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

Dana Moshkovitz [*]

July 2, 2014

**Abstract**

We establish a tight $\mathcal{NP}$-hardness result for the Set-Cover problem based on a strong PCP theorem. Our work implies that 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 $\alpha$ and the running time $\exp(N^{\Omega(\alpha)})$ assuming Sat requires exponential time.

The reduction is obtained by modifying Feige's reduction. The latter provides a lower bound of $\exp(N^{\Omega(\alpha/\log\log N)})$ 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 all their neighbors colored in different colors.

In the conference version of this work, the Set-Cover result was conditioned on a conjecture we call "The Projection Games Conjecture" (PGC), an instantiation of the Sliding Scale Conjecture of Bellare, Goldwasser, Lund and Russell to projection games. More precisely, the prerequisite was a quantitative version of this conjecture that was slightly beyond what was known at the time of the paper's writing. Shortly afterwords, Dinur and Steurer, based on a result by the author and Raz, proved the quantitative version of the conjecture sufficient for the Set-Cover result.

More broadly, in this paper we discuss the Projection Games Conjecture and its applications to hardness of approximation, e.g., to polynomial hardness factors for Closest-Vector-Problem and to studying the behavior of CSPs around their approximability threshold.

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

[*] dmoshkov@csail.mit.edu. Department of Electrical Engineering and Computer Science, MIT. This material is based upon work supported by the National Science Foundation under Grant Number 1218547.

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 $\mathcal{NP} \not\subseteq DTIME(n^{\text{poly}\log n})$ [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 prior to this work 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 $1/(\log n)^{O(1)}$ for PCP, 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. Moreover, the blow-up in our reduction is essentially optimal.

**Theorem 2.** *[With [DS13]] For every $0 < \alpha < 1$, (exact)* Sat *on inputs of size $n$ can be reduced in polynomial time to approximating* Set-Cover *to within $(1-\alpha)\ln N$ on inputs of size $N = n^{O(1/\alpha)}$.*

The theorem proves that approximating Set-Cover on inputs of size $N$ better than $(1-\alpha)\ln N$ is $\mathcal{NP}$-hard. 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 $2^{\Omega(n)}$ ("The Exponential Time Hypothesis" [IP99]). This follows from a sub-exponential $2^{O(N^{\alpha}+\text{poly}\log N)}$-time approximation algorithm for $(1-\alpha)\ln N$ approximating Set-Cover [CKW09].

## 2 Projection Games and the Projection Games Conjecture

In the conference version of this work [Mos12], Theorem 2 was conditioned on a conjecture we call "The Projection Games Conjecture" (PGC), or, more precisely, on a quantitative version of this conjecture that was slightly beyond what was known at the time of the paper's writing. Shortly afterward, Dinur and Steurer [DS13], based on a result by the author and Raz [MR08], proved the quantitative version of the conjecture sufficient for Theorem 2. Shortly after that, the author [Mos14] gave an alternative, and arguably simpler, proof of the required theorem. In this section we discuss 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 $\Sigma_A$, $\Sigma_B$; (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 $n = |A| + |B| + |E|$. The size of the alphabet of the projection game is $\max\{|\Sigma_A|, |\Sigma_B|\}$. A PCP Theorem for projection games with soundness error $\varepsilon$ and alphabet size $k$ (where $\varepsilon$ and $k$ may depend on $n$) states the following:

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

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

The author and Raz proved the following:

**Theorem 3** ([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 is $N$ for $N = n^{1+o(1)} \operatorname{poly}(1/\varepsilon)$. The projection game is over an alphabet of size exponential in $1/\varepsilon$, and has soundness error $\varepsilon$. The reduction can be computed in linear time in the size and the alphabet size of the projection game. The projection game is on a bi-regular graph whose degrees are* $\operatorname{poly}(1/\varepsilon)$.

