Matrix Rigidity

Rigidity of Codes

Sasha Golovnev

November 2, 2020

Linear Codes

- A linear code C is a k-dimensional subspace of $\mathbb{F}^n$

Linear Codes

- A linear code C is a k-dimensional subspace of $\mathbb{F}^n$
- The distance of C is

$$d(C) = \mathbf{min}\left(\|w\|_0 : w \in C, w \neq \mathbf{0}\right)$$

Linear Codes

- A linear code C is a k-dimensional subspace of $\mathbb{F}^n$
- The distance of C is

$$d(C) = \mathbf{min}\left(\|w\|_0 : w \in C, w \neq \mathbf{0}\right)$$

- A generator matrix $G \in \mathbb{F}^{n \times k}$ is a matrix whose columns form a basis of C

Explicit Linear Codes

Proposition

For any finite field $\mathbb{F}$, there exists an explicit family of linear error correcting codes over $\mathbb{F}$ of dimension $k = n/4$ and minimum distance $d = \delta n$ for a constant $\delta > 0$.

Explicit Linear Codes

Proposition

For any finite field $\mathbb{F}$, there exists an explicit family of linear error correcting codes over $\mathbb{F}$ of dimension $k = n/4$ and minimum distance $d = \delta n$ for a constant $\delta > 0$.

Such codes are called good.

Rigidity of Codes

- Friedman, PR, SSS: every generator matrix G of a good code has rigidity

$$\mathcal{R}_G^{\mathbb{F}}(r) \geq \Omega\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right) .$$

Rigidity of Codes

- Friedman, PR, SSS: every generator matrix G of a good code has rigidity

$$\mathcal{R}_G^{\mathbb{F}}(r) \geq \Omega\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right) .$$

- Every good code has a generator matrix G

$$\mathcal{R}_G^{\mathbb{F}}(\varepsilon n) \geq \Omega(n^2) .$$

Rigidity of Codes

- Friedman, PR, SSS: every generator matrix G of a good code has rigidity

$$\mathcal{R}_G^{\mathbb{F}}(r) \geq \Omega\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right).$$

tight

- Every good code has a generator matrix G ✓

$$\mathcal{R}_G^{\mathbb{F}}(\varepsilon n) \geq \Omega(n^2).$$

- Some good codes have a generator matrix G ✓

$$\mathcal{R}_G^{\mathbb{F}}(r) \leq O\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right).$$

Rigidity of Codes

- Friedman, PR, SSS: every generator matrix G of a good code has rigidity

$$\mathcal{R}_G^{\mathbb{F}}(r) \geq \Omega\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right).$$

cannot improve

- Every good code has a generator matrix G

$$\mathcal{R}_G^{\mathbb{F}}(\varepsilon n) \geq \Omega(n^2).$$

- Some good codes have a generator matrix G

$$\mathcal{R}_G^{\mathbb{F}}(r) \leq O\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right).$$

- Thus, we cannot improve the known explicit bound for all generator matrices of good codes

Rigid Generators

Lemma

Let $C \subseteq \mathbb{F}^n$ be a subspace of dimension $k = \Theta(n)$ ~~and distance $d = \Theta(n)$~~. There exists a generator matrix $A \in \mathbb{F}^{n \times k}$ of C of rigidity

$$\mathcal{R}_A^{\mathbb{F}}(\varepsilon k) \geq \Omega(n^2)$$

for a constant $\varepsilon > 0$.

$\dim(C) = k.$ $\quad C \subseteq \mathbb{F}^n$

$G \in \mathbb{F}^{n \times k}$ – basis of C s.t

$R_G^{\mathbb{F}}(\varepsilon k) \geq \Omega(k^2)$

$$\boxed{\begin{matrix} x_1 + x_2 + x_7 + x_n = 0 \\ x_3 + x_8 + x_{n-1} = 0 \end{matrix}}$$

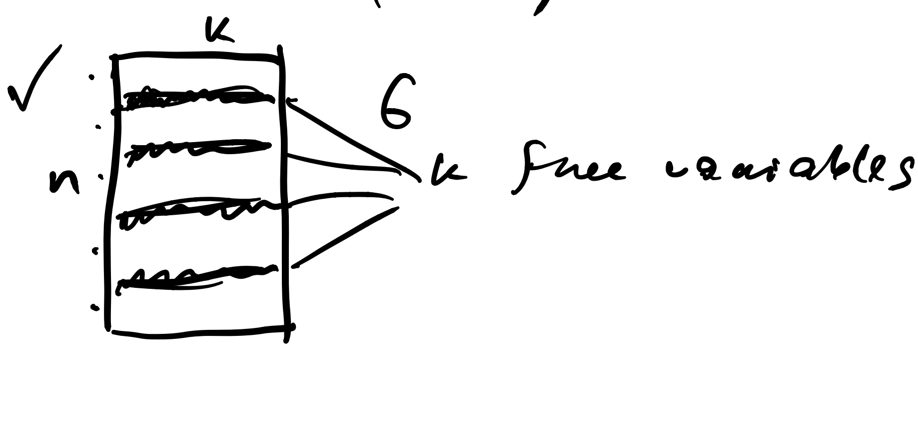
$(n-2)$-dim subspace

Free variables:

$x_1, x_2, \dots, x_{n-2}$ – free variables $\in \mathbb{F}^{n-2}$

$$\boxed{\begin{cases} x_n = -x_1 - x_2 - x_7 \\ x_{n-1} = -x_3 - x_8 \end{cases}}$$

k-dim subspace has k free variables ($\in \mathbb{F}^k$)



Whatever I write in these

k rows, I can always extend it

to a basis $\mathbb{F}^{n\times k}$.

For every $B \in \mathbb{F}^{k\times k}$, I can

find a generator matrix

$G \in \mathbb{F}^{n\times k}$ s.t. G projected

to k free coordinates equals B.

$$\mathcal{R}_G^{\mathbb{F}}(r) \geq \mathcal{R}_B^{\mathbb{F}}(r) \qquad \forall r$$

For example, in HW2, Problem 1, we'll show

that a random B has rigidity

$$\mathcal{R}_B^{\mathbb{F}}(\varepsilon k) \geq \Omega(k^2)$$

$$\Rightarrow \mathcal{R}_G^{\mathbb{F}}(\varepsilon k) \geq \Omega(k^2) \qquad \square$$

Limitation for Codes

$0 \leq d \leq n(\frac{1}{2}-\delta)$

Theorem (Dvi16)

For every $\delta > 0$ and large enough $n \in \mathbb{N}$, there exists a generator matrix $M \in \mathbb{F}_2^{n \times k}$ of a linear code $C \subseteq \mathbb{F}_2^n$ of dimension $k = \Theta(n)$ and distance $d = (1/2 - \delta)n$ of rigidity

optimal dim

optimal distance

$$\mathcal{R}_M^{\mathbb{F}_2}(r) \leq O\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right)$$

for every $\Omega(\log n) \leq r \leq O(n)$.

Distance of a code = min Ham dist between two points in the code.

$\mathbb{F}_2^n$

Can I have distance $\underbrace{\frac{n}{2}+1}$?

$0^n \in C$

$x \in C \quad \|x\|_0 \geq \frac{n}{2}+1$

$1^n \notin C \quad \|1^n - x\|_0 \leq \frac{n}{2}-1 < d$

at most 2 points

Can have distance $\frac{n}{2}$, then $k = O(\log n)$

The best what you can have over $\mathbb{F}_2$ for reasonable params

$d = n(\frac{1}{2}-\delta)$,

δ is constant.

Explicit codes

Odd Sum of Bernoullis

Lemma

For any $n \in \mathbb{N}$ and $p \in [0,1]$, let X be a sum of n independent Bernoulli random variables with mean p, then

$$\Pr[X \text{ is odd}] = \frac{1}{2} - \frac{1}{2}(1-2p)^n \geq \frac{1}{2} - \frac{1}{2}e^{-2pn}.$$

Lemma

For any $n \in \mathbb{N}$ and $p \in [0,1]$, let X be a sum of n independent Bernoulli random variables with mean p, then

$$\Pr[X \text{ is odd}] = \frac{1}{2} - \underline{\frac{1}{2}(1-2p)^n} \geq \frac{1}{2} - \frac{1}{2}e^{-2pn}.$$

$$\Pr[X \text{ is even}] \geq \frac{1}{2} + \frac{1}{2}(1-2p)^n$$

Proof:

$$\overbrace{\square \quad \square \quad \square \quad \cdots \quad \square}^{n \text{ coins}}$$

$$\Pr[X=1] = \binom{n}{1} \cdot p \cdot (1-p)^{n-1}$$

$$\Pr[X=3] = \binom{n}{3} \cdot p^3 \cdot (1-p)^{n-3}$$

$$\cdots$$

$$\Pr[X \text{ is odd}] =$$

$$= \sum_{k=1}^{\lceil \frac{n}{2} \rceil} \Pr[X = 2k-1] =$$

$$= \sum_{k=1}^{\lceil \frac{n}{2} \rceil} \binom{n}{2k-1} \cdot p^{2k-1} \cdot (1-p)^{n-(2k-1)}$$

$$=$$

$$((1-p)+p)^n = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

$$((1-p)-p)^n = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} (-1)^i$$

$$\sum_{\substack{i=0 \\ i=\text{odd}}}^{n} \binom{n}{i} p^i (1-p)^{n-i} \cdot 2$$

$$= \frac{((1-p)+p)^n - ((1-p)-p)^n}{2}$$

$$= \frac{1-(1-2p)^n}{2} = \frac{1}{2} - \frac{(1-2p)^n}{2}$$

$x \geq -1$ $\quad \underbrace{1+x} \leq e^x$

$$\geq \frac{1}{2} - \frac{1}{2}\left(e^{-2p}\right)^n = \frac{1}{2} - \frac{1}{2} e^{-2pn}$$

