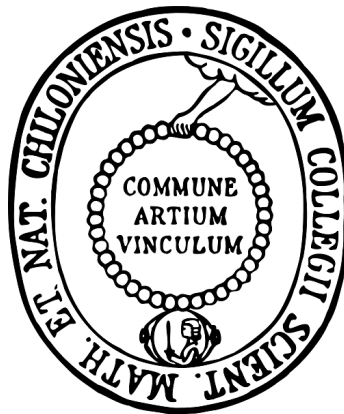


The smallest singular value of tensor matrices

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel



vorgelegt von
Colin Matthias Kleinschmidt

Kiel, Juli 2016

Erstgutachter:
Zweitgutachter:

Prof. Dr. Carsten Schütt
Prof. Dr. Uwe Rösler

Tag der mündlichen Prüfung:
Zum Druck genehmigt:

17.10.2016
Kiel, 17.10.2016

gez. Prof. Dr. Natascha Oppelt, Dekanin

Danksagung

Mein herzlicher Dank gilt meinem Betreuer Herrn Prof. Dr. Schütt für seine intensive Betreuung sowie die stets angenehmen und lehrreichen Diskussionen. Ebenso möchte ich mich bei Herrn Prof. Dr. Rösler bedanken für meine Stelle als wissenschaftlicher Mitarbeiter, die Möglichkeit zahlreiche Konferenzen und Workshops zu besuchen und die freundliche Aufnahme in die Stochastikgruppe an der Christian-Albrechts-Universität zu Kiel. Ich danke Herrn Prof. Dr. König für seine Anmerkungen und Ratschläge und ich danke Olaf Mordhorst für seine hilfreichen Hinweise und das Korrekturlesen meiner Arbeit.

Ich möchte mich außerdem bei meinen Mitdoktoranden für die nette Gemeinschaft und die vielen netten Gespräche bedanken. Zu guter Letzt danke ich meiner Familie und meiner Freundin Jennifer für ihre Liebe und Unterstützung.

Abstract

In this thesis we investigate the smallest singular value of random square matrices whose entries are independent subgaussian random variables. Statements about the behaviour of the singular values of a random matrix are expressed in stochastic terms as singular values of a random matrix are random variables themselves.

We estimate the order of the expectation of the smallest singular value of a Gaussian square matrix with disturbed last column, i.e. all entries of the random matrix are independent centered Gaussian random variables, but the variance of the entries of the last column have variance ε for an $\varepsilon > 0$ whereas the other entries are standard normal distributed random variables.

This example of a random square matrix with not necessarily identically distributed entries is crucial to understand the behaviour of the smallest singular value. In the more general case of tensor square matrices $(a_i b_j X_{i,j})_{i,j \leq n}$, where $a_1, \dots, a_n, b_1, \dots, b_n \neq 0$ are reals and $X_{i,j}$ are subgaussian random variables, we estimate the probability that the smallest singular value is smaller/greater than some bound. To do so we use a technique called *decomposition of the sphere* due to M. Rudelson and R. Vershynin, where the Euclidean sphere is decomposed in compressible vectors, i.e. unit vectors that carry most of their mass only in a few coordinates, and incompressible vectors, which are unit vectors that are not compressible. This represents a new approach in this context, as we use the properties of this decomposition to get in a position, where we can use Hölder's inequality and get rid of the factors $a_1, \dots, a_n, b_1, \dots, b_n$. We also generalize a result of M. Rudelson and R. Vershynin to estimate the tail probability of the smallest singular value explicitly and we show that if we lower the assumptions one can generalize the tail estimate to almost square matrices.

In the papers of M. Rudelson and R. Vershynin dealing with the invertibility of random square matrices whose entries are independent subgaussian random variables with variances at least one, it was a crucial step that compressible vectors do not have an great impact on the smallest singular value. We show that in our setting the situation changes completely.

Zusammenfassung

In dieser Arbeit beschäftigen wir uns mit der kleinsten singulären Zahl von quadratischen Zufallsmatrizen, deren Einträge unabhängige subgaußsche Zufallsvariable sind. Singuläre Zahlen von Zufallsmatrizen sind selbst wieder Zufallsvariablen, weswegen Aussagen über sie stochastischer Natur sind.

Wir bestimmen die Größenordnung des Erwartungswertes der kleinsten singulären Zahl einer quadratischen Gaußmatrix mit gestörter letzter Spalte, d.h. alle Einträge der Zufallsmatrix sind unabhängige, zentrierte Gaußvariablen, allerdings ist die Varianz der Einträge der letzten Spalte gleich ε für ein $\varepsilon > 0$, wohingegen die anderen Einträge standard normalverteilt sind.

Dieses Beispiel einer quadratischen Zufallsmatrix mit nicht notwendigerweise identisch verteilten Einträgen ist entscheidend um das Verhalten der kleinsten singulären Zahl zu verstehen. In dem allgemeineren Fall der quadratischen Tensormatrizen $(a_i b_j X_{i,j})_{i,j \leq n}$, wobei $a_1, \dots, a_n, b_1, \dots, b_n \neq 0$ reelle Zahlen und $X_{i,j}$ subgaußsche Zufallsvariablen sind, schätzen wir die Wahrscheinlichkeit, dass die kleinste singuläre Zahl kleiner/ größer als eine gewisse Schranke ist, ab. Wir benutzen dazu eine Technik namens *decomposition of the sphere*, die auf M. Rudelson und R. Vershynin zurückgeht, in welcher die Euklidische Sphäre in komprimierbare Vektoren, d.h. Einheitsvektoren, die den größten Teil ihrer Maße in wenigen Koordinaten tragen und nicht komprimierbare Vektoren zerlegt wird. Dies ist in diesem Zusammenhang ein neuartiges Vorgehen, da wir die Eigenschaften der Zerlegung nutzen, um uns in die Lage zu versetzen, die Hölder Ungleichung einsetzen zu können um die Faktoren $a_1, \dots, a_n, b_1, \dots, b_n$ zu eliminieren. Wir haben auch das Vorgehen von M. Rudelson und R. Vershynin verallgemeinert, um die Tail Wahrscheinlichkeit der kleinsten singulären Zahl explizit zu berechnen. Wir zeigen außerdem, dass man diese Abschätzungen auf Zufallsmatrizen, die geringfügig von der quadratischen Form abweichen, ausweiten kann, wenn man die Voraussetzungen lockert.

In den Arbeiten von M. Rudelson und R. Vershynin, die sich mit der Invertierbarkeit von quadratischen Zufallsmatrizen beschäftigen, deren Einträge unabhängige subgaußsche Zufallsvariablen mit Varianzen größer oder gleich eins sind, war es ein entscheidender Punkt, dass die komprimierbaren Vektoren keinen großen Einfluss auf die kleinste singuläre Zahl haben. Wir zeigen, dass dies in unserer Situation nicht der Fall sein muss.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Notation and preliminaries | 9 |
| 2.1 | Singular values | 9 |
| 2.2 | Subgaussian random variables | 15 |
| 2.3 | Central limit theorem | 18 |
| 3 | A first result and estimates for compressible vectors | 19 |
| 3.1 | Expected value of the smallest singular value of a Gaussian square random matrix with disturbed last column | 24 |
| 3.2 | Estimate for compressible vectors | 29 |
| 4 | Upper bound for the least singular value | 37 |
| 4.1 | Upper bound for tensor matrices | 40 |
| 4.2 | Upper bound for rectangular matrices | 50 |
| 5 | Lower bound for the least singular value | 59 |
| 5.1 | Estimate for Gaussian random matrices | 59 |
| 5.2 | General estimate | 66 |
| 5.3 | A new look at compressible vectors | 70 |
| 6 | Random simplices and connections to singular values of random matrices | 75 |
| 7 | Bibliography | 83 |

Chapter 1

Introduction

In this thesis we study the smallest singular value of real-valued random matrices, i.e. matrices whose entries are real valued random variables. Please note that throughout this text the shape of the matrices will be deterministic.

The theory of random matrices and their singular values was first studied by John von Neumann and his associates in their research concerning numerical inversion of large matrices. In these studies random matrices were used to test algorithms. Von Neumann speculated that with high probability the condition number

$$\sigma(X) = \sup_{\|x\|_2=1, \|y\|_2=1} \frac{\|Xx\|_2}{\|Xy\|_2}$$

of an $n \times n$ random matrix X is of order n , i.e. there exist absolute constants $c, C > 0$ such that

$$cn \leq \sigma(X) \leq Cn$$

with high probability (see [47], pp. 14, 477, 555). This problem can be reduced to a problem of estimating the *hard edge* of the spectrum, i.e. the largest singular value

$$s_1(X) = \sup_{\|x\|_2=1} \|Xx\|_2 = \|X\|_{2 \rightarrow 2}$$

and the smallest singular value

$$s_n(X) = \inf_{\|x\|_2=1} \|Xx\|_2,$$

since we have the equality

$$\sigma(X) = \frac{s_1(X)}{s_n(X)}.$$

We want to remind the reader of the fact that since X is a random matrix these quantities are random variables, too.

There are plenty of results on the largest singular value. We will mention just a few of them to give a short overview: 1988 Z. D. Bai, P. R. Krishnaiah, J. W. Silverstein and Y. Q. Yin showed in [48] and [1] that for an $n \times n$ random matrix X whose entries are independent identical distributed random variables with bounded fourth moment one has for some absolute constants $C, c > 0$ that

$$C\sqrt{n} \leq s_1(X) \leq c\sqrt{n}$$

with high probability. In [16] R. Latala showed the following for random matrices whose entries are independent normal distributed random variables, which are not necessarily identically distributed: For all $i, j = 1, \dots, n$ let $a_{i,j} \in \mathbb{R}_{\geq 0}$ and let $g_{i,j}$ be independent standard normal distributed random variables, then

$$\mathbb{E} \|(a_{i,j} g_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \lesssim \max_{1 \leq i \leq n} \|(a_{i,j})_{j \leq n}\|_2 + \max_{1 \leq j \leq n} \|(a_{i,j})_{i \leq n}\|_2 + \|(a_{i,j})_{i,j \leq n}\|_4.$$

A recent result by S. Riemer and C. Schütt is in the same spirit. In [22] they showed the following: For all $i, j = 1, \dots, n$ let $a_{i,j} \in \mathbb{R}_{\geq 0}$ and let $g_{i,j}$ be independent standard normal distributed random variables, then

$$\mathbb{E} \max_{1 \leq i \leq n} \|(a_{i,j} g_{i,j})_{j \leq n}\|_2 + \mathbb{E} \max_{1 \leq j \leq n} \|(a_{i,j} g_{i,j})_{i \leq n}\|_2 \lesssim \mathbb{E} \|(a_{i,j} g_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2}$$

and

$$\mathbb{E} \|(a_{i,j} g_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \lesssim (\ln(n))^{\frac{3}{2}} \left(\mathbb{E} \max_{1 \leq i \leq n} \|(a_{i,j} g_{i,j})_{j \leq n}\|_2 + \mathbb{E} \max_{1 \leq j \leq n} \|(a_{i,j} g_{i,j})_{i \leq n}\|_2 \right).$$

S. Riemer and C. Schütt also managed to generalize this result for $n \times n$ random matrices whose entries are independent centered random variables with finite second moment. We want to remark that the bounds above are dimension free, so one is able to estimate low-dimensional matrices in a high-dimensional space without the need of changing the norm. The results of S. Riemer and C. Schütt show that the expectation of the maximum column or the expectation of the maximum row of the random matrix X seem to be a good candidate, if one wants to estimate the expectation of the largest singular value of X .

Much less is known about the smallest singular value of random matrices. In 1988 and 1991 A. Edelman and S. Szarek showed in [7] and [40] independently from each other that for an $n \times n$ random matrix X whose entries are independent standard normal distributed random variables there exist absolute constants $C, c > 0$ such that

$$c \frac{1}{\sqrt{n}} \leq s_n(X) \leq C \frac{1}{\sqrt{n}}.$$

As well, A. Edelman as S. Szarek used the rotational invariance of the Gaussian distribution to come to their results. In fact, A. Edelman showed even more than the estimate above: He used the exact joint density function for the n eigenvalues of the Wishart matrix XX^T which is given by

$$K_{n,n} \exp\left(-\frac{1}{2} \sum_{i \leq n} \lambda_i\right) \prod_{i \leq n} \lambda_i^{-\frac{1}{2}} \prod_{i < j} (\lambda_i - \lambda_j) d\lambda_1 \dots d\lambda_n,$$

where

$$K_{n,n}^{-1} = \left(\frac{2^n}{\pi}\right)^{\frac{n}{2}} \prod_{i \leq n} \left(\Gamma\left(\frac{n-i+1}{2}\right)\right)^2$$

and derived the exact probability density function of the smallest eigenvalue λ_{\min} of the Wishart matrix XX^T which is in fact the square of the smallest singular value of X . The probability density function of λ_{\min} is given by

$$f_{\lambda_{\min}}(x) = \frac{n}{2^{n-\frac{1}{2}}} \frac{\Gamma(n)}{\Gamma(\frac{n}{2})} x^{-\frac{1}{2}} \exp\left(-x \frac{n}{2}\right) U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{x}{2}\right),$$

where, for $a > 0$ and $b < 1$, the Tricomi function, $U(a, b, z)$ is the unique solution to Kummer's equation

$$z \frac{d^2 w}{dz^2} + (b - z) \frac{dw}{dz} - aw = 0$$

satisfying $U(a, b, 0) = \frac{\Gamma(1 - b)}{\Gamma(1 + a - b)}$ and $U(a, b, \infty) = 0$.

A consequence of the above is that for an $n \times n$ random matrix X_n whose entries are independent standard normal distributed random variables the random variable $n\lambda_{\min}$ converges in distribution (as n tends to infinity) to a random variable whose probability density function is given by

$$f(x) = \frac{1 + \sqrt{x}}{2\sqrt{x}} \exp\left(-\frac{x}{2} - \sqrt{x}\right)$$

and therefore,

$$\left| \mathbb{P}\left(n(s_n(X_n))^2 \leq t\right) - \left(1 - \exp\left(-\frac{t}{2} - \sqrt{t}\right)\right) \right| \xrightarrow{n \rightarrow \infty} 0.$$

In 2010 T. Tao and V. Vu were able to generalize the results of A. Edelman by showing that the law of $n(s_n(X))^2$ is universal with respect to the choice of the entries $X_{i,j}$ ($i, j \leq n$). They proved that for random matrices X whose entries are independent copies of a normalized random variable ξ , i.e. in the real-valued case that $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = 1$, that under the condition that $\mathbb{E}|\xi|^{C_0} < \infty$ for some sufficiently large absolute constant C_0 , we have that for all $t > 0$

$$\left| \mathbb{P}\left(n(s_n(X))^2 \leq t\right) - \int_0^t \frac{1 + \sqrt{x}}{2\sqrt{x}} \exp\left(-\frac{x}{2} - \sqrt{x}\right) dx \right| \leq Cn^{-c},$$

where $c > 0$ is an absolute constant and $C > 0$ depends only on $\mathbb{E}|\xi|^{C_0}$ but is uniformly in t . For more information see [43].

