PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 135, Number 5, May 2007, Pages 1293–1303
S 0002-9939(06)08615-1
Article electronically published on November 14, 2006

SAMPLING CONVEX BODIES: A RANDOM MATRIX APPROACH

GUILLAUME AUBRUN

(Communicated by N. Tomczak-Jaegermann)

ABSTRACT. We prove the following result: for any $\varepsilon > 0$, only $C(\varepsilon)n$ sample points are enough to obtain $(1+\varepsilon)$-approximation of the inertia ellipsoid of an unconditional convex body in $\mathbf{R}^n$. Moreover, for any $\rho > 1$, already ρn sample points give isomorphic approximation of the inertia ellipsoid. The proofs rely on an adaptation of the moments method from Random Matrix Theory.

1. INTRODUCTION AND THE MAIN RESULTS

Notation kept throughout the paper. The letters $C, c, C', \ldots$ denote absolute positive constants, notably independent of the dimension. The value of such constants may change from line to line. Similarly, $C(\varepsilon)$ denotes a constant depending only on the parameter ε. The canonical basis of $\mathbf{R}^n$ is $(e_1, \ldots, e_n)$, and the Euclidean norm and scalar product are denoted by $|\cdot|$ and $\langle \cdot, \cdot \rangle$. The operator norm of a matrix is denoted by $\|\cdot\|$. For a real symmetric matrix A, we write $\lambda_{\max}(A)$ (respectively $\lambda_{\min}(A)$) for the largest (respectively smallest) eigenvalue of A. Also, $\operatorname{diag}(A)$ denotes the diagonal matrix obtained from A by assigning the value 0 to all non-diagonal coefficients. A *convex body* is a convex compact subset of $\mathbf{R}^n$ with non-empty interior. A convex body K is said to be *unconditional* if it is invariant under sign flips of the coordinates: for any $\eta = (\eta_1, \ldots, \eta_n) \in \{-1, 1\}^n$,

$$(x_1, \ldots, x_n) \in K \Longleftrightarrow (\eta_1 x_1, \ldots, \eta_n x_n) \in K.$$

We reserve the letters X, Y to denote an $\mathbf{R}^n$-valued random vector; $X_1, \ldots, X_N$ are i.i.d. copies of X. If $\mathbf{E}X = 0$, X is said to be *centered*. The random vector X is said to be *isotropic* if it is centered and for all $y \in \mathbf{R}^n$,

$$\mathbf{E}\langle X, y\rangle^2 = |y|^2.$$

This is equivalent to the *inertia matrix* $\mathbf{E}X \otimes X$ being the identity matrix. We will consider the special case when X is uniformly distributed on a convex body K. We will then say "inertia matrix of K", "K is isotropic"[1]... for "inertia matrix of X", "X is isotropic"... . The *inertia ellipsoid* of K is the unique ellipsoid with the same

Received by the editors December 1, 2005 and, in revised form, December 21, 2005.
2000 *Mathematics Subject Classification.* Primary 15A52, 52A20.
Key words and phrases. Isotropic, random matrix, convex body, moments method.
This research was supported in part by the European Network PHD, FP6 Marie Curie Actions, MCRN-511953 and was done in part while the author was visiting the University of Athens.

[1] This terminology slightly differs from [17, 9] where isotropic convex bodies are normalized to have volume 1.

inertia matrix as K. For recent results on isotropic convex bodies, a good reference is the survey [9]. Any random vector has an affine image which is isotropic, and this image is unique up to orthogonal transformation. Thus, for affinely invariant problems we can restrict ourselves to isotropic random vectors. If we do not know the law but only N samples of the random vector X, we can only consider the *empirical inertia matrix*

$$A^N(X) := \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i.$$

The matrices $A^N(X)$ tend almost surely to the identity matrix when N tends to infinity; a natural question is to quantify this convergence. This problem was considered with algorithmic motivations by Kannan, Lovász and Simonovits [12] in the case when X is uniformly distributed on a convex body. It was proved in [12] that $\|A^N(X) - \mathrm{Id}\| \leqslant \varepsilon$ with probability larger than $1-\varepsilon$ provided $N \geqslant C(\varepsilon)n^2$. This was improved by Bourgain [6] to $N \geqslant C(\varepsilon)n\log^3 n$ and later by Rudelson [21] to $N \geqslant C(\varepsilon)n\log^2 n$. Rudelson proved actually the following inequality, valid for a general random vector.

Theorem (Rudelson's inequality)**.** *For any isotropic random vector X we have*

$$\mathbf{E}\left\|A^N(X) - \mathrm{Id}\right\| \leqslant C\sqrt{\frac{\log n}{N}}(\mathbf{E}|X|^{\log N})^{1/\log N} \tag{1}$$

provided the right-hand side is smaller than 1.

