Random matrices with independent rows or columns

Nicole Tomczak-Jaegermann

Phenomena in High Dimensions
in geometric analysis, random matrices,
and computational geometry

Roscoff, June 25–29, 2012

Project on random matrices with indepeendent rows or columns, norms and condition numbers of their submatrices.

Involved in this project:
Alain Pajor and various subsets of:

Radosław Adamczak,
Olivier Guédon,
Rafał Latała,
Alexander Litvak,
Krzysztof Oleszkiewicz,
Nicole Tomczak-Jaegermann

Let $X = (X(1), \dots, X(N))$ be a random vector in $\mathbb{R}^N$ with full dimensional support. We say that the distribution of X is

- *logaritmically concave*, if X has density of the form $e^{-h(x)}$ with $h\colon \mathbb{R}^N \to (-\infty, \infty]$ convex (one of equivalent definitions by C. Borell)
- *isotropic*, if $\mathbb{E}X(i) = 0$ and $\mathbb{E}X(i)X(j) = \delta_{i,j}$.

For $x \in \mathbb{R}^N$ we put

- $|x| = \|x\|_2 = \left(\sum_{i=1}^N x_i^2\right)^{1/2}$
- $P_I x$ - canonical projection of x onto $\{y \in \mathbb{R}^N : \ \text{supp}(y) \subset I\}$, $I \subset \{1, \dots, N\}$.

For integers $k \leqslant \ell$ we we use the shorthand notation $[k, \ell] = \{k, \dots, \ell\}$.

Examples

1. Let $K \subset \mathbb{R}^n$ be a convex body ($=$ compact convex, with non-empty interior) (symmetric means $-K = K$).
X a random vector uniformly distributed in K. Then the corresponding probability measure on $\mathbb{R}^n$

$$\mu_K(A) = \frac{|K \cap A|}{|K|}$$

is log-concave (by **Brunn-Minkowski**).
Moreover, for every convex body K there exists an affine map T such that μ_{TK} is isotropic.

2. The Gaussian vector $G = (g_1, ..., g_n)$, where g_i's have $\mathcal{N}(0, 1)$ distribution, is isotropic and log-concave.

3. Similarly the vector $X = (\xi_1, ..., \xi_n)$, where ξ_i's have exponential distribution (i.e., with density $f(t) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|t|)$, for $t \in \mathbb{R}$)
is isotropic and log-concave.

Let $n, N \geqslant 1$ be integers (a priori, no relation between them); fixed throughout.
Our interest in behaviour of invariants as functions of n, N

Random matrix: A is $n \times N$ matrix, defined either by a sequence of rows or columns, which will be independent random vectors

$$A = \begin{bmatrix} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{bmatrix}, \qquad A = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Difference with RMT where entries are independent,
limiting behaviour of invariants when the size $\to \infty$

Let $k \leqslant n$ and $m \leqslant N$ integers.

$A_{k,m} =$ maximal operator norm over submatrices of A with k rows, m columns.

“operator norm” means the norm $A : \mathbb{R}^m \to \mathbb{R}^k$ with the Euclidean norms.

Example: Let $X_1, \dots, X_n \in \mathbb{R}^N$ be independent random vectors.
Let A be $n \times N$ random matrix with rows $X_1, \dots, X_n$. This acts as an operator

$$A : \mathbb{R}^N \to \mathbb{R}^n \qquad Ax = \Big(\langle X_j, x\rangle\Big)_{j=1}^{n} \in \mathbb{R}^n, \qquad \text{for } x \in \mathbb{R}^N.$$

$$A_{k,m} = \sup_{\substack{J \subset [1,n] \\ |J| = k}} \sup_{x \in U_m} \left(\sum_{j \in J} |\langle X_j, x\rangle|^2 \right)^{1/2}.$$

$U_m = \{x \in S^{N-1} : |\operatorname{supp} x| \leqslant m\}$.

