PCP theorem: exponential-size proofs

CS 6810—David Steurer

Spring 2016

1 Introduction

Let $\mathbb{F}_2$ be the field with two elements (the set $\{0,1\}$ with addition and multiplication modulo 2). For $n \in \mathbb{N}$, let $\mathbb{F}_2^n$ be the vector space of n-dimensional tuples over this field.

Recall the following NP-complete problem:

Problem (QuadEq). Given a matrix $A \in \mathbb{F}_2^{m \times n^2}$ and a vector $b \in \mathbb{F}_2^m$, find an assignment $w \in \mathbb{F}_2^n$ such that

$$A(w \otimes w) = b\,. \tag{1}$$

We can state the PCP theorem in terms of randomized query-efficient verifiers for this problem. Here, $V^\pi(x)$ denotes the output of a randomized algorithm V on input x with query access to a string π. (At any point during the computation of V, it can access an element of the string π. The elements of π are addressed by natural numbers encoded in binary.)

PCP theorem (verifier view). There exists a randomized poly-time algorithm V such that

- **YES:** if $x \in$ QuadEq, then

$$\exists \pi.\ \mathbb{P}\Big\{V^\pi(x) = 1\Big\} = 1 \tag{2}$$

- **NO:** if $x \notin$ QuadEq, then

$$\forall \pi.\ \mathbb{P}\Big\{V^\pi(x) = 1\Big\} < 0.99 \tag{3}$$

- **Queries:** The computation $V^\pi(x)$ queries only $O(1)$ positions in π (independent of the length of x).
- **Randomness:** The computation $V^\pi(x)$ uses at most $O(\log|x|)$ random bits.

In these notes, we prove a relaxed version of this theorem, where the verifier is allowed to use $O(|x|)$ random bits. This amount of randomness allows the verifier to access a range of positions in π that is exponential in the length of the input instance $|x|$. We will refer to this relaxed version of the verifier as an "exponential-size PCP system". The exponential blowup in the length of the proofs π prevents these verifier from being directly useful for hardness of approximation reductions. However, it turns out that they play a key role as "local gadgets" in the construction for the final PCP theorem (with a logarithmic number of random bits).

2 PCP systems and error-correcting codes

Error-correcting codes turn out to play a key role for our exponential-size PCP system. In this section, we explain why they are useful here.

Let (A, b) with $A \in \mathbb{F}_2^{m \times n^2}$ and $b \in \mathbb{F}_2^m$ be an instance of QuadEq. We are to decide if there exists an assignment $w \in \mathbb{F}_2^n$ such that

$$A(w \otimes w) = b\,. \tag{4}$$

In our exponential-size PCP system, every assignment $w \in \mathbb{F}_2^n$ corresponds to some proof $\pi = \pi(w)$. (This property is common for NP-hardness reductions.) We would like the verifier V to have the following properties:

- **YES:** if $w \in \mathbb{F}_2^n$ satisfies the system $x = (A, b)$, then $\mathbb{P}\{V^\pi(x) = 1\} = 1$ for $\pi = \pi(w)$.
- **NO:** if $w' \in \mathbb{F}_2^n$ does not satisfies the system $x = (A, b)$, then $\mathbb{P}\{V^{\pi'}(x) = 1\} < 0.99$ for $\pi' = \pi(w')$.

Since V makes only a constant number of queries to the purported proof, its acceptance probability stays approximately the same if the proof is perturbed at a small fraction of the positions.[1] Therefore, we would like that satisfying assignments get mapped to proofs that far in Hamming distance from proofs that non-satsifying assignments get mapped to. In fact, our PCP system satisfies the following stronger property:

> For our exponential-size PCP system, the map $w \mapsto \pi(w)$ is an error correcting code in the sense that any two different assignemnts $w \neq w'$ get mapped to proofs $\pi(w)$ and $\pi(w')$ that differ in a constant fraction of positions.

In fact, our PCP system is based on the Walsh-Hadamard code and it exploits several remarkable properties of this error-correcting code (see the sections on local testing and correcting below).

3 Walsh-Hadamard code

The *Walsh-Hadmard encoding* of $x \in \mathbb{F}_2^n$ is a bit string $\mathrm{WH}[x] \in \mathbb{F}_2^{\mathbb{F}_2^n}$ of length 2^n such that for all $r \in \mathbb{F}_2^n$

$$\mathrm{WH}[x](r) = r^T x\,. \tag{5}$$

In words, $\mathrm{WH}[u]$ consists of all parity checks of x.

