Approximate Polymorphisms

Gilad Chase, **Yuval Filmus**, Dor Minzer, Nitin Saurabh

May 26, 2021

Abstract

An n-bit function f is a polymorphism of an m-bit function g if $f \circ g^n = gof^m$. For example, an n-bit function f is a polymorphism of the 2-bit function $g = \mathsf{XOR}$ if $f(x+y) = f(x) + f(y)$ for all x, y.

It is known that all exact polymorphisms of XOR are XORs, and furthermore, all approximate polymorphisms of XOR are close to XORs — this is the classical linearity testing.

We determine all approximate polymorphisms of g for an arbitrary function g.

In addition, we consider "list decoding" variants of this question.

1 Introduction: linearity testing

Linearity testing is one of the prototypical examples of property testing:

- 100% regime: If $f\colon \{0,1\}^n \to \{0,1\}$ satisfies $f(x \oplus y) = f(x) \oplus f(y)$ for all x, y then f is an XOR.
- 99% regime: If $f\colon \{0,1\}^n \to \{0,1\}$ satisfies $f(x \oplus y) = f(x) \oplus f(y)$ with probability $1-\epsilon$ then f is $O(\epsilon)$-close to an XOR.
- 51% regime: If $f\colon \{0,1\}^n \to \{0,1\}$ satisfies $f(x \oplus y) = f(x) \oplus f(y)$ with probability $1/2+\epsilon$ then f has $\Omega(\epsilon)$ correlation with some XOR.

In this work, we ask the following question:

What happens when we replace $\oplus$ with another function $g\colon \{0,1\}^m \to \{0,1\}$?

Concretely, let's take as an example the AND function.

- 100% regime: Which functions $f\colon \{0,1\}^n \to \{0,1\}$ satisfy $f(x \wedge y) = f(x) \wedge f(y)$ for all x, y?
- 99% regime: Which functions $f\colon \{0,1\}^n \to \{0,1\}$ satisfy $f(x \wedge y) = f(x) \wedge f(y)$ with probability $1-\epsilon$?
- 51% (?) regime: What can we say about functions $f\colon \{0,1\}^n \to \{0,1\}$ which satisfy $f(x \wedge y) = f(x) \wedge f(y)$ with "non-trivial" probability?

The answer to the first question is well-known: either $f = 0$ or f is an AND. In previous work (Filmus, Lifshitz, Minzer, Mossel), we answered the second question: f is close to 0 or to an AND (see also a simplified version on my homepage). The third question was left open.

2 Exact polymorphisms

A function $f\colon \{0,1\}^n \to \{0,1\}$ is a *polymorphism* of $g\colon \{0,1\}^m \to \{0,1\}$ if

$$f \circ g^n = g \circ f^m.$$

In other words, the following diagram "commutes", in the sense that the two ways to compute $*$ result in the same value:

$$\begin{array}{|c|c|c|c|c|}
\hline
\cdot & \cdots & \cdot & \xrightarrow{g} & \cdot \\
\hline
\vdots & \ddots & \vdots & \xrightarrow{g} & \vdots \\
\hline
\cdot & \cdots & \cdot & \xrightarrow{g} & \cdot \\
\hline
\downarrow f & \downarrow f & \downarrow f & & \downarrow f \\
\hline
\cdot & \cdots & \cdot & \xrightarrow{g} & * \\
\hline
\end{array}$$

For any function g, the functions $f = x_i$ are always polymorphisms of g. Other "trivial" polymorphisms include $f = 1 - x_i$ when g is odd, and $f = b$ when $g(b, \ldots, b) = b$.

Dokow and Holzman classified all nontrivial polymorphisms for all g. Suppose that g depends on all of its inputs (which is without loss of generality), and that $m \geq 2$. Then g has nontrivial polymorphisms in the following cases:

- $g = \mathsf{XOR}$ or $g = \mathsf{NXOR}$. The nontrivial polymorphisms are XORs and NXORs.
- $g = \mathsf{AND}$. The nontrivial polymorphisms are ANDs.
- $g = \mathsf{OR}$. The nontrivial polymorphisms are ORs.

Dokow and Holzman actually solved a more general problem, in which the various f in the figure are allowed to be different functions. This will show up later on.