One cannot hope for a soundness error that is lower than $1/N$. Hence, the dependence of $\varepsilon$ in $N$ is as low as possible up to the identity of the constant $c$. On the other hand, the alphabet size in Theorem 3 is not known to be tight. It can be shown that the alphabet size must be at least $1/\varepsilon$ where $\varepsilon$ is the soundness error (assuming $\mathcal{P} \neq \mathcal{NP}$). Moreover, certain PCP constructions – while deficient in other parameters – have alphabets of size $\operatorname{poly}(1/\varepsilon)$, see, e.g., [Raz98]. This motivates the conjecture that an alphabet size of $\operatorname{poly}(1/\varepsilon)$ could be achieved in Theorem 3 as well:

**Conjecture 1** (Projection Games Conjecture[1], 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 $N = n^{1+o(1)} \operatorname{poly}(1/\varepsilon)$ over an alphabet of size $\operatorname{poly}(1/\varepsilon)$ that has soundness error $\varepsilon$.*

In almost all applications, one wishes the size and the alphabet size to be at most polynomial in $n$. This happens in Theorem 3 only when $\varepsilon \geq 1/(\log N)^b$ for a constant $b > 0$. The PGC, on the other hand, gives polynomial size and alphabet size for any $\varepsilon \geq 1/N^c$.

The PGC is the Sliding Scale Conjecture of Bellare, Goldwasser, Lund and Russell [BGLR93] instantiated for projection games (of almost-linear size). 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. Today, focusing on the projection games version of the conjecture as in Conjecture 1 is folklore in the PCP community.

Approximation algorithms for projection games were researched [Pel07, CHK09, MM13], and the conjecture is consistent with the state of the art algorithm, giving $1/\varepsilon = O(\sqrt[4]{Nk})$ [MM13]. For PCPs with more than two queries (corresponding to games on hypergraphs, where the edges carry general predicates rather than projections), 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 $1/\epsilon$.

[1] A slightly weaker version of the Projection Games Conjecture is one in which the size of the projection game is polynomial $N = \operatorname{poly}(n, 1/\varepsilon)$ rather than almost-linear.

Dinur and Steurer show how to achieve soundness error that is poly-logarithmic in $N$ (for *any* poly-logarithm) simultaneously with polynomial-sized alphabet, at the cost of increasing the size. This suffices for the reduction to Set-Cover in Theorem 2 to go through. The idea is to apply parallel repetition on Theorem 3, and Dinur and Steurer were the first to successfully analyze parallel repetition for the relevant parameters:

**Theorem 4** ([DS13]). *There exists $c > 0$, such that for every $\varepsilon \geq 1/N^c$ and every $k \geq 1$,* Sat *on input of size $n$ can be reduced to a projection game on a bi-regular graph whose size is $N^k$ for $N = n^{1+o(1)} \operatorname{poly}(1/\varepsilon)$. The projection game is over an alphabet of size exponential in $1/\varepsilon$ and $k$, and has soundness error $(2\varepsilon)^{k/2}$. The reduction can be computed in linear time in the size and the alphabet size of the projection game. The projection game is on a bi-regular graph whose degrees are* $\operatorname{poly}(1/\varepsilon^k)$.

A similar theorem was later proved by the author [Mos14] via a combinatorial analysis of parallel repetition.

The Projection Games Conjecture has a similar flavor to the Unique Games Conjecture (UGC) of Khot [Kho02]: both assert that low soundness error[2] 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 $\pi_e$ are one-to-one. Unique games are 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[3] and bounded soundness error. The two conjectures, UGC and PGC, seem unrelated: neither would imply the other.

The following variant of the PGC is useful for hardness of approximation:

**Definition 5** (Linear projection game). *A linear projection game is a projection game in which the alphabets are of the form $\Sigma_A = \mathbb{F}^a$, $\Sigma_B = \mathbb{F}^b$, where $\mathbb{F}$ is a finite field, and $a \geq b$ are natural numbers. The projections in the game are affine transformations $\mathbb{F}^a \to \mathbb{F}^b$.*

**Conjecture 2** (Linear PGC). *There exists $c > 0$, such that for every $\varepsilon \geq 1/n^c$,* Sat *on inputs of size $n$ can be efficiently reduced to a linear projection game of size $N = n^{1+o(1)} \operatorname{poly}(1/\varepsilon)$ and alphabet size $\operatorname{poly}(1/\varepsilon)$. Satisfiable instances of* Sat *are mapped to projection games where $1 - \varepsilon$ fraction of the edges can be satisfied, while unsatisfiable instances of* Sat *are mapped to projection games where at most $\varepsilon$ fraction of the edges can be satisfied.*

Note that for linear projection games, one can efficiently distinguish the case where all edges can be satisfied from the case where not all edges can be satisfied – by Gaussian elimination. Therefore, it was necessary to modify the statement of Conjecture 1.

