A Small Gap in the Gap Amplification of Assignment Testers*

Oded Goldreich† Or Meir‡

28th October 2007

Abstract

An important extension of the proof of the PCP theorem by Irit Dinur (J. ACM 54(3), also ECCC TR05-046) is a gap amplification theorem for Assignment Testers. Specifically, this theorem states that the rejection probability of an Assignment Tester can be amplified by a constant factor, at the expense of increasing the output size of the Assignment Tester by a constant factor. We point out a gap in the proof of this theorem, and show that this gap can be filled.

In this note we discuss a gap in one of the proofs in the work of Dinur [D05, D07], and show how it can be filled. The gap refers to the amplification of Assignment Testers, and the underlying issue does not occur in the case of standard PCPs. We refer both to the journal version of the work [D07] and to the version posted on ECCC [D05], since both of them are cited in the literature.

1 Background

We begin by recalling the definition of [D05, D07] of the notions of Assignment Testers, also known as PCPs of Proximity (see also [BGHSV04, DR06]):

Definition 1 ([D05, Definition 3.1], [D07, Definition 2.8]). An Assignment Tester with alphabet Σ_0 and rejection probability $\varepsilon > 0$ is a polynomial-time transformation $\mathcal{P}$ whose input is a circuit Φ

*This research was partially supported by the Israel Science Foundation (grant No. 460/05).

†Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100 Israel. Email: oded.goldreich@weizmann.ac.il

‡Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100 Israel. Email: or.meir@weizmann.ac.il

over Boolean variables X, and whose output is a constraint graph $G = \langle (V, E), \Sigma_0, \mathcal{C} \rangle$ such that $X \subseteq V$ (where the elements of X are viewed both as variables and as vertices), and such that the following hold. Let $V' = V \backslash X$, and let $a : X \to \{0, 1\}$ be an assignment, then

- (Completeness) If $a \in \mathrm{SAT}(\Phi)$, there exists $b : V' \to \Sigma_0$ such that $\mathrm{UNSAT}_{a \cup b}(G) = 0$.
- (Soundness) If a $a \notin \mathrm{SAT}(\Phi)$ then for all $b : V' \to \Sigma_0$, $\mathrm{UNSAT}_{a \cup b}(G) \geq \varepsilon \cdot \mathrm{dist}(a, \mathrm{SAT}(\Phi))$[1].

The main technical result of [D05, D07] is a gap amplification theorem for PCPs. The following important extension of this theorem to Assignment Testers is also provided in [D05, D07]:

Theorem 2 ([D05, Theorem 8.1], [D07, Theorem 9.1]). *There exists $t \in \mathbb{N}$ such that given an assignment-tester with constant-size alphabet Σ and rejection probability ε, one can construct an assignment-tester with the same alphabet and rejection probability at least* $\min\{2\varepsilon, 1/t\}$*, such that the output size of the new reduction is bounded by at most by a constant factor times the output size of the given reduction.*

The assignment tester of Theorem 2 is constructed in two steps: First, an intermediate assignment tester with alphabet $\Sigma^{d^{t/2}}$ and rejection probability $p = \Omega\left(\min\left\{\sqrt{t} \cdot \varepsilon, 1/t\right\}\right)$ for a constant $d \in \mathbb{N}$ and an arbitrary $t \in \mathbb{N}$ is constructed. Then, a composition theorem of Dinur and Reingold [DR06] is applied to the intermediate assignment tester in order to reduce its alphabet's size, resulting in an assignment tester with alphabet Σ and rejection probability $\Omega(p) = \Omega\left(\min\left\{\sqrt{t} \cdot \varepsilon, 1/t\right\}\right)$. The number t is then fixed to some sufficiently large natural number that yields the desired rejection probability.

