The Projection Games Conjecture and The NP-Hardness of $\ln n$-Approximating Set-Cover

Dana Moshkovitz *

June 11, 2012

Abstract

We suggest the research agenda of establishing new hardness of approximation results based on the "projection games conjecture", i.e., an instantiation of the Sliding Scale Conjecture of Bellare, Goldwasser, Lund and Russell to projection games.

We pursue this line of research by establishing a tight $\mathcal{NP}$-hardness result for the Set-Cover problem. Specifically, we show that under the projection games conjecture (in fact, under a quantitative version of the conjecture that is only slightly beyond the reach of current techniques), it is $\mathcal{NP}$-hard to approximate Set-Cover on instances of size N to within $(1-\alpha)\ln N$ for arbitrarily small $\alpha > 0$. Our reduction establishes a tight trade-off between the approximation accuracy α and the time required for the approximation $2^{N^{\Omega(\alpha)}}$, assuming Sat requires exponential time.

The reduction is obtained by modifying Feige's reduction. The latter only provides a lower bound of $2^{N^{\Omega(\alpha/\log\log N)}}$ on the time required for $(1-\alpha)\ln N$-approximating Set-Cover assuming Sat requires exponential time (note that $N^{1/\log\log N} = N^{o(1)}$). The modification uses a combinatorial construction of a bipartite graph in which any coloring of the first side that does not use a color for more than a small fraction of the vertices, makes most vertices on the other side have their neighbors all colored in different colors.

1 Introduction

1.1 Projection Games and The Projection Games Conjecture

Most of the $\mathcal{NP}$-hardness of approximation results known today (e.g., all of the results in Håstad's paper [Hås01]) are based on a PCP Theorem for *projection games* (also known as Label-Cover) [AS98, ALM+98, Raz98, MR10]. The input to a projection game consists of: (i) a bipartite graph $G = (A, B, E)$; (ii) finite alphabets Σ_A, Σ_B; (iii) constraints (also called *projections*) $\pi_e : \Sigma_A \to \Sigma_B$ for every edge $e \in E$. The goal is to find assignments to the vertices $\varphi_A : A \to \Sigma_A$, $\varphi_B : B \to \Sigma_B$ that *satisfy* as many of the edges as possible. We say that an edge $e = (a, b) \in E$ is satisfied, if the projection constraint holds, i.e., $\pi_e(\varphi_A(a)) = \varphi_B(b)$. We denote the size of a projection game by $n = |A| + |B| + |E|$. A PCP Theorem for projection games with soundness error ε and alphabet size k (where ε and k may depend on n) states the following:

> *Given a projection game of size n with alphabets of size k, it is $\mathcal{NP}$-hard to distinguish between the case where all edges can be satisfied and the case where at most ε fraction of the edges can be satisfied.*

*dmoshkov@csail.mit.edu. Department of Electrical Engineering and Computer Science, MIT.

We can refine this statement by saying that there is a reduction from (exact) SAT to projection games, and the reduction maps instances of SAT of size n to projection games of size $N = n^{1+o(1)} poly(1/\varepsilon)$. Such PCPs are referred to as "almost-linear size PCP" because of the exponent of n, although for small ε the blow-up may be super-linear.

The state of the art today for PCP Theorems for projection games is the following:

Theorem 1 ([MR10]). *There exists $c > 0$, such that for every $\varepsilon \geq 1/N^c$,* SAT *on input of size n can be reduced to a projection game of size $N = n^{1+o(1)} poly(1/\varepsilon)$ over alphabet of size $exp(1/\varepsilon)$ that has soundness error ε. The reduction is computed in polynomial time in N.*

Note that one cannot hope for ε that is lower than $1/N$ (polynomially small). The $exp(1/\varepsilon)$ in the statement is not tight. It can be shown that $|\Sigma| \geq 1/\varepsilon$, and we conjecture that an alphabet size of $poly(1/\varepsilon)$ could be achieved:

Conjecture 1 (Projection games conjecture, PGC). *There exists $c > 0$, such that for every $\varepsilon \geq 1/N^c$,* SAT *on input of size n can be efficiently reduced to a projection game of size $N = n^{1+o(1)} poly(1/\varepsilon)$ over alphabet of size $poly(1/\varepsilon)$ that has soundness error ε.*

In almost all applications, one wishes the alphabet size to be at most polynomial in n. This happens in Theorem 1 only when $\varepsilon \geq 1/(\log N)^b$ for some constant $b > 0$. The PGC, on the other hand, gives polynomial alphabet for any $\varepsilon \geq 1/N^c$.

