# COMP760, LECTURES 4-5: FOURIER UNIFORMITY

HAMED HATAMI

## 1. Linearity test

Blum, Luby, and Rubinfeld [1] made a beautiful observation that given a function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, it is possible to inquire the value of $f$ on a few random points, and accordingly probabilistically distinguish between the case that $f$ is a linear function and the case that $f$ has to be modified on at least $\epsilon > 0$ fraction of points to become a linear function. Inspired by this observation, Rubinfeld and Sudan [3] defined the concept of property testing which is now a major area of research in theoretical computer science. Roughly speaking to test a function for a property means to examine the value of the function on a few random points, and accordingly (probabilistically) distinguish between the case that the function has the property and the case that it is not too close to any function with that property. Interestingly and to some extent surprisingly these tests exist for various basic properties. The first substantial investigation of property testing occurred in Goldreich, Goldwasser, and Ron [2] who showed that several natural combinatorial properties are testable. Since then there has been a significant amount of research on classifying the testable properties in combinatorial and algebraic settings.

In this section, we will state and analyze the BLR linearity test. We start by formally defining a linear function.

**Definition 1.1.** *A function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ is called* linear *if $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{Z}_2^n$.*

Consider $\epsilon, \delta > 0$. Given a function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, we want to query the value of $f$ on few points to distinguish correctly with probability at least $1 - \delta$ between the following two cases

(1) $f$ is linear.
(2) $f$ is $\epsilon$-far from every linear function. I.e. for every linear $g : \mathbb{Z}_2^n \to \mathbb{Z}_2$,

$$\Pr[f(x) \neq g(x)] \geq \epsilon.$$

In order to apply Fourier analysis, we need the range of $f$ to be $\mathbb{C}$ rather than $\mathbb{Z}_2$. To achieve this we can compose $f$ with an injective homomorphism from $\mathbb{Z}_2$ to $\mathbb{C}$ (which is basically $\chi_1 : x \to (-1)^x$, the only non-principal character of $\mathbb{Z}_2$). In other words $f$ is linear if and only if $(-1)^f$ is multiplicative (i.e. $(-1)^{f(x+y)} = (-1)^{f(x)}(-1)^{f(y)}$) which is equivalent to being a character of $\mathbb{Z}_2^n$. So linearity test is can be reformulated as "character test":

Consider $\epsilon, \delta > 0$. Given a function $f : \mathbb{Z}_2^n \to \{-1, 1\}$, we want to query the value of $f$ on few points to distinguish correctly with probability at least $1 - \delta$ between the following two cases

(1) $f$ is a character.
(2) $f$ is $\epsilon$-far from every character. I.e. for every $a \in \mathbb{Z}_2^n$,

$$\Pr[f(x) \neq \chi_a(x)] \geq \epsilon.$$

Now let us state the BLR test.

```
Blum, Luby, and Rubinfeld's linearity test:
  • Set $N = \lceil \frac{\ln \delta}{\ln(1-\epsilon)} \rceil$.
  • For $i = 1, \ldots, N$
  •     Pick two random points $x, y \in \mathbb{Z}_2^n$.
  •     If $f(x)f(y) \neq f(x+y)$, then Reject.
  • Accept.
```

1.1. **Analysis of the BLR test.** First note that if $f$ is a character then the BLR test always succeeds, that is, it never rejects a character. The bulk of the analysis lies in proving that if $f$ is $\epsilon$-far from every character, then $f$ is rejected with probability at least $1 - \delta$.

