$ we have

$$\mathbb{P}[(A \times B \times C)(D_h) = 1] = \mathbb{P}[(x, y) \in A \times B] \cdot \mathbb{P}[(x \cdot y)^{-1} \cdot h \in C | (x, y) \in A \times B], \tag{3}$$

where (x, y) is uniform in G^2. If either A or B is small, that is $\mathbb{P}[x \in A] \leq \epsilon$ or $\mathbb{P}[y \in B] \leq \epsilon$, then also $\mathbb{P}[(x, y) \in A \times B] \leq \epsilon$ and hence (3) is at most ϵ as well. This holds for every h, so we also have $|\mathbb{P}(A \times B \times C)(D_1) = 1] - \mathbb{P}[(A \times B \times C)(D_h) = 1]| \leq \epsilon$. We will choose ϵ later.

Otherwise, A and B are large: $\mathbb{P}[x \in A] > \epsilon$ and $\mathbb{P}[y \in B] > \epsilon$. Let (x', y') be the distribution of (x, y) conditioned on $(x, y) \in A \times B$. We have that x' and y' are independent and each is uniform over at least $\epsilon|G|$ elements. By Lemma 14 this implies $\|x' \cdot y' - U\|_2 \leq \|x'\|_2 \cdot \|y'\|_2 \cdot \sqrt{\frac{|G|}{d}}$, where U is the uniform distribution. As mentioned after the lemma, by Cauchy–Schwarz we obtain

$$\|x' \cdot y' - U\|_1 \leq |G| \cdot \|x'\|_2 \cdot \|y'\|_2 \cdot \sqrt{\frac{1}{d}} \leq \frac{1}{\epsilon} \cdot \frac{1}{\sqrt{d}},$$

where the last inequality follows from the fact that $\|x\|_2, \|y\|_2 \leq \sqrt{\frac{1}{\epsilon|G|}}$.

This implies that $\|(x' \cdot y')^{-1} - U\|_1 \leq \frac{1}{\epsilon} \cdot \frac{1}{\sqrt{d}}$ and $\|(x' \cdot y')^{-1} \cdot h - U\|_1 \leq \frac{1}{\epsilon} \cdot \frac{1}{\sqrt{d}}$, because taking inverses and multiplying by h does not change the distance to uniform. These two last inequalities imply that

$$|\mathbb{P}[(x' \cdot y')^{-1} \in C] - \mathbb{P}[(x' \cdot y')^{-1} \cdot h \in C]| \leq O(\frac{1}{\epsilon\sqrt{d}});$$

and thus we get that

$$|\mathbb{P}[(A \times B \times C)(D_1) = 1] - \mathbb{P}[(A \times B \times C)(D_h) = 1]| \leq O(\frac{1}{\epsilon\sqrt{d}}).$$

Picking $\epsilon = 1/d^{1/4}$ completes the proof. □

Returning to arbitrary deterministic protocols P (as opposed to rectangles), write P as a union of 2^c disjoint rectangles by Lemma 16. Applying Lemma 17 and summing over all rectangles we get that the distinguishing advantage of P is at most $2^c/d^{1/4}$. For $c \leq (1/100)\log d$ the advantage is at most $1/100$, concluding the proof.

5 Number-on-forehead

In number-on-forehead (NOH) communication complexity [CFL83] with k parties, the input is a k-tuple $(x_1, \ldots, x_k)$ and each party i sees all of it except x_i. The seminal work [BNS92] proved lower bounds when k is logarithmic in the input length, and it remains a major open problem to prove lower bounds for larger k.

Theorem 2 can be extended to the multiparty setting as follows (enabling an application in cryptography from [MV13], where the question is asked).

Theorem 18. *Let $G = SL(2, q)$ and let $h \neq 1_G$. Let M be a $k \times t$ matrix of elements of G. Suppose party $i = 1, 2, \ldots, k$ knows all rows except row i. Let g be the product of the entries of M column by column: $g = \prod_{j=1}^{t} M_{j,1} M_{j,2} \cdots M_{j,k}$. Suppose the parties are promised that g either equals 1_G or h. Deciding which case it is requires randomized communication $\geq t/b^{2^k} \log |G|$ for all $t \geq b^{2^k}$ where b is a universal constant.*

In particular, for every fixed k the communication is $\Omega(\log |G|)$, which is tight. The lower bound above is only meaningful when k is doubly-logarithmic in t.

Problem 19. Is there a lower bound with k logarithmic in t? More ambitiously, is there a lower bound with k super-logarithmic in t? We know of no non-trivial upper bounds. For $k = 3$, what is the smallest t for which lower bounds hold? The proof in [GV] gives a large constant for t. Does $t = 2$ suffice?

