Electron. Commun. Probab. **19** (2014), no. 83, 1–10.
DOI: 10.1214/ECP.v19-3807
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Lower bounds on the smallest eigenvalue of a sample covariance matrix*

Pavel Yaskov†

Abstract

We provide tight lower bounds on the smallest eigenvalue of a sample covariance matrix of a centred isotropic random vector under weak or no assumptions on its components.

Keywords: Covariance matrices; Gram matrices; Random matrices.
AMS MSC 2010: 60B20.
Submitted to ECP on September 17, 2014, final version accepted on December 2, 2014.

1 Introduction

Lower bounds on the smallest eigenvalue of a sample covariance matrix (or a Gram matrix) play a crucial role in the least squares problems in high-dimensional statistics (see, for example, [5]). These problems motivate the present work.

For a random vector X_p in $\mathbb{R}^p$, consider a random $p \times n$ matrix $\mathbf{X}_{pn}$ with independent columns $\{X_{pk}\}_{k=1}^n$ distributed as X_p and the Gram matrix

$$\mathbf{X}_{pn}\mathbf{X}_{pn}^\top = \sum_{k=1}^n X_{pk}X_{pk}^\top.$$

If X_p is centred, then $n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top$ is the sample covariance matrix corresponding to the random sample $\{X_{pk}\}_{k=1}^n$. For simplicity, we will further assume that X_p is isotropic, i.e. $\mathbb{E}X_pX_p^\top = I_p$ for a $p \times p$ identity matrix I_p, and consider only those p which are not greater than n (otherwise $\mathbf{X}_{pn}\mathbf{X}_{pn}^\top$ would be degenerate).

In this paper we derive sharp lower bounds for $\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top)$, where $\lambda_p(A)$ is the smallest eigenvalue of a $p \times p$ matrix A. We try to impose as few restrictions on the components of X_p as possible. In proofs we use the same strategy as in [6].

2 Main results

Put $c_p(a) = \inf \mathbb{E}\min\{(X_p, v)^2, a\}$, $C_p(a) = \sup \mathbb{E}(X_p, v)^2 \min\{(X_p, v)^2, a\}$,

$$L_p(\alpha) = \sup \mathbb{E}|(X_p, v)|^{2+\alpha} \quad \text{and} \quad K_p = \inf \mathbb{E}|(X_p, v)|$$

*Supported by RNF grant 14-21-00162 from the Russian Scientific Fund.
†Steklov Mathematical Institute of RAS, Russia. E-mail: yaskov@mi.ras.ru

for given $a, \alpha > 0$, where all suprema and infima are taken over $v \in \mathbb{R}^p$ with $\|v\| = 1$, and $\|v\| = (\sum_{i=1}^p v_i^2)^{1/2}$ is the Euclidean norm of $v = (v_1, \dots, v_p)$. Denote also by $M_p(\alpha)$ the infimum over all $M > 0$ such that

$$\mathbb{P}(|(X_p, v)| > t) \leqslant \frac{M}{t^{2+\alpha}} \quad \text{for all } t > 0 \text{ and } v \in \mathbb{R}^p,\ \|v\| = 1.$$

Our main lower bounds are as follows.

Theorem 2.1. *If X_p is an isotropic random vector in $\mathbb{R}^p$ and $p/n \leqslant y$ for some $y \in (0,1)$, then, for all $a > 0$,*

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant c_p(a) - \frac{C_p(a)}{a} - 5ay + \frac{\sqrt{C_p(2a)}Z}{\sqrt{n}}$$

for a centred random variable $Z = Z(p,n,a)$ with $\mathbb{P}(Z < -t) \leqslant e^{-t^2/2}$, $t > 0$.

Theorem 2.2. *Let X_p be an isotropic random vector in $\mathbb{R}^p$, $p/n \leqslant y$ for some $y \in (0,1)$. If $L_p(2) < \infty$, then*

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant 1 - 4C\sqrt{y} + \frac{CZ}{\sqrt{n}}$$

for $C = \sqrt{L_p(2)}$ and some $Z = Z(p,n)$ with $\mathbb{E}Z = 0$ and $\mathbb{P}(Z < -t) \leqslant e^{-t^2/2}$, $t > 0$. Moreover, there are universal constants $C_0, C_1, C_2 > 0$ such that

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant C_0 K_p^2 + \frac{C_1 Z}{\sqrt{n}}$$

whenever $y \leqslant C_2 K_p^2$ and $Z = Z(p,n)$ as above.

Useful bounds for $c_p(a)$ and $C_p(a)$ in terms of $L_p(\alpha)$ and $M_p(\alpha)$ are given in the following proposition.