$A_{k,m}$ – examples: independent columns

I: A is $n \times N$ matrix defined by independent isotropic log-concave **columns**

$$A = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Main application to approximation of a covariance matrix by empirical covariance matrices
$A_{n,m}$ (i.e., $k = n$) was sufficient. It corresponds to submatrices of full columns, thus preserving the structure of the matrix.

Approximation of a covariance matrix

Let $X \in \mathbb{R}^n$ isotropic and log-concave,
$(X_i)_{i \leqslant N}$ independent copies of X.
By isotropicity, $\mathbb{E} X \otimes X = \mathrm{Id}$.

By the law of large numbers, the empirical covariance matrix converges to Id.

$$\frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i \longrightarrow \mathrm{Id} \qquad \text{as } N \to \infty..$$

Kannan-Lovász-Simonovits asked (around 1995), motivated by a problem of complexity in computing volume in high dimension:
Under the above assumptions, estimate the size N *for which, given* $\varepsilon \in (0,1)$*,*

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| \leqslant \varepsilon$$

holds with high probability.

Typical "translation" of a limit law into a quantitative statement in the non-limit theory.

KLS question

KLS showed that for any $\varepsilon, \delta \in (0,1)$ (under a finite third moment assumption), $N \geqslant (C/\varepsilon\delta) n^2$ gives the required approximation, with probability $1-\delta$.

Bourgain (1996): for any $\varepsilon, \delta \in (0,1)$, there exists $C(\varepsilon, \delta) > 0$ such that $N = C(\varepsilon, \delta) n \log^3 n$ gives the approximation with probability $1-\delta$.

Rudelson:

- using non-commutative Khinchine inequalities of Pisier and Lust-Piquard/Pisier;
- by majorazing measure approach of Talagrand.

Several other authors improved powers of logarithm, from late 1990's to 2010

ALPT: N proportional to n is sufficient (JAMS 2010), improved in CRAS 2011. Let $X \in \mathbb{R}^n$ be isotropic log-concave, $X_1, \ldots, X_N$ be independent copies of X.

$$\mathbb{P}\left(\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| \leqslant C\sqrt{n/N} \right) \geqslant 1 - e^{-c\sqrt{n}}.$$

So letting $\varepsilon = C\sqrt{n/N}$ we get $N = Cn/\varepsilon^2$.

As the corollary of **ALPT** we get
a quantitative version of Bai-Yin theorem for matrices of a fixed size:

$A_{k,m}$ – examples: independent rows

II: A is $n \times N$ matrix, defined by independent (isotropic log-concave) **rows**

$$A = \begin{bmatrix} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{bmatrix}$$

$A_{k,m}$ is harder to tackle if $m < N$; because the row structure of A is destroyed.

$$A_{k,m} = \sup_{\substack{J \subset [1,n] \\ |J| = k}} \sup_{x \in U_m} \left(\sum_{j \in J} |\langle X_j, x \rangle|^2 \right)^{1/2}.$$

Applicable for studies of reconstruction problems and in particular RIP; for uniform versions of some geometric questions on large deviation estimates....

Large Deviation for $A_{k,m}$

Intuition: A – a random matrix with independent isotropic log-concave rows. Then for a submatrix $A_{J,I}$ with k rows and m columns,

$$\left(\mathbb{E}\|A_{I,J}\|^2\right)^{1/2} \geqslant \sqrt{\max\{k,m\}}.$$

One of main results by ALLPT is a large deviation theorem for $A_{k.m}$.

Let n, N, $k \leqslant n$, *and* $m \leqslant N$, *let* A *be a* $n \times N$ *matrix with independent isotropic log-concave rows. For* $t \geqslant 1$ *we have*

$$\mathbb{P}\left(A_{k,m} \geqslant C\, t\, \lambda\right) \leqslant \exp(-t\lambda/\sqrt{\log(3m)}),$$

where

$$\lambda = \sqrt{\log\log(3m)}\sqrt{m}\log\left(\frac{e\max\{N,n\}}{m}\right) + \sqrt{k}\log\left(\frac{en}{k}\right)$$

and C *is a universal constant.*

The bound is essentially optimal up to $\sqrt{\log\log m}$ factor.