Consider a character $\chi_a$, and note that

$$\Pr[f(x) \neq \chi_a(x)] = \frac{1}{2} - \frac{1}{2}\mathbb{E}[f(x)\chi_a(x)] = \frac{1}{2} - \frac{1}{2}\widehat{f}(a).$$

So if $f$ is $\epsilon$-far from every character, then

$$\epsilon \leq \frac{1}{2} - \frac{1}{2}\max_a \widehat{f}(a),$$

or equivalently

$$\max_a \widehat{f}(a) \leq 1 - 2\epsilon. \tag{1}$$

Now let us analyze the probability that $f$ is not rejected in an iteration of the for-loop in the BLR algorithm. Note that

$$\Pr_{x,y}[f(x)f(y) = f(x+y)] = \Pr_{x,y}[f(x)f(y)f(x+y) = 1] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[f(x)f(y)f(x+y)].$$

Replacing $f$ with its Fourier expansion, we get

$$\begin{aligned}\mathbb{E}[f(x)f(y)f(x+y)] &= \mathbb{E}\left[\sum_{a,b,c} \widehat{f}(a)\widehat{f}(b)\widehat{f}(c)\chi_a(x)\chi_b(y)\chi_c(x+y)\right] \\ &= \sum_{a,b,c} \widehat{f}(a)\widehat{f}(b)\widehat{f}(c)\mathbb{E}_x[\chi_{a+c}(x)]\mathbb{E}_y[\chi_{a+b}(y)].\end{aligned}$$

Note that $a + c = 0$ if and only if $a = c$, and thus (see Lemma 1.2 from Lecture 2),

$$\mathbb{E}[\chi_{a+c}(x)] = \begin{cases} 1 & a = c \\ 0 & a \neq c \end{cases}$$

Similarly

$$\mathbb{E}[\chi_{b+c}(y)] = \begin{cases} 1 & b = c \\ 0 & a \neq c \end{cases}$$

Hence

$$\mathbb{E}[f(x)f(y)f(x+y)] = \sum_a \widehat{f}(a)^3,$$

which shows that

$$\Pr_{x,y}[f(x)f(y) = f(x+y)] = \frac{1}{2} + \frac{1}{2}\sum_a \widehat{f}(a)^3 \leq \frac{1}{2} + \frac{1}{2}\left(\max_a \widehat{f}(a)\right)\sum_a \widehat{f}(a)^2. \tag{2}$$

By the Parseval identity

$$\sum_{a \in G} \widehat{f}(a)^2 = \|f\|_2^2 = 1.$$

So

$$\Pr_{x,y}[f(x)f(y) = f(x+y)] \leq \frac{1}{2} + \frac{1}{2}\max_a \widehat{f}(a). \tag{3}$$

Now to finish the proof note that by (1) and (3) if $f$ is $\epsilon$-far from every character, then

$$\Pr_{x,y}[f(x)f(y) = f(x+y)] \leq 1-\epsilon.$$

Thus the probability that BLR makes a mistake and accepts $f$ is at most $(1-\epsilon)^N = (1-\epsilon)^{\lceil \frac{\ln \delta}{\ln(1-\epsilon)} \rceil} \leq \delta$.

## 2. Pseudorandomness

The concept of pseudo-randomness is extremely important to theoretical computer science and many areas of mathematics. Pseudo-randomness refers to the property of behaving similar to a typical random structure or process in a certain specified way. For example, a random sequence of zeros and ones typically have roughly the same number of zeros and ones. So any zero-one sequence which has roughly the same number of zeros and ones is pseudo-random in this sense. However, even for this very weak notion of pseudo-randomness, there is a sequence such that establishing its pseudo-random behavior is one of the most notorious open problems of mathematics.

Let us be more precise here. Consider a sequence $a = (a_1, a_2, a_3, \ldots)$ where each $a_i$ is chosen randomly and independently from the set $\{-1, 0, 1\}$. Fix an arbitrary $\epsilon > 0$. For every $n \geq 1$, let $S_n(a) = |\sum_{i=1}^n a_i|$ be the absolute value of the sum of the first $n$ elements of the sequence. It follows from the Chernoff bound (See Lecture 1, Section 4) that

$$\Pr[|S_n(a)| \geq n^{\frac{1}{2}+\epsilon}] \leq 2e^{-\frac{n^{2\epsilon}}{3}}. \tag{4}$$