If X is uniformly distributed on an isotropic convex body, then it satisfies

$$(\mathbf{E}|X|^p)^{1/p} \leqslant C\sqrt{n} \quad \text{for} \quad 2 \leqslant p \leqslant c\sqrt{n}. \tag{2}$$

This estimate was proved by Bobkov and Nazarov [4] for unconditional convex bodies and recently extended by Paouris [19] to general isotropic bodies. When plugged in Rudelson's inequality, it yields that if $N \geqslant C(\varepsilon)n\log n$, we have $\|A^N(X)-\mathrm{Id}\| \leqslant \varepsilon$ with probability larger than $1-\varepsilon$ (see [10, 19]). On the other hand, when X is isotropic we have $\mathbf{E}|X|^2 = n$ and consequently we must take N larger than $cn\log n$ to use Rudelson's inequality. Note that this value $N \sim n\log n$ is sharp for some discrete examples. The simplest is given by the isotropic random vector Y uniformly distributed on the (properly normalized) vertices of the cross-polytope

$$\{\pm\sqrt{n}e_1, \dots, \pm\sqrt{n}e_n\}.$$

The matrix $A^N(Y)$ is then diagonal and its diagonal coefficients are distributed as

$$\frac{n}{N}\left(p_1, \dots, p_n\right),$$

where p_i denote the number of balls falling in the ith urn when we put randomly, uniformly and independently N balls in n urns. This problem, known as the random allocation problem, is well-studied and it is known (see [13], chapter 2.6) that we must take $N \geqslant c(\varepsilon)n\log n$ in order to get $\max(p_i) \leqslant (1+\varepsilon)\min(p_i)$ with probability larger than $\frac{1}{2}$.

We prove that for the class of unconditional convex bodies, it is possible to go below this bound of $n\log n$ and to take N proportional to n.

Theorem 1. *There are absolute constants C, c such that the following holds. Let $0 < \varepsilon \leqslant 1$. Let X be uniformly distributed on an unconditional isotropic convex body in $\mathbf{R}^n$. Then for $N \geqslant Cn/\varepsilon^2$, we have with probability larger than $1 - C\exp(-cn^{1/5})$,*

$$\|A^N(X) - \mathrm{Id}\| \leqslant \varepsilon.$$

In other words, for every y in $\mathbf{R}^n$,

$$(1-\varepsilon)|y|^2 \leqslant \frac{1}{N}\sum_{i=1}^{N}\langle X_i, y\rangle^2 \leqslant (1+\varepsilon)|y|^2. \tag{3}$$

We can also obtain the isomorphic analogue of Theorem 1, using as few sample points as possible

Theorem 2. *Let $\rho > 1$ and $N \geqslant \rho n$. Let X be uniformly distributed on an unconditional isotropic convex body in $\mathbf{R}^n$. Then, with probability larger than $1 - 2\exp(-c(\rho)n^{1/5})$ we have for every y in $\mathbf{R}^n$,*

$$\frac{1}{C(\rho)}|y|^2 \leqslant \frac{1}{N}\sum_{i=1}^{N}\langle X_i, y\rangle^2 \leqslant C(\rho)|y|^2.$$

Question 1. Do both results extend to all isotropic convex bodies?

Remark. With slight modifications of the proofs, one can prove the same results for all isotropic random vectors with a law being log-concave and unconditional.

2. The random matrix approach: Auxiliary results and the structure of the proof

Our proof uses standard techniques from Random Matrix Theory (RMT). A part of the classical random matrix theory deals with random vectors X with i.i.d. coordinates. Here is a reformulation of a result of Bai and Yin [2].

Theorem (Bai–Yin)**.** *Let Z be a random variable with mean 0, variance 1 and finite fourth moment. Let $X^{(n)}$ be a random vector on $\mathbf{R}^n$ whose coordinates are i.i.d. copies of Z. We consider a sequence of integers (N_n) tending to infinity in such a way that the ratio n/N_n tends to a limit $\beta \in (0,1)$. Then, almost surely,*

$$\lim_{n\to\infty} \lambda_{\max}\left(A^{N_n}(X^{(n)})\right) = (1+\sqrt{\beta})^2,$$
$$\lim_{n\to\infty} \lambda_{\min}\left(A^{N_n}(X^{(n)})\right) = (1-\sqrt{\beta})^2.$$

This theorem is restricted to random vectors with independent coordinates. Moreover, as often happens in RMT, this is a limit-result. Therefore it may be hard to use to get a result in a fixed dimension. In a few cases, quantified analogues (sometimes called *localized*) of limit-theorems from RMT have been proved; see [15]. For the Bai–Yin theorem, this has been done by S. Sodin in the special case of random signs [22], but we still lack a quantified version for general entries. Bai and Yin proved their theorem as a consequence of the following result:

$$\limsup_{n\to\infty} \left\|A^{N_n}(X^{(n)}) - (1+\beta)\mathrm{Id}\right\| \leqslant 2\sqrt{\beta} \quad \text{a.s.} \tag{4}$$

We follow a similar approach, except that we would rather estimate

$$\left\|A^{N_n}(X^{(n)}) - \mathrm{Id}\right\|. \tag{5}$$

When $\beta \ll 1$, the quantities involved in (4) and (5) are comparable. This is no longer true when β becomes large (i.e., when the matrices involved become close to being square matrices). Because of $\lambda_{\max}$, it is even impossible to obtain from (5) any non-trivial information on $\lambda_{\min}$ when β is larger than $3-2\sqrt{2} \approx 0.172$.

