1 The PCP Theorem

The usual formulation of the PCP theorem is equivalent to:

Theorem 1. There exists a finite constraint language $\mathcal{A}$ such that the following problem is NP-hard: given a CSP($\mathcal{A}$)-instance $\Phi$ decide if $\Phi$ is satisfiable or if every subset of the constraints of size $|\Phi|/2$ is unsatisfiable.

The problem in the statement is also called GapCSP($\mathcal{A}$).

2 The Label Cover Problem

Let $\Sigma$ be a finite alphabet. An instance of the label cover problem for $\Sigma$ is:

- a bipartite graph $(V_1, V_2, E)$,
- for every $(u, v) \in E$, a function $\pi_{u,v} : \Sigma \rightarrow \Sigma$.

A solution to an instance is a pair of mappings $s_1 : V_1 \rightarrow \Sigma$ and $s_2 : V_2 \rightarrow \Sigma$ such that for all $(u, v) \in E$, $\pi_{u,v}(s_1(u)) = s_2(v)$.

Proposition 1. Label cover over a 3-element alphabet is NP-hard.

Definition 1. A (strong) Mal’cev condition $(\mathcal{L}, \Sigma)$ where $\mathcal{L}$ is a finite functional language and $\Sigma$ is a finite set of identities over $\mathcal{L}$.

Example 1. $m(x, x, y) \approx m(y, x, x) \approx y$ with $\mathcal{L} = \{m\}$.

Definition 2. A Mal’cev condition $(\mathcal{L}, \Sigma)$ is satisfied in a clone $\mathcal{C}$ if there is a mapping $\xi : \mathcal{L} \rightarrow \mathcal{C}$ such that $\xi(\mathcal{L}) \models \Sigma$.

A Mal’cev condition is said to be trivial if one of the following equivalent statements hold:

1. it is satisfied in the clone of projections,
2. it is satisfied in every clone,
3. it is satisfied in $\text{Pol}(1\text{-in-3-SAT})$.

A Mal’cev condition is said to be bipartite if there is a partition $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ such that every identity is of the form $f(\text{variables}) \approx g(\text{variables})$ with $f \in \mathcal{L}_1$ and $g \in \mathcal{L}_2$. It is special bipartite if it is of the form $f(\text{variables}) \approx g(x_1, \ldots, x_n)$ (all variables are different in the right-hand side).

**Proposition 2.** It is NP-hard to decide whether a height 1 special bipartite Mal’cev condition is trivial.

We first prove that the two problems are equivalent up to polytime reductions.

The translation from the Mal’cev problem to LC($\Sigma$) is as follows. The vertices $(V_1, V_2)$ correspond to $(\mathcal{L}_1, \mathcal{L}_2)$. A set of identities $\mathcal{E}$ gives an edge relation on $V_1 \cup V_2$. Given an edge $f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) \approx g(x_1, \ldots, x_n)$, we take $\pi_{f,g}$ to be $\pi$.

Was the other reduction given?

**Proof of Proposition 2.** We even prove that the even more special case of height 1 bipartite special ternary in two variables is NP-hard.

We reduce 1-in-3-SAT to this problem. Let $V, C$ be an instance of 1-in-3-SAT. We take $\mathcal{L}_2$ to be the set of variables. We take $\mathcal{L}_1$ to contain one symbol $f_C$ for each constraint $C(x_{i_1}, x_{i_2}, x_{i_3})$ in $C$. We add the equalities $f_C(x, x, y) \approx g_{i_3}(x, y), f_C(x, y, x) \approx g_{i_2},$ and $f_C(y, x, x) \approx g_{i_1}$.

**Definition 3.** The GapLabelCover$_\epsilon(\Sigma)$ problem is as follows: given an LC($\Sigma$)-instance, decide if it is solvable, or if no $\epsilon$-part is solvable.

**Theorem 2.** For all $\epsilon > 0$, there exists $\Sigma$ such that GapLabelCover$_\epsilon(\Sigma)$ is NP-hard.

Be convinced that PCP implies that there exist $\epsilon \in (0, 1)$ and $\Sigma$ such that GapLabelCover$_\epsilon(\Sigma)$ is NP-hard. Parallel repetition (no time to formalize) gives a way to turn $\epsilon$ into any number in $(0, 1)$.

### 3 Expander graphs

“I only have two hours to talk so let’s get on with it.”