In Section 5 we discuss applications of the PGC and the linear PGC to proving polynomial hardness factors for the Closest-Vector-Problem and to studying the behavior of Max-3LIN and other CSPs around their approximability thresholds.

# 3 Preliminaries

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

[2] The unique games conjecture only asks for arbitrarily small constant soundness error $\varepsilon$, while the PGC asks for polynomially small error.

[3] 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").

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.

# 4 Set-Cover Hardness

## 4.1 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 6** (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 7** (Agreement soundness)**.** *Assume a projection game $(G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ for deciding whether a boolean formula $\phi$ is satisfiable. We say that $G$ has* agreement soundness error *$\varepsilon$, if for unsatisfiable $\phi$, 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 4.1** (Combinatorial construction)**.** *For $0 < \varepsilon < 1$, for a prime power $D$, and $n$ that is a power of $D$, there is an explicit construction of a regular graph $H = (U, V, E)$ with $|U| = n$, $V$-degree $D$, and $|V| \leq n^{O(1)}$ that satisfies the following. For every partition $U_1, \ldots, U_l$ of $U$ into sets, such that $|U_i| \leq \varepsilon |U|$ 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 $\varepsilon D^2$.*

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 4.1) 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

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

□

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 4.1.

**Lemma 4.2.** *Let $D \ge 2$ be a prime power and let $n$ be a power of $D$. Let $\varepsilon > 0$. From a projection game with soundness error $\varepsilon^2 D^2$ and $B$-degree $n$, 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 power.*

*Proof.* Let $\mathcal{G} = (G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ be the original projection game. Let $H = (U, V, E_H)$ be the graph from Lemma 4.1, where $n$, $D$ and $\varepsilon$ are as given in the current lemma. Let us 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 $b$.

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 an $\varepsilon D^2$ 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 $\varepsilon D^2$.

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 goodness of $b$ and the property of $H$, there must be $\sigma \in \Sigma_B$ such that $|U_\sigma| > \varepsilon\,|U|$. We call $\sigma$ 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 $\varphi_A, \varphi_B$ 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 $\ell$ values, and an agreement is interpreted as agreement on *one* of the assignments in the list:

**Definition 8** (List total disagreement). *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 9** (List agreement soundness). *Assume a projection game* $(G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ *for deciding membership whether a boolean formula* $\phi$ *is satisfiable. We say that* $G$ *has* list agreement soundness error $(\ell, \varepsilon)$, *if for unsatisfiable* $\phi$, *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 $\varepsilon$, then even when the prover is allowed to assign each $A$ vertex $\ell$ values, the game is still sound. This is argued in the next corollary:

**Lemma 4.3** (Projection game with list agreement soundness). *Let* $\ell \geq 1$, $0 < \varepsilon' < 1$. *A projection game with agreement soundness error* $\varepsilon'$ *has list agreement soundness error* $(\ell, \varepsilon'\ell^2)$.

*Proof.* Assume by 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 $\ell$ 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 $\varphi_A$ is at least $1/\ell^2$. Thus, under $\varphi_A$, the expected fraction of the $B$ vertices that have at least two neighbors that agree on them, is more than $\varepsilon'$. In particular, there exists an assignment to the $A$ vertices, such that more than $\varepsilon'$ fraction of the $B$ vertices have two neighbors that agree on them. This contradicts the agreement soundness of the game. □

Summarizing the above:

**Corollary 4.4.** *For any* $\ell = \ell(n) = \text{poly}\log n$, *for any constant prime power* $D \geq 2$ *and constant* $0 < \alpha < 1$, Sat *on input of size* $n$ *can be reduced to a projection game of size* $N = poly(n)$ *with alphabet size* $poly(n)$, *where the* $B$*-degree is* $D$, *and the list agreement soundness error is* $(\ell, \alpha)$.

*Proof.* Our starting point is the game from Theorem 4 with soundness error $(2\varepsilon)^{k/2}$ so $\sqrt{(2\varepsilon)^{k/2}} \leq \alpha/2(D\ell)^2$. We apply Lemma 4.2 and Lemma 4.3. □

## 4.2 Following Feige's Reduction

In the remainder, we will show how to use Corollary 4.4 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 uses at most one set out of each partition, is necessarily large:

**Lemma 4.5** (Partition system, [NSS95]). *For natural numbers* $m$, $D$ *and* $0 < \alpha < 1$, *for all* $u \geq (D^{O(\log D)} \log m)^{1/\alpha}$, *there is an explicit construction of a universe* $U$ *of size* $u$ *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* $\mathcal{P}_{i_j}$.