Proposition 2.3. *Let X_p be an isotropic random vector in $\mathbb{R}^p$. Then, for all $a, \alpha > 0$,*

$$c_p(a) \geqslant 1 - \frac{L_p(\alpha)}{a^{\alpha/2}} \quad \text{and} \quad c_p(a) \geqslant 1 - \frac{2\alpha^{-1}M_p(\alpha)}{a^{\alpha/2}}.$$

In addition, for all $\alpha \in (0,2]$ and each $a > 0$, $C_p(a)$ is bounded from above by

$$a^{1-\alpha/2}L_p(\alpha) \quad \text{and} \quad (1+2/\alpha)M_p(\alpha)a^{1-\alpha/2} + \begin{cases} 2M_p(\alpha)a^{1-\alpha/2}/(1-\alpha/2), & \alpha \in (0,2), \\ 2M_p(2)\log\max\{a,1\} + 1, & \alpha = 2. \end{cases}$$

3 Applications

We now describe different corollaries of Theorem 2.1 and Theorem 2.2. The next corollary extends Theorem 1.3 in [4] and Theorem 3.1 in [5] (for $A_i = X_{pi}X_{pi}^\top$).

Corollary 3.1. *Let X_p be an isotropic random vector in $\mathbb{R}^p$, $p/n \leqslant y$ for some $y \in (0,1)$ and $L_p(\alpha) < \infty$ for some $\alpha \in (0,2]$. Then, with probability at least $1 - e^{-p}$,*

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant 1 - C_\alpha y^{\alpha/(2+\alpha)},$$

where

$$C_\alpha = \begin{cases} 9(L_p(\alpha))^{2/(2+\alpha)}, & \alpha \in (0,2), \\ (4+\sqrt{2})\sqrt{L_p(2)}, & \alpha = 2. \end{cases}$$

Remark 3.2. One may further weaken assumptions in Corollary 3.1. Namely, one may assume that $M_p(\alpha) < \infty$ for some $\alpha \in (0,2)$. The conclusion of Corollary 3.1 will still hold with some $C_\alpha > 0$ that depends only on α and $M_p(\alpha)$. In the case $\alpha = 2$, one would have a lower bound of the form $1 - C_2\sqrt{y\log(e/y)}$ with $C_2 > 0$ depending only on $M_p(2)$.

Theorems 2.1 and 2.2 improve Theorem 2.1 in [6] as the next corollary shows.

Corollary 3.3. *Let X_p be an isotropic random vector in $\mathbb{R}^p$. If $L_p(\alpha) < \infty$ for some $\alpha \in (0,2)$ and $p/n \leqslant \varepsilon^{1+2/\alpha}/(10(4L_p(\alpha))^{2/\alpha})$, then*

$$\mathbb{E}\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant 1-\varepsilon.$$

The same conclusion holds if $L_p(2) < \infty$ and $n \geqslant 16L_p(2)\varepsilon^{-2}p$.

Let us formulate the final corollary that improves Theorem 3.1 in [4] for small K_p.

Corollary 3.4. *Let X_p be an isotropic random vector in $\mathbb{R}^p$. Then there are universal constants $C_0^*, C_1^*, C_2^* > 0$ such that, with probability at least $1-\exp\{-C_1^* K_p^4 n\}$,*

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant C_0^* K_p^2$$

when $p/n \leqslant C_2^* K_p^2$.

The range of applicability of Corollary 3.4 is very wide. Namely, there exist some universal constant $K > 0$ such that $K_p \geqslant K$ for a very large class of isotropic random vectors X_p. By Corollary 3.4, this means that $\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top)$ is separated from zero by an universal constant.

The existence of K follows from results related to Kashin's decomposition theorem. The infinite dimensional version of this theorem is given in Kashin [2] (for a proof, see [3]). It states the following.

> There is an universal constant $K > 0$ such that $L_2(0,1) = H_1 \oplus H_2$ for some linear subspaces of $H_i \subset L_2(0,1)$, $i = 1,2$, such that $\|x\|_1 \geqslant K\|x\|_2$ for all $x \in H_1 \cup H_2$, where $\|x\|_d$ is the standard norm in $L_d(0,1)$, $d = 1,2$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an underlying probability space. Assume that $\Omega = (0,1)$, $\mathcal{F}$ is the Borel σ-algebra and $\mathbb{P}$ is the Lebesgue measure. If all components of $X_p = (x_1, \dots, x_p)$ are in H_1, or all components of X_p are in H_2, then $K_p \geqslant K$.

If we consider only discrete random vectors X_p, we may say more. Namely, Kashin [1] proved that, for any $\delta > 0$ and all $N \in \mathbb{N}$, $\mathbb{R}^N$ contains a linear subspace H with $\dim H \geqslant (1-\delta)N$ such that $|e|_1 \geqslant K|e|_2$ for some $K = K(\delta) > 0$ not depending on N and all $e = (e_1, \dots, e_N) \in H$,[1] where

$$|e|_d = \Big(\frac{1}{N}\sum_{i=1}^N |e_i|^d\Big)^{1/d}, \quad d = 1,2.$$

In particular, if $\{e^{(k)}\}_{k=1}^p$ is *any* orthonormal system in H and $\{x^{(i)}\}_{i=1}^N$ are columns of the $p \times N$ matrix with rows $\{(e^{(k)})^\top\}_{k=1}^p$, then, for all $v = (v_1, \dots, v_p) \in \mathbb{R}^p$ with $\|v\| = \sqrt{\sum_{j=1}^p v_j^2} = 1$,

$$K = K\Big(\frac{1}{N}\sum_{i=1}^N |(x^{(i)}, v)|^2\Big)^{1/2} = K\Big|\sum_{k=1}^p v_k e^{(k)}\Big|_2 \leqslant \Big|\sum_{k=1}^p v_k e^{(k)}\Big|_1 = \frac{1}{N}\sum_{i=1}^N |(x^{(i)}, v)|.$$

If X_p is such that $\mathbb{P}(X_p = x^{(i)}) = 1/N$, $1 \leqslant i \leqslant N$, then $K_p \geqslant K = K(\delta)$.

