

PSRGs via Random Walks on Graphs

Daniel A. Spielman

October 3, 2012

11.1 Overview

There has been a lot of work on the design of Pseudo-Random Number Generators (PSRGs) that can be proved to work for particular applications. In this class, we will see one of the foundational results in this area, namely Impagliazzo and Zuckerman's [IZ89] use of random walks on expanders to run the same algorithm many times. We are going to perform a very crude analysis that is easy to present in class. Rest assured that much tighter analyses are possible and much better PSRGs have been constructed since.

11.2 Expander Graphs

Expander graphs are an incredibly useful family of graphs that we will study quite a bit in the next two weeks. An expander graph is a graph with constant vertex degree and constant conductance. To define that formally, one really needs to talk about a sequence of graphs of increasing numbers of vertices. We will call such a sequence an *expander family* if there exist constants d and ϕ such that every graph in the family has degree d and conductance at least ϕ . Random d -regular graphs are expanders with high probability [Bol88, Fri08].

In these lectures, we will use the spectral characterization of expanders. We will consider a family of graphs to be expanders if

$$|\lambda_i - d| < \epsilon d, \quad \forall i \geq 2$$

for some constant ϵ . In the adjacency eigenvalues, this is equivalent to

$$|\mu_i| < \epsilon d, \quad \forall i \geq 2.$$

Cheeger's inequality provides a rough equivalence between these two definitions.

Explicit constructions of such expanders are known. Spectrally speaking, the best expanders are those produced by Margulis [Mar88] and Lubotzky, Phillips and Sarnak [LPS88]. These expanders are extremely explicit: given the name of a vertex v , one can produce a list of the neighbors of v in time proportional to the number of bits used to write v (see also the work of Alon, Bruck, Naor, Naor and Roth [ABN⁺92] for even more explicit constructions). This allows one to take a step in a random walk on a graph with 2^n vertices in time polynomial in n . That is, we can study random walks on such graphs without having to write down the entire graph.

11.3 Why Study PSRGs?

Pseudo-random number generators take a seed which is presumably random (or which has a lot of randomness in it), and then generate a long string of random bits that are supposed to act random. We should first discuss why we would actually want such a thing. I can think of two reasons.

1. Random bits are scarce. This might be surprising. After all, if you look at the last few bits of the time that I last hit a key, it is pretty random. Similarly, the low-order bits of the temperature of the processor in my computer seem pretty random. While these bits are pretty random, there are not too many of them.

Many randomized algorithms need *a lot* of random bits. Sources such as these just do not produce random bits with a frequency sufficient for many applications.

2. If you want to re-run an algorithm, say to de-bug it, it is very convenient to be able to use the same set of random bits by re-running the PSRG with the same seed. If you use truly random bits, you can't do this.

You may also wonder how good the standard pseudo-random number generators are. The first answer is that the default ones, such as `rand` in C, are usually terrible. There are many applications, such as those in my thesis, for which these generators produce behavior that is very different from what one would expect from truly random bits (yes, this is personal). On the other hand, one can use cryptographic functions to create bits that will act random for most purposes, unless one can break the underlying cryptography [HILL99]. But, the resulting generators are usually much slower than the fastest pseudo-random generators. Fundamentally, it comes down to a time-versus-quality tradeoff. The longer you are willing to wait, the better the pseudo-random bits you can get.

11.4 Today's Application : repeating an experiment

Today, we are going to consider a very simple situation in which one wants to run the same randomized algorithm many times. Imagine that you have a randomized algorithm A , that returns “yes”/“no” answers, and which is usually correct. In particular, imagine that you can prove that it is correct at least 99% of the time, regardless of the input. Sometimes 99% accuracy is enough, but sometimes it is not. We can boost the accuracy by running the algorithm a few times, and taking the majority vote of the answers we get back¹. But, this seems to require a new set of random bits for each run. In this lecture, we will prove that 9 new bits per run suffice.

Since we will not make any assumptions about the algorithm, we will use truly random bits the first time we run it. But, we will show that we only need 9 new random bits for each successive run. In particular, we will show that if we use our PSRG to generate bits for $t + 1$ runs, then the probability that majority answer is wrong decreases exponentially in t .

