Derandomization by random walks

Kent Quanrud

November 10, 2020

1 Introduction

Recall that a language L is in the **class P** if there is a deterministic poly(n)-time algorithm that decides if an input x of size n is in L.

Definition 1. *A language L is in the* **class RP** *if there is a randomized polynomial time algorithm deciding L with the following probabilistic "one-sided" error guarantee:*

1. *Given an input $x \in L$, the algorithm decides that $x \in L$ with constant probability (say, $1/2$).*
2. *Given an input $x \notin L$, the algorithm always decides that $x \notin L$.*

A language L is in the **class BPP** *if there is a randomized polynomial time algorithm deciding L with the following probabilistic "two-sided" error guarantee:*

1. *Given an input $x \in L$, the algorithm decides that $x \in L$ with probability $2/3$.*
2. *Given an input $x \notin L$, the algorithm decides that $x \notin L$ with probability $2/3$.*

We have the containments

$$\mathbf{P} \subseteq \mathbf{RP} \subseteq \mathbf{BPP},$$

since of course, no error (**P**) is better than one-sided error (**RP**), which is better than two-sided error (**BPP**). It is a major open question if these are equal. It is often believed that $\mathbf{P} = \mathbf{BPP}$. More functionally, many people are believe that a randomized algorithm with two-sided error is a strong indicator that we should be able to find a deterministic one as well. Right now we do not know if **P** equals **RP** or if **RP** equals **BPP**.

There is a interest in the theory community, sometimes under the heading of *pseudorandomness*, in a refined understanding of how much randomness is required for certain problems. While the holy grail is **P** vs **BPP** is hard to attack, there is a rich body of literature or results moving towards this conclusion, that has also produced many algorithmic ideas of independent interest.

Let $L \in \mathbf{RP}$, and fix an input size n. Suppose that an algorithm for L requires m random bits to decided L with one-sided error $1/2$. If we want to decrease the error to δ, for some δ, then we could independently repeat the algorithm $\lceil \log 1/\delta \rceil$ times, taking the disjunction of responses. Often times we deemphasize the logarithmic overhead incurred from repetition, particularly since it is easily made up for by the convenience of randomization. Still it is a natural and profound question to ask if one can do better. Surprisingly, one *can*.

Theorem 2. *Given an algorithm in **RP** that uses m random bits to achieve error $1/2$, one can decrease the error probability to δ while increasing the running time by a factor of $\log(1/\delta)$ and using a* total *of $m + O(\log(1/\delta))$ random bits.*

We can do the same for algorithms in **BPP**. The algorithm will be the same as for **RP**, though the analysis is different.

Theorem 3. *Given an algorithm in **BPP** that uses m random bits to achieve error $1/3$ (on both sides), one can decrease the error probability to δ while increasing the running time by a factor of $\log(1/\delta)$ and using a* total *of $m + O(\log(1/\delta))$ random bits.*

The algorithm obtain the better-than-repitition amplification above is conceptually very clean. Let G be a constant degree expander with vertex set $V = \{0,1\}^m$ - that is, a vertex for every possible bit string of m bits. (We will have to address how to build such a G later, but for now let us assume it exists.) Start from a uniformly random vertex $v_0 \in V$. (This takes m random bits). Then take a random walk in G for $O(\log(1/\delta))$ steps. For each vertex v_i we visit, use v_i as the input for a new instance of the algorithm. For algorithms in **RP**, take the disjunction (the "or") of all the outputs. For algorithms in **BPP**, take the mean.

The above construction suggests that random bit strings generated by expanders are almost as good as completely random bit strings. We have two basic issues to address.

1. We need to show that the bit strings generated by an expander are useful, for both the **RP** and **BPP** case.

2. We need to show how to efficiently make such a graph G.

Note that for the second point, since G is exponentially sized, this construction has to be implicit.

Let us first consider the first point - why bit strings generated by an expander graph on bit strings seems to be almost as good as totally random bit strings. We will have different lemma's for either case. The first lemma is for amplifying **RP**.

Lemma 4. *Let $G = (V, E)$ be a regular undirected graph whose random walk has spectral gap γ. Consider a t-step random walk $v_1, v_2, \ldots, v_t \in V$ where $v_1 \in V$ is chosen uniformly at random. For any set $B \subset V$, the probability that the entire random walk stays in B is*

$$(\mu + (1-\gamma)(1-\mu))^t,$$

where $\mu = |B|/|V|$.