We will use the contrapositive of the lemma: if $U$ has a cover of size at most $\ell$, then this cover must contain at least two sets from the same partition. The choice of parameters of interest to us is: $m$ is at most polynomial in $n$ ($m$ will be $|\Sigma_B|$ of the projection game), $D$ is a sufficiently large constant, and $\alpha$ is a small constant.

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 of "anti-universal sets" in [NSS95] de-randomizes this randomized construction. This is the mapping from our notation to the notation in [NSS95]: $m \to n$, $D \to b$, $\ell \to k$, $U$ is the anti-universal set.

We now describe the reduction from a projection game $\mathcal{G}$ as in Corollary 4.4, to a Set-Cover instance $\mathcal{SC}_\mathcal{G}$.

Apply Lemma 4.5 for $m = |\Sigma_B|$ and $D$ which is the $B$-degree of the projection game. The parameter $u$ will be determined later. 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 $\Sigma_B$ symbols $\sigma_1, \ldots, \sigma_m$. The elements of the Set-Cover instance are $B \times U$. Equivalently, each vertex $b \in B$ has a copy of the universe $U$. Covering this universe corresponds to satisfying the edges that touch $b$. There are $m$ ways to satisfy the edges that touch $b$ – one for every possible assignment $\sigma \in \Sigma_B$ to $b$. The different partitions covering $U$ correspond to those different assignments.

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. Taking $S_{a,\sigma}$ to the cover would correspond to assigning $\sigma$ to $a$. Notice that a cover might consist of several sets of the form $S_{a,\cdot}$ for the same $a \in A$, which is the reason we consider list agreement. 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 $\{b\} \times S$, where $S$ is the $i$'th subset of the partition $\mathcal{P}_{\varphi_e(\sigma)}$.

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. Formally, we prove:

**Claim 4.6.** *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 4.5. If $\mathcal{G}$ has agreement soundness $(\ell, \alpha)$, 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 by 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 $\alpha$. By the contrapositive of Lemma 4.5 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 $\ell$ 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 $\alpha$ fraction of the $b \in B$, the $A$ vertices will not totally disagree. □

**Proof of Theorem 2:** Fix a constant $0 < \alpha < 1$ and a prime power $D$. For a sufficiently large $\ell' = \Theta(\log n)$, let $\mathcal{G} = (G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ be the projection game with list agreement soundness $(\ell', \alpha)$ obtained from Corollary 4.4. We take $u = |U| = \Theta(|B|^{1/\alpha})$, so $u \geq (D^{O(\log D)} \log |\Sigma_B|)^{1/\alpha}$ as required for Lemma 4.5. Let $\ell = D \ln u (1-\alpha) \leq \ell'$. The inapproximability ratio we get for Set-Cover from Claim 4.6 is $(1-2\alpha)\ln |U|$. Let $N = |U|\,|B|$ be the number of elements in $\mathcal{SC}_{\mathcal{G}}$. We have $\ln N = (1+\alpha)\ln |U|$. The inapproximability ratio is at least $(1-3\alpha)\ln N$. Note that the reduction is polynomial in $|A|$, $|\Sigma_A|$, $|B|$, $|\Sigma_B|$ and $|U|$. Hence, the reduction is polynomial in $n$. This proves Theorem 2.

# 5 Applications of the Projection Games Conjecture

In this section we describe a few applications of the PGC to hardness of approximation.

## 5.1 The Closest Vector Problem

The Closest-Vector-Problem (CVP) is to find, given a basis $b_1, \ldots, b_n \in \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, the closest point to $x$ – with respect to the $\ell_2$ distance – in the lattice spanned by $b_1, \ldots, b_n$, i.e., in $\{\sum_{i=1}^n \alpha_i b_i \mid \alpha_1, \ldots, \alpha_n \in \mathbb{Z}\}$.

Lattice problems like CVP are quite natural and have been researched a lot. One of the motivations for studying them comes from cryptography, where encryption systems believed to be secure even against quantum adversaries were built assuming the worst-case hardness of approximating lattice problems. The inapproximability factors known to be useful for cryptography are as large as $\tilde{\Theta}(n)$ for constructing collision resistant hash functions and one way functions [MR07], and $\Theta(n^2)$ for public-key cryptography [Reg09], but it is unlikely that such an approximation is $\mathcal{NP}$-hard, as it (and in fact any approximation to within roughly $\sqrt{n}$) would result in a collapse of the polynomial hierarchy [AR05]. For more details see [MG02, Reg10].