¹Check for yourself that running it twice doesn't help

11.5 The Random Walk Generator

Let r be the number of bits that our algorithm needs for each run. So, the space of random bits is $\{0, 1\}^r$. For some input which is fixed throughout this lecture, let $X \subset \{0, 1\}^r$ be the settings of the random bits on which the algorithm gives the wrong answer, and let Y be the settings on which it gives the correct answer. These sets will typically differ for different inputs.

Our pseudo-random generator will use a random walk on an expander graph whose vertex set is $\{0, 1\}^r$. The expander graph will be d -regular, for some constant d . Let \mathbf{A} be the adjacency matrix of the graph, and let its eigenvalues be $d = \mu_1 > \mu_2 \geq \dots \geq \mu_n$. One property that we will require of the expander graph is that

$$\frac{|\mu_i|}{d} \leq 1/10, \quad (11.1)$$

for all i . We will see how to construct such graphs next week. In the meantime, let me promise you that such graphs exist with $d = 400$. Degree 400 seems sort of big, but we will see that it is reasonable for our application.

For the first run of our algorithm, we will require r truly random bits. We treat these bits as a vertex of our graph. For each successive run, we choose a random neighbor of the present vertex, and feed the corresponding bits to our algorithm. That is, we choose a random i between 1 and 400, and move to the i th neighbor of the present vertex. Note that we only need $\log_2 400 \approx 9$ random bits to choose the next vertex. So, we will only need 9 new bits for each run of the algorithm.

One thing to be careful about here is that we will need a very concise description of the graph. As the graph has 2^r vertices, and r may be large, we certainly wouldn't want to store the whole graph, let alone the adjacency matrix. Instead, we need a fast procedure that given the name of a vertex $v \in \{0, 1\}^r$ and an $i \leq d$ will quickly produce the name of the i th neighbor of v . To see how such a thing might be possible, consider the hypercube graph on $\{0, 1\}^r$. It has degree $d = r$, and given a $v \in \{0, 1\}^r$ and an $i \leq d$, we can obtain the name of the i th neighbor of v by just flipping the i th bit of v . For now, I promise you that we know expanders with similarly concise descriptions.

11.6 Formalizing the problem

Assume that we are going to run our algorithm $t + 1$ times. So, our pseudo-random generator will begin at a truly random vertex v , and then take t random steps. Recall that we defined X to be the set of vertices on which the algorithm gets the answer wrong, and we assume that $|X| \leq 2^r/100$. If we report the majority of the outcomes of the $t + 1$ runs of the algorithm, we will return the correct answer as long as the random walk is inside X less than half the time. To analyze this, let v_0 be the initial random vertex, and let v_1, \dots, v_t be the vertices produced by the t steps of the random walk. Let $T = \{0, \dots, t\}$ be the time steps, and let $S = \{i : v_i \in X\}$. We will prove

$$\Pr [|S| > t/2] \leq \left(\frac{2}{\sqrt{5}} \right)^{t+1}.$$

To begin our analysis, recall that the initial distribution of our random walk is $\mathbf{p}_0 = \mathbf{1}/n$. Let

χ_X and χ_Y be the characteristic vectors of X and Y , respectively, and let $D_X = \text{diag}(\chi_X)$ and $D_Y = \text{diag}(\chi_Y)$. Let

$$\mathbf{W} = \frac{1}{d} \mathbf{A} \quad (11.2)$$

be the transition matrix of the ordinary random walk on G . Last lecture I said that I would always use \mathbf{W} to denote the transition matrix of the lazy random walk. But, this is not the lazy random walk. It would be silly to use the lazy random walk for this problem, as there is no benefit to re-running the experiment with the same random bits as before. So, I lied. Let $\omega_1, \dots, \omega_n$ be the eigenvalues of \mathbf{W} . By (11.2) and assumption (11.1), $|\omega_i| \leq 1/10$.

