Approximate Polymorphisms (BGU)

Yuval Filmus

2 May 2022

Abstract

We extend linearity testing to functions beyond XOR.
Joint work with Gilad Chase, Noam Lifshitz, Dor Minzer, Elchanan Mossel, and Nitin Saurabh.

1 Introduction

Suppose that $f\colon \{0,1\}^n \to \{0,1\}$.

- If $f(x \oplus y) = f(x) \oplus f(y)$ for all x, y, then f is *linear*, that is, of the form $x_{i_1} \oplus \cdots \oplus x_{i_m}$.
- If $f(x \oplus y) = f(x) \oplus f(y)$ with probability $1 - \epsilon$, then f is $O(\epsilon)$-close to a linear function (“linearity testing”).
- If $f(x \oplus y) = f(x) \oplus f(y)$ with probability at least $1/2 + \epsilon$, then f is $\Omega(\epsilon)$-correlated with some linear function (“list-decoding regime”).

Linearity testing is instrumental in PCP-related results, such as the optimal inapproximability of 3XOR due to Håstad.

What happens when we replace $\oplus$ by another function, such as $\wedge$?

2 Polymorphisms

Let’s start by considering the case $\epsilon = 0$, in greater generality. Let $P \subseteq \{0,1\}^M$ be an arbitrary *predicate*. A function $f\colon \{0,1\}^n \to \{0,1\}$ is a *polymorphism* of P if for any $n \times M$ table whose rows belong to P, if we apply f to each column then we get another row belonging to P:

$$\begin{array}{|c|c|c|l} \hline x_{11} & \cdots & x_{1M} & \in P \\ \hline \vdots & \vdots & \vdots & \\ \hline x_{n1} & \cdots & x_{nM} & \in P \\ \hline \multicolumn{1}{c}{\downarrow f} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{\downarrow f} & \\ \hline f(x_{\star 1}) & \cdots & f(x_{\star M}) & \in P \\ \hline \end{array}$$

Linearity testing is an example of a *truth-functional predicate*. Every $g\colon \{0,1\}^m \to \{0,1\}$ gives rise to the $(m+1)$-ary predicate

$$P_g = \{(x_1, \ldots, x_m, z) : x_1, \ldots, x_m \in \{0,1\}, z = g(x_1, \ldots, x_m)\}.$$

The corresponding picture is:

$$\begin{array}{|c|c|c|c|c|} \hline x_{11} & \cdots & x_{1m} & & g(x_{11}, \ldots, x_{1m}) \\ \hline \vdots & \vdots & \vdots & & \vdots \\ \hline x_{n1} & \cdots & x_{nm} & & g(x_{n1}, \ldots, x_{nm}) \\ \hline \multicolumn{1}{c}{\downarrow f} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{\downarrow f} & & \multicolumn{1}{c}{\downarrow f} \\ \hline y_1 & \cdots & y_m & & g(y_1, \ldots, y_m) \\ \hline \end{array}$$

In other words, f satisfies the equation

$$f(g(x_{11},\dots,x_{1m}),\dots,g(x_{n1},\dots,x_{nm})) = g(f(x_{11},\dots,x_{n1}),\dots,f(x_{1m},\dots,x_{nm})),$$

which we abbreviate by $f \circ g^n = g \circ f^m$.

Here are some particular examples:

- Linearity testing ($g = x \oplus y$): $f(x \oplus y) = f(x) \oplus f(y)$.
- AND testing ($g = x \wedge y$): $f(x \wedge y) = f(x) \wedge f(y)$.
- When P is 3NAE, we can think of each row as encoding a ranking of three candidates: the columns encode whether $1 > 2, 2 > 3, 3 > 1$. Each ranking corresponds to one of the six satisfying assignments of P. Arrow's theorem states that the only polymorphisms of 3NAE are dictators $f(x) = x_i$ and anti-dictators $f(x) = 1 - x_i$.
- When P is NAND, a polymorphism is the same as an intersecting family: the condition states that if x, y are disjoint (when interpreted as subsets of $[n]$), then not both belong to the family of sets encoded by f.