We define the Hamming distance between bit strings.

> **Definition:** The (relative) Hamming distance of two strings $f, g \in \mathbb{F}_2^{\mathbb{F}_2^n}$ is
>
> $$\mathrm{dist}(f, g) \stackrel{\mathrm{def}}{=} \mathop{\mathbb{P}}_{r \in \mathbb{F}_2^n}\Big\{ f(r) \neq g(r) \Big\}\,. \tag{6}$$

The minimum distance of the Walsh-Hadmard code is 1/2.

Theorem: Every pair $x \neq y$ of bit strings satisfies

$$\mathrm{dist}\Big(\mathrm{WH}[x], \mathrm{WH}[y]\Big) = 1/2\,. \tag{7}$$

4 Exponential-size PCP system I

We define a *correct satisfiability proof* for a QuadEq instance $x = (A, b)$ to be a string π of the form

$$\pi = \mathrm{WH}[w]\mathrm{WH}[u] \text{ such that } u = w \otimes w \text{ and } Au = b\,. \tag{8}$$

Theorem: There exists a polynomial-time verifier V_0 that uses 4 queries and $O(N)$ random bits of inputs of length N such that for every QuadEq instance $x = (A, b)$ with n variables, and every string $\pi = \mathrm{WH}[w]\mathrm{WH}[u]$ for some $w \in \mathbb{F}_2^n$ and $u \in \mathbb{F}_2^{n^2}$, either

- **YES:** $\pi = \mathrm{WH}[w]\mathrm{WH}[u]$ is a correct satisfiability proof for x and

$$\mathbb{P}\{V_0^\pi(x) = 1\} = 1\,. \tag{9}$$

- **NO:** $\pi = \mathrm{WH}[w]\mathrm{WH}[u]$ is not a correct satisfiability proof for x and

$$\mathbb{P}\{V_0^\pi(x) = 1\} \leq 3/4\,. \tag{10}$$

The verifier V_0 consists of two parts: The first part is to check that $Au = b$ and the second part is to check that $u = w \otimes w$. For the first part, we use the observation that $Au \neq b$ implies that $r^T Au \neq r^T b$ for half of the choices $r \in \mathbb{F}_2^n$. Since $\mathrm{WH}[u](A^T r) = r^T Au$, the verifier can check that condition by querying one bit in the proof if it has the form $\pi = \mathrm{WH}[w]\mathrm{WH}[u]$. For the second part, we use a similar observation.

4.1 Verifier V_0

input: QuadEq instance $x = (A, b)$ with $A \in \mathbb{F}_2^{m \times n^2}$ and $b \in \mathbb{F}_2^m$
proof: $\pi = fg \in \mathbb{F}_2^{2^n + 2^{n^2}}$ with $f \in \mathbb{F}_2^n \to \mathbb{F}_2$ and $g \in \mathbb{F}_2^{n^2} \to \mathbb{F}_2$
operation:

- pick $r \in \mathbb{F}_2^m$ at random
- check $g(A^T r) = r^T b$
- pick $r, s \in \mathbb{F}_2^n$ at random
- check $g(r \otimes s) = f(r)g(s)$

Proof. Let $u \in \mathbb{F}_2^{n^2}$ and $w \in \mathbb{F}_2^n$ be such that $f = \mathrm{WH}[w]$ and $g = \mathrm{WH}[u]$. Suppose $A(w \otimes w) \neq b$. Then, either $u \neq w \otimes w$ or $Au \neq b$. We are to show that in either case the verifier accepts with probability at most 3/4.

In case $Au \neq b$, the first test fails with probability 1/2: Since $g(A^T r) = (A^T r)^T u = r^T A u$,

$$\mathop{\mathbb{P}}_{r}\{g(A^T r) \neq r^T b\} = \mathrm{dist}(\mathrm{WH}[Au], \mathrm{WH}[b]) = 1/2\,. \tag{11}$$