Marcello Mamino

We want to prove: fix finite domain $\Sigma$ and let $\Gamma$ be the full binary constraint language on $\Sigma$. We prove: the classical problem CSP($\Gamma$) reduces in polynomial-time to GapCSP($\Gamma$).

More precisely. Let $\Phi$ be an instance of CSP($\Gamma$) that is unsatisfiable, but a big portion of the constraints are simultaneously satisfiable. We want to construct a new instance $\Psi$ such that at most half of the constraints are satisfiable. Let $n$ be the number of constraints of $\Phi$. The gap of $\Phi$ is the smallest number
$\alpha \in [0, 1]$ such that a part of $\Phi$ of size $(1 - \alpha)n$ is satisfiable. If $\Phi$ is not satisfiable, the gap is $> 0$. We amplify this gap by multiplying it repeatedly by a constant $k$, which is meant to be independent of the size of $\Phi$.

At each step, suppose that our instance grows by a factor of $h$ (independent of $|\Phi|$ as well). This means that the size of the instance grows as $h^{\text{number of steps}} \sim n^{1 \log h(k)}$.

Strategic goal: $\exists \Sigma, k > 1, h, \alpha$ such that we can effectively translate an instance of CSP(Γ) of gap $x < \alpha$ to one of gap $xk$ and size not larger than $h \times$ original size.

We define the product of two instances $\Phi_1(\pi)$ and $\Phi_2(\eta)$. It contains as variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ and constraints $R(\pi') \land S(\eta')$ (of arity $\text{len}(\pi') + \text{len}(\eta')$) for every pair of constraint $R(\pi')$ and $S(\eta')$ from $\Phi_1$ and $\Phi_2$.

Claim. If $\alpha$ is the gap of $\Phi_1$ and $\Phi_2$, then gap of the product is $1 - (1 - \alpha)^2$.

Proof. Easy, just project on each coordinate, this gives subsets of each instance with relative size at most $(1 - \alpha)$. ♦

This proves that once we manage to go from a gap that depends on $\Phi$ to the $\alpha$ from the statement (which is a constant), then we can also go to $\frac{1}{2}$ by iterated squaring (a constant number of times).

Let’s now go back to the strategic goal. One step to reduce the gap is composed of the following substeps:

1. Regularize the graph (i.e., turn it into an expander).
2. Increase the gap (this causes the domain size of increase).
3. Shrink the domain.

Expanders are used at every step of this strategy.

**Definition 4** (Expansion ratio). Let $G$ be a graph. The edge expansion ratio of $G$, denoted by $h(G)$, is defined as

$$\min_{S \subseteq G, |S| \leq \frac{|V(G)|}{2}} \frac{|E(S, S')|}{|S|},$$

where $E(S, S')$ is the set of edges with one endpoint in $S$ and the other endpoint in $S$.

Little game of the expanders: find infinite families of $d$-regular graphs with $h(G) \geq \alpha > 0$. It’s easy to build such a family by a probabilistic argument, but we would like to compute them efficiently and deterministically.

Let $M(G)$ be the adjacency matrix of a $d$-regular graph $G$. This is a symmetric matrix with nonnegative integer entries, where the entry $(v, w)$ for $v, w \in G$ counts the number of edges between $v$ and $w$. Take $\frac{1}{d} M(G)$, which is a bistochastic matrix. Being a symmetric real matrix, it is diagonalisable as $A \times \text{diag}(1, \lambda_2, \ldots, \lambda_{|G|}) \times A^{-1}$ with $1 \geq \lambda_2 \geq \cdots \geq \lambda_{|G|}$ being the eigenvalues of $\frac{1}{d} M(G)$. 3
Exercise. Remember that the spectral radius of a stochastic matrix is precisely 1. Prove that \(-1\) is an eigenvalue iff the graph is bipartite.

**Definition 5.** The spectral gap of \(G\) is defined to be \(1 - \lambda_2\). We also define it as \(\min\{1 - |\lambda_2|, 1 - |\lambda_1|\}\).

Remark: if \(v\) is a probability distribution on \(V(G)\), a random walk with \(n\) steps gives the distribution \((\frac{1}{2}M(G))^n v\), and the matrix \((\frac{1}{2}M(G))^n\) looks like \(A \text{ diag}(1, \lambda_2, \ldots, \lambda_{n(G)}) A^{-1}\), where the diagonal converges quickly to diag\((1, 0, \ldots, 0)\).