Paouris' large deviation theorem

Paouris' large deviation (2005):
There exists $c > 0$ *such that if* X *is an isotropic log-concave random vector in* $\mathbb{R}^N$*, then for all* $t \geqslant 1$*,*

$$\mathbb{P}\left\{|X| \geqslant c\, t\, \sqrt{N}\right\} \leqslant \exp(-t\sqrt{N}).$$

Equivalent formulations, for pth moments, etc....

Weak parameter
For a vector X in $\mathbb{R}^N$ we define

$$\sigma_X(p) := \sup_{t \in S^{N-1}} \left(\mathbb{E}|\langle t, X\rangle|^p\right)^{1/p} \quad p \geqslant 1.$$

Examples

- For isotropic log-concave vectors X, $\sigma_X(p) \leqslant p/\sqrt{2}$.
- For subgaussian vectors X, $\sigma_X(p) \leqslant C\sqrt{p}$.

Paouris' theorem with weak parameter

For any log-concave random vector X,

$$(\mathbb{E}|X|^p)^{1/p} \leqslant C\Big(\mathbb{E}|X| + \sigma_X(p)\Big) \quad \text{for } p \geqslant 2,$$

and if X is isotropic,

$$\mathbb{P}(|X| \geqslant t) \leqslant \exp\left(-\sigma_X^{-1}\left(\frac{t}{C}\right)\right) \quad \text{for } t \geqslant C(\mathbb{E}|X|^2)^{1/2}.$$

Uniform large deviation theorem

Uniform Paouris-type theorem [ALLPT]:
For $1 \leqslant m \leqslant N$ *and an isotropic log-concave vector* X *in* $\mathbb{R}^N$ *we have, for* $t \geqslant 1$,

$$\mathbb{P}\Big(\sup_{\substack{I\subset[1,N]\\|I|=m}} |P_I X| \geqslant ct\sqrt{m}\log\Big(\frac{eN}{m}\Big)\Big) \leqslant \exp\Big(-\sigma_X^{-1}\Big(\frac{t\sqrt{m}}{\sqrt{\log(em)}}\log\Big(\frac{eN}{m}\Big)\Big)\Big).$$

If X is isotropic log-concave in $\mathbb{R}^N$, then so is $P_I X$, for every $I \subset [1, N]$. However the probability is too high to beat the complexity of the family of subsets (which is $\binom{N}{m}$). So a direct union bound argument cannot be used.

Trade off of an extra logarithm in the threshold; is based on new non-trivial estimates for order statistics.

Order Statistics

For an N–dimensional random vector X by $X_1^* \geqslant X_2^* \geqslant \ldots \geqslant X_N^*$ we denote the nonincreasing rearrangement of $|X(1)|, \ldots, |X(N)|$.

In particular, $X_1^* = \max\{|X(1)|, \ldots, |X(N)|\}$ and $X_N^* = \min\{|X(1)|, \ldots, |X(N)|\}$. Random variables X_k^*, $1 \leqslant k \leqslant N$, are called order statistics of X.

Problem Find upper bound for $\mathbb{P}\left(X_k^* \geqslant t\right)$.

Order Statistics for isotropic log-concave vectors

Let X be N-dimensional log-concave isotropic vector. Then

$$\mathbb{P}\,(X_k^* \geqslant t) \leqslant \exp\Big(-\sigma_X^{-1}\Big(\frac{1}{C}t\sqrt{k}\Big)\Big) \quad \text{for } t \geqslant C\log\Big(\frac{eN}{k}\Big).$$

The weak parameter is needed for a better control of a probability for random vectors which are sums of independent random vectors, in terms of sequences of coefficients in these sums.
Latala (2010) proved a version without the weak parameter.
ALLPT (2012) the present version

The approach is based on the suitable estimate of moments of the process $N_X(t)$

$$N_X(t) := \sum_{i=1}^{n} \mathbf{1}_{\{X(i)\geqslant t\}}, \quad t \geqslant 0.$$

That is, $N_X(t)$ is equal to the number of coordinates of X larger than or equal to t.

