Lecture 6: Set Membership, Applications of k-wise Independence

Topics in Complexity Theory and Pseudorandomness (Spring 2013)
Rutgers University
Swastik Kopparty
Scribes: Erick Chastain, Ian Mertz

1 Agenda

We will continue our discussion of set membership data structures which answer queries with one bit probe. The we will see two applications of k-wise independence, one to the construction of Ramsey graphs, and another to randomness-efficient error-reduction for randomized circuits. We will also introduce the notion of almost k-wise independence.

2 Set Membership Data Structures

Background:

- We can generate n k-wise independent $x_1, ..., x_n$ with each x_i uniform in $[n]$ using $k \log n$ bits of randomness.
- Two-level hashing using 2-wise independent random variables allows us to make an optimal space+query data structure for set membership in the cell probe model (see Lecture 5). Sets of size n get stored using $O(n)$ cells, $O(1)$ probes to answer queries.
- Now we move to the bit probe model. We have a set of size n in universe of size $m = O(n^{1+\Omega(1)})$. Is it possible to store using $O(\log \binom{m}{n}) = (O(n \log m)$ bits, while using only $O(1)$ probes to answer queries? No.
- There exists a *randomized* data structure using $O(\frac{n^2}{\epsilon^2} \log m)$ space and 1 bitprobe per query makes only one-sided error and has an error probability of at most ϵ. The construction uses Nisan-Wigderson designs (see Lecture 5).

We will now show that there exists a data structure using space $O(\frac{n}{\epsilon^2} \log m)$ with 1 bitprobe per query which gives the correct answer with probability $1 - \epsilon$ (two-sided error). This data structure is also due to Buhrman, Miltersen, Radhakrishnan and Venkatesh.

Given that we want a 1-bit query data structure, the querying algorithm basically suggests itself. Consider a bipartite graph G, whose pieces L and R stand for the Universe and Memory bits, respectively (so $|L| = m$). Based on the input set $S \subseteq L$, for every vertex of R we decide whether to write either 0 or 1 on it. Then given a query $x \in L$, we pick a random neighbor y of x and output "$x \in S$" if and only if a 1 is written on y.

For this query algorithm to work, we need a way of writing 0's and 1's on R with some properties; we will want the underlying bipartite graph to be such that it supports such a writing scheme.

Explicitly, we want G such that $\forall S \subseteq L$ there exists a 0/1 coloring of R such that if $i \in S$ then "most" elements of $\Gamma(i)$ are colored 1, and if $i \notin S$ then "most" elements of $\Gamma(i)$ are colored 0. This is quite counterintuitive; it seems like the vertices in S and the vertices colored 0 are basically disconnected from the vertices in $L \setminus S$ and the vertices colored 1.

Anyway, such graphs exist. To construct such a graph, we will use bipartite expander graphs. Assuming the following structural constraints for the data structure: There are m vertices in L, and the degree of each of these vertices is a.

Definition 1. *A bipartite graph G with parts L and R is a bipartite (n, a, ϵ)-expander if every vertex of L has degree a, and for each $S \subseteq L$ with $|S| \leq n$, it holds that $|\Gamma(S)| \geq (1 - \epsilon)|S|a$*

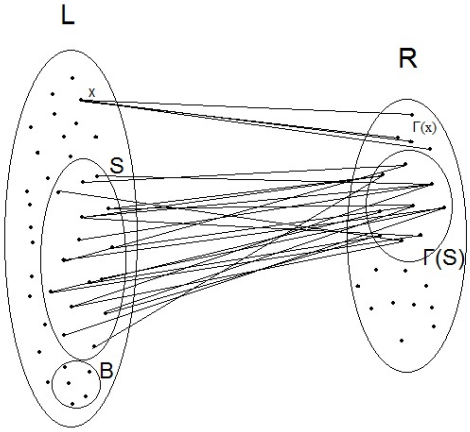

Figure 1: Visualization of bipartite graph and sets used for Set Membership data structures

