LAST TIME

Proof complexity

Proof system P for L. Binary, poly-time computable predicate $P(x, \pi)$ such that

- $x \in L \Rightarrow \exists \pi$ s.t. $P(x, \pi) = 1$
- $x \notin L \Rightarrow \forall \pi'$ $P(x, \pi') = 0$

Proof system is polynomially bounded if in addition proofs can be chosen of size $|\pi| \leq poly(|x|)$

NP is exactly the set of languages with polynomially bounded proof systems

Propositional proof system proof system for $\overline{\text{CNFSAT}}$.

Reasons for proof cplx (among others):

(1) Prove stronger and stronger lower bounds
$\rightsquigarrow$ No p-bounded propositional proof system
$\rightsquigarrow$ $P \neq NP$

Fascinating line of research, but gets stuck pretty early

(2) Connections to SAT solving
Study proof systems used by state-of-the-art SAT solvers
Understand potential and limitations of these solvers.

Most popular by far: RESOLUTION

Proved exponential lower bounds on resolution proofs for the pigeonhole principle.

PROPERTY TESTING

"Given some large object $\mathcal{O}$, want to look at small part of it and decide whether it has some property $\mathcal{P}$"

What is a property?

Ex 1 Given an (undirected) graph G, is is bipartite? I.e. can we partition $V(G)$ into $V_1 \dot\cup V_2$ so that all edges go between V_1 and V_2?

Encode graph G on n vertices as

$$f_G : [n] \times [n] \to \{0,1\}$$

$$f_G(i,j) = \begin{cases} 1 & \text{if } (i,j) \in E(G) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{BIPARTITENESS} = \{ f \mid f \text{ encodes a bipartite graph} \}$$

Ex 2 Fix vector space V over field $\mathbb{F}$ (Think of ~~$\mathbb{R}$~~ $\mathbb{R}^n$ for now; will get specific soon)

f is LINEAR if

$$\forall x,y \in V \;\; \forall \alpha, \beta \in \mathbb{F} \quad f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

$$\text{LINEARITY} = \{ f : V \to \mathbb{F} \mid f \text{ is a linear function} \}$$

DEF 3 Let D_n and R_n, $n \in \mathbb{N}^+$, be some (growing with n) domains and ranges of finite size. Suppose for every n that $\mathcal{P}_n$ is some subset of $\{ f : D_n \to R_n \}$

Then $\mathcal{P} = \bigcup_{n \in \mathbb{N}^+} \mathcal{P}_n$ is a PROPERTY

(Often have $R_n = R$ for all $n \in \mathbb{N}^+$)

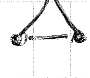
A property is really just a language, but where we think of the strings as representing function tables

PROPERTY TESTING, ATTEMPT 1
Given property P and function f, look at a few function values of f and decide (whether $f \in P$ correctly)

Impossible.
Construct a graph that looks bipartite on the edges inspected, then add other (not inspected) edges making graph non-bipartite

PROPERTY TESTING, ATTEMPT 2
Given property P and function f, RANDOMLY look at few function values of f and output 0 / 1

If $f \in P$, then $\Pr[\text{output is } 1] \geq 2/3$
If $f \notin P$, then $\Pr[\text{output is } 0] \geq 2/3$

Also impossible.
Take bipartite graph G. Add just one triangle 
Vanishingly small probability of seeing this triangle.
Otherwise everything looks fine.

<u>More realistic goal</u>: distinguish between
a) $f \in P$
b) f is <u>far</u> from being in P

What does it mean to be "far"?

