Let X be a random variable s.t. $E[\exp(\lambda|X|)] \leq \exp\left(\frac{\nu\lambda^2}{2}\right)$ $\forall\, \lambda$ real.

What can X be?

Lecture 14 :- Concentration with random matrices.

An 'analog' coding problem :-

Message : $x \in \mathbb{R}^n$

(Linear) Encoder : $x \mapsto \underset{\overset{\shortparallel}{y}}{Ax}$ where $A : \mathbb{R}^{m \times n}$
$m = f(n)$.

- Encoded vector Ax is transmitted along a channel which introduces noise

Received vector is $z = Ax + e$, $e \in \mathbb{R}^m$.

Decoder : $z \mapsto D(z) \in \mathbb{R}^n$.

Want :- Minimize $\frac{\|D(z) - x\|_2}{\|e\|_1}$, m/n
(Don't want to increase the dimension).

Assumption :- The 'adversary's' cost for perturbing each entry is proportional to the perturbation.

Trivial soln. :- $Ax = Tx \quad T \in \mathbb{R}^+$.

$D(z) = \frac{z}{T}$.

$$\|D(z) - x\|_2 = \left\|\frac{Tx + e}{T} - T\right\|_2 = \frac{\|e\|_2}{T} \leq \frac{\|e\|_1}{T}$$

So can get $\frac{\|D(z) - x\|_2}{\|e\|_1} \leq \frac{1}{T}$.

But just increasing power is not a reasonable solution.

Power constraint :- $\|Ax\|_2^2 \leq \|x\|_2^2$

Decoding strategy : Given z, find the 'closest' vector $\tilde{z}$ in Range(A). [in $\|\cdot\|_1$-norm].

$\forall\ \tilde{z} \in$ Range(A) there is a unique $\tilde{y} := \underset{\sim}{A^{-1}}(\tilde{z})$ s.t. $A\tilde{y} = \tilde{z}$.

inverse on the range of A.

$$D(z) = A^{-1}(\tilde{z}). \text{ where } \tilde{z} = \underset{y \in \text{Range}(A)}{\operatorname{argmin}} \|y - z\|_1$$

$\|Dz - x\|_2 \leq ?$

Q1:- How far is $\tilde{z}$ from z? (in terms of e).

We know $\exists\ y = Ax \in$ range (A) s.t.

$\|z - y\|_1 = \|e\|_1$

So, we must have $\|z-\tilde{z}\|_1 \leq \|e\|_1$.

Q2:- How large is $\|\tilde{z}-Ax\|_1$?.

$$\|\tilde{z}-Ax\|_1 = \|\tilde{z}-z+z-Ax\|_1$$
$$\leq \|\tilde{z}-z\|_1 + \|z-Ax\|_1$$
$$\leq 2\|e\|_1.$$

So therefore:- We have $\|A\tilde{x}-Ax\|_1 \leq 2\|e\|_1$.

$(\because \tilde{z} = A\tilde{x}$, where $\tilde{x} = D(z))$.

$$\|A(\tilde{x}-x)\|_1 \leq 2\|e\|_1$$

$$\|\tilde{x}-x\|_2 = \|\underbrace{A^{-1}}_{\substack{\text{inverse of } A \\ \text{on its range}}}(A(\tilde{x}-x))\|_2$$

$$\leq \|A^{-1}\|_{1\to 2} \cdot \|A(\tilde{x}-x)\|_1$$

$$\leq 2\|A^{-1}\|_{1\to 2} \cdot \|e\|_1$$

where for a matrix M, $\|M\|_{1\to 2} = \sup_{\|v\|:\|v\|_1\leq 1} \|Mv\|_2$

If we have A : $\|Ax\|_2 \leq \|x\|_2$ then the above decoding process achieves

$$\frac{\|D(Ax+e)-x\|_2}{\|e\|_1} \leq 2\|A^{-1}\|_{1\to 2}.$$

In general Suppose we had an $m \times n$ matrix B with some given $\|B\|_{2\to 2}$ then we can get a decoding process D and a matrix

$$A = \frac{B}{\|B\|_{2\to 2}} \quad \text{s.t.}$$

$$\frac{\|D(Ax+e) - x\|_2}{\|e\|_1} \leq 2\|B\|_{2\to 2} \cdot \|B^{-1}\|_{1\to 2}.$$

$$\boxed{\text{So the question is: How small can } \|B\|_{2\to 2} \cdot \|B^{-1}\|_{1\to 2} \text{ be for an } m\times n \text{ matrix?}}$$