A central question is whether one can show that lattice problems are $\mathcal{NP}$-hard to approximate to within some polynomial factors $\ll \sqrt{n}$. The best existing $\mathcal{NP}$-hardness result for CVP is for a factor of $\exp((\log n)^{1-\alpha})$ for any constant $\alpha > 0$ (and even for certain $\alpha = o(1)$) [DKS98]. Assuming the PGC, we can obtain hardness of approximating CVP up to polynomial factors by a reduction of Arora, Babai, Stern and Sweedyk [ABSS97]. We will state the theorem as in [Kho10]:

**Theorem 10** (CVP Hardness [ABSS97])**.** *Given a projection game $\mathcal{G} = (G = (A, B, E), \Sigma_A, \Sigma_B, \Phi)$ one can construct in poly$(N)$ time a lattice $\mathcal{L}$ in $\mathbb{R}^N$ and a point $x \in \mathbb{R}^N$ where $N = |A|\,|\Sigma_A| + |B|\,|\Sigma_B|$, such that:*

- *Completeness: If there is an assignment to the vertices of $G$ that satisfies all of its edges, then the distance between $x$ and $\mathcal{L}$ is at most $\sqrt{2\,|A|\,|B|}$.*

- *Soundness: If there is no assignment to the vertices of $G$ that satisfies even $\varepsilon$ fraction of its edges, then the distance of $x$ and $\mathcal{L}$ is at least $0.1\sqrt{|A|\,|B|\,/\varepsilon}$.*

*Hence, assuming the PGC, there exists $c > 0$, such that approximating* Closest-Vector-Problem *to within $N^c$ on an $N$-dimensional lattice is $\mathcal{NP}$-hard.*

### 5.2 Around the Approximability Thresholds of CSPs

Constraint Satisfaction Problems (CSP) are defined by a set of variables $v_1, \dots, v_n$, an alphabet $\Sigma$, and constraints $\varphi_1, \dots, \varphi_m$, each depending on $q$ variables. The number $q = O(1)$ is called the *arity* of the CSP. The task is to find an assignment to the variables that maximizes the number of satisfied constraints. One obtains specific CSPs by restricting the type of constraints. Examples include Max-3Sat, where one is given 3CNF clauses on Boolean variables, and Max-qLin, where one is given linear equations over $GF(2)$.

CSPs were studied a lot in hardness of approximation, and for many of them we know sharp approximability thresholds. In fact, assuming the Unique Games Conjecture, we know that all CSPs over constant-sized alphabets have thresholds, where they pass from admitting polynomial time algorithms to being $\mathcal{NP}$-hard [Rag08]. For specific problems like Max-3Lin, we know even sharper results:

**Theorem 11** (Hardness of Max-3Lin [Hås01, Kho01]). *Linear projection games on inputs of size $n$ and soundness/completeness error $\varepsilon$ can be reduced to distinguishing, given a* Max-3Lin *instance of size $N = n\,\mathrm{poly}(1/\varepsilon)$, between the case that $(1 - \varepsilon')$ fraction of the equations can be satisfied, and the case where no assignment satisfies more than $(1/2 + \varepsilon')$ fraction of the equations, where $\varepsilon = \mathrm{poly}(\varepsilon')$. The reduction is linear in $N$.*

*Hence, assuming the linear PGC, approximating* Max-3Lin *to within $1/2 + 1/N^c$ for some constant $c > 0$ is $\mathcal{NP}$-hard.*

Note that a random assignment to the variables satisfies half of the equations in expectation, and one can always find in deterministic polynomial time an assignment that satisfies at least half of the equations. The theorem says that approximating Max-3Lin transitions from being easy to being hard within a window of $\varepsilon'$ at 1/2. The width $\varepsilon'$ determines how sharp the phase transition is. Note that at $1/2+1/N^{o(1)}$ the approximation problem is (essentially) exponentially hard assuming the exponential time hypothesis and the linear PGC.

Theorem 11 is proved by using the Hadamard code as in [Kho01] instead of the long code as in [Hås01]. The advantage of the reduction in [Hås01] is that it allows one to start with (non-linear) projection games. Its disadvantage is that it incurs a blow-up of $N = n \exp(1/\varepsilon)$. Using [Hås01] and Theorem 3, the current record, not assuming the linear PGC, is $\varepsilon' = 1/(\log\log N)^{O(1)}$.