Consider again Theorem **??** for amplifying **RP** with expanders. We apply the above lemma to the expander graph of bit strings where we set B to be the set of all random strings causing the algorithm to err. Amplifying the original algorithm a constant number of times at needed, we can make μ arbitrarily small; say, $1/2$. By the above lemma the probability that *all* the bit strings cause error at a rate or $(1 - \gamma/2)^t$.

We now present the lemma that is important for **BPP**.

Lemma 5. *Let $\epsilon \in (0,1)$ be fixed. Let $G = (V, E)$ be a regular undirected graph whose random walk has spectral gap $\gamma \geq 1 - \epsilon$. Let $f : V \to [0,1]$ be a fixed function of the vertices. Let*

$$\mu = \mathrm{E}[f(v)] \text{ where } v \in V \text{ uniformly at random.}$$

Consider a random walk $v_1, v_2, \ldots, v_k \in V$ where $v_1 \in V$ is chosen uniformly at random. Then for all $\beta > 0$,

$$\mathrm{P}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu\right| \geq \epsilon\mu + \beta\right] \leq ce^{\epsilon k((1+\epsilon)\epsilon - \beta)}$$

for a universal constant $c > 0$. In particular, for (say) $\epsilon \le \mu/4$, we have

$$\mathrm{P}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu\right| \ge \epsilon\mu\right] \le ce^{-\frac{\epsilon k\mu}{c}}$$

for a universal constant $c > 0$.

To *make* the expander G, we apply the following theorem, which is a tweak on our previous construction for derandomizing random walks.

Lemma 6. *Let H be a graph with d^8 vertice, degree d, and spectral gap $\ge 7/8$. Define graphs $G_1, G_2, \ldots$ by*

$$G_1 = H^2$$
$$G_{t+1} = \mathcal{Z}\Big((G_t \otimes G_t)^2 \,\Big|\, H\Big)$$

Let n_t be the number of vertices in G_t. Then $n_t \approx c^{2^t}$, and simulating one step of a random walk in G_t takes $\mathrm{poly}(\log(n_t))$ time.

2 Amplifying RP

Lemma 4. *Let $G = (V, E)$ be a regular undirected graph whose random walk has spectral gap γ. Consider a t-step random walk $v_1, v_2, \ldots, v_t \in V$ where $v_1 \in V$ is chosen uniformly at random. For any set $B \subset V$, the probability that the entire random walk stays in B is*

$$(\mu + (1-\gamma)(1-\mu))^t,$$

where $\mu = |B|/|V|$.

Proof. Let R be the random walk matrix on G. Let $P : \mathbb{R}^V \to \mathbb{R}^V$ be the projection subspace spanned by $\mathbb{R}^S$; that is,

$$(Px)_v = \begin{cases} x_v & \text{if } v \in S \\ 0 & \text{otherwise.} \end{cases}$$

Note that P is a linear function with $P = P^T$ and $P^2 = P$. Consider the product PRP. Given a nonnegative vector x, we drop all the mass outside of B, take a step according to R, and again drop all the probability mass. In particular, the probability that the entire walk stays in B is

$$\big\langle \mathbb{1}, (PRP)^t(\mathbb{1}/n)\big\rangle.$$

We claim that PRP has maximum eigenvalue $\le \mu + 1 - \gamma$. If so, then

$$\big\langle \mathbb{1}, (PRP)^t(\mathbb{1}/n)\big\rangle \le (\gamma\mu + 1 - \gamma)$$