4 Proofs.

In proofs of Theorem 2.1 and Theorem 2.2, we follow the strategy of Srivastava and Vershynin [6]. The key step is the following lemma.

[1] In fact, the Haar measure of such orthogonal matrices C that $H = CH_1$ satisfies this property is greater than $1 - 2^{-N}$ for some $K = K(\delta) > 0$, where $H_1 = \{(e_1, \dots, e_N) \in \mathbb{R}^N : e_i = 0,\ i \geqslant (1-\delta)N + 1\}$ (see [1]).

Lemma 4.1. *Let A be a $p \times p$ symmetric matrix with $A \succ 0$, $v \in \mathbb{R}^p$, $l \geqslant 0$, $\varphi > 0$,*

$$Q(l,v) = v^\top (A - lI_p)^{-1} v \quad \text{and} \quad q(l,v) = \frac{v^\top (A - lI_p)^{-2} v}{\operatorname{tr}(A - lI_p)^{-2}}, \tag{4.1}$$

hereinafter $A \succ 0$ means that A is positive definite. If $A - lI_p \succ 0$, $\operatorname{tr}(A - lI_p)^{-1} \leqslant \varphi$ and

$$\Delta = \frac{q(l,v)}{1 + 3\varphi q(l,v) + Q(l,v)},$$

then $A - (l + \Delta)I_p \succ 0$ and $\operatorname{tr}(A + vv^\top - (l + \Delta)I_p)^{-1} \leqslant \varphi$.

The proof of Lemma 4.1 is given in Appendix.

The strategy itself consists in the following. Let A_0 be a $p \times p$ zero matrix and

$$A_k = \sum_{j=1}^k X_{pj} X_{pj}^\top, \quad 1 \leqslant k \leqslant n.$$

Consider some $\varphi > 0$ and take $l_0 = -p/\varphi$ that satisfies $\operatorname{tr}(A_0 - l_0 I_p)^{-1} = \varphi$.

Put $l_k = l_{k-1} + \Delta_k$ for $1 \leqslant k \leqslant n$, where

$$\Delta_k = \frac{q_k(l_{k-1}, X_{pk})}{1 + 3\varphi q_k(l_{k-1}, X_{pk}) + Q_k(l_{k-1}, X_{pk})},$$

$Q_k(l_{k-1}, X_{pk})$ and $q_k(l_{k-1}, X_{pk})$ are defined as $Q(l,v)$ and $q(l,v)$ in (4.1) with $A = A_{k-1}$ and $v = X_{pk}$. Applying Lemma 4.1 iteratively, we infer that $\operatorname{tr}(A_k - l_k I_p)^{-1} \leqslant \varphi$ and $A_k - l_k I_p \succ 0$ for all $1 \leqslant k \leqslant n$. Therefore,

$$\lambda_p(\mathbf{X}_{pn} \mathbf{X}_{pn}^\top) = \lambda_p(A_n) \geqslant l_n = l_0 + \Delta_1 + \ldots + \Delta_n.$$

Let $\mathbb{E}_k = \mathbb{E}(\,\cdot\,| X_{p1}, \ldots, X_{pk})$, $1 \leqslant k \leqslant n$, and $\mathbb{E}_0 = \mathbb{E}$. We have

$$\lambda_p(n^{-1} \mathbf{X}_{pn} \mathbf{X}_{pn}^\top) \geqslant -\frac{p}{n\varphi} + \frac{1}{n} \sum_{k=1}^n \mathbb{E}_{k-1} \Delta_k + \frac{Y}{\sqrt{n}}, \tag{4.2}$$

where $Y = n^{-1/2} \sum_{k=1}^n (\Delta_k - \mathbb{E}_{k-1} \Delta_k)$.

To apply estimate (4.2), we need to choose φ and obtain good lower bounds for $\mathbb{E}_{k-1} \Delta_k$ as well as upper bounds for $\mathbb{P}(Y < -t)$, $t < 0$. The next lemmata which proofs are given in Appendix provide such bounds.

