Electronic Colloquium on Computational Complexity, Comment 2 on Report No. 46 (2005)

Gap amplification using lazy random walks

Jaikumar Radhakrishnan
jaikumar@tti-c.org

November 2, 2005

Abstract

This note is based on the original version of Irit Dinur's paper (ECCC TR05-046). It contains two suggestions concerning the product construction. First, instead of using paths of a fixed length t, one can use paths with varying lengths in order to simplify some calculations. Second, one can view Proposition 2.4 as guaranteeing some sort of pairwise independence, and use the second-moment method instead of explicitly bounding the overcount (as in Lemma 5.3).

The product graph: The vertices of G^t are the same as the vertices of G. The alphabet of G^t is Σ^{d^t}, where every vertex specifies values (its opinion) for all vertices reachable within t steps. We will describe the constraints of G^t indirectly using a verifier who probes two assignments.

> The verifier picks a random vertex a in $V(G)$. She then performs a random walk in G, but *after* each step she stops with probability $\frac{1}{t}$. Let the sequence of vertices generated by the walk be
> $$a = v_0, v_1, \ldots, v_T = b,$$
> where $T \geq 1$ is a random variable with mean t. If for some edge $e = (v_i, v_{i+1})$, vertex a has an opinion for v_i and b has an opinion for v_{i+1}, and these opinions violate the constraint c^e, then reject. Otherwise, accept.

Remark: At present there is no bound on the length of the walks. We will argue that the required amplification can be achieved even if the verifier's walks are truncated to $T^* = \lceil 5t \ln |\Sigma| \rceil$, and she just accepts if her walk has not stopped within these many steps. Then, the number of edges in G^t is at most $2|V(G)| \cdot (dt)^{T*}$. (We can label edges coming out of any vertex by distinct sequences of length T^* from the set $[d] \times [t]$). With this modification, we now have the following variant of Lemma 3.4.

Lemma 0.1 *Let $\lambda < d$, d and $|\Sigma|$ be arbitrary constants. There exists a constant $\beta = \beta(\lambda, d) > 0$ such that for every d-regular constraint graph $G = \langle V, E, \Sigma, \mathcal{C} \rangle$ with self-loops with $\lambda(G) \leq \lambda$*

$$\overline{\text{SAT}}(G^t) \geq \frac{\beta}{|\Sigma|^4} \cdot t \cdot \min\left\{\overline{\text{SAT}}(G), \frac{1}{t}\right\}.$$

Fact about the random walk: We will need the following fact about the lazy random walk (without truncation) described above. Consider the verifier's walks conditioned on the event that it uses (u, v) exactly k times (for some $k \geq 1$), that is, the number of i's for which $(v_i, v_{i+1}) = (u, v)$ is exactly k.

ISSN 1433-8092

[F1] Let the starting vertex of the random walk be a and the ending vertex of this walk be b (these are random variables). Then, a and b are independent. Furthermore, a has the same distribution as the random vertex obtained by the following random process.

> Start the random walk at u, but stop with probability $\frac{1}{t}$ *before* making each move (so we stop at u itself with probability $\frac{1}{t}$). Output the final vertex.

Note that the above random walk could use the move (v, u) any number of times, but the distribution on the destination does not depend on the number of times (v, u) is used; in particular, it is the same if we condition on the event that the walk does not use (v, u) even once. Similarly, we can generate the distribution for endpoint b by a random walk starting from v, stopping with probability $\frac{1}{t}$ before each step.

The new assignment for G: Fix an assignment $\vec{\sigma}$ for G^t. Motivated by Fact [F1], we will construct the new assignment σ for G. To obtain $\sigma(u)$, we use the random walk starting at u mentioned above (stopping with probability $\frac{1}{t}$ before each step). This generates a distribution on the vertices of G. Restrict attention to those vertices that have an opinion for u. For each letter in the alphabet determine the probability (under this distribution) that the vertex assigns that letter to u. Then, pick the letter that has the highest probability. With probability at least $1 - (1 - \frac{1}{t})^t \geq \frac{1}{2}$, the walk stops within t steps, so the vertices that do have an opinion have total probability at least $\frac{1}{2}$. Thus, with probability at least

$$\frac{1}{2|\Sigma|}. \tag{1}$$

the random walk from u will reach a vertex whose opinion on u is the same as $\sigma(u)$.

Notation: Let F be the set of edges of G that are not satisfied by this assignment σ. From now on, when we say an edge (u, v) we mean the ordered pair; so we think of F and E as sets of ordered pairs. Each undirected edge will give rise to two edges in E (but self-loops will appear only once). As in the paper, we will assume that $\frac{|F|}{|E|} \leq \frac{1}{t}$. We will refer to (v_i, v_{i+1}) as the i-th edge of the walk. We care about the orientation, so unless this is a self-loop, (v_{i+1}, v_i) is *not* the i-th edge of the walk. We say that the i-th edge (say $(v_i, v_{i+1}) = (u, v)$) of the verifier's walk is *faulty* if

1. $(u, v) \in F$ and
2. the starting vertex of the walk a has an opinion for v_i and it agrees with $\sigma(u)$ and
3. the final vertex b has an opinion for v and it agrees with $\sigma(v)$.