Then it follows from the Borel-Cantelli theorem (See Lecture 1, Theorem 2.8) that with probability 1, we have

$$S_n(a) = O(n^{\frac{1}{2}+\epsilon}). \tag{5}$$

Recall that the Möbius function $\mu : \mathbb{N} \to \{-1, 0, 1\}$ is defined as

$$\mu(n) = \begin{cases} 0 & p^2 | n \text{ for some prime } p; \\ (-1)^k & n = p_1 \ldots p_k \text{ for distinct primes } p_1, \ldots, p_k. \end{cases}$$

Then the famous Riemann hypothesis conjecture is equivalent to saying that the Möbius function is pseudo-random in the sense of (5). That is, $S_n(\mu) = O(n^{\frac{1}{2}+\epsilon})$ for every $\epsilon > 0$.

For a notion of pseudo-randomness to be useful, it must imply that a pseudo-random structure behaves similar to a random structure in various ways (not just what's required in the definition of the notion). For example, this is the case with the Riemann hypothesis conjecture. If it is true, it would imply that the set of primes behave similar to certain random sets in various different ways. This is exactly where the importance of the conjecture lies. It would enable us to settle various important conjectures about the prime numbers.

## 3. Fourier Uniformity

In this section we study one of the important notions of the pseudo-randomness called Fourier uniformity. Consider the group $G = \mathbb{Z}_2^n$ and a random subset $A \subseteq \mathbb{Z}_2^n$ (equivalently a random function $f : \{0,1\}^2 \to \{0,1\}$). By Hoeffding's Inequality (See Lecture 1, Section 4), for $a \in \mathbb{Z}_2^n$ with $a \neq 0$, we have

$$\Pr\left[|\widehat{1_A}(a)| > \alpha\right] = \Pr\left[\left|\sum_{x \in G} 1_A(x)\chi_a(x)\right| > \alpha|G|\right] \leq 2e^{\frac{-\alpha^2|G|^2}{2|G|}} = 2e^{-\alpha^2 2^{n-1}}.$$

Then the union bound implies that

$$\Pr\left[\max_{a \neq 0} |\widehat{1_A}(a)| > \alpha\right] \leq 2^{n+1}e^{-\alpha^2 2^{n-1}}.$$

By setting $\alpha = 2^{-n/2}\sqrt{2n}$, we have

$$\Pr\left[\max_{a \neq 0} |\widehat{1_A}(a)| > 2^{-n/2}\sqrt{2n}\right] = o_{n \to \infty}(1). \tag{6}$$

So typically all non-principal Fourier coefficients of a random subset of $\mathbb{Z}_2^n$ have smaller magnitudes than $2^{-n/2}\sqrt{2n}$. Note that $2^{-n/2}\sqrt{2n}$ is quite small. Indeed by Parseval's identity, for any arbitrary $A \subseteq G$ with $0 < |A|/|G| = c < 1$ where $c$ is a fixed constant, we have

$$c = \|1_A\|_2^2 = \sum_{a \in G} |\widehat{1_A}(a)|^2 = |\widehat{1_A}(0)|^2 + \sum_{a \neq 0} |\widehat{1_A}(a)|^2 = c^2 + \sum_{a \neq 0} |\widehat{1_A}(a)|^2$$

which shows that

$$\max_{a \neq 0} |\widehat{1_A}(a)| \geq 2^{-n/2}c(1-c).$$

Then our upper bound (6) of $2^{-n/2}\sqrt{2n}$ for the maximum non-principal Fourier coefficient of a random $A$ is not very far from the lower bound $2^{-n/2}c(1-c)$.

We established that typically all non-principal Fourier coefficients of a random subset of $G \subseteq Z_2^n$ are of small magnitude. This fact can be easily generalized to all finite Abelian groups. Inspired by this, we would like to to call a subset $A \subseteq G$ pseudo-random if all of its non-principal Fourier coefficients are of small magnitude.

# Lecture 5

It is more constructive to formalize this in terms of the $U^2$ norm.