Let's see how we can use these matrices to understand the probabilities under consideration. For a probability vector \mathbf{p} on vertices, the probability that a vertex chosen according to \mathbf{p} is in X may be expressed

$$\chi_X^T \mathbf{p} = \mathbf{1}^T D_X \mathbf{p}.$$

The second form will be more useful, as

$$D_X \mathbf{p}$$

is the vector obtained by zeroing out the events in which the vertices are not in X . If we then want to take a step in the graph G , we multiply by \mathbf{W} . That is, the probability that the walk starts at vertex in X , and then goes to a vertex i is $q(i)$ where

$$\mathbf{q} = \mathbf{W} D_X \mathbf{p}_0.$$

Continuing this way, we see that the probability that the walk is in X at precisely the times $i \in R$ is

$$\mathbf{1}^T D_{Z_t} \mathbf{W} D_{Z_{t-1}} \mathbf{W} \cdots D_{Z_2} \mathbf{W} D_{Z_0} \mathbf{p}_0,$$

where

$$Z_i = \begin{cases} X & \text{if } i \in R \\ Y & \text{otherwise.} \end{cases}$$

We will prove that this probability is at most $(1/5)^{|R|}$. It will then follow that

$$\begin{aligned} \Pr[|S| > t/2] &\leq \sum_{|R| > t/2} \Pr[\text{the walk is in } X \text{ at precisely the times in } R] \\ &\leq 2^{t+1} \left(\frac{1}{5}\right)^{(t+1)/2} \\ &= \left(\frac{2}{\sqrt{5}}\right)^{t+1}. \end{aligned}$$

11.7 Matrix Norms

Recall that the operator norm of a matrix M (also called the 2-norm) is defined by

$$\|M\| = \max_v \frac{\|M\mathbf{v}\|}{\|\mathbf{v}\|}.$$

The matrix norm measures how much a vector can increase in size when it is multiplied by \mathbf{M} . When \mathbf{M} is symmetric, the 2-norm is just the largest absolute value of an eigenvalue of \mathbf{M} (prove this for yourself). It is also immediate that

$$\|\mathbf{M}_1 \mathbf{M}_2\| \leq \|\mathbf{M}_1\| \|\mathbf{M}_2\|.$$

You should also verify this yourself. As \mathbf{D}_X , \mathbf{D}_Y and \mathbf{W} are symmetric, it is easy to show that each has norm at most 1.

Warning 11.7.1. *While the largest eigenvalue of a walk matrix is 1, the norm of an asymmetric walk matrix can be larger than 1. For instance, consider the walk matrix of the path on 3 vertices. Verify that it has norm $\sqrt{2}$.*

Our analysis rests upon the following bound on the norm of $\mathbf{D}_X \mathbf{W}$.

Lemma 11.7.2.

$$\|\mathbf{D}_X \mathbf{W}\| \leq 1/5.$$

Let's see why this implies the theorem. For any set R , let Z_i be as defined above. As $\mathbf{p}_0 = \mathbf{W} \mathbf{p}_0$, we have

$$\mathbf{1}^T \mathbf{D}_{Z_t} \mathbf{W} \mathbf{D}_{Z_{t-1}} \mathbf{W} \cdots \mathbf{D}_{Z_2} \mathbf{W} \mathbf{D}_{Z_0} \mathbf{p}_0 = \mathbf{1}^T (\mathbf{D}_{Z_t} \mathbf{W}) (\mathbf{D}_{Z_{t-1}} \mathbf{W}) \cdots (\mathbf{D}_{Z_0} \mathbf{W}) \mathbf{p}_0.$$

Now,

$$\|\mathbf{D}_{Z_{t-1}} \mathbf{W}\| \leq \begin{cases} 1/5 & \text{for } i \in R, \text{ and} \\ 1 & \text{for } i \notin R. \end{cases}$$

So,

$$\|(\mathbf{D}_{Z_t} \mathbf{W}) (\mathbf{D}_{Z_{t-1}} \mathbf{W}) \cdots (\mathbf{D}_{Z_0} \mathbf{W})\| \leq (1/5)^{|R|}.$$