**Theorem 3.** For every \(d\)-regular finite graph \(G\) with second greatest associated eigenvalue \(\lambda_2\), we have

\[
d\frac{1 - \lambda_2}{2} \leq h(G) \leq d\sqrt{2(1 - \lambda_2)}.
\]

**Proof.** Part 1: We prove the second inequality. Construct the matrix \(N\) with rows \(V\) and columns \(E\). Fix an arbitrary orientation on \(G\). For each edge \(e = (v, w)\), we put \(-1\) in \(N\) at position \(n\), and 1 at position \(n\).

Let \(f: V \rightarrow \mathbb{R}\) be a column vector. Then \(\|f^T N\|^2 = \sum_{(i,j) \in E(G)} |f(i) - f(j)|^2\). Define \(B_f := \sum_{(i,j) \in E(G)} |f(i) - f(j)|^2\). We prove that

\[
\|f\|^2 h(G) \leq B_f \leq \sqrt{2d} \|f^T N\| \cdot \|f\|
\]

holds when the support of \(f\) has size at most \(\frac{d}{2}\) and is nonnegative. Then by choosing \(f\) appropriately, this proves the second inequality.

We prove \(\|f\|^2 h(G) \leq B_f\). Let \(\beta_0, \ldots, \beta_r\) be an enumeration of the image of \(f\) where \(\beta_0 < \cdots < \beta_r\). Note that \(\beta_0 = 0\) since the support of \(f\) is not the whole \(V\). Let \(V_i\) be the set of \(x \in V(G)\) such that \(f(x) \geq \beta_i\). We have \(V(G) = V_0 \supseteq V_1 \supseteq \cdots \supseteq V_r\).

Claim. \(B_f\) is equal to \(\sum_{i=1}^r |E(V_i, V_{i-1})| \times (\beta_i^2 - \beta_{i-1}^2)\).

Proof. Quite clear. \(\diamondsuit\)

Now, \(\sum_{i=1}^r |E(V_i, V_{i-1})| \times (\beta_i^2 - \beta_{i-1}^2) \geq h(G) \sum_{i=1}^r |V_i| \times (\beta_i^2 - \beta_{i-1}^2)\), where we use the fact that \(|V_i| \leq \frac{|V(G)|}{2}\). We can rewrite this as \(h(G) \sum_{i=1}^r \beta_i^2(|V_i| - |V_{i+1}|)\), if we define \(V_{r+1} := \emptyset\). Finally, we note that this is simply equal to \(h(G)\|f\|^2\).

On to prove \(B_f \leq \sqrt{2d} \|f^T N\| \cdot \|f\|\). But this is just stupid computations:

\[
B_f = \sum_{(x,y) \in E} |f(x) - f(y)| \cdot |f(x) + f(y)|
\]

\[
\leq \sqrt{\sum (f(x) - f(y))^2} \cdot \sqrt{\sum (f(x) + f(y))^2}
\]

\[
\leq \|f^T N\| \cdot \sqrt{2 \sum f(x)^2 + f(y)^2}
\]

\[
\leq \|f^T N\| \cdot \sqrt{2d} \|f\|.
\]
What we just proved is that

\[ h(G) \leq \sqrt{2d} \frac{||f^TN||}{||f||} \]

holds for every nonnegative function \( f : V \rightarrow \mathbb{R} \) with support of size at most \( |V|/2 \). Let \( g \) be an eigenvector associated with the eigenvalue \( \lambda_2 \). Define \( f := g^+ \), i.e., the vector that is like \( g \) when it is nonnegative and is 0 otherwise. If the support of \( f \) is too big, just pick \( g' := -g \) and do the same. Define another matrix \( L := N \times N^T \) of size \( |V(G)|^2 \). The matrix \( L \) has \( d 's \) on the diagonal and at position \( |i,j| \) is \(-1\times\) number of edges connecting \( i \) and \( j = -dm_{i,j} \). Therefore, \( L = dl - dM \).

Let \( x \) be in the support of \( f \). We compute

\[
(Lf)(x) = d \cdot f(x) - \sum_{y \in V(G)} d \cdot M_{x,y} f(y)
= d \cdot g(x) - \sum_{y \in \text{supp}(f)} dM_{x,y} g(y)
\leq (Lg)(x)
= (dg - d\lambda_2 g)(x)
= d(1 - \lambda_2) g(x).
\]

This is amazing, because it gives \( ||f^TN||^2 = f^TN^T f = f^TLf = (f^TL|\text{supp}(f)) \cdot f = ||f^TL|\text{supp}(f)|| \cdot ||f|| \leq d(1 - \lambda_2) ||f||^2 \). So, in the inequality at the top of the page, we obtain \( h(G) \leq ... \).