Lemma 4.2. *Let U and V be non-negative random variables. Then, for all $a > 0$,*

$$\mathbb{E}\frac{U}{1+V} \geqslant \frac{|\mathbb{E}\min\{U,a\}|^2}{\mathbb{E}\min\{U,a\} + \mathbb{E}V\min\{U,a\}}.$$

In addition, if $\mathbb{E}U = 1$, then $\mathbb{E}U/(1+V) \geqslant 1/(1 + \mathbb{E}UV)$. Moreover,

$$\mathbb{E}\frac{U}{1+V} \geqslant \frac{|\mathbb{E}\sqrt{U}|^2}{1 + \mathbb{E}V}.$$

Lemma 4.3. *Let X_p be an isotropic random vector in $\mathbb{R}^p$, $A, B \succ 0$ be a $p \times p$ symmetric matrices with $\operatorname{tr}(A) = 1$ and $\operatorname{tr}(B) \leqslant 1$ that are simultaneously diagonalisable. If*

$$\Delta = \frac{X_p^\top A X_p}{1 + b^{-1}(X_p^\top A X_p + X_p^\top B X_p/3)}$$

for some $b > 0$*, then, for any* $a > 0$*,*

$$\mathbb{E}\Delta \geqslant c_p(a) - \frac{5C_p(a)}{3b} \quad \textit{and} \quad \mathbb{E}\Delta^2 \leqslant C_p(b).$$

In addition, if $L_p(2) < \infty$*, then* $\mathbb{E}\Delta \geqslant 1 - 4L_p(2)b^{-1}/3$ *and* $\mathbb{E}\Delta^2 \leqslant L_p(2)$*. Moreover,*

$$\mathbb{E}\Delta \geqslant \frac{K_p^2}{1+4(3b)^{-1}}.$$

Lemma 4.4. *Let* $(D_k)_{k=1}^n$ *be a sequence of non-negative random variables adapted to a filtration* $(\mathcal{F}_k)_{k=1}^n$ *such that* $\mathbb{E}(D_k^2|\mathcal{F}_{k-1}) \leqslant 1$ *a.s. for* $k = 1, \dots, n$*, where* $\mathcal{F}_0$ *is the trivial* σ*-algebra. If*

$$Z = \frac{1}{\sqrt{n}} \sum_{k=1}^n (D_k - \mathbb{E}(D_k|\mathcal{F}_{k-1})),$$

then $\mathbb{P}(Z < -t) \leqslant \exp\{-t^2/2\}$ *for all* $t > 0$*.*

Proof of Theorem 2.1. Take in Lemma 4.3 $X_p = X_{pk}$,

$$A = \frac{(A_{k-1} - l_{k-1}I_p)^{-2}}{\operatorname{tr}(A_{k-1} - l_{k-1}I_p)^{-2}}, \quad B = (A_{k-1} - l_{k-1}I_p)^{-1}/\varphi, \quad a = \frac{1}{5\varphi}, \quad b = \frac{5a}{3} = \frac{1}{3\varphi}. \tag{4.3}$$

Clearly A and B commute hence they are simultaneously diagonalizable. Additionally, we have $\operatorname{tr}(A) = 1$ and $\operatorname{tr}(B) = \operatorname{tr}(A_{k-1} - l_{k-1}I_p)^{-1}/\varphi \leqslant 1$. Using Lemma 4.3, we arrive at the lower bounds

$$\mathbb{E}_{k-1}\Delta_k \geqslant c_p(a) - \frac{C_p(a)}{a}, \quad 1 \leqslant k \leqslant n,$$

hereinafter all inequalities with conditional mathematical expectations hold almost surely. By (4.2), the latter implies that

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant c_p(a) - \frac{C_p(a)}{a} - \frac{5ap}{n} + \frac{\sqrt{C_p(2a)}Z}{\sqrt{n}},$$

where

$$Z = \frac{1}{\sqrt{C_p(2a)n}} \sum_{k=1}^n (\Delta_k - \mathbb{E}_{k-1}\Delta_k).$$

Note that $(\Delta_k - \mathbb{E}_{k-1}\Delta_k)_{k=1}^n$ is a martingale difference sequence with respect to the natural filtration of $(X_{pk})_{k=1}^n$. Obviously, $\mathbb{E}Z = 0$. By Lemma 4.3, $\mathbb{E}_{k-1}\Delta_k^2 \leqslant C_p(b) \leqslant C_p(2a)$. Therefore, Lemma 4.4 with $D_k = \Delta_k/\sqrt{C_p(2a)}$ yields that $\mathbb{P}(Z < -t) \leqslant \exp\{-t^2/2\}$, $t > 0$. Thus we have proven Theorem 2.1. □

Proof of Theorem 2.2. The proof follows the same line as the proof of Theorem 2.1.