Clearly, the verifier rejects whenever she sees a faulty edge on her walk. *If a walk has been truncated no edge on it is considered faulty.*

Goal: We wish to show that $\Pr[\text{Verifier rejects}] \geq \dfrac{\beta}{|\Sigma|^4} t \cdot \dfrac{|F|}{|E|}$.

Amplification: Let us first see how the calculation works when we don't truncate the walk. Consider the random walk chosen by the verifier. Say, the starting vertex is a and the final vertex is b. Let the random variable N_F denote the number of faulty edges on the walk (if the same edge (u, v) appears faulty several times, each occurrence contributes once to N_F). We wish to show that the probability that $N_F > 0$ is large. We have two claims.

Claim 1: $\mathbf{E}[N_F] \geq t\frac{|F|}{|E|}\frac{1}{4|\Sigma|^2}$.

Claim 2: $\mathbf{E}[N_F^2] \leq Ct\frac{|F|}{|E|}$, where C is a constant depending on the d and λ of the constraint graph G.

Our goal follows from Claims 1 and 2, by using the inequality (see, e.g., Alon & Spencer (2000), Section 4.8, Exercise 1)

$$\Pr[N_F > 0] \geq \frac{\mathbf{E}[N_F]^2}{\mathbf{E}[N_F^2]}.$$

Proof of Claim 1: We will estimate the expected number of faulty occurrences for each in F. Fix one such edge $e = (u, v)$. The expected number of occurrences of (u, v) in the walk is exactly $t/|E|$. Condition on the event that the walk has exactly k occurrences of (u, v) in it. Now, the starting vertex a can be generated by the same process that was used to construct σ (by Fact [F1]). The final vertex b can also be similarly generated. Further, a and b are independent. Thus, with probability at least $\frac{1}{4|\Sigma|^2}$ (see (1) above) both a and b have opinions that agree with σ for u and v respectively. In this case, (u, v) is declared faulty. So, overall, the expected number of faulty occurrences of (u, v) in the verifier's walk is at least

$$\frac{t}{|E|} \cdot \frac{1}{4|\Sigma|^2}.$$

Claim 1 now follows using linearity of expectation by summing over all $|F|$ possibilities for (u, v).

Proof of Claim 2: Let χ_i be the random variable indicating whether the i-th edge $(i = 0, 1, \ldots)$ of the walk is in F. Then, $N_F \leq \sum_i \chi_i$. Also, $\Pr[\chi_i = 1] = \frac{|F|}{|E|}\left(1 - \frac{1}{t}\right)^i$. Furthermore,

$$\begin{aligned}
\mathbf{E}[N_F^2] &\leq 2 \sum_{0 \leq i \leq j < \infty} \mathbf{E}[\chi_i \chi_j] \\
&\leq 2 \sum_i \Pr[\chi_i = 1] \sum_{j \geq i} \Pr[\chi_j = 1 \mid \chi_i = 1] \\
&\leq 2 \sum_i \Pr[\chi_i = 1] \left[1 + \sum_{\ell \geq 1} \left(1 - \frac{1}{t}\right)^\ell \left(\frac{|F|}{|E|} + (\frac{\lambda}{d})^{\ell-1}\right)\right] \qquad (2) \\
&\leq 2 \sum_i \Pr[\chi_i = 1] \left[1 + \sum_{\ell \geq 1} \left(1 - \frac{1}{t}\right)^\ell \frac{|F|}{|E|} + (\frac{\lambda}{d})^{\ell-1}\right] \\
&\leq 2t\frac{|F|}{|E|}\left(t\frac{|F|}{|E|} + c\right).
\end{aligned}$$

We use Proposition 2.4 to justify (2). The claim follows because we have assumed that $\frac{|F|}{|E|} \leq \frac{1}{t}$.

Verifier with truncated walks: When the verifier uses truncated walks, we only need to redo Claim 1, removing the contribution to $\mathbf{E}[N_F]$ from long walks. The contribution to $\mathbf{E}[N_F]$ from walks of length ℓ is at most $\ell|F|/|E|$ times the probability that $T = \ell$. So, the contribution to $\mathbf{E}[N_F]$ from walks of length at least $k + 1$ is

$$\Pr[T \geq k + 1] \cdot \mathbf{E}[T \mid T \geq k + 1] \cdot \frac{|F|}{|E|}.$$

The first factor is at most $(1-\frac{1}{t})^k$, the second is $k+t$. So, the quantity we need to remove from our previous lower bound for $\mathbf{E}[N_F]$ (in Claim 1) is at most

$$\exp(-\frac{k}{t})(k+t)\cdot\frac{|F|}{|E|}.$$

We let $k = \lceil 5t\ln|\Sigma|\rceil$, so that this quantity is much less than $t\cdot\frac{|F|}{|E|}\cdot\frac{1}{8|\Sigma|^2}$. Thus, we have the following.

Revised Claim 1: Let N'_F be the random variable that counts the number of faulty edges on the walk when the verifier truncates the walks at $\lceil 5t\ln|\Sigma|\rceil$. Then,

$$\mathbf{E}[N'_F] \geq t\cdot\frac{|F|}{|E|}\cdot\frac{1}{8|\Sigma|^2}.$$

Note that $N'_F \leq N_F$, so Claim 2 applies to N'_F as well, and the goal is established as before using Chebyshev's inequality.

Acknowledgments. Thanks to Eli Ben-Sasson, Irit Dinur, Prahladh Harsha, Adam Kalai and Nanda Raghunathan for their comments.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'