The proof of Theorem 18 is obtained from the following mixing result. Let X be a probability distribution on G^m such that any two coordinates are uniform in G^2. Then, a pointwise product of s independent copies of X is nearly uniform in G^m, where s depends on m only. Again, such a result is false for abelian groups. For more on this proof we refer the reader to [GV].

In the remainder of this survey we shall instead focus on the problem of separating deterministic and randomized communication. For $k = 2$, we know the optimal separation: The equality function requires $\Omega(n)$ communication for deterministic protocols, but can be solved using $O(1)$

communication if we allow the protocols to use randomness. (Randomness is public, here and elsewhere in this survey.) For $k = 3$, the best known separation between deterministic and randomized protocol is $\Omega(\log n)$ vs $O(1)$ [BDPW10]. In the following we give a new proof of this result, for a different function: $f(x, y, z) = 1_G$ if and only if $x \cdot y \cdot z = 1$ for $x, y, z \in SL(2, q)$. As is true for some functions in [BDPW10], a stronger separation could hold for f. For context, let us state and prove the upper bound for randomized communication.

Claim 20. f has randomized communication complexity $O(1)$.

Proof. In the number-on-forehead model, computing f reduces to two-party equality with no additional communication: Alice computes $y \cdot z =: w$ privately, then Alice and Bob check if $x = w^{-1}$. □

To prove the lower bound for deterministic protocols we reduce the communication problem to a combinatorial problem.

Definition 21. A *corner* in a group G is a set $\{(x, y), (xz, y), (x, zy)\} \subseteq G^2$, where x, y are arbitrary group elements and $z \neq 1_G$.

For intuition, if G is the abelian group of real numbers with addition, a corner becomes $\{(x, y), (x + z, y), (x, y + z)\}$ for $z \neq 0$, which are the coordinates of an isosceles triangle. We now state the theorem that connects corners and lower bounds.

Lemma 22. *Let G be a group and δ a real number. Suppose that every subset $A \subseteq G^2$ with $|A|/|G^2| \geq \delta$ contains a corner. Then the deterministic communication complexity of f (defined as $f(x, y, z) = 1 \iff x \cdot y \cdot z = 1_G$) is $\Omega(\log(1/\delta))$.*

Proof. We saw already twice that a number-in-hand c-bit protocol can be written as a disjoint union of 2^c rectangles (Lemmas 9, 16). Likewise, a number-on-forehead c-bit protocol P can be written as a disjoint union of 2^c *cylinder intersections* $C_i := \{(x, y, z) : f_i(y, z)g_i(x, z)h_i(x, y) = 1\}$ for some $f_i, g_i, h_i \colon G^2 \to \{0, 1\}$:

$$P(x, y, z) = \sum_{i=1}^{2^c} f_i(y, z)g_i(x, z)h_i(x, y).$$

The proof idea of the above fact is to consider the 2^c transcripts of P, then one can see that the inputs giving a fixed transcript are a cylinder intersection.

Let P be a c-bit protocol. Consider the inputs $\{(x, y, (xy)^{-1})\}$ on which P accepts. Note that at least a 2^{-c} fraction of them are accepted by some cylinder intersection $C = f \cdot g \cdot h$. Let $A := \{(x, y) : (x, y, (xy)^{-1}) \in C\} \subseteq G^2$. Since the first two elements in the tuple determine the last, we have $|A|/|G^2| \geq 2^{-c}$.

Now suppose A contains a corner $\{(x, y), (xz, y), (x, zy)\}$. Then

$$\begin{aligned}
(x, y) \in A &\implies (x, y, (xy)^{-1}) \in C &&\implies h(x, y) = 1, \\
(xz, y) \in A &\implies (xz, y, (xzy)^{-1}) \in C &&\implies f(y, (xyz)^{-1}) = 1, \\
(x, zy) \in A &\implies (x, zy, (xzy)^{-1}) \in C &&\implies g(x, (xyz)^{-1}) = 1.
\end{aligned}$$

This implies $(x, y, (xzy)^{-1}) \in C$, which is a contradiction because $z \neq 1$ and so $x \cdot y \cdot (xzy)^{-1} \neq 1_G$. □

It is known that $\delta \geq 1/\text{polyloglog}|G|$ implies a corner for certain abelian groups G, see [LM07] for the best bound and pointers to the history of the problem. For $G = SL(2,q)$ a stronger result, presented in the next section, is known: $\delta \geq 1/\text{polylog}|G|$ implies a corner [Aus16]. This in turn implies communication $\Omega(\log\log|G|) = \Omega(\log n)$.

