Global vs local obstructions for the regularization of the norm of random matrices

Liza Rebrova

joint with Roman Vershynin

University of Michigan

CMS Sectional meeting, Edmonton, June 25-27, 2016

Global vs local obstructions

Setting: Object lacking some good property
Question: can we gain this property by a local change of an object?

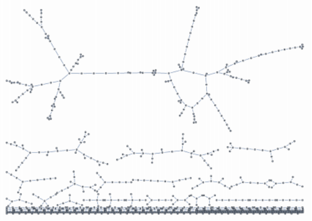
We can ask this for various structures/objects and various properties.

Example

Object - Erdös-Rényi random graph $G(n, p)$;
property - connectivity; local change - in $o(n)$ edges
When $p \sim \frac{1}{n}$ structure properties change: a giant component appear:

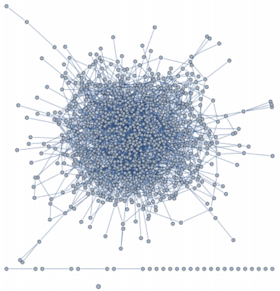
1. before threshold - lots of connected components
2. after threshold - a giant component + a few other components

Global vs local obstructions - Example



$p < \frac{1}{n}$

There are $O(n)$ small components, we cannot connect them all by $o(n)$ edges – obstructions to connectivity are "global"



$p > \frac{1}{n}$

A giant component and $log(n)$ other components, we can connect everything by a short cycle – obstructions are "local"

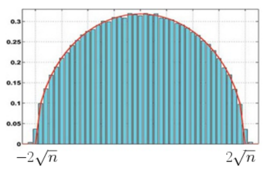
(Pictures are taken from A. Novozhilov's course in Mathematics of Networks, NDSU)

Another example - norm regularization problem

Object - $n \times n$ random matrix A with i.i.d. entries a_{ij}
Property - $\|A\| \lesssim C\sqrt{n}$ with high probability
Local change - in a small $\varepsilon n \times \varepsilon n$ submatrix $A_0 \subset A$

Example

- If $a_{ij} \sim N(0,1)$, then $\|A\| \simeq 2\sqrt{n}$.
- If a_{ij} are subgaussian, then also $\|A\| \simeq O(\sqrt{n})$



But: if just $\mathbb{E}a_{ij}^2 = 1$, then there are examples $\|A\| \sim O(n^{2/\alpha})$ for any $\alpha \geq 2$ with probability at least $1/2$ (A.Litvak, S.Spector)

Question: If $\|A\| \gg \sqrt{n}$ w.h.p, is it a global or local obstruction?

Norm regularization problem

Example

If a_{ij} are not mean zero: $\mathbb{E}a_{ij} \sim 1$, then $\|A\| \geq n$, and the problem is global.

So, we assume $\mathbb{E}a_{ij} = 0$.
Can we improve the norm of a random matrix by deletion of its small sub-matrix?

Theorem (L.R-R.Vershynin, informal statement)

A is a random square matrix with i.i.d. centered elements a_{ij}.

- *if a_{ij} have finite variance $\therefore$ local obstructions*
- *if not $\therefore$ global obstructions*

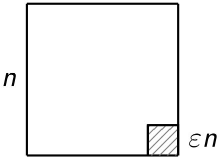
Theorem (Part 1: local obstructions)

Let A be a random $n \times n$ matrix with i.i.d. elements, $\mathbb{E}a_{ij} = 0$, $\mathbb{E}a_{ij}^2 = 1$. Then for any $\varepsilon \in (0, \frac{1}{2}]$ with probability

$$1 - 11\exp(-\varepsilon n/6)$$

there exists an $\varepsilon n \times \varepsilon n$ sub-matrix $A_0 \subset A$, such that

$$\|A \setminus A_0\| \leq C_\varepsilon \sqrt{n} = O\left(\frac{\ln(\varepsilon^{-1})}{\sqrt{\varepsilon}}\right)\sqrt{n}.$$



We need to delete just a small sub-matrix

Application to the random graphs

Let $G(n,p)$ – Erdös-Rényi random graph. What do we know about the norm of its (scaled) adjacency matrix $A = \frac{1}{\sqrt{p}} \cdot$Adjacency ?

- If $p \geq \frac{\log n}{n}$, then $\|A - \mathbb{E}A\| \lesssim \sqrt{n}$. This is good! ($\mathbb{E}A \sim n$, i.e. we have concentration)
- If $p < \frac{\log n}{n}$, especially if $p \lesssim \frac{1}{n}$, then $\|A - \mathbb{E}A\| \gg \sqrt{n}$ (sparse graphs do not concentrate)

We want to regularize the graph, such that new adjacency matrix satisfies

$$\|A' - \mathbb{E}A\| \lesssim \sqrt{n}.$$

What are the obstructions for the regularization?