The subject of this note is a gap in the first step of the foregoing construction, namely, the construction of the intermediate assignment tester. Specifically, we show that under certain circumstences, the intermediate assignment tester has output size which is quadratic in the output size of the input assignment tester, failing to establish Theorem 2. Such an increase in the output size can not be afforded by the applications of Theorem 2 presented in [D05] and [D07]. We comment that those circumstences do not seem to occur in the applications of Theorem 2 presented in of [D05]. In this note we show that the proof of Theorem 2 can be corrected so the theorem holds under any circumstences.

We recall the way in which the intermediate assignment tester is constructed: Let Φ be a circuit over Boolean variables X.

1. First, the intermediate assignment tester runs the input assignment tester on input Φ, yielding a constraint graph $G = \langle (V, E), \Sigma, \mathcal{C} \rangle$. For any vertex $v \in V$, let $\deg_G(v)$ denote the degree of v in G.

[1]Note that [D07] denotes the relative Hamming distance by rdist, and therefore the foregoing inequality is phrased as $\mathrm{UNSAT}_{a \cup b}(G) \geq \varepsilon \cdot \mathrm{dist}(a, \mathrm{SAT}(\Phi))$.

2. Next, the intermediate assignment tester constructs the constraint graph $H = (\text{prep}(G))^t$. We denote the set of vertices of H by V_H. Recall that prep(G) is the graph in which every vertex v of G is replaced by an expander graph $[v]$ of $\deg_G v$ vertices, whose vertices represent "copies" of v and whose edges correspond to equality constraints. Note that the $X \nsubseteq V_H$, since each $x \in X$ was replaced by $[x]$.

3. Finally, the intermediate assignment tester constructs and outputs a constraint graph H', whose set of vertices is $V_H \cup X$ and whose edges consist of the edges of H and of "consistency edges" that check consistency between V_H and X. The edges are reweighted such that the consistency edges form half of the edges of H'. For every $v \in V_H \cup X$, let $\deg_{H'}(v)$ denote the degree of v in H'.

2 The gap

The gap in the proof arises in the way the consistency edges between X and V_H are defined. Specifically, we show that if the graph G is highly non-regular, the construction of H' may contain too many consistency edges. For simplicity, let us assume that $t = 0$, but note that the argument holds for any value of t. For $t = 0$, it holds that $H = \text{prep}(G)$ and that $V_H = \bigcup_{v \in V} [v]$, where $[v]$ is the set of vertices that represent "copies" of the vertex v of G.

The work of [D05, D07] defined the consistency edges based on a randomized testing procedure. This procedure is given oracle access to an assignment $A : V_H \cup X \to \Sigma$ to H', and is allowed to make two queries to A. The procedure then decides whether to accept or reject A.

The consistency edges are defined using the procedure as follows: For every possible coin tosses ω, let v_1^ω and v_2^ω denote the vertices that the procedure queries on coin tosses ω. For every possible coin tosses ω, a consistency edge is placed between v_1^ω and v_2^ω, and this edge accepts an assignment $A : V_H \cup X \to \Sigma$ if and only if the procedure accepts on coin tosses ω when given oracle access to A. Under the assumption that $t = 0$, the aforementioned procedure is as follows:

1. Select $x \in X$ uniformly at random.

2. Select $z \in [x]$ uniformly at random (recall that $[x]$ is the set of vertices in H that are copies of x).

3. Accept if and only if $A(x) = A(z)$.

Note that for every $x \in X$, the number $\deg_{H'}(x)$ is equal to the number of consistency edges connected to x using the foregoing procedure. The problem is now as follows:

- Since the procedure chooses $x \in X$ uniformly at random (at Step 1), it follows that every variable $x \in X$ must have the same degree in H'. That is, for every two variables $x, y \in X$, it holds that $\deg_{H'}(x) = \deg_{H'}(y)$.
- Since the procedure chooses $z \in [x]$ uniformly at random (at Step 2), every variable $x \in X$ must satisfy $\deg_{H'}(x) \geq |[x]| = \deg_G(x)$.
- Combining the previous two items, it follows that the degree of every variable $x \in X$ is at least $\max_{x \in X} \{\deg_G(x)\}$, and therefore the number of consistency edges added by the foregoing procedure is at least $|X| \cdot \max_{x \in X} \{\deg_G(x)\}$.