Post classified in 1941 all polymorphisms of all binary predicates ("Post's lattice"). When the predicate is truth-functional, Dokow and Holzman determined the possible polymorphisms independently:

- The projections $f(x) = x_i$ are always polymorphisms.
- The anti-projections $f(x) = 1 - x_i$ are polymorphisms whenever g is odd.
- The constant function $f(x) = b$ is a polymorphism if $g(b,\dots,b) = b$.
- If g is an XOR or a negated XOR, then XORs and negated XORs are polymorphisms of g (subject to a parity constraint).
- If g is an OR, then ORs are polymorphisms of g.
- If g is an AND, then ANDs are polymorphisms of g.

If g is not one of XOR, NXOR, OR, AND then it has only trivial polymorphisms (depending on at most one coordinate).

3 Approximate polymorphisms

A function f is an ϵ-*approximate polymorphism* of a function g if the defining condition of polymorphism holds with probability $1 - \epsilon$ over a uniformly random choice of the nm variables x_{ij}.

When P is a predicate, we can similarly define approximate polymorphisms once we specify a probability distribution on the rows of the $n \times M$ table (different rows are sampled independently). The natural choice is the uniform distribution over P (this generalizes the definition for functions).

Here are some classical results:

- Linearity testing states that an ϵ-approximate polymorphism of $x \oplus y$ is $O(\epsilon)$-close to an exact polymorphism of $x \oplus y$.
- Kalai's approximate Arrow's theorem (in a stronger version due to Mossel) states that an ϵ-approximate polymorphism of 3NAE is $O(\epsilon)$-close to an exact polymorphism of 3NAE.
- Friedgut and Regev showed that approximate intersecting families (suitably defined) are close to intersecting families. Their definition corresponds to approximate polymorphisms of NAND with respect to a non-uniform distribution over the rows.

It is natural to conjecture the following:

Conjecture. *An ϵ-approximate polymorphism of g is $O_g(\epsilon)$-close to an exact polymorphism of g.*

Together with Lifshitz, Minzer and Mossel, we proved a weaker version of this conjecture for $g = x \wedge y$, in which $O(\epsilon)$ is replaced with another function $\delta = \delta(\epsilon)$ which satisfies

$$\frac{1}{\delta} \approx \operatorname{polylog} \frac{1}{\epsilon}.$$

It turns out that the conjecture, as stated, is wrong. Consider $g = \overline{x \wedge y}$:

x_1	y_1		$\overline{x_1 \wedge y_1}$
x_2	y_2		$\overline{x_2 \wedge y_2}$
$\downarrow f$	$\downarrow f$		$\downarrow h$
$x_1 \wedge x_2$	$y_1 \wedge y_2$		$\overline{x_1 \wedge y_1} \vee \overline{x_2 \wedge y_2}$

where $f(z_1, z_2) = z_1 \wedge z_2$ and $h(z_1, z_2) = z_1 \vee z_2$. The final line satisfies the predicate P_g since

$$g(x_1 \wedge x_2, y_1 \wedge y_2) = \overline{x_1 \wedge x_2 \wedge y_1 \wedge y_2} = \overline{x_1 \wedge y_1} \vee \overline{x_2 \wedge y_2}.$$

This gives what we call a *skew* polymorphism, which is a pair of n-ary functions (f, h) satisfying $h \circ g^n = g \circ f^m$.

Skew polymorphisms are relevant here since we can fit both of them in a single function which behaves like f on a balanced input and like h on a biased input. Concretely, if we define $\phi(x)$ as $f(x)$ if $|x| \approx \frac{1}{2}n$ and as $h(x)$ if $|x| \approx \frac{3}{4}n$, then ϕ is an $o(1)$-approximate polymorphisms of g.

Adapting the conjecture accordingly, we are able to prove the following theorem:

Theorem. *If f is an ϵ-approximate polymorphism of g then f is δ-close to F, where (F, h) is a skew polymorphism of g, and*

$$\frac{1}{\delta} \approx \log\log \frac{1}{\epsilon}.$$

Every polymorphism f of g gives rise to the skew polymorphism (f, f). The only other examples are extensions of the above example: when g is NAND, we get skew polymorphisms of the form $(\bigwedge_{i \in S} x_i, \bigvee_{i \in S} x_i)$, and when g is NOR, we get skew polymorphisms of the form $(\bigvee_{i \in S} x_i, \bigwedge_{i \in S} x_i)$; this also follows from the work of Dokow and Holzman.

