# PCP theorem: hardness of quadratic equations and verifier view

*CS 6810—David Steurer*

*Spring 2016*

## 1 Hardness of approximation for systems of quadratic equations

**Problem (QuadEq).** Given a matrix $A \in \{0,1\}^{m \times n^2}$ and a vector $b \in \{0,1\}^m$, find an assignment $w \in \{0,1\}^n$ such that

$$A(w \otimes w) = b \qquad \mod 2\,. \tag{1}$$

Here, $\otimes$ is the Kronecker product for matrices and vectors. This operator is often pronounced "tensor". In our setting, $w \otimes w$ is a $n^2$-dimensional vector indexed by pairs $(i,j)$ with $i,j \in [n]$ such that

$$(w \otimes w)_{ij} = w_i \cdot w_j\,. \tag{2}$$

Therefore, (1) is equivalent to a list of $k$ homogeneous quadratic equations in the variables $w_1, \ldots, w_n$,

$$\forall k \in [m]. \quad \sum_{i,j \in [n]} A_{k,ij} w_i w_j = b_k \qquad \mod 2\,. \tag{3}$$

It is NP-complete to decide if for a given instance of QuadEq all equations can be satisfied.

*Can we efficiently find an assignment that satisfies many equations of a given satisfiable system?*

**Notation.** Let MaxQuadEq be the problem of finding an assignment that satisfies as many equations as possible. Let $\mathrm{opt}(A,b)$ denote the maximum fraction of satisfied equations over all assignments $w$.

The following theorem shows that it is NP-hard to achieve an approximation ratio of 0.51 for MaxQuadEq.

**Theorem.** There exists a (randomized) polynomial-time function $f$ that maps every MaxQuadEq instance $(A,b)$ to a MaxQuadEq instance $(A',b')$ such that

- **YES:** If $\mathrm{opt}(A,b) = 1$, then $\mathrm{opt}(A',b') = 1$.
- **NO:** If $\mathrm{opt}(A,b) < 1$, then $\mathrm{opt}(A',b') < 0.51$ with probability 0.99 over the randomness of the function $f$.

**Aside:** A random assignment achieves approximation ratio 1/4. It is NP-hard to achieve a strictly larger approximation ratio. For satisfiable instances, approximation ratio 3/8 is possible and it is NP-hard to achieve a strictly larger approximation ratio.

## 1.1 Proof of theorem

Let $(A, b)$ be an instance of QuadEq with $n$ variables and $m$ equations. Let $R \in \{0,1\}^{d \times m}$ be a random matrix for $d$ to be determined later. (We can choose $d = 1000n$.)

We define the randomized function $f$ by choosing $A' = RA$ and $b' = Rb$. In other words, the resulting system of quadratic equations $RA(w \otimes w) = Rb$ consists of $d$ random linear combinations of the original system $A(w \otimes w) = b$.

*YES case:* If $\mathrm{opt}(A, b) = 1$, then there exists an assignment $w$ that satisfies the quadratic system $A(w \otimes w) = b$. The same assignment also satisfies $RA(w \otimes w) = b$. Therefore, $\mathrm{opt}(A', b') = 1$.

*NO case:* Suppose $\mathrm{opt}(A, b) < 1$. We will show the following claim.

**Claim:** For every assignment $w \in \{0,1\}^n$,

$$\mathop{\mathbb{P}}_{R}\Big\{w \text{ satisfies at least } 0.51d \text{ equations of } (A', b')\Big\} < 0.01 \cdot 2^{-n}\,. \tag{4}$$

This claim implies the NO case of the theorem by the union bound.

$$\begin{aligned}\mathop{\mathbb{P}}_{R}\Big\{\mathrm{opt}(A', b') \geq 0.51\Big\} &\leq \sum_{w \in \{0,1\}^n} \mathop{\mathbb{P}}_{R}\Big\{w \text{ satisfies at least } 0.51d \text{ equations of } (A', b')\Big\} \\ &< 2^n \cdot 0.01 \cdot 2^{-n} \quad \text{by Claim} \\ &\leq 0.01\,.\end{aligned} \tag{5}$$

It remains to prove the claim.