Problem 23. Does $\delta \geq 1/|G|^{\Omega(1)}$ imply a corner for some group G? This would give the optimal separation.

6 The corners theorem for quasirandom groups

In this section we prove the corners theorem for quasirandom groups, following Austin [Aus16]. Our exposition has several minor differences with that in [Aus16], which may make it more computer-science friendly. Possibly a proof can also be obtained via certain local modifications and simplifications of Green's exposition [Gre05b, Gre05a] of an earlier proof for the abelian case. We focus on the case $G = SL(2,q)$ for simplicity, but the proof immediately extends to other quasirandom groups (with corresponding parameters).

Theorem 24. *Let $G = SL(2,q)$. Every subset $A \subseteq G^2$ of density $|A|/|G|^2 \geq 1/\log^{\alpha}|G|$ contains a corner, where $\alpha > 0$ is a universal constant.*

6.1 Proof idea

For intuition, suppose A is a rectangle, i.e., $A = B \times C$ for $B, C \subseteq G$. Let's look at the quantity

$$\mathbb{E}_{x,y,z\leftarrow G}[A(x,y)A(xz,y)A(x,zy)]$$

where $A(x,y) = 1$ iff $(x,y) \in A$. Note that the random variable in the expectation is equal to 1 exactly when $\{(x,y),(xz,y),(x,zy)\} \subseteq A$. However, z could be 1, which cannot be the case for a corner. We'll show that this expectation is greater than $1/|G|$, which implies that A does contain a corner, that is, a set $\{(x,y),(xz,y),(x,zy)\}$ with $z \neq 1$. Since we are taking $A = B \times C$, we can rewrite the above quantity as

$$\begin{aligned}\mathbb{E}_{x,y,z\leftarrow G}[B(x)C(y)B(xz)C(y)B(x)C(zy)] &= \mathbb{E}_{x,y,z\leftarrow G}[B(x)C(y)B(xz)C(zy)] \\ &= \mathbb{E}_{x,y,z\leftarrow G}[B(x)C(y)B(z)C(x^{-1}zy)]\end{aligned}$$

where the last line follows by replacing z with $x^{-1}z$ in the uniform distribution. If $|A|/|G|^2 \geq \delta$, then both $|\text{B}|/|\text{G}| \geq \delta$ and $|B|/|G| \geq \delta$. Condition on $x \in B$, $y \in C$, $z \in B$. Then the distribution $x^{-1}zy$ is a product of three independent distributions, each uniform on a set of density $\geq \delta$. (In fact, two distributions would suffice for what follows.) By Lemma 14, $x^{-1}zy$ is $\delta^{-1}/|G|^{\Omega(1)}$ close to uniform in statistical distance. This implies that the above expectation equals

$$\frac{|A|}{|G|^2}\cdot\frac{|B|}{|G|}\cdot\left(\frac{|C|}{|G|} \pm \frac{\delta^{-1}}{|G|^{\Omega(1)}}\right) \geq \delta^2\left(\delta - \frac{1}{|G|^{\Omega(1)}}\right) \geq \delta^3/2 > 1/|G|,$$

for $\delta > 1/|G|^c$ for a small enough constant c. Hence, rectangles of density polynomial in $1/|G|$ contain corners.

Given the above, it is natural to try to decompose an arbitrary set A into rectangles. We will make use of a more general result.

6.2 Weak regularity lemma

Let U be some universe (we will take $U = G^2$) and let $f : U \to [-1, 1]$ be a function (for us, $f = 1_A$). Let $D \subseteq \{d : U \to [-1, 1]\}$ be some set of functions, which can be thought of as "easy functions" or "distinguishers" (these will be rectangles or closely related to them). The next theorem shows how to decompose f into a linear combination g of the d_i up to an error which is polynomial in the length of the combination. More specifically, f will be indistinguishable from g by the d_i.

Lemma 25. *Let $f : U \to [-1, 1]$ be a function and $D \subseteq \{d : U \to [-1, 1]\}$ a set of functions. For all $\epsilon > 0$, there exists a function $g := \sum_{i \leq s} c_i \cdot d_i$ where $d_i \in D$, $c_i \in \mathbb{R}$ and $s = 1/\epsilon^2$ such that for all $d \in D$*

$$|\mathbb{E}_{x \leftarrow U}[f(x) \cdot d(x)] - \mathbb{E}_{x \leftarrow U}[g(x) \cdot d(x)]| \leq \epsilon.$$

A different way to state the conclusion, which we will use, is to say that we can write $f = g + h$ so that $\mathbb{E}[h(x) \cdot d(x)]$ is small.