Assume first that $C^2 = L_p(2) < \infty$ and $p/n \leqslant y$ for some $y > 0$. Define $X_p^\top A X_p$ and $X_p^\top B X_p$ in the same way as in (4.3). Then, by Lemma 4.3 (with $\varphi = 1/(3b)$),

$$\mathbb{E}_{k-1}\Delta_k \geqslant 1 - 4C^2\varphi, \quad 1 \leqslant k \leqslant n.$$

Taking $\varphi = \sqrt{y}/(2C)$ in (4.2), we get $p/(n\varphi) \leqslant y/\varphi = 2C\sqrt{y}$ and

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant 1 - 4C\sqrt{y} + \frac{CZ}{\sqrt{n}},$$

where

$$Z = \frac{1}{C\sqrt{n}} \sum_{k=1}^n (\Delta_k - \mathbb{E}_{k-1}\Delta_k).$$

As in the proof of Theorem 2.1, it follows from Lemma 4.3 that $\mathbb{E}_{k-1}\Delta_k^2 \leqslant L_p(2) = C^2$, $1 \leqslant k \leqslant n$. Therefore, by Lemma 4.4, $\mathbb{P}(Z < -t) \leqslant \exp\{-t^2/2\}$, $t > 0$.

Finally, consider the case with $K_p > 0$ (the case with $K_p = 0$ is trivial). By Lemma 4.3 with $b = (3\varphi)^{-1}$ and $\varphi = 1/4$,

$$\mathbb{E}_{k-1}\Delta_k \geqslant \frac{K_p^2}{1+4\varphi} = \frac{K_p^2}{2}, \quad 1 \leqslant k \leqslant n.$$

Taking $p/n \leqslant y = K_p^2/16$ in (4.2), we get

$$\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant \frac{K_p^2}{4} + \frac{\sqrt{C_p(4/3)}Z}{\sqrt{n}}$$

for some Z with $\mathbb{P}(Z < -t) \leqslant \exp\{-t^2/2\}$, $t > 0$ (see the end of the proof of Theorem 2.1). Since $C_p(4/3) \leqslant 4/3$, the variable

$$Z_0 = \frac{\sqrt{C_p(4/3)}}{\sqrt{4/3}}Z$$

satisfies $\mathbb{P}(Z_0 < -t) \leqslant \exp\{-t^2/2\}$, $t > 0$. Replacing Z by Z_0, we get the result. $\square$

Proof of Proposition 2.3. If U is non-negative random variable with $\mathbb{E}U = 1$, then

$$\mathbb{E}\min\{U, a\} = \mathbb{E}U - \mathbb{E}(U-a)\mathbb{1}(U > a) \geqslant 1 - \mathbb{E}U\mathbb{1}(U > a) \geqslant 1 - \frac{\mathbb{E}U^{1+\alpha/2}}{a^{\alpha/2}},$$

$$\mathbb{E}\min\{U, a\} = \mathbb{E}U - \int_a^\infty \mathbb{P}(U > t)\,dt \geqslant 1 - \int_a^\infty \frac{M}{t^{1+\alpha/2}}\,dt \geqslant 1 - \frac{2M}{\alpha a^{\alpha/2}},$$

$$\mathbb{E}U\min\{U, a\} \leqslant \mathbb{E}U^{1+\alpha/2}a^{1-\alpha/2},$$

$$\begin{aligned}
\mathbb{E}U\min\{U, a\} \leqslant & a\mathbb{E}(U-a)I(U > a) + a^2\mathbb{P}(U > a) + \mathbb{E}\min\{U^2, a^2\} \\
= & a\int_a^\infty \mathbb{P}(U > t)\,dt + a\mathbb{P}(U > a) + \int_0^{a^2} \mathbb{P}(U^2 > t)\,dt \\
\leqslant & a\int_a^\infty \frac{M}{t^{1+\alpha/2}}\,dt + Ma^{1-\alpha/2} + \int_0^{a^2} f(t,\alpha)\,dt \\
\leqslant & (1+2/\alpha)Ma^{1-\alpha/2} + \begin{cases} 2Ma^{1-\alpha/2}/(1-\alpha/2), & \alpha \in (0,2), \\ 2M\log\max\{a,1\} + 1, & \alpha = 2, \end{cases}
\end{aligned}$$

where $M = \sup\{t^{1+\alpha/2}\mathbb{P}(U > t) : t > 0\}$, $f(t,\alpha) = Mt^{-1/2-\alpha/4}$ for $\alpha \in (0,2)$ and

$$f(t,2) = \begin{cases} Mt^{-1}, & t > 1, \\ 1, & t \in [0,1]. \end{cases}$$

Putting $U = (X_p, v)^2$ for given $v \in \mathbb{R}^p$ with $\|v\| = 1$ and taking the infimum or the supremum over such v in the above inequalities, we finish the proof. $\square$

Proof of Corollary 3.1. Consider the case $\alpha \in (0,2)$. Set $L = L_p(\alpha)$ and $y = p/n$. By Proposition 2.3,

$$c_p(a) - \frac{C_p(a)}{a} \geqslant 1 - \frac{2L}{a^{\alpha/2}} \quad \text{and} \quad C_p(2a) \leqslant L\,(2a)^{1-\alpha/2} \leqslant 2La^{1-\alpha/2}.$$

By Theorem 2.1,

$$\begin{aligned}\mathbb{P}(\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) < 1 - 4La^{-\alpha/2} - 5ay) &\leqslant \mathbb{P}\Big(\sqrt{C_p(2a)}Z/\sqrt{n} < -2La^{-\alpha/2}\Big)\\ &\leqslant \mathbb{P}(\sqrt{2La^{1-\alpha/2}}Z/\sqrt{n} < -2La^{-\alpha/2})\\ &\leqslant \exp\{-La^{-1-\alpha/2}n\}.\end{aligned}$$

Taking $y = La^{-1-\alpha/2}$, we get the desired inequality.