In 2004 D.A. Spielman and S. Teng introduced the smoothed analysis of algorithms to investigate the phenomenon that some algorithms are known to work well in practice, but have a poor worst-case and average-case performance. Smoothed analysis can be seen as a hybrid of worst-case and average-case analysis, see [38] and [32] for more information on this method. In 2006 A. Sankar, D.A. Spielman and S. Teng proved in [32] that perturbations of arbitrary matrices are unlikely to have large condition numbers. Therefore, they showed that the smallest singular value of $n \times n$ matrices $A + X$, where X is an $n \times n$ random matrix, whose entries are identical standard normal distributed random variables and A is an arbitrary $n \times n$ matrix satisfies:

$$\mathbb{P}(s_n(A + X) \leq \varepsilon) \leq 2.35\varepsilon\sqrt{n}.$$

Studying the result of A. Sankar, D.A. Spielman and S. Teng it is natural to ask if one can show similar results for nongaussian matrices. In fact, surprisingly the answer to that is negative. This is due to T. Tao and V. Vu (see [44]). More precisely they showed the following: There exist constants $c, C > 0$ such that for every $2n \times 2n$ random matrix X_{2n} whose entries are ± 1 valued Bernoulli variables and for any $L \geq 2n$, there exist a $2n \times 2n$ deterministic matrix M such that $\|M\|_{2 \rightarrow 2} = L$ and

$$\mathbb{P}\left(s_{2n}(M + X_{2n}) \leq c \frac{2n}{L}\right) \geq \frac{C}{(2n)^{\frac{1}{2}}}.$$

Another approach was made in 2008 by M. Rudelson and R. Vershynin who studied the invertibility of random matrices. Again, as the smallest singular value $s_{\min}(A)$ is equal to the distance from the matrix A to the set of non-invertible matrices in the operator norm, these studies lead to

a need of estimating the probability that the smallest singular value of a random matrix lies within a small distance to the origin. In [28] the two authors showed that for an $n \times n$ random matrix X whose entries are independent real-valued random variables with variances at least one and fourth moments bounded by an absolute constant B , it follows that for every $\delta > 0$ there exist $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ depending only on δ and B , such that

$$\mathbb{P}(s_n(X) \leq \varepsilon\sqrt{n}) \leq \delta$$

for all $n \geq n_0$. Furthermore, they showed for the class of subgaussian random variables an even sharper estimate. First, recall that a random variable ξ is called subgaussian if its tail decays super-exponentially: There exists a $B > 0$ such that for all $t > 0$

$$\mathbb{P}(|\xi| > t) \leq 2 \exp\left(-\frac{t^2}{B^2}\right).$$

We will resume some more properties of subgaussian random variables in the sequel. For now, it is enough to mention that the minimal B , satisfying the inequality above, is called subgaussian moment of the random variable ξ .

One could think that the bound for Gaussian matrices somehow expands to random matrices with independent subgaussian entries. This is not true, since $n \times n$ random sign matrices, i.e. matrices with independent ± 1 -valued symmetric Bernoulli variables as entries, are not invertible with positive probability. Even the fact that the probability of a random sign matrix being singular converges to zero is nontrivial. The first one to prove this was J. Komlós in [15] in 1967. In 1995 J. Kahn, J. Komlós and E. Szemerédi bounded the probability exponentially by c^n , where $c \in (0, 1)$ is a universal constant (see [13]). Finding the optimal value of c is still an open problem.

In [28] M. Rudelson and R. Vershynin proved the following statement for random matrices with subgaussian entries: Let ξ_1, \dots, ξ_n be independent centered random variables with variances at least one and subgaussian moments bounded by an absolute constant $B > 0$. Define A to be the $n \times n$ matrix whose rows are independent copies of the random vector (ξ_1, \dots, ξ_n) . Then for every $\varepsilon \geq 0$ one has that

$$\mathbb{P}\left(s_n(A) \leq \varepsilon \frac{1}{\sqrt{n}}\right) \leq C\varepsilon + c^n,$$

where the constants $C > 0$ and $c \in (0, 1)$ depend only on B . Since the determinant of the matrix A equals the product of its singular values, one easily checks that for $\varepsilon = 0$ the statement above yields an exponential bound for the probability that A is singular and thus generalizes the result of J. Kahn, J. Komlós and E. Szemerédi. On the other hand the term $C\varepsilon$ is necessary if one thinks of the result of A. Sankar, D.A. Spielman and S. Teng for Gaussian matrices.

In [27] M. Rudelson and R. Vershynin also proved the reverse estimate: Let X be an $n \times n$ matrix whose entries are independent identical distributed centered subgaussian random variables with unit variance and subgaussian moment bounded by $B > 0$. Then for every $K \geq 2$ one has

$$\mathbb{P}\left(s_n(X) > K \frac{1}{\sqrt{n}}\right) \leq \frac{C}{K} \ln K + c^n,$$

where $C > 0$ and $c \in (0, 1)$ depend only on B . Again, this result can be generalized to the case of random square matrices whose entries are independent copies of a centered random variable with unit variance and bounded fourth moment. Together with the results in [28] they proved that the median of the least singular value is of order $\frac{1}{\sqrt{n}}$.

Altogether the results for the largest singular value and the results for the smallest singular value prove the conjecture of J. von Neumann: the condition number of random matrices whose entries are independent copies of a random variable with unit variance and bounded fourth moment is of order n with high probability.

We want to point out that the main difference when dealing with the smallest / largest singular value of an $n \times n$ random matrix X is that we do not have any deterministic candidate for a unit vector x which minimizes $\|Xx\|_2$ whereas the norm seems to be maximized for some vector of the standard basis $(e_i)_{i \leq n}$. Indeed, if one wants to estimate the least singular value of a random matrix one either has to use strong algebraic properties like A. Edelman did when he uses the rotational invariance of the Gaussian distribution to determine the joint law of the eigenvalues of the Wishart matrix or if one is interested in more general results one has to use a technique called *decomposition of the sphere* and has to work with random unit vectors. This is the main difficulty of estimating the least singular value.

Last but not least we want to mention the results from Y. Gordon, A. Litvak, C. Schütt and E. Werner in [11]. They proved the following: Let $\alpha, \beta > 0$. Let $p > 0$ and $(x_i)_{i \leq n}$ be a finite sequence of real numbers and ξ_1, \dots, ξ_n be random variables such that

$$\mathbb{P}(|\xi_i| \leq t) \leq \alpha t \text{ for every } t \geq 0$$

is satisfied. Then

$$\frac{1}{1+p} \alpha^{-p} \left(\sum_{i \leq n} \frac{1}{|x_i|} \right)^{-p} \leq \mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p.$$

Moreover, if the ξ_i 's are independent and their tails suffice the inequality

$$\mathbb{P}(|\xi_i| > t) \leq \exp(-\beta t) \text{ for every } t \geq 0$$

then

$$\mathbb{E} \min_{1 \leq i \leq n} |x_i \xi_i|^p \leq \beta^{-p} \Gamma(1+p) \left(\sum_{i \leq n} \frac{1}{|x_i|} \right)^{-p}.$$

For a sequence of real numbers x_1, \dots, x_n we define $k - \min_{1 \leq i \leq n} x_i$ to be the k -th smallest one, so one has $1 - \min_{1 \leq i \leq n} x_i = \min_{1 \leq i \leq n} x_i$ and $n - \min_{1 \leq i \leq n} x_i = \max_{1 \leq i \leq n} x_i$. Once again, let $\alpha, \beta > 0$ and $p > 0$. Let $2 \leq k \leq n$ and $0 < x_1 \leq x_2 \leq \dots \leq x_n$. Then Y. Gordon, A. Litvak, C. Schütt and E. Werner proved that under the condition that ξ_1, \dots, ξ_n are independent random variables, satisfying the inequalities from above, one has

$$\begin{aligned} \frac{1}{2e\alpha} \left(1 - \frac{1}{4\sqrt{\pi}} \right)^{\frac{1}{p}} \max_{1 \leq j \leq k} \frac{k+1-j}{\sum_{i=j}^n \frac{1}{x_i}} &\leq \left(\mathbb{E} k - \min_{1 \leq i \leq n} |x_i \xi_i|^p \right)^{\frac{1}{p}} \\ &\leq \beta^{-1} 4\sqrt{2} \max\{p, \ln(1+k)\} \max_{1 \leq j \leq k} \frac{k+1-j}{\sum_{i=j}^n \frac{1}{x_i}}. \end{aligned}$$

Clearly, for $p = 1$ these results from above can be seen as estimates for the expectation of the singular values of a diagonal matrix with diagonal entries $x_i \xi_i$.

Motivated by the results of S. Riemer and C. Schütt for the largest singular value of a random matrix with non-identically distributed entries we are interested in non-asymptotic estimates for

the least singular value of random matrices with non-identically distributed entries. As mentioned above there is no deterministic candidate for a unit vector to minimize the norm $\|Xx\|_2$. Therefore, we need some structure in the matrix. As they are important examples, whose largest singular values are well studied (see for example [6] or [22]), we study tensor matrices $(a_i b_j X_{i,j})_{i,j \leq n}$, where $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R} \setminus \{0\}$ are sequences of real numbers and the $X_{i,j}$ are independent copies of a centered subgaussian random variable with unit variance. To handle the problem of estimating the smallest singular value of tensor matrices we first considered the particular case of a Gaussian random matrix whose last column is disturbed. More specific we consider the following: For any $\varepsilon > 0$ we want to estimate the expectation of the smallest singular value of the matrix

$$G_\varepsilon = \begin{pmatrix} g_{1,1} & \cdots & g_{1,n-1} & \varepsilon g_{1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ g_{n,1} & \cdots & g_{n,n-1} & \varepsilon g_{n,n} \end{pmatrix},$$

where for $i, j \in \{1, \dots, n\}$ the $g_{i,j}$ are i.i.d. standard normal distributed random variables. In chapter 3 we will proof that

$$\left(\frac{1}{2}\right)^3 \min \left\{ \sqrt{\frac{\pi}{2}} \varepsilon, \frac{1}{\sqrt{n}} \right\} \leq \mathbb{E} s_n(G_\varepsilon) \leq \sqrt{2} \min \left\{ \frac{\varepsilon}{\sqrt{\pi}}, \frac{1}{\sqrt{n-1}} \right\}$$

which is quiet remarkable, as it shows that even for arbitrarily small $\varepsilon > 0$ it is not enough to estimate $\mathbb{E} s_n(G_\varepsilon)$ by the expectation of its last column since

$$\sqrt{\frac{2}{\pi}} \varepsilon \sqrt{n} \leq \mathbb{E} \varepsilon \|(g_{i,n})_{i \leq n}\|_2 \leq \varepsilon \sqrt{n}.$$

In the second part of chapter 3 we review results of M. Rudelson and R. Vershynin from [28] which are crucial to understand the difference between the well known setting in [28] and our situation.

Furthermore, in chapter 4 we will prove the following statement for random tensor matrices: For every $\lambda \geq 2$ and for all $a_1, \dots, a_n, b_1, \dots, b_n \in (0, 1]$ we have

$$\mathbb{P} \left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > 2\lambda \ln(\lambda^2) \min \left\{ \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(1/a_i)_{i \leq n}\|_2}, \frac{\|(a_i)_{i \leq n}\|_2}{\sqrt{n} \|(1/b_i)_{i \leq n}\|_2} \right\} \right) \leq \frac{6}{\lambda} + Qn^{-q}$$

where $X = (X_{i,j})_{i,j \leq n}$ is a random matrix, whose entries are independent copies of a centered subgaussian random variable with unit variance and $q, Q > 0$ are suitable constants. In the particular case that $a_1 = \dots = a_n = b_1 = \dots = b_n = 1$ our result coincides with the result from M. Rudelson and R. Vershynin.

We want to point out that the estimates for $\mathbb{E} s_n(G_\varepsilon)$ coincide with

$$\mathbb{P} \left(s_n(G_\varepsilon) > \lambda \min \left\{ \frac{1}{\sqrt{n-1 + \frac{1}{\varepsilon^2}}}, \frac{\sqrt{n-1 + \varepsilon^2}}{n} \right\} \right) \leq \left(\frac{C}{\lambda} \right) \ln(\lambda) + c^n + \exp(-c'n),$$

since

$$\min \left\{ \frac{1}{\sqrt{n-1 + \frac{1}{\varepsilon^2}}}, \frac{\sqrt{n-1 + \varepsilon^2}}{n} \right\} \sim \frac{1}{\sqrt{n-1 + \frac{1}{\varepsilon^2}}}.$$

The proofs of the formulas rely on the fact that for an invertible matrix X one has the following identity:

$$s_n(X) = \min_{\|x\|_2=1} \|Xx\|_2 = \frac{1}{\|X^{-1}\|_{2 \rightarrow 2}}.$$

This is a problem, if one wants to get analogue estimates for rectangular matrices. Still we were able to prove the following: Let $n \in \mathbb{N}$ and let $1 \leq k \leq n$ be a natural number. Let $b_1, \dots, b_{n-k} \neq 0$ be a sequence of real numbers and $g_{i,j}$ ($i = 1, \dots, n$ and $j = 1, \dots, n-k$) be i.i.d. standard normal distributed random variables. Define G as the $n \times (n-k)$ random matrix $G = (b_j g_{i,j})_{i \leq n, j \leq n-k}$. Then

$$\mathbb{P}\left(s_{n-k}(G) > 2\lambda \ln(\lambda^4) \frac{1}{\|(\frac{1}{b_j})_{j \leq n-k}\|_2}\right) \leq \frac{2(1 + 2^{\frac{k+1}{2}})}{\lambda}.$$