The projection games conjecture is in fact the Sliding Scale Conjecture of Bellare, Goldwasser, Lund and Russell [BGLR93] instantiated for projection games. By "sliding scale" we refer to the idea that the error can be decreased as we increase the alphabet size. Bellare et al. conjectured that polynomially small error could be achieved simultaneously with polynomial alphabet, even for two queries. They did not formulate their conjecture for projection games – the importance of projection games was not fully recognized when they published their work in 1993.

1.2 Previous Work

Approximation algorithms for projection games were researched, and the conjecture is consistent with the state of the art algorithm, giving $1/\varepsilon = O(\sqrt[3]{Nk})$ [CHK09] (Note that the formulation in [CHK09] is slightly different than ours – they have a vertex per pair (vertex, assignment) in our formulation).

The existing hardness results for projection games include two results: the one mentioned in Theorem 1 and another result that is based on parallel repetition [Raz98]:

Theorem 2 ([Raz98]). *There exists $c > 0$, such that for every $\varepsilon \geq 1/N^{c/\log n}$,* SAT *on input of size n can be efficiently reduced to a projection game of size N over alphabet of size $O(1/\varepsilon)$ that has soundness error ε.*

Note that when the reduction is polynomial, i.e., $N = n^{O(1)}$, the soundness error is constant. Better soundness error ε can be obtained for larger N. For instance, for $N = n^{O(\log n)}$, one obtains $\varepsilon = 1/n$. Unfortunately, polynomially small error $1/N^c$ cannot be obtained from Theorem 2.

For PCPs with more than two queries, soundness error approaching polynomial, $\varepsilon = 2^{-(\log N)^{1-\epsilon}}$ for every $\epsilon > 0$, is known [DFK+11]. Alas, these PCPs are not projection games, and the number of queries depends on $1/\epsilon$.

The projection games conjecture has a similar flavor to the unique games conjecture (UGC) of Khot [Kho02]: both assert that low soundness error[1] for a special kind of 2-prover games can be obtained for sufficiently large alphabets. Unique games are the special case of projection games in which the projections π_e are 1-1. Unique games appear to be much easier than general projection games. In particular, while there are constructions of projection games with low soundness error for Sat, we do not know of any constructions of unique games with almost-perfect completeness[2] and bounded soundness error. The two conjectures, UGC and PGC, seem unrelated: neither would imply the other.

1.3 The Potential Influence of The PGC

We believe that the projection games conjecture provides a stable foundation on which many new hardness of approximation results can be based. In particular, for several central approximation problems, achieving tight hardness results seems to require projection games with low soundness error; a few examples follow.

In a work in progress with Gopal we research the approximability of Max-3Sat and Max-3Lin just above their approximation thresholds, which are 7/8 and 1/2, respectively. For context, Håstad discusses hardness beyond any *constant* larger than the thresholds [Hås01], and Moshkovitz-Raz improve this to $1/(\log\log n)^{O(1)}$ beyond the threshold [MR10], which is still quite large for reasonable n's. Researching the range of $1/n^{O(1)}$ beyond the threshold is possible assuming projection games with polynomially small error.

Other results we hope could be achieved (but would require further ideas) are:

- Tight lower bound for $n^{1-o(1)}$-approximation of Clique [Hås99, Kho01].
- Tight lower bound for $n^{\Omega(1)}$-approximation of the Shortest-Vector-Problem (SVP) in lattices [Kho05].

In this paper, we show a tight lower bound on $(1-\alpha)\ln n$-approximation of Set-Cover assuming the projection games conjecture. (Of course, all the lower bounds are conditioned on a lower bound for Sat.)