*Proof of claim:* Since the system $(A, b)$ is not satisfiable, the vector $y = A(w \otimes w) \mod 2$ is not the 0 vector. Therefore, $Ry \mod 2$ is a uniformly random vector in $\{0,1\}^d$. (**Exercise.**) Thus, the random variables $X_1, \ldots, X_d$ with $X_i = 1 - (Ry)_i$ are independent Bernoulli variables with probability 1/2 of being equal to 1. Note that $X_i = 1$ if assignment $w$ satisfies the $i$-th constraint of the system $(A', b')$ and $X_i = 0$ otherwise. Therefore, $\sum_{i=1}^{d} X_i$ is the number of constraints of $(A', b')$ satisfied by $w$. By the Chernoff bound,

$$\mathbb{P}\Big\{\sum_{i=1}^{d} X_i \geq (1+\varepsilon)d/2\Big\} \leq e^{-\varepsilon^2 d/6}\,. \tag{6}$$

We choose $\varepsilon = 2/100$ and $d = 12n/\varepsilon^2$. Then for $n \geq 6$,

$$\mathbb{P}\Big\{\sum_{i=1}^{d} X_i \geq 0.51d\Big\} \leq e^{-\varepsilon^2 d/6} = e^{-2n} \leq 2^{-n} \cdot 2^{-n} \leq 0.01 \cdot 2^{-n}\,. \tag{7}$$

This bound proves the claim because by our choice of $X_1, \ldots, X_d$, the event $\sum_{i=1}^{d} X_i \geq 0.51d$ is the same as the event that $w$ satisfies at least $0.51d$ equations of $(A', b')$. □

## 2 Verifier view of the PCP theorem

A useful definition of NP is in terms of polynomial time verifiers.

Let $L \subseteq \{0,1\}^*$ be a language. Then, $L \in$ NP if and only if there exists a polynomial-time algorithm $V$ and a polynomial $p$ such that

$$L = \left\{ x \mid \exists \pi.\ |\pi| \leq p(|x|) \wedge V(x,\pi) = 1 \right\} \tag{8}$$

In words, NP consists of all decision problems such that every YES instance of the problem has a proof for being a YES instance that can be checked in time polynomial in the length of the instance.

Recall the statement of the PCP theorem.

**PCP theorem:** There exists a polynomial-time function $f$ that maps every 3Sat instance $x$ to a Max3Sat instance $\varphi$ with the following properties:

- **YES:** if $x$ is satisfiable then $opt(\varphi) = 1$
- **NO:** if $x$ is not satisfiable then $opt(\varphi) < 0.99$

Using the function $f$ from the statement of this theorem we can construct a randomized verifier for 3Sat.

**Randomized verifier $V_{PCP}$ for 3Sat:**
Given: 3Sat instance $x$ and a purported proof $\pi$ of satisfiability

- compute Max3Sat instance $\varphi = f(x)$,
- choose clauses $C_1, \ldots, C_t$ uniformly at random from $\varphi$ for $t = 1000$,
- check that $\pi$ satisfies clauses $C_1, \ldots, C_t$.[1]

This randomized verifier has the following remarkable properties:

**Properties of $V_{PCP}$:**

- **YES:** if $x$ is satisfies, then there exists $\pi$ such that $V_{PCP}(x,\pi) = 1$ with probability 1.
- **NO:** if $x$ is not satisfiable, then every $\pi$ satisfies $V_{PCP}(x,\pi) = 1$ with probability at most 0.001.
- **Efficiency:** $V_{PCP}$ runs in polynomial time and in addition satisfies:
- **Queries:** $V_{PCP}$ on inputs $x$ and $\pi$ queries at most $O(1)$ positions of $\pi$. (The number of queries is independent of the length of the inputs.)
- **Randomness:** $V_{PCP}$ on inputs $x$ and $\pi$ uses at most $O(1) \cdot \log|x|$ random bits.

The YES and NO case properties of the verifier $V_{PCP}$ are similar to the properties of a polynomial-time verifier for 3Sat in the sense of (8). The difference is that in the NO case $V_{PCP}$ is allowed to answer incorrectly with low probability over the internal randomness of $V_{PCP}$.

The key property $V_{PCP}$ is about query efficiency. This randomized can verify the correctness of a purported satisfiability proof $\pi$ by examining only a constant number of bits in the proof.

It turns out that any randomized verifier with the properties above can be used to construct a function $f$ as described in the statement of the PCP theorem above.

*Informal statement of the PCP theorem:* There exists a randomized polynomial-time algorithm that can verify the correctness of a purported satisfiability proof by examining only a constant number of bits of it.

## Footnotes

1. Here, we view $\pi$ as an assignment for the variables of $\varphi$. Therefore, in order to check whether $\pi$ satisfies the clauses $C_1, \ldots, C_t$ we only need to query the assigned values for the $3t$ variables that appears in the clauses $C_1, \ldots, C_t$.