Interactions between high dimensional geometry and random matrices

Alain Pajor

University of Paris Est Marne-la-Vallée

Hong Kong January 2015

Model

We consider $n \times N$ random matrices of the form

$$A = A(Y_1, \ldots, Y_N) = \Big[Y_1 \cdots\cdots\cdots Y_N \Big] = \Big[\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \Big]$$

with columns $Y_1, \ldots, Y_N$ where $Y, Y_1, \ldots, Y_N$ are i.i.d. random vectors in $\mathbb{R}^n$.
We assume that Y is **isotropic** with the normalization of R.M.T.

$$\mathbb{E}[Y] = 0 \quad \text{and} \quad \mathbb{E}[YY^\top] = \frac{I_n}{n}.$$

We define

$$M = \frac{n}{N} \sum_{i=1}^{N} Y_i Y_i^\top = \frac{n}{N} A A^\top$$

$$\text{and denote} \quad 0 \leq \cdots \leq \lambda_i(M) \leq \cdots \leq \lambda_1(M)$$

the eigenvalues of M which are the square of the singular values $\Big(s_i\big(\sqrt{\frac{N}{n}}\, A\big)\Big)_{i \geq 1}$ where the operator A is considered from the Euclidean spaces $\mathbb{R}^N$ to $\mathbb{R}^n$,

$$A : \ell_2^N \longrightarrow \ell_2^n.$$

Model

$$\mathbb{E}[Y] = 0 \quad \text{and} \quad \mathbb{E}[YY^\top] = \frac{I_n}{n}$$

$$M = \frac{n}{N} \sum_{i=1}^{N} Y_i Y_i^\top = \frac{n}{N} AA^\top$$

with eigenvalues $\quad 0 \leq \cdots \leq \lambda_i(M) \leq \cdots \leq \lambda_1(M)$.

In Random Matrix Theory one is mostly interested in asymptotic properties (local or global) of the spectrum as $n \to \infty$ and $n \sim dN$.

In high dimensional geometry and in asymptotic geometric analysis (contrary to what is written) one is interested in **quantitative** estimates, in terms of n and N and properties of Y and of the ambient space, of quantities $f(A)$ that may depend (or not) on the spectrum of M. "Quantitative" means that we ask for estimates of

$$\mathbb{P}(f(A) \in I)$$

<u>for fixed large n and N</u> and some interval I.

Quantitative (non-asymptotic) problems

Problems in high dimensional geometry

- Geometry of polytopes
- Neighborliness and R.I.P.

Problems in asymptotic geometric analysis

Let E be an N-dimensional normed space and B_E be its unit ball.

- Estimates of geometric parameters of the unit ball $\ker A \cap B_E$ ($N \geq n$) of the random subspace defined by $\ker A$: Euclidean radius and inradius, mean width...
- Approximation theory (Kolmogorov diameter)

Interactions between compressed sensing, random matrices and high dimensional geometry

by

Djalil Chafaï, Olivier Guédon, Guillaume Lecué, Alain Pajor

Panorama et synthèse (2013) SMF

Problems in high dimensional geometry

- Random polytopes

Let

$$K(A) = K(Y_1, \ldots, Y_N)$$

be the convex hull (or symmetric convex hull) of the columns Y_i, $1 \leq i \leq N$.

Thus $K(A)$ is a random polytope.

- $(0,1)$-random polypotes : Y is a random vertex of the hypercube.
- Y is a random point uniformly distributed in a convex body K.
  - How the geometry of the convex body is reflected in $K(Y_1, \ldots, Y_N)$ (vice versa), concerning the structure of faces, geometric parameters, volume...
  - How many points N, picked at random do you need to approximate geometric parameters of the body, such as center of mass, mean width $\ldots$ Problems of algorithmic complexity in geometry.
  - Approximation of the inertia matrix (covariance matrix of Y). Measure of the degree of accuracy in terms of n and N (developed below).

Approximating the inertia matrix or the covariance matrix

In a paper on the algorithmic complexity of computing volume in high dimensions, Kannan, Lovász and Simonovits asked for the following question (1996):

Question (KLS)

Let K be a convex body in $\mathbb{R}^n$. Given $\varepsilon > 0$, how many independent points Y_i uniformly distributed on K are needed for the empirical covariance matrix to approximate the covariance matrix up to ε with overwhelming probability?