There are several types of gains that can obtained from the PGC:

- **Better lower bounds.** For some problems (e.g., Set-Cover) the soundness error obtained from parallel repetition (Theorem 2) is sufficient, but the blow-up in the reduction translates into weak lower bounds. For Set-Cover, this lower bound is $2^{n^{\Omega(\alpha/\log\log n)}}$ which is much lower than the exponential lower bound one could a-priori hope for (note that $n^{1/\log\log n} = n^{o(1)}$).
- **Minimal assumptions.** The parallel-repetition based hardness result for Set-Cover can equivalently stated by saying: we can rule out polynomial-time $(1-\alpha)\ln n$-approximation for Set-Cover assuming $\mathcal{NP} \not\subseteq DTIME(n^{O(\log\log n)})$ [Fei98]. The PGC lets one see what results can potentially be obtained relying only on the minimal assumption $\mathcal{P} \neq \mathcal{NP}$.

[1] The unique games conjecture only asks for arbitrarily small constant soundness error ε, while the PGC asks for polynomially small error.

[2] For unique games, if all the edges can be satisfied simultaneously, then one can find a satisfying assignment in polynomial time. Hence, we consider the case where *almost* all edges can be satisfied simultaneously ("almost perfect completeness").

- **Improved inapproximability factors.** For many problems (such as the other problems mentioned above: SVP, Clique, etc), one seems to need polynomially small soundness error to obtain the best inapproximability factor.

In all the aforementioned examples, the existing reductions have super-polynomial blow-up, not only in order to achieve low error for a projection game, but also to facilitate the reduction. For instance, Håstad's reductions use the long code on top of a projection game. For low error ε, the long code incurs a large blow-up $2^{(1/\varepsilon)^{O(1)}}$ [Hås01]. Basing hardness results on the PGC, would require reductions that do not resort to large blow-ups.

1.4 Set-Cover

We demonstrate the application of the PGC to the $\mathcal{NP}$-hardness of approximating Set-Cover. In Set-Cover, given a collection of sets over the same base set, such that the sets cover all of the base set, the goal is to find as few sets as possible that cover the entire base set:

Definition 3 (Set-Cover)**.** *The input to* Set-Cover *consists of a base set U, $|U| = n$ and subsets $S_1, \dots S_m \subseteq U$, $\bigcup_{j=1}^{m} S_j = U$, $m \leq poly(n)$. The goal is to find as few sets $S_{i_1}, \dots, S_{i_k}$ as possible that cover U, i.e., $\bigcup_{j=1}^{k} S_{i_j} = U$.*

Set-Cover is a classic $\mathcal{NP}$-hard optimization problem. It is equivalent to the Hitting-Set, Hypergraph-Vertex-Cover and Dominating-Set problems, and is a special case of many other problems, e.g., Group-Steiner-Tree and Group-Traveling-Salesman-Problem.

The greedy algorithm was shown to give a $(\ln n + 1)$-approximation for Set-Cover [Chv79]. Slavík analyzed the low order terms of the greedy algorithm, and showed that it in fact obtains an approximation to within $\ln n - \ln \ln n + O(1)$ [Sla96]. Set-Cover also has a linear programming based algorithm that gives approximation to within similar factors [Sri99].

Lund and Yannakakis proved that Set-Cover cannot be approximated in polynomial time to within any factor better than $(\log_2 n)/4$, assuming $NP \not\subseteq DTIME(n^{poly \log n})$ [LY93]. By adapting their construction, Feige changed the leading constant to the right constant, and showed that Set-Cover cannot be approximated in polynomial time to within $(1 - \alpha) \ln n$ for any $\alpha > 0$, assuming $\mathcal{NP} \not\subseteq DTIME(n^{O(\lg \lg n)})$ [Fei98] (the improvement in the assumption is due to the proof of the parallel repetition theorem [Raz98] in the time between the two results). Under $\mathcal{P} \neq \mathcal{NP}$, the best hardness factor known is about $0.2 \ln n$ [AMS06], based on the PCP of [RS97, AS03].

The assumption $\mathcal{NP} \not\subseteq DTIME(n^{O(\lg \lg n)})$ in Feige's work comes from the use of the parallel repetition theorem. Parallel repetition is used by Feige not only to ensure very low error $1/(\log n)^{O(1)}$, but also for its unique structure. It was assumed by some that the blow-up incurred by parallel repetition was inherent to the problem. We show that this is not the case, assuming the PGC. Moreover, the blow-up in our reduction is essentially optimal.