Recall that

$$R = \frac{\gamma}{n}(\mathbb{1} \otimes \mathbb{1}) + R'$$

where R' has all of its eigenvalues in the range $[1-\gamma, \gamma-1]$. Then

$$PRP = \frac{\gamma}{n} P(\mathbb{1} \otimes \mathbb{1})P + PR'P.$$

$PR'P$ has all its eigenvalues in the range $[1-\gamma, \gamma-1]$ too. Consider the first term. We claim that it has maximum eigenvalue $\gamma\mu = |S|/n$, which would give the overall bound of

$$\gamma\mu + 1 - \gamma = \mu + (1-\mu)(1-\gamma)$$

that we seek. To this end, we have

$$\frac{\gamma}{n} P(\mathbb{1} \otimes \mathbb{1})P = \frac{\gamma}{n}(P\mathbb{1} \otimes P\mathbb{1})$$

and the maximum eigenvector of the outer product $(P\mathbb{1} \otimes P\mathbb{1})$ is (proportional to) $P\mathbb{1}$ with eigenvalue

$$\frac{\langle P\mathbb{1}, P\mathbb{1}\rangle^2}{\langle P\mathbb{1}, P\mathbb{1}\rangle} = \langle P\mathbb{1}, P\mathbb{1}\rangle = |S|.$$

This gives the desired bound. ■

3 Efficiently amplifying BPP

Lemma 5. *Let $\epsilon \in (0,1)$ be fixed. Let $G = (V, E)$ be a regular undirected graph whose random walk has spectral gap $\gamma \geq 1 - \epsilon$. Let $f : V \to [0,1]$ be a fixed function of the vertices. Let*

$$\mu = \mathrm{E}[f(v)] \text{ where } v \in V \text{ uniformly at random.}$$

Consider a random walk $v_1, v_2, \ldots, v_k \in V$ where $v_1 \in V$ is chosen uniformly at random. Then for all $\beta > 0$,

$$\mathrm{P}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu\right| \geq \epsilon\mu + \beta\right] \leq ce^{\epsilon k((1+\epsilon)\epsilon - \beta)}$$

for a universal constant $c > 0$. In particular, for (say) $\epsilon \leq \mu/4$, we have

$$\mathrm{P}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu\right| \geq \epsilon\mu\right] \leq ce^{-\frac{\epsilon k\mu}{c}}$$

for a universal constant $c > 0$.

Proof. Initially the proof proceeds similarly to the Chernoff bound. We let us prove the inequality on the upper tail. The lower tail follows similarly, and then we can take the the union bound over both. We have

$$\mathrm{P}\left[\sum_{i=1}^{k} f(v_i) \geq k\mu + \beta\right] \leq \mathrm{E}\left[e^{t(f(v_1)+\cdots+f(v_k))}\right] e^{-\epsilon k\mu - \epsilon\beta}. \tag{1}$$

The key identity is

$$\mathrm{E}\left[e^{\epsilon(f(v_1)+\cdots+f(v_k))}\right] = \frac{1}{n}\left\langle \mathbb{1}, F(RF)^k\mathbb{1}\right\rangle \text{ where } F = \mathrm{diag}\left(e^{\epsilon f(v_1)}, \ldots, e^{\epsilon f(v_n)}\right)$$

is the diagonal matrix with the exponentiated values along the diagonal. One way to interpret the above is to first recall that R^k models k steps of a random walk. Then inserting F in between the R's is like collecting the values $e^{\epsilon f(v)}$ along the walk. We claim the following.

Claim 1. $\frac{1}{n}\left\langle F^{1/2}\mathbb{1}, F(RF)^k F^{1/2}\mathbb{1}\right\rangle \leq e^{\epsilon + (k(1+\epsilon)\epsilon(\mu+\epsilon))}$

Assuming claim 1 holds, we have

$$(1) \leq e^{\epsilon} \cdot e^{\epsilon k((1+\epsilon)\epsilon - \beta)},$$

as desired.

Let us now prove Claim 1. Since R has spectral gap $\geq 1 - \epsilon$, and R is symmetric, we can pull out a $(1-\epsilon)$-fraction of its leading eigenvector (corresponding to the uniform distribution), writing

$$R = \frac{1-\epsilon}{n}(\mathbb{1} \otimes \mathbb{1}) + R'$$

where R' has eigenvalues between ϵ and $-\epsilon$. Thereby

$$F^{1/2}RF^{1/2} = \frac{1-\epsilon}{n}F^{1/2}(\mathbb{1} \otimes \mathbb{1})F^{1/2} + F^{1/2}R'F^{1/2}.$$

We claim the following.

Claim 2. $F^{1/2}R'F^{1/2}$ *has its eigenvalues between* $[\epsilon e^\epsilon, -\epsilon e^\epsilon]$.