In case $u \neq w \otimes w$, the second test fails with probability $\geq 1/4$: Let Q be the vector $u - w \otimes w$ arranged as an $n \times n$ matrix, so that $Q_{i,j} = u_{ij} - w_i w_j$. Since $f(r)f(s) - g(r \otimes s) = (r \otimes s)^T (u - w \otimes w) = r^T Q s$,

$$\begin{aligned}
\mathop{\mathbb{P}}_{r,s\in\mathbb{F}_2^n}\Big\{g(r \otimes s) = f(r)g(s)\Big\} &= \mathop{\mathbb{P}}_{r,s\in\mathbb{F}_2^n}\{r^T Q s \neq 0\} \\
&= \mathop{\mathbb{P}}_{s\in\mathbb{F}_2^n}\{Qs \neq 0\} \cdot \mathop{\mathbb{P}}_{r\in\mathbb{F}_2^n}\{r^T Q s \neq 0 \mid Qs \neq 0\} \\
&= 1/2 \cdot \mathop{\mathbb{P}}_{s\in\mathbb{F}_2^n}\{Qs \neq 0\} \\
&\geq 1/2\,.
\end{aligned} \tag{12}$$

The last step uses that every $Q \neq 0$ satisfies $\mathbb{P}_{s\in\mathbb{F}_2^n}\{Qs \neq 0\} \geq 1/2$. x

We conclude that in both cases, the acceptance probability of V_0 is at most 3/4. □

It remains to compose V_0 with other verifiers such that resulting verifier accepts with high probability only if π is close in Hamming distance to some correct satisfiability proof. The key component is a way of testing whether testing whether a string is close to a Walsh-Hadamard encoding by querying only three positions of the string.

5 Local testing for Walsh-Hadamard

The following theorem shows that by querying three positions of a string it is possible to check whether it is close to a Walsh-Hadamard encoding.

Theorem (linearity test): Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a function. Suppose that for $\varepsilon \geq 1/2$,

$$\mathop{\mathbb{P}}_{r,s\in\mathbb{F}_2^n}\Big\{f(r+s) = f(r) + f(s)\Big\} \geq 1 - \varepsilon\,. \tag{13}$$

Then, $\mathrm{dist}(f, \mathrm{WH}[x]) \leq \varepsilon$ for some $x \in \mathbb{F}_2^n$.

The proof of this theorem uses Fourier analysis.

6 Interlude: Fourier analysis

The linearity test theorem is about $\mathbb{F}_2$-valued functions. It turns out that it is useful to consider the corresponding real-valued function.

Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a function. Consider the corresponding real-valued function $F\colon \mathbb{F}_2^n \to \mathbb{R}$, defined as

$$F(r) = (-1)^{f(r)}\,. \tag{14}$$

Suppose that the linearity test accepts f with probability $1-\varepsilon$, so that

$$\mathop{\mathbb{P}}_{r,s\in\mathbb{F}_2^n}\Big\{f(r+s)=f(r)+f(s)\Big\}=1-\varepsilon\,. \tag{15}$$

The acceptance probability of the linearity test corresponds to the value of the following cubic form in F,

$$F\mapsto \mathop{\mathbb{E}}_{r,s\in\mathbb{F}_2^n} F(r)F(r+s)F(s)\,. \tag{16}$$

Indeed,

$$\mathop{\mathbb{E}}_{r,s\in\mathbb{F}_2^n} F(r)F(r+s)F(s)=\mathop{\mathbb{E}}_{r,s\in\mathbb{F}_2^n}(-1)^{f(r)+f(s)+f(r+s)}=\mathbb{P}\{\text{accept}\}-\mathbb{P}\{\text{reject}\}=1-2\epsilon\,. \tag{17}$$

Next we introduce a particular basis for real-valued functions on $\mathbb{F}_2^n$. This basis turns out to diaognalize the cubic form (16). Let $\chi_x\colon \mathbb{F}_2^n\to\mathbb{F}_2^n$ be defined as

$$\chi_x(r)=(-1)^{WH[x](r)}=(-1)^{r^Tx}\,. \tag{18}$$

The set of functions $\{\chi_x \mid x\in\mathbb{F}_2^n\}$ is called the *Fourier basis*. The functions $\{\chi_x\}$ are also called *characters*.

Lemma (properties of characters):

1. every character χ_x is a group homomorphism from $(\mathbb{F}_2^n,+)$ to $(\mathbb{R},\cdot)$,

$$\forall r,s\in\mathbb{F}_2^n.\ \chi_x(r+s)=\chi_x(r)\chi_x(s)\,. \tag{19}$$

2. characters are pairwise orthogonal (w.r.t. to the inner product $\langle f,g\rangle = \mathbb{E}_{r\in\mathbb{F}_2^n} f(r)g(r)$),

$$\forall x,y\in\mathbb{F}_2^n.\ \mathop{\mathbb{E}}_r \chi_x(r)\chi_y(r)=\begin{cases}1 & x=y\\ 0 & x\neq y\end{cases} \tag{20}$$

3. characters form a basis of the real-valued functions on $\mathbb{F}_2^n$.