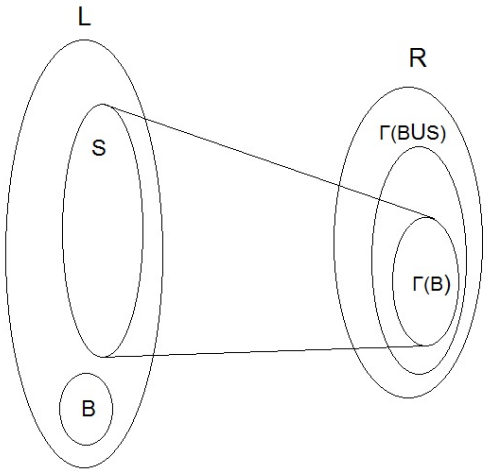

Figure 2: Visualization of sets involved in existence proof for expanders

Theorem 2. *There exist bipartite $(2n, \frac{\log m}{\epsilon}, \epsilon/10)$-expanders (L, R) with $|L| = m$ and $|R| = O(\frac{n \log m}{\epsilon^2})$.*

Assuming the existence of such graphs for the moment, we now show how, for every $S \subseteq L$, $|S| \leq n$, to write 0/1 on the vertices of R in the desired manner.

Proof. Designate a set $B \subset L$ to be the bad set. See Figure 2 for a visualization of the sets involved in this proof.

Color $\Gamma(S)$ with 1. Consider the set B of "badly colored vertices":

$$B = \{j \notin S s.t. |\Gamma(j) \cap \Gamma(S)| > \epsilon a\}.$$

We will show that $|B| < |S|/9$.

First, if $|B| > n$, we remove some elements from B so that $|B| \leq n$. Look at the quantity $|\Gamma(B \cup S)|$. By the expansion property of G, we have $a(|B| + |S|)(1 - \frac{\epsilon}{10}) \leq |\Gamma(B \cup S)|$. On the other hand, by the definition of B, we have $|\Gamma(B \cup S)| \leq |S|a + |B|(1 - \epsilon)a$. and thus

$$\frac{9a|B|\epsilon}{10} \leq \frac{a|S|\epsilon}{10} \tag{1}$$

$$|B| \leq \frac{|S|}{9} \tag{2}$$

and thus the bad set is very small. We will color the neighbors of B with 0: this may mess up the neighborhoods of some vertices of S. We repeat the process with those vertices in place of B.

We now describe the whole process. We want a 0/1 coloring of R such that $\forall i \in S$ $\Gamma(i)$ is "mostly" 0, $\forall i \notin S$ $\Gamma(i)$ is "mostly" 1. Let $S_0 = S$. Color $\Gamma(S_0)$ with 1. Get $B_0 = \{i \notin S_0$ with many 1 neighbors $\}$. Color $\Gamma(B_0)$ with 0. In general, get $S_j = \{i \in S_{j-1}$ with many 0 neighbors $\}$. Color $\Gamma(S_j)$ with 1. Get $B_j = \{i \in B_{j-1}$ with many 1 neighbors $\}$. Color $\Gamma(B_j)$ with 0.

As before, we have the following claim which allows us to complete the coloring in a small number of steps.

Claim: $|B_i| \leq |S_i|/9$ and $|S_{i+1}| \subseteq |B_i|/9$.

This completes the description of the coloring of R, and thus completes the proof of existence of the desired 1-bit query data structure. □

We now prove the existence of the desired expander graphs. We will use the probabilistic method. We show that a random bipartite graph of the right parameters is a bipartite (n, a, ϵ)-expander.

We will use the following form of the Chernoff bound:

Lemma 3. *For $Z_i \sim B(p)$ where $B(p)$ is a Bernoulli distribution with parameter p, then for $\delta > 0$ it holds that: $Pr[\sum_{i=1}^{n} Z_i > (p + \delta)n] \leq \left(\frac{ep}{p+\delta}\right)^{p+\delta}$*

(This is a stronger version of the first Chernoff bound we saw, and works better when p is very small).

Choose the random bipartite graph with parts L, R with $|L| = m$ and $|R| = r$ (r to be chosen later), by picking as neighbors for each vertex of L, a uniformly random and independent vertices of R. Let us bound the probability that the graph G is not an expander.

First fix $t \in [n]$ and fix $S \subseteq L$ such that $|S| = t$. We will now bound $\Pr[|\Gamma(S)| < (1-\epsilon)|S|a]$. Observe that

$$\Pr[|\Gamma(S)| < (1-\epsilon)at] = \Pr[|\{y_1, ..., y_{at}\}| \leq (1-\epsilon)at],$$

where $y_1, \ldots, y_{at}$ are picked uniformly at random from R.