The lemma is due to Frieze and Kannan [FK96], see also [RTTV08, Gow10, Skó17]. It is called "weak" because it came after Szemerédi's regularity lemma, which has a stronger distinguishing conclusion. However, the lemma is also "strong" in the sense that Szemerédi's regularity lemma has s as a tower of $1/\epsilon$ whereas here we have s polynomial in $1/\epsilon$. The weak regularity lemma is also simpler. There also exists a proof [Tao17] of Szemerédi's theorem (on arithmetic progressions), which uses weak regularity as opposed to the full regularity lemma used initially.

Proof. We will construct the approximation g through an iterative process producing functions $g_0, g_1, \ldots, g$. We will show that $||f - g_i||_2^2$ decreases by $\geq \epsilon^2$ each iteration.

Start: Define $g_0 = 0$ (which can be realized setting $c_0 = 0$).

Iterate: If not done, there exists $d \in D$ such that $|\mathbb{E}[(f - g) \cdot d]| > \epsilon$. Assume without loss of generality $\mathbb{E}[(f - g) \cdot d] > \epsilon$.

Update: $g' := g + \lambda d$ where $\lambda \in \mathbb{R}$ shall be picked later. Let us analyze the progress made by the algorithm. We have

$$\begin{aligned}
||f - g'||_2^2 &= \mathbb{E}_x[(f - g')^2(x)] \\
&= \mathbb{E}_x[(f - g - \lambda d)^2(x)] \\
&= \mathbb{E}_x[(f - g)^2] + \mathbb{E}_x[\lambda^2 d^2(x)] - 2\mathbb{E}_x[(f - g) \cdot \lambda d(x)] \\
&\leq ||f - g||_2^2 + \lambda^2 - 2\lambda \mathbb{E}_x[(f - g)d(x)] \\
&\leq ||f - g||_2^2 + \lambda^2 - 2\lambda\epsilon \\
&\leq ||f - g||_2^2 - \epsilon^2,
\end{aligned}$$

where the last line follows by taking $\lambda = \epsilon$. Therefore, there can only be $1/\epsilon^2$ iterations because $||f - g_0||_2^2 = ||f||_2^2 \leq 1$. □

6.3 Getting more for rectangles

Returning to the main proof, we will use the weak regularity lemma to approximate the indicator function of an arbitrary set A by rectangles. That is, we take D to be the collection of indicator functions for all sets of the form $S \times T$ for $S, T \subseteq G$. The weak regularity lemma shows how to decompose A into a linear combination of rectangles. These rectangles may overlap. However, we

ideally want A to be a linear combination of *non-overlapping* rectangles. In other words, we want a *partition* of rectangles. It is possible to achieve this at the price of exponentiating the number of rectangles. Note that an exponential loss is necessary even if $S = G$ in every $S \times T$ rectangle; or in other words in the uni-dimensional setting. This is one step where the terminology "rectangle" may be misleading – the set T is not necessarily an interval. If it was, a polynomial rather than exponential blow-up would have sufficed to remove overlaps.

Claim 26. Given a decomposition of A into rectangles from the weak regularity lemma with s functions, there exists a decomposition with $2^{O(s)}$ rectangles which don't overlap.

Proof. Exercise. □

In the above decomposition, note that it is natural to take the coefficients of rectangles to be the density of points in A that are in the rectangle. This gives rise to the following claim.

Claim 27. The weights of the rectangles in the above claim can be the average of f in the rectangle, at the cost of doubling the error.

Consequently, we have that $f = g + h$, where g is the sum of $2^{O(s)}$ non-overlapping rectangles $S \times T$ with coefficients $\mathbb{P}_{(x,y)\in S\times T}[f(x, y) = 1]$.