Consider the case $\alpha = 2$. By Theorem 2.2 with $y = p/n$ and $C = \sqrt{L_p(2)}$,

$$\mathbb{P}(\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) < 1-(4+\sqrt{2})C\sqrt{y}) \leqslant \mathbb{P}(CZ/\sqrt{n} < -\sqrt{2}C\sqrt{y}) \leqslant \exp\{-yn\} = \exp\{-p\}.$$

□

Proof of Corollary 3.3. Set $L = L_p(\alpha)$ for given $\alpha \in (0, 2)$. By Proposition 2.3,

$$c_p(a) - \frac{C_p(a)}{a} \geqslant 1 - \frac{2L}{a^{\alpha/2}}.$$

Therefore, taking in Theorem 2.1

$$a = (4L/\varepsilon)^{2/\alpha} \quad \text{and} \quad p/n \leqslant y = \frac{\varepsilon^{1+2/\alpha}}{10(4L)^{2/\alpha}},$$

we derive the first bound

$$\mathbb{E}\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant 1 - \frac{2L}{a^{\alpha/2}} - 5ay \geqslant 1 - \varepsilon.$$

Similarly, taking $y = \varepsilon^2/(16C^2)$ for $C = \sqrt{L_p(2)}$ in Theorem 2.2, we get that

$$\mathbb{E}\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) \geqslant 1 - 4C\sqrt{y} \geqslant 1 - \varepsilon.$$

□

Proof of Corollary 3.4. Let $C_0, C_1, C_2 > 0$ be such that the second bound in Theorem 2.2 holds. Then, for $p/n \leqslant C_2K_p^2$,

$$\mathbb{P}(\lambda_p(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top) < C_0K_p^2/2) \leqslant \mathbb{P}(C_1Z/\sqrt{n} < -C_0K_p^2/2) \leqslant \exp\{-C_0^2K_p^4n/(8C_1^2)\}.$$

Putting $C_0^* = C_0/2$, $C_1^* = C_0^2/(8C_1^2)$ and $C_2^* = C_2$, we finish the proof. □

5 Appendix

Proof of Lemma 4.1. By Lemma 2.2 in Srivastava and Vershynin [6], if $A - (l+\Delta)I_p \succ 0$ and $q(l+\Delta, v)/[1 + Q(l+\Delta, v)] \geqslant \Delta$, then

$$\operatorname{tr}(A + vv^\top - (l+\Delta)I_p)^{-1} \leqslant \operatorname{tr}(A - lI_p)^{-1}.$$

In addition, by Lemma 2.4 in Srivastava and Vershynin [6], if $A - lI_p \succ 0$, $\Delta < 1/\varphi$ and $\operatorname{tr}(A - lI_p)^{-1} \leqslant \varphi$, then $A - (l+\Delta)I_p \succ 0$ and

$$\frac{q(l+\Delta, v)}{1 + Q(l+\Delta, v)} \geqslant \frac{q(l, v)(1-\varphi\Delta)^2}{1 + Q(l, v)(1-\varphi\Delta)^{-1}}.$$

Therefore, we only need to show that

$$\frac{q(l, v)(1-\varphi\Delta)^2}{1 + Q(l, v)(1-\varphi\Delta)^{-1}} \geqslant \Delta = \frac{q(l, v)}{1 + 3\varphi q(l, v) + Q(l, v)},$$

since $\Delta \leqslant 1/(3\varphi)$ by construction.

By Bernoulli's inequality, $(1-x)^3 \geqslant 1-3x$ whenever $x \in [0,1]$. Hence,

$$\frac{q(l,v)(1-\varphi\Delta)^2}{1+Q(l,v)(1-\varphi\Delta)^{-1}} = \frac{q(l,v)(1-\varphi\Delta)^3}{1-\varphi\Delta+Q(l,v)} \geqslant \frac{q(l,v)(1-\varphi\Delta)^3}{1+Q(l,v)} \geqslant \frac{q(l,v)(1-3\varphi\Delta)}{1+Q(l,v)} = \Delta,$$

where the last equality holds by the definition of Δ. □

Proof of Lemma 4.2. We have

$$\mathbb{E}\frac{U}{1+V} \geqslant \mathbb{E}\frac{\min\{U,a\}}{1+V}$$

for all $a>0$. By the Cauchy-Schwartz inequality,

$$\mathbb{E}\frac{\min\{U,a\}}{1+V}\mathbb{E}(1+V)\min\{U,a\} \geqslant \left|\mathbb{E}\frac{\sqrt{\min\{U,a\}}}{\sqrt{1+V}}\sqrt{(1+V)\min\{U,a\}}\right|^2 = |\mathbb{E}\min\{U,a\}|^2.$$

This gives the first inequality. Tending a to infinity, we get the second inequality.