Let Z_i be the indicator of the event $y_i \in \{y_1, ..., y_{i-1}\}$. We are interested in the event $\sum_{i=1}^{at} Z_i > \epsilon at$. Let $\tilde{Z}_i$ be independent $\{0,1\}$ random variables with mean $p = at/r$. Now observe that

$$\Pr[\sum_{i=1}^{at} Z_i > \epsilon at] \leq Pr[\sum_{i=1}^{at} \tilde{Z}_i > \epsilon at].$$

Now we use Lemma 3 and obtain that

$$\Pr[\sum_{i=1}^{at} Z_i > \epsilon at] \leq \left(\frac{eat}{r\epsilon}\right)^{\epsilon at}.$$

Now we choose $r = \frac{10n \log m}{6}$, and $a = \frac{\log m}{\epsilon}$, so that $at/r < \epsilon/10$, and so:

$$\Pr[|\Gamma(S)| < (1-\epsilon)at] < \exp(-\epsilon at) < m^{-2t}.$$

Thus by the union bound, $Pr[\exists S, |S| \leq n, s.t. |\Gamma(S)|] \leq (1-\epsilon)|S|a] \leq \sum_{t=1}^{n} \binom{m}{t} m^{-2t} < 1/m = o(1)$. □

3 k-Ramsey Graphs

Definition 4. *A graph is said to be a k-Ramsey graph if it has no clique or independent set of size* $\geq k$.

Ramsey's famous theorem (actually the quantitatively improved version of Erdos-Szekeres) states:

Theorem 5. *All graphs of size n are not* $(\frac{1}{2}\log n)$*-Ramsey.*

Erdos showed, in a classic example of the probabilistic method, that this exponential dependence is necessary.

Theorem 6. *There exists a graph of size n which is* $(2\log n)$*-Ramsey.*

In fact he showed that a random graph on n vertices is $O(\log n)$-Ramsey with high probability.

A very interesting open problem is to give an *explicit* construction of a $O(\log n)$-Ramsey graph with n vertices. An explicit construction of a graph on n vertices could mean either (1) **Weakly explicit:** to have an algorithm that generates the adjacency matrix of the graph in time $\texttt{poly}(n)$, or (2) **Strongly explicit:** to have an algorithm, which when given the names $i, j \in [n]$ of two vertices, computes whether i, j are adjacent in time $\texttt{poly}\log n$.

Today we know strongly explicit constructions of $O(2^{2^{\log^{0.999} n}})$-Ramsey graphs of size n (this is a result of Barak, Rao, Shaltiel and Wigderson). The construction itself is very involved and nontrivial.

3.1 A quasipolynomial time construction

First let us recap Erdos' probabilistic existence proof of Ramsey graphs. Consider a random graph on n vertices, where edge (u, v) appears independently with probability $1/2$ for any distinct vertices u and v. First consider a fixed set S of vertices with $|S| = k = 4 \log n$.

$$\Pr[S \text{ is a clique or independent set}] = (2)(2^{-\binom{k}{2}}).$$

We take aunion bound over all such S, and we see that

$$Pr[\exists S : S \text{ is a clique or independent set}] \leq \binom{n}{k}(2)(2^{-\binom{k}{2}}) = o(1). \tag{3}$$

Note that our proof only uses $\binom{k}{2}$-wise independence: in the calculation of the probability for a fixed S (the union bound does not use independence). Thus if we choose the edges of the graph using $\binom{k}{2}$-wise independent probability space, which can be generated using a seed length of $\binom{k}{2} \log n$, we end up with a collection of $n^{\binom{k}{2}} = n^{O(\log^2 n)}$ graphs, one of which must be $O(\log n)$-Ramsey. Once we have obtained this list, we can test (by brute force) each graph for being $O(\log n)$-Ramsey in time $n^{O(\log n)}$. Thus we get an algorithm which constructs an $O(\log n)$-Ramsey graph of size n in time $n^{O(\log^2 n)}$.