3 Approximate polymorphisms

A function $f\colon \{0,1\}^n \to \{0,1\}$ is an *ϵ-approximate polymorphism* of $g\colon \{0,1\}^m \to \{0,1\}$ if

$$\Pr[f \circ g^n = g \circ f^m] \geq 1 - \epsilon.$$

It is natural to conjecture that any approximate polymorphism of g is close to an exact polymorphism of g. This is the case for all nontrivial cases listed above: $\mathsf{XOR}, \mathsf{NXOR}, \mathsf{AND}, \mathsf{OR}$. Is it true in general?

When g is XOR or NXOR, linearity testing tells us that all approximate polymorphisms are close to exact polymorphisms. When $g \neq \mathsf{XOR}, \mathsf{NXOR}$, we will be able to show that any approximate polymorphism of g is close to a junta (more on this, later). This allows us to analyze approximate polymorphisms of g using a very simple argument.

Suppose that f is δ-close to a junta F, which depends on the first t coordinates. Recall our table above. We choose the last $n - t$ rows at random. After fixing the values of the last $n - t$ coordinates, we get $m + 1$ new functions $f_0, \ldots, f_m$, where $f_1, \ldots, f_m$ are all δ-close to F (the remaining function f_0 is also δ-close to F, but with respect to a biased measure):

$$\begin{array}{|c|c|c|c|c|}
\hline
\cdot & \cdots & \cdot & \xrightarrow{g} & \cdot \\
\hline
\vdots & \ddots & \vdots & \xrightarrow{g} & \vdots \\
\hline
\cdot & \cdots & \cdot & \xrightarrow{g} & \cdot \\
\hline
\downarrow f_1 & \downarrow f_j & \downarrow f_m & & \downarrow f_0 \\
\hline
\cdot & \cdots & \cdot & \xrightarrow{g} & * \\
\hline
\end{array}$$

The diagram above fails to commute with probability ϵ. We choose the parameters δ, t so that $\epsilon < 2^{-mt}$. This means that the diagram *always* commutes, that is,

$$f_0 \circ g^n = g \circ (f_1, \ldots, f_m).$$

We say that $(f_0, \ldots, f_m)$ is a *multi-sorted polymorphism* of g.

Dokow and Holzman determined all multi-sorted polymorphisms for all functions g. In our case, we know that $f_1, \ldots, f_m$ are all close to F, and so to each other. Given the explicit classification of all multi-sorted polymorphisms, this implies that $f_1 = \ldots = f_m$, and so (f_0, f_1) is a *skew polymorphism*, that is

$$f_0 \circ g^n = g \circ f_1^m.$$

Apart from the nontrivial solutions listed above, we get two more cases in which $f_0 \neq f_1$:

- $g = \mathsf{NAND}$. The nontrivial skew polymorphisms are $f_0 = \mathsf{OR}$ and $f_1 = \mathsf{AND}$.
- $g = \mathsf{NOR}$. The nontrivial skew polymorphisms are $f_0 = \mathsf{AND}$ and $f_1 = \mathsf{OR}$.

These correspond to functions f which look like f_1 around the middle slice, and like f_0 around the $\mathbb{E}[g]n$-slice.

4 Closeness to junta

Let us now explain why approximate polymorphisms are close to juntas. We will concentrate on the case $g = \mathsf{AND}$, and then indicate the minor changes needed for the general case.

Our starting point is Jones' regularity lemma, which states that for *every* function f we can find a small set T of coordinates such that for most $y \in \{0,1\}^{\overline{T}}$, the function $f_{\overline{T} \to y}$ is pseudorandom (technically, has small low-degree influences).

Suppose that f is an ϵ-approximate polymorphism of AND, that is

$$f(x \wedge y) = f(x) \wedge f(y)$$

for most $x, y \in \{0,1\}^n$. In order to show that f is close to a T-junta, it suffices to show that for most $y \in \{0,1\}^{\overline{T}}$, the function $f_{\overline{T} \to y}$ is nearly constant. We do so by contradiction: assuming that

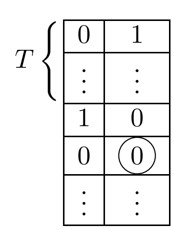
- f is δ-far from constant, and
- with probability at least δ, the restriction $f_{\overline{T} \to y}$ is pseudorandom and δ-far from constant,

we will reach a contradiction (it suffices to only explicitly assume the second property, of course).

We construct two coupled pairs of input $(x, y), (x, z)$, each of which is individually uniformly distributed over $\{0,1\}^n \times \{0,1\}^n$, in the following way:

1. Sample x, y at random.
2. Set $z = y$. For each index $i \notin T$ such that $x_i = 0$, resample z_i.

In pictures:



The circled part is resampled in z.