Furthermore, in chapter 5 we will show the reverse estimates. We will prove for an $n \times n$ random matrix $X = (a_j g_{i,j})_{i,j \leq n}$, where $a_1, \dots, a_n \in \mathbb{R} \setminus \{0\}$ and the $g_{i,j}$ are i.i.d. standard normal distributed random variables that for all $\varepsilon > 0$ we have

$$\mathbb{P}\left(s_n(X) \leq \frac{\varepsilon}{\|(\frac{1}{a_j})_{j \leq n}\|_2}\right) \leq 2.35\varepsilon,$$

where the constant 2.35 is due to the fact that we used some techniques of A. Sankar, D.A. Spielman and S. Teng (see [32, page 8]). Please notice, that for random matrices $X = (a_j g_{i,j})_{i,j \leq n}$ as above the lower bound and the upper bound coincide, since

$$\begin{aligned} & \mathbb{P}\left(s_n((a_j g_{i,j})_{i,j \leq n}) > 2\lambda \ln(\lambda^2) \min\left\{\frac{1}{\|(\frac{1}{a_j})_{j \leq n}\|_2}, \frac{\|(a_j)_{j \leq n}\|_2}{n}\right\}\right) \\ &= \mathbb{P}\left(s_n((a_j g_{i,j})_{i,j \leq n}) > 2\lambda \ln(\lambda^2) \frac{1}{\|(\frac{1}{a_j})_{j \leq n}\|_2}\right) \\ &\leq \frac{6}{\lambda} + Qn^{-q}, \end{aligned}$$

for suitable constants $q, Q > 0$. We will discuss this in detail in chapter 5. In chapter 5 we will also discuss a more general statement: Let $X = (X_{i,j})_{i,j \leq n}$ be a random matrix whose entries are independent centered real-valued random variables with variance at least 1 and subgaussian moments bounded by $B > 0$. Then for every $\varepsilon > 0$ there exists a constant $C > 0$ and a $c \in (0, 1)$ depending on B such that for all $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R} \setminus \{0\}$ and all $\delta, \rho \in (0, 1)$ we have

$$\begin{aligned} & \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\ &\leq C\varepsilon + 2c^n + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\ &+ \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n}^T x\|_2 \leq \lambda \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right), \end{aligned}$$

where $\text{Comp}(\delta, \rho)$ is a subset of the Euclidean sphere, which we will define more accurately later. For now it suffices to think of an element $x \in \text{Comp}(\delta, \rho)$, which we call a compressible vector, as an unit vector, who carries most of his mass only in a few coordinates. Again for $b_1 = \dots = b_n = 1$ one gets the familiar bound

$$\frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2} = \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}.$$

Coming from this theorem it is a natural question to bound the term

$$\begin{aligned} & \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\ &+ \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n}^T x\|_2 \leq \lambda \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right). \end{aligned}$$

In the previous work it seemed as if the compressible vectors don't have a big impact on the least singular value. We found out that if one considers tensor matrices as above, the situation changes completely.

Finally in chapter 6, we deal with random simplices, i.e. simplices whose vertices are chosen randomly. We study diverse quantities of random simplices and show a connection to the study of singular values of random matrices. Even though we did not manage to find a concrete geometric representation of the smallest singular value in a random simplex, the study of these objects inspired us regarding our main theorems.

Chapter 2

Notation and preliminaries

In this thesis $n \in \mathbb{N}$ denotes a large integer, i.e. an integer bigger than a suitable absolute constant. By $(e_k)_{k \leq n}$ we denote the canonical basis of the Euclidean space \mathbb{K}^n equipped with the canonical inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\|\cdot\|_2 = (\sum_{k \leq n} |x_k|^2)^{\frac{1}{2}}$, where $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. We denote the distance of a point $x \in \mathbb{K}^n$ to a set $A \subset \mathbb{K}^n$ as $\text{dist}(x, A) = \inf_{y \in A} \|x - y\|_2$.

More generally, we denote the l_p norm on \mathbb{K}^n as $\|x\|_p = (\sum_{k \leq n} |x_k|^p)^{\frac{1}{p}}$ for $1 \leq p < \infty$ and $\|x\|_\infty = \max_{k \leq n} |x_k|$. Furthermore, we denote the Euclidean ball and the Euclidean sphere as B_2^n and S^{n-1} . For $1 \leq p \leq \infty$ we also use the notation $\partial B_p^n(0, r) = \{x \in \mathbb{R}^n \mid \|x\|_p = r\}$.

Let $p, q \in [1, \infty]$. For a linear operator $T : l_p^n \rightarrow l_q^n$ we denote its operator norm as $\|T\|_{p \rightarrow q} = \sup_{\|x\|_p=1} \|Tx\|_q$. Furthermore, we write $\mathbb{1}_{n \times n}$ for the $n \times n$ identity matrix and $0_{n \times n}$ for the $n \times n$ matrix whose entries are all identical to zero.

In this thesis we denote by c, c', C, C' absolute constants independent from the underlying dimension. The values of the constants may change from line to line. We write $x \sim y$ if there exist absolute constants c, C such that $cx \leq y \leq Cx$.

Let X, Y be random variables, then we denote the probability with respect to X by \mathbb{P}^X . Furthermore, we denote the conditional expectation of X with respect to Y as $\mathbb{E}(X|Y)$.

2.1 Singular values

In the beginning we want to review some well known results from linear algebra. Our review relies on chapter one of [5] and chapter one of [41]. Let $m, n \in \mathbb{N}$ be natural numbers. A fundamental tool whenever one deals with matrices is the singular value decomposition, which expresses a diagonalization of the matrix up to unitary transformations of the space. The theory of singular values of $m \times n$ (non-)Hermitian matrices is a generalization of the theory of eigenvalues of $n \times n$ Hermitian matrices, in which the Singular Value Decomposition can be seen as a counterpart to the spectral theorem.

Furthermore, the extremal singular values, i.e. the largest and smallest/least singular value, bear essential information on the invertibility of the matrix. These two singular values are very natural geometric quantities: The $m \times n$ matrix A can be seen as a map from $l_2^m = (\mathbb{R}^m, \|\cdot\|_2)$ to l_2^n and therefore maps the Euclidean ball B_2^m to an ellipsoid \mathcal{E} . In this setting the largest and the least singular value are half of the length of the longest and the shortest axes of \mathcal{E} .

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ be either the complex or the real field. For an arbitrary $m \times n$ matrix A we denote by \overline{A}^T the conjugate transpose of A . We say that an $n \times n$ \mathbb{K} -valued matrix A is

\mathbb{K} -unitary if it suffices the equality

$$A\bar{A}^T = \mathbf{1}_{n \times n}.$$

We denote for every $n \times n$ matrix with \mathbb{C} -valued entries by $\lambda_1(A), \dots, \lambda_n(A)$ the eigenvalues of A which are the roots in \mathbb{C} of the characteristic polynomial $\det(A - X\mathbf{1}_{n \times n}) \in \mathbb{C}[X]$. In the sequel we label the eigenvalues of a matrix A so that $|\lambda_1(A)| \geq \dots \geq |\lambda_n(A)|$. The following version of the singular value decomposition theorem can be found in [5, page 2].

Theorem 1 (Singular Value Decomposition) *Let $m, n \in \mathbb{N}$. For every $m \times n$ matrix A with entries in \mathbb{K} , there exists a \mathbb{K} -unitary $m \times m$ matrix U , a \mathbb{K} -unitary $n \times n$ matrix V and a sequence of non-negative real numbers $s_1 \geq \dots \geq s_{m \wedge n} \geq 0$ such that*

$$\bar{U}^T AV = \text{diag}(s_1, \dots, s_{m \wedge n}),$$

where $\text{diag}(s_1, \dots, s_{m \wedge n})$ is an $m \times n$ diagonal matrix. Furthermore the sequence $s_1, \dots, s_{m \wedge n}$ does not depend on the particular choice of the matrices U, V .

Proof: Choose $v_1 \in \mathbb{K}^n$ with $\|v_1\|_2 \leq 1$ such that

$$\|Av_1\|_2 = \max_{\|x\|_2 \leq 1} \|Ax\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2 = \|A\|_{2 \rightarrow 2} = s_1(A).$$

Such a vector exists as $B_2^n = \{x \in \mathbb{K}^n \mid \|x\|_2 \leq 1\}$ is a compact set. If we can choose $v_1 = 0$, then it follows that $A = 0_{n \times n}$ is the matrix whose entries are all identical zero. In this case the stated result is trivial.

Therefore, we assume w.l.o.g. that $s_1(A) > 0$ and $\|v_1\|_2 = 1$. We define $u_1 = \frac{Av_1}{s_1(A)}$. Then we can find an $m \times m$ \mathbb{K} -unitary matrix U_1 with first column equal to u_1 and an $n \times n$ \mathbb{K} -unitary matrix V_1 whose first column equals v_1 . It follows that

$$\bar{U}_1^T AV_1 = \bar{U}_1^T \underbrace{\begin{pmatrix} Av_1 & AV_1' \end{pmatrix}}_{\in \mathbb{K}^{n \times n}},$$

where V_1' is the $n \times (n-1)$ matrix which we obtain from V by removing the first column. Furthermore, since the columns of \bar{U}_1^T form an orthonormal basis we obtain

$$\bar{U}_1^T AV_1 = \begin{pmatrix} \frac{\langle Av_1, Av_1 \rangle}{s_1(A)} & w \\ 0 & B_1 \end{pmatrix} = A_1$$

where $w = (w_1 \dots w_{n-1})$ is an $(n-1) \times 1$ matrix with entries in \mathbb{K} and B_1 is an $(m-1) \times (n-1)$ matrix with entries in \mathbb{K} . It follows that

$$\left\| A_1 \begin{pmatrix} s_1(A) \\ w_1 \\ \dots \\ w_{n-1} \end{pmatrix} \right\|_2^2 \geq \left(s_1(A)^2 + \left| \left\langle \begin{pmatrix} w_1 \\ \dots \\ w_{n-1} \end{pmatrix}, \begin{pmatrix} w_1 \\ \dots \\ w_{n-1} \end{pmatrix} \right\rangle \right|^2.$$

In particular, we have that

$$\|A_1\|_{2 \rightarrow 2}^2 \geq s_1(A)^2 + \left| \left\langle \begin{pmatrix} w_1 \\ \dots \\ w_{n-1} \end{pmatrix}, \begin{pmatrix} w_1 \\ \dots \\ w_{n-1} \end{pmatrix} \right\rangle \right| \geq \|A\|_{2 \rightarrow 2}^2$$

and since A and A_1 are unitary equivalent we even obtain equality

$$\|A_1\|_{2 \rightarrow 2}^2 = \|A\|_{2 \rightarrow 2}^2.$$

It follows that $w = (0 \dots 0)$. Therefore we have shown that

$$A_1 = \begin{pmatrix} s_1(A) & 0 \\ 0 & B_1 \end{pmatrix}.$$

Repeating the argument $m \wedge n$ times we obtain the desired result. \square

Definition 1 We call the non-negative real numbers $s_k(A) := s_k$, $k \in \{1, \dots, m \wedge n\}$ from Theorem 1 the singular values of the matrix A .

Remark: Let A be an $m \times n$ matrix and let U, V be \mathbb{K} -unitary matrices as in Theorem 1, then the columns of U are the eigenvectors of the $m \times m$ matrix $A\bar{A}^T$ and the columns of V are the eigenvectors of the $n \times n$ matrix $\bar{A}^T A$. $A\bar{A}^T$ and $\bar{A}^T A$ are both positive semidefinite Hermitian matrices, which share the same sequence of eigenvalues, up to the multiplicity of the eigenvalue 0. Furthermore we have

$$s_k(A) = \lambda_k(\sqrt{A\bar{A}^T}) = \sqrt{\lambda_k(A\bar{A}^T)} = \sqrt{\lambda_k(\bar{A}^T A)} = \lambda_k(\sqrt{\bar{A}^T A}) = s_k(\bar{A}^T), \quad (2.1)$$

for all $k \in \{1, \dots, m \wedge n\}$. Actually, if one thinks of the diagonal matrix $D = \text{diag}(s_1(A)^2, \dots, s_{m \wedge n}(A)^2)$ as an $\begin{cases} m \times m \\ n \times n \end{cases}$ matrix by appending as much zeros as needed, we have $\begin{cases} \bar{U}^T A\bar{A}^T U = D \\ \bar{V}^T \bar{A}^T A V = D. \end{cases}$

When A is a normal matrix, i.e. $A\bar{A}^T = \bar{A}^T A$, then we have that $m = n$ and $s_k(A) = |\lambda_k(A)|$ for $k \in \{1, \dots, n\}$.

Furthermore we want to remind the reader, that for any $m \times n$ matrix A , the matrices $A, \bar{A}, A^T, \bar{A}^T, UA, AU'$ share the same sequence of singular values, where U, U' denote arbitrary \mathbb{K} -unitary matrices.

The following minimax-formulas are very helpful if one wants to compute eigenvalues/ singular values explicitly, as they show that the i -th singular value functional $A \mapsto s_i(A)$ is a minimax expression of linear functionals.

We state two versions of the Courant-Fischer minimax theorem: The first one makes a statement for eigenvalues of an Hermitian matrix and can be found in [41, page 42]. The other one is a version for singular values, which we learned from Chafaï ([5, page 3]).

Theorem 2 (Courant-Fischer minimax theorem) Let $n, m \in \mathbb{N}$.

1. Let A be an $n \times n$ Hermitian matrix and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ the ordered eigenvalues of A . Then we have

$$\lambda_i(A) = \max_{\dim(V)=i} \min_{\substack{v \in V: \\ \|v\|_2=1}} \langle v, Av \rangle$$

and

$$\lambda_i(A) = \min_{\dim(V)=n-i+1} \max_{\substack{v \in V: \\ \|v\|_2=1}} \langle v, Av \rangle$$

for all $1 \leq i \leq n$, where V ranges over all subspaces of \mathbb{K}^n with the indicated dimension.

2. Let A be an $m \times n$ matrix, then we have for the singular values of A

$$s_i(A) = \max_{\dim(V)=i} \min_{\substack{x \in V: \\ \|x\|_2=1}} \|Ax\|_2$$

and

$$s_i(A) = \min_{\dim(V)=n-i+1} \max_{\substack{x \in V: \\ \|x\|_2=1}} \|Ax\|_2,$$

for all $1 \leq i \leq m \wedge n$ where V ranges over all subspaces of \mathbb{K}^n with the indicated dimension. In particular, we have

$$s_1(A) = \max_{\|x\|_2=1} \|Ax\|_2 \text{ and } s_{m \wedge n}(A) = \max_{\dim(V)=m \wedge n} \min_{\substack{x \in V: \\ \|x\|_2=1}} \|Ax\|_2,$$

where V ranges over all subspaces of \mathbb{K}^n with the indicated dimension.