In this framework, the expected norm of a self-adjoint random matrix A is usually bounded from above by $(\mathbf{E}\mathrm{tr}A^k)^{1/k}$, for an even integer k (usually large, but not very large). We are led to some combinatorial problems to estimate $\mathbf{E}\mathrm{tr}A^k$. This is the so-called *moments method* initiated by Wigner. The main advantage of considering (5) rather than (4) is that the combinatorics involved are simpler. We prove the following proposition.

Proposition 1. *Let X be a random vector uniformly distributed on an unconditional convex body in $\mathbf{R}^n$. We write A for $A^N(X)$, the empirical inertia matrix of X with $N \geqslant n$ sample points. For $k = n^{1/5}$, we have*

$$\left(\mathbf{E}\|A - \mathrm{diag}(A)\|^k\right)^{1/k} \leqslant C\sqrt{\frac{n}{N}}, \tag{6}$$

where C is a universal constant.

We postpone the proof of Proposition 1 to section 4. We show in section 3 how to derive Theorem 1 and the upper estimate in Theorem 2 from Proposition 1. Then, the lower estimate in Theorem 2 is proved using the Laplace transform technique from [16] (see section 5).

Remark. A natural question is whether the Bai–Yin theorem can be extended to random vectors uniformly distributed on convex bodies. This question needs sharper tools than the problem which is treated here. The author obtained an affirmative answer for the case of the unit ball of ℓ_p^n [1].

3. Proof of the main results

Let Z be a random variable. For $p \geqslant 1$ we define $\|Z\|_p = (\mathbf{E}|Z|^p)^{1/p}$ and for $\alpha > 0$, the norm Ψ_α by

$$\Psi_\alpha(Z) = \sup_{p \geqslant 1} \frac{\|Z\|_p}{p^{1/\alpha}}.$$

The Ψ_α norm is usually defined for $\alpha \geqslant 1$ via the Orlicz function $\psi_\alpha(x) = \exp(x^\alpha) - 1$. The following lemma will be used. This is an analogue of the Bernstein inequalities for Ψ_α random variables with $0 < \alpha \leqslant 1$.

Lemma 1. *For any $0 < \alpha \leqslant 1$, there are constants $C(\alpha), c(\alpha)$ such that the following holds. Let $(Z_j)_{1 \leqslant j \leqslant m}$ be i.i.d. random variables such that $\mathbf{E}Z_1 = 0$ and $\Psi_\alpha(Z_1) \leqslant K$. For any $t \geqslant C(\alpha)K/\sqrt{m}$,*

$$\mathbf{P}\left(\left|\frac{1}{m}\sum_{j=1}^m Z_j\right| > t\right) \leqslant \exp\left(-c(\alpha)\min\left(m\frac{t^2}{K^2}, m^\alpha \frac{t^\alpha}{K^\alpha}\right)\right). \tag{7}$$

Proof. It was proved by Latała ([14], Theorem 2 and Remark 2) that for any i.i.d. random variables $Z_1, \dots, Z_m$ such that $\mathbf{E}Z_1 = 0$, and any $p \geqslant 2$,

$$\|Z_1 + \cdots + Z_m\|_p \sim \sup\left\{\frac{p}{s}\left(\frac{m}{p}\right)^{1/s}\|Z_1\|_s;\ \max\left(2, \frac{p}{m}\right) \leqslant s \leqslant p\right\},$$

where $\sim$ means equality up to a universal multiplicative constant. With the hypotheses of the lemma, $\|Z_1\|_s \leqslant K s^{1/\alpha}$ and so

$$\|Z_1 + \cdots + Z_m\|_p \leqslant \begin{cases} C(\alpha) K \sqrt{mp} & \text{if } 2 \leqslant p \leqslant m^{\frac{\alpha}{2-\alpha}}, \\ C(\alpha) K p^{1/\alpha} & \text{if } p \geqslant m^{\frac{\alpha}{2-\alpha}}. \end{cases}$$

