Chapter 1

Lecture 1, hour 2

scribed by Artur Jeż and Łukasz Jeż

1.1 Introduction

We are given a function f and some property $\mathcal{F}$. Function f is given as a 'black box'—we can query it with an input x and receive $f(x)$ as a result. We are interested in testing, whether $f \in \mathcal{F}$, we would like to check it using few queries, preferably a constant amount. We focus on properties $\mathcal{F}$ that are interesting—we are not interested in value at some specific point or so other local question. We rather want to inspect some global property. The use of randomization is inevitable—we can check only few values of f, so it is rather easy to fool us if we proceed in a deterministic manner. The property we shall focus on now, is being a homomorphism. In the following we denote by $\mathcal{F}$ the class of group homomorphisms. We define the *distance* δ between two functions as

$$\delta(f, g) = \mathbb{P}\mathrm{r}_x \left[f(x) \neq g(x) \right] .$$

The distance from the set is defined in the usual way:

$$\delta(f, A) = \min_{g \in A} \mathbb{P}\mathrm{r}_x \left[f(x) \neq g(x) \right] .$$

1.2 Test for homomorphism

We give a simple test for checking, whether a given function $f : G \to G$ is a group homomorphism. We assume, that G is abelian and hence write its operation as additions.

Test 1. *Let $x, y \leftarrow_u G$. If $f(x) + f(y) = f(x + y)$ then **accept**, else **reject**.*

To prove, that Test 1 is 'correct' we need to show that if $f \in \mathcal{F}$ then f is accepted with 'high' probability and if $f \notin \mathcal{F}$ then f is rejected with 'high' probability. What should be noted here, is that if $\delta(f, \mathcal{F})$ is very small then

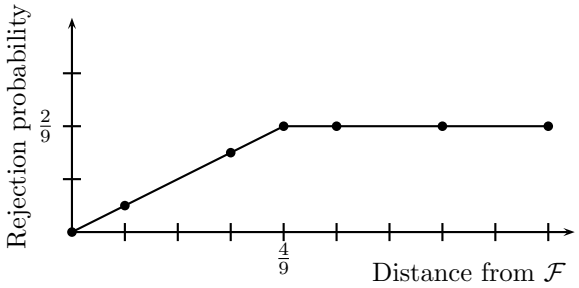

Figure 1.1: The dependency between distance from $\mathcal{F}$ and rejection probability. The area below the line is forbidden by the Theorem 3. There are examples, that show that we cannot improve the bound: there are functions f such that $(\delta(f,\mathcal{F}), \Pr_{x,y}[f(x)+f(y) \neq f(x+y)])$ is exactly on the line. They are represented by dots on the line

it is hard for us to find it out—for example if it differs from homomorphism φ at only one point then we have to at least read this unique place. Hence the rejection probability should depend on the distance from $\mathcal{F}$.

Fact 2. *If $f \in \mathcal{F}$, then Test 1 accepts with probability* 1.

The 'converse' statement is:

Theorem 3. *If Test 1 fails with probability $\delta_0 < \frac{2}{9}$, that is $\Pr[f(x)+f(y) \neq f(x+y)] \leq \delta_0 < \frac{2}{9}$, then $\delta(f,\mathcal{F}) \leq 2\delta_0$.*

There is something special about the constant $\frac{2}{9}$. It cannot be strengthened, cf. Figure 1.1. The interpretation of the Theorem is as follows: the area below the line is forbidden—the Theorem says there are no functions there. Still the shape of this line is rather peculiar, one might think it can be improved. This is not the case, as there are known examples exactly on the line, in both parts of it (the sloped and the horizontal one).

Failed approach

First we try the straightforward approach. It fails, but provides us some knowledge.

Let φ be the homomorphism closest to f and let $\epsilon = \delta(f,\varphi)$. We assume that φ is far from f and then try to prove that f is rejected with high probability, basing on its differences with φ. This approach seems to make sense—if f is not a homomorphism, it has a lot of 'bad values' and we will exploit it in the proof.

Define the following events

$$\begin{aligned} E_1 &= \{x : \varphi(x) \neq f(x)\} & \mathbb{P}\mathrm{r}_x[E_1] &= \epsilon \\ E_2 &= \{y : \varphi(y) \neq f(y)\} & \mathbb{P}\mathrm{r}_y[E_2] &= \epsilon \\ E_3 &= \{x, y : \varphi(x+y) \neq f(x+y)\} & \mathbb{P}\mathrm{r}_{x,y}[E_3] &= \epsilon \end{aligned}$$

Obtaining probabilities of E_1 and E_2 is straightforward, as $\delta(f, \varphi) = \epsilon$. To see that $\mathbb{P}\mathrm{r}_{x,y}[E_3] = \epsilon$, notice that regardless of the choice of x, if y has the uniform distribution over G, then $x + y$ also has uniform distribution over G.

Moreover

$$\begin{aligned} \mathbb{P}\mathrm{r}_{x,y}[E_1 \wedge E_3] &= \epsilon^2 \,, \\ \mathbb{P}\mathrm{r}_{x,y}[E_2 \wedge E_3] &= \epsilon^2 \,, \end{aligned}$$

since E_1 is independent from E_3, same as E_2 from E_3. Then

$$\mathbb{P}\mathrm{r}_{x,y}[\neg E_1 \wedge \neg E_2 \wedge E_3] \geq \mathbb{P}\mathrm{r}_{x,y}[E_3] - \mathbb{P}\mathrm{r}_{x,y}[E_1 \wedge E_3] - \mathbb{P}\mathrm{r}_{x,y}[E_2 \wedge E_3] = \epsilon\,(1 - 2\epsilon\,.)$$

This result turns out quite good for small ϵ, but miserable for larger ϵ. Furthermore it is far worse than the linear bound that we have promised. In overall it is not that impressive. It is time to think what is wrong with the approach?