By construction,

$$f(x \wedge y) = f(x \wedge z),$$

and so with probability $1 - 2\epsilon$,

$$f(x) \wedge f(y) = f(x) \wedge f(z).$$

On the other hand, $f(x) = 1$ with probability δ, and $f' = f_{T \to y|_T}$ is pseudorandom and δ-far from constant with probability δ.

Consider now the following alternative way of sampling y, z:

1. Choose half of the coordinates in $\overline{T}$ at random, and fix them to random values $y^{(1)} = z^{(1)}$, obtaining a function f''.

2. Choose two random inputs $y^{(2)}, z^{(2)}$ for f''.

According to the celebrated "It Ain't Over Till It's Over" theorem, with some probability γ the function f'' is γ-far from constant, and so $f''(y^{(2)}) \neq f''(z^{(2)})$ with probability roughly 2γ. Altogether, $f(x) \wedge f(y) \neq f(x) \wedge f(z)$ with probability at least $\delta^2\gamma^2$. If ϵ is small enough as a function of δ, we reach a contradiction.

The argument only used two features of AND:

- $0 \wedge b$ doesn't depend on b.

- $1 \wedge b$ does depend on b.

The first property states that AND has a *non-trivial certificate*, that is, there is some partial input which determines the output of the function. We can find such a certificate as long as g depends on all inputs and is not XOR or NXOR.

Assuming that the partial input doesn't specify coordinate j, the second property states that there is some other partial input, setting values to all coordinates apart from j, in which the output is not determined. This is true as long as g depends on all inputs.

5 List-decoding versions

What can we say about functions f which satisfy $f \circ g^n = g \circ f^m$ with "non-trivial probability"? We are looking for a result of the following form:

If $f \circ g^n = g \circ f^m$ with probability $s_g + \epsilon$ then f has some non-trivial structure.

The "strength" of the structure can deteriorate with ϵ. We think of s_g as the optimal value for this type of structure.

When $g = \mathsf{XOR}$, we can show such a result for $s_g = 1/2$, the structure being correlation with some character. The value $1/2$ is optimal since a random function will satisfy $f(x \oplus y) = f(x) \oplus f(y)$ with probability close to $1/2$ but will not have any structure.

What happens when $g = \mathsf{AND}$? If f is a random function then $f(x \wedge y) = f(x) \wedge f(y)$ with probability $1/2$, so we can aim for $s_\wedge = 1/2$. But in fact we can improve this: we can choose f to be random around the middle slice, and equal to zero around the quarter slice. This suggests aiming at $s_\wedge = 3/4$.

It turns out that this conjecture can be further improved: we can let f be the majority function around the middle slice, and an appropriate thresholds function around the quarter slice, achieving a success probability of roughly 0.815.

Is this the correct value of $s_\wedge$? To answer this question, we first have to specify a notion of structure. We choose the following: f correlates with a *low-degree* character. For this notion of structure, we are able to show that $s_\wedge$ is at most roughly 0.866; we conjecture that 0.815 is optimal.

How does one analyze this problem? Suppose that f does not correlate with any low-degree character. We will try to bound the probability that $f(x \wedge y) = f(x) \wedge f(y)$.

The idea is to apply the invariance principle. However, the invariance principle only applies to pseudorandom functions, and only after applying a bit of noise. We use Jones' regularity lemma to partition f into a bunch of pseudorandom functions. Using a result of Mossel on "connected spaces", we show that as long as $g \neq \mathsf{XOR}, \mathsf{NXOR}$, applying noise doesn't affect the probability that $f(x \wedge y) = f(x) \wedge f(y)$ by much.

Applying the invariance principle to restrictions of f, we obtain functions F_0, F_1, F_2 on Gaussian space with

$$\Pr[F_0(x \wedge y) = F_1(x) \wedge F_2(y)] \approx \Pr[f(x \wedge y) = f(x) \wedge f(y)],$$

where $(x \wedge y, x, y)$ on the left is a multivariate Gaussian with the appropriate mean vector and covariance matrix.

Since f doesn't correlate with any low-degree character, the functions F_1, F_2 are balanced, and we obtain the upper bound 0.866 using Borell's isoperimetric inequality.

The lower bound 0.815 is also obtained in the same way, using a construction in Gaussian space. We can only realize the construction when $F_1 = F_2$, leaving a gap between the upper bound and the lower bound even in their Gaussian space forms.