Proof. Let g be a partition decomposition with arbitrary weights. Let g' be a partition decomposition with weights being the average of f. It is enough to show that for all rectangle distinguishers $d \in D$

$$|\mathbb{E}[(f - g')d]| \leq |\mathbb{E}[(f - g)d]|.$$

By the triangle inequality, we have that

$$|\mathbb{E}[(f - g')d]| \leq |\mathbb{E}[(f - g)d]| + |\mathbb{E}[(g - g')d]|.$$

To bound $\mathbb{E}[(g - g')d]|$, note that the error is maximized for a d that respects the decomposition in non-overlapping rectangles, i.e., d is the union of some non-overlapping rectangles from the decomposition. This can be argued using that, unlike f, the value of g and g' on a rectangle $S \times T$ from the decomposition is fixed. But, from the point of "view" of such d, $g' = f$! More formally, $\mathbb{E}[(g - g')d] = \mathbb{E}[(g - f)d]$, which concludes the proof. □

We need to get still a little more from this decomposition. In our application of the weak regularity lemma above, we took the set of distinguishers to be characteristic functions of rectangles. That is, distinguishers that can be written as $U(x) \cdot V(y)$ where U and V map $G \to \{0, 1\}$. We will use that the same guarantee holds for U and V with range $[-1, 1]$, up to a constant factor loss in the error. Indeed, let U and V have range $[-1, 1]$. Write $U = U_+ - U_-$ where U_+ and U_- have range $[0, 1]$, and the same for V. The error for distinguisher $U \cdot V$ is at most the sum of the errors for distinguishers $U_+ \cdot V_+$, $U_+ \cdot V_-$, $U_- \cdot V_+$, and $U_- \cdot V_-$. So we can restrict our attention to distinguishers $U(x) \cdot V(y)$ where U and V have range $[0, 1]$. In turn, a function $U(x)$ with range $[0, 1]$ can be written as an expectation $\mathbb{E}_a U_a(x)$ for functions U_a with range $\{0, 1\}$, and the same for V. We conclude by observing that

$$\mathbb{E}_{x,y}[(f - g)(x, y)\mathbb{E}_a U_a(x) \cdot \mathbb{E}_b V_b(y)] \leq \max_{a,b} \mathbb{E}_{x,y}[(f - g)(x, y)U_a(x) \cdot V_b(y)].$$

6.4 Proof

Let us now finish the proof by showing a corner exists for sufficiently dense sets $A \subseteq G^2$. We'll use three types of decompositions for $f : G^2 \to \{0,1\}$, with respect to the following three types of distinguishers, where U_i and V_i have range $\{0,1\}$:

1. $U_1(x) \cdot V_1(y)$,
2. $U_2(xy) \cdot V_2(y)$,
3. $U_3(x) \cdot V_3(xy)$.

The first type is just rectangles, what we have been discussing until now. The distinguishers in the last two classes can be visualized over $\mathbb{R}^2$ as parallelograms with a 45-degree angle. The same extra properties we discussed for rectangles can be verified for them too.

Write f for the characteristic function of A. Recall that we want to show

$$\mathbb{E}_{x,y,g}[f(x,y)f(xg,y)f(x,gy)] > \frac{1}{|G|}.$$

We'll decompose the i-th occurrence of f via the i-th decomposition listed above. We'll write this decomposition as $f = g_i + h_i$. We apply this in a certain order to produce sums of products of three functions. The inputs to the functions don't change, so to avoid clutter we do not write them, and it is understood that in each product of three functions the inputs are, in order $(x,y),(xg,y),(x,gy)$. The decomposition is:

$$\begin{aligned} &fff \\ =&ffg_3 + fffh_3 \\ =&fg_2g_3 + fh_2g_3 + fffh_3 \\ =&g_1g_2g_3 + h_1g_2g_3 + fh_2g_3 + fffh_3. \end{aligned}$$

We first show that the expectation of the first term is big. This takes the next two claims. Then we show that the expectations of the other terms are small.

Claim 28. For all $g \in G$, the expectations $\mathbb{E}_{x,y}[g_1(x,y)g_2(xg,y)g_3(x,gy)]$ are the same up to $2^{O(s)}/|G|^{\Omega(1)}$.

Proof. We just need to get error $1/|G|^{\Omega(1)}$ for any product of three functions from the three decomposition types. We have:

$$\begin{aligned} &\mathbb{E}_{x,y}[c_1U_1(x)V_1(y) \cdot c_2U_2(xgy)V_2(y) \cdot c_3U_3(x)V_3(xgy)] \\ =&c_1c_2c_3\mathbb{E}_{x,y}[(U_1 \cdot U_3)(x)(V_1 \cdot V_2)(y)(U_2 \cdot V_3)(xgy)] \\ =&c_1c_2c_3 \cdot \mathbb{E}_x[(U_1 \cdot U_3)(x)] \cdot \mathbb{E}_y[(V_1 \cdot V_2)(y)] \cdot \mathbb{E}_z[(U_2 \cdot V_3)(z)] \pm \frac{1}{|G|^{\Omega(1)}}. \end{aligned}$$

This is similar to what we discussed in the overview, and is where we use mixing. Specifically, if $\mathbb{E}_x[(U_1 \cdot U_3)(x)]$ or $\mathbb{E}_y[(V_1 \cdot V_2)(y)]$ are at most $1/|G|^c$ for a small enough constant c than we are done. Otherwise, conditioned on $(U_1 \cdot U_3)(x) = 1$, the distribution on x is uniform over a set of density $1/|G|^c$, and the same holds for y, and the result follows by Lemma 14. □

Recall that we start with a set of density $\geq 1/\log^{\alpha}|G|$.