For simplicity we write C instead of $C(\alpha)$. Let $X = \frac{1}{m}(Z_1 + \cdots + Z_m)$. Using the Markov inequality we get

$$\mathbf{P}(|X| \geqslant t) \leqslant \begin{cases} \left(\frac{CK\sqrt{p}}{t\sqrt{m}}\right)^p & \text{if } 2 \leqslant p \leqslant m^{\frac{\alpha}{2-\alpha}}, \\ \left(\frac{CKp^{1/\alpha}}{tm}\right)^p & \text{if } p \geqslant m^{\frac{\alpha}{2-\alpha}}. \end{cases}$$

If $C'K/\sqrt{m} \leqslant t \leqslant 2CKm^{-\frac{1-\alpha}{2-\alpha}}$ we use the first bound with $p = mt^2/(4C^2K^2)$; if $t \geqslant 2CKm^{-\frac{1-\alpha}{2-\alpha}}$ we use the second bound with $p = (mt/(2CK))^\alpha$. In both cases we obtain $\mathbf{P}(|X| \geqslant t) \leqslant 1/2^p$, and the lemma is proved. □

We turn now to the proof of the main theorems.

Proof of Theorem 1. We write $A = A^N(X)$ and use the inequality

$$\|A - \mathrm{Id}\| \leqslant \|A - \mathrm{diag}(A)\| + \|\mathrm{diag}(A) - \mathrm{Id}\|.$$

To bound the first term, we use Proposition 1 for $N \geqslant 16C^2 n/\varepsilon^2$ and the Markov inequality to get

$$\mathbf{P}(\|A - \mathrm{diag}(A)\| \geqslant \varepsilon/2) \leqslant 2^{-n^{1/5}}.$$

For $1 \leqslant i \leqslant N$, we write $X_i = (x_{i1}, \ldots, x_{in})$. The second term can be estimated as follows:

$$\|\mathrm{diag}(A) - \mathrm{Id}\| = \max_{1 \leqslant j \leqslant n} \left| \frac{1}{N} \sum_{i=1}^{N} x_{ij}^2 - 1 \right|.$$

It is a direct consequence of Borell's lemma (see [5], or [18], p.135) that for some constant C,

$$\Psi_{\frac{1}{2}}(x_{ij}^2 - 1) \leqslant C,$$

and we get using Lemma 1 that for $0 < \varepsilon \leqslant 1$,

$$\mathbf{P}(\|\mathrm{diag}(A) - \mathrm{Id}\| \geqslant \varepsilon/2) \leqslant n \exp(-c\sqrt{n}).$$

This proves Theorem 1. □

Proof of Theorem 2. We get as a consequence of Proposition 1 a good estimate for the largest eigenvalue of A using the inequality $\|A\| \leqslant \|A - \mathrm{Id}\| + 1$ (or alternatively, using the moments method directly on A instead of $A - \mathrm{Id}$): for any $N \geqslant n$,

$$\mathbf{P}(\|A\| \geqslant C) \leqslant C \exp(-cn^{1/5}).$$

The smallest eigenvalue is automatically controlled by the following general proposition, proved in section 5. □

Proposition 2. *For every $M > 0$ and $\rho > 1$ there are constants $c = c(M, \rho)$ and $\kappa = \kappa(M, \rho)$ such that, for every random vector X uniformly distributed on an isotropic convex body and any $N \geqslant \rho n$, the empirical inertia matrix $A = A^N(X)$ automatically satisfies*

$$\mathbf{P}\left(\lambda_{\min}(A) \leqslant c\right) \leqslant \mathbf{P}\left(\lambda_{\max}(A) \geqslant M\right) + \exp(-\kappa n).$$

Remark. Usually in random matrix theory, it is substantially harder to deal with the smallest eigenvalue than with the largest; hence the above proposition may be surprising. We emphasize that this is only an isomorphic result.

4. Proof of Proposition 1: The moments method

Proof. Let us introduce the random matrix Γ defined as

$$\Gamma = \frac{1}{\sqrt{N}} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}. \tag{8}$$

We have the equality $A = \Gamma^t \Gamma$. We will use the moments method, with k an even integer. For $1 \leqslant i \leqslant N$, we write the coordinates of X_i as $(x_{ij})_{1 \leqslant j \leqslant n}$.

$$\begin{aligned} \mathbf{E}\|A - \operatorname{diag}(A)\|^k &\leqslant \mathbf{E}\operatorname{tr}(A - \operatorname{diag}(A))^k \\ &= \sum_{s_1 \neq s_2 \neq \cdots \neq s_k \neq s_1} \mathbf{E} A_{s_1 s_2} \dots A_{s_{k-1} s_k} A_{s_k s_1} \\ &= \frac{1}{N^k} \sum_{\substack{s_1 \neq s_2 \neq \cdots \neq s_k \neq s_1 \\ r_1, r_2, \dots, r_k}} \mathbf{E} x_{r_1 s_1} x_{r_1 s_2} x_{r_2 s_2} \dots x_{r_k s_k} x_{r_k s_1}, \end{aligned}$$

