TAMU SUMIFRAS workshop

"Constructive regularization of the random matrix norm"

08/18/18

---

## ① Optimal norm order

$A$ $n\times n$ random matrix
independent identically distr. entries

$\|A\| = \|A\|_{2\to 2} = \sup_{x\in S^{n-1}} \|Ax\|_2 = \sup_{u,v\in S^{n-1}} |\sum A_{ij}u_iv_j|$

$= s_1(A)$ - operator norm

$\sqrt{n}$ - for large class of matrices

- 

Wigner semicircle law
$A_{ij} \sim N(0,1)$

- $\|A\| = (2+o(1))\sqrt{n}$ w.h.p. as $n\to\infty$
$\mathbb{E} A_{ij} = 0$, $\mathbb{E} A_{ij}^4 < \infty$ Bai-Yin

- $\mathbb{P}\{\|A\| \le t\sqrt{n}\} \ge 1 - \exp(-c_0t^2n)$ $\forall t \ge C_0$
$A_{ij}$ has gaussian-like tails
(non-asymptotic-type bounds)

$\sqrt{n}$ - best possible
Indeed, $\mathbb{E}A_{ij}^2 = c \Rightarrow \mathbb{E}(\sum A_{ij}^2) = cn$
$\|A\| \ge \max \|A_i\|_2 \ge \mathbb{E}\|A_i\|_2 \sim \sqrt{n}$

$\sqrt{n}$ is violated if $\mathbb{E}A_{ij}^4 = \infty$

- Silverstein example
- Litvak-Spector: constructed family of distributions $\mathbb{E}A_{ij}^2 = 1$
s.t. $\|A\| \sim O(n^\alpha)$ $\forall \alpha \le 1$
with prob $\sim 1/2$)
- Ex: $\begin{bmatrix} c & c & c & c \\ c & c & c & c \\ c & c & c & c \end{bmatrix} \to$ $\mathbb{E}A_{ij} = c +$ tight concentration
$\|A\| \sim cn$

## ② Norm regularization

What is in the structure of a typical matrix causes blow up of the norm?

Local regularization:

- make changes in a $\varepsilon$-fraction of the entries of the matrix
($\varepsilon n \times \varepsilon n$ sub-block) $\longrightarrow \tilde{A}$
- $\|\tilde{A}\| \sim O(\sqrt{n})$ with high probability

$\boxed{\text{Thm 1}}$ (R + Vershynin)

Local regularization is possible

$\Updownarrow$

$\mathbb{E} A_{ij} = 0$ and $\mathbb{E} A_{ij}^2 = 1$

If $A$ is $n \times n$ RM with iid entries $A_{ij}$: $\mathbb{E} A_{ij} = 0$, $\mathbb{E} A_{ij}^2 = 1$. Then $\forall \varepsilon \in (0, \frac{1}{6}]$ w/prob $1 - 11 \exp(-\frac{\varepsilon n}{12})$ there exists an $\varepsilon n \times \varepsilon n$ submatrix replacing with zeros $\to \tilde{A}$:

$$\|\tilde{A}\| \le C \cdot \sqrt{c_\varepsilon \cdot n} \qquad c_\varepsilon = \frac{\log \varepsilon^{-1}}{\varepsilon}$$

* optimal order in $n$
* almost optimal order in $\varepsilon$ ($\varepsilon^{-1}$ is optimal)
* inconstructive

reason: first $\|\tilde{A}\|_{\infty \to 2}$ is estimated

$\downarrow$ Grothendieck-Pietsch factorization for matrices

then $\|\tilde{\tilde{A}}\|$

③ Towards constructive regularization

What can we change (zero out) to restore $\|\tilde{A}\| \sim \sqrt{n}$?

Idea ①: can we remove $\varepsilon n$ largest entries?