**Part 2:** let us now prove the first inequality. By rearranging, we want to prove \( d\lambda_2 \geq d - 2h(G) \). First, it suffices to prove that there exists \( f : V \rightarrow \mathbb{R} \) that is orthogonal to \( \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \) and such that we have \( \frac{d^TML}{||f||^2} \geq d - 2h(G) \).

Indeed, suppose that we have such an \( f \) where the value of the fraction is \( \alpha \). Let \( A \) be orthogonal and such that \( A^TMA \) is diagonal. Let \( g := A^T f \). The first observation is that the first component of \( g \) is 0 (trivial). Then just compute.

Let \( S \) be a set witnessing the edge-expansion ratio. Define \( f := |S|1_S - |S|1_\overline{S} \). It is indeed orthogonal to \( 1_V \) and then it is simple computation. The denominator of the fraction is \( ||f||^2 = |S|^2 \cdot |S| + |S|^2 \cdot |\overline{S}| = (|S| + |\overline{S}|) \times |S||\overline{S}| = n|S||\overline{S}| \). Note that \( f \) can be written as \( n1_S - |S|1_V \) and also as \(-n1_\overline{S} + |S|1_V \). The numerator of the fraction is then

\[
f^T dM f = (n1_S - |S|1_V)^T M (-n1_\overline{S} + |S|1_V)
= -n^2 1_S^T dM 1_S + n|S|1_S^T dM 1_V + n|S|1_V^T dM 1_\overline{S} - |S||\overline{S}| 1_S^T dM 1_\overline{S}
= -n^2 |E(S, \overline{S})| + nd|S||S| + nd|S||\overline{S}| - |S||\overline{S}|n \cdot d
= nd|S||\overline{S}| - n^2 h(G)|S|.
\]
The fraction is then
\[
d^{-\frac{f^T M f}{\|f\|_2}} = \frac{n d |S|S| - n^2 |E(S, \overline{S})|}{n |S||\overline{S}|}
\]
\[
= d - \frac{n h(G)}{|S|}
\]
\[
\geq d - 2h(G).
\]

This concludes the proof.

Let us now see a construction of an infinite family of expanders!

**Definition 6.** A graph \(G = (V, E)\) is an \((n, d, \alpha)\)-expander if it is \(d\)-regular, \(|V| = n\) and all the associated eigenvalues, except for 1 are bounded above in absolute value by \(\alpha\).

We define the zigzag product of two graphs. Let \(G\) be an orange graph that is \((n, m, \alpha)\) and \(H\) be a blue graph that is \((m, d, \beta)\). We build \(G \odot H\) by first exploding every orange vertex by \(m\) vertices and then pasting arbitrarily a copy of \(H\) inside. The zigzag product is then the graph on this vertex set and where there is an edge \((x, y)\) if there is a blue edge from \(x\) to \(z\), an orange edge from \(z\) to \(w\), and a blue edge from \(w\) to \(y\).

**Theorem 4.** \((n, m, \alpha) \odot (m, d, \beta) = (n \cdot m, d^2, f(\alpha, \beta))\), where \(f(\alpha, \beta) \leq \alpha + \beta + \beta^2\). Moreover, if \(\alpha, \beta < 1\), then \(f(\alpha, \beta) < 1\).

**Proof.** We only prove the claim about \(\alpha + \beta + \beta^2\). Let \(M\) be the normalized (or nonnormalized, but probably normalized) adjacency matrix of the resulting zigzag product. We can write \(M\) as \(HGH\), where \(H\) is the adjacency matrix of the \((m, d, \beta)\)-expander and \(G\) is the adjacency matrix of the \((n, m, \alpha)\)-expander.

