Restricted isometry property for random matrices with heavy tailed columns

Olivier GUÉDON[a] Alexander E. LITVAK[b] Alain PAJOR[a]
Nicole TOMCZAK-JAEGERMANN[b,1]

[a] *Université Paris-Est, Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), UPEMLV, F-77454 Marne-la-Vallée, France*

[b] *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

Received *****; accepted after revision +++++

Presented by œœœœœ

Abstract

Let A be a matrix whose columns $X_1, \ldots, X_N$ are independent random vectors in $\mathbb{R}^n$. Assume that p-th moments of $\langle X_i, a\rangle$, $a \in S^{n-1}$, $i \le N$, are uniformly bounded. For $p > 4$ we prove that with high probability A has the Restricted Isometry Property (RIP) provided that Euclidean norms $|X_i|$ are concentrated around $\sqrt{n}$ and that the covariance matrix is well approximated by the empirical covariance matrix provided that $\max_i |X_i| \le C(nN)^{1/4}$. We also provide estimates for RIP when $\mathbb{E}\,\phi\left(|\langle X_i, a\rangle|\right) \le 1$ for $\phi(t) = (1/2)\exp(t^\alpha)$, with $\alpha \in (0, 2]$.

Soit A une matrice dont les colonnes $X_1, \ldots, X_N$ sont des vecteurs indépendants de $\mathbb{R}^n$. On suppose que les moments d'ordre p des $\langle X_i, a\rangle$, $a \in S^{n-1}$, $1 \le i \le N$ sont uniformément bornés pour un $p > 4$. On démontre que si les normes euclidiennes des $|X_i|$ se concentrent autour de $\sqrt{n}$, la matrice A vérifie une propriété d'isométrie restreinte avec grande probabilité et que si $\max_i |X_i| \le C(nN)^{1/4}$ la matrice de covariance empirique est une bonne approximation de la matrice de covariance. On démontre aussi une propriété d'isométrie restreinte quand $\mathbb{E}\,\phi\left(|\langle X_i, a\rangle|\right) \le 1$ pour tout $a \in S^{n-1}$, $1 \le i \le N$ avec $\phi(t) = (1/2)\exp(t^\alpha)$ et $\alpha \in (0, 2]$.

1. Introduction. Our two main results go in two parallel directions: the Restricted Isometry Property abbreviated as RIP and a question of Kannan-Lovász-Simonovits about an approximation of a covariance matrix by empirical covariance matrices referred below as KLS problem.

In this note $X_1, \ldots, X_N$ denote independent random vectors in $\mathbb{R}^n$ satisfying for some function ϕ

$$\forall 1 \le i \le N \quad \forall a \in S^{n-1} \quad \mathbb{E}\,\phi\left(|\langle X_i, a\rangle|\right) \le 1. \tag{1}$$

Email addresses: olivier.guedon@u-pem.fr (Olivier GUÉDON), aelitvak@gmail.com (Alexander E. LITVAK), alain.pajor@u-pem.fr (Alain PAJOR), nicole.tomczak@ualberta.ca (Nicole TOMCZAK-JAEGERMANN).

[1] This author holds the Canada Research Chair in Geometric Analysis.

We will focus on two choices of the function ϕ: $\phi(t) = t^p$, with $p > 4$, or $\phi(t) = (1/2)\exp(t^\alpha)$, with $\alpha \in (0, 2]$. The $n \times N$ matrix whose columns are X_1, ..., X_N will be denoted by A. As usual, C, C_1, ..., c, c_1, ...will always denote absolute positive constants, whose values may change from line to line.

2. Restricted Isometry Property (RIP). We first recall the definition of the RIP, which was introduced in [6], in order to study the exact reconstruction problem by ℓ_1 minimization. It is noteworthy that the problem of reconstruction can be reformulated in terms of convex geometry, namely in terms of neighborliness of the symmetric convex hull of $X_1, \dots, X_N$, as was shown in [7].