Claim 29. $\mathbb{E}_{x,y}[g_1(x,y)g_2(x,y)g_3(x,y)] > 1/\log^{4\alpha}|G|$.

Proof. We will relate the expectation over x, y to f using the Hölder inequality: For random variables $X_1, X_2, \ldots, X_k$,

$$\mathbb{E}[X_1 \ldots X_k] \leq \prod_{i=1}^{k} \mathbb{E}[X_i^{c_i}]^{1/c_i} \text{ such that } \sum 1/c_i = 1.$$

To apply this inequality in our setting, write

$$f = (f \cdot g_1 g_2 g_3)^{1/4} \cdot \left(\frac{f}{g_1}\right)^{1/4} \cdot \left(\frac{f}{g_2}\right)^{1/4} \cdot \left(\frac{f}{g_3}\right)^{1/4}.$$

By the Hölder inequality the expectation of the right-hand side is

$$\leq \mathbb{E}[f \cdot g_1 g_2 g_3]^{1/4} \mathbb{E}\left[\frac{f}{g_1}\right]^{1/4} \mathbb{E}\left[\frac{f}{g_2}\right]^{1/4} \mathbb{E}\left[\frac{f}{g_3}\right]^{1/4}.$$

The last three terms equal to 1 because

$$\mathbb{E}_{x,y}\frac{f(x,y)}{g_i(x,y)} = \mathbb{E}_{x,y}\frac{f(x,y)}{\mathbb{E}_{x',y' \in Cell(x,y)}[f(x',y')]} = \mathbb{E}_{x,y}\frac{\mathbb{E}_{x',y' \in Cell(x,y)}[f(x',y')]}{\mathbb{E}_{x',y' \in Cell(x,y)}[f(x',y')]} = 1.$$

where $Cell(x,y)$ is the set in the partition that contains (x,y). Putting the above together we obtain

$$\mathbb{E}[f] \leq \mathbb{E}[f \cdot g_1 g_2 g_3]^{1/4}.$$

Finally, because the functions are positive, we have that $\mathbb{E}[f \cdot g_1 g_2 g_3]^{1/4} \leq \mathbb{E}[g_1 g_2 g_3]^{1/4}$. This concludes the proof. □

It remains to show the other terms are small. Let ϵ be the error in the weak regularity lemma with respect to distinguishers with range $\{0,1\}$. Recall that this implies error $O(\epsilon)$ with respect to distinguishers with range $[-1,1]$. We give the proof for one of the terms and then we say little about the other two.

Claim 30. $|\mathbb{E}[f(x,y)f(xg,y)h_3(x,gy)]| \leq O(\epsilon)^{1/4}$.

The proof involves changing names of variables and doing Cauchy-Schwarz to remove the terms with f and bound the expectation above by $\mathbb{E}[h_3(x,g)U(x)V(xg)]$, which is small by the regularity lemma (with Type 3 decomposition).

Proof. Replace g with gy^{-1} in the uniform distribution to get

$$\begin{aligned}
&\mathbb{E}^4_{x,y,g}[f(x,y)f(xg,y)h_3(x,gy)] \\
&= \mathbb{E}^4_{x,y,g}[f(x,y)f(xgy^{-1},y)h_3(x,g)] \\
&= \mathbb{E}^4_{x,y}[f(x,y)\mathbb{E}_g[f(xgy^{-1},y)h_3(x,g)]] \\
&\leq \mathbb{E}^2_{x,y}[f^2(x,y)]\mathbb{E}^2_{x,y}\mathbb{E}^2_g[f(xgy^{-1},y)h_3(x,g)] \\
&\leq \mathbb{E}^2_{x,y}\mathbb{E}^2_g[f(xgy^{-1},y)h_3(x,g)] \\
&= \mathbb{E}^2_{x,y,g,g'}[f(xgy^{-1},y)h_3(x,g)f(xg'y^{-1},y)h_3(x,g')],
\end{aligned}$$

where the first inequality is by Cauchy-Schwarz.