* works only if $\mathbb{E} A_{ij}^{2+\delta} < \infty$ for some $\delta > 0$

One way to check this:

$\boxed{\text{Thm}}$ (Bandeira-Van Handel)

$$\mathbb{P}\left(\|X\| \ge (1+\gamma) \cdot \sigma + t\right) \le n \cdot \exp\left(-\frac{t^2}{c_\gamma \cdot \sigma_*^2}\right)$$

any $\gamma > 0$

$\sigma^2 = \max_i \sum_j \mathbb{E}(x_{ij}^2)$ — max expected row/col norm

$\sigma_*^2 = \max_{ij} \|X_{ij}\|_\infty$ — max entry

truncation on the level $|A_{ij}| \sim \frac{\sqrt{n}}{\sqrt{\log n}}$

* 2 finite moments — counterexample:

~~[illegible]~~ iid $0, \pm\sqrt{n}$ entries

$\mathbb{P}(A_{ij} = \sqrt{n}) = \frac{1}{2n}$, $\mathbb{P}(A_{ij} = -\sqrt{n}) = \frac{1}{2n}$

$\to \|A_{ij}\| \gg \sqrt{n}$ (direct computation: there is a "heavy" row $\|A_i\|_2^2 = n \cdot \#(\text{non-zeros}) \sim n \frac{\log n}{\log\log n}$)

$\to$ # non-zeros in the matrix $\sim n$

(cannot be placed in $\varepsilon n \times \varepsilon n$ submatrix with high prob)

$\to$ we cannot zero out entries "by size"; need to find the most dense part of the matrix (in realization)

* matrix Bernstein inequality

$\|\tilde{A}\| \le \frac{\sqrt{n}}{\sqrt{\varepsilon}} \cdot \ln n$ for 2 moments

Can we do better?

Not only to improve extra factor, but make regularization that addresses real obstructions to the good norm (which are not only in element size, but location)

**Thm 2**

$A$ - $n\times n$ random matrix with iid symmetrically distributed entries $\mathbb{E}A_{ij}^2 = 1$
For $\forall \varepsilon \in (0, \frac{1}{6}]$, $r \ge 1$ with prob $\ge 1 - n^{0.1-r}$
if we zero out $\varepsilon n$ rows & columns with largest $L_2$-norms $\to \tilde{A}$:

$\|\tilde{A}\| \le Cr\sqrt{C_\varepsilon n \log\log n}$ $\quad C_\varepsilon = \frac{\log \varepsilon^{-1}}{\varepsilon}$

* almost optimal in $n$
* requires additional symmetry assump.
* gives simple regularization procedure + description of the obstructions

Remark Equivalent regularization way:
zero out any product subset of entries, such that all rows & col's will have $L_2$-norm $\lesssim \sqrt{C\varepsilon n}$
(by Thm 1 this is a local regularization)
(allows to zero out only a submatrix - See ⑥)

④ Bernoulli random matrices

motivated "heavy rows & cols" criterium

Consider $n\times n$ 0-1 Bernoulli matrix $B$: $\mathbb{P}(B_{ij}=1) = p$

$\mathbb{E}(B_{ij} - \mathbb{E}B_{ij})^2 \sim p \to$ optimal norm $\|B - \mathbb{E}B\| \sim \sqrt{np}$

Known results:

if $p \gtrsim \sqrt{\ln n}$ $\quad \|B - \mathbb{E}B\| \sim \sqrt{np}$ whp [Feige - Ofek]

if $p \ll \sqrt{\ln n}$ No (Krievelevich - Sudakov counterexample)

! case $p \sim \frac{1}{n}$ $\longleftrightarrow$ exactly 2 finite moments

! # of non-zeros in a row/col $\longleftrightarrow$ large $L_2$ norm of a row/col

$e_i^{row/col}(B)$ := number of non-zeros in a row/column

$\mathbb{E}\, e_i^{row} = np$

More known results:

[FO] zero out rows & cols such that $e_i^{row/col}(B) > C\cdot\mathbb{E}\, e_i^{row}(B) \to \tilde{B}$

$\|\tilde{B} - \mathbb{E}B\| \sim \sqrt{np}$ w. high prob ⊛

(Feige - Ofek)