Proof. The first property is by definition,

$$\chi_x(r+s)=(-1)^{(r+s)^Tx}=(-1)^{r^Tx}(-1)^{r^Tx}=\chi_x(r)\chi_x(s)\,. \tag{21}$$

The second property is equivalent to the fact that different Walsh-Hadamard encodings have distance 1/2,

$$\mathop{\mathbb{E}}_r \chi_x(r)\chi_y(r)=\mathop{\mathbb{E}}_r(-1)^{r^T(x+y)}=\begin{cases}1 & x=y\\ 0 & x\neq y\end{cases} \tag{22}$$

The third property follows because $\operatorname{span}\{\chi_x\}\subseteq\mathbb{R}^{\mathbb{F}_2^n}$ and $\dim\operatorname{span}\chi_x=\dim\mathbb{R}^{\mathbb{F}_2^n}$ imply $\operatorname{span}\{\chi_x\}=\mathbb{R}^{\mathbb{F}_2^n}$. □

Since the characters form a basis for real-valued functions on $\mathbb{F}_2^n$, we can decompose F as a linear combination of characters,

$$F = \sum_x c_x \chi_x \,. \tag{23}$$

The numbers $\{c_x \mid x \in \mathbb{F}_2^n\}$ are called the *Fourier coefficients* of F.

Lemma (properties of $\{c_x\}$).

1. Normalization: $\sum_{x \in \mathbb{F}_2^n} c_x^2 = 1$.
2. Linearity test diagonalization: $\sum_x c_x^3 = \mathbb{E}_{r,s \in \mathbb{F}_2^n} F(r)F(r+s)F(s)$.
3. Hamming distance to Walsh-Hadamard code: $c_x = 1 - 2 \cdot \mathrm{dist}(f, WH[x])$.

Proof. The first property holds because

$$1 = \mathop{\mathbb{E}}_r F(r)^2 = \sum_{x,y} c_x c_y \mathop{\mathbb{E}}_r \chi_x(r)\chi_y(r) = \sum_x c_x^2 \,. \tag{24}$$

The second property holds because

$$1 - 2\epsilon = \mathop{\mathbb{E}}_{r,s} F(r)F(r+s)F(s) = \sum_{x,y,z} c_x c_y c_z \mathop{\mathbb{E}}_{r,s} \chi_x(r)\chi_y(r+s)\chi_z(s) = \sum_x c_x^3 \,. \tag{25}$$

The proof of the third property is an exercise. □

7 Proof of linearity test theorem

Suppose that the linearity test accepts the function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ with probability $1 - \varepsilon$. Let $\{c_x \mid x \in \mathbb{F}_2^n\}$ be the Fourier coefficients of the function $(-1)^f$. Since the Fourier basis diagonalizes the cubic form that corresponds to the linearity test,

$$1 - 2\varepsilon = \sum_{x \in \mathbb{F}_2^n} c_x^3 \le \max_{x \in \mathbb{F}_2^n} c_x \cdot \sum_{x \in \mathbb{F}_2^n} c_x^2 = \max_{x \in \mathbb{F}_2^n} c_x \,. \tag{26}$$

The last step uses the normalization property of the Fourier coefficients. By the relationship between Fourier coefficients and Hamming distance to Walsh-Hadmard encodings, it follows that

$$\min_{x \in \mathbb{F}_2^n} \mathrm{dist}(f, \mathrm{WH}[x]) = 1/2 - \max_{x \in \mathbb{F}_2^n} c_x/2 \le \varepsilon \,. \quad \square \tag{27}$$

8 Local correcting for Walsh-Hadamard

The verifier V_0 (see above) assumed that the supplied proof is a Hadamard encoding. The linearity test (see above) allows us to say that without loss of generality the supplied proof is close to a Hadamard encoding. Unfortunately, it's not clear that V_0 correctly works on

proofs that are close to Hadamard encodings because some positions of the supplied proof are much more likely to be queried by V_0 than other positions. (Exercise.)