Let T be an arbitrary $n \times N$ matrix. For $1 \le m \le N$ the isometry constant of T is defined as the smallest number $\delta_m = \delta_m(T)$ satisfying

$$(1 - \delta_m)|z|^2 \le |Tz|^2 \le (1 + \delta_m)|z|^2, \qquad \text{for } z \in \mathbb{R}^N \text{ with } |\mathrm{supp}(z)| \le m. \tag{2}$$

Let $\delta \in (0, 1)$. The matrix T is said to satisfy the RIP of order m with parameter δ if $0 \le \delta_m(T) < \delta$.

Returning to the vectors X_1, ..., X_N (independent and satisfying (1)) the concentration of $|X_i|$'s around their average is controlled by the function

$$P(\theta) := \mathbb{P}\left(\max_{i \le N}\left|\frac{|X_i|^2}{n} - 1\right| \ge \theta\right) \qquad \text{for } \theta \in (0, 1). \tag{3}$$

Note that in order to have RIP we need $P(\theta) < 1$, as the maximum under the probability equals to $\delta_1(A/\sqrt{n})$, which is less than or equal to $\delta_m(A/\sqrt{n})$.

Conditions saying that a random matrix satisfies RIP were investigated in many works. We refer to [8] and references therein. In [3] the authors studied the model when a random matrix consist of independent columns. It was proved that if X_i's are centered of variance 1 and satisfy assumption (1) with $\phi(t) = (1/2)\exp(t^\alpha)$, $\alpha \ge 1$ then A satisfies RIP with high probability. However, due to technical reasons, the case $\alpha < 1$ was left open. Moreover, it was not clear if RIP can hold under assumptions on moments of X_i's. In this note we show that A satisfies RIP not only in the case $\alpha < 1$ but also when the marginals of X_i's satisfy moments condition only, that is condition (1) with $\phi(t) = t^p$, $p > 4$. Note that in view of a result by Bai, Silverstein and Yin [5] it seems that one can't expect similar bounds when $p < 4$.

Theorem 1 **Case 1.** *Let $p > 4$ and $\phi(t) = t^p$. Let $0 < \varepsilon < \min\{1, (p-4)/4\}$ and $0 < \theta < 1$. Assume that $2^8/(\varepsilon\,\theta) \le N \le c\,\theta\,(c\,\varepsilon\,\theta)^{p/2}\, n^{p/4}$ and set*

$$m = \left[C(\varepsilon, p)\, \theta^{2p/(p-4-2\varepsilon)}\, n \left(\frac{N}{n}\right)^{-2(2+\varepsilon)/(p-4-2\varepsilon)}\right].$$

Case 2. *Let $\phi(t) = (1/2)\exp(t^\alpha)$. Assume that $8/\theta \le N \le c\,\theta\, \exp\left((1/2)\,(c\,\theta\sqrt{n})^\alpha\right)$ and set*

$$m = \left[C^{-2/\alpha}\,\theta^2\, n \left(\ln(C^{2/\alpha}\, N/(\theta^2\, n))\right)^{-2/\alpha}\right].$$

In both cases we have $\mathbb{P}\left(\delta_m(A/\sqrt{n}) \le \theta\right) \ge 1 - 2^{-9}\theta - P(\theta/2)$.

3. Kannan-Lovász-Simonovits problem (KLS). Let X_i's and A be as above and assume additionally that X_i's are identically distributed as a centered random vector X. KLS problem asks how fast the empirical covariance matrix $T := (1/N)AA^\top$ converges to the covariance matrix $\Sigma := (1/N)\mathbb{E}AA^\top$ (originally it was asked about so-called log-concave random vectors). In particular, is it true that with high probability the operator norm $\|T - \Sigma\| \le \varepsilon\|\Sigma\|$ for N being proportional to n? The corresponding important question in Random Matrix Theory is about the limit behavior of smallest and largest singular values. In the case of Wishart matrices, that is when the coordinates of X are i.i.d. random variables with finite fourth moment, the Bai-Yin theorem [4] states that the limits of minimal and maximal singular numbers of T are $(1 \pm \sqrt{\beta})^2$ as $n, N \to \infty$ and $n/N \to \beta \in (0, 1)$. Moreover, it is known [5] that boundedness of fourth moment is needed in order to have the convergence of the largest singular value. The asymptotic non-limit behavior (also called "non-asymptotic" in Statistics), i.e., sharp upper and lower bounds for singular values in terms of n and N, when n and N are sufficiently large were studied in several works. To keep the notation more compact and clear we set

$$M := \max_{i \le N} |X_i|, \qquad S := \sup_{a \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, a \rangle^2 - \mathbb{E} \langle X_i, a \rangle^2 \right) \right|.$$