$\square$

Codes with Optimal Redundancy

Lemma

For every $d < n/2$, there exists a linear code $C \subseteq \mathbb{F}_2^n$ of dimension k and distance d such that the redundancy of C is

tight.

$$r = n - k = O\left(d \log\left(\frac{n}{d}\right)\right).$$

Linear code

$C \subseteq \mathbb{F}^n$ $\quad$ $\dim(C) = k.$

distance d

Redundancy $\boxed{R} = n - k.$

Linear codes: you can take a k-bit message, encode it in $n > k$ bits, send these n bits, even if there are a few mistakes in the sent message, one can correctly decode the k-bit message.

$\boxed{R} = \overline{n-k}$ redundancy

Let's prove $\exists$ linear codes in $\mathbb{F}_2^n$ of dim k s.t.

$$R = n-k \leq O(d \log(\tfrac{n}{d}))$$

Proof:

Iteratively construct basis with k vectors.

x First vector — any non-zero from $\mathbb{F}_2^n$

y 2nd vector — any vector on distance to the pr sp $\geq d$.

In my space I have 4 vectors: $0, x, y, x+y$.

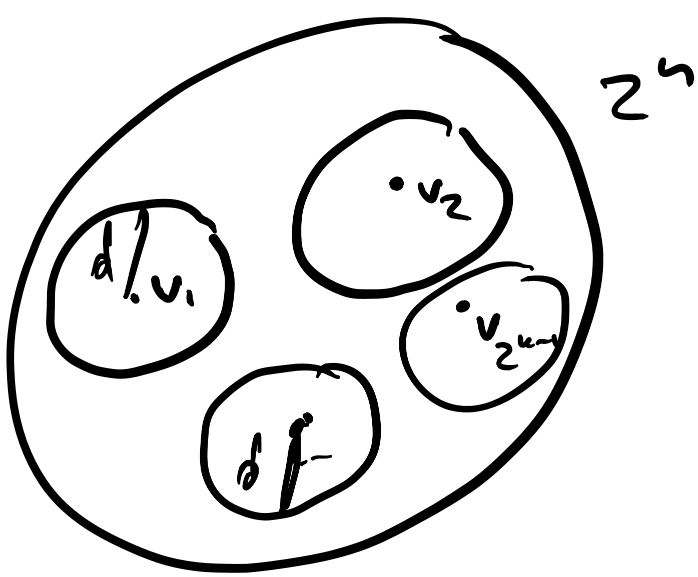
z 3rd — dist from z to 4 vectors $\geq d$.

8 vectors

keep doing, want k basis vectors, i.e., 2^k vectors in my space.

2^{k-1} vectors in the current space.

Be sure that among 2^n vectors $\exists$ a vector $> d$ away from all 2^{k-1} vectors..



$2^{k-1} \cdot$ Volume of ball of radius $d < 2^n$

$$2^{k-1} \cdot \binom{n}{\leq d} < 2^n$$

If this holds, then we have a code that we want.

we can k s.t.

$$2^k \cdot \binom{n}{\leq d} = 2^n$$

$$n-k = \log_2 \binom{n}{\leq d}$$

$$R = n-k = \log_2 \binom{n}{\leq d} \approx$$

$$\approx \log \binom{n}{d} \leq \log \left(\frac{ne}{d}\right)^d =$$

$$= d \cdot \log\left(\frac{ne}{d}\right) = O\left(d \log\left(\frac{n}{d}\right)\right)$$

$\square$

Limitation for Codes

Theorem (Dvi16)

For every $\delta > 0$ and large enough $n \in \mathbb{N}$, there exists a generator matrix $M \in \mathbb{F}_2^{n \times k}$ of a linear code $C \subseteq \mathbb{F}_2^n$ of dimension $k = \Theta(n)$ and distance $d = (1/2 - \delta)n$ of rigidity

$$\mathcal{R}_M^{\mathbb{F}_2}(r) \leq O\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right)$$

for every $\Omega(\log n) \leq r \leq O(n)$.