An alternative formulation of the formulas above is

$$s_k(A) = \max_{\substack{\dim(V)=k \\ \dim(W)=k}} \min_{\substack{(x,y) \in V \times W \\ \|x\|_2=1=\|y\|_2}} \langle Ax, y \rangle .$$

Proof:

1. It is enough to prove

$$\lambda_i(A) = \max_{\dim(V)=i} \min_{v \in V: \|v\|_2=1} \langle v, Av \rangle$$

as

$$\lambda_i(A) = \min_{\dim(V)=n-i+1} \max_{v \in V: \|v\|_2=1} \langle v, Av \rangle$$

follows by replacing A by $-A$, since we have the identity

$$\lambda_i(-A) = -\lambda_{n-i+1}(A).$$

First, assume that $i = 1$, i.e. we have to verify

$$\lambda_1(A) = \max_{\dim(V)=1} \min_{v \in V: \|v\|_2=1} \langle v, Av \rangle = \max_{v \in \mathbb{K}^n: \|v\|_2=1} \langle v, Av \rangle .$$

By the spectral theorem for Hermitian matrices, we know that A has an orthonormal eigenbasis. W.l.o.g. we may assume that this eigenbasis is the standard basis of \mathbb{K}^n e_1, \dots, e_n . It follows that for every $v = (v_1, \dots, v_n)$ we have

$$\langle v, Av \rangle = \sum_{i \leq n} \lambda_i |v_i|^2.$$

Therefore, we have

$$\max_{v \in \mathbb{K}^n: \|v\|_2=1} \langle v, Av \rangle = \max_{v \in \mathbb{K}^n: \|v\|_2=1} \sum_{i \leq n} \lambda_i |v_i|^2 = \lambda_1.$$

Now we prove the other cases. Therefore, let $i = 2, \dots, n$. Again, by the spectral theorem we may assume that the eigenbasis of A is the standard basis of \mathbb{K}^n . Then we get

$$\begin{aligned} \lambda_i(A) &= \min_{v \in \text{span}\{e_1, \dots, e_i\}: \|v\|_2=1} \sum_{j \leq i} \lambda_j(A) |v_j|^2 \\ &= \min_{v \in \text{span}\{e_1, \dots, e_i\}: \|v\|_2=1} \langle v, Av \rangle \\ &\leq \max_{\dim(V)=i} \min_{v \in V: \|v\|_2=1} \langle v, Av \rangle . \end{aligned}$$

Now, we need to show the reverse inequality. Therefore, we show that every i -dimensional subspace V of \mathbb{K}^n contains a unit vector v such that

$$\langle v, Av \rangle \leq \lambda_i(A).$$

Let $W = \text{span}\{e_i, \dots, e_n\}$. Then W has dimension $n - i + 1$. It follows

$$V \cap W \neq \emptyset$$

for every subspace $V \subset \mathbb{K}^n$ with $\dim(V) = i$. Let $v \in V \cap W$ with $\|v\|_2 = 1$, then

$$\langle v, Av \rangle = \sum_{j=i}^n \lambda_j(A) |v_j|^2 \leq \lambda_i(A).$$

2. Again, it suffices to prove one of the stated inequalities, since they both follow directly from the related minimax formulae for Hermitian matrices. Let $1 \leq i \leq m \wedge n$. Then we have

$$\begin{aligned} s_i(A) &= \sqrt{\lambda_i(\overline{A}^T A)} \\ &= \sqrt{\max_{\dim(V)=i} \min_{v \in V: \|v\|_2=1} \langle v, \overline{A}^T Av \rangle} \\ &= \max_{\dim(V)=i} \min_{v \in V: \|v\|_2=1} \sqrt{\langle Av, Av \rangle} \\ &= \max_{\dim(V)=i} \min_{v \in V: \|v\|_2=1} \|Ax\|_2, \end{aligned}$$

which finishes the proof. \square

From the Courant-Fischer variational formulas, one can easily see, that the largest singular value of A is the operator norm of A for the Euclidean norm $\|\cdot\|_2$, namely

$$s_1(A) = \|A\|_{2 \rightarrow 2}.$$

Furthermore the map $A \mapsto s_1(A)$ is Lipschitz and convex. If A is an invertible $n \times n$ square matrix, then the singular values of A^{-1} are the inverses of the singular values of A . More precisely we have

$$s_k(A^{-1}) = (s_{n-k+1}(A))^{-1}, \text{ for all } k = 1, \dots, n.$$

This is due to the following consideration, where U and V denote the \mathbb{K} -unitary matrices of the singular value decomposition (see Theorem 1)

$$\begin{aligned} \mathbf{1}_{n \times n} &= AA^{-1} = \overline{U}^T AA^{-1}U = \overline{U}^T AV\overline{V}^T A^{-1}U = (\overline{U}^T AV)(\overline{U}^T A^{-1}V)^{-1} \\ &= \text{diag}(s_1(A), \dots, s_n(A))(\overline{U}^T A^{-1}V)^{-1}. \end{aligned}$$

It follows that

$$(\overline{U}^T AV)^{-1} = (\text{diag}(s_1(A), \dots, s_n(A)))^{-1} = \text{diag}(s_1^{-1}(A), \dots, s_n^{-1}(A)).$$

Moreover, an $n \times n$ square matrix A is invertible iff $s_n(A) > 0$ and in this case we have the following equality

$$s_n(A) = s_1(A^{-1})^{-1} = \|A^{-1}\|_{2 \rightarrow 2}^{-1}.$$

Again, the map $A \mapsto s_n(A)$ is Lipschitz, but it is not convex. We want to show the first property: Let A, B be two $m \times n$ matrices. Then

$$\begin{aligned} s_n(A) &= \min_{\|x\|_2=1} \|Ax\|_2 \\ &\leq \min_{\|x\|_2=1} (\|Ax - Bx\|_2 + \|Bx\|_2) \\ &\leq \|A - B\|_{2 \rightarrow 2} + \min_{\|x\|_2=1} \|Bx\|_2 \\ &= \|A - B\|_{2 \rightarrow 2} + s_n(B). \end{aligned}$$

A consequence of the Courant-Fischer variational formulas is the following theorem, which we learned from Chafaï in [5, page 4].

Theorem 3 (Cauchy interlacing by rows deletion) *Let A be an $m \times n$ matrix and $k \in \mathbb{N}$ with $1 \leq k \leq m \leq n$ and let B be an $(m - k) \times n$ matrix obtained from A by deleting k rows. Then for every $i \in \{1, \dots, m - k\}$,*

$$s_i(A) \geq s_i(B) \geq s_{i+k}(A).$$

Proof: W.l.o.g. we delete the last k rows. From the Courant-Fischer minimax theorem we know that for every $i = 1 \leq m - k$

$$\begin{aligned} s_i(A) &= \max_{\dim(V)=i} \min_{\substack{x \in V: \\ \|x\|_2=1}} \|Ax\|_2 \\ &= \max_{\dim(V)=i} \min_{\substack{x \in V: \\ \|x\|_2=1}} \left(\sum_{i \leq m} | \langle A_i, x \rangle |^2 \right)^{\frac{1}{2}} \\ &\geq \max_{\dim(V)=i} \min_{\substack{x \in V: \\ \|x\|_2=1}} \left(\sum_{i \leq m-k} | \langle A_i, x \rangle |^2 \right)^{\frac{1}{2}} \\ &= \max_{\dim(V)=i} \min_{\substack{x \in V: \\ \|x\|_2=1}} \|Bx\|_2 \\ &= s_i(B), \end{aligned}$$

where A_i denote the i -th row of A . Furthermore, we get

$$\begin{aligned} s_{i+k}(A) &= s_{i+k}(A^T) \\ &= \min_{\substack{\dim(V)=m-(i+k)+1 \\ V \subset \mathbb{K}^m}} \max_{\substack{x \in V: \\ \|x\|_2=1}} \|A^T x\|_2 \\ &\leq \min_{\substack{\dim(V)=(m-k)-i+1 \\ V \subset \mathbb{K}^{m-k}}} \max_{\substack{x \in V: \\ \|x\|_2=1}} \|A^T x\|_2 \\ &= \min_{\substack{\dim(V)=(m-k)-i+1 \\ V \subset \mathbb{K}^{m-k}}} \max_{\substack{x \in V: \\ \|x\|_2=1}} \|B^T x\|_2 \\ &= s_i(B^T) \\ &= s_i(B), \end{aligned}$$

which proves the theorem. □

In particular, we have $[s_{m-k}(B), s_1(B)] \subset [s_m(A), s_1(A)]$. So one can say row deletions causes a compression of the singular values interval. Alternatively, one can describe the phenomenon by saying that if we add a row to B then the largest singular value increases, while the least singular value decreases.

In random matrix theory one is often in a situation, where it is easier to work with a perturbation $A + B$ of the matrix A instead of working with A . In this context the matrix B is often considered to be *small* in some sense. In this situation it is often useful to have eigenvalue/ singular value inequalities relating A , B and $A + B$. One example of such an inequality is Weyl's inequality, which we deduce from the Courant-Fischer variational formulas for singular values. Weyl's inequality can be found in [41, page 53]

Theorem 4 (Weyl inequality) *Let A, B be $m \times n$ complex matrices for some $1 \leq m \leq n$. Then*

$$s_{i+j-1}(A + B) \leq s_i(A) + s_j(B)$$

whenever $1 \leq i, j, i + j - 1 \leq m$.

Proof: From Theorem 2 it follows that it is enough to show that every $i + j - 1$ -dimensional subspace V contains a unit vector v such that

$$\|(A + B)v\|_2 \leq s_i(A) + s_j(B).$$

But since

$$s_i(A) = \max_{\dim(V)=i} \min_{x \in V: \|x\|_2=1} \|Ax\|_2$$

by Theorem 2, there is a subspace U of codimension $i - 1$ such that

$$\|Av\|_2 \leq \max_{\dim(V)=i} \min_{x \in V: \|x\|_2=1} \|Ax\|_2 = s_i(A)$$

for all unit vectors $v \in U$. Analogue there exists a subspace $W \subset V$ of codimension $j - 1$ such that

$$\|Bv\|_2 \leq s_j(B)$$

for all unit vectors $v \in W$. Then $U \cap W$ has codimension less or equal than $i + j - 2$. Therefore, the intersection of $U \cap W$ and V is non-trivial, which finishes the proof. \square

2.2 Subgaussian random variables

In this section we will shortly introduce subgaussian random variables. Our review is based on [46]. Subgaussian random variables are a class of random variables, which generalize centered Gaussian random variables. To understand their defining properties, we briefly recall some basic properties of the standard normal random variable X .

If one looks at the tail of X , one has the following estimate

$$\mathbb{P}(|X| > t) = \frac{2}{\sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{x^2}{2}\right) dx \leq 2 \exp\left(-\frac{t^2}{2}\right),$$

for all $t \geq 0$. So the tail of a standard normal distributed random variable X decays super-exponentially. Another well known fact is about the absolute moments of a standard normal distributed random variable X :

$$c_p \sqrt{p} \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} = \sqrt{2} \left(\frac{\Gamma(\frac{1+p}{2})}{\Gamma(\frac{1}{2})} \right)^{\frac{1}{p}} \leq C_p \sqrt{p}$$

for all $p \geq 1$ and suitable constants $c_p, C_p > 0$. Furthermore, the moment generating function of a standard normal distributed random variable can be computed as

$$\mathbb{E} \exp(tX) = \exp\left(\frac{t^2}{2}\right).$$

As we will see by lowering these properties we will come to a more general class of random variables, which we will call subgaussian random variables.

Definition 2 (Subgaussian random variable) *A random variable X is subgaussian if there exists a $B > 0$ such that*

$$\mathbb{P}(|X| > t) \leq 2 \exp\left(\frac{-t^2}{B^2}\right)$$

for all $t \geq 0$. The minimal B here is called the subgaussian moment of X .

So subgaussian random variables are a class of random variables, whose tails are dominated by the tail of a centered Gaussian random variable. An interesting fact about the properties of subgaussian random variables is the following lemma, which can be found in [46] on page 9 ff. and in [45] lectures 3 and 5.

Lemma 1 *Let X be a random variable. Then the following properties are equivalent with parameters $K_i > 0$ differing from each other by at most an absolute constant factor.*

1. $\mathbb{P}(|X| > t) \leq \exp\left(1 - \frac{t^2}{K_1^2}\right)$ for all $t \geq 0$ (Tails).

2. $(\mathbb{E}|X|^p)^{\frac{1}{p}} \leq K_2\sqrt{p}$ for all $p \geq 1$ (Moments).

3. $\mathbb{E} \exp\left(\frac{X^2}{K_3^2}\right) \leq e$ (Super-exponential moment).

Moreover, if $\mathbb{E}X = 0$ then the properties above are also equivalent to the following one:

4. $\mathbb{E} \exp(tX) \leq \exp(t^2 K_4^2)$ for all $t \in \mathbb{R}$ (Moment generating function).

Proof: 1. \Rightarrow 2. Let us assume that property 1 holds. By rescaling X to $\frac{X}{K_1}$ we can assume w.l.o.g. that $K_1 = 1$. Then we have by change of variables

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}(|X|^p \geq t) dt = \int_0^\infty \mathbb{P}(|X| > s) p s^{p-1} ds.$$

Using property 1 we get that

$$\mathbb{E}|X|^p \leq \int_0^\infty \exp(1 - s^2) p s^{p-1} ds = \left(\frac{ep}{2}\right) \Gamma\left(\frac{p}{2}\right)$$

by the definition of the Gamma function. Using Stirling's formula we obtain

$$\mathbb{E}|X|^p \leq \left(\frac{ep}{2}\right) \left(\frac{p}{2}\right)^{\frac{p}{2}}.$$

Taking the p -th root yields property 2 with a suitable constant K_2 .

2. \Rightarrow 3. Let us assume that property 2 holds. Again by a rescaling argument we can assume w.l.o.g. that $K_2 = 1$. Now let $c > 0$ be a sufficiently small absolute constant, which we will determine later. We're using the Taylor series for the exponential function and the monotone convergence theorem to obtain

$$\mathbb{E} \exp(cX^2) = \mathbb{E} \sum_{k=0}^{\infty} \frac{c^k X^{2k}}{k!} = \sum_{k=0}^{\infty} \frac{c^k \mathbb{E} X^{2k}}{k!}.$$

Using property 2 and Stirling's formula we get that

$$\mathbb{E} \exp(cX^2) \leq \sum_{k=0}^{\infty} \frac{c^k (2k)^k}{k!} \leq \sum_{k=0}^{\infty} \left(\frac{2c}{e}\right)^k.$$

Let now $c \leq \frac{e-1}{2}$, then we have $\frac{2c}{e} \leq \frac{e-1}{e}$ and we can compute the geometric series as

$$\mathbb{E} \exp(cX^2) \leq \frac{1}{1 - \frac{2c}{e}} \leq e.$$

Therefore, for $c \leq \frac{e-1}{2}$ we have shown property 3 with $K_3 = \frac{1}{\sqrt{c}}$.