Results analogous to Theorem 11 hold for other CSPs as well, e.g., for Max-3Lin over larger finite fields, for Max-3Sat and for other problems from Håstad's paper [Hås01].

## 6 Open Problems

The main open problem is to prove (or disprove) the Projection Games Conjecture.

We believe that many more hardness of approximation results could be proved based on the PGC. Several concrete open problems are:

1. Prove a theorem similar to Theorem 11 for *satisfiable* instances of Max-3Sat.
2. Prove PGC-based hardness results for large families of CSPs similar to what is known under the Unique Games Conjecture for all CSPs [Rag08]. A significant step in this direction was recently taken by Chan [Cha13].
3. Prove a PGC-based hardness result for approximating Shortest-Vector-Problem to within polynomial factors. Note that there is a quasi-polynomial reduction from Closest-Vector-Problem to Shortest-Vector-Problem [Kho05, HR12] (see survey [Kho10]).
4. Prove a PGC-based hardness result for approximating Clique to within $N/\operatorname{poly}\log N$. Note that there is a quasi-polynomial reduction from Max-3Lin to Clique [Kho01, KP06].

Another open problem is to show equivalence between the PGC and the linear PGC.

## 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

[ABSS97] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. *Journal of Computer and System Sciences*, 54(2):317–331, 1997.

[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.

[AR05] D. Aharonov and O. Regev. Lattice problems in np $\cap$ conp. *Journal of the ACM*, 52(5):749–765, September 2005.

[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.

[Cha13] S. O. Chan. Approximation resistance from pairwise independent subgroups. In *Proc. 45th ACM Symp. on Theory of Computing*, pages 447–456, 2013.

[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.

[DKS98] I. Dinur, G. Kindler, and S. Safra. Approximating-CVP to within almost-polynomial factors is NP-hard. In *FOCS*, pages 99–111, 1998.

[DS13] I. Dinur and D. Steurer. Analytical approach to parallel repetition. *CoRR*, abs/1305.1979, 2013.

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

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

[HR12] I. Haviv and O. Regev. Tensor-based hardness of the shortest vector problem to within almost polynomial factors. *Theory of Computing*, 8(1):513–531, 2012.

[IP99] R. Impagliazzo and R. Paturi. The complexity of k-SAT. In *2012 IEEE 27th Conference on Computational Complexity*, pages 237–240, 1999.

[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.

[Kho10] S. Khot. Inapproximability results for computational problems on lattices. In *The LLL Algorithm*, pages 453–473. 2010.

[KP06] S. Khot and A. K. Ponnuswami. Better inapproximability results for maxclique, chromatic number and min-3lin-deletion. In *Proc. Automata, Languages and Programming, 33rd International Colloquium*, pages 226–237, 2006.

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

[MG02] D. Micciancio and S. Goldwasser. *Complexity of Lattice Problems: a cryptographic perspective*, volume 671 of *The Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers, 2002.

[MM13] P. Manurangsi and D. Moshkovitz. Improved approximation algorithms for projection games - (extended abstract). In *ESA*, pages 683–694, 2013.

[Mos12] D. Moshkovitz. The projection games conjecture and the NP-hardness of lnn-approximating set-cover. In *APPROX-RANDOM*, pages 276–287, 2012.

[Mos14] D. Moshkovitz. Parallel repetition from fortification. In *Proc. 55th IEEE Symp. on Foundations of Computer Science*, 2014.

[MR07] D. Micciancio and O. Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM Journal on Computing*, 37(1):267–302, 2007.

[MR08] D. Moshkovitz and R. Raz. Sub-constant error low degree test of almost-linear size. *SIAM Journal on Computing*, 38(1):140–180, 2008.

[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.

[Pel07] D. Peleg. Approximation algorithms for the label-cover$_{\mathrm{max}}$ and red-blue set cover problems. *J. Discrete Algorithms*, 5(1):55–64, 2007.

[Rag08] P. Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proc. 40th ACM Symp. on Theory of Computing*, pages 245–254, 2008.

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

[Reg09] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):34:1–34:40, 2009.

[Reg10] O. Regev. On the complexity of lattice problems with polynomial approximation factors. In *The LLL Algorithm*, pages 475–496. 2010.

[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.