Now, suppose that $|X| = \Omega(\text{size}(G))$ and that there exists $x_0 \in X$ for which $\deg_G(x_0) = \Omega(\text{size}(G))$ (observe that this can be the case if G is highly non-regular). In such a case, the number of consistency edges that will be added in the construction of H' will be at least $|X| \cdot \deg_G(x_0) = \Omega\left(\text{size}(G)^2\right)$, and therefore we will have $\text{size}(H') = \Omega\left(\text{size}(G)^2\right)$, contradicting the claim of Theorem 2. Note that this problem does not occur if G is a regular graph, since in such case we have that

$$|X| \cdot \max_{x \in X} \{\deg_G(x)\} = \sum_{x \in X} \deg_G(x) \leq \text{size}(G)$$

and therefore we will have $\text{size}(H') = O(\text{size}(G))$, as required.

3 Filling the gap

We turn to describe how the gap can be filled. In order to fill the gap, we modify the foregoing randomized procedure as follows. For every $x \in [x]$, fix $[x]'$ to be an arbitrary subset of $[x]$ of size $\min\{|[x]|, \text{size}(H)/|X|\}$. The modified procedure is the same as the original procedure, except for that in Step 2, it chooses z uniformly at random from *the set* $[x]'$ *instead of* $[x]$. Observe that this modification indeed solves the problem, since now the degree of every variable $x \in X$ in H' is bounded by $\text{size}(H)/|X|$, and therefore the total number of consistency edges is at most $\text{size}(H) = O(\text{size}(G))$.

The reason that the modified procedure works is roughly as follows: Consider some given assignment to X. Ideally, we would like that if a variable $x \in X$ is assigned a value that is inconsistent with most of $[x]$, then this variable violates $\Omega(1/|X|)$-fraction of the edges of H'. Suppose now that some variable $x \in X$ is assigned a value that is inconsistent with most of the vertices in $[x]$. Then, either that x is inconsistent with most of the set $[x]'$, or most of the set $[x]'$ is inconsistent with most of the set $[x]$. In the first case, at least $\Omega(1/|X|)$-fraction of the edges are violated, since the modified procedure chooses x with probability $1/|X|$ and then chooses with probability at least $\frac{1}{2}$ a vertex $z \in [x]'$ that is inconsistent with x.

The case where x is consistent with most of $[x]'$ is more problematic, since the procedure is likely to choose $z \in [x]'$ that is consistent with x. Not that such a case is only possible if $[x]' \neq [x]$ (since x is inconsistent with most of $[x]$), and therefore the set $[x]'$ is of size at least $s = \text{size}(H) / |X|$. Thus, there is a subset of $[x]$ of size $\Omega(s)$ that is inconsistent with most of $[x]$, and therefore by the mixing properties of the expander $[x]$, about $\Omega(s)$ inner edges of $[x]$ are violated. It follows that the fraction of violated edges is at least

$$\frac{\Omega(s)}{\text{size}(H')} = \frac{\Omega(s)}{O(\text{size}(H))} = \Omega\left(\frac{1}{|X|}\right)$$

as required. Below we give a rigorous proof of this argument.

We describe the modified procedure for an arbitrary value of t (rather than just $t = 0$):

1. Select $x \in X$ uniformly at random.
2. Select $z \in [x]'$ uniformly at random (recall that $[x]'$ is an arbitrary subset of $[x]$ of size $\min\{|[x]|, \text{size}(H) / |X|\}$).
3. Take a $t/2$-step random walk in $\text{prep}(G)$ starting from z, and let w be the endpoint of the walk. Accept if and only if $A(w)_z = A(x)$.

We now use the procedure to define the consistency edges as before, and then reweight the edges of H' such that the consistency edges form half of the edges of H'. It is not hard to see that this modification *solves the problem*: Indeed, this construction requires placing at most $\text{size}(H) / |X|$ consistency edges on H' for every variable in X, which sums up to only $O(\text{size}(H)) = O(\text{size}(G))$ consistency edges.