where the sum is taken over all indices $r_1, \dots, r_k$ in $\{1, \dots, N\}$ and $s_1, \dots, s_k$ in $\{1, \dots, n\}$, consecutively distinct. Since the random vectors $(X_1, \dots, X_N)$ are independent, each expectation appearing in the sum can be factorized as the product of N factors of the form

$$\mathbf{E} x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n} \tag{9}$$

for some integers $\alpha_1, \dots, \alpha_n$, where $X = (x_1, \dots, x_n)$. Since the vector X is unconditional, it is invariant under sign flips of coordinates, and this shows that the expectation (9) is zero if one of the α_i is odd. We are led to the following inequality:

$$\mathbf{E}\|A - \operatorname{diag}(A)\|^k \leqslant \frac{1}{N^k} \sum_{[(r_1,s_1),\dots,(r_{2k},s_{2k})] \text{V-graph}} \mathbf{E} x_{r_1 s_1} \dots x_{r_{2k} s_{2k}},$$

where a *V-graph* is a $2k$-tuple of pairs $(r_i, s_i) \in \{1, \dots, N\} \times \{1, \dots, n\}$ such that

- (V1) $s_{2i+1} = s_{2i}$ (and $s_1 = s_{2k}$),
- (V2) $r_{2i} = r_{2i-1}$,
- (V3) each couple (r_i, s_i) appears an even number of times,
- (V4) $s_{2i} \neq s_{2i-1}$.

We associate several parameters to a V-graph $G = [(r_1, s_1), \dots, (r_{2k}, s_{2k})]$, following the standard combinatorial techniques of [8, 23]. Let $r(G) = \#\{r_i\}$, $c(G) = \#\{s_i\}$ and $\ell(G) = r(G) + c(G)$. The number $r(G)$ (resp. $c(G)$) is the number of distinct row indices (resp. column indices) from Γ that appear in G. Let also $d(G) = \#\{(r_i, s_i)\}$ be the number of distinct couples of indices that appear in G. Let $n_2(G)$ be the number of indices i such that the couple (r_i, s_i) appears exactly 2 times in G and $n_+(G)$ be the number of indices i such that the couple (r_i, s_i) appears 4 times or more in G. We clearly have $n_2(G) + n_+(G) = 2k$.

Lemma 2. *If $G = [(r_1, s_1), \dots, (r_{2k}, s_{2k})]$ is a V-graph, then*

$$\mathbf{E} x_{r_1 s_1} \dots x_{r_{2k} s_{2k}} \leqslant C^k k^{n_+(G)}.$$

Proof. First, we use the fact that the rows of the matrix Γ are independent to write the whole expectation as a product of N factors of the form (9). Note that the sum of all exponents α_i appearing in the factors (9) is exactly $2k$, while the sum restricted to the exponents satisfying $\alpha_i \geqslant 4$ is exactly $n_+(G)$. We use the following comparison theorem by Bobkov and Nazarov. Let $B_1^n = \{x \in \mathbf{R}^n \text{ s.t. } \sum |x_j| \leqslant 1\}$ denote the unit ball of ℓ_1^n; the convex body $\alpha_n B_1^n$ is isotropic for $\alpha_n = \sqrt{(n+1)(n+2)/2}$. It has been proved in [4] that for some absolute constant C, if $X = (x_1, \ldots, x_n)$ is uniformly distributed on an isotropic unconditional convex body, and if $Y = (y_1, \ldots, y_n)$ is uniformly distributed on CnB_1^n, then for any increasing functions $f_i : \mathbf{R}^+ \to \mathbf{R}$,

$$\mathbf{E} \prod_{i=1}^n f_i(|x_i|) \leqslant \mathbf{E} \prod_{i=1}^n f_i(|y_i|).$$

On each factor (9), we use this result with $f_i(x) = x^{\alpha_i}$ (recall that α_i is necessarily even). The resulting expectation for Y can then be estimated using subindependence [3], a special property of ℓ_p^n-balls asserting that for any increasing functions $f_i : \mathbf{R}^+ \to \mathbf{R}$,

$$\mathbf{E} \prod_{i=1}^n f_i(|y_i|) \leqslant \prod_{i=1}^n \mathbf{E} f_i(|y_i|).$$

Now, by Borell's lemma (see [5], or [18] p.135), there is an absolute constant C such that $\Psi_1(y_i) \leqslant C$, which means that $\mathbf{E}|y_i|^p \leqslant (Cp)^p$. If $\alpha_i \geqslant 4$, we use the following bound:

$$\mathbf{E} y_i^{\alpha_i} \leqslant \left(\mathbf{E} y_i^{n_+(G)}\right)^{\alpha_i / n_+(G)} \leqslant (Cn_+(G))^{\alpha_i} \leqslant (2Ck)^{\alpha_i}.$$

Using this method to bound separately all the factors of the form (9), we are led to

$$\mathbf{E} x_{r_1 s_1} \ldots x_{r_{2k} s_{2k}} \leqslant (\mathbf{E} y_1^2)^{n_2(G)/2} (2Ck)^{n_+(G)},$$

and the lemma is proved. □

Applying Lemma 2, we obtain the inequality

$$\mathbf{E}\|A - \operatorname{diag}(A)\|^k \leqslant \left(\frac{C}{N}\right)^k \sum_{G \text{ V-graph}} k^{n_+(G)}. \tag{10}$$

We now state some bounds on the parameters associated to a V-graph.