Application - random graphs

What are the obstructions for the regularization?

1. Local/global - ?

   Obstructions are local (known, Feige-Ofek)

2. What causes the obstructions (in terms of graph)?

   Idea: obstructions are in high-degree vertices.
   For the regularization it is enough to

   - U.Feige-E.Ofek: delete all high-degree vertices ($> 10\cdot$expected degree)
   - C.Le-R.Vershynin: reweight or delete some of the edges adjacent to high-degree vertices (to make all the degrees bounded)
   - L.R-R.Vershynin (Bernoulli case corollary): delete a small $\varepsilon n \times \varepsilon n$ sub-graph

Dependence on ε

Optimal dependence would be $C_\varepsilon = O(\frac{1}{\sqrt{\varepsilon}})$.
To see this, consider

- $\varepsilon \lesssim \frac{1}{n}$ and $\|A\| = O(n) = O(\sqrt{n}/\sqrt{\varepsilon})$ with probability 1/2, or
- any $\varepsilon \leq 1/2$ and Bernoulli matrix A with rare $\frac{\sqrt{n}}{\sqrt{\varepsilon}}$ spikes

So, our argument gives log-optimal dependence $C_\varepsilon = O(\frac{\ln(\varepsilon^{-1})}{\sqrt{\varepsilon}})$

Example (Bernoulli case)

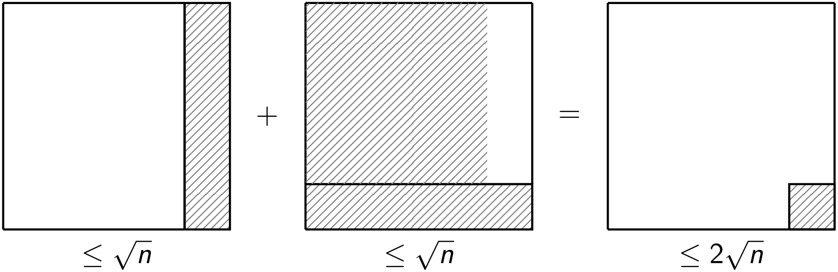
However, in "good Bernoulli case" dependence is better. Let A be a square matrix with i.i.d. elements distributed like

$$a_{ij} = \begin{cases} \frac{1}{\sqrt{p}} & \text{with probability } p \\ 0 & \text{otherwise} \end{cases}, \quad p \cdot n = d = O(1) \geq 4$$

Then the theorem is hold with $C_\varepsilon = O(\ln(\varepsilon^{-1}))$.

Observation 0: εn columns cut

It is enough to show that εn -columns cut regularizes the norm:



Shaded sub-matrices are deleted, white sub-blocks have good norm

Step 1: Regularization of $\|.\|_{\infty\to 2}$ norm

Proposition (L.R-K.Tikhomirov)

Let $\varepsilon \in (0,1]$ and A be a matrix as above. Then with high probability there exists a diagonal matrix $D = (d_{ij})_{i,j=1}^n$, such that

(1) $\|AD\|_{\infty\to 2} \leq \frac{C}{\sqrt{\varepsilon}} n$

(2) $d_{ii} \in (0,1)$ *and* $\prod_{i=1}^{n} d_{ii} \geq \exp(-\varepsilon n)$.

Idea: (1) d_{ii} is a weight we put on i-th column of the matrix and (2) there are few small d_{ii}: all but εn columns have weights larger than a small constant β.
We delete these εn columns to get:

$$\|A \setminus \ (\varepsilon n \text{ "bad" columns})\ \|_{\infty\to 2} \leq \frac{1}{\beta} \cdot \frac{C}{\sqrt{\varepsilon}} n.$$

Step 2: Grothendieck-Pietsch factorization

Improving the standard $\frac{1}{\sqrt{n}}\|B\|_{\infty\to 2} \le \|B\| \le \|B\|_{\infty\to 2}$, we use

Theorem (Grothendieck-Pietsch, sub-matrix version)

Let B be a $n \times n_1$ real matrix and $\delta > 0$. Then there exists $J \subset [n_1]$ with $|J| \ge (1-\varepsilon)n_1$ such that

$$\|B_{[n]\times J}\| \le \frac{2\|B\|_{\infty\to 2}}{\sqrt{\varepsilon n_1}}.$$

We use it with $n_1 = (1-\varepsilon)n$ to find $|J| \ge (1-2\varepsilon)n$, such that

$$\|A_{[n]\times J}\| \le \frac{2\|A \setminus A'\|_{\infty\to 2}}{\sqrt{\varepsilon n}} \le \frac{C' n/\sqrt{\varepsilon}}{\sqrt{\varepsilon n}} \le C_\varepsilon \sqrt{n}.$$