The last inequality also follows from the Cauchy-Schwartz inequality. Namely,

$$\mathbb{E}\frac{U}{1+V}\,\mathbb{E}(1+V) \geqslant \left|\mathbb{E}\frac{\sqrt{U}}{\sqrt{1+V}}\sqrt{1+V}\right|^2 = |\mathbb{E}\sqrt{U}|^2.$$

□

Proof of Lemma 4.3. Let $\{v_1,\dots,v_p\}$ be an orthonormal basis of $\mathbb{R}^p$ such that

$$A = \sum_{i=1}^p a_i v_i v_i^\top \quad \text{and} \quad B = \sum_{i=1}^p b_i v_i v_i^\top,$$

where $a_1,\dots,a_p,b_1,\dots,b_p>0$ are eigenvalues of A and B. Since $\mathrm{tr}A = \sum_{i=1}^p a_i = 1$, $X_p^\top A X_p = \sum_{i=1}^p a_i (X_p,v_i)^2$ and the function $f(x) = x/(1+c(x+d))$ is concave on $\mathbb{R}_+$ for any $c,d \geqslant 0$, we have (for Δ defined in Lemma 4.3)

$$\Delta \geqslant \sum_{i=1}^p a_i \Delta_i \quad \text{for} \quad \Delta_i = \frac{(X_p,v_i)^2}{1+b^{-1}((X_p,v_i)^2 + X_p^\top B X_p/3)}.$$

Fix $j \in \{1,\dots,p\}$ and $b>0$. By Lemma 4.2,

$$\mathbb{E}\Delta_j \geqslant \frac{|\mathbb{E}\min\{(X_p,v_j)^2,a\}|^2}{\mathbb{E}\min\{(X_p,v_j)^2,a\}+b^{-1}C} \quad \text{and} \quad \mathbb{E}\Delta_j \geqslant \frac{(\mathbb{E}|(X_p,v_j)|)^2}{1+b^{-1}(1+\mathrm{tr}B/3)} \geqslant \frac{K_p^2}{1+4/(3b)},$$

where $C = \mathbb{E}((X_p,v_j)^2 + X_p^\top B X_p/3)\min\{(X_p,v_j)^2,a\}$. By the second inequality,

$$\mathbb{E}\Delta \geqslant \sum_{i=1}^p a_i \frac{K_p^2}{1+4/(3b)} = \frac{K_p^2}{1+4/(3b)}.$$

We have $x^2/(x+c) \geqslant x-c$ for all $x,c \geqslant 0$. This yields that

$$\frac{|\mathbb{E}\min\{(X_p,v_j)^2,a\}|^2}{\mathbb{E}\min\{(X_p,v_j)^2,a\}+b^{-1}C} \geqslant \mathbb{E}\min\{(X_p,v_j)^2,a\} - b^{-1}C.$$

We need to bound C from above. Obviously, $\mathbb{E}(X_p,v_j)^2\min\{(X_p,v_j)^2,a\} \leqslant C_p(a)$. In addition, since $x\min\{y,a\} \leqslant x\min\{x,a\} + y\min\{y,a\}$ for all $x,y,a \geqslant 0$, we have

$$\mathbb{E}(X_p^\top B X_p)\min\{(X_p,v_j)^2,a\} = \sum_{i=1}^p b_i \mathbb{E}(X_p,v_i)^2\min\{(X_p,v_j)^2,a\} \leqslant 2\mathrm{tr}B\cdot C_p(a) \leqslant 2C_p(a).$$

Hence, $C \leqslant 5C_p(a)/3$. Combining all estimates together yields

$$\mathbb{E}\Delta \geqslant c_p(a) - \frac{5C_p(a)}{3b}.$$

Let us now prove that $\mathbb{E}\Delta^2 \leqslant C_p(b)$. We have

$$\Delta^2 \leqslant \frac{(X_p^\top AX_p)^2}{(1+b^{-1}X_p^\top AX_p)^2} \leqslant \frac{(X_p^\top AX_p)^2}{1+b^{-1}X_p^\top AX_p}.$$

Consider the function $f(x) = x^2/(1+b^{-1}x)$, $x \geqslant 0$. Its derivative

$$f'(x) = \frac{2x}{1+b^{-1}x} - \frac{b^{-1}x^2}{(1+b^{-1}x)^2} = \frac{2x+b^{-1}x^2}{(1+b^{-1}x)^2} = b\frac{2bx+x^2}{(b+x)^2} = b\Big(1 - \frac{b^2}{(b+x)^2}\Big)$$

is increasing on $\mathbb{R}_+$. This means that $f = f(x)$ is convex and

$$\mathbb{E}\frac{(X_p^\top AX_p)^2}{1+a^{-1}X_p^\top AX_p} \leqslant \sum_{i=1}^p a_i \mathbb{E}\frac{(X_p, v_i)^4}{1+b^{-1}(X_p,v_i)^2} \leqslant \sum_{i=1}^p a_i \mathbb{E}(X_p,v_i)^2 \min\{(X_p,v_i)^2, b\}.$$

The latter gives the desired inequality $\mathbb{E}\Delta^2 \leqslant \mathrm{tr}A \cdot C_p(b) = C_p(b)$.