DEF 4 For $f, g : D_n \to R_n$, the <u>distance</u> $\delta(f,g)$ is

$$\delta(f,g) = \Pr_{x \in_R D_n}\left[f(x) \neq g(x) \right]$$

f and g are <u>ε-close</u> if $\delta(f,g) \leq \varepsilon$

<u>ε-far</u> if $\delta(f,g) > \varepsilon$

The distance from f to $\mathcal{P}$ is

$$\delta(f, \mathcal{P}) = \min_{g \in \mathcal{P}} \delta(f,g)$$

DEF 5 A PROPERTY TESTER for a property $\mathcal{P}$ with query complexity $q(n)$ is an algorithm T that given $f : D_n \to R_n$ $\boxed{\text{and } \varepsilon > 0}$ $\boxed{\text{makes}}$ $q(n)$ random access queries* $f(x_1), f(x_2), \ldots, f(x_{q(n)})$ and outputs $1/0$ (accepts/reject) and that satisfies

(a) If $f \in \mathcal{P}$, then $\Pr[\,T \text{ accepts } f\,] \geq 2/3$

(b) If ~~$f \notin \mathcal{P}$~~ f is ε-far from $\mathcal{P}$, then $\Pr[\,T \text{ accepts } f\,] \leq 1/3$

T has ONE-SIDED ERROR if $\Pr[T \text{ accepts}] = 1$ for $f \in \mathcal{P}$.

T is NON-ADAPTIVE if all query points $x_1, x_2, \ldots$ are decided in advance without seeing any function values $f(x_i)$

(*) This is called that T has <u>oracle access</u> to f (can't see all of f, but can ask an "oracle" for function values $f(x_i)$

Note that there are no restrictions on the __running time__ of T - the focus is on the __query complexity__.

__DEF 6__ A property $\mathcal{P}$ is STRONGLY TESTABLE if it has a one-sided tester with query complexity $q(n) = O(1)$.

That is, # queries can (and will) depend on distance parameter ε, but is independent of the size of the domain and range!

__PROP 7__ If $\mathcal{P} = \bigcup_{n \in \mathbb{N}^+} \mathcal{P}_n$, $\mathcal{P}_n \subseteq \{f : D_n \to R\}$ is strongly testable, then $\mathcal{P}$ has a non-adaptive one-sided tester with constant query complexity.

__Proof__ Straightforward exercise.

PROPERTY TESTING

- Whole (sub) area of TCS with lots of research and papers since mid-90s
- Applications to / relations with
  - Probabilistically checkable proofs PCPs
  - Coding theory
  - Computational learning theory
  - Computational geometry
  - Combinatorics
  - Statistics [et cetera]
- Also relevant setting for Internet-age "big data"

<u>Goal for today</u> Show that LINEARITY is strongly testable

Focus on $f : \{0,1\}^n \to \{0,1\}$ $\quad D_n = \{0,1\}^n$, $R_n = \{0,1\}$

Vector space over $GF(2)$ with addition mod 2

For $x, y \in \{0,1\}^n$ $\quad x+y = (x_1+y_1, \ldots, x_n+y_n)$

$0+0=0$
$0+1=1$
$1+0=1$
$1+1=0$

f is LINEAR if $\forall x \forall y$ $\boxed{f(x+y) = f(x) + f(y)}$

<u>PROP 8</u>

$f : \{0,1\}^n \to \{0,1\}$ is linear iff $\exists \alpha \subseteq [n]$ such that $f(x) = \sum_{i \in \alpha} x_i$

<u>Proof</u> Also exercise.

How to test linearity?

Naive idea: Pick $x, y \in \{0,1\}^n$ uniformly and independently at random and check if $f(x+y) = f(x) + f(y)$

Clearly one-sided test — will always accept linear functions

But is it any good at accepting non-linearity?

Distance from linearity = fraction of bits needed to be flipped in function table to get to closest linear function

Warm-up

(1) If tester accepts with probability 1, then f is linear. (Why?)

(2) Suppose f uniformly random function (all function values chosen at random. Then acceptance with prob 1/2 (although f probably $\frac{1}{2}$-far from linear)

But what if the test accepts with prob $1-\varepsilon$? Want to argue that f must be ε'-close to linear, since otherwise the tester should be likely to reject

THEOREM 9 [Blum, Luby, & Rubinfeld '90] [Bellare, Coppersmith, Håstad, Kiwi & Sudan '96]

If $\Pr_{x,y \in_R \{0,1\}^n}\left[f(x+y) = f(x)+f(y)\right] \geq 1-\varepsilon$ for $\varepsilon < 1/2$, then f is ε-close to a linear function.