Theorem 4. *For every $0 < \alpha < 1$, there is $c = c(\alpha)$, such that if the projection games conjecture holds with error $\varepsilon = \frac{c}{\lg^4 n}$, then (exact)* Sat *on inputs of size n can be reduced in polynomial time to approximating* Set-Cover *on inputs of size $N = n^{O(1/\alpha)}$ better than $(1 - \alpha) \ln N$.*

The theorem proves that approximating Set-Cover on inputs of size N better than $(1 - \alpha) \ln N$ is NP-hard, assuming the PGC. Interestingly, the blow-up of the reduction $N = n^{O(1/\alpha)}$ is optimal (up to the constant in the $O(\cdot)$), assuming that Sat requires exponential time $2^{\Omega(n)}$

and the PGC. This follows from a sub-exponential $2^{O(n^{\alpha}\log n)}$-time approximation algorithm for $(1-\alpha)\ln N$ approximating Set-Cover [CKW09].

Another interesting point about the theorem is that the quantitative version of the PGC that we need, namely, $\varepsilon = \frac{c}{\lg^4 n}$ for sufficiently small constant $c > 0$, is much weaker than the full conjecture, and it is just outside the reach of current techniques.

1.5 Preliminaries

For a set S and a natural number ℓ we denote by $\binom{S}{\ell}$ the family of all sets of ℓ elements from S.

We assume without loss of generality that the projection game in Conjecture 1 is bi-regular, i.e., all the A vertices have the same degree, which we call the A-degree, and all the B vertices have the same degree, which we call the B-degree. We note that any projection game can be converted to bi-regular using a technique developed in [MR10] ("right degree reduction – switching sides – right degree reduction"), and the cost in the soundness error and graph size does not change the parameters as stated in Conjecture 1.

2 Set-Cover Hardness

2.1 The New Component

Feige uses the structure obtained from parallel repetition to achieve a projection game in which the soundness guarantee is that very few B vertices have any two of their neighbors agree on a value for them:

Definition 5 (Total disagreement)**.** *Assume a projection game $(G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$. Let $\varphi_A : A \to \Sigma_A$ be an assignment to the A vertices. We say that the A vertices* totally disagree *on a vertex $b \in B$ if there are no two neighbors $a_1, a_2 \in A$ of b, $e_1 = (a_1, b), e_2 = (a_2, b) \in E$, for which*

$$\pi_{e_1}(\varphi_A(a_1)) = \pi_{e_2}(\varphi_A(a_2)).$$

Definition 6 (Agreement soundness)**.** *Assume a projection game $(G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ for deciding whether a boolean formula ϕ is satisfiable. We say that G has* agreement soundness error *ε, if for unsatisfiable ϕ, for any assignment $\varphi_A : A \to \Sigma_A$, the A vertices are in total disagreement on at least $1-\varepsilon$ fraction of the $b \in B$.*

Feige used parallel repetition together with a coding theoretic "trick" to achieve agreement soundness. We show a different way to achieve agreement soundness. Our construction centers around the following combinatorial construction:

Lemma 2.1 (Combinatorial construction)**.** *For $0 < \varepsilon < 1$, for infinitely many n, D, there is an explicit construction of a regular graph $H = (U, V, E)$ with $|U| = n$, V-degree D, and $|V| \le n^{O(1)}$ that satisfies the following. For every partition $U_1, \ldots, U_l$ of U into sets, such that $|U_i| \le \varepsilon |U|$ for $i = 1, \ldots, l$, the fraction of vertices $v \in V$ with more than one neighbor in any single set U_i, is at most εD^2.*

Note that the combinatorial property could be achieved by a randomized construction, or by a construction that has a V vertex per every possible set of D neighbors in U. However, the first construction is randomized and the second – too wasteful with a size of $\approx |U|^D$. The lemma can therefore be thought of as a *derandomization* of the randomized/full constructions.