Now consider the case with $L_p(2) < \infty$. By Lemma 4.2,

$$\mathbb{E}\Delta \geqslant 1/[1+b^{-1}(\mathbb{E}(X_p^\top AX_p)^2 + \mathbb{E}(X_p^\top AX_p)(X_p^\top BX_p)/3)].$$

Since the function $f(x) = x^2$ is convex on $\mathbb{R}$, $X_p^\top AX_p = \sum_{i=1}^n a_i(X_p,v_i)^2$ and $\mathrm{tr}A = 1$, we get that

$$\mathbb{E}(X_p^\top AX_p)^2 \leqslant \sum_{i=1}^n a_i \mathbb{E}(X_p,v_i)^4 \leqslant L_p(2).$$

Similarly,

$$\mathbb{E}(X_p^\top BX_p)^2 \leqslant (\mathrm{tr}B)^2 \mathbb{E}\Big(\frac{X_p^\top BX_p}{\mathrm{tr}B}\Big)^2 \leqslant L_p(2),$$

where we have used that $\mathrm{tr}B \leqslant 1$. Applying the Cauchy-Schwartz inequality yields that

$$\mathbb{E}(X_p^\top AX_p)(X_p^\top BX_p) \leqslant \sqrt{\mathbb{E}(X_p^\top AX_p)^2 \mathbb{E}(X_p^\top BX_p)^2} \leqslant L_p(2).$$

To finish the proof, we only need to note that

$$1/[1+b^{-1}(\mathbb{E}(X_p^\top AX_p)^2 + \mathbb{E}(X_p^\top AX_p)(X_p^\top BX_p)/3)] \geqslant \frac{1}{1+4L_p(2)b^{-1}/3} \geqslant 1 - \frac{4L_p(2)}{3b}.$$

□

Proof of Lemma 4.4. Since $e^{-x} \leqslant 1 - x + x^2/2$ for all $x \geqslant 0$, we have

$$\begin{aligned}\mathbb{E}(e^{-\lambda D_k}|\mathcal{F}_{k-1}) &\leqslant 1 - \lambda\mathbb{E}(D_k|\mathcal{F}_{k-1}) + \frac{\lambda^2\mathbb{E}(D_k^2|\mathcal{F}_{k-1})}{2} \\ &\leqslant 1 - \lambda\mathbb{E}(D_k|\mathcal{F}_{k-1}) + \frac{\lambda^2}{2} \\ &\leqslant \exp\{-\lambda\mathbb{E}(D_k|\mathcal{F}_{k-1}) + \lambda^2/2\}\end{aligned}$$

for any $\lambda > 0$. Therefore, $\mathbb{E}(e^{-\lambda(D_k - \mathbb{E}(D_k|\mathcal{F}_{k-1}))}|\mathcal{F}_{k-1}) \leqslant \exp\{\lambda^2/2\}$ and

$$\begin{aligned}\mathbb{P}\Big(\sum_{k=1}^n (D_k - \mathbb{E}(D_k|\mathcal{F}_{k-1})) < -t\sqrt{n}\Big) &\leqslant e^{-\lambda t\sqrt{n}}\mathbb{E}\exp\Big\{-\lambda\sum_{k=1}^n(D_k - \mathbb{E}(D_k|\mathcal{F}_{k-1}))\Big\} \\ &\leqslant \exp\{n\lambda^2/2 - \lambda t\sqrt{n}\},\end{aligned}$$

where the last bound could be obtained iteratively by the law of iterated mathematical expectations. Putting $\lambda = t/\sqrt{n}$, we derive that $\mathbb{P}(Z < -t) \leqslant \exp\{-t^2/2\}$, $t > 0$. □

References

[1] Kašin, B.S.: Section of some finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, **41**, (1977), 334–351. MR-0481792

[2] Kashin, B.S.: On a special orthogonal decomposition of the space $L_2(0,1)$. *Math. Notes*, **95**, (2014), 570–572.

[3] Krivine, J.-L.: On a theorem of Kashin. (French) *Seminaire d'Analyse Fonctionnelle 1983-1984*. Publ. Math. Univ. Paris VII, **20**, Univ. Paris VII, Paris, (1984), 21–26. MR-0825300

[4] Koltchinskii, V. and Mendelson, S.: Bounding the smallest singular value of a random matrix without concentration. arXiv:1312.3580.

[5] Oliveira, R.I.: The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. arXiv:1312.2903.

[6] Srivastava, N. and Vershynin, R.: Covariance estimation for distributions with $2+\varepsilon$ moments. *Ann. Probab.*, **41**, (2013), 3081–3111. MR-3127875

Acknowledgments. We thank B.S. Kashin and the referee for useful suggestions.