COROLLARY 10

LINEARITY (for functions $f: \{0,1\}^n \to \{0,1\}$) is strongly testable.

Proof If f is ε-far from linear, then randomly testing $f(x+y) = f(x)+f(y)$ detects a violation with probability at least ε. ~~Repeating~~ the test k times will lead to acceptance with prob $\leq (1-\varepsilon)^k$. Choose k constant so that $(1-\varepsilon)^k \leq 1/3$. $\boxed{\square}$

How to prove something like Thm 9? Discrete Fourier analysis. Why? It works...

FOURIER TRANSFORM OVER $\{0,1\}^n$ (a.k.a. $GF(2)^n$)

First small conceptual shift

Instead of $\{0,1\}$, consider $\{+1,-1\}$

Mapping $\quad b \mapsto (-1)^b \qquad \begin{matrix} 0 \mapsto 1 \\ 1 \mapsto -1 \end{matrix}$

Addition $\rightsquigarrow$ Multiplication

$b_1 + b_2 \qquad (-1)^{b_1} \cdot (-1)^{b_2} = (-1)^{b_1+b_2}$

$f : \{\pm 1\}^n \to \{\pm 1\}$ is "LINEAR" iff $\exists\, \alpha \subseteq [n]$ such that $f(x) = \prod_{i \in \alpha} x_i$

Set of functions $f : \{\pm 1\}^n \to \mathbb{R}$ define a 2^n-dimensional vector space

- $(f+g)(x) = f(x) + g(x)$
- $(cf)(x) = c f(x) \qquad c \in \mathbb{R}$

Inner product

$$\langle f, g \rangle = E_{x \in_R \{\pm 1\}^n} \left[f(x)\, g(x) \right]$$

$$= \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x) g(x)$$

The standard basis for this vector space are the functions $\{e_x\}_{x \in \{\pm 1\}^n}$ where

$$e_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}$$

Orthogonal basis. Every f can be written

$$f = \sum_{x \in \{\pm 1\}^n} a_x e_x$$

where $a_x = f(x)$

Fourier basis

For all $\alpha \subseteq [n]$ $\quad \chi_\alpha(x) = \prod_{i \in \alpha} x_i$

$$\left[\chi_\emptyset(x) = 1\right]$$

[All linear functions.]

PROPOSITION 11

$\{\chi_\alpha\}_{\alpha \subseteq [n]}$ is an orthonormal basis for the vector space.

Proof $\quad \{\chi_\alpha\}_{\alpha \subseteq [n]}$ contain 2^n elements = dimension.

They are all linearly independent since

$$\langle \chi_\alpha, \chi_\beta \rangle = \delta_{\alpha,\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{otherwise} \end{cases}$$

Let $\gamma = \alpha \Delta \beta = (\alpha \cup \beta) \setminus (\alpha \cap \beta)$

Then if $\gamma \neq \emptyset$ it holds that

$$\langle \chi_\alpha, \chi_\beta \rangle = \frac{1}{2^n} \sum_x \prod_{i \in \gamma} x_i = 0 \qquad (\dagger)$$

Proving $(\dagger)$ is left as an exercise.

Hence, every $f : \{\pm 1\}^n \to \mathbb{R}$ can be represented in the Fourier basis as

Fourier coefficients

$$f = \sum_{\alpha \subseteq [n]} \hat{f}_\alpha \cdot \chi_\alpha \qquad (*)$$

LEMMA 12

(1) $\quad \langle f, g \rangle = \sum_\alpha \hat{f}_\alpha \hat{g}_\alpha$

(2) $\quad \langle f, f \rangle = \sum_\alpha \hat{f}_\alpha^2 \qquad$ (Parseval's identity)