We give an upper bound of \(||f^T HGH f||/||f||^2\), for \(f\) orthogonal to \(1_V\). Write \(f := f'' + f^\perp\), where \(f''\) is constant in each town and \(f^\perp\) is such that the sum on each town is 0. Expand the product in the numerator, which gives

\[
|f''^T HGH f'' + 2f'' HGH f^\perp + f^\perp^T HGH f^\perp| \leq |f''^T HGH f''| + 2 |f'' HGH f^\perp| + |f^\perp^T HGH f^\perp|
\]
\[
= |f''^T G f''| + 2 |f'' HGH f^\perp| + |f^\perp^T HGH f^\perp|
\]
\[
\leq \alpha |f''|^2 + 2 ||G|| ||f''|| ||H f^\perp|| + ||G|| \cdot ||H f^\perp||^2
\]
\[
\leq \alpha |f|^2 + 2 ||G|| |f||f^\perp|| + \beta^2 ||f||^2
\]
\[
\leq \alpha |f|^2 + \beta |f|^2 + \beta^2 ||f||^2.
\]
How to build such an $H$? Let $q$ be a prime number. Consider $V := (\mathbb{F}_q)^2$. The edges are $\{(a, b), (c, d)\}$ such that $ac = b + d$. The intuition is that of $(a, b)$ being on the line $(c, d)$, in a projective setting.

**Proposition 3.** $G = (V, E)$ is a $(q^2, q, \frac{1}{\sqrt{q}})$ expander.

**Proof.** It is clearly a $q$-regular graph. Let $M$ be the adjacency matrix. We have

$$(qM)^2 = \begin{pmatrix} q & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & q & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & q & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & q & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & q & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & q & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$  

Write the matrix as $q \cdot I + 1_q \otimes (1_q - I)$, where $1_q$ is the square matrix of size $q$ with all 1’s. Thus, the eigenvalues of $(qM)^2$ are $q, 0, q^2$. Therefore, after normalizing the second greatest eigenvalue is $\frac{1}{\sqrt{q}}$. \[\square\]

Now define:

- $AP_q := G$,
- $AP^1_q := AP_q \otimes AP_q$,
- $AP^{n+1}_q := AP^n_q \otimes AP_q$

**Proposition 4.** $AP^n_q$ is $(q^{2(n+1)}, q^2, O(\frac{1}{\sqrt{q^n}}))$ (where the big-Oh notation is with respect to $q$, not $n$!).

Pick $n = 3$, $q$ large enough, this gives the $H$ that we wanted above.

## 4 Preprocessing

Recall that our goal is to go from an instance $\Phi$ of CSP($\Sigma$) to an instance $\Psi$ of CSP($\Sigma$) whose gap is at least $\alpha$, for some constant $\alpha$ that does not depend on $\Phi$. Moreover, we want this transformation to be doable in polynomial time. The formal statement is as follows:

**Theorem 5.** There exists a finite domain $\Sigma_0$ such that for every finite domain $\Sigma$, there are $C > 0$ and $0 < \alpha < 1$ such that, given an input $\Phi$ of CSP($\Sigma$), there exists $\Psi$ an instance of CSP($\Sigma_0$) that is computable in polynomial time and such that:
• $|\Psi| \leq C \cdot |\Phi|$, 
• if $\Phi$ is not satisfiable then $\Psi$ is not satisfiable, 
• $\text{gap}(\Psi)$ is at least $\min(2 \cdot \text{gap}(\Phi), \alpha)$.

To prove this theorem, we iterate a three-step procedure:

Preprocessing Turn the input sentence $\Phi$ into a $d$-regular expander $\Theta_1$,

Amplification Compute an instance $\Theta_2$ of CSP($\Sigma'$) with bigger gap, but also $|\Sigma'| = |\Sigma|^{\text{poly}(d)}$,

Reduction Compute an instance $\Theta_3$ of CSP($\Sigma_0$) with controlled gap.

Here, we deal with the first step. It might be useful to recall that we view instances of the CSP as graphs $(V,E)$ where each edge $(u,v)$ is labelled by a subset of $\Sigma \times \Sigma$ that gives the acceptable values for an assignment to the pair $(u,v)$.

We first need the following refinement of Marcello’s expanders. (Note: in the following, I consider the eigenvalues of the non-normalized adjacency matrix. Deal with it.)

Lemma 1. There exists $d'$ and $0 < \lambda_0 < d'$ such that one can construct in polynomial time a family $(H_n)$ of $d'$-regular expanders such that for all $n \geq 0$, the graph $H_n$ has $n$ vertices and $\lambda_2(H_n) < \lambda_0$.