Proof. (of Lemma 2.1) Associate U with a space $\mathbb{F}^m$ where $\mathbb{F}$ is a finite field of size $|\mathbb{F}| = D$, and m is a natural number. Let V be the set of all lines in $\mathbb{F}^m$. Hence, $|V| = \binom{|U|}{2}/\binom{|\mathbb{F}|}{2}$. We connect a line $v \in V$ with a point $u \in U$ if u lies in v.

Let us show this construction satisfies the desired property. Fix a partition $U_1, \ldots, U_l$ of U into tiny sets, $|U_i| \leq \varepsilon |U|$ for $i = 1, \ldots, l$. For every $1 \leq i \leq l$, the number of V lines that have at least two neighbors in U_i is at most $\binom{|U_i|}{2}$. Thus the total number of V vertices with more than one neighbor in a single U_i is at most

$$\begin{aligned}
\sum_{i=1}^{l} \binom{|U_i|}{2} &\leq \sum_{i=1}^{l} \frac{|U_i|^2}{2} \\
&\leq \max\{|U_i| \mid 1 \leq i \leq l\} \cdot \sum_{i=1}^{l} \frac{|U_i|}{2} \\
&\leq \varepsilon |U| \cdot \frac{|U|}{2} \\
&\leq \varepsilon |\mathbb{F}|^2 |V|.
\end{aligned}$$

□

We show how to take a projection game with standard soundness and convert it to a projection game with total disagreement soundness, by combining it with the graph from Lemma 2.1.

Lemma 2.2. *Let $D \geq 2$, $\varepsilon > 0$. From a projection game with soundness error $\varepsilon^2 D^2$, we can construct a projection game with agreement soundness error $2\varepsilon D^2$ and B-degree D. The transformation preserves the alphabets of the game. The size is raised to a constant factor.*

Proof. Let $\mathcal{G} = (G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ be the original projection game. Assume that the B-degree is $|U|$, and we use U to enumerate the neighbors of a B vertex, i.e., there is a function $E^{\leftarrow} : B \times U \to A$ that given a vertex $b \in B$ and $u \in U$, gives us the A vertex which is the u neighbor of b.

Let $H = (U, V, E_H)$ be the graph from Lemma 2.1. We create a new projection game $(G = (A, B \times V, E'), \Sigma_A, \Sigma_B, \Phi')$. The intended assignment to every vertex $a \in A$ is the same as its assignment in the original game. The intended assignment to a vertex $\langle b, v\rangle \in B \times V$ is the same as the assignment to b in the original game. We put an edge $e' = (a, \langle b, v\rangle)$ if $E^{\leftarrow}(b, u) = a$ and $(u, v) \in E_H$. We define $\pi_{e'} \equiv \pi_{(a,b)}$.

If there is an assignment to the original game that satisfies c fraction of its edges, then the corresponding assignment to the new game satisfies c fraction of its edges.

Suppose there is an assignment for the new game $\varphi_A : A \to \Sigma_A$ in which more than $2\varepsilon D^2$ fraction of the vertices in $B \times V$ do not have total disagreement.

Let us say that $b \in B$ is "good" if for more than εD^2 of the vertices in $\{b\} \times V$ the A vertices do not totally disagree. Note that the fraction of good $b \in B$ is at least εD^2.

Focus on a good $b \in B$. Consider the partition of U into $|\Sigma_B|$ sets, where the set corresponding to $\sigma \in \Sigma_B$ is:

$$U_\sigma = \{u \in U \mid a = E^{\leftarrow}(b, u) \wedge e = (a, b) \wedge \pi_e(\varphi_A(a)) = \sigma\}.$$

By the property of H, there must be $\sigma \in \Sigma_A$ such that $|U_\sigma| > \varepsilon |U|$. We call σ the "champion" for b.