Proof (2) follows from (1). To prove (1), just expand in the Fourier basis

$$\langle f, g \rangle = \left\langle \sum_\alpha \hat{f}_\alpha \chi_\alpha , \sum_\beta \hat{g}_\beta \chi_\beta \right\rangle \quad \left(\begin{array}{c}\text{linearity of}\\ \text{inner product}\end{array}\right)$$

$$= \sum_\alpha \sum_\beta \hat{f}_\alpha \hat{g}_\beta \langle \chi_\alpha , \chi_\beta \rangle \quad (\text{orthonormality}\ \cancel{\text{orthogonality}})$$

$$= \sum_\alpha \hat{f}_\alpha \hat{g}_\alpha \qquad \boxed{\text{LL}}$$

EXAMPLE 13

$$MAJ(u_1, u_2, u_3) = \begin{cases} +1 & \text{if at least } \cancel{\text{two}} \text{ of } u_1, u_2, u_3 = 1 \\ -1 & \text{otherwise} \end{cases}$$

$$MAJ(u_1, u_2, u_3) = \frac{1}{2} u_1 + \frac{1}{2} u_2 + \frac{1}{2} u_3 - \frac{1}{2} u_1 u_2 u_3$$

So $\widehat{MAJ}_{\{1\}} = \widehat{MAJ}_{\{2\}} = \widehat{MAJ}_{\{3\}} = \frac{1}{2}$

$\widehat{MAJ}_{\{1,2,3\}} = -\frac{1}{2}$

For all other $\alpha \subseteq [n]$ $\quad \widehat{MAJ}_\alpha = 0$

Back to proof of Thm 9, but in $\{\pm 1\}$-setting

Linear functions = Fourier basis functions

Let $x \cdot y = (x_1 y_1, x_2 y_2, \dots, x_n y_n)$

$f(x+y)$ in $\{0,1\}$-setting $=$ $f(xy)$ in $\{\pm 1\}$-setting

Note that for basis functions we have

$$\chi_\alpha(xy) = \chi_\alpha(x) \cdot \chi_\alpha(y) \quad \left(\begin{array}{c}\text{This is what}\\ \text{linearity means}\end{array}\right)$$

Clearly, if function f is linear, then it has some Fourier coefficient $= 1$ (namely $\cancel{\chi_\alpha}$ $\hat{f}_\alpha$ if $f = \chi_\alpha$)

What if f is <u>somewhat</u> close to linear? Does it have a somewhat large Fourier coefficient?

$$\langle f, g\rangle = \frac{1}{2^n}\sum_{x\in\{\pm 1\}^n} f(x)g(x)$$

$$= \begin{pmatrix}\text{fraction of inputs on} \\ \text{which } f \,\&\, g \text{ \underline{agree}}\end{pmatrix} - \begin{pmatrix}\text{fraction of inputs on} \\ \text{which } f \,\&\, g \text{ \underline{disagree}}\end{pmatrix} \quad (\dagger)$$

Let's say that f & g <u>have agreement $1-\gamma$</u> if f & g are at distance exactly γ

[1-agreement = same function = 0-close]

By $(\dagger)$ f has agreement $\frac{1}{2}+\frac{\varepsilon}{2}$ with g iff $\langle f, g\rangle = \varepsilon$ $(**)$

But Fourier coefficients measure agreement with linear functions!

$$\hat{f}_\beta = \langle f, \chi_\beta\rangle = \left\langle \sum_\alpha \hat{f}_\alpha \chi_\alpha, \chi_\beta\right\rangle$$

$$= \sum_\alpha \hat{f}_\alpha \langle \chi_\alpha, \chi_\beta\rangle$$

and use orthonormality

Thus, f has significant agreement with a linear function if f has a large Fourier coefficient.

The fact that we are using Fourier basis (and not standard basis) helps us see from the coordinates in the Fourier expansion how close to linear a function is!

More concretely

f is γ-close to a linear function – let's write $\gamma = \frac{1-\varepsilon}{2}$ –

IFF

f has agreement $\geq 1-\gamma = \frac{1}{2} + \frac{\varepsilon}{2}$

with some linear function

IFF (by (**))

$\exists\, \alpha$ s.t. $\widehat{f_\alpha} = \langle f, \chi_\alpha \rangle \geq \varepsilon$

Thus, we can rewrite Thm 9 as follows.