**Definition 3.1.** *Let $G$ be a finite Abelian group, and $f : G \to \mathbb{C}$ be a function. The $U^2$ norm of $f$ is defined as*

$$\|f\|_{U^2} = \left(\mathbb{E} f(x)\overline{f(x+y)f(x+z)}f(x+y+z)\right)^{1/4}, \tag{7}$$

*where $x, y, z$ are random elements of $G$.*

We need to establish that the $U^2$ norm is actually a norm. Note that a priori it is not even clear that the expected value in the right-hand side of (7) is a non-negative real number. Replacing $f$

by its Fourier expansion and expanding, we have

$$\begin{aligned}
\|f\|_{U^2}^4 &= \mathbb{E} f(x)\overline{f(x+y)f(x+z)}f(x+y+z) \\
&= \sum_{a,b,c,d} \widehat{f}(a)\overline{\widehat{f}(b)\widehat{f}(c)}\widehat{f}(d)\mathbb{E}\left[\chi_a(x)\chi_{-b}(x+y)\chi_{-c}(x+z)\chi_d(x+y+z)\right] \\
&= \sum_{a,b,c,d} \widehat{f}(a)\overline{\widehat{f}(b)\widehat{f}(c)}\widehat{f}(d)\mathbb{E}\left[\chi_{a-b-c+d}(x)\chi_{-b+d}(y)\chi_{-c+d}(z)\right].
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{E}\left[\chi_{a-b-c+d}(x)\chi_{-b+d}(y)\chi_{-c+d}(z)\right] &= \begin{cases} 1 & a-b-c+d=0, -b+d=0, -c+d=0; \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} 1 & a=b=c=d; \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}$$

we conclude that $\|f\|_{U^2}^4 = \sum_{a\in G} |\widehat{f}(a)|^4$ which in turn implies

$$\|f\|_{U^2} = \left(\sum_{a\in G} |\widehat{f}(a)|^4\right)^{1/4} = \left\|\widehat{f}\right\|_4 = \|f * f\|_2^{1/2}. \tag{8}$$

So $U^2$ norm is nothing but the $L_4$ norm of the Fourier transform of $f$.

**Exercise 3.2.** Prove the identity $\|f\|_{U^2} = \|f * f\|_2^{1/2}$ directly without using the Fourier expansion of $f$. ■

Now that we introduced the $U^2$ norms we can define the relevant notion of pseudo-randomness called *Fourier uniformity* (or *uniformity* for short).

**Definition 3.3.** *A function $f : G \to \mathbb{C}$ is called $\delta$-uniform if $\|f\|_{U^2} \le \delta$. A subset $A \subseteq G$ is called $\delta$-uniform if and only if $1_A - \mathbb{E}1_A$ is $\delta$-uniform.*

How does this notion of uniformity relate to our discussion about the maximum non-principal Fourier coefficient? The following simple lemma relates the $U^2$ norm to the largest Fourier coefficient of $f$.

**Lemma 3.4.** *Let $G$ be a finite Abelian group, and $f : G \to \mathbb{C}$ satisfy $|f| \le 1$. Then*

$$\|\widehat{f}\|_\infty \le \|f\|_{U^2} \le \sqrt{\|\widehat{f}\|_\infty}.$$

*Proof.* By (8)

$$\|f\|_{U^2}^4 = \sum_{a\in G} |\widehat{f}(a)|^4 \ge \max_{a\in G} |\widehat{f}(a)|^4 = \|\widehat{f}\|^4,$$

which establishes the first inequality. To prove the second inequality, note that by the Parseval identity $\sum_{a\in G} |\widehat{f}(a)|^2 = \|f\|_2 \le 1$. Hence

$$\|f\|_{U^2}^4 = \sum_{a\in G} |\widehat{f}(a)|^4 \le \left(\max_{a\in G} |\widehat{f}(a)|^2\right) \sum_{a\in G} |\widehat{f}(a)|^2 \le \max_{a\in G} |\widehat{f}(a)|^2 = \|\widehat{f}\|_\infty^2.$$