Estimate for N_X

For any isotropic log-concave vector X *and* $p \geqslant 1$ *we have*

$$\mathbb{E}(t^2 N_X(t))^p \leqslant (C\sigma_X(p))^{2p} \quad \text{for } t \geqslant C\log\left(\frac{Nt^2}{\sigma_X^2(p)}\right).$$

To get estimate for order statistics we observe that $X_k^* \geqslant t$ implies that $N_X(t) \geqslant k/2$ or $N_{-X}(t) \geqslant k/2$ and vector $-X$ is also isotropic and log-concave. Estimates for N_X and Chebyshev's inequality give

$$\mathbb{P}\left(X_k^* \geqslant t\right) \leqslant \left(\frac{2}{k}\right)^p \left(\mathbb{E}N_X(t)^p + \mathbb{E}N_{-X}(t)^p\right) \leqslant 2\left(\frac{Cp}{t\sqrt{k}}\right)^{2p}$$

provided that $t \geqslant C\log(Nt^2/p^2)$. We take $p = \frac{1}{eC}t\sqrt{k}$ and notice that the restriction on t follows by the assumption that $t \geqslant C\log(eN/k)$.

Estimate for $N_X(t)$

Proof of estimate for $N_X(t)$ is based on two ideas.

- the restriction of a log-concave vector X to a convex set is log-concave;
- Paouris' large deviation theorem.

Uniform Paouris-type estimate

For any $m \leqslant N$ and any isotropic log-concave vector X in $\mathbb{R}^N$ we have for $t \geqslant 1$,

$$\mathbb{P}\Big(\sup_{\substack{I\subset[1,N]\\ |I|=m}} |P_I X| \geqslant ct\sqrt{m}\log\Big(\frac{eN}{m}\Big)\Big) \leqslant \exp\Big(-\sigma_X^{-1}\Big(\frac{t\sqrt{m}}{\sqrt{\log(em)}}\log\Big(\frac{eN}{m}\Big)\Big)\Big).$$

Idea of the proof. It is easy.

$$\sup_{\substack{I\subset[1,N]\\ |I|=m}} |P_I X| = \Big(\sum_{k=1}^{m} |X_k^*|^2\Big)^{1/2} \leqslant 2\Big(\sum_{i=0}^{s-1} 2^i |X_{2^i}^*|^2\Big)^{1/2},$$

where $s = \lceil \log_2 m \rceil$.

Applications – reconstruction, compressed sensing

Let $n, N \geqslant 1$. Let $T \subset \mathbb{R}^N$ and Γ be an $n \times N$ matrix.

Consider any vector $x \in T$. Assuming that Γx is known, the problem is to reconstruct x with a fast algorithm.

Hypothesis on T and on Γ. The common hypothesis is that $T = U_m$.

the Restricted Isometry Property (RIP) of order m: for all m-sparse vectors x,

$$(1 - \delta)|x| \leqslant |\Gamma x| \leqslant (1 + \delta)|x|.$$

The RIP parameter:

$$\delta_m = \delta_m(\Gamma) = \sup_{x \in U_m} \left| |\Gamma x|^2 - \mathbb{E}|\Gamma x|^2 \right|$$

Introduced by E. Candes, J. Romberg and T. Tao around 2006.
If δ_{2m} is appropriately small then every m-sparse vector x can be reconstructed from Γx by the ℓ_1-minimization method.