We define an assignment $\varphi_B : B \to \Sigma_B$ that assigns good b's their champions, and other b's arbitrary values. The fraction of edges that φ_A, φ_B satisfy in the original game is at least $\varepsilon^2 D^2$. □

Next we consider a variant of projection games that is relevant for the reduction to SET-COVER. In this variant the prover is allowed to assign each vertex ℓ values, and an agreement is interpreted as agreement on *one* of the assignments in the list:

Definition 7 (List agreement)**.** *Assume a projection game $(G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$. Let $\ell \geq 1$. Let $\hat{\varphi}_A : A \to \binom{\Sigma_A}{\ell}$ be an assignment that assigns each A vertex l alphabet symbols. We say that the A vertices* totally disagree *on a vertex $b \in B$ if there are no two neighbors $a_1, a_2 \in A$ of b, $e_1 = (a_1, b), e_2 = (a_2, b) \in E$, for which there exist $\sigma_1 \in \hat{\varphi}_A(a_1)$, $\sigma_2 \in \hat{\varphi}_A(a_2)$, such that*

$$\pi_{e_1}(\sigma_1) = \pi_{e_2}(\sigma_2).$$

Definition 8 (List agreement soundness)**.** *Assume a projection game $(G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ for deciding membership whether a boolean formula ϕ is satisfiable. We say that G has* agreement soundness error *(ℓ, ε), if for unsatisfiable ϕ, for any assignment $\hat{\varphi}_A : A \to \binom{\Sigma_A}{\ell}$, the A vertices are in total disagreement on at least $1 - \varepsilon$ fraction of the $b \in B$.*

If a projection game has low error ε, then even when the prover is allowed to assign each A vertex ℓ values, the game is still sound. This is argued in the next corollary:

Lemma 2.3 (Projection game with list agreement soundness)**.** *Let $\ell \geq 1$, $0 < \varepsilon' < 1$. A projection game with agreement soundness error ε' has agreement soundness error $(\ell, \varepsilon' \ell^2)$.*

Proof. Assume on way of contradiction that the projection game has an assignment $\hat{\varphi}_A : A \to \binom{\Sigma_A}{\ell}$ such that on more than $\varepsilon' \ell^2$ fraction of the B vertices, the A vertices do not totally disagree. Define an assignment $\varphi_A : A \to \Sigma_A$ by assigning every vertex $a \in A$ a symbol picked uniformly at random from the ℓ symbols in $\hat{\varphi}_A(a)$. If a vertex $b \in B$ has two neighbors $a_1, a_2 \in A$ that agree on b under the list assignment $\hat{\varphi}_A$, then the probability that they agree on b under the assignment φ_A is at least $1/\ell^2$. Thus, under φ_A, the expected fraction of the B vertices that have at least two neighbors that agree on them, is more than ε'. In particular, there exists an assignment to the A vertices, such that more than ε' fraction of the B vertices have two neighbors that agree on them. This contradicts the agreement soundness of the game. □

By applying Lemma 2.2 and then Lemma 2.3 on the game from Conjecture 1, we get:

Corollary 2.4. *Assuming Conjecture 1, for any $\ell \geq 1$, for infinitely many D, for any $\varepsilon \geq 1/n^c$, given a projection game with alphabet size poly$(1/\varepsilon)$ and B-degree D, it is $\mathcal{NP}$-hard to distinguish between the case where all edges can be satisfied, and the case where the agreement soundness error is $(\ell, 2D\ell^2\sqrt{\varepsilon})$.*

2.2 Following Feige's Reduction

In the remainder, we will show how to use Corollary 2.4 to obtain the desired hardness result for SET-COVER. The reduction is along the lines of Feige's original reduction.

For the reduction we rely on a combinatorial construction of a universe together with partitions of it. Each partition covers the universe, but any cover that takes at most one set out of each partition, is necessarily large:

Lemma 2.5 (Partition system, [NSS95])**.** *For natural numbers m, D, for $\alpha \leq 2/D$, there is an explicit construction of a universe U, $|U| \leq$ poly$(D^{\log D}, \log m)$ and partitions $\mathcal{P}_1, \ldots, \mathcal{P}_m$ of U into D sets that satisfy the following: there is no cover of U with $\ell = D \ln |U| (1 - \alpha)$ sets $S_{i_1}, \ldots, S_{i_\ell}$, $1 \leq i_1 < \cdots < i_\ell \leq m$, such that set S_{i_j} belongs to partition $\mathcal{P}_{i_j}$.*

To see why $\ell = D \ln |U| (1-\alpha)$ is to be expected (this later determines the hardness factor we get), think of the following randomized construction: each element in U corresponds to a vector in $[D]^m$, specifying for each of the m partitions, to which of its D sets it belongs. Consider a uniformly random choice of such a vector. Fix any $S_{i_1}, \ldots, S_{i_\ell}$. The probability that a random element is not covered by $S_{i_1}, \ldots, S_{i_\ell}$ is $(1-1/D)^\ell \approx e^{-\ell/D}$. When $\ell = D \ln |U| (1-\alpha)$, we have $e^{-\ell/D} \geq 1/|U|$, and we expect one of the $|U|$ elements in U not to be covered by $S_{i_1}, \ldots, S_{i_\ell}$. The construction in [NSS95] de-randomizes this randomized construction.