□

If $A \subseteq G$, then $f = 1_A - \mathbb{E}1_A = \sum_{a \neq 0} \widehat{1_A}(a)\chi_a$ satisfies the condition $|f| \leq 1$ of Lemma 3.4. Hence $A$ is uniform if and only if all non-principal Fourier coefficients of $1_A$ are of small magnitude. Note that this is a proper notion of pseudo-randomness in the sense that random subsets of $G$ are typically very uniform. For example (6) shows that with high probability a random subset of $\mathbb{Z}_2^n$ is $2^{-n/4}(2n)^{1/4}$-uniform.

3.1. **Testing Uniformity.** The averages similar to $\mathbb{E}f(x)\overline{f(x+y)f(x+z)}f(x+y+z)$ appear quite naturally in the area of algebraic property testing. It is possible to estimate them by randomly and independently selecting a number of random points and computing the corresponding empirical average. For example, a second moment argument (Lemma 4.2, Lecture 1) implies the following observation about the $U^2$ norm.

**Lemma 3.5.** *Let $G$ be a finite Abelian group, and $f : G \to \mathbb{C}$ satisfy $|f| \leq 1$. Suppose that $x_1, \ldots, x_N, y_1, \ldots, y_N, z_1, \ldots, z_N \in G$ are chosen independently uniformly at random. Then denoting*

$$\eta = \frac{1}{N}\sum_{i=1}^{N} f(x_i)\overline{f(x_i+y_i)f(x_i+z_i)}f(x_i+y_i+z_i),$$

*we have for every $\epsilon > 0$,*

$$\Pr[\big|\|f\|_{U^2}^4 - \eta\big| > \epsilon] \leq \frac{1}{N\epsilon^2} = o_{N\to\infty}(1).$$

So from the point of view of the property testing, by Lemma 3.4 it is possible to distinguish between the following two cases: $\|\widehat{f}\|_\infty$ is tiny *vs* $\|\widehat{f}\|_\infty$ is non-negligible.

3.2. **3-term Arithmetic Progressions.** As we mentioned earlier for a notion of pseudo-randomness to be useful, it must imply that a pseudo-random structure behaves similar to a random structure in various ways. To illustrate the usefulness of $\delta$-uniformity, let us mention one example.

Define $t_3(A)$ the density of 3-term arithmetic progressions in a set $A \subseteq \mathbb{Z}_N$ as the probability that $x, x+y, x+2y \in A$ for randomly chosen $x, y \in \mathbb{Z}_N$. Note that

$$t_3(A) = \mathbb{E}_{x,y}1_A(x)1_A(x+y)1_A(x+2y) \tag{9}$$

$$= \sum_{a,b,c}\widehat{1}_A(a)\widehat{1}_A(b)\widehat{1}_A(c)\mathbb{E}[\chi_a(x)\chi_b(x+y)\chi_c(x+2y)] \tag{10}$$

$$= \sum_{a,b,c}\widehat{1}_A(a)\widehat{1}_A(b)\widehat{1}_A(c)\mathbb{E}[\chi_{a+b+c}(x)\chi_{b+2c}(y)]. \tag{11}$$

Since

$$\begin{aligned}\mathbb{E}\left[\chi_{a+b+c}(x)\chi_{b+2c}(y)\right] &= \begin{cases} 1 & a+b+c=0, b+2c=0; \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & a=b=-2c; \\ 0 & \text{otherwise,} \end{cases}\end{aligned}$$

we get

$$t_3(A) = \sum_c \widehat{1_A}(c)^2\widehat{1_A}(-2c) = \widehat{1_A}(0)^3 + \sum_{c\neq 0}\widehat{1_A}(c)^2\widehat{1_A}(-2c).$$