[LLV] reweight or partially zero out some rows & cols s.t. $e_i^{row/col}(\tilde{B}) \le C np$

⊛ (Le-Levina-Vershynin)

! More delicate regularization is better for graph applications, where "heavy" rows might mean "more important"

B) Thm 2 proof ideas

High-level idea: split

$$|A_{ij}| \sim \sum_k 2^k \cdot \mathbb{1}\{|A_{ij}| \in (2^{k-1}, 2^k]\} = \sum 2^k B^k$$

and apply Bernoulli results at each "level"

Three complications:

1) regularization of rows/cols of A does not directly controls rows/cols of different "levels"

**Proposition** $B$ - Bernoulli 0-1 iid entries $B_{ij}$: $\mathbb{P}\{B_{ij}=1\} = p$

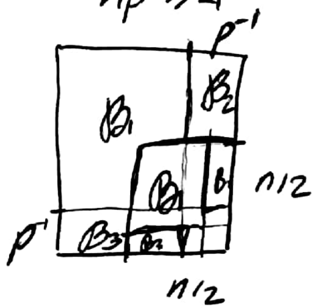
$\forall n \ge 4$ $\forall p \ge 0$ $\forall r \ge 1$ w/prob $1 - 3n^{-r}$ all entries of $B = \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \mathcal{B}_3$:

- $e_i^{row}(\mathcal{B}_1) \lesssim rnp$, $e_i^{col}(\mathcal{B}_1) \lesssim rnp$
- $e_i^{row}(\mathcal{B}_2) \lesssim r$
- $e_i^{col}(\mathcal{B}_3) \lesssim r$

(follows from LLV)

recursive decomposition by Chernoff's inequality

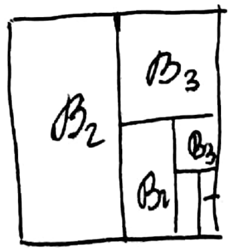
$np \ge 4$



- in every $n2^{-k} \times n2^{-k}$ submatrix of $B$ there are at most $\frac{1}{p}2^{-k}$ rows & cols with $> C_1 np$ entries

$np \ge 4$

- in every $n2^{-k} \times \frac{1}{p}2^{-k}$ submatrix of $B$ there are at most $\frac{1}{4}n2^{-k}$ columns with $> C_2 r$ non-zero entries



$np < 4$

$\boxed{\text{Result}}$ on $\mathcal{B}_1$ apply Bernoulli results

on $\mathcal{B}_2$ ($\mathcal{B}_3$)

$\|X\| \le \max_i \|X_i\|_2 \cdot [\max_j e_j^{row}(X)]^{1/2}$

! # of levels need to be small

2) We cannot estimate $\| |A| \|$ - too large
How to pass to absolute values?

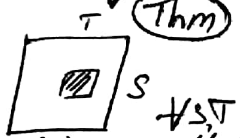
Idea (Feige-Ofek): $B$-Bernoulli

$$\Big|\sum_{ij} (B_{ij} - \mathbb{E}B_{ij})\, u_i v_j\Big| \le \qquad u,v \in S^{n-1}$$

$$\Big|\sum_{\text{light } ij} (B_{ij} - \mathbb{E}B_{ij}) u_i v_j\Big| + \Big|\sum_{\text{non-light}} B_{ij} u_i v_j\Big| + \Big|\sum_{\text{non-light}} \mathbb{E}B_{ij} u_i v_j\Big|$$

Bernstein

Thm



$\forall S,T$
$e(S,T)$ are well bdd

$p$

$\sum u_i^2 v_j^2 = 1 \Rightarrow$
$\#$ non-light $\le \frac{n}{p}$
$\frac{\sqrt{n}}{\sqrt{p}} \cdot p \le \sqrt{np}$

$$\boxed{\text{light indices} := \Big\{(i,j) : |u_i v_j| \le \frac{\sqrt{p}}{\sqrt{n}}\Big\}}$$