It remains to show that the intermediate assignment tester that uses the modified randomized procedure has rejection probability $\Omega\left(\min\left\{\sqrt{t} \cdot \varepsilon, 1/t\right\}\right)$. In order to do it, we prove a result analogous to [D05, Lemma 8.2] and [D07, Lemma 9.2]. The reason that we prove again such a result is that [D05, D07] proves the result for her construction of H', while we prove it for the modified version of this construction. The following lemma also differs from [D05, Lemma 8.2] and [D07, Lemma 9.2] in some (hidden) constant factors.

Lemma 3. *Assume that $\varepsilon < 1/t$ and fix an assignment $a : X \to \{0, 1\}$. Then*

- *If $a \in SAT(\Phi)$ then there exists $b : V_H \to \Sigma^{d^t}$ such that $UNSAT_{a \cup b}(H') = 0$.*
- *If $\delta = dist(a, SAT(\Phi)) > 0$ then for every $b : V_H \to \Sigma^{d^t}$ it holds that $UNSAT_{a \cup b}(H') = \Omega(\sqrt{t} \cdot \varepsilon) \cdot \delta$.*

Proof The first item of the lemma can be proved using the same proof as in [D05, D07]. Turning to the second item, assume that $\delta = \mathrm{dist}\,(a, \mathrm{SAT}\,(\Phi)) > 0$ and fix an assignment $b : V_H \to \Sigma^{d^t}$ to H. We prove that $\mathrm{UNSAT}_{a\cup b}\,(H') = \Omega(\sqrt{t}\cdot\varepsilon)\cdot\delta$. As in [D05, D07], let b_1 be the assignment to $\mathrm{prep}\,(G)$ decoded from b using a plurality vote, and let b_0 the assignment to G decoded from b_1 using plurality vote. The case where $\mathrm{dist}\,(b_0|_X, a) \le \delta/2$ can be proved using the same proof as in [D05, D07], which roughly says as follows: If $\mathrm{dist}\,(b_0|_X, a) \le \delta/2$, then using the triangle inequality it can be shown that $\mathrm{dist}\,(b_0|_X, \mathrm{SAT}\,(\Phi)) \ge \delta/2$, and therefore by the definition of G it holds that $\mathrm{UNSAT}_{b_0}\,(G) \ge \varepsilon\cdot\delta/2$. Thus, by the properties of preprocessing and graph powering proved in [D05, D07], it holds that $\mathrm{UNSAT}_b\,(H) = \Omega(\sqrt{t}\cdot\varepsilon)\cdot\delta$. Finally, since the edges of H form half of the edges of H', it follows that $\mathrm{UNSAT}_{a\cup b}\,(H') = \Omega(\sqrt{t}\cdot\varepsilon)\cdot\delta$, as required.

We turn to handle the case where $\mathrm{dist}\,(b_0|_X, a) > \delta/2$. Assume that $\mathrm{dist}\,(b_0|_X, a) > \delta/2$. We prove that in such case it holds that $\mathrm{UNSAT}_{a\cup b}\,(H') = \Omega(\delta)$, which implies the required result. Let b_0' be an assignment for G constructed as follows: For every $v \in V$, the value $b_0'\,(v)$ is decided according to a plurality vote among the values assigned by b_1 to the vertices in $[v]'$, i.e., $b_0'\,(v)$ is the value that maximizes the probability $\Pr_{u\in[v]}\,[b_1(u) = b_0'(v)]$. Recall that, in contrast, b_0 is defined by plurality vote in among $[v]$. We consider two possible cases: $\mathrm{dist}\,(b_0|_X, b_0'|_X) \le \delta/4$ and $\mathrm{dist}\,(b_0|_X, b_0'|_X) > \delta/4$.