We now describe the reduction from a projection game $\mathcal{G}$ as in Corollary 2.4, to a Set-Cover instance $\mathcal{SC}_\mathcal{G}$.

Apply Lemma 2.5 for $m = |\Sigma_B|$ and D which is the B-degree of the projection game. Let U be the universe, and $\mathcal{P}_{\sigma_1}, \ldots, \mathcal{P}_{\sigma_m}$ be the partitions of U. We index the partitions by Σ_B symbols $\sigma_1, \ldots, \sigma_m$. The elements of the Set-Cover instance are $B \times U$.

For every vertex $a \in A$ and an assignment $\sigma \in \Sigma_A$ to a we have a set $S_{a,\sigma}$ in the Set-Cover instance. The intuition is that whether we take $S_{a,\sigma}$ to the cover would correspond to assigning σ to a. The set $S_{a,\sigma}$ is a union of subsets, one for every edge $e = (a, b)$ touching a. Suppose e is the i'th edge coming into b ($1 \leq i \leq D$), then the subset associated with e is the i'th subset of the partition $\mathcal{P}_{\varphi_e(\sigma)}$. Note that if we have an assignment to the A vertices, such that all of b's neighbors agree on one value for b, then the D subsets corresponding to those neighbors and their assignments form a partition that covers b's universe. On the other hand, if one uses only sets that correspond to totally disagreeing assignments to the neighbors, then by the definition of the partitions, covering U requires $\approx \ln |U|$ times more sets.

Claim 2.6. *The following hold:*

- *Completeness: If all the edges in $\mathcal{G}$ can be satisfied, then $\mathcal{SC}_\mathcal{G}$ has a set cover of size $|A|$.*
- *Soundness: Let $\ell \doteq D \ln |U| (1-\alpha)$ be as in Lemma 2.5. If $\mathcal{G}$ has agreement soundness (ℓ, α), then every set cover of $\mathcal{SC}_\mathcal{G}$ is of size more than $|A| \ln |U| (1-2\alpha)$.*

Proof. Completeness follows from taking the set cover corresponding to each of the A vertices and its satisfying assignment.

Let us prove soundness. Assume on way of contradiction that there is a set cover C of $\mathcal{SC}_\mathcal{G}$ of size at most $|A| \ln |U| (1-2\alpha)$. For every $a \in A$ let s_a be the number of sets in C of the form $S_{a,\cdot}$. Hence, $\sum_{a \in A} s_a = |C|$. For every $b \in B$ let s_b be the number of sets in C that participate in covering $\{b\} \times U$. Then, denoting the A-degree of G by D_A,

$$\sum_{b \in B} s_b = \sum_{a \in A} s_a D_A \leq D_A |A| \ln |U| (1-2\alpha) = D |B| \ln |U| (1-2\alpha).$$

In other words, on average over the $b \in B$, the universe $\{b\} \times U$ is covered by at most $D \ln |U| (1-2\alpha)$ sets. Therefore, by Markov's inequality, the fraction of $b \in B$ whose universe $\{b\} \times U$ is covered by at most $D \ln |U| (1-\alpha) = \ell$ sets is at least α. By Lemma 2.5 and our construction, for such $b \in B$, there are two edges $e_1 = (a_1, b), e_2 = (a_2, b) \in E$ with $S_{a_1,\sigma_1}, S_{a_2,\sigma_2} \in C$ where $\pi_{e_1}(\sigma_1) = \pi_{e_2}(\sigma_2)$.