3. \Rightarrow 1. Assume property 3 holds. Again, w.l.o.g. we may assume that $K_3 = 1$. Using Markov's inequality and property 3 we obtain

$$\mathbb{P}(|X| \geq t) = \mathbb{P}(\exp(X^2) \geq \exp(t^2)) \leq \exp(-t^2) \mathbb{E} \exp(X^2) \leq \exp(1 - t^2).$$

2. \Rightarrow 4. Let X be a centered random variable and assume that property 2 holds. As before, we can assume that $K_2 = 1$. Again, we use Taylor's expansion:

$$\mathbb{E} \exp(tX) = 1 + t\mathbb{E}X + \sum_{k=2}^{\infty} t^k \frac{\mathbb{E}X^k}{k!} \leq 1 + \sum_{k=2}^{\infty} \frac{t^k \sqrt{k}^k}{k!},$$

where we used property 2 and $\mathbb{E}X = 0$. By Stirling's formula we obtain

$$\mathbb{E} \exp(tX) \leq 1 + \sum_{k=2}^{\infty} \left(\frac{e|t|}{\sqrt{k}} \right)^k.$$

Now, we have to differ two cases:

1) $|t| \leq \frac{1}{e}$, then the series is smaller than a geometric series and therefore

$$\mathbb{E} \exp(tX) \leq 1 + Ct^2 \leq \exp(ct^2)$$

for suitable constants $C, c > 0$.

2) $|t| > \frac{1}{e}$. Since, X suffices property 2 it also suffices the equivalent property 3 and we obtain that $\mathbb{E} \exp\left(\frac{X^2}{K_3^2}\right) \leq e$. Since

$$tX - \frac{K_3^2}{4}t^2 = -\frac{K_3^2}{4}\left(t - \frac{2X}{K_3^2}\right)^2 + \frac{X^2}{K_3^2} \leq \frac{X^2}{4c}$$

we get that

$$\mathbb{E} \exp\left(tX - \frac{K_3^2}{4}t^2\right) \leq \mathbb{E} \exp\left(\frac{X^2}{K_3^2}\right) \leq e.$$

Since $|t| > \frac{1}{e}$ it follows that

$$\mathbb{E} \exp(tX) \leq \exp\left(1 + \frac{K_3^2}{4}t^2\right) \leq \exp(ct^2)$$

for suitable $c > 0$.

4. \Rightarrow 1. Assume that property 4 holds. W.l.o.g. we may assume $K_4 = 1$. Let $u > 0$ be a positive real number, then we have by Markov's inequality

$$\mathbb{P}(X \geq t) = \mathbb{P}(\exp(uX) \geq \exp(tu)) \leq \frac{\mathbb{E} \exp(uX)}{\exp(tu)},$$

for all $t > 0$. Using property 4 we get

$$\mathbb{P}(X > t) \leq \exp(u^2 - tu).$$

We optimize in t and thus choose $t = \frac{u}{2}$. It follows that

$$\mathbb{P}(X \geq t) \leq \exp\left(\frac{-t^2}{4}\right).$$

Repeating the argument for the random variable $-X$, we can also conclude that

$$\mathbb{P}(X \leq -t) \leq \exp\left(\frac{-t^2}{4}\right).$$

Altogether, we get that

$$\mathbb{P}(|X| \geq t) \leq 2 \exp\left(\frac{-t^2}{4}\right) \leq \exp\left(1 - \frac{t^2}{4}\right).$$

Thus property 1 follows. □

Remark: We chose the constants 1 and e in the properties 1 and 3 in the lemma above for convenience: The constant 1 can be replaced by an arbitrary positive real number and the constant e can be replaced by any real number greater than 1.

Important examples of subgaussian random variables are of course Gaussian random variables, which follows easily from the definition. Another not so obvious, but yet important example are symmetric signed Bernoulli random variables, i.e. random variables with distribution

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

Since we have

$$\mathbb{E}|X|^p = 1 \text{ for all } p \geq 1$$

we get by Lemma 1 that X is subgaussian. With the same argument we also obtain that every bounded random variable (i.e. $|X| \leq M$ almost surely for some $M > 0$) is a subgaussian random variable.

Subgaussian random variables have many interesting properties, for example they suffice an approximate rotational invariance, but we do not get more detailed here. The interested reader may find more information about subgaussian random variables in [46] on page 9 ff. and in [26] on page 4 ff.

2.3 Central limit theorem

In stochastics central limit theorems are known as powerful tools. There are certain different versions of central limit theorems known, but we are mainly interested in a quantitative form, namely the Berry-Esséen theorem. The following central limit theorem can be found in [43] on page 31 Proposition D.1 and in [39] on page 71 Theorem 2.1.30.

Theorem 5 (Berry-Esséen Central Limit Theorem) *Let $v = (v_1, \dots, v_n) \in S^{n-1}$. Furthermore let X_1, \dots, X_n be i.i.d. centered random variables with variance one and finite third moments. Then for every $t > 0$:*

$$\left| \mathbb{P}\left(\sum_{k \leq n} v_k X_k \leq t\right) - \mathbb{P}(g \leq t) \right| \leq C \sum_{k \leq n} |v_k|^3,$$

where C is a constant depending on the third moments $\mathbb{E}|X_1|^3 < \infty$ and g is a standard normal distributed random variable. In particular we have

$$\left| \mathbb{P}\left(\sum_{k \leq n} v_k X_k \leq t\right) - \mathbb{P}(g \leq t) \right| \leq C' \max_{1 \leq j \leq n} |v_j|,$$

where again C' is a constant depending on the third moments $\mathbb{E}|X_1|^3 < \infty$.

Chapter 3

A first result and estimates for compressible vectors

Let $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R} \setminus \{0\}$ be real numbers and let X be a centered random variable with unit variance and subgaussian moment bounded by B . In the next chapter we are interested in the smallest singular value of $n \times n$ random tensor matrices $(a_i b_j X_{i,j})_{i,j \leq n}$, where for every $i, j = 1, \dots, n$ $X_{i,j}$ are independent copies of X . More precisely we want to estimate the probability

$$\mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \min \left\{ \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, \frac{\|(a_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{b_i})_{i \leq n}\|_2} \right\}\right)$$

from above. To get a feeling for matrices of that kind we first look at the expected value of the $n \times n$ matrix

$$\begin{pmatrix} g_{1,1} & \cdots & g_{1,n-1} & \varepsilon g_{1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ g_{n,1} & \cdots & g_{n,n-1} & \varepsilon g_{n,n} \end{pmatrix},$$

where for $i, j \in \{1, \dots, n\}$ the $g_{i,j}$ are i.i.d. standard normal distributed random variables. This can be done by a direct computation and the results of A. Edelman.

To prove the main theorem of chapter 4 we need a decomposition of the sphere into two sets: the so called *compressible* and *incompressible* vectors. To get familiar with the set of compressible vectors we show the following estimate in section 3.2:

$$\mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|Ax\|_2 \leq c\sqrt{n}, \|A\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \leq \exp(-c'n),$$

where A is an $n \times n$ matrix whose entries are independent centered random variables with variances at least 1 and bounded fourth moments, $c, c' > 0$ are constants and $K \geq 1$. This estimate is used by M. Rudelson and R. Vershynin in [28] to show that the random matrix A is invertible with high probability. One can learn from the estimate for compressible vectors that in the known situations these vectors do not seem to have a great impact on the least singular value of A .

Definition 3 Let $n, m \in \mathbb{N}$. Let $a_1, \dots, a_n, b_1, \dots, b_m \in \mathbb{R} \setminus \{0\}$ be real numbers. Then we call $(a_i b_j X_{i,j})_{i \leq n, j \leq m}$ a random tensor matrix, where for all $i \leq n$ and $j \leq m$ $X_{i,j}$ is a random variable.

Our next lemma states a useful identity for the operator norm of the inverse operator. Therefore, we generalize the result

$$s_n(A) = \min_{\|x\|_2=1} \|Ax\|_2 = \frac{1}{s_1(A^{-1})} = \frac{1}{\max_{\|x\|_2=1} \|A^{-1}x\|_2}$$

from chapter 2 for the least singular value of an invertible $n \times n$ matrix A .

Lemma 2 *Let X, Y be Banach spaces and $T \in L(Y, X)$ be an invertible linear operator. Then we have the following equality*

$$\|T^{-1}\|_{X \rightarrow Y} = \left(\min_{\|y\|_Y=1} \|Ty\|_X \right)^{-1},$$

where for an operator $S : X \rightarrow Y$ $\|S\|_{X \rightarrow Y} = \max_{\|x\|_X=1} \|Sx\|_Y$ denotes the operator norm.

Proof: Let $T \in L(Y, X)$ be invertible, then we can compute the operator norm of its inverse as follows

$$\begin{aligned} \|T^{-1}\|_{X \rightarrow Y} &= \max_{x \neq 0} \frac{\|T^{-1}x\|_Y}{\|x\|_X} \\ &= \max_{Ty \neq 0} \frac{\|y\|_Y}{\|Ty\|_X} \\ &= \left(\min_{Ty \neq 0} \frac{\|Ty\|_X}{\|y\|_Y} \right)^{-1} \\ &= \left(\min_{y \neq 0} \frac{\|Ty\|_X}{\|y\|_Y} \right)^{-1} \\ &= \left(\min_{\|y\|_Y=1} \|Ty\|_X \right)^{-1}. \end{aligned}$$

□

A main ingredient of our proofs are biorthogonal systems, which will be used to connect the columns of a matrix A with the rows of its inverse A^{-1} .

Definition 4 *Let H be an n -dimensional Hilbert space and let $(X_k)_{k \leq n}$ and $(X_k^*)_{k \leq n}$ be sequences of vectors in H . We call the system $(X_k, X_k^*)_{k \leq n}$ a biorthogonal system in H if for all $j, k \in \{1, \dots, n\}$ the equality*

$$\langle X_j^*, X_k \rangle = \delta_{j,k}$$

holds, where

$$\delta_{j,k} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}, \quad j, k \in \{1, \dots, n\}$$

denotes the Kronecker delta.

If $(X_k, X_k^*)_{k \leq n}$ is a biorthogonal system, then $(X_k)_{k \leq n}$ is a linearly independent system: Assume there exists an $i \in \{1, \dots, n\}$ such that

$$X_i = \sum_{j \neq i} a_j X_j$$

for some a_j not all equal to 0, then

$$1 = \langle X_i^*, X_i \rangle = \sum_{j \neq i} a_j \langle X_i^*, X_j \rangle = 0,$$

which contradicts the properties of a biorthogonal system.

In the sequel we use the following notation:

$$H_k = \text{span}\{X_i \mid i \in \{1, \dots, n\}, i \neq k\}.$$

The importance of biorthogonal systems in the study of the least singular value of random matrices lies in the following lemma which can be found in [27].

Lemma 3 *Biorthogonal systems have the following properties*

1. Let A be an $n \times n$ invertible matrix with columns $X_k = Ae_k$ for $k \in \{1, \dots, n\}$. Define $X_k^* = (A^{-1})^T e_k$ as the k -th row of the inverse matrix A^{-1} , then $(X_k, X_k^*)_{k \leq n}$ is a biorthogonal system in \mathbb{R}^n .
2. Let $(X_k)_{k \leq n}$ be a linearly independent system in an n -dimensional Hilbert space H . Then there exist unique vectors $(X_k^*)_{k \leq n}$ such that $(X_k, X_k^*)_{k \leq n}$ is a biorthogonal system in H .
3. Let $(X_k, X_k^*)_{k \leq n}$ be a biorthogonal system in an n -dimensional Hilbert space H . For all $k \in \{1, \dots, n\}$ we can compute the norm of the k -th row as

$$\|X_k^*\|_2 = \frac{1}{\text{dist}(X_k, H_k)},$$

where $\text{dist}(X_k, H_k)$ denotes the distance from the vector X_k to the subspace H_k .

Proof:

1. From the equality

$$\mathbf{1}_{n \times n} = A^{-1}A = (\langle X_i^*, X_j \rangle)_{i,j \leq n}$$

it follows that

$$\langle X_i^*, X_j \rangle = \delta_{i,j}, \text{ for all } i, j \leq n.$$

Therefore $(X_k, X_k^*)_{k \leq n}$ is a biorthogonal system.

2. Let A' be the matrix with respect to $\{e_1, \dots, e_n\}$ whose k -th column is given by X_k for $k \in \{1, \dots, n\}$. Then A' is invertible and the inverse is unique. From 1. it follows that the rows X_k^* of $(A')^{-1}$ are a sequence of vectors such that $(X_k, X_k^*)_{k \leq n}$ is a biorthogonal system. This sequence of vectors is unique: If $(Y_k)_{k \leq n}$ is a sequence of vectors in H such that for all $j, k \in \{1, \dots, n\}$ the equality $\langle X_k, Y_j \rangle = \delta_{k,j}$ holds, then the matrix which j -th row is Y_j is inverse to A' . Therefore $X_k^* = Y_k$.
3. From $\langle X_k^*, X_j \rangle = \delta_{k,j}$ ($k, j \in \{1, \dots, n\}$) it follows that $X_k^* \in \text{span}\{X_j \mid j \neq k\}^\perp = H_k^\perp$ and therefore, since H_k is a hyperplane, we get that

$$\text{dist}(X_k, H_k) = \inf_{y \in H_k} \|X_k - y\|_2 = \|X_k - P_k X_k\|_2$$

where P_k denotes the orthogonal projection onto H_k . We prove the last equality: We can write every $y \in H_k$ as $y = P_k X_k - z$ for a $z \in H_k$. Therefore, we obtain

$$\begin{aligned} \|X_k - y\|_2^2 &= |\langle X_k - y, X_k - y \rangle| \\ &= |\langle X_k - (P_k X_k - z), X_k - (P_k X_k - z) \rangle| \\ &= |\langle X_k - P_k X_k, X_k - P_k X_k \rangle + 2\langle X_k - P_k X_k, z \rangle + \langle z, z \rangle| \\ &= \|X_k - P_k X_k\|_2^2 + \|z\|_2^2 \\ &\geq \|X_k - P_k X_k\|_2^2 \end{aligned}$$

where the last equality is due to the fact that $X_k - P_k X_k \in H_k^\perp$. We check this: Every $y \in H_k$ can be written as $y = \sum_{j \neq k} \lambda_j X_j$ where $\lambda_j \in \mathbb{R}$ for all $j = 1, \dots, n$ following the

definition of H_k . Therefore, we get that

$$\begin{aligned}
\langle y, X_k - P_k X_k \rangle &= \left\langle \sum_{j \neq k} \lambda_j X_j, X_k - P_k X_k \right\rangle \\
&= \sum_{j \neq k} \lambda_j (\langle X_j, X_k \rangle - \langle X_j, P_k X_k \rangle) \\
&= \sum_{j \neq k} \lambda_j (\langle X_j, X_k \rangle - \langle P_k X_j, X_k \rangle) \\
&= \sum_{j \neq k} \lambda_j (\langle X_j, X_k \rangle - \langle X_j, X_k \rangle) \\
&= 0.
\end{aligned}$$

From the Cauchy-Schwarz inequality it follows that

$$\text{dist}(X_k, H_k) = \left| \left\langle \frac{X_k^*}{\|X_k^*\|_2}, X_k - P_k X_k \right\rangle \right|$$

since $\dim(H_k^\perp) = 1$ and therefore, $\frac{X_k^*}{\|X_k^*\|_2}$ and $X_k - P_k X_k$ have to be linearly dependent. We get that

$$\text{dist}(X_k, H_k) = \left| \left\langle \frac{X_k^*}{\|X_k^*\|_2}, X_k \right\rangle - \left\langle P_k \frac{X_k^*}{\|X_k^*\|_2}, X_k \right\rangle \right| = \left| \left\langle \frac{X_k^*}{\|X_k^*\|_2}, X_k \right\rangle \right| = \frac{1}{\|X_k^*\|_2}$$

and the statement is proved. \square

It is often easier to work with the operator norm of the inverse matrix than doing the computations directly for the smallest singular value. Therefore, we want to use Lemma 3 to compute the inverse of a tensor matrix, which will be helpful many times. This illustrates the usefulness of biorthogonal systems, if one has to deal with inverse matrices.

