Today

- Low Density Parity Check Codes.
- Linear Time Decoding.

Decoding from Parity Check & Syndrome

- Parity check matrix $H$ is $n \times m \ (k = n - m)$.
- $y$ codeword iff $yH = 0$.
- If $y$ is close to codeword, then can $yH$ give any info?
  - Idea: let $(yH)_j \neq 0$, then one of the bits $i$ such that $H_{ij} \neq 0$ is corrupt.
  - Usually: This is not useful. Too many such bits.
  - Low-Density Parity Check Idea: But may be useful if $H$ has low weight.

Low Density Parity Check Matrices

- Defn: $H$ has sparsity $c$ if every column has at most $d$ non-zeroes.
- Defn: $\{H_{n,m}\}_{n,m}$ defines a LDPC Code if there exists $d$ such that every matrix in family is $c$-sparse.
- Theorem: [Gallager '63] LDPC codes achieve Gilbert-Varshamov bound.
- Theorem: [Gallager '63] $\exists$ LDPC codes that correct constant fraction in linear time (efficiently)!
- Subsequent work: [Tanner] (composition + explicit directions); [Sipser-Spielman]
Graph-theoretic view

- $n \times m$ 0/1 Matrices $\equiv$ Bipartite Graphs $(L, R, E)$ with $|L| = n$, and $|R| = m$.
- Left vertex = coordinate of (code)words.
- Right vertex = constraint
- $c_1, \ldots, c_n$ codeword if parity of neighbors of every right vertex is even.
- When/Why is this an error-correcting code (of large minimum distance)?

Bad graphs

- If there exists a subset $S \subseteq L$ of small size such that $S$ has neighbors of only even degree on right. Then $1_S$ is a codeword (necessary and sufficient).
- How to rule this out?
- Suppose know that no small set $S$ has neighbors of degree $\geq 2$. Or ... Every small set has some "Unique neighbors".
- Then $G$ leads to good code.

Unique neighbors in graphs

- How can we prove existence of unique neighbors for small sets?
- Well studied in context of expansion: If graph is a very good expander then small sets have unique neighbors.
- Defn: $G$ is $(c, d)$-regular if every left vertex has degree $c$ and every right vertex has degree $d$.
- Expansion: $G = (L, R, E)$ is a $\gamma, \delta$ expander if every set $S \subseteq L$ with $|S| \leq \delta n$ has $|\Gamma(S)| \geq \gamma|S|$. ($\Gamma(S) = \{j \in R | \exists i \in S, (i, j) \in E\}$).

Folklore theorem about unique neighbors

- $\gamma > c/2$ implies, $S$ of size less than $\delta n$ has unique neighbor.
- $\gamma$ and $c$?
  - Note trivially $\gamma \leq c$.
  - Should scale linearly with $c$ for $\delta = o(1)$.
  - For random $(c, d)$-regular graph, can get $\gamma = c - 1$ for some $\delta > 0$
Formal folklore claim & proof

Claim: $G(c,d)$-regular and $(\gamma, \delta)$-expander implies $S$ of cardinality $\leq \delta n$ has at least $(2\gamma - c)|S|$ unique neighbors.

Proof: Let $U$ be unique neighbors and $D$ be degree two or greater neighbors. We have $U + 2D \leq \#$ edge into $S = cS$. $U + D \geq \gamma S$. Combining, get bound.

- Leads to following algorithm.

Decoding?

- Once again boils down to unique neighbors .... How?
- Lets start with a simple hope: Pick violated constraint and flip some variable in it.
- Not such a good idea - since most likely violated constraint has a unique flipped neighbor and mostly correct neighbors. So we are more likely to flip good guy instead of bad!
- Better idea: Take a violated constraint and try to figure out which one of its neighbors is the error. How to detect this? Erroneous bit hopefully participates in many violated constraints.

Decoding algorithm

- While $\exists$ left vertex with more violated neighbors than unviolated ones, FLIP this vertex.

Note: Alg. can be implemented to take $O(1)$ time per iteration.
Analysis

- # iterations \( \leq \) # initially violated constraints.
- \( \Rightarrow \) Alg. must terminate.
- Termination possibilities:
  1. Terminates with right codeword.
  2. Terminates with wrong codeword.
  3. Terminates at non-codeword.

Analysis: Ruling out (2)

Claim 1: If # errors \( \leq \delta n/(2c) \) then Case 2 can’t happen.

Proof: If # errors as above, then initial # violated constraints is less than \( \delta n/2 \). So alg. terminates in \( \delta n/2 \) steps. At this point distance from transmitted word \( \leq \) #errors + # steps \( \leq \delta n/(2c) + \delta n/2 < \delta n \). But if rec’d vector is distinct from transmitted word, then distance \( \geq \delta n \).

Analysis: Ruling out (3)

Claim 2: At final iteration, say \( S \) is the set of indices that are in error. Then if \( 0 < |S| \leq \delta n \), then there exists \( i \in S \) with more violated neighbors than unviolated, provided \( \gamma > 3c/4 \).

Proof: Actually will prove more unique neighbors than non-unique. Say # unique neighbors > \((c/2)|S|\). (True if \(2\gamma - c > c/2\) or \(\gamma > 3c/4\)). Then some vertex in \( S \) has more than \((c/2)\) unique neighbors. QED.

Conclusion

- LDPC code based on very good expander leads to Linear time decoding.
- Can we find such good expanders?
- For long time, answer was NO. Random graph was this good, but couldn’t even pick one at random and test. Big bottleneck exactly at \(\gamma = c/2\). The unique neighbor property can not be guaranteed by the eigenvalue method ...
- Recent breakthroughs: Capalbo, Reingold, Vadhan, and Wigderson. Can build such graphs; and techniques quite familiar. Might do some of this next time.
• What did we know to construct? Graphs with $\gamma < c/2$.

• Can we do anything with these? Yes [Tanner,SipserSpielman].

Tanner products

• Suppose $\gamma > c/\Delta$.

• Can we use this to do anything?

• Can’t prove neighborhood of $S$ has unique neighbor.

• But can prove has low-degree neighbor (into $S$).

• Claim: $|S| \leq \delta n$ implies $(\Delta \gamma - c)|S|/(\Delta - 1)$ neighbors of degree less than $\Delta$ into $S$.

• Proof as usual.

• So what?

• Now insist that neighbors of constraint vertex come from code $C$ of min. dist. $\Delta$.

• Gives explicit construction of $\Omega(1)$ rel. dist. code.

• Sipser-Spielman give linear time decoding algorithm.