Kannan, Lovász, Simonovits question reformulated

Question

Let $Y \in \mathbb{R}^n$ be a random vector. Given ε, what should be the size N of a sample $(Y_i)_{i \le N}$ in order that the <u>*empirical covariance matrix*</u> *Σ_N approximates the* <u>*covariance matrix*</u> *Σ of Y up to ε with overwhelming probability? That is*

$$\|\Sigma_N - \Sigma\| = \left\| \frac{1}{N} \sum_1^N Y_i Y_i^\top - \Sigma \right\| \le \varepsilon \|\Sigma\| \quad \textit{with high probability.}$$

We may assume that Y is isotropic and the question becomes

Question

Assume that $Y \in \mathbb{R}^n$ is isotropic, then we would like that with overwhelming probability

$$\|M - I\| = \left\| \frac{n}{N} \sum_1^N Y_i Y_i^\top - I \right\| = \max(|1 - \lambda_n|, |\lambda_1 - 1|) \le \varepsilon.$$

Kannan, Lovász, Simonovits question reformulated

Equivalently, the question is about a **quantitative estimate** on the edges of the support of the spectrum:

$$(1-\varepsilon) \leq \lambda_n \leq \lambda_1 \leq (1+\varepsilon).$$

In the regime $n \sim dN$, $d \in (0,1)$, the asymptotics are known when **the entries are i.i.d.** (Bai-Yin, under fourth moment condition)

$$\lim \lambda_n(M) = (1-\sqrt{d})^2 = \lambda_- \quad a.s. \quad \text{and} \quad \lim \lambda_1(M) = (1+\sqrt{d})^2 = \lambda_+ \quad a.s.$$

Conjecture:

Let Y be an isotropic random vector uniformly distributed in a convex body of dimension n. Let $\varepsilon > 0$. Then a sample of size

$$N = C(\varepsilon) n$$

is sufficient to approximate M up to ε (w.h.p.), that is $\|M - I\| \leq \varepsilon$ w.h.p.

Result in the convex or log-concave setting

- KLS (96): $N \sim c(\varepsilon) n^2$
- Breakthrough, Bourgain (96) : $N \sim c(\varepsilon) n \log^3 n$
- Rudelson (99): $N \sim c(\varepsilon) n \log^2 n$
- ...

Theorem [ALPT (2009-2010)] :

Let $N \geq n \geq 1$. Let $(Y_i)_{i \leq N}$ be independent isotropic log-concave (*) random vectors in $\mathbb{R}^n$, then with high probability

$$\|M - I\| = \left\| \frac{n}{N} \sum_{1}^{N} Y_i Y_i^\top - I \right\| \leq c \sqrt{\frac{n}{N}}.$$

Thus, let $\varepsilon \in (0,1)$. A log-concave isotropic random vector satisfies the conjecture with $C(\varepsilon) = O\left(\frac{1}{\varepsilon^2}\right)$ (which is optimal):

a sample of size $N = Cn/\varepsilon^2$ is sufficient
to approximate the covariance matrix up to ε w.h.p.

ALPT=R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann

(*) **log-concave random vector**= on its support, its distribution has a density whose log is a concave

Question

Which type of vector Y satisfies the conjecture ?

That is, which random vector $Y \in \mathbb{R}^n$ has the property that for N proportional to n, the empirical covariance matrix of a sample of size N approximates the covariance matrix of Y up to ε with overwhelming probability?

For instance Y log-concave, Y Bernoulli, subgaussian...
More generally, it was proved in ALPT that if for some $C, c > 0$,

- Y is isotropic
- $|Y| \leq c$ with high probability
- $\forall t > 1 \ \forall \theta \in S^{n-1} \quad \mathbb{P}(|\langle Y, \theta\rangle| > t(\mathbb{E}|\langle Y, \theta\rangle|^2)^{1/2}) \leq C \exp(-ct)$

Then Y satisfies the conjecture.

$|\,.\,|$ denotes the Euclidean norm

Let $p \geq 1$, define the weak p-th moment as

$$\sigma_p(Y) = \sup_{|\theta| \leq 1} \left(\mathbb{E}|\langle Y, \theta \rangle|^p\right)^{1/p}.$$

Conjecture (Srivastava and Vershynin, 2011)

Assuming

Y isotropic, $|Y| \leq c\,(\mathbb{E}|Y|^2)^{1/2}$ and $\sigma_p(Y) \leq \sigma.\sigma_2(Y)$ with $p > 2$

then *a sample of size $N = Cn$ is sufficient to approximate the covariance matrix up to ε w.h.p., where $C = C(\varepsilon, p, c, \sigma)$.*

In the case $p = 2$, $N \sim n \log n$ is needed when Y is uniformly distributed on the set of n vectors of Euclidean norm 1. Indeed, the size $N \sim n \log n$ is needed for the sample $\{Y_1, ..., Y_N\}$ to contain all these vectors, which is required for a nontrivial covariance approximation.