Lemma 4 *Let $a_1, \dots, a_n, b_1, \dots, b_n \neq 0$ and let $X = (X_{i,j})_{i,j \leq n}$ be an invertible matrix. Then the inverse of the matrix $(a_i b_j X_{i,j})_{i,j \leq n}$ is given by*

$$(a_i b_j X_{i,j})_{i,j \leq n}^{-1} = \left(\frac{1}{b_i} \frac{1}{a_j} X_{i,j}^* \right)_{i,j \leq n},$$

where $(X_i^* = (X_{i,j}^*)_{j \leq n})_{i \leq n}$ is the unique biorthogonal system to the system of column vectors $(X_i = (X_{j,i})_{j \leq n})_{i \leq n}$ of X .

Proof: As a consequence of Lemma 3 we only have to check whether the system

$$\left(\left(a_j b_i X_{j,i} \right)_{j \leq n}, \left(\frac{1}{b_i} \frac{1}{a_j} X_{i,j}^* \right)_{j \leq n} \right)_{i \leq n}$$

is a biorthogonal system. For $k, l \in \{1, \dots, n\}$ we compute

$$\begin{aligned}
\left\langle \left(a_j b_k X_{j,k} \right)_{j \leq n}, \left(\frac{1}{b_l} \frac{1}{a_j} X_{l,j}^* \right)_{j \leq n} \right\rangle &= \sum_{j \leq n} a_j b_k \frac{1}{b_l} \frac{1}{a_j} X_{j,k} X_{l,j}^* = \frac{b_k}{b_l} \sum_{j \leq n} X_{j,k} X_{l,j}^* \\
&= \frac{b_k}{b_l} \langle X_k, X_l^* \rangle = \delta_{k,l}.
\end{aligned}$$

\square

The next lemma is an useful identity linking vectors with independent standard normal distributed entries to vectors that are uniformly distributed on the Euclidean sphere.

Lemma 5 *Let g be an n -dimensional random vector whose entries are independent standard normal distributed random variables. Then $\|g\|_2$ is stochastically independent to $\frac{g}{\|g\|_2}$ and $\frac{g}{\|g\|_2}$ is uniformly distributed on the euclidean sphere.*

Remark: $\mathbb{P}(g = 0) = 0$ and hence $\frac{g}{\|g\|_2}$ is defined almost surely.

Proof: Let $A \subset \mathbb{R}^n$ be an arbitrary Borel set and let $B \subset \mathbb{R}$ be a Borel set such that

$$\mathbb{P}(\|g\|_2 \in B) > 0.$$

Then by the use of polar coordinates we can write the Gaussian measure γ_n of the set A as

$$\begin{aligned} \gamma_n(A) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_A \exp\left(-\frac{\|x\|_2^2}{2}\right) dx \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_0^\infty \int_{A \cap \partial B_2^n(0,r)} \exp\left(-\frac{r^2}{2}\right) d\lambda_{\partial B_2^n(0,r)} dr \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_0^\infty \int_{\frac{1}{r}A \cap \partial B_2^n(0,1)} 1 d\lambda_{\partial B_2^n(0,1)} \exp\left(-\frac{r^2}{2}\right) r^{n-1} dr. \end{aligned}$$

Now, let g be the random vector described above, then we get that

$$\begin{aligned} &\mathbb{P}\left(\left(\frac{g_1}{\|g\|_2}, \dots, \frac{g_n}{\|g\|_2}\right) \in A \mid \|g\|_2 \in B\right) \\ &= \frac{\mathbb{P}\left(\left(\frac{g_1}{\|g\|_2}, \dots, \frac{g_n}{\|g\|_2}\right) \in A, \|g\|_2 \in B\right)}{\mathbb{P}(\|g\|_2 \in B)} \\ &= \frac{\gamma_n\left(\left(\frac{x_1}{\|x\|_2}, \dots, \frac{x_n}{\|x\|_2}\right) \in A, \|x\|_2 \in B\right)}{\gamma_n(\|x\|_2 \in B)} \\ &= \frac{\int_{\{(x_1, \dots, x_n) \in \|x\|_2 A, \|x\|_2 \in B\}} \exp\left(-\frac{\|x\|_2^2}{2}\right) dx}{\int_{\{(x_1, \dots, x_n) \in \mathbb{R}^n: \|x\|_2 \in B\}} \exp\left(-\frac{\|x\|_2^2}{2}\right) dx} \\ &= \frac{\int_0^\infty \int_{A \cap \partial B_2^n(0,1)} 1 d\lambda_{\partial B_2^n(0,1)} \mathbf{1}_{\{r \in B\}} \exp\left(-\frac{r^2}{2}\right) r^{n-1} dr}{\int_0^\infty \int_{\partial B_2^n(0,1)} 1 d\lambda_{\partial B_2^n(0,1)} \mathbf{1}_{\{r \in B\}} \exp\left(-\frac{r^2}{2}\right) r^{n-1} dr} \\ &= \frac{\int_{A \cap \partial B_2^n(0,1)} 1 d\lambda_{\partial B_2^n(0,1)}}{\int_{\partial B_2^n(0,1)} 1 d\lambda_{\partial B_2^n(0,1)}}, \end{aligned}$$

which ends the proof. □

By Lemma 3 one gets the identity $\|X_k^*\|_2 = \frac{1}{\text{dist}(X_k, H_k)}$. This identity is very useful if one deals with Gaussian matrices as the following lemma shows.

Lemma 6 *Let g be a standard normal distributed random variable. For $i \in \{1, \dots, n\}$ let X_i be stochastically independent random vectors, whose entries are independent copies of g . For $k \in \{0, \dots, n-1\}$ denotes $H_1^{n-k} = \text{span}\{X_i \mid i \in \{2, \dots, n-k\}\}$. Then we have the following equality*

$$\mathbb{P}(\text{dist}(X_1, H_1^{n-k}) < \varepsilon) = \mathbb{P}(\|(g_i)_{i \leq k+1}\|_2 < \varepsilon) \text{ for all } \varepsilon > 0,$$

where the g_i for $i \leq k+1$ are also independent copies of g .

Proof: We want to use the fact that the vectors X_1, \dots, X_n are linearly independent almost surely. We prove this: Since the length of each vector does not affect the linear dependency it suffices to

prove that $\frac{X_1}{\|X_1\|_2}, \dots, \frac{X_n}{\|X_n\|_2}$ are linearly independent almost surely. As we know from Lemma

5 the stochastically independent vectors $\frac{X_1}{\|X_1\|_2}, \dots, \frac{X_n}{\|X_n\|_2}$ are defined almost surely and are uniformly distributed on the sphere. Of course, $\left\{ \frac{X_1}{\|X_1\|_2} \right\}$ is a linear independent set of vectors.

Assume that for $1 < m \leq n$ the vectors $\frac{X_1}{\|X_1\|_2}, \dots, \frac{X_{m-1}}{\|X_{m-1}\|_2}$ are linearly independent. Therefore, they form a subspace of dimension $m - 1$ which intersects the Euclidean sphere. This intersection is an $m - 2$ -dimensional subsphere which we denote by S . Since $\sigma_{n-1}(S) = 0$, where σ_{n-1} is the unique normalized rotational invariant measure on S^{n-1} we have that the event $\{X_m \in S\}$ has probability zero. Therefore, we obtain that X_1, \dots, X_m are almost surely linearly independent for all $1 < m \leq n$.

Since the vectors X_1, \dots, X_n are linearly independent almost surely, we can assume w.l.o.g. that $\dim(H_1^{n-k}) = n - k - 1$. Let Q be an orthonormal matrix such that $QH_1^{n-k} = \mathbb{R}^{n-k-1}$. We denote by P_1^{n-k} the orthogonal projection onto H_1^{n-k} . Then we get

$$\begin{aligned} \text{dist}(X_1, H_1^{n-k}) &= \|X_1 - P_1^{n-k} X_1\|_2 \\ &= \|Q(X_1 - P_1^{n-k} X_1)\|_2 \\ &= \|QX_1 - QP_1^{n-k} X_1\|_2 \\ &= \|(0, \dots, 0, QX_1(n-k-1), \dots, QX_1(n))^T\|_2, \end{aligned}$$

where we used that Q is an isometry. Since QX_1 equals X_1 in distribution we get that

$$QX_1(i) \stackrel{\mathcal{D}}{=} g \text{ for all } i \leq n$$

and it follows that

$$\text{dist}(X_1, H_1^{n-k}) \stackrel{\mathcal{D}}{=} \left(\sum_{j \leq k+1} |g_j|^2 \right)^{\frac{1}{2}},$$

where the equality is an equality in distribution and the g_j 's are independent copies of g . \square

In particular since $H_1 = H_1^n$ Lemma 6 says that for Gaussian vectors $(X_i)_{i \leq n}$ we have following equality for all $\varepsilon > 0$

$$\mathbb{P}(\text{dist}(X_1, H_1) < \varepsilon) = \mathbb{P}(|g| < \varepsilon).$$

So we have that $\text{dist}(X_1, H_1)$ is equal in distribution to $|g|$, where g is a standard normal distributed random variable.

In [43] T. Tao and V. Vu proved that for n -dimensional random vectors X_i ($i = 1, \dots, n$) whose coordinates are independent copies of a normalized random variable X (i.e. $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$) one has a similar statement. In this case $\text{dist}(X_k, H_k)$ is approximately Gaussian distributed: For all $t > 0$ we have

$$|\mathbb{P}(\text{dist}(X_k, H_k) \leq t) - \mathbb{P}(|g| \leq t)| \leq Cn^{-\frac{c}{4}}$$

for suitable constants $C, c > 0$. This is due to a Berry-Esséen argument.

3.1 Expected value of the smallest singular value of a Gaussian square random matrix with disturbed last column

As mentioned in the introduction A. Edelman used the rotational invariance of the Gaussian distribution in his paper [7] to compute the probability density function of the smallest singular

value of particular random matrices. The following two theorems can be found in [8] on page 41 and page 43.

Theorem 6 (Edelman) *The probability density function of the square of the smallest singular value of an $(n \times (n + 1))$ random matrix A , whose entries are i.i.d. standard normal distributed, is given by*

$$f_{s_n^2(A)} : \mathbb{R} \rightarrow \mathbb{R}, \quad f_{s_n^2(A)}(x) = \begin{cases} \frac{n}{2} \exp\left(-\frac{xn}{2}\right) & x \in [0, \infty) \\ 0 & \text{else} \end{cases}.$$

Theorem 7 *The probability density function of the square of the smallest singular value of an $(n \times n)$ random matrix A , whose entries are i.i.d. standard normal distributed, is given by*

$$f_{s_n^2(A)} : \mathbb{R} \rightarrow \mathbb{R}, \quad f_{s_n^2(A)}(x) = \begin{cases} \frac{n}{2^{n-\frac{1}{2}}} \frac{\Gamma(n)}{\Gamma(\frac{n}{2})} x^{-\frac{1}{2}} \exp\left(-x\frac{n}{2}\right) U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{x}{2}\right) & x \in [0, \infty) \\ 0 & \text{else} \end{cases},$$

where, for $a > 0$ and $b < 1$, the Tricomi function, $U(a, b, z)$ is the unique solution to Kummer's equation

$$z \frac{d^2 w}{dz^2} + (b - z) \frac{dw}{dz} - aw = 0$$

satisfying $U(a, b, 0) = \frac{\Gamma(1-b)}{\Gamma(1+a-b)}$ and $U(a, b, \infty) = 0$.

In [32] it was stated that the theorem above has the following consequence.

Theorem 8 (Edelman) *Let A be an $n \times n$ random matrix, whose entries are i.i.d. standard normal distributed random variables, then for any $\varepsilon \geq 0$*

$$\mathbb{P}(s_n(A) \leq \varepsilon) \leq \sqrt{n}\varepsilon.$$

For any $\varepsilon > 0$ we are now in the position to analyze the least singular value of the matrix

$$G_\varepsilon = \begin{pmatrix} g_{1,1} & \cdots & g_{1,n-1} & \varepsilon g_{1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ g_{n,1} & \cdots & g_{n,n-1} & \varepsilon g_{n,n} \end{pmatrix},$$

where for $i, j \in \{1, \dots, n\}$ the $g_{i,j}$ are i.i.d. standard normal distributed random variables.