More notation

Upper estimates for

$$\delta_m = \delta_m(\Gamma) = \sup_{x \in U_m} \left| |\Gamma x|^2 - \mathbb{E}|\Gamma x|^2 \right|$$

More generally, for any $T \subset S^{N-1}$,

$$\delta_T(\Gamma) = \sup_{x \in T} \left| |\Gamma x|^2 - \mathbb{E}|\Gamma x|^2 \right|.$$

Let $X_1, \ldots, X_n \in \mathbb{R}^N$ independent; Γ the $n \times N$ matrix with rows X_i. (In reconstruction problems – we look for vectors given by their *measurements*)

Let $1 \leqslant k \leqslant n$ and define the parameter $\Gamma_k(T)$ by

$$\Gamma_k(T)^2 = \sup_{y \in T} \sup_{\substack{I \subset \{1,\ldots,n\} \\ |I| = k}} \sum_{i \in I} | \langle X_i, y \rangle |^2.$$

We write $\Gamma_{k,m} = \Gamma_k(U_m)$.
It agrees with the definition introduced earlier – of $A_{k,m}$

Fundamental Lemma:

[ALPT, CRAS], [ALLPT]:
Let $X_1, \dots, X_n \in \mathbb{R}^N$ *be independent isotropic,* $T \subset S^{N-1}$ *finite. Let* $0 < \theta < 1$ *and* $B \geqslant 1$*. Then with probability at least* $1 - |T| \exp\left(-3\theta^2 n/8B^2\right)$,

$$\begin{aligned}
\delta_T\left(\frac{\Gamma}{\sqrt{n}}\right) &= \sup_{y\in T}\left|\frac{1}{n}\sum_{i=1}^{n}\left(|\langle X_i, y\rangle|^2 - \mathbb{E}|\langle X_i, y\rangle|^2\right)\right| \\
&\leqslant \theta + \frac{1}{n}\left(\sup_{y\in T}\sum_{i=1}^{n}|\langle X_i, y\rangle|^2 \mathbf{1}_{\{|\langle X_i, y\rangle| \geqslant B\}}\right. \\
&\qquad \left. + \sup_{y\in T}\mathbb{E}\sum_{i=1}^{n}|\langle X_i, y\rangle|^2 \mathbf{1}_{\{|\langle X_i, y\rangle| \geqslant B\}}\right) \\
&\leqslant \theta + \frac{1}{n}\left(\Gamma_k(T)^2 + \mathbb{E}\Gamma_k(T)^2\right).
\end{aligned}$$

where $k \leqslant n$ *is the largest integer satisfying* $k \leqslant (\Gamma_k(T)/B)^2$*.*

Corollary for RIP:

Let X_i, Γ, $0 < \theta < 1$ *and* $B \geqslant 1$, *as before. Assume that* $m \leqslant N$ *satisfies*

$$m \log \frac{11eN}{m} \leqslant \frac{3\theta^2 n}{16B^2}.$$

Then with probability at least $1 - \exp\left(-\frac{3\theta^2 n}{16B^2}\right)$ *one has*

$$\delta_m\left(\frac{\Gamma}{\sqrt{n}}\right) = \sup_{y \in U_m} \left| \frac{1}{n} \sum_{i=1}^{n} \left(|\langle X_i, y\rangle|^2 - \mathbb{E}|\langle X_i, y\rangle|^2\right) \right|$$
$$\leqslant 2\,\theta + \frac{2}{n}\left(\Gamma_{k,m}^2 + \mathbb{E}\Gamma_{k,m}^2\right),$$

where $k \leqslant n$ *is the largest integer satisfying* $k \leqslant (\Gamma_{km}/B)^2$.

RIP Theorem for matrices with independent rows:

Let $n, N \geqslant 1$ *and* $0 < \theta < 1$. *Let* Γ *be an* $n \times N$ *matrix, whose rows are independent isotropic log-concave random vectors* X_i, $i \leqslant n$.
There exists an absolute constant $c > 0$, *such that if* $m \leqslant N$ *satisfies*

$$m \log\log 3m \left(\log \frac{3\max\{N, n\}}{m}\right)^2 \leqslant c \left(\frac{\theta}{\log(3/\theta)}\right)^2 n$$

then

$$\delta_m(\Gamma/\sqrt{n}) \leqslant \theta$$

with high probability.

Optimal up to a log log factor.
For unconditional distributions we know that this factor can be removed; we conjecture that in general can be removed as well.