As $\|\mathbf{p}_0\| = 1/\sqrt{n}$ and $\|\mathbf{1}\| = \sqrt{n}$, we may conclude

$$\begin{aligned} \mathbf{1}^T (\mathbf{D}_{Z_t} \mathbf{W}) (\mathbf{D}_{Z_{t-1}} \mathbf{W}) \cdots (\mathbf{D}_{Z_0} \mathbf{W}) \mathbf{p}_0 &\leq \|\mathbf{1}^T\| \|(\mathbf{D}_{Z_t} \mathbf{W}) (\mathbf{D}_{Z_{t-1}} \mathbf{W}) \cdots (\mathbf{D}_{Z_0} \mathbf{W}) \mathbf{p}_0\| \\ &\leq \|\mathbf{1}^T\| (1/5)^{|R|} \|\mathbf{p}_0\| \\ &= (1/5)^{|R|}. \end{aligned}$$

11.8 The norm of $\mathbf{D}_X \mathbf{W}$

Proof of Lemma 11.7.2. Let \mathbf{x} be any non-zero vector, and write

$$\mathbf{x} = c_1 \mathbf{1} + \mathbf{y},$$

where $\mathbf{1}^T \mathbf{y} = 0$. So,

$$\|\mathbf{x}\| = \sqrt{(c_1 \sqrt{n})^2 + \|\mathbf{y}\|^2} \geq \max(c_1 \sqrt{n}, \|\mathbf{y}\|).$$

We also have

$$\mathbf{D}_X \mathbf{W} \mathbf{x} = c_1 \mathbf{D}_X \mathbf{W} \mathbf{1} + \mathbf{D}_X \mathbf{W} \mathbf{y}.$$

We know that $\mathbf{W}\mathbf{1} = \mathbf{1}$, so

$$\mathbf{D}_X \mathbf{W}\mathbf{1} = \chi_X,$$

and

$$\|c_1 \mathbf{D}_X \mathbf{W}\mathbf{1}\| = c_1 \|\chi_X\| = c_1 \sqrt{X} \leq c_1 \sqrt{n}/10 \leq \|\mathbf{x}\|/10.$$

To bound the norm of the other term, recall that \mathbf{W} is symmetric, and that its leading eigenvector ϕ_1 is a constant vector. So, we may expand \mathbf{y} in the other eigenvectors ϕ_2, \dots, ϕ_n as

$$\mathbf{y} = \sum_{i \geq 2} c_i \phi_i, \quad \text{where } c_i = \phi_i^T \mathbf{y}.$$

We find

$$\|W\mathbf{y}\| = \left\| \sum_{i \geq 2} c_i \omega_i \phi_i \right\| = \sqrt{\sum_{i \geq 2} (c_i \omega_i)^2} \leq (1/10) \sqrt{\sum_{i \geq 2} (c_i)^2} = (1/10) \|\mathbf{y}\|.$$

So,

$$\|\mathbf{D}_X \mathbf{W}\mathbf{y}\| \leq \|\mathbf{D}_X\| \|W\mathbf{y}\| \leq (1/10) \|\mathbf{y}\| \leq (1/10) \|\mathbf{x}\|.$$

Summing these two terms, we find

$$\|\mathbf{D}_X \mathbf{W}\mathbf{x}\| \leq \|\mathbf{D}_X \mathbf{W} c_1 \mathbf{1}\| + \|\mathbf{D}_X \mathbf{W}\mathbf{y}\| \leq (1/10) \|\mathbf{x}\| + (1/10) \|\mathbf{x}\| = (1/5) \|\mathbf{x}\|.$$

□

References

- [ABN⁺92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38(2):509–516, March 1992.
- [Bol88] Béla Bollobás. The isoperimetric number of random regular graphs. *Eur. J. Comb.*, 9(3):241–244, May 1988.
- [Fri08] Joel Friedman. *A Proof of Alon’s Second Eigenvalue Conjecture and Related Problems*. Number 910 in Memoirs of the American Mathematical Society. American Mathematical Society, 2008.
- [HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.
- [IZ89] R. Impagliazzo and D. Zuckerman. How to recycle random bits. In *30th annual IEEE Symposium on Foundations of Computer Science*, pages 248–253, 1989.
- [LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
- [Mar88] G. A. Margulis. Explicit group theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problems of Information Transmission*, 24(1):39–46, July 1988.