Vector space view: Consider the case where B has orthonormal columns. $\|B\|_{2\to 2} = 1$, B, B^{-1} are 'rotations': rotating a copy of $\mathbb{R}^n$ sitting inside $\mathbb{R}^m$ into a different space. So the preserve 2-norms of vectors. Then this $\|B^{-1}\|_{1\to 2}$ is just

$$\max_{\substack{x \in \text{range}(B) \\ x \neq 0}} \frac{\|B^{-1}x\|_2}{\|x\|_1}$$

$$= \max_{\substack{x \in \text{range}(B) \\ x \neq 0}} \frac{\|x\|_2}{\|x\|_1}, \text{ because } B^{-1} \text{ is a 'rotation' so it preserves 2-norms.}$$

$$= \frac{1}{\underbrace{\min_{\substack{x \in \text{range}(B) \\ x \neq 0}} \frac{\|x\|_1}{\|x\|_2}}_{\text{Could potentially be as large as } \sqrt{m}.}}$$

So, the question is: with $m = \Theta(n)$ can we get.

$$\min_{\substack{x \in \text{range}(B) \\ x \neq 0}} \frac{\|x\|_1}{\|x\|_2} = \Omega(\sqrt{n}) \, ?$$

- 'Euclidean section': This means that in range(B), which a subspace of $\mathbb{R}^m$ of dimension $\Theta(m)$, we have

$$\underbrace{\Theta(\sqrt{m}) \|x\|_2}_{} \leq \|x\|_1 \underset{\text{Trivially.}}{\leq} \sqrt{m} \|x\|_2$$

On this space, upto constant factors, $\frac{\|x\|_1}{\|x\|_2} \sim \sqrt{m}$.

['Dvoretzky theorem' (This is perhaps the simplest case): Inside $(\mathbb{R}^m, \ell_1)$ sits a subspace of dimension $\Theta(m)$ when ℓ_1 and ℓ_2 are essentially similar.]

(Q) How do we find these sections / such a B?

$$\|B\|_{2\to 2} \leq 1 \longleftarrow$$

and $\|B^{-1}\|_{1\to 2} \leq \frac{1}{c_2\sqrt{n}}$. $B \in \mathbb{R}^{m\times n}$ with $m \leq c_1 n$.

with $c_1, c_2 > 0$ constants.

<u>Idea</u>:- Maybe a 'random' B works?

$$B_{ij} = \frac{\pm 1}{\sqrt{m}} \quad \text{with probability } 1/2 \text{ independently.}$$

Let's fix $u \in \mathbb{R}^n$, $\|u\|_2 = 1$.

$$E\left[\|Bu\|_2^2\right] = E\left[\sum_{j=1}^{m} |(Bu)_j|^2\right]$$
$$= \sum_{j=1}^{m} E\left[\left(\sum_{i=1}^{n} B_{ji} u_i\right)^2\right]$$
$$= \frac{1}{m}\sum_{j=1}^{m} \sum_{i=1}^{n} u_i^2 = 1. \quad \text{—} (*)$$

Goal is to prove that with this choice
$$\|B\|_{2\to 2} \leq c \quad \text{with prob} \geq 1/4.$$
for some constant c.

$$\iff \quad \|Bu\|_2^2 \leq c \quad \forall\, u \in \mathbb{R}^n, \|u\|_2 = 1.$$

We do have. $\forall\, u \in \mathbb{R}^n$, $\|u\|_2 = 1$.

$(*)$ $$\Pr\left[\|Bu\|_2 \geq t\right] \leq \frac{1}{t^2} \quad \forall\, t \in \mathbb{R}.$$

But we cannot take a union bound over all u. There are infinitely many u's.

- Also $(*)$ is not the best → This can be made into a Hoeffding like bound using the fact that for any fixed u, each entry of Bu is Subgaussian (and also, different entries of Bu are independent), with parameter $\Theta(\frac{1}{m})$

: $E[\exp(\lambda (Bu)_i)] \leq \exp(\frac{c\lambda^2}{m})$

for some fixed constant c.

From this, and an argument similar to Problem 4 of HW 1, we get that. $\forall u$, $\|u\| = 1$.

(1) $\Pr[\|Bu\|_2 > c_1 + t] \leq \exp(-c_2 m \cdot t^2) \quad \forall t \geq 0$

for some c_1 and c_2 positive constants.

But still, we need to take an union bound over infinitely many u.

ε-net. Find a finite collection F_ε of u, $\|u\| \leq 1$ s.t. every u with $\|u\| \leq 1$ is close (ε-close) to some element of F_ε.

Why does this help? Suppose F_ε is a finite s.t.

(1) $\forall u$ s.t. $\|u\|_2 \leq 1$, there exists $v \in F_\varepsilon$ s.t. $\|u - v\|_2 \leq \varepsilon$.

(2) $\exists c > 0$ s.t. $\|Bv\|_2 \leq c \quad \forall v \in F_\varepsilon$.