Definition 7. Let $G$ be a constraint graph over $\Sigma$. The constraint graph $G' := \text{prep}_1(G)$ is defined as follows:

• $V'$ is the set of all elements of the form $v := \{(v,e)| e \in E, v \in e\}$, for $v \in V$.
• $E'$ is the following set. First, all the edges of the form $\{(v,e),(v',e)\}$ where $e = \{v,v'\} \in E$ are in $E'$. Secondly, for every $v$ let $H_v$ be a $d$-regular expander on $[v]$ (with constant expansion ratio $h$). Add all the edges of $H_v$ to $E'$.

• For every edge of the second kind, the constraint-label is “equality” (i.e., $\{(a,a) : a \in \Sigma\}$). For every edge $\{(v,e),(v',e)\}$, the constraint is the constraint of $e$.

Clearly, $\text{prep}_1(G)$ is a $d + 1$ regular graph, $|V'| \leq 2|E|$, and $|E'| = |E| + \sum_v |E(H_v)| \leq |E| + d|E| = (d + 1)|E|$. We show that the gap of $\text{prep}_1(G)$ is controlled by the gap of $G$.

Proposition 5. We have $O(1) \cdot \text{gap}(G) \leq \text{gap}(\text{prep}_1(G)) \leq \text{gap}(G)$. Moreover, suppose we have an assignment (not necessarily satisfying) $f : V' \rightarrow \Sigma$. Define $g : V \rightarrow \Sigma$ by looking at the values $f(v,e)$ for $v \in E$ and $(v,e) \in [v]$ and assigning $g(v)$ to the one that appears the most often (in case of tie, decide arbitrarily). Then $O(1) \cdot \text{gap}_g(G) \leq \text{gap}_f(\text{prep}_1(G))$. 

8
Proof. Let $G' := \text{prep}_1(G)$. We start by the bound $\text{gap}(G') \leq \text{gap}(G)$. Let $f : V \to \Sigma$ be an assignment witnessing $\text{gap}(G)$. Define $g : V' \to \Sigma$ by $g(v, e) := f(v)$. Clearly $g$ satisfies the equality constraints, and if $\{ (v, e), (v', e) \}$ is an edge, we have $(g(v, e), g(v', e)) = (f(v), f(v'))$. Therefore, if $f$ satisfies the constraint associated with $e$, then $g$ satisfies the constraint associated with $\{ (v, e), (v', e) \}$. So $g$ violates exactly $\text{gap}(G)|E|$ constraints of this type. Therefore, $\text{gap}(\text{prep}_1(G)) \leq \frac{|\text{gap}(G)|}{|E|} \leq \text{gap}(G)$.

We prove the second part of the statement (which implies the first inequality). Let $f : V' \to \Sigma$ and $g : V \to \Sigma$ be as in the statement. Let $F \subseteq E$ be the set of edges $e = (u, v)$ such that $(g(u), g(v))$ does not satisfy the constraint associated with $e$ in $G$. Similarly, let $F' \subseteq E'$ be the corresponding set of edges for $f$. Thus, $\text{gap}_g(G) = \frac{|F|}{|E|}$ and $\text{gap}_f(G') = \frac{|F'|}{|E'|}$. If $|F'| \geq \frac{\text{gap}_g(G)}{2}|E|$, we have

$$\text{gap}_f(G') \cdot |E'| \geq \frac{\text{gap}_g(G)}{2},$$

so that

$$\text{gap}_f(G') \geq \frac{\text{gap}_g(G)}{2} \cdot \frac{|E'|}{|E|} \geq \frac{\text{gap}_g(G)}{2(d+1)},$$

and hence $\text{gap}_g(G) \leq 2(d+1) \text{gap}_f(G')$ as required.