Now replace $g \to x^{-1}g, g' \to x^{-1}g$ and reason in the same way:

$$\begin{aligned}
&= \mathbb{E}^2_{x,y,g,g'}[f(gy^{-1},y)h_3(x,x^{-1}g)f(g'y^{-1},y)h_3(x,x^{-1}g')] \\
&= \mathbb{E}^2_{g,g',y}[f(gy^{-1},y)\cdot f(g'y^{-1},y)\mathbb{E}_x[h_3(x,x^{-1}g)\cdot h_3(x,x^{-1}g')]] \\
&\leq \mathbb{E}_{x,x',g,g'}[h_3(x,x^{-1}g)h_3(x,x^{-1}g')h_3(x',x'^{-1}g)h_3(x',x'^{-1}g')].
\end{aligned}$$

Replace $g \to xg$ to rewrite the expectation as

$$\mathbb{E}[h_3(x,g)h_3(x,x^{-1}g')h_3(x',x'^{-1}xg)h_3(x',x'^{-1}g')].$$

We want to view the last three terms as a distinguisher $U(x)\cdot V(xg)$. First, note that h_3 has range $[-1,1]$. This is because $h_3(x,y) = f(x,y) - \mathbb{E}_{x',y'\in Cell(x,y)}f(x',y')$ and f has range $\{0,1\}$, where recall that $Cell(x,y)$ is the set in the partition that contains (x,y). Fix x',g'. The last term in the expectation becomes a constant $c \in [-1,1]$. The second term only depends on x, and the third only on xg. Hence for appropriate functions U and V with range $[-1,1]$ this expectation can be rewritten as

$$\mathbb{E}[h_3(x,g)U(x)V(xg)],$$

which concludes the proof. □

There are similar proofs to show that the remaining terms are small, see [Aus16]. For fh_2g_3, we can perform simple manipulations and then reduce to the above case. For $h_1g_2g_3$, we have a slightly easier proof than above.

To conclude the proof, suppose our set has density $\delta \geq 1/\log^{\alpha}|G|$, and the error in the regularity lemma is ϵ. By the above results we can bound

$$\mathbb{E}_{x,y,g}[f(x,y)f(xg,y)f(x,gy)] \geq 1/\log^{4\alpha}|G| - 2^{O(1/\epsilon^2)}/|G|^{\Omega(1)} - \epsilon^{\Omega(1)},$$

where the terms in the right-hand size come, left-to-right, from Claim 29, 28, and 30. Picking $\epsilon = 1/\log^{1/3}|G|$ the proof is completed for sufficiently small α.

References

[AL00] Andris Ambainis and Satyanarayana V. Lokam. Improved upper bounds on the simultaneous messages complexity of the generalized addressing function. In *Latin American Symposium on Theoretical Informatics (LATIN)*, pages 207–216, 2000. (document)

[Amb96] Andris Ambainis. Upper bounds on multiparty communication complexity of shifts. In *Symp. on Theoretical Aspects of Computer Science (STACS)*, pages 631–642, 1996. (document)

[AMS99] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. of Computer and System Sciences*, 58(1, part 2):137–147, 1999. 1

[Aus16] Tim Austin. Ajtai-Szemerédi theorems over quasirandom groups. In *Recent trends in combinatorics*, volume 159 of *IMA Vol. Math. Appl.*, pages 453–484. Springer, [Cham], 2016. (document), 5, 6, 6.4

[BC92] Michael Ben-Or and Richard Cleve. Computing algebraic formulas using a constant number of registers. *SIAM J. on Computing*, 21(1):54–58, 1992. (document)

[BDPW10] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel. Separating deterministic from randomized multiparty communication complexity. *Theory of Computing*, 6(1):201–225, 2010. (document), 5

[BGKL03] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication complexity of simultaneous messages. *SIAM J. on Computing*, 33(1):137–166, 2003. (document)

[BNP08] László Babai, Nikolay Nikolov, and László Pyber. Product growth and mixing in finite groups. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 248–257, 2008. (document), 14

[BNS92] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. of Computer and System Sciences*, 45(2):204–232, 1992. 5

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *15th ACM Symp. on the Theory of Computing (STOC)*, pages 94–99, 1983. 5

[CP10] Arkadev Chattopadhyay and Toniann Pitassi. The story of set disjointness. *SIGACT News*, 41(3):59–85, 2010. 1

[DHKP97] Martin Dietzfelbinger, Torben Hagerup, Jyrki Katajainen, and Martti Penttonen. A reliable randomized algorithm for the closest-pair problem. *J. Algorithms*, 25(1):19–51, 1997. 3.2