N. Srivastava and R. Vershynin (2011)

Assuming

- Y is isotropic
- $|Y| \leq c\left(\mathbb{E}|Y|^2\right)^{1/2}$
- for some $p > 2$, $\left(\mathbb{E}|QY|^p\right)^{1/p} \leq C\left(\mathbb{E}|QY|^2\right)^{1/2}$, for all orthogonal projection Q,

then a sample of size $N \approx n/\varepsilon^\alpha$ is sufficient to approximate Σ, where $\alpha > 2$ depends p.

S. Mendelson - G. Paouris (2012)

Let $p > 8$. Assuming

$$Y \text{ isotropic, } |Y| \leq c\left(\mathbb{E}|Y|^2\right)^{1/2} \text{ and } \sigma_p(Y) \leq \sigma.\sigma_2(Y)$$

then *a sample of size $N = Cn$ is sufficient to approximate the covariance matrix up to ε w.h.p., where $C = C(\varepsilon, p, c, \sigma) = C(p, c, \sigma)/\varepsilon^2$.*

The accuracy (dependence in ε) is optimal.

S. Mendelson - G. Paouris (2012)

Let $p > 8$. Assuming

$$Y \text{ isotropic}, \ |Y| \leq c\left(\mathbb{E}|Y|^2\right)^{1/2} \ \text{ and } \sigma_p(Y) \leq \sigma.\sigma_2(Y)$$

then *a sample of size $N = Cn$ is sufficient to approximate the covariance matrix up to ε w.h.p., where $C = C(\varepsilon, p, c, \sigma) = C(p, c, \sigma)/\varepsilon^2$.*

Using the same method Guédon-Litvak-Pajor-Tomczak-Jaegermann (2013) completed the result for $8 \geq p > 4$, with $C = C(\varepsilon, p, c, \sigma) = C(p, c, \sigma)/\varepsilon^\alpha$ and $\alpha > 2$ (probably not optimal?).

Remains the case $4 \geq p > 2$ and optimality (dependence in ε).

Important remark: In all these results the hypothesis are satisfied when the entries are independent with bounded p-moment ($p > 4$), but the dependence in ε is probably not optimal.

Quantitative Tracy-Widom

The methods of the above results show that for some $c > 1$

$$\|M\| \leq 1 + c\sqrt{\frac{n}{N}}$$

but do not give what is expected for the asymptotic ($c \sim 1$?). By an other approach Pillai and Yin got a "quantitative" Tracy-Widom estimate:

N. S. Pillai, J. Yin (2012)

Assuming that

- one-dimensional marginals have sub-exponential tail behavior
- concentration of quadratic forms $\langle Y, BY \rangle$, for every $n \times n$ matrix B
  - off-diagonal part of quadratic forms satisfies a sub-exponential tail behavior
  - $\langle Y, BY \rangle$ highly concentrates around their expectation

then w.h.p.

$$\lambda_- - N^{-2/3}\varphi \leq \lambda_n \leq \lambda_1 \leq \lambda_+ + N^{-2/3}\varphi$$

where $\varphi = (\log N)^{c \log \log N}$

Remark: The hypothesis are satisfied when the entries are i.i.d., centered with variance 1 and have sub-exponential tail behavior. What else? Log-concave?

The lower edge

The difference of moment hypothesis between the lower and upper edges to bound the eigenvalues appeared first in a paper of Srivastava and Vershynin.

N. Srivastava and R. Vershynin (2011)

Assuming that

$$Y \text{ isotropic, } |Y| \leq c\left(\mathbb{E}|Y|^2\right)^{1/2} \text{ and } \sigma_p(Y) \leq \sigma.\sigma_2(Y) \text{ for some } p > 2$$

Then, for $Nd \geq n$, we have

$$\mathbb{E}\lambda_n^2 \geq 1 - Cd^{1/\alpha}$$

where $\alpha > 2$ depends on p and C depends on c and σ.