Recall the result for $A_{k,m}$.
For $n \leqslant N$, $k \leqslant n$, $m \leqslant N$,

$$A_{k,m} = \sup_{\substack{J \subset [1,n] \\ |J| = k}} \sup_{x \in U_m} \left(\sum_{j \in J} |\langle X_j, x \rangle|^2 \right)^{1/2}.$$

Then for $t \geqslant 1$ we have

$$\mathbb{P}\left(A_{k,m} \geqslant C\, t\, \lambda\right) \leqslant \exp(-t\lambda / \sqrt{\log(3m)}),$$

where

$$\lambda = \sqrt{\log\log(3m)}\sqrt{m}\log(eN/m) + \sqrt{k}\log(en/k).$$

$A_{k,m}$, idea of proof

To bound $A_{k,m}$ one has then to prove uniformity with respect to two families of different character:
one being $\{I \subset [1, N] : \ |I| = k\}$; and the other equal to $U_m(\mathbb{R}^N)$.

$X_1, \ldots, X_n$ independent isotropic N-dimensional log-concave vectors.
$x = (x_i) \in \mathbb{R}^n$ with some structural assumptions, like sparsity.... we consider $Y = \sum_{i=1}^n x_i X_i$.
By duality we need to estimate probability that

$$\left\{ \sup_{\substack{J \subset [1,N] \\ |J| = m}} \left| P_J \left(\sum_{i=1}^n x_i X_i \right) \right| \geqslant t \right\} = \left\{ \sup_{\substack{J \subset [1,N] \\ |J| = m}} |P_J Y| \geqslant t \right\}$$

for every $t \geqslant 0$, depending on the norms $|x|$ and $\|x\|_\infty$.
Complexity of these families are too high for using a union bound argument, and so we need to come up with some chaining.

$A_{k,m}$, continuation

This leads us to distinguishing two cases, depending on the relation between k and k':

$$k' = \inf\{\ell \geqslant 1 \colon m \log(eN/m) \leqslant \ell \log(en/\ell)\}.$$

Step 1. when $k \geqslant k'$. We reduce to the case $k \leqslant k'$.
Step 2. Case $k \leqslant k'$.

To build intuition we may take $k' \sim k$.

Step 1. We take only the family of k-sparse vectors, but do not need projections.

$$\left\{ \sup_{x \in U_k} \left| \sum_{i=1}^{n} x_i X_i \right| \geqslant t \right\}.$$

Assume first that x is a "flat" vector: $x_i = \pm a$ or 0 and $a = \tilde{k}^{-1/2}$, where $\tilde{k} = |\,\mathrm{supp}(x)|$. That is, $|x| = 1$ and $\|x\|_\infty = \tilde{k}^{-1/2}$.
Direct argument shows that the estimate is right for such vectors..

We may have $0 < |x_1| \leqslant |x_2| \leqslant \ldots |x_{\tilde{k}}|$ and $x_j = 0$ for $j > \tilde{k}$, which may consist of some number of "flat" vectors.
The first natural try is to consider separately each flat vector and then add the results together. This works but may produce an extra logarithmic factor.

Chaining:
let $k_1 \sim k/2$, $k_2 \sim k/4$, ..., $k_s \sim k/2^s \sim k'$. So $\sum_{j=1}^{s} k_j = k'$.
Given $x \in U_k$, let x^1 be the restriction of x to the k_1 smallest coordinates; x^2 be the restriction of x to the next k_2 smallest coordinates, etc.

This way,

$$x = \sum_{i=1}^{s} x^i$$

where x^i's have mutually disjoint supports, each of cardinality $\leqslant k_i$, and coordinates of x^i are larger than coordinates of x^j if $i < j$.
We use Paouris-type estimates for each x^i...
This is similar to ALPT (JAMS).

$A_{k,m}$, Step 2

Step 2. Another chaining argument, more delicate in definitions of ε-nets. We use the uniform estimate for projections of sums, which in general is weaker than in Case 1.
At this step we lose $\log\log m$.

Congratulations Alain!