Suppose that $\|B\|_{2\to 2} = M$. With ① and ② we get the following: Take any u s.t. $\|u\|_2 \leq 1$.

Then $\exists v \in F_\varepsilon$ s.t. $\|u-v\|_2 \leq \varepsilon$. Now,

$$\|Bu\|_2 = \|B(v+u-v)\|_2$$

$$\leq \|Bv\|_2 + \|B(u-v)\|_2$$

$$\leq \underbrace{c}_{\text{from ②}} + \underbrace{M\|u-v\|_2}_{\because \|B\|_{2\to 2} = M.}$$

$$\leq c + M\varepsilon \quad (\because \|u-v\|_2 \leq \varepsilon.$$

So, we see that

$$\|Bu\|_2 \leq c + M\varepsilon \quad \forall\ u,\ \|u\|_2 \leq 1.$$

Then

$$M = \sup_{u : \|u\|_2 \leq 1} \|Bu\|_2 \leq c + M\varepsilon.$$

$$\Rightarrow \quad M(1-\varepsilon) \leq c$$

$$\Rightarrow \quad M \leq \frac{c}{1-\varepsilon} \qquad (\varepsilon < 1).$$

So we have :-

Lemma :- Let F_ε be an ε-net of $\{x \mid \|x\|_2 \leq 1\}$. If B satisfies $\|Bv\| \leq c \ \forall\ v \in F_\varepsilon$ then

$$\|B\|_{2\to 2} \leq \frac{c}{1-\varepsilon}, \text{ assuming } \varepsilon < 1.$$

Thus if we find F_ε of size $\exp(O(m))$, we will get $\|B\|_{2\to 2} \leq c$ wp > 0, using ① above.

Q :- How does one find F_ε ? (Or show existence)

Process Suppose we have picked so far points $S = \{p_1, \dots p_k\}$ To pick the next point (if necessary) we pick a point in the ball whose distance to all the previously picked points is greater than ε.

— If we can't find such a point, our collection is already an ε-net?

Invariant that we maintain.

$$\|p-q\|_2 > \varepsilon \quad \forall \text{ two distinct points } p, q \in S$$

when $S \subseteq \{x \mid \|x\| \leq 1\}$.

So, (1) the balls of radius $\frac{\varepsilon}{2}$ around distinct points in S cannot intersect!

(2) Since $p \in S \Rightarrow \|p\| \leq 1$, the ball of radius $\frac{\varepsilon}{2}$ around p lies in $\{x \mid \|x\| \leq 1 + \frac{\varepsilon}{2}\}$.

So (1) and (2) give.

$$\mathrm{Vol}\left(\bigcup_{p \in S} \mathrm{Ball}(p, \varepsilon/2)\right)$$

$$= \sum_{p \in S} \mathrm{Vol}\left(\mathrm{Ball}(p, \varepsilon/2)\right) = \underbrace{|S| \cdot \mathrm{Vol}\left(\mathrm{Ball}\left(0, \tfrac{\varepsilon}{2}\right)\right)}_{\because \text{ the balls are just translations.}}$$

$$\leq \mathrm{Vol}\left(\mathrm{Ball}\left(0, 1 + \tfrac{\varepsilon}{2}\right)\right).$$

$$|S| \cdot \mathrm{Vol}\left(\mathrm{Ball}\left(0, \frac{\varepsilon}{2}\right)\right) \leqslant \mathrm{Vol}\left(\mathrm{Ball}\left(0, 1+\frac{\varepsilon}{2}\right)\right)$$

$$\therefore \quad |S| \leqslant \frac{\mathrm{Vol}\left(\mathrm{Ball}\left(0, 1+\frac{\varepsilon}{2}\right)\right)}{\mathrm{Vol}\left(\mathrm{Ball}\left(0, \frac{\varepsilon}{2}\right)\right)}.$$

Now

$$\mathrm{Vol}(\mathrm{Ball}(0, r)) = \mathrm{Vol}(\mathrm{Ball}(0,1)) \cdot r^m$$

when Vol is in $\mathbb{R}^m$.

So, we get $|S| \leqslant \left(\frac{2}{\varepsilon} + 1\right)^m \leqslant \left(\frac{3}{\varepsilon}\right)^m$

when $\varepsilon \leqslant 1$.

This, along with ①, establishes the easy part that $\|B\|_{2\to 2} \leqslant C$ with prob. > 0 when $m > C_1 n$ (where C, C_1 are absolute constants).

- $\|B^{-1}\|_{1\to 2}$ can be handled by giving a high probability lower bound on $\|Bv\|_1$, $\|v\|_2 = 1$, and then a slightly different ε-net argument.