This was then improved by V. Koltchinskii and S. Mendelson (2013) where the optimal behavior of λ_n (up to a multiplicative constant) was obtained for $p > 4$.
Assuming that

$$Y \text{ isotropic, } |Y| \leq c\left(\mathbb{E}|Y|^2\right)^{1/2} \text{ and } \sigma_p(Y) \leq \sigma.\sigma_2(Y) \text{ for some } p > 4$$

then for $Nd \geq n$, w.h.p. one has

$$\lambda_n \geq 1 - C\sqrt{d}.$$

A lower bound is also obtained when $2 < p \leq 4$.

The lower edge :Tikhomirov

Finally very recently K. Tikhomirov proved that no moment assumption is needed.

K. Tikhomirov (2014)

Let $\beta, d \in (0,1)$. Assume that $A = (a_{ij})$ has i.i.d. entries satisfying

$$\sup_{\lambda} \mathbb{P}\{|a_{11} - \lambda| \leq 1\} \leq 1 - \beta$$

then for $Nd \geq max(N_0, n)$,

$$s_n(A) > c\sqrt{N}$$

with probability larger than $1 - \exp(-C\sqrt{N})$, where c, C, N_0 depend on β and d.

Remark: The hypothesis is satisfies for a zero median random variable satisfying $\mathbb{E}|a_{11}|^p \geq m$, $\mathbb{E}|a_{11}|^q \leq M$ for some $0 < p < q$ and $m, N > 0$.

The lower edge: asymptotics

Very recently K. Tikhomirov proved that moment 2 is sufficient to get a Bai-Yin lower edge asymptotic.

K. Tikhomirov (2014)

Assume that $A = (a_{ij})$ has i.i.d. entries with zero mean and variance $1/n$. Then with probability 1,

$$\lambda_n \longrightarrow (1 - \sqrt{d})^2 = \lambda_-.$$

where $d = n/N \in (0, 1)$.

Problems of high dimensional geometry

- Neighborliness and R.I.P.

Let $1 \leq s \leq n \leq N$. For every $J \subset \{1, \ldots, N\}$ with cardinality s, denote by

$$A(Y_j; j \in J) = A^J = (Y_j)_{j \in J}$$

the $n \times s$ matrix with columns $(Y_j)_{j \in J}$.

In **compressed sensing**, one is interested in the so-called **Restricted Isometry Property (R.I.P.)**. We ask that for every $J \subset \{1, \ldots, N\}$ with cardinality s, A^J acts almost isometrically, that is, A acts almost like an isometry on **s-sparse** vectors. This is quantified as follows.

Let $\theta \in (0, 1)$. The matrix satisfies $RIP_s(\theta)$ if for ever $J \subset \{1, \ldots, N\}$ with cardinality s, the spectrum of $M^J = (A^J)^\top A^J$ lies in the interval $[1 - \theta, 1 + \theta]$. (Candés-Tao). Equivalently, w.h.p.

$$\forall J \subset \{1, \ldots, N\}\ |J| = s \qquad 1 - \theta \leq \lambda_s(M^J) \leq \lambda_1(M^J) \leq 1 + \theta$$

If this is true for a small θ then every s-sparse vector of $\mathbb{R}^N$ can be reconstructed exactly from the data Ax by a fast algorithm (Candés-Romberg-Tao, Donoho).

Problems of high dimensional geometry

- Neighborliness and R.I.P.

$\mathrm{RIP}_s(\theta)$, $1 \le s \le n \le N$:

$$\forall J \subset \{1,\dots,N\}\ |J| = s \qquad 1-\theta \le \lambda_s(M^J) \le \lambda_1(M^J) \le 1+\theta$$

Of course we want the a priori hypothesis of sparsity to be the less restrictive possible, that is, s the largest possible, say almost proportional to n (w.h.p.). Thus the problem becomes to find s the largest possible s.t. w.h.p. we have a good <u>uniform</u> control of the edges of the spectrum of the restricted matrices M^J with $|J| = s$.

Indeed, s is proportional to n up to $\log^{1/2}(N/n)$ for Gaussian random matrices (CRT,Do) or for Bernoulli matrices (MPT, BDDeW). The case of "structured matrices" when the entries of the columns are not independent is an other problem (see the books FR).

An other approach developed by D. Donoho and his collaborators is the study of the structure of faces of the random polytope $K(A)$.

CRT(2006)= Candés-Romberg-Tao, Do(2006)= Donoho,

MPT(2007)=Mendelson-P.-Tomczak-Jaegermann, BDDeW(2008)=Baraniuk-Davenport-DeVore-Wakin,

FR(2014)= Foucart-Rauhut

THANK YOU