Lemma 3. *Let $G = [(r_1, s_1), \ldots, (r_{2k}, s_{2k})]$ be a V-graph. Then*

(a) $d(G) \leqslant k$.
(b) $\ell(G) \leqslant d(G) + 1 \leqslant k + 1$.
(c) $r(G) \leqslant k/2$.
(d) $n_+(G) \leqslant 4(k - \ell(G) + 1)$.

Proof. Assertion (a) is an immediate consequence of property (V3). To prove (b), read the V-graph from (r_1, s_1) to (r_{2k}, s_{2k}). Each of the $d(G)$ first occurrences (r_i, s_i) of some couple of indices may bring a new row index or a new column index, but not both (except for $i = 1$) because of properties (V1-V2). This shows that $\ell(G) \leqslant d(G) + 1$, and (b) follows from (a). For (c), we use the following property which is a consequence of (V2) and (V4): whenever a row index r appears in G, we can find i and j such that $r_i = r_j = r$ and $s_i \neq s_j$. Hence, by property (V3) every row index that appears in G appears at least 4 times, and $r(G) \leqslant k/2$. For (d), note that $d(G) \leqslant \frac{1}{2}n_2(G) + \frac{1}{4}n_+(G)$ (with equality iff no couple (r,s) appears 6 times or more). The result then follows from (b) and the equality $n_2(G) + n_+(G) = 2k$. □

We now need a bound on the number of V-graphs with given r and c. This can be done using standard combinatorial techniques developed in [8, 23]. We present here a different approach, suggested to us by S. Szarek, based on the facts that the combinatorics of V-graphs do not depend on the specific entries of the random matrix, and that for Gaussian random matrices very precise information is available.

Lemma 4. *Let $I \subset \{1,\dots,N\}$ and $J \subset \{1,\dots,n\}$, such that $r + c \leqslant k+1$, with $r = \#I$ and $c = \#J$. Then the number of V-graphs $G = [(r_1,s_1),\dots,(r_{2k},s_{2k})]$ such that $\{r_i\} \subset I$ and $\{s_i\} \subset J$ is bounded from above by $(Ck)^k$, for some absolute constant C.*

Proof. Let $G_{r,c} = (g_{ij})$ be a random $r \times c$ matrix with entries being independent $N(0,1)$ random variables. It is well known (as a consequence of Slepian's lemma, see [7], chapter 2.3, or using a net argument) that for some absolute constant C, $\mathbf{E}\|G_{r,c}\| \leqslant C(\sqrt{r} + \sqrt{c}) \leqslant C'\sqrt{k}$. Moreover, the operator norm is a 1-Lipschitz function with respect to the entries of the matrix (in the Hilbert–Schmidt metric), and a standard concentration property of the Gaussian measure ([7], chapter 2.2) implies that for any $t \geqslant 0$,

$$\mathbf{P}(\|G_{r,c}\| \geqslant \mathbf{E}\|G_{r,c}\| + t) \leqslant \exp(-t^2/2).$$

This in turn implies that

$$\mathbf{E}\|G_{r,c}\|^{2k} = \int_0^\infty 2kt^{2k-1}\mathbf{P}(\|G_{r,c}\| \geqslant t)dt \leqslant (C\sqrt{k})^{2k}.$$

Also, since $\mathrm{tr}(G_{r,c}^t G_{r,c})^k$ is the sum of the $2k$th powers of singular values of $G_{r,c}$, we have

$$\mathbf{E}\mathrm{tr}(G_{r,c}^t G_{r,c})^k \leqslant \min(r,c)\mathbf{E}\|G_{r,c}\|^{2k} \leqslant (C'k)^k.$$

On the other hand, the quantity $\mathbf{E}\mathrm{tr}(G_{r,c}^t G_{r,c})^k$ itself can also be expanded as

$$\mathbf{E}\mathrm{tr}(G_{r,c}^t G_{r,c})^k = \sum \mathbf{E} g_{r_1 s_1} \dots g_{r_{2k} s_{2k}},$$

where the sum is taken over $2k$-tuples of pairs $(r_i, s_i) \subset I \times J$ satisfying conditions (V1), (V2), (V3). Notably, all V-graphs considered in the statement of the lemma enter this setting. Moreover, all terms in the sum are positive, and even larger than 1 since $\mathbf{E}g_{ij}^p \geqslant 1$ for p an even integer. This shows that the number of V-graphs with indices contained in $I \times J$ is bounded from above by $\mathbf{E}\mathrm{tr}(G_{r,c}^t G_{r,c})^k$, which proves the lemma. □

Lemma 4 implies that the number of V-graphs G such that $r(G) = r$ and $c(G) = c$ is bounded by

$$\binom{N}{r}\binom{n}{c}(Ck)^k \leqslant C_1^k N^r n^c \frac{k^k}{\ell(G)^{\ell(G)}} \leqslant C_2^k N^r n^c k^{k-\ell(G)},$$

where we used the inequalities $\binom{B}{b} \leqslant (Be/b)^b$, $r^r c^c \geqslant (\ell/2)^\ell$ and $k^k/\ell^\ell \leqslant (ke)^{k-\ell}$. By Lemma 3(c), for any V-graph G we have

$$\begin{aligned} N^{r(G)} n^{c(G)} &\leqslant \begin{cases} N^{\ell(G)} & \text{if } \ell(G) \leqslant k/2, \\ N^{k/2} n^{\ell(G)-k/2} & \text{if } \ell(G) \geqslant k/2 \end{cases} \\ &\leqslant N^k \left(\frac{n}{N}\right)^{k/2} \frac{1}{n^{k-\ell(G)}}. \end{aligned}$$

Gathering the V-graphs with the same ℓ, we get as a consequence of inequality (10) and Lemma 3(d) that (setting $m = l - 1$)

$$\begin{aligned} \mathbf{E}\|A - \operatorname{diag}(A)\|^k &\leqslant \left(\frac{C}{N}\right)^k \sum_{l=2}^{k+1} k^{4(k-l+1)} k C_2^k k^{k-l} N^k \left(\frac{n}{N}\right)^{k/2} \frac{1}{n^{k-l}} \\ &\leqslant \left(C\sqrt{\frac{n}{N}}\right)^k n \sum_{m=1}^{k} \left(\frac{k^5}{n}\right)^{k-m}. \end{aligned}$$

We now choose k to be the smallest even integer such that $k \geqslant n^{1/5}$. We then use the inequality $n \sum (k^5/n)^{k-m} \leqslant C^k$ to finish the proof (note that the l.h.s. in (6) is an increasing function of k). □

5. Proof of Proposition 2: The role of log-concavity

Proof. The proof is similar to the proof of the main theorem in [16], but here the log-concavity makes things much easier. We introduce the matrix Γ defined by (8), which we think of as an operator from ℓ_2^n to ℓ_2^N. We write $s_{\min}(\Gamma)$ for the smallest singular value of Γ, which equals $\sqrt{\lambda_{\min}(A)}$. For $\varepsilon > 0$ to be determined later, let $\mathcal{N}$ be an ε-net in S^{n-1} with cardinality smaller than $(3/\varepsilon)^n$ (existence of such a net is proved using volumetric arguments, see Lemma 4.10 in [20]). Set also $t = \varepsilon\sqrt{M}$. Let $\overline{\Omega}$ be the event "$\|\Gamma\| < \sqrt{M}$". By the standard approximation argument, the event

$$\overline{\Omega} \cap \{\exists x \in S^{n-1} \text{ s.t. } |\Gamma x| \leqslant t\}$$

is contained in the event

$$\overline{\Omega} \cap \{\exists x \in \mathcal{N} \text{ s.t. } |\Gamma x| \leqslant 2t\}.$$

Consequently,

$$\mathbf{P}(s_{\min}(\Gamma) \leqslant t) \leqslant \mathbf{P}(\overline{\Omega}^c) + \#\mathcal{N} \max_{x \in S^{n-1}} \mathbf{P}(|\Gamma x| \leqslant 2t).$$

For fixed x in the sphere S^{n-1} and j between 1 and N, let f_j be the random variable $\langle X_j, x\rangle$. It is well known [11, 9] that when K is an isotropic convex body in $\mathbf{R}^n$, the $(n-1)$-dimensional volume of hyperplane sections is controlled: for any affine hyperplane H, we have $\operatorname{vol}_{n-1}(K \cap H) \leqslant C$ for a universal constant C. Consequently, for any $s \geqslant 0$ we have $\mathbf{P}(|f_j| \leqslant s) \leqslant Cs$. Calculations are now straightforward:

$$\begin{aligned}
\mathbf{P}(|\Gamma x| \leqslant 2t) &= \mathbf{P}\left(\sum_{i=1}^{N} f_i^2 \leqslant 4t^2 N\right) \\
&= \mathbf{P}\left(N - \sum_{i=1}^{N} f_j^2/4t^2 \geqslant 0\right) \\
&\leqslant \mathbf{E}\exp\left(N - \sum_{i=1}^{N} f_j^2/4t^2\right) \\
&= \left(e\mathbf{E}\exp(-f_1^2/4t^2)\right)^N \\
&= e^N\left(\int_0^1 \mathbf{P}(\exp(-f_1^2/4t^2) > s)ds\right)^N \\
&= e^N\left(\int_0^1 \mathbf{P}(f_1 \leqslant 2t\sqrt{\log(1/s)})ds\right)^N \\
&\leqslant e^N\left(2Ct\int_0^1 \sqrt{\log(1/s)}ds\right)^N \\
&= (Ce\sqrt{\pi}t)^N,
\end{aligned}$$

and consequently,

$$\mathbf{P}(s_{\min}(\Gamma) \leqslant t) \leqslant \mathbf{P}(\|A\| \geqslant M) + \left(\frac{3\sqrt{M}}{t}\right)^n (Ce\sqrt{\pi}t)^N.$$

Thus for any $\rho > 1$, we can choose t (and thus ε) such that the conclusion of the Proposition holds. □

Acknowledgement

I want to thank Alain Pajor, Olivier Guédon and Sasha Sodin for many fruitful discussions, and Staszek Szarek for useful remarks on a preliminary version of this paper.

References

[1] G. Aubrun. Random points in the unit ball of ℓ_p^n. Positivity, to appear.
[2] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 1993. MR1235416 (94j:60060)
[3] K. Ball and I. Perissinaki. The subindependence of coordinate slabs in ℓ_p^n balls. *Israel J. Math.*, 107:289–299, 1998. MR1658571 (99k:52012)
[4] S. G. Bobkov and F. L. Nazarov. Large deviations of typical linear functionals on a convex body with unconditional basis. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 3–13. Birkhäuser, Basel, 2003. MR2073422 (2005f:52013)
[5] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216, 1975. MR0399402 (53:3246)
[6] J. Bourgain. Random points in isotropic convex sets. In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Math. Sci. Res. Inst. Publ.*, pages 53–58. Cambridge Univ. Press, Cambridge, 1999. MR1665576 (99m:60021)
[7] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 317–366. North-Holland, Amsterdam, 2001. MR1863696 (2004f:47002a)

[8] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980. MR0566592 (81m:60046)
[9] A. Giannopoulos. Notes on isotropic convex bodies (preprint). http://eudoxos.math.uoa.gr/~apgiannop/isotropic-bodies.ps, 2003.
[10] A. Giannopoulos, M. Hartzoulaki, and A. Tsolomitis. Random points in isotropic unconditional convex bodies. *J. London Math. Soc.* (2) 72(3):779–798, 2005. MR2190337
[11] D. Hensley. Slicing convex bodies—bounds for slice area in terms of the body's covariance. *Proc. Amer. Math. Soc.*, 79(4):619–625, 1980. MR0572315 (81j:52008)
[12] R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms*, 11(1):1–50, 1997. MR1608200 (99h:68078)
[13] V. F. Kolchin, B. A. Sevast′yanov, and V. P. Chistyakov. *Random allocations*. V. H. Winston & Sons, Washington, D.C., 1978. Translated from Russian, translation edited by A. V. Balakrishnan, Scripta Series in Mathematics. MR0471016 (57:10758b)
[14] R. Latała. Estimation of moments of sums of independent real random variables. *Ann. Probab.*, 25(3):1502–1513, 1997. MR1457628 (98h:60021)
[15] M. Ledoux. Deviation inequalities on largest eigenvalues. http://www.lsp.ups-tlse.fr/Ledoux/Jerusalem.pdf, 2005.
[16] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491–523, 2005. MR2146352
[17] V. D. Milman and A. Pajor. Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 64–104. Springer, Berlin, 1989. MR1008717 (90g:52003)
[18] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov. MR0856576 (87m:46038)
[19] G. Paouris. Concentration of mass on isotropic convex bodies. *C. R. Math. Acad. Sci. Paris*, 342(3):179–182, 2006. MR2198189
[20] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989. MR1036275 (91d:52005)
[21] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999. MR1694526 (2000c:60059)
[22] S. Sodin. On the smallest singular value of a Bernoulli random matrix (following Bai and Yin). Appendix to a preprint by Artstein–Friedland–Milman "More geometric applications of Chernoff inequalities".
[23] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields*, 78(4):509–521, 1988. MR0950344 (89g:60117)

Institut de Mathématiques de Jussieu, Projet Analyse Fonctionnelle, Université de Paris 6, 175 rue du Chevaleret, 75013 Paris, France

Current address: Institut Camille Jordan, Université de Lyon 1, 21 Avenue Claude Bernard, 69622 Villeurbanne, Cedex France