Otherwise, $|F'| < \frac{\text{gap}_g(G)}{2}|E|$. Let $S = \bigcup_{v \in V} \{ (v, e) \mid g(v) \neq f'(v, e) \}$. Observe that for every $e = \{v, v'\} \in F$, either $\{ (v, e), (v', e) \} \in F'$, or we have $f(v, e) \neq g(v')$ or $f(v', e) \neq g(v')$, in which case $v$ or $v'$ is in $S$. Hence, $|F'| + |S| \geq |F| = \text{gap}_g(G) \cdot |E|$. With the bound on $|F'|$ that we have, this gives $\text{gap}_g(G) \cdot |E| < \frac{\text{gap}_g(G)}{2}|E| + |S|$, i.e., $|S| > \frac{\text{gap}_g(G)}{2}|E|$. Fix one vertex $v$. Partition naturally the set $S_v := [v] \cap S$ into $\bigcup_{u \in \Sigma} S_v^u$. By the very definition of $g$, it is the case that $|S_v^u| < \frac{|v|}{2}$ holds for all $a \in \Sigma$. Since $H_v$ is an expander on $[v]$, we have that $|E(S_v^u, S_v^u) \cup u| \geq |S_v^u| \cdot h$. Note that the constraint of every edge between $S_v^u$ and $S_v^u$ is violated by $f$, because this constraint is an equality constraint. Therefore, $f$ violates at least $\sum_{u \in V} \frac{1}{2} \sum_{a \in \Sigma} |S_v^u| \cdot h \geq \frac{h}{2} \sum_{v \in V} |S_v \cap [v]| > \frac{\text{gap}_g(G)}{4}|E|$. That is to say, $\frac{h}{3(d+1)} \text{gap}_f(G) \leq \text{gap}_f(G')$.}

Now, we have a regular graph with bounds on the gap. We seek to make this graph into an expander with loops.

**Definition 8.** Let $G$ be a graph, and let $H$ be a $d'$-regular expander over the same vertices as $G$ and with second eigenvalue $\lambda_2(H) < \lambda_0 < d'$. Define $\text{prep}_2(G)$ to be the graph over the same vertex set defined as follows. The edges of $\text{prep}_2(G)$ are: edges of $G$, edges of $H$, and a loop at every vertex (we allow parallel edges here). The constraints are: the constraints of $G$ for edges of $G$, and $\Sigma \times \Sigma$ for all other edges.

It is clear that $\text{prep}_2(G)$ has $|V|$ vertices, $|E(G)| + |E(H)| + |V|$ edges. Moreover, if $G$ is $d$-regular, then $\text{prep}_2(G)$ is $d + d' + 1$ regular.

**Proposition 6.** Let $G$ be a graph and let $G' := \text{prep}_2(G)$. Then
\[ \lambda_2(G') \leq d + \lambda_0 + 1, \]

• for every \( f : V \to \Sigma \), we have
\[
\frac{d}{d + d'_0 + 1} \cdot \text{gap}_f(G) \leq \text{gap}_f(G') \leq \text{gap}_f(G).
\]

**Proof.** The thing about the gap is obvious: the new constraints are not really constraints, so we only “diluted” the unsatisfiability. The first item is less obvious. Since the adjacency matrix of an undirected graph is symmetric, we have the equality
\[
\lambda_2(G) = \max_{x \in \mathbb{R}^V, x \perp V} \frac{||x^TM(G)x||}{||x||} = \max_{||x||=1, x \perp V} ||x^TMx||.
\]

Note that
\[
M(G') = M(G) + M(H) + I,
\]
so that
\[
||x^TM(G')x|| \leq ||x^TM(G)x|| + ||x^T M(H)x|| + ||x||^2.
\]
Maximizing over \( x \) we obtain
\[
\lambda_2(G') \leq \lambda_2(G) + \lambda_2(H) + 1 < \lambda_2(G) + \lambda_0 + 1 \leq d + \lambda_0 + 1.
\]

Putting together the two constructions, we obtain the preprocessing step:

**Corollary 1.** There are \( 0 < \lambda_0 < d \) such that for every constraint graph \( G \), the graph \( G' := \text{prep}_2(\text{prep}_1(G)) \) is constructible in polynomial time and such that:

• \( G' \) is \( d \)-regular with self-loops, \( \lambda(G') \leq \lambda_0 < d \),

• \( O(1) \cdot \text{gap}(G) \leq \text{gap}(G') \leq \text{gap}(G) \).

## 5 Amplification

### 6 Alphabet Reduction by composition

Suppose that we have a polytime reduction \( P \) from any language \( L \) in NP to the gap constraint satisfaction problem over \( \Sigma_0 \), where \( x \in L \) is mapped to a satisfiable instance and \( x \notin L \) is mapped to an instance with gap at least \( \alpha \), for \( \alpha > 0 \).

We use \( P \) to build a reduction from the gap satisfaction problem over \( \Sigma \) to the gap satisfaction problem over \( \Sigma_0 \) by “composition”, i.e., given an input constraint graph \( G \) over \( \Sigma \), we build a new constraint graph \( G \circ P \) with the right properties.