Our case: $\boxed{\text{light} := \Big\{(i,j) : |u_i v_j|\, |A_{ij}| \le \frac{2}{\sqrt{n}}\Big\}}$

$$\Big|\sum A_{ij} u_i v_j\Big| \le$$

$$\Big|\sum_{\text{light}} A_{ij} u_i v_j\Big| + \sum_{\text{non-light}} |A_{ij}| |u_i v_j|$$

Bernstein

$$\le \sum_{ij} \sum_k 2^k B^k_\varepsilon u_i v_j \le$$

$$\sum_k 2^k \sqrt{n p_k} \le$$

$$\sqrt{n} \cdot \underbrace{\sum 2^{2k} p_k}_{\le 1} \cdot \boxed{\#\text{ levels}}$$

3) Consider as few levels as possible

$$2^k \in \left[\sqrt{\frac{cn}{\ln n}},\ \sqrt{nC_\varepsilon}\right]$$

bigger entries — by Thm 1 zeroed
smaller entries — by Bandeira-vanHandel
deletion is not needed

$$k := k_1 - k_0 \le \frac{1}{2}\Big(\log_2(C_\varepsilon n) - \log_2 \frac{cn}{\ln n}\Big) \le$$

$$\le \frac{1}{2} \log \frac{C_\varepsilon n \cdot \ln n}{cn} \sim \log\log n$$

! Symmetry is needed only to keep zero on various truncations

## ⑥ Regularization on sub-matrix

Need to find the most "dense" part of submatrix

* Enough to find a column subset (only)

Idea (R+Tikhomirov): to find a small column subset s.t. all other rows & cols have bounded $L_2$-norms
+
Apply Thm 2

**Alg1**

- assign weights to the entries $w \in [0,1]$ row-wise:

  if $\begin{cases} e_i^{row} > L n p_k \\ B_{ij} \neq 0 \end{cases} \Rightarrow w_{ij} := \frac{L n p_k}{e_i^{row}}$

  otherwise $\Rightarrow w_{ij} := 1$

- assign weights to columns — multiply column-wise
- zero out columns with the weights < const

By (R-Tikhomirov) there are at most $\varepsilon n p_k$ such columns w/prob $1 - \exp(-n \exp(-L p_k n))$

For convergence, need to choose $p_k = 2^{-k}$

$$p_k = \underset{(k_0)}{\frac{\varepsilon}{4n}} \ldots \underset{(k_1)}{\frac{\ln n}{16 C_\varepsilon n}}$$

Quantiles. $q_k := \min\{t : \mathbb{P}\{A_{ij}^2 > q_k\} = 2^{-k}\}$

- $e_i^{row}(B^k) = \#$ entries of $A$ in row $i$ : $A_{ij}^2 \in (q_{k-1}, q_k]$

This algorithm requires the knowledge of $q_k$ (quantiles of $A_{ij}^2$)

**Algorithm:**

1. for every level $k = k_0 \ldots k_1$ **Alg 1**
2. $J = \bigcup_k$ ($\varepsilon n p_k$ exceptional cols from level $k$)

   $|J| \le \varepsilon n$

- for $q_k \le \frac{\sqrt{n}}{\sqrt{\ln n}}$ — Bernstein inequality implies ~~bdd~~ rows & cols

3. for $q_k > \frac{\sqrt{n}}{\sqrt{\varepsilon}}$ — zero out everything (at most $\varepsilon n$ w.h.p. – Markov)

- norm is OK – THM 2

4. Repeat for transpose — get $I$: $|I| \le \varepsilon n$
5. Zero out $\boxed{A_{I \times J}}$

**Corollary 1** $A$ – $n \times n$ random matrix with iid symmetrically distributed entries $\mathbb{E} A_{ij}^2 = 1$

For $\forall \varepsilon \in (0, \frac{1}{6}]$, $r \ge 1$, $\tilde{A}$ obtained via algorithm with probability at least $1 - C n^{0.1 - r}$

$$\|\tilde{A}\| \le r^{3/2} \sqrt{c_\varepsilon \, n \ln \ln n}$$