Note that the bound $S \le \varepsilon$ is equivalent to bounds $1 \pm \varepsilon$ for minimal/maximal singular values. For Gaussian matrices it is known that singular values satisfy with probability close to one

$$S := \sup_{a \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, a \rangle^2 - \mathbb{E} \langle X_i, a \rangle^2 \right) \right| \le C \sqrt{n/N}. \tag{4}$$

In [1], [2] the same estimates were obtained for large class of random matrices, which in particular does not require that entries of the columns are independent or that X_i's are identically distributed. In particular it solves the original KLS problem. More precisely, under assumptions that X_i's satisfy condition (1) with $\phi(t) = e^t/2$ and that $M \le C(Nn)^{1/4}$ with high probability (both conditions hold for log-concave vectors). Of course, in view of Bai-Yin theorem the question raises if one can substitute the function $\phi(t) = e^t/2$ with the function $\phi = t^p$ for the "right" restriction $p \ge 4$. The first attempt in this direction was done in [10], where the bound $S \le C(p, K)(n/N)^{1/2-2/p}(\ln \ln n)^2$ was obtained for every $p > 4$ provided that $M \le K\sqrt{n}$. Clearly, $\ln \ln n$ is a "parasitic" term, which, in particular, does not allow to solve the KLS problem with N proportional to n. Very recently, the "right" upper bound $S \le C(n/N)^{1/2}$ was proved for $p > 8$ provided that $M \le C(Nn)^{1/4}$ ([9]). The purpose of our note is to show that one can solve the KLS problem in the case $p > 4$. Thus only the case $p = 4$ is left open.

Theorem 2 *Let $4 < p \le 8$ and $\phi(t) = t^p$. Let $\varepsilon \in (0, 1)$ and $\gamma = p - 4 - 2\varepsilon$. Then*

$$S \le C \left(\frac{1}{N} M^2 + C(p, \varepsilon) \left(\frac{n}{N} \right)^{\gamma/p} \right),$$

with probability larger than $1 - 8e^{-n} - 4\varepsilon^{-p/2} \max\{N^{-3/2}, n^{-(p/4-1)}\}$.

In particular Theorem 2 implies that if $M \le C_1(p, \varepsilon) n^{\gamma/(2p)} N^{1/2-\gamma/(2p)}$ with large probability, then $S \le C_2(p, \varepsilon) \left(n/N \right)^{\gamma/p}$ with large probability.

Proofs of both Theorems are based on estimates for fundamental parameters of sequences of independent random vectors. Let $1 \le k \le N$ and let $X_1, \cdots, X_N$ be random vectors in $\mathbb{R}^n$. We define A_k and B_k by

$$A_k := \sup_{\substack{a \in S^{N-1} \\ |\mathrm{supp}(a)| \le k}} \left| \sum_{i=1}^{N} a_i X_i \right| \quad \text{and} \quad B_k^2 := \sup_{\substack{a \in S^{N-1} \\ |\mathrm{supp}(a)| \le k}} \left| \left| \sum_{i=1}^{N} a_i X_i \right|^2 - \sum_{i=1}^{N} a_i^2 |X_i|^2 \right|. \tag{5}$$

The main technical result gives estimates for A_k and B_k.

Theorem 3 *Let $X_1, \dots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying (1). Let $p > 4$, $\sigma \in (2, p/2)$, $\alpha \in (0, 2)$, $t > 0$ and $\lambda \ge 1$. Define additional parameters M_1 and β in two cases.*

Case 1. $\phi(x) = x^p$. *We assume that $\lambda \le p$ and set*

$$M_1 := C_1(\sigma, \lambda, p) \sqrt{k} \left(\frac{N}{k} \right)^{\sigma/p} \quad \text{and} \quad \beta := C_2(\sigma, \lambda) N^{-\lambda} + C_3(\sigma, \lambda) \frac{N^2}{t^p}.$$