Theorem 9 *For all $n \geq 2$ we have*

$$\left(\frac{1}{2}\right)^3 \min\left\{\sqrt{\frac{\pi}{2}}\varepsilon, \frac{1}{\sqrt{n}}\right\} \leq \mathbb{E}s_n(G_\varepsilon) \leq \sqrt{2} \min\left\{\frac{\varepsilon}{\sqrt{\pi}}, \frac{1}{\sqrt{n-1}}\right\}$$

Proof: From the singular value decomposition it follows that

$$\det((g_{i,j})_{i,j \leq n}) = \det(\text{diag}(s_1((g_{i,j})_{i,j \leq n}), \dots, s_n((g_{i,j})_{i,j \leq n}))) = \prod_{i \leq n} s_i((g_{i,j})_{i,j \leq n}).$$

Therefore, the determinant of $(g_{i,j})_{i,j \leq n}$ is zero, if and only if $s_n((g_{i,j})_{i,j \leq n})$ equals zero which means that $(g_{i,j})_{i,j \leq n}$ is not invertible if and only if $s_n((g_{i,j})_{i,j \leq n}) = 0$. By Theorem 8 it follows that $(g_{i,j})_{i,j \leq n}$ is almost surely invertible and therefore by Lemma 4 G_ε is almost surely invertible and we can estimate the operator norm of its inverse by

$$\|G_\varepsilon^{-1}\|_{2 \rightarrow 2} = \|(G_\varepsilon^{-1})^T\|_{2 \rightarrow 2} = \max_{\|x\|_2=1} \|(G_\varepsilon^{-1})^T x\|_2 \geq \|(G_\varepsilon^{-1})^T e_n\|_2.$$

Again we use Lemma 4 to obtain

$$\|G_\varepsilon^{-1}\|_{2 \rightarrow 2} \geq \frac{1}{\varepsilon} \|X_n^*\|_2,$$

where X_n^* denotes the n -th row of $(g_{i,j})_{i,j \leq n}^{-1}$. By Lemma 2 and Lemma 3 it follows that

$$s_n(G_\varepsilon) = \min_{\|x\|_2=1} \|G_\varepsilon x\|_2 = \frac{1}{\|G_\varepsilon^{-1}\|_{2 \rightarrow 2}} \leq \frac{\varepsilon}{\|X_n^*\|_2} = \varepsilon \text{dist}(X_n, H_n).$$

On the other hand we can estimate $s_n(G_\varepsilon)$ by

$$s_n(G_\varepsilon) = s_n(G_\varepsilon^T) = \min_{\|x\|_2=1} \|G_\varepsilon^T x\|_2 \leq \min\{\|G_\varepsilon^T x\|_2 \mid x = (x_1, \dots, x_n) \in S^{n-1}, x_n = 0\} = s_{n-1}(C^T),$$

where

$$C^T = \begin{pmatrix} g_{1,1} & \cdots & g_{1,n-1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ g_{n,1} & \cdots & g_{n,n-1} \end{pmatrix}.$$

We can now use Theorem 6 to estimate the expected value of $s_{n-1}(C^T)$

$$\begin{aligned} \mathbb{E}s_n(C^T) &\stackrel{\text{Jensen}}{\leq} (\mathbb{E}s_n^2(C^T))^{\frac{1}{2}} = \left(\int_0^\infty x \frac{n-1}{2} \exp\left(-\frac{x(n-1)}{2}\right) dx \right)^{\frac{1}{2}} \\ &= \left(-x \exp\left(-\frac{x(n-1)}{2}\right) \Big|_0^\infty + \int_0^\infty \exp\left(-\frac{x(n-1)}{2}\right) dx \right)^{\frac{1}{2}} \\ &= \left(\int_0^\infty \exp\left(-\frac{x(n-1)}{2}\right) dx \right)^{\frac{1}{2}} = \left(\frac{2}{n-1} \right)^{\frac{1}{2}}, \end{aligned}$$

where we used partial integration. It follows that

$$\mathbb{E}s_n(G_\varepsilon) \leq \mathbb{E} \min\{\varepsilon \text{dist}(X_n, H_n), s_{n-1}(C^T)\} \leq \min\{\varepsilon \mathbb{E} \text{dist}(X_n, H_n), \mathbb{E}s_{n-1}(C^T)\}.$$

We use Lemma 6 to obtain that $\text{dist}(X_n, H_n) \stackrel{\mathcal{D}}{=} |g|$, where $|g|$ is a standard normal distributed random variable. It follows that

$$\mathbb{E}s_n(G_\varepsilon) \leq \min \left\{ \varepsilon \mathbb{E}|g|, \sqrt{\frac{2}{n-1}} \right\} = \sqrt{2} \min \left\{ \frac{\varepsilon}{\sqrt{\pi}}, \frac{1}{\sqrt{n-1}} \right\}.$$

Now we want to estimate the expected value of $s_n(G_\varepsilon)$ from below. In the sequel we will denote by X_k^* the k -th row of $(g_{i,j})_{i,j \leq n}^{-1}$ and write $(g_{i,j})_{i,j \leq n}^{-1} = (X_{i,j}^*)_{i,j \leq n}$ following the notation of Lemma

3. Then by Lemma 4

$$\begin{aligned}
\mathbb{E}s_n(G_\varepsilon) &= \mathbb{E} \frac{1}{\|(G_\varepsilon^{-1})^T\|_{2 \rightarrow 2}} = \mathbb{E} \frac{1}{\left\| \begin{pmatrix} X_{1,1}^* & \cdots & X_{n-1,1}^* & \frac{1}{\varepsilon} X_{n,1}^* \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ X_{1,n}^* & \cdots & X_{n-1,n}^* & \frac{1}{\varepsilon} X_{n,n}^* \end{pmatrix} \right\|_{2 \rightarrow 2}} \\
&\geq \mathbb{E} \frac{1}{\left\| \begin{pmatrix} X_{1,1}^* & \cdots & X_{n-1,1}^* & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ X_{1,n}^* & \cdots & X_{n-1,n}^* & 0 \end{pmatrix} \right\|_{2 \rightarrow 2} + \left\| \begin{pmatrix} 0 & \cdots & 0 & \frac{1}{\varepsilon} X_{n,1}^* \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \frac{1}{\varepsilon} X_{n,n}^* \end{pmatrix} \right\|_{2 \rightarrow 2}} \\
&= \mathbb{E} \frac{1}{\max_{\|x\|_2=1} \left\| \left(\sum_{i \leq n} X_{1,i}^* x_i, \dots, \sum_{i \leq n} X_{n-1,i}^* x_i, 0 \right) \right\|_2 + \frac{1}{\varepsilon} \|X_n^*\|_2} \\
&\geq \mathbb{E} \frac{1}{\max_{\|x\|_2=1} \left\| \left(\sum_{i \leq n} X_{1,i}^* x_i, \dots, \sum_{i \leq n} X_{n-1,i}^* x_i, \sum_{i \leq n} X_{n,i}^* x_i \right) \right\|_2 + \frac{1}{\varepsilon} \|X_n^*\|_2} \\
&\geq \mathbb{E} \frac{1}{2 \max \left\{ \|(g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2}, \frac{1}{\varepsilon} \|X_n^*\|_2 \right\}} \\
&= \frac{1}{2} \mathbb{E} \min \{s_n((g_{i,j})_{i,j \leq n}), \varepsilon \text{dist}(X_n, H_n)\},
\end{aligned}$$

where we used Lemma 3 in the last step. To analyze the last term any further we first need an estimate of the distribution of the random variable $\min\{s_n((g_{i,j})_{i,j \leq n}), \varepsilon \text{dist}(X_n, H_n)\}$.

It follows from Lemma 6 that

$$\mathbb{P}(\varepsilon \text{dist}(X_n, H_n) \leq t) = \mathbb{P}\left(|g| \leq \frac{t}{\varepsilon}\right) = \sqrt{\frac{2}{\pi}} \int_0^{\frac{t}{\varepsilon}} \underbrace{\exp\left(-\frac{x^2}{2}\right)}_{\leq 1} dx \leq \sqrt{\frac{2}{\pi}} \frac{t}{\varepsilon},$$

where g is a standard normal distributed random variable. By Theorem 8 we conclude that

$$\begin{aligned}
\mathbb{P}(\min\{s_n((g_{i,j})_{i,j \leq n}), \varepsilon|g|\} \leq t) &= 1 - \mathbb{P}(\min\{s_n((g_{i,j})_{i,j \leq n}), \varepsilon|g|\} > t) \\
&= 1 - \mathbb{P}(\{s_n((g_{i,j})_{i,j \leq n}) > t\} \cap \{\varepsilon|g| > t\}) \\
&= 1 - \left(1 - \mathbb{P}\left(\{s_n((g_{i,j})_{i,j \leq n}) > t\} \cap \{\varepsilon|g| > t\}^c\right)\right) \\
&= 1 - \left(1 - \mathbb{P}\left(\{s_n((g_{i,j})_{i,j \leq n}) \leq t\} \cup \{|g| \leq \frac{t}{\varepsilon}\}\right)\right) \\
&\leq \mathbb{P}\left(s_n((g_{i,j})_{i,j \leq n}) \leq t\right) + \mathbb{P}\left(|g| \leq \frac{t}{\varepsilon}\right) \\
&\leq \left(\sqrt{n} + \sqrt{\frac{2}{\pi}} \frac{1}{\varepsilon}\right)t,
\end{aligned}$$

where we used the union bound in the first inequality. By the use of the formula

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t) dt$$

for a non-negative real valued random variable it finally follows that

$$\begin{aligned}
\mathbb{E}s_n(G) &\geq \frac{1}{2} \int_0^\infty \mathbb{P}(\min\{s_n((g_{i,j})_{i,j \leq n}), \varepsilon|g|\} > t) dt \\
&= \frac{1}{2} \int_0^\infty (1 - \mathbb{P}(\min\{s_n((g_{i,j})_{i,j \leq n}), \varepsilon|g|\} \leq t)) dt \\
&\geq \frac{1}{2} \int_0^\infty \left(1 - \left(\sqrt{n} + \sqrt{\frac{2}{\pi} \frac{1}{\varepsilon}} t\right)_+ \right) dt,
\end{aligned}$$

where $(x)_+ = \max\{x, 0\}$. It follows that

$$\begin{aligned}
\mathbb{E}s_n(G) &\geq \frac{1}{2} \int_0^{(\sqrt{\frac{2}{\pi} \frac{1}{\varepsilon}} + \sqrt{n})^{-1}} \left(1 - \left(\sqrt{n} + \sqrt{\frac{2}{\pi} \frac{1}{\varepsilon}} t\right) \right) dt \\
&= \frac{1}{2} \left[t - \left(\sqrt{n} + \sqrt{\frac{2}{\pi} \frac{1}{\varepsilon}} \frac{t^2}{2}\right) \right]_0^{(\sqrt{\frac{2}{\pi} \frac{1}{\varepsilon}} + \sqrt{n})^{-1}} \\
&= \left(\frac{1}{2}\right)^2 \frac{1}{\sqrt{\frac{2}{\pi} \frac{1}{\varepsilon}} + \sqrt{n}} \\
&= \left(\frac{1}{2}\right)^2 \frac{1}{2 \max\{\sqrt{\frac{2}{\pi} \frac{1}{\varepsilon}}, \sqrt{n}\}} \\
&= \left(\frac{1}{2}\right)^3 \min\left\{\sqrt{\frac{\pi}{2}} \varepsilon, \frac{1}{\sqrt{n}}\right\},
\end{aligned}$$

which ends the proof. \square

As mentioned in the introduction the estimates in Theorem 9 are surprising as these bounds are way better than taking the minimum of the expectation of the last (small) column

$$\sqrt{\frac{2}{\pi}} \varepsilon \sqrt{n} \leq \mathbb{E}\|(\varepsilon g_{i,n})_{i \leq n}\|_2 \leq \varepsilon \sqrt{n},$$

and the least singular value of

$$\begin{pmatrix} g_{1,1} & \cdots & g_{1,n-1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ g_{n,1} & \cdots & g_{n,n-1} \end{pmatrix},$$

which at first sight may be a reasonable approach for an upper bound. We give a short proof for the bounds of the expectation of the last column. The upper bound is a consequence of Jensen's inequality:

$$\mathbb{E}\|(\varepsilon g_{i,n})_{i \leq n}\|_2 \leq \mathbb{E}\left(\sum_{i=1}^n |\varepsilon|^2 |g_{i,n}|^2\right)^{\frac{1}{2}} \leq \varepsilon \left(\sum_{i=1}^n \mathbb{E}|g_{i,n}|^2\right)^{\frac{1}{2}} = \varepsilon \sqrt{n},$$

where we used that the function $x \mapsto \sqrt{x}$ is a concave function. The lower bound is a consequence of the triangle inequality.

$$\mathbb{E}\|(\varepsilon g_{i,n})_{i \leq n}\|_2 = \varepsilon \mathbb{E}\|(|g_{i,n}|)_{i \leq n}\|_2 \geq \varepsilon \|(\mathbb{E}|g_{i,n}|)_{i \leq n}\|_2 = \sqrt{\frac{2}{\pi}} \varepsilon \sqrt{n}.$$

With almost analogue considerations as for Theorem 9 we can bound the expected value of the least singular value of

$$G'_\varepsilon = \begin{pmatrix} \varepsilon g_{1,1} & \cdots & \varepsilon g_{1,n-1} & g_{1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \varepsilon g_{n,1} & \cdots & \varepsilon g_{n,n-1} & g_{n,n} \end{pmatrix},$$

where the $g_{i,j}$ ($i, j \in \{1, \dots, n\}$) are i.i.d. standard normal distributed random variables.

Corollary 1 *For all $n \geq 2$ we have*

$$\left(\frac{1}{2}\right)^3 \frac{\varepsilon}{\sqrt{n}} \leq \mathbb{E}s_n(G'_\varepsilon) \leq \varepsilon \sqrt{\frac{2}{n-1}}.$$

3.2 Estimate for compressible vectors

In order to prove our main results we use a technique called decomposition of the sphere, which is due to Mark Rudelson and Roman Vershynin (see [28, page 14 ff.]).

Definition 5 *Let $\delta, \rho \in (0, 1)$. A vector $x \in \mathbb{R}^n$ is called sparse if $x \in \text{Sparse}(\delta) = \{x \in \mathbb{R}^n \mid |\text{supp}(x)| \leq \delta n\}$. A vector $x \in S^{n-1}$ is called compressible if $x \in \text{Comp}(\delta, \rho) = \{x \in S^{n-1} \mid d(x, \text{Sparse}(\delta)) \leq \rho\}$, where d denotes the Euclidean distance. Respectively we call a vector x incompressible if $x \in \text{Incomp}(\delta, \rho) = S^{n-1} \setminus \text{Comp}(\delta, \rho)$.*

In the sequel the parameters δ, ρ can be seen as small constants. We learned from M. Rudelson and R. Vershynin (see [28] on page 14 and [45] lecture 16) that the terminology of compressible and incompressible vectors is borrowed from signal processing and sparse approximation theory. In signal processing one studies signals $x \in \mathbb{R}^n$. One can think of x as the vector of the Fourier or wavelet coefficients of the signal. Let $\bar{x} \in \mathbb{R}^n$ denote the vector that consists of the coordinates of x arranged in non increasing order, i.e. $\bar{x}_1 \geq \bar{x}_2 \geq \dots \bar{x}_n$. We say that the signal x is compressible if and only if it has coefficients that decay fast in the sense that for every $k = 1, \dots, n$ the k -th coordinate can be estimated by $|\bar{x}_k| \leq Ck^{-\frac{1}{p}}$ where $C > 0$ is an absolute constant and $p \in (0, 1)$. Therefore, most of the information carried by the signal x lies in just δn coefficients for some small constant $\delta \in (0, 1)$. So x can be approximated by a vector who is carried in only δn coordinates and therefore is compressible in the sense of the definition above.