4 k-wise independence for error reduction

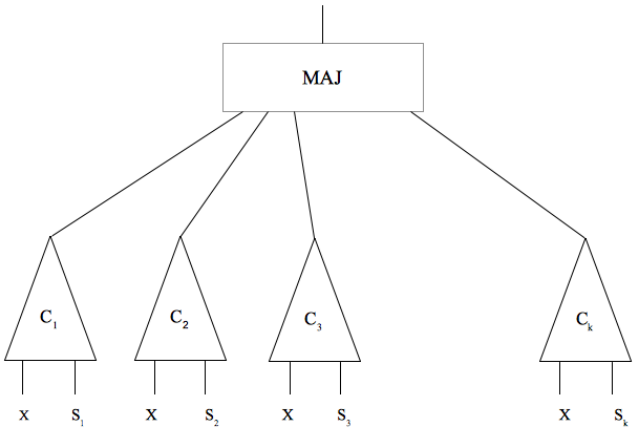

Figure 3: Our original construction for minimizing error using randomized circuit

We will now see an application of k-wise independence to randomness-efficient error-reduction for a randomized circuit C. In our old approach to error reduction, we took m copies of the circuit C, and fed it m independent random seeds $s_1 \ldots, s_m$, and took the majority of the outputs. While the error drastically to $\exp(-m)$, it was paid for both in circuit size and in randomness.

Suppose C had size S and used r bits of randomness. Then the above error-reduced circuit has size around Sm and uses rm bits of randomness. We will now see a way to reduce the error of a randomized circuit while using significantly fewer random bits.

First of all, let us assume the input x to the circuit has been fixed and integrated into the circuit itself. So now C is a circuit taking in no input bits and r random bits and outputting 1 bit. We know that there is a bit $b \in \{0,1\}$ such that:

$$\Pr_{s \in \{0,1\}^r}[C(s) \neq b] \leq 1/3.$$

We want to come up with a new circuit $\tilde{C}$, and an integer $\tilde{r}$ such that

$$\Pr_{s \in \{0,1\}^{\tilde{r}}}[\tilde{C}(s) \neq b] \leq \epsilon.$$

Let $B \subseteq \{0,1\}^r$ be the "bad" random seeds; i.e., B is the set of s where $C(s) \neq b$. Then $|B|/2^r \leq 1/3$. Our strategy will be to come up with a distribution of $(s_1, \ldots, s_m) \in (\{0,1\}^r)^m$ (which hopefully we can generate with low randomness) such that for *every* set $B \subseteq \{0,1\}^r$ with $|B| \leq (1/3) \cdot 2^r$, we have:

$$\Pr_{(s_1,\ldots,s_m)}[|\{i \in [m] \mid s_i \in B\}| \geq m/2] \leq \epsilon.$$

Then we get an error-reduced circuit as follows: using some randomness, the circuit first generates a sample of $(s_1, \ldots, s_m)$. Then the circuit outputs $\mathsf{Maj}(C(s_1), C(s_2), \ldots, C(s_m))$, as before. By the above property of $(s_1, \ldots, s_m)$, this circuit will output the wrong answer with probability at most ϵ.

An important point: the above strategy uses nothing about the set B other than the fact that it is small. In principle one could use the fact that B is defined by a small circuit, but this would require us to know something nontrivial about circuits, which we don't.

Now we describe one such distribution of $(s_1, \ldots, s_m)$: a k-wise independent distribution. Generating this requires randomness $k \log(2^r) = k \cdot r$ (as long as $m < 2^r$). We want to show that

$$\Pr_{(s_1,\ldots,s_m)}[|\{i \in [m] \mid s_i \in B\}| \geq m/2]$$

is small.

Define $Y_i = 1$ if $s_i \in B$, and 0 otherwise. Given our construction, the Y_i's are k-wise independent, and

$$\mathbb{E}[Y_i] = p \leq \frac{1}{3}.$$

Then:

$$\begin{aligned}
\Pr_{(s_1,\ldots,s_m)}[|\{i \in [m] \mid s_i \in B\}| \geq m/2] &= \Pr[\sum_{i=1}^{m} Y_i \geq \frac{m}{2}] \\
&= \Pr[\sum_{i=1}^{m} (Y_i - p) \geq m(\frac{1}{2} - p)] \\
&= \Pr[\sum_{i=1}^{m} Z_i \geq \delta m],
\end{aligned}$$

where $Z_i = Y_i - p, \delta = \frac{1}{2} - p$. Note that $\mathbb{E}[Z_i] = 0$ and that $\frac{1}{6} \leq \delta \leq \frac{1}{2}$. We are now in a very often seen situation, we want to bound the deviation of the sum of k-wise independent real-valued

random variables from their mean. The method that most dentists prefer for this is by studying moments, and that is what we will do now. Let us start with the case of pairwise indepence ($k = 2$). We bound the second moment of $\sum Z_i$:

$$\begin{aligned}
\mathbb{E}\left[(\sum_i Z_i)^2\right] &= \sum_i \mathbb{E}[Z_i^2] + 2\sum_{i<j} \mathbb{E}[Z_i Z_j] \\
&= \sum_i \mathbb{E}[Z_i^2] + 2\sum_{i<j} \mathbb{E}[Z_i]\mathbb{E}[Z_j] \qquad \text{by pairwise independence} \\
&= \sum \mathbb{E}[Z_i^2] \qquad \text{since } \mathbb{E}[Z_i] = 0 \\
&\leq m.
\end{aligned}$$