- Suppose that $\mathrm{dist}\,(b_0|_X, b_0'|_X) \le \delta/4$. We show that in such case $a\cup b$ violates at least $\delta/16$ of the consistency edges of H', by considering the action of the modified randomized procedure defined above. Using the triangle inequality, it holds that $\mathrm{dist}(b_0'|_X, a) > \delta/4$. It follows that with probability at least $\delta/4$, the procedure chooses in Step 1 a vertex $x \in [x]$ such that $b_0'(x) \ne a(x)$. The value $b_0'(x)$ is defined to be the most popular value assigned by b_1 to the vertices of $[x]'$, and therefore with probability at least $\frac{1}{2}$ the procedure chooses in Step 2 a vertex $z \in [x]$ such that $b_1(z) \ne a(x)$. Similiarly, conditioned on $b_1(z) \ne a(x)$, with probability at least $\frac{1}{2}$ the procedure chooses in Step 3 a vertex w such that $b(w)_z \ne a(x)$. Thus, it follows that in this case the randomized procedure rejects $a \cup b$ with probability at least $\frac{\delta}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} = \delta/16$, and therefore $\mathrm{UNSAT}_{a\cup b}\,(H') = \Omega(\delta)$, as required.

- Suppose that $\mathrm{dist}\,(b_0|_X, b_0'|_X) > \delta/4$. We show that in such case $\mathrm{UNSAT}_b\,(H) = \Omega(\delta)$, due to the violation of the equality constraints of $\mathrm{prep}\,(G)$. Recall that $\mathrm{prep}\,(G)$ is constructed by replacing every vertex v of G with a set of copies $[v]$ of size $\deg_G\,(v)$, placing the edges of an expander on $[v]$ and associating those edges with equality constraints. Observe that the inequality $b_0(x) \ne b_0'(x)$ can only hold for variables $x \in X$ for which $[x]' \ne [x]$, since for other variables x the definitions of $b_0\,(x)$ and $b_0'\,(x)$ coincide. Thus, for every $x \in X$ such that $b_0(x) \ne b_0'(x)$, it holds that $\left|[x]'\right| = \mathrm{size}\,(H)\,/\,|X|$, by definition of $[x]'$.
  Now, observe for every $x \in X$ that satisfies $b_0(x) \ne b_0'(x)$, it holds that $\Omega(\left|[x]'\right|) = \Omega\,(\mathrm{size}\,(H)\,/\,|X|)$ equality edges of $[x]$ are violated by b_1, due to the mixing properties of the expander that was used for the construction of $\mathrm{prep}\,(G)$. It follows that in this case the number of edges of

prep (G) that are violated by b_1 is at least

$$(\mathrm{dist}\,(b_0|_X, b_0'|_X) \cdot |X|) \cdot \Omega\left(\frac{\mathrm{size}\,(H)}{|X|}\right) = \Omega\,(\delta \cdot \mathrm{size}\,(H))$$

The latter equality implies that $\mathrm{UNSAT}_b\,(H) = \Omega(\delta)$, and therefore $\mathrm{UNSAT}_{a\cup b}\,(H') = \Omega\,(\delta)$, as required.

References

[BGHSV04] E. Ben-Sasson, O. Goldreich, P. Harsham, M.Sudan and S. Vadhan, *Robust PCPs of Proximity, Shorter PCPs and Applications to Coding*, SIAM Journal of Computing 36(4), 2006, pages 889-974. Preliminary version in STOC 2004, pages 120-134.

[D05] Irit Dinur, *The PCP theorem by gap amplification*, ECCC TR05-046.

[D07] Irit Dinur, *The PCP theorem by gap amplification*, Journal of ACM 54(3), 2007. Preliminary version in STOC 2006, pp. 241-250.

[DR06] I. Dinur and O. Reingold, *Assignment testers: Towards combinatorial proofs of the PCP theorem*, SIAM Journal of Computing 36(4), 2006, pp. 975-1024. Preliminary version in FOCS 2004, pp. 155-164.