Step 2: Grothendieck-Pietsch factorization

Improving the standard $\frac{1}{\sqrt{n}}\|B\|_{\infty\to 2} \le \|B\| \le \|B\|_{\infty\to 2}$, we use

Theorem (Grothendieck-Pietsch, sub-matrix version)

Let B be a $n \times n_1$ real matrix and $\delta > 0$. Then there exists $J \subset [n_1]$ with $|J| \ge (1-\varepsilon)n_1$ such that

$$\|B_{[n]\times J}\| \le \frac{2\|B\|_{\infty\to 2}}{\sqrt{\varepsilon n_1}}.$$

We use it with $n_1 = (1-\varepsilon)n$ to find $|J| \ge (1-2\varepsilon)n$, such that

$$\|A_{[n]\times J}\| \le \frac{2\|A \setminus A'\|_{\infty\to 2}}{\sqrt{\varepsilon n}} \le \frac{C' n/\sqrt{\varepsilon}}{\sqrt{\varepsilon n}} \le C_\varepsilon \sqrt{n}.$$

Problem! We got $C_\varepsilon \sim \frac{1}{\sqrt{\varepsilon}}\frac{1}{\sqrt{\varepsilon}} = \frac{1}{\varepsilon} \gg \frac{1}{\sqrt{\varepsilon}}$ - optimal constant

Step 3: fighting for a good C_ε

Solution (for bounded entries): consider only "small" entries of the matrix $|a_{ij}| \lesssim \sqrt{n}$, then on Step 1 $\|A \setminus A'\|_{\infty\to 2} \le C\sqrt{\ln(\varepsilon^{-1})}n$.

Hence, for a matrix A such that $\mathbb{E}a_{ij} = 0, \mathbb{E}{a_{ij}}^2 \le 1$, $|a_{ij}| \le \frac{\sqrt{n}}{2}$ a.s.:

$$\|A \setminus A'\| \le C\frac{\sqrt{\ln(\varepsilon^{-1})}}{\sqrt{\varepsilon}}\sqrt{n}.$$

General case:

$$A = A \cdot \mathbb{1}_{\{|a_{ij}| \lesssim \sqrt{n}\}} + A \cdot \mathbb{1}_{\{\sqrt{n} \lesssim |a_{ij}| \lesssim \frac{\sqrt{n}}{\sqrt{\varepsilon}}\}} + A \cdot \mathbb{1}_{\{\frac{\sqrt{n}}{\sqrt{\varepsilon}} \lesssim |a_{ij}|\}}$$

↓ sparsity and size (most non-zero elements belong to sparse rows)

↓ very sparse (εn non-zero elements)

Theorem (Part 2: global obstructions)

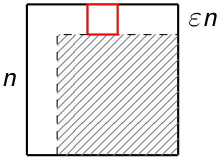
Let A is an $n \times n$ matrix with i.i.d. entries, such that

- $\mathbb{E}a_{ij}^2 \geq M$,
- $|a_{ij}| \leq \sqrt{n}$ *almost surely.*

If $M = M(C, \varepsilon)$ is a large enough constant, then any $\varepsilon n \times \varepsilon n$ sub-matrix A_0 has large Frobenius (and, hence, operator) norm

$$\sqrt{n}\|A_0\| \geq \|A_0\|_F \geq Cn,$$

with probability at least $1 - \exp(-\varepsilon n)$.



Any $\varepsilon n \times \varepsilon n$ sub-matrix blows up the norm

Proof idea:

- split elements onto levels "by size":

$$\|A_0\|_F^2 = \sum_{a_{ij}\in A_0} a_{ij}^2 = \sum_{k=0}^{\infty} \sum_{a_{ij}\in A_0} a_{ij}^2 \mathbb{1}_{\{2^k \le a_{ij}^2 < 2^{k+1}\}}$$

- argue that the majority of the levels in any $\varepsilon n \times \varepsilon n$ sub-block contain many non-zero elements (use Chernoff's inequality).

■

Conclusion:

Theorem (informal statement)

A is a random square matrix with i.i.d. centered elements a_{ij},

- *if $\mathbb{E}a_{ij}^2 < M$ $\therefore$ there are local obstructions*
- *if not, and entries are $\sqrt{n}$-bounded $\therefore$ there are global obstructions*

for the regularization of the operator norm $\|A\|$.

Thanks for your attention! :)