The last ingredient of the exponential-size PCP system is a randomized procedure that is able to reliably reconstruct any particular position of a Walsh-Hadamard encoding given query-access to a string close to it.

Lemma (local correcting for Walsh-Hadamard).
Let $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ and $x \in \mathbb{F}_2^n$. Suppose $\mathrm{dist}(f, \mathrm{WH}[x]) \leq \varepsilon$. Then for every $r \in \mathbb{F}_2^n$,

$$\mathop{\mathbb{P}}_{s\in\mathbb{F}_2^n}\Big\{\mathrm{WH}[x](r) = f(r+s) + f(s)\Big\} \geq 1 - 2\varepsilon\,. \tag{28}$$

In particular, there exists a randomized algorithm Q that given input r and query access to f outputs $\mathrm{WH}[x](r)$ with probability $1 - 2\varepsilon$ by making only 2 queries to f,

$$\mathbb{P}\Big\{Q^f(r) = \mathrm{WH}[x](r)\Big\} \geq 1 - 2\varepsilon\,. \tag{29}$$

9 Exponential-size PCP system II

We prove now the following theorem which gives an exponential-size PCP system. In fact we will prove a stronger property of the verfier:

Theorem (Exponential-size PCP of proximity). There exists a randomized poly-time algorithm V such that for every instance x of QuadEq,

- **YES:** If π is a correct satisfiability proof for x (in the sense of (8)), then

$$\mathbb{P}\Big\{V^\pi(x) = 1\Big\} = 1\,, \tag{30}$$

- **NO:** If π is not 0.01-close to some correct satisfiability proof for x (in the sense of (8)), then

$$\mathbb{P}\Big\{V^\pi(x) = 1\Big\} < 0.99\,. \tag{31}$$

- **Queries:** The computation $V^\pi(x)$ queries only $O(1)$ positions in π (independent of the length of x).
- **Randomness:** The computation $V^\pi(x)$ uses at most $O(|x|)$ random bits.

Here is the final construction of the verifier V.

9.1 Verifier V

input: QuadEq instance $x = (A, b)$ with $A \in \mathbb{F}_2^{m \times n^2}$ and $b \in \mathbb{F}_2^m$
proof: $\pi = fg \in \mathbb{F}_2^{2^n + 2^{n^2}}$ with $f \in \mathbb{F}_2^n \to \mathbb{F}_2$ and $g \in \mathbb{F}_2^{n^2} \to \mathbb{F}_2$
operation:

1. Run linearity tests on both f and g in order to check that both f and g are close to Hadmard encodings of strings u and w (total of 6 queries).
2. Run the verifier V_0 in order to check whether u and w are satisfying assignments for the instance x. Here, V_0 accesses the supplied proof π indirectly through the local correcting algorithm Q. (Each of the four queries of V_0 gets translated by Q to two queries in π, which makes a total of 8 queries.)

Proof. Suppose V given x and π accepts with probability at least 0.99. We are to show that π is 0.01-close to a correct satisfiability proof for x.

Let $\varepsilon = 0.01$. If V accepts with probability at least $1 - \varepsilon$, then in particular the linearity tests accept with probability at least $1 - \varepsilon$. Therefore, both f and g are ε-close to Walsh-Hadamard encodings WH[w] and WH[u] for some $w \in \mathbb{F}_2^n$ and $u \in \mathbb{F}_2^{n^2}$. We can lower bound the acceptance probability of V_0 on the proof $\pi' = \mathrm{WH}[w]\mathrm{WH}[u]$,

$$\mathbb{P}\left\{V_0^{\pi'}(x)\right\} \geq \mathbb{P}\left\{V_0^{Q^f Q^g}(x)\right\} - 4 \cdot 2\varepsilon \geq 1 - 9\varepsilon > 3/4\,. \tag{32}$$

Here, $V_0^{Q^f Q^g}(x)$ denotes the output of V_0 on input x when accessing the proof $\pi = fg$ indirectly through the local correcting algorithm Q. The first step uses that Q answers all queries of V_0 correctly with probability at least $1 - 4 \cdot 2\varepsilon$ because each of the four queries of V_0 fails with probability at most 2ε.

By the **NO** case property of V_0, the lower bound (32) on the acceptance probability of V_0 implies that π' is a correct satisfiability proof, which means that x is satisfiable and that π is ε-close to a correct satisfiability proof. □

Footnotes

1. See the second exercise of homework 2 for a formalization of this statement.