Denoting $\alpha = \widehat{1_A}(0) = |A|/N$, we conclude that

$$\begin{aligned}
\left|t_3(A) - \alpha^3\right| &= \left|\sum_{c\neq 0} \widehat{1_A}(c)^2\widehat{1_A}(-2c)\right| \leq \left(\max_{c\neq 0}|\widehat{1_A}(c)|\right)\sum_{c\neq 0}\left|\widehat{1_A}(c)\widehat{1_A}(-2c)\right| \\
&\leq \left(\max_{c\neq 0}|\widehat{1_A}(c)|\right)\left(\sum_{c\neq 0}|\widehat{1_A}(c)|^2\right)^{1/2}\left(\sum_{c\neq 0}|\widehat{1_A}(-2c)|^2\right)^{1/2} \\
&\leq \left(\max_{c\neq 0}|\widehat{1_A}(c)|\right)\left(\sum_{c\neq 0}|\widehat{1_A}(c)|^2\right)^{1/2}\left(2\sum|\widehat{1_A}(c)|^2\right)^{1/2} \\
&\leq \left(\max_{c\neq 0}|\widehat{1_A}(c)|\right)\|1_A\|_2\|1_A\|_2 \leq \sqrt{2}\max_{c\neq 0}|\widehat{1_A}(c)|,
\end{aligned}$$

where the first inequality is by Cauchy-Schwarz, and the constant 2 appears in the third line because $-2c_1 = -2c_2$ has at most two solutions in $\mathbb{Z}_N$ (when $N$ is a prime, this constant can be removed).

In particular when $A$ is $\delta$-uniform we have

$$\left|t_3(A) - \alpha^3\right| \leq \sqrt{2}\max_{c\neq 0}|\widehat{1_A}(c)| \leq \sqrt{2}\|1_A - \mathbb{E}[1_A]\|_{U^2} \leq \sqrt{2}\delta.$$

This shows that a sufficiently uniform subset of $\mathbb{Z}_N$ has the same density of 3-term arithmetic progressions (which is $\alpha^3$) as a random subset of $\mathbb{Z}_N$ with density $\alpha$.

## 4. Probably approximately correct learning of Fourier Coefficients

In computational learning theory, probably approximately correct learning (PAC learning) is a framework for mathematical analysis of machine learning. It was proposed in 1984 by Leslie Valiant. We will not formally define the PAC learning. We just mention that unlike in property testing, here we are not allowed to inquire the value of the function on the points that we want. Instead the points are drawn from the domain according to a fixed distribution (usually uniform) and we are only allowed to observe the value of the function on these points. Note that we do not have any control over the distribution.

Consider a function $f : G \to \mathbb{C}$ which satisfies $|f| \leq 1$. Suppose that points $x$ are drawn from $G$ according to the uniform distribution, and for a fixed $a \in G$, we would like to estimate $\widehat{f}(a)$. This can be easily done by the following lemma which is a direct consequence of (Lemma 4.2, Lecture 1).

**Lemma 4.1.** *Let $G$ be a finite Abelian group, and $f : G \to \mathbb{C}$ satisfy $|f| \leq 1$. Suppose that $x_1, \ldots, x_N \in G$ are chosen independently uniformly at random. Then denoting*

$$\eta = \frac{1}{N}\sum_{i=1}^{N} f(x_i)\overline{\chi_a(x_i)},$$

*we have for every $\epsilon > 0$,*

$$\Pr[\left|\widehat{f}(a) - \eta\right| > \epsilon] \leq \frac{1}{N\epsilon^2}.$$

**Remark 4.2.** Applying Hoeffding's Inequality instead of (Lemma 4.2, Lecture 1) would improve the bound in Lemma 4.1 from $\frac{1}{N\epsilon^2}$ to $2e^{\frac{-N}{2\epsilon^2}}$. ■

## References

[1] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. In *STOC '90: Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 73–83, New York, NY, USA, 1990. ACM.

[2] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[3] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials and their applications to program testing. Technical report, Ithaca, NY, USA, 1993.

School of Computer Science, McGill University, Montréal, Canada
*E-mail address*: hatami@cs.mcgill.ca