In this section we want to study the impact of compressible vectors on the order of the least singular value. It turns out that studying the term

$$\inf_{x \in \text{Comp}(\delta, \rho)} \|Ax\|_2$$

for an $n \times n$ random matrix A is strongly connected to the study of the least singular value of rectangular random matrices. This is due to an ε -net argument that we will explain in the sequel. The main lemma of this section and its proof can be found in [28, page 15]. We present it here in detail for the seek completeness and to get a better understanding of the problems that come up when we are dealing with tensor matrices.

Lemma 7 *Let $n \in \mathbb{N}$ be a natural number. Let A be an $n \times n$ matrix whose entries are independent centered random variables with variances at least 1 and fourth moments bounded by B , and let $K \geq 1$. Then there exist $\delta, \rho, c, c' > 0$ only depending on B and K and such that*

$$\mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|Ax\|_2 \leq c\sqrt{n}, \|A\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \leq \exp(-c'n).$$

We want to point out that the estimate we get in Lemma 7 is much stronger than we actually need. We explain this in the setting of an $n \times n$ random matrix whose entries are independent copies of a centered subgaussian random variable ξ with variance one and subgaussian moment bounded by B' . By Lemma 1 one easily sees that a subgaussian random variable has a bounded fourth moment (and thus one can get an equivalent formulation of Lemma 7, where $\delta, \rho, c, c' > 0$ are only depending on the subgaussian moments). As already explained in the introduction one gets for every $\varepsilon > 0$ that

$$\mathbb{P}(s_n(A) \leq \varepsilon n^{-\frac{1}{2}}) \leq C\varepsilon + c^n,$$

where $C > 0$ and $c \in (0, 1)$ depend on B' . If one chooses ε such that one obtains

$$\mathbb{P}(s_n(A) \geq \varepsilon n^{-\frac{1}{2}}) \geq \frac{1}{2}$$

it follows that

$$\mathbb{M}s_n(A) \geq \varepsilon n^{-\frac{1}{2}},$$

where $\mathbb{M}s_n(A)$ denotes the median of the random variable $s_n(A)$. But Lemma 7 ensures us the existence of $\delta, \rho, c, c' > 0$ only depending on B such that

$$\mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|Ax\|_2 \leq c\sqrt{n}\right) \leq \exp(-c'n).$$

Here we used the following lemma which can be found in [28] on page 10:

Lemma 8 *Let A be an $n \times n$ random matrix whose entries are independent centered random variables with variances at least one and subgaussian moments bounded by $B > 0$. Then*

$$\mathbb{P}(\|A\|_{2 \rightarrow 2} > K\sqrt{n}) \leq 2 \exp(-n),$$

where K depends only on B .

This enables us to remove the condition on the operator norm of the random matrix A from the estimate. It follows that

$$\mathbb{M}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|Ax\|_2\right) \geq cn^{\frac{1}{2}}.$$

That shows that compressible vectors do not have a great impact on the order of the least singular value in this setting.

Before we can give a proof of Lemma 7 we need to prove a statement for rectangular random matrices.

Lemma 9 *Let G be an $n \times k$ random matrix whose entries are independent centered random variables with variances at least 1 and fourth moments bounded by B . Let $K \geq 1$. Then there exist $c, d > 0$ and $\delta_0 \in (0, 1)$ which depends only on B and K and so that $k < \delta_0 n$. Then we have*

$$\mathbb{P}\left(\inf_{x \in S^{k-1}} \|Gx\|_2 \leq c\sqrt{n}, \|G\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \leq \exp(-dn).$$

For the proof of Lemma 9 we need some tools. We begin with the following general form of the Payley-Zygmund inequality.

Lemma 10 *Let $p \in (1, \infty)$ and let $q = \frac{p}{p-1}$ the conjugate index. Let $X \geq 0$ be a random variable with $\mathbb{E}X^{2p} < \infty$. Then we have*

$$\mathbb{P}(X > \lambda) \geq \frac{(\mathbb{E}X^2 - \lambda^2)^q}{(\mathbb{E}X^{2p})^{\frac{q}{p}}},$$

for every $0 \leq \lambda \leq \sqrt{\mathbb{E}X^2}$.

Proof: We get by Hölder's inequality

$$\begin{aligned} \mathbb{E}X^2 &= \mathbb{E}X^2 \mathbf{1}_{\{X > \lambda\}} + \mathbb{E}X^2 \mathbf{1}_{\{X \leq \lambda\}} \\ &\leq (\mathbb{E}X^{2p})^{\frac{1}{p}} (\mathbb{E} \mathbf{1}_{\{X > \lambda\}})^{\frac{1}{q}} + \mathbb{E}X^2 \mathbf{1}_{\{X \leq \lambda\}} \\ &\leq (\mathbb{E}X^{2p})^{\frac{1}{p}} (\mathbb{E} \mathbf{1}_{\{X > \lambda\}})^{\frac{1}{q}} + \lambda^2 \\ &= (\mathbb{E}X^{2p})^{\frac{1}{p}} \mathbb{P}(X > \lambda)^{\frac{1}{q}} + \lambda^2. \end{aligned}$$

Therefore we get that

$$\mathbb{P}(X > \lambda) \geq \frac{(\mathbb{E}X^2 - \lambda^2)^q}{(\mathbb{E}X^{2p})^{\frac{q}{p}}}.$$

□

The following inequality is known as standard-symmetrization inequality and can be found in [18] as Lemma 6.3.

Lemma 11 *Let $F : [0, \infty) \rightarrow [0, \infty)$ be a convex, non-decreasing function. Furthermore for any $m \in \mathbb{N}$ let $(X_i)_{i \leq m}$ be a sequence of independent centered random variables in a Banach space $(B, \|\cdot\|)$ with $\mathbb{E}F(\|X_i\|) < \infty$ for every $i \leq m$, then we have*

$$\mathbb{E}F\left(\frac{1}{2}\left\|\sum_{i \leq m} \varepsilon_i X_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_{i \leq m} X_i\right\|\right) \leq \mathbb{E}F\left(2\left\|\sum_{i \leq m} \varepsilon_i X_i\right\|\right),$$

where $(\varepsilon_i)_{i \leq m}$ is a sequence of independent $\{\pm 1\}$ -valued $\text{Ber}(\frac{1}{2})$ variables that are independent of $(X_i)_{i \leq m}$.

Proof: We denote with X'_i an independent copy of the random variable X_i and define $\overline{X}_i := X_i - X'_i$. Then by construction \overline{X}_i is a symmetric random variable, i.e. \overline{X}_i has the same distribution as $-\overline{X}_i$. Let $(\varepsilon_i)_{i \leq m}$ be a sequence of independent $\{\pm 1\}$ -valued $\text{Ber}(\frac{1}{2})$ variables that are independent of $(X_i)_{i \leq m}$ and $(X'_i)_{i \leq m}$. Then we have for the conditional expectation that

$$\begin{aligned} \mathbb{E}\left(F\left(\left\|\sum_{i \leq m} \overline{X}_i\right\|\right) \mid \sum_{i \leq m} X_i\right) &\stackrel{\text{Jensen}}{\geq} F\left(\mathbb{E}\left(\left\|\sum_{i \leq m} \overline{X}_i\right\| \mid \sum_{i \leq m} X_i\right)\right) \\ &\stackrel{\Delta\text{-inequality}}{\geq} F\left(\left\|\mathbb{E}\left(\sum_{i \leq m} \overline{X}_i \mid \sum_{i \leq m} X_i\right)\right\|\right) \\ &= F\left(\left\|\mathbb{E}\left(\sum_{i \leq m} X_i - X'_i \mid \sum_{i \leq m} X_i\right)\right\|\right) \\ &= F\left(\left\|\mathbb{E}\left(\sum_{i \leq m} X_i \mid \sum_{i \leq m} X_i\right) - \mathbb{E}\left(\sum_{i \leq m} X'_i \mid \sum_{i \leq m} X_i\right)\right\|\right) \\ &= F\left(\left\|\sum_{i \leq m} X_i - \mathbb{E}\left(\sum_{i \leq m} X'_i\right)\right\|\right) \\ &= F\left(\left\|\sum_{i \leq m} X_i\right\|\right). \end{aligned}$$

Therefore by the law of total expectation we get that

$$\mathbb{E}F\left(\left\|\sum_{i \leq m} \overline{X}_i\right\|\right) = \mathbb{E}\left(\mathbb{E}\left(F\left(\left\|\sum_{i \leq m} \overline{X}_i\right\|\right) \mid \sum_{i \leq m} X_i\right)\right) \geq \mathbb{E}F\left(\left\|\sum_{i \leq m} X_i\right\|\right).$$

Since the \overline{X}_i are symmetric and F is non-decreasing we get that

$$\mathbb{E}F\left(\left\|\sum_{i \leq m} X_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_{i \leq m} \varepsilon_i \overline{X}_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_{i \leq m} \varepsilon_i X_i\right\| + \left\|\sum_{i \leq m} \varepsilon_i X'_i\right\|\right)$$

and finally since F is a convex function

$$\mathbb{E}F\left(\left\|\sum_{i \leq m} X_i\right\|\right) \leq \frac{1}{2}\mathbb{E}F\left(2\left\|\sum_{i \leq m} \varepsilon_i X_i\right\|\right) + \frac{1}{2}\mathbb{E}F\left(2\left\|\sum_{i \leq m} \varepsilon_i X'_i\right\|\right) \leq \mathbb{E}F\left(2\left\|\sum_{i \leq m} \varepsilon_i X_i\right\|\right).$$

Analogue it follows that

$$\mathbb{E}F\left(\frac{1}{2}\left\|\sum_{i \leq m} \varepsilon_i X_i\right\|\right) \leq \mathbb{E}F\left(\frac{1}{2}\left\|\sum_{i \leq m} \varepsilon_i \bar{X}_i\right\|\right) = \mathbb{E}F\left(\frac{1}{2}\left\|\sum_{i \leq m} \bar{X}_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_{i \leq m} X_i\right\|\right).$$

□

The next lemma transfers a one dimensional probability estimate for a general class of random variables to the n -dimensional case.

Lemma 12 *Let $n \in \mathbb{N}$ be a natural number. Let X_1, \dots, X_n be independent, non-negative random variables with variances at least 1. Consider there exist $\lambda > 0$ and $\mu \in (0, 1)$, with*

$$\mathbb{P}(X_k < \lambda) \leq \mu \text{ for all } k \in \{1, \dots, n\}.$$

Then there exist $\lambda' > 0$ and $\mu' \in (0, 1)$, depending only on λ and μ , with

$$\mathbb{P}\left(\sum_{k \leq n} X_k^2 < \lambda' n\right) \leq (\mu')^n.$$

Proof: Let $\lambda > 0$ and $\mu \in (0, 1)$ such that the assumptions of the lemma are satisfied. We use Markov's inequality:

$$\begin{aligned} \mathbb{P}\left(\sum_{k \leq n} X_k^2 < \lambda' n\right) &= \mathbb{P}\left(\lambda' n - \sum_{k \leq n} X_k^2 > 0\right) \\ &= \mathbb{P}\left(\exp\left(\lambda' n - \sum_{k \leq n} X_k^2\right) > 1\right) \\ &\leq \mathbb{E} \exp\left(\lambda' n - \sum_{k \leq n} X_k^2\right). \end{aligned}$$

Now we use the independence and then go on by estimating $\mathbb{E} \exp(-X_k^2)$. Again, we use the formula

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t) dt$$

for real valued and non negative random variables X and obtain

$$\begin{aligned} \mathbb{P}\left(\sum_{k \leq n} X_k^2 < \lambda' n\right) &\leq \exp(\lambda' n) \prod_{k=1}^n \mathbb{E} \exp(-X_k^2) \\ &= \exp(\lambda' n) \prod_{k=1}^n \int_0^\infty \mathbb{P}(\exp(-X_k^2) > t) dt \\ &= \exp(\lambda' n) \prod_{k=1}^n \int_0^\infty \mathbb{P}\left(\exp(X_k^2) < \frac{1}{t}\right) dt. \end{aligned}$$

Since $\exp(X_k^2) \geq 1$ we have

$$\mathbb{P}\left(\exp(X_k^2) < \frac{1}{t}\right) = 0$$

for all $t > 1$. It follows that

$$\begin{aligned} \mathbb{P}\left(\sum_{k \leq n} X_k^2 < \lambda' n\right) &= \exp(\lambda' n) \prod_{k=1}^n \int_0^1 \mathbb{P}\left(\exp(X_k^2) < \frac{1}{t}\right) dt \\ &= \exp(\lambda' n) \prod_{k=1}^n \int_0^1 \mathbb{P}(X_k^2 < -\ln(t)) dt. \end{aligned}$$

Now, we make the following change of variables:

$$t = \exp(-s^2\lambda^2) \text{ and } dt = -2s\lambda^2 \exp(-s^2\lambda^2)ds$$

and get that

$$\begin{aligned} & \mathbb{P}\left(\sum_{k \leq n} X_k^2 < \lambda' n\right) \\ &= \exp(\lambda' n) \prod_{k=1}^n \int_0^\infty 2s\lambda^2 \exp(-s^2\lambda^2) \mathbb{P}(X_k < s\lambda) ds \\ &= \exp(\lambda' n) \prod_{k=1}^n \left\{ \int_0^1 2s\lambda^2 \exp(-s^2\lambda^2) \mathbb{P}(X_k < s\lambda) ds + \int_1^\infty 2s\lambda^2 \exp(-s^2\lambda^2) \mathbb{P}(X_k < s\lambda) ds \right\} \\ &\leq \exp(\lambda' n) \prod_{k=1}^n \left\{ \int_0^1 2s\lambda^2 \exp(-s^2\lambda^2) \mathbb{P}(X_k < \lambda) ds + \int_1^\infty 2s\lambda^2 \exp(-s^2\lambda^2) \mathbb{P}(X_k < s\lambda) ds \right\} \\ &\leq \exp(\lambda' n) \prod_{k=1}^n \underbrace{\left\{ \mu \int_0^1 2s\lambda^2 \exp(-s^2\lambda^2) ds + \int_1^\infty 2s\lambda^2 \exp(-s^2\lambda^2) \mathbb{P}(X_k < s\lambda) ds \right\}}_{=