Afterthought—the only information we have used is the existence of a homomorphism φ that is distant from f that is $\delta(f, \varphi)$ is big. But if we take a homomorphism f, then it has a lot of distant homomorphisms... So we should use different approach.

Second approach

Now we try to prove the Theorem 3 in the other way around: assuming that the rejection probability is low we construct a nearby homomorphism φ. What is important, if f is a homomorphism, there is no other homomorphism close to it, consult Exercise 2.

Knowing that there is a homomorphism ϵ-close to f, we try to construct such a homomorphism φ. Please note that this construction is just a 'mental experiment': given f we are not going to calculate its close homomorphism. We will use it just in the analysis of the Test 1. Firstly fix some $x \in G$. For all $y \in G$ the equality

$$\varphi(x) = \varphi(x + y) - \varphi(y)$$

holds. Since f and φ are ϵ-close

$$\begin{aligned} \mathbb{P}\mathrm{r}_y\left[\varphi(y) = f(y)\right] &= 1 - \epsilon \\ \mathbb{P}\mathrm{r}_y\left[\varphi(x + y) = f(x + y)\right] &= 1 - \epsilon, \end{aligned}$$

and so

$$\mathbb{P}\mathrm{r}_y\left[\varphi(x) = f(x + y) - f(y)\right] \geq 1 - 2\epsilon.$$

This means that with high probability the correct value of $\varphi(x)$ is $f(x+y)-f(y)$. Since we would like to define φ using f the natural attempt is to set $\varphi(x)$ to the mode of $f(x+y)-f(y)$:

$$\varphi(x) := \text{Plurality}\{f(x+y)-f(y) : y\} \ .$$

Now we prove the following:

1. $\delta(f,\varphi) \leq 2\delta_0$
2. $\forall_x \mathbb{P}\mathrm{r}_{y_1,y_2}[f(x+y_1)-f(y_1) \neq f(x+y_2)-f(y_2)] \leq 2\delta_0$
3. φ is a well defined homomorphism

Proof of 1. Define the 'bad' set of x'es as

$$B := \{x : \mathbb{P}\mathrm{r}_y[f(x) \neq f(x-y)-f(y)] \geq \frac{1}{2}\} \ .$$

We want to show that B is a small set. Notice, that if $x \notin B$ then

$$\mathbb{P}\mathrm{r}_y[f(x) = f(x-y)-f(y)] > \frac{1}{2}$$

and hence $f(x)=\varphi(x)$, as the majority of $f(x-y)-f(y)$ is equal to $f(x)$.

The only thing that we know about f is its probability of being rejected by the Test 1, hence we use this information:

$$\mathbb{P}\mathrm{r}_{x,y}[f(x+y) \neq f(x)+f(y) \wedge x \in B] \leq \delta_0$$

as only $f(x+y) \neq f(x)+f(y)$ has probability at most δ_0. Going into conditional probability

$$\mathbb{P}\mathrm{r}_{x,y}[f(x+y) \neq f(x)+f(y) | x \in B] \cdot \mathbb{P}\mathrm{r}_x[x \in B] \leq \delta_0$$

As B defines exactly those x'es, for which $\mathbb{P}\mathrm{r}_y[f(x+y) \neq f(x)+f(y)] \geq \frac{1}{2}$ we obtain that $\mathbb{P}\mathrm{r}_{x,y}[f(x+y) \neq f(x)+f(y) | x \in B] \geq \frac{1}{2}$ and hence $\mathbb{P}\mathrm{r}_x[x \in B] \leq 2\delta_0$, and so B is a small set.

Now if $x \notin B$ then $\varphi(x)=f(x)$, therefore

$$\mathbb{P}\mathrm{r}_x[\varphi(x)=f(x)] \geq \mathbb{P}\mathrm{r}_x[x \notin B] = 1-\mathbb{P}\mathrm{r}_x[x \in B] \geq 1-2\delta_0 \ .$$

We conclude that $\delta(f,\varphi) \leq 2\delta_0$. □

1.3 Exercises

1. Give $f : \mathbb{Z}_n \to \mathbb{Z}_n$ such that

$$\mathbb{P}\mathrm{r}_x[f(x+1)=f(x)+1] \to 1 \quad \text{and} \quad \delta(f,\mathcal{F}) \to 1 \ .$$

2. If $\varphi, \psi : G \to H$ are homomorphisms then $\delta(\varphi, \psi) \geq \frac{1}{2}$

3. Prove that if

$$\mathbb{P}\mathrm{r}_{y_1,y_2}\left[f(x+y_1) - f(y_1) \neq f(x+y_2) - f(y_2)\right] < \frac{4}{9}$$

then

$$\mathbb{P}\mathrm{r}_{y}\left[\varphi(x) \neq f(x+y) - f(y)\right] < \frac{1}{3}\,.$$

4. Prove that for every function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ it holds that $\delta(f, \mathcal{F}) \leq \frac{1}{2}$.

5. Assume that $\{X_i\}_{i=1}^n$ are iiu (independent identically distributed) and $\mathbb{P}\mathrm{r}\left[X_i = 1\right] = 1-p$ and $\mathbb{P}\mathrm{r}\left[X_i = 0\right] = p$. What is the value of $\mathbb{P}\mathrm{r}\left[\oplus_{i=1}^n X_i = 1\right]$?
Hint: Consider $Y_i = (-1)^{X_i}$.