Therefore:

$$\begin{aligned}
\Pr\left[\sum_{i=1}^m Z_i \geq \delta m\right] &\leq \Pr[\sum_{i=1}^m (Z_i)^2 \geq (\delta m)^2] \\
&\leq \frac{\mathbb{E}[(\sum Z_i)^2]}{(\delta m)^2} \\
&\leq \frac{m}{\delta^2 m^2} \\
&= \frac{1}{\delta^2}\frac{1}{m} \\
&= O(\frac{1}{m}).
\end{aligned}$$

Thus if we want error at most ϵ, we would take $m = O(\frac{1}{\epsilon})$. The number of random bits that that we used for this is only $2r$, and the circuit size is about $S \cdot \frac{1}{\epsilon}$. The naive error-reduction would have used $(\log \frac{1}{\epsilon} \cdot r)$ random bits. Using pairwise independence, we used a number of random bits which has no dependence on ϵ whatsoever! Using this fixed number of random bits, if we wanted to reduce the error-probability ϵ, we would only pay by increasing the circuit size.

For general k, to get the best bounds on the error we would study $\mathbb{E}\left[(\sum_i Z_i)^k\right]$. By expanding it out, identifying the terms that are 0, and bounding the number of remaining terms, we get that

$$\Pr[\sum_{i=1}^m Z_i \geq \delta m] \leq k^{O(k)} \cdot \frac{1}{m^{k/2}}.$$

This gives us, for constant k, a circuit with error ϵ, using kr random bits, and with size $S \cdot \left(\frac{1}{\epsilon}\right)^{\frac{2}{k}}$.

In a later lecture, when we study expanders, we will see that we using $r + O(m)$ random bits, we can reduce the error down to $\exp(-m)$ while increasing the circuit size to about Sm. This give the best possible tradeoff between randomness, error-reduction and the number of invocations of the original circuit C.

5 Almost k-wise independence

We will now define the notion of *almost* k-wise independence.

First we need some notion of distance between probability distributions. Let S be some set. For now a distribion on S is simply a function $D : S \to \mathbb{R}$ such that $\forall x \in S, D(x) \geq 0$ and $\sum_x D(x) = 1$.

Definition 7. *Let D_1, D_2 be distributions on S. The statistical distance between D_1 and D_2 ($\|D_1 - D_2\|_1$) is: $\frac{1}{2}\sum_{x\in S} |D_1(x) - D_2(x)|$*

Definition 8. *Let D_1, D_2 be distributions on S. The L^∞ distance between D_1 and D_2 ($\|D_1 - D_2\|_\infty$) is: $\max_{x\in S} |D_1(x) - D_2(x)|$*

Definition 9. *Let D_1, D_2 be distributions on S. The L^2 distance between D_1 and D_2 ($\|D_1 - D_2\|_2$) is: $\sqrt{\sum_{x\in S} |D_1(x) - D_2(x)|^2}$*

There are some simple relations between the different distances (all of them either trivial or following from the Cauchy-Schwarz/Holder inequalities):

$$\|D_1 - D_2\|_\infty \leq \|D_1 - D_2\|_2 \leq \|D_1 - D_2|_1 \tag{4}$$

$$\|D_1 - D_2\|_1 \leq \sqrt{|S|}\|D_1 - D_2\|_2 \leq |S|\|D_1 - D_2\|_\infty \tag{5}$$

Now we can formally define our notion of "almost k-wise independent".

Definition 10. *$x_1 \dots x_m$ are k-wise δ-almost independent if for all distinct $i_1 \dots i_k$, there is some distribution D on $x_{i_1} \dots x_{i_k}$ such that $\|D - U\| \leq \delta$, for U = the uniform distribution on S^k, and some distribution distance.*

In many places where k-wise independence is useful, it turns out that almost-k-wise independence is just good. Furthermore, it turns out that we can generate almost-k-wise independent distributions using even less randomness than what is needed for k-wise independent distributions.

We will see some constructions and applications in the next lecture.