We define an assignment $\hat{\varphi}_A : A \to \binom{\Sigma_A}{\ell}$ to the A vertices as follows. For every $a \in A$ pick ℓ different symbols $\sigma \in \Sigma_A$ from those with $S_{a,\sigma} \in C$ (add arbitrary symbols if there are not enough). As we showed, for at least α fraction of the $b \in B$, the A vertices will not totally disagree. □

Fix a constant $0 < \alpha < 1$. The inapproximability ratio we get for Set-Cover from Claim 2.6 is $(1-2\alpha)\ln|U|$, assuming agreement soundness (ℓ, α). The latter is obtained from Corollary 2.4 for $\varepsilon = c/\log^4 n$ for a certain constant $c = c(\alpha)$. Let $N = |U|\,|B|$ be the number of elements in $\mathcal{SC}_{\mathcal{G}}$. We take $|U| = \Theta(|B|^{1/\alpha})$ (we might need to duplicate elements for that), so $\ln N = (1+\alpha)\ln|U|$, and the inapproximability ratio is at least $(1-3\alpha)\ln N$. Note that the reduction is polynomial in n. This proves Theorem 4.

3 Open Problems

The main open problem is to prove the projection games conjecture. We believe that many more hardness of approximation results could be proved based on the PGC. Two concrete open problems are to prove results for Clique and SVP. It will be interesting to show equivalence between certain strong hardness results and the PGC. Another very interesting open problem is to find better approximation algorithms for projection games.

Acknowledgments

The motivation to prove the Set-Cover result came from discussions with Ran Raz. The author would also like to thank Scott Aaronson, Zach Friggstad, Ryan O'Donnell, Muli Safra and anonymous reviewers for useful comments.

References

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[AMS06] N. Alon, D. Moshkovitz, and S. Safra. Algorithmic construction of sets for k-restrictions. *ACM Trans. Algorithms*, 2:153–177, 2006.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[AS03] S. Arora and M. Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003.

[BGLR93] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs and applications to approximations. In *Proc. 25th ACM Symp. on Theory of Computing*, pages 294–304, 1993.

[CHK09] M. Charikar, M. Hajiaghayi, and H. Karloff. Improved approximation algorithms for label cover problems. In *ESA*, pages 23–34, 2009.

[Chv79] V. Chvatal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4(3):233–235, 1979.

[CKW09] M. Cygan, L. Kowalik, and M. Wykurz. Exponential-time approximation of weighted set cover. *Inf. Process. Lett.*, 109(16):957–961, 2009.

[DFK+11] I. Dinur, E. Fischer, G. Kindler, R. Raz, and S. Safra. PCP characterizations of NP: Toward a polynomially-small error-probability. *Computational Complexity*, 20(3):413–504, 2011.

[Fei98] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.

[Hås99] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[Hås01] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[Kho01] S. Khot. Improved inapproximability results for maxclique, chromatic number and approximate graph coloring. In *Proc. 42nd IEEE Symp. on Foundations of Computer Science*, pages 600–609, 2001.

[Kho02] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symp. on Theory of Computing*, pages 767–775, 2002.

[Kho05] S. Khot. Hardness of approximating the shortest vector problem in lattices. *Journal of the ACM*, 52(5):789–808, 2005.

[LY93] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. In *Proc. 25th ACM Symp. on Theory of Computing*, 1993.

[MR10] D. Moshkovitz and R. Raz. Two query PCP with sub-constant error. *Journal of the ACM*, 57(5), 2010.

[NSS95] M. Naor, L. J. Schulman, and A. Srinivasan. Splitters and near-optimal derandomization. In *Proc. 36th IEEE Symp. on Foundations of Computer Science*, pages 182–191, 1995.

[Raz98] R. Raz. A parallel repetition theorem. In *SIAM Journal on Computing*, volume 27, pages 763–803, 1998.

[RS97] R. Raz and S. Safra. A sub-constant error-probability low-degree test and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th ACM Symp. on Theory of Computing*, pages 475–484, 1997.

[Sla96] Petr Slavík. A tight analysis of the greedy algorithm for set cover. In *Proc. 28th ACM Symp. on Theory of Computing*, pages 435–441, 1996.

[Sri99] A. Srinivasan. Improved approximations guarantees for packing and covering integer programs. *SIAM Journal on Computing*, 29(2):648–670, 1999.