THEOREM 14

If $f : \{\pm 1\}^n \to \{\pm 1\}$ satisfies $(0 < \varepsilon \leq 1/2)$

$$\Pr_{x,y \in_R \{\pm\}^n} \left[f(xy) = f(x) f(y) \right] \geq \frac{1}{2} + \varepsilon,$$

then there is some $\alpha \subseteq [n]$ such that $\widehat{f_\alpha} \geq 2\varepsilon$.

If we can prove this it follows that the only reason the linearity test succeeds for f with significant probability is that f is close to a linear function.

So if f is <u>far</u> from a linear function, the test must <u>reject</u> with decent probability.

Now all that remains is the calculations needed to prove Thm 14.

Proof of Thm 14

$$\mathcal{E}_{x,y}\left[f(x y)f(x)f(y)\right] =$$

$$= 1 \cdot \Pr\left[f(xy) = f(x)f(y)\right] - 1 \cdot \Pr\left[f(xy) \neq f(x)f(y)\right]$$

$$\geq \left(\frac{1}{2}+\varepsilon\right) - \left(\frac{1}{2}-\varepsilon\right) = 2\varepsilon$$

$$\boxed{\chi_\alpha(xy) = \chi_\alpha(x)\chi_\alpha(y)}$$

Since f is ± 1-valued, $\langle f, f\rangle = 1$.

$$2\varepsilon \leq \mathcal{E}_{x,y}\left[f(xy)\,f(x)\,f(y)\right]$$ (Fourier-expand)

$$= \mathcal{E}_{x,y}\left[\left(\sum_\alpha \hat{f}_\alpha \chi_\alpha(xy)\right)\left(\sum_\beta \hat{f}_\beta \chi_\beta(x)\right)\left(\sum_\gamma \hat{f}_\gamma \chi_\gamma(y)\right)\right]$$

$$= \mathcal{E}_{x,y}\left[\sum_{\alpha,\beta,\gamma} \hat{f}_\alpha \hat{f}_\beta \hat{f}_\gamma \chi_\alpha(x)\chi_\alpha(y)\chi_\beta(x)\chi_\gamma(y)\right]$$ (linearity of expectation)

$$= \sum_{\alpha,\beta,\gamma} \hat{f}_\alpha \hat{f}_\beta \hat{f}_\gamma \mathcal{E}_{xy}\left[\chi_\alpha(x)\chi_\alpha(y)\chi_\beta(x)\chi_\gamma(y)\right]$$ (independence)

$$= \sum_{\alpha,\beta,\gamma} \hat{f}_\alpha \hat{f}_\beta \hat{f}_\gamma \underbrace{\mathcal{E}_x\left[\chi_\alpha(x)\chi_\beta(x)\right]}_{\text{inner product}} \underbrace{\mathcal{E}_y\left[\chi_\alpha(y)\chi_\gamma(y)\right]}_{\text{inner product}}$$

[Because of orthogonality, vanishes except when $\alpha = \beta = \gamma$]

$$= \sum_\alpha \hat{f}_\alpha^{\,3} \leq$$

$$\leq \left(\max_\alpha \hat{f}_\alpha\right) \cdot \sum_\alpha \hat{f}_\alpha^{\,2}$$

By Parseval
$$\sum_\alpha \hat{f}_\alpha = \langle f, f\rangle = 1$$

$$= \max_\alpha \hat{f}_\alpha \,.$$

Hence

$$\max_\alpha \hat{f}_\alpha \geq 2\varepsilon$$

and the theorem follows. $\boxed{}$

TAKE-HOME MESSAGE

- Property testing: Yet another exciting subarea of TCS
- Lots and lots of research — we only saw one example, namely
- Linearity of functions $f: \{0,1\}^n \to \{0,1\}$ is strongly testable (incidentally, important for PCP constructions).
- Proved by Fourier analysis: very powerful tool, but not very easy to understand why it works and how...