[FK96] Alan M. Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 12–20, 1996. 6.2

[Gow08] W. T. Gowers. Quasirandom groups. *Combinatorics, Probability & Computing*, 17(3):363–387, 2008. (document), 4, 3.3, 14

[Gow10] W. T. Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *Bull. Lond. Math. Soc.*, 42(4):573–606, 2010. 6.2

[Gre05a] Ben Green. An argument of Shkredov in the finite field setting, 2005. Available at people.maths.ox.ac.uk/greenbj/papers/corners.pdf. 6

[Gre05b] Ben Green. Finite field models in additive combinatorics. *Surveys in Combinatorics, London Math. Soc. Lecture Notes 327, 1-27*, 2005. 6

[GV] W. T. Gowers and Emanuele Viola. Interleaved group products. *SIAM J. on Computing*. To appear. Special issue of FOCS 2016. (document), 1, 1, 2, 2.2, 2.3, 4.1, 19, 5

[GV15] W. T. Gowers and Emanuele Viola. The communication complexity of interleaved group products. In *ACM Symp. on the Theory of Computing (STOC)*, 2015. 2

[IL95] Neil Immerman and Susan Landau. The complexity of iterated multiplication. *Inf. Comput.*, 116(1):103–116, 1995. (document)

[KMR66] Kenneth Krohn, W. D. Maurer, and John Rhodes. Realizing complex Boolean functions with simple groups. *Information and Control*, 9:190–195, 1966. (document)

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997. (document)

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992. 1

[LM07] Michael T. Lacey and William McClain. On an argument of Shkredov on two-dimensional corners. *Online J. Anal. Comb.*, (2):Art. 2, 21, 2007. 5

[LW54] Serge Lang and André Weil. Number of points of varieties in finite fields. *American Journal of Mathematics*, 76:819–827, 1954. 2.3

[Mil14] Eric Miles. Iterated group products and leakage resilience against NC^1. In *ACM Innovations in Theoretical Computer Science conf. (ITCS)*, 2014. (document)

[Mix89] David A. Mix Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *J. of Computer and System Sciences*, 38(1):150–164, 1989. (document)

[MV13] Eric Miles and Emanuele Viola. Shielding circuits with groups. In *ACM Symp. on the Theory of Computing (STOC)*, 2013. (document), 1, 1, 5

[PRS97] Pavel Pudlák, Vojtěch Rödl, and Jiří Sgall. Boolean circuits, tensor ranks, and communication complexity. *SIAM J. on Computing*, 26(3):605–633, 1997. (document)

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992. 1

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000. (document)

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Dense subsets of pseudorandom sets. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 76–85, 2008. 6.2

[RY19] Anup Rao and Amir Yehudayoff. *Communication complexity*. 2019. https://homes.cs.washington.edu/ anuprao/pubs/book.pdf. (document)

[Sha16] Aner Shalev. Mixing, communication complexity and conjectures of Gowers and Viola. *Combinatorics, Probability and Computing*, pages 1–13, 6 2016. arXiv:1601.00795. 1, 1, 2

[She14] Alexander A. Sherstov. Communication complexity theory: Thirty-five years of set disjointness. In *Symp. on Math. Foundations of Computer Science (MFCS)*, pages 24–43, 2014. 1

[Skó17] Maciej Skórski. A cryptographic view of regularity lemmas: Simpler unified proofs and refined bounds. In *Theory and Applications of Models of Computation - 14th Annual Conference, TAMC 2017, Bern, Switzerland, April 20-22, 2017, Proceedings*, pages 586–599, 2017. 6.2

[Tao17] Terence Tao. Szemerédi's proof of Szemerédi's theorem, 2017. https://terrytao.files.wordpress.com/2017/09/szemeredi-proof1.pdf. 6.2

[Vio14] Emanuele Viola. The communication complexity of addition. *Combinatorica*, pages 1–45, 2014. 3.2

[Vio16a] Emanuele Viola. Thoughts: Mixing in groups, 2016. https://emanueleviola.wordpress.com/2016/10/21/mixing-in-groups/. (document), 4.1

[Vio16b] Emanuele Viola. Thoughts: Mixing in groups ii, 2016. https://emanueleviola.wordpress.com/2016/11/15/mixing-in-groups-ii/. (document), 4.1

[Vio17] Emanuele Viola. Special topics in complexity theory. Lecture notes of the class taught at Northeastern University. Available at http://www.ccs.neu.edu/home/viola/classes/spepf17.html, 2017. (document)

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *11th ACM Symp. on the Theory of Computing (STOC)*, pages 209–213, 1979. (document)