On the proof The proof is based on the following technical lemma:

Lemma. *For every g other than XOR and negated XOR, and every $\delta > 0$, there exist $\eta > 0$ and $L \in \mathbb{N}$ such that every η-approximate polymorphisms of g is δ-close to an L-junta.*

(XOR and negated XOR can be handled directly using the standard linearity testing argument.)

Given this lemma (whose proof we do not describe), we prove our main theorem, in the following version: for every $\delta > 0$ there is $\epsilon > 0$ such that if f is an ϵ-approximate polymorphism of g then f is δ-close to F, where (F, h) is a skew polymorphism of g.

Given $\delta > 0$, we apply the lemma to obtain parameters η and L, and choose $\epsilon \ll \eta, 2^{-mL}$. Now suppose that f is an ϵ-approximate polymorphism of g. Apply the lemma to get a L-junta F which is δ-close to f. Suppose that F depends on the first L coordinates. For random $y_1, \ldots, y_m \in \{0,1\}^{n-L}$:

- $f_i := f(\cdot, y_i)$ is δ-close to F for all $i \in [m]$.
- $\Pr[h \circ g^L \neq g \circ (f_1, \ldots, f_m)] \approx \epsilon$, where $h := f(\cdot, g(y_1, \ldots, y_m))$.

Since $\epsilon \ll 2^{-mL}$, in fact $h \circ g^L = g \circ (f_1, \ldots, f_m)$. Dokow and Holzman found all solutions to this equation, and from their classification and the known relations $f_i \approx f_j$ we can conclude that $f_1 = \cdots = f_m$, and so (f_1, h) is a skew polymorphism of g. Finally, $f_1 \approx F \approx f$.

4 List-decoding regime

How large can $\Pr[f(x \wedge y) = f(x) \wedge f(y)]$ be if f is quasirandom? The answer depends on how we define quasirandomness. Let us say that f is quasirandom if it does not correlate with any junta. Here are the two obvious quasirandom constructions:

- Choose f at random. The "test" $f(x \wedge y) = f(x) \wedge f(y)$ succeeds with probability 1/2.
- Choose f at random on balanced inputs, and let $f(x) = 0$ on inputs x where $|x| \approx \frac{1}{4}n$. The test now succeeds with probability 3/4.

In the second example, f correlates with a junta with respect to the biased measure $\mu_{1/4}$, but not with respect to the unbiased measure $\mu_{1/2}$.

Can we beat the second example? Indeed we can. The following function passes the test with probability roughly 0.814975:

$$f(x) = \begin{cases} \mathsf{majority}(x) & |x| \approx \frac{1}{2}n, \\ \left[x_1 + \cdots + x_n \geq \frac{1}{4}n + \theta\sqrt{\frac{3}{16}n}\right] & |x| \approx \frac{1}{4}n, \end{cases}$$

where $\theta \approx 0.908$. Moreover, this is optimal.

The idea behind this construction is that for random x, y, the three values

$$\sum_i x_i, \quad \sum_i y_i, \quad \sum_i x_i y_i$$

are closely approximated by a multivariate Gaussian $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, where $\mathbf{x}$ and $\mathbf{y}$ are independent, and $\mathbf{z}$ is positively correlated with each of them. Normalizing these values to

$$\frac{\sum_i x_i - \frac{1}{2}n}{\sqrt{\frac{1}{4}n}}, \quad \frac{\sum_i y_i - \frac{1}{2}n}{\sqrt{\frac{1}{4}n}}, \quad \frac{\sum_i x_i y_i - \frac{1}{4}n}{\sqrt{\frac{3}{16}n}},$$

each of the marginals $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ is a standard Gaussian. The function f is defined as the sign function on $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ and as an appropriate threshold function on $\hat{\mathbf{z}}$. The threshold is the exact point at which

$$\Pr[\mathsf{sign}(\hat{\mathbf{x}}) = \mathsf{sign}(\hat{\mathbf{y}}) = 1 \mid \hat{\mathbf{z}}] = \frac{1}{2}.$$

The proof that this f is optimal combines the invariance principle with an extension of Borell's isoperimetric theorem due to Joe Neeman.

5 Open questions

Our work suggests several avenues for future research:

1. Extend the characterization of approximate polymorphisms to predicates.
2. Determine the correct dependence of δ on ϵ.
3. Extend the entire framework to larger alphabets.
4. Determine the quasirandomness threshold for all g.