**Definition 9.** An assignment tester with alphabet \( \Sigma_0 \) and rejection probability \( \epsilon > 0 \) is an algorithm \( P \) whose input is a circuit \( \Phi \) over Boolean variables \( X \), and whose output is a constraint graph \( G = (V,E) \) over \( \Sigma_0 \) and with constraints \( \mathcal{C} \) such that \( V \supseteq X \) and such that the following holds: let \( V' = V \setminus X \) and \( a : X \to \{0,1\} \).

**Completeness** if \( a \) is a valid assignment for \( \Phi \), there exists \( b : V' \to \Sigma_0 \) such that \( \text{gap}_{a \cup b}(G) = 0 \),

**Soundness** if \( a \) is not valid for \( \Phi \), then for all \( b : V' \to \Sigma_0 \) we have \( \text{gap}_{a \cup b}(G) \geq \epsilon \cdot \text{rdist}(a, \text{SAT}(\Phi)) \).
where for \( h, k \in \Sigma^\ell \), we write \( \text{rdist}(h,k) := \frac{\text{dist}(h,k)}{\ell} \).

**Theorem 6.** There exists \( \epsilon > 0 \) and an explicit construction of an assignment tester \( \mathcal{P} \) with alphabet \( \Sigma = \{0,1\}^3 \) and rejection probability \( \epsilon \).

**Proof.**

We now show how to define the composition \( \mathcal{G} \circ \mathcal{P} \). This makes use of error correcting codes (maps \( \Sigma \to \{0,1\}^\ell \)).

**Definition 10.** An error correcting code \( e \) is said to have

- linear dimension if \( \exists C > 0 \) such that \( \ell \leq C \cdot \log(\Sigma) \).
- relative distance \( \rho > 0 \) if for every \( a_1 \neq a_2 \in \Sigma \), we have \( \text{rdist}(e(a_1), e(a_2)) \geq \rho \).

Two \( \ell \)-bits strings \( s_1, s_2 \) are \( \delta \)-far (resp. \( \delta \)-close) if their relative distance is at most (resp. at least) \( \delta \).

**Definition 11.** Let \( \mathcal{G} = (V,E) \) be a constraint graph over \( \Sigma \) with constraints \( C \), and let \( \mathcal{P} \) be an assignment tester, \( \text{ECC}: \Sigma \to \{0,1\}^\ell \) arbitrary encoding with linear dimension and relative distance \( \rho > 0 \). We define \( \mathcal{G} \circ \mathcal{P} \) to be the constraint graph \( (V',E') \) over \( \Sigma_0 \) with constraints \( C' \) as follows:

**Robustization** convert very constraint \( c(e) \in C \) to a circuit \( \tilde{c}(e) \) using \( \text{ECC} \):

- for every variable \( v \in V \), let \( [v] \) be a fresh set of \( \ell \) boolean variables, for each \( e = (v,w) \in E \), \( \tilde{c}(e) \) is a circuit on \( 2\ell \) variables \([v] \cup [w]\) that outputs 1 iff the assignment is the image under \( \text{ECC} \) of a valid assignment for the constraint \( c(e) \).

**Composition** Run \( \mathcal{P} \) on each \( \tilde{c}(e) \), and let \( \mathcal{G}_e \) be the constraint graph \( (V_e,E_e) \) that \( \mathcal{P} \) (over \( \Sigma \)) outputs. Without loss of generality, we can assume that every \( |E_e| \) is the same. The vertex set of \( \mathcal{G} \circ \mathcal{P} \) is \( \bigcup_{e \in E} V_e \), the edge set is \( \bigcup_{e \in E} E_e \), and \( C' = \bigcup_{e \in E} C_e \).

**Lemma 2.** Assume the existence of an assignment tester with constant rejection probability \( \epsilon \) and alphabet \( \Sigma_0 \). There exist a constant \( \beta_3 > 0 \) that depends only on \( \mathcal{P} \) and another constant \( c(\mathcal{P},|\Sigma|) \) such that given any constraint graph \( \mathcal{G} = ((V,E),\Sigma,C) \) we can compute in time linear in the size of \( \mathcal{G} \) the graph \( \mathcal{G}' := \mathcal{G} \circ \mathcal{P} \) such that: the size of \( \mathcal{G}' \) is \( c(\mathcal{P},\Sigma) \cdot |\mathcal{G}| \) and \( \beta_3 \text{gap}(\mathcal{G}) \leq \text{gap}(\mathcal{G}') \leq \text{gap}(\mathcal{G}) \).