Theorem (Dvi16)

For every $\delta > 0$ and large enough $n \in \mathbb{N}$, there exists a generator matrix $M \in \mathbb{F}_2^{n \times k}$ of a linear code $C \subseteq \mathbb{F}_2^n$ of dimension $k = \Theta(n)$ and distance $d = (1/2 - \delta)n$ of rigidity

$$\mathcal{R}_M^{\mathbb{F}_2}(r) \leq O\left(\frac{n^2}{r} \cdot \log \frac{n}{r}\right)$$

for every $\Omega(\log n) \leq r \leq O(n)$.

$C' \subseteq \mathbb{F}_2^m$ be a code with optimal redundancy

$\dim(C') = k = \Theta(m)$

distance of C' is $\underline{\underline{d}}$.

$R = m - k = \Theta(d \log \frac{m}{d})$

Generator matrix

$$\underline{G} = \begin{array}{c} k \\ \boxed{\begin{array}{c} I_k \\ \hline G' \end{array}} \end{array} \begin{array}{l} k \\ m-k = R \end{array}$$

Let $\underline{B \in \mathbb{F}_2^{n \times m}}$ be a random matrix, each entry is 1 ind. w.p.

distance $\frac{c}{d}$, $c = \Theta(\log(1/\delta))$ - constant.

Study $\underline{BG} \in \mathbb{F}_2^{n \times k}$

(1) $\underline{BG}$ is non-rigid w.p. $\geq \frac{3}{4}$.

(2) $\underline{BG}$ is a generator of good ECC with distance $(\frac{1}{2}-\delta)n$. w.p. $\geq \frac{3}{4}$

$\Rightarrow$ w.p. $\geq \frac{1}{2}$ BG is a non-rigid ECC matrix.

$\Rightarrow$ $\exists$ exists $M = BG$

(1) non-rigid

(2) generates ECC

$B \in \mathbb{F}_2^{n \times m}$ - random $\quad p = \frac{c}{d}$

$G = \begin{bmatrix} I_k \\ G' \end{bmatrix}$ $\quad I_k$ - $k \times k$

G' - $R \times k$.

$B = n\left[\overset{k}{B_1} \mid \overset{R}{B_2}\right]$

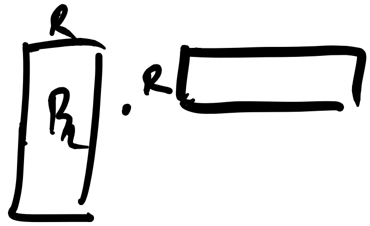
$$B \cdot G = [B_1 \ \ B_2] \cdot \begin{bmatrix} I_k \\ G' \end{bmatrix} =$$

$$= B_1 \cdot I_k + B_2 \cdot G' =$$

$$= B_1 + B_2 \cdot G' \text{ - non-rigid.}$$

B_1 is sparse w. high prob

and $\underline{B_2 \cdot G'}$ is low rank



$\operatorname{rk}(B_2 \cdot G') \le \underbrace{R}$ – low rank.

$B_1 \in \underbrace{\mathbb{F}_2}^{n \times k}$ where each entry is 1 w.p. $\frac{c}{d}$

Expect to see $n \cdot k \cdot \frac{c}{d}$

Markov's ineq: w.p. $\frac{3}{4}$

$$\# \text{ of ones} \le n \cdot k \cdot \frac{c}{d} \cdot 4 = O\left(\frac{nk}{d}\right)$$

$$R^{\mathbb{F}_2}_{BG}(R) \le O\left(\frac{nk}{d}\right) \quad k = \Theta(n)$$

$$R = O\left(d \log\left(\frac{n}{d}\right)\right) \Rightarrow d = R/\log\left(\frac{n}{R}\right)$$

$$R^{\mathbb{F}_2}_{BG}(R) \le \left(\frac{n^2}{R} \log\left(\frac{n}{R}\right)\right)$$

It remains to show that

BG generates a good code.

Every non-zero lin combination of cols of BG has high Hamming weights.

$\forall \boxed{x} \in \mathbb{F}_2^k \setminus \{0^k\}$

$\boxed{BGx}$ has Hamming w. $\geq n(\frac{1}{2}-\delta)$

Gx has high Hamming weight.

d non-zeros

$B \cdot (Gx)$ - in every coordinate has sum $\geq d$ random Bernoullis

Every coordinate is 1 w.p. $\frac{1}{2} - e^{-\Omega(d)}$

$\boxed{n \text{ coordinates}}$, each of them is 1 w.p. $\frac{1}{2} - e^{-\Omega(d)}$

By Chernoff, almost $\frac{n}{2}$ coordinates must be ones whp.

Formally, $n\left(\frac{1}{2}-\delta\right)$ coordinates must be ones w.p. $1-2^{\Omega(-n)}$

Union bound over $x \in \{0,1\}^k$

Prob of success $1-2^{\Omega(-n)} \cdot 2^k$

$$= 1-2^{\Omega(-n)} > \frac{3}{4} \quad \square$$