Case 2. $\phi(x) = (1/2)\exp(x^\alpha)$. *Assume that $\lambda > 2$ and set*

$$M_1 := (C\lambda)^{1/\alpha}\sqrt{k}\left(\ln\frac{N}{k}+\frac{1}{\alpha}\right)^{1/\alpha} \qquad \textit{and} \qquad \beta := 4N^{-\lambda}+\frac{N^2}{2\exp((2t)^\alpha)}.$$

In both cases assume that $\beta < 1/32$. Then with probability at least $1-\sqrt{\beta}$ one has

$$A_k \le (1-4\sqrt{\beta})^{-1}\left(M+18\sqrt{t}\sqrt{M}+M_1\right)$$

and

$$B_k^2 \le (1-4\sqrt{\beta})^{-2}\left(M^2+(730t+M_1)\,M+2M_1^2\right).$$

Combining definitions (2), (3) and (5) we see that the RIP is controlled by B_m and $P(\theta)$, and that Theorem 1 immediately follows from Theorem 3, with an appropriate choice of parameters ($\sigma = 2+\varepsilon$).

The proof of Theorem 2 requires several steps. Symmetrization and use of formulas for sums of k smallest order statistics of independent non-negative random variables with heavy tails reduce the problem of estimating S with large probability to estimates for A_k given in Theorem 3.

The proof of Theorem 3 is based on the study of suprema of bilinear forms of independent random vectors as developed in [9]. Let $X_1,\dots,X_N$ be independent random vectors in $\mathbb{R}^n$. We let for $1<k\le N$ and $I\subset\{1,...,N\}$,

$$Q_k(I) = \sup_{\substack{E\subset\{1,\dots,N\}\\ |E|\le k}}\ \sup_{a\in B_2^E}\left\langle \sum_{i\in E\cap I} a_iX_i,\ \sum_{j\in E\cap I^c} a_jX_j\right\rangle. \tag{6}$$

It turns out that if the X_i's satisfy (1) for $\phi(x)=x^p$ ($p>4$), then there is a recursive inequality for $Q_k(I)$: given $\varepsilon\in(0,1/2)$, $\gamma\in(1/2,1)$, and any $t>0$, we have, with large probability,

$$Q_k(I) \le \frac{Q_{[\gamma k]}(I)+tA_k}{1-2\varepsilon} \tag{7}$$

Here ε represents the size of an ε-net in B_2^E, over all E of dimension $\le k$, so it naturally appears in the estimate of probability. Iterating (7) we get an upper bound of the form $Q_k(I)\le C\,(tM+M_1A_k)$. This in turn leads to the required upper bounds by methods already used in earlier papers (see e.g., [1], [3],[9]).

References

[1] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave Ensembles,* J. Amer. Math. Soc. 23 (2010), no. 2, 535–561

[2] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Sharp bounds on the rate of convergence of empirical covariance matrix*, C.R. Math. Acad. Sci. Paris, 349 (2011), 195–200.

[3] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling,* Constructive Approximation, 34 (2011), 61–88.

[4] Z. D. Bai and Y. Q. Yin, *Limit of the smallest eigenvalue of a large dimensional sample covariance matrix,* Ann. Probab. 21 (1993), 1275–1294.

[5] Z. D. Bai, J. W. Silverstein and Y. Q. Yin, *A note on the largest eigenvalue of a large dimensional sample covariance matrix*, J. Mult. Anal. 26 (1988), 166–168.

[6] E.J. Candés and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), 4203–4215.

[7] D.L. Donoho, *Compressed sensing,* IEEE Trans. Information Theory 52 (2006), 1289–1306.

[8] D. Chafaï, O. Guédon, G. Lecué, A. Pajor, *Interactions between compressed sensing random matrices and high dimensional geometry*, Panoramas et Synthèses [Panoramas and Syntheses], 37. Soc. Math. de France, Paris, 2012.

[9] S. Mendelson, G. Paouris, *On the singular values of random matrices,* J. Eur. Math. Soc., to appear.

[10] R. Vershynin, *How close is the sample covariance matrix to the actual covariance matrix?*, J. Theoret. Probab. 25 (2012), 655–686.