Claim 3. $F^{1/2}(\mathbb{1}\otimes\mathbb{1})F^{1/2}$ *has maximum eigenvalue* $\mathrm{E}\left[e^{\epsilon f(v)}\right]$.

For the first claim regarding $F^{1/2}R'F^{1/2}$, for any vector x

$$\begin{aligned}\left\langle x, F^{1/2}R'F^{1/2}x\right\rangle &= \left\langle F^{1/2}x, R'\left(F^{1/2}x\right)\right\rangle \le \epsilon\left\|F^{1/2}x\right\|^2 \\ &= \epsilon\langle x, Fx\rangle \le \epsilon e^t\|x\|^2,\end{aligned}$$

where we repeatedly invoke the fact that the maximum eigenvalue of a symmetric matrix is given by the Rayleigh quotient.

For the second claim, we have

$$F^{1/2}(\mathbb{1}\otimes\mathbb{1})F^{1/2} = \left(F^{1/2}\mathbb{1}\otimes F^{1/2}\mathbb{1}\right)$$

which has maximum eigenvalue

$$\left\|F^{1/2}\mathbb{1}\right\|^2 = \langle\mathbb{1}, F\mathbb{1}\rangle = \sum_v e^{\epsilon f(v)} = n\,\mathrm{E}\left[e^{\epsilon f(v)}\right].$$

This establishes the second claim.

Combining the two claims above, we have that $F^{1/2}RF^{1/2}$ has maximum (absolute) eigenvalue

$$(1-\epsilon)\,\mathrm{E}\left[e^{\epsilon f(x)}\right] + \epsilon e^\epsilon.$$

Thus $F^{1/2}RF^{1/2}$ has maximum eigenvalue at most

$$\left\|F^{1/2}RF^{1/2}\right\| \le \epsilon e^\epsilon + (1-\epsilon)\,\mathrm{E}\left[e^{\epsilon f(v)}\right].$$

We expand the right hand side by the inequality $e^\epsilon \le 1+\epsilon+\epsilon^2$ for small $|t|$, giving the upper bound

$$\begin{aligned}\left\|F^{1/2}RF^{1/2}\right\| &\le \epsilon\left(1+\epsilon+\epsilon^2\right) + (1-\epsilon)\,\mathrm{E}\left[1+\epsilon f(v)+\epsilon^2 f(v)\right] \\ &= \epsilon\left(1+\epsilon+\epsilon^2\right) + (1-\epsilon)\left(1+\epsilon\mu+\epsilon^2\mu\right) \\ &= 1+(1+t)\epsilon(\mu+\epsilon) \\ &\le e^{(1+\epsilon)\epsilon(\mu+\epsilon)}\end{aligned}$$

In turn, for the kth power, we have

$$\left\|\left(F^{1/2}RF^{1/2}\right)^k\right\| = \left\|F^{1/2}RF^{1/2}\right\|^k \le e^{k(1+\epsilon)\epsilon(\mu+\epsilon)}.$$

Finally, returning to the original quantity we wanted to sum, we have

$$\frac{1}{n}\left\langle F^{1/2}\mathbb{1}, F(RF)^kF^{1/2}\mathbb{1}\right\rangle \le e^{k(1+t)\epsilon(\mu+\epsilon)}\,\mathrm{E}\left[e^{\epsilon f(v)}\right] \le (1+O(\epsilon))e^{(k(1+\epsilon)\epsilon(\mu+\epsilon))}.$$

as desired for Claim 1. ■

4 Efficiently making large expanders

It remains to be shown that large expanders can be constructed efficiently. Previously, in the interest of deterministic connectivity, we studied the amplification

$$G \mapsto \mathcal{Z}\big(G^2 \,\big|\, H\big),$$

where the degrees and sizes of G and H are appropriately set. The primary goal of the goal of that exercise was to increase the spectral gap given an input graph (with bad spectral gap) in a space efficient manner.

Here our goal is slightly different, because simply want to make an expander over 2^m vertices without the burden of some bad input graph. That is, we simply want to make a large - very, very large - expander. Note that we want to make this graph implicitly and be able to take a step in the graph in $O(\text{polylog}(m))$ time per step – importantly, this is *doubly* logarithmic in the number of vertices, 2^m. If we apply the construction form connectivity starting from a constant sized expander, we will end up needing $O(m)$ iterations to get up to 2^m vertices, since each iteration increases the number of vertices of a constant factor. Thus, in contrast to before, the goal is to *increase the number of vertices given an expander* as efficiently as possible.

Let us now restate the main lemma that we need to prove.

Lemma 6. *Let H be a graph with d^8 vertice, degree d, and spectral gap $\geq 7/8$. Define graphs $G_1, G_2, \ldots$ by*

$$G_1 = H^2$$
$$G_{t+1} = \mathcal{Z}\Big((G_t \otimes G_t)^2 \,\Big|\, H\Big)$$

Let n_t be the number of vertices in G_t. Then $n_t \approx c^{2^t}$, and simulating one step of a random walk in G_t takes $\text{poly}(\log(n_t))$ time.

We first recall the first two lemma's that we proved previously.

Lemma 7. *Let $G = (V, E)$ be a regular undirected graph n vertices and degree d. Then G^k is a regular undirected graph on V with degree d^k, with random walk map R^k. If R has spectral gap γ, then R^k has spectral gap $1 - (1-\gamma)^k$.*

Lemma 8. *Let $G = (V, E)$ be a regular undirected graph with n vertices and degree d, with spectral gap γ_G. Let H be a regular undirected graph with d vertices and degree d_0. Then $\mathcal{Z}(G \,|\, H)$ is a regular undirected graph with nd vertices, degree d_0^2 and spectral gap $\gamma_G \gamma_H^2$.*

The second lemma, regarding the zig-zag product, required the following structural lemma about the tensor product of undirected graphs and their random walks.

Lemma 9. *Let G_1 and G_2 be regular undirected graphs with degrees d_1 and d_2 and random walk matrices R_1 and R_2 respectively. Then $G_1 \otimes G_2$ is a regular undirected graph with degree $d_1 d_2$, Then the random walk matrix of $G_1 \otimes G_2$, denoted $R_1 \otimes R_2 : \mathbb{R}^{V_1 \times V_2} \to \mathbb{R}^{V_1 \times V_2}$, is also symmetric. The map*

$$(v_1, v_2) \in \mathbb{R}^{V_1} \times \mathbb{R}^{V_2} \mapsto v_1 \otimes v_2 \in \mathbb{R}^{V_1 \times V_2}$$

gives a one-to-one correspondance between pairs of eigenvectors from G_1 to G_2, where an eigenvector v_1 with eigenvalue λ_1 of G_1 and an eigenvector v_2 with eigenvalue λ_2 of G_2 maps to an eigenvector $v_1 \otimes v_2$ of $G_1 \otimes G_2$ with eigenvalue $\lambda_1 \lambda_2$.

Let $d \in \mathbb{N}$ be a fixed constant and let H be an undirected regular graph d^4 vertices, degree d, and spectral gap $7/8$. Let $G_0 = H^2$. We now generate graphs $G_1, G_2, \ldots$ iteratively by

$$G_{i+1} = \mathcal{Z}\Big((G_i \otimes G_i)^2 \,\Big|\, H\Big).$$

The various parameters of interest develop as follows.

Graph	G	$\to$	$G \otimes G$	$\to$	$(G \otimes G)^2$	$\to$	$\mathcal{Z}\Big((G \otimes G)^2 \,\Big|\, H\Big)$
Vertices	n	$\to$	n^2	$\to$	n^2	$\to$	$n^2 d^4$
Degree	d^2	$\to$	d^4	$\to$	d^8	$\to$	d^2
$\boldsymbol{\gamma}$	γ	$\to$	γ	$\to$	$2\gamma - \gamma^2$	$\to$	$\big(2\gamma - \gamma^2\big)(7/8)^2$

We note that $\gamma \geq 1/2$ implies $\big(2\gamma - \gamma^2\big) \geq 1/2$, so the spectral gap never drops below $1/2$.

References

[1] Omer Reingold, Salil Vadhan, and Avi Wigderson. "Entropy Waves, the Zig-Zag Graph Product, and New Constant-Degree Expanders". In: *Annals of Mathematics*. Second Series 155.1 (2002), pp. 157–187. URL: `http://www.jstor.org/stable/3062153`.

[2] Salil P. Vadhan. "Pseudorandomness". In: *Found. Trends Theor. Comput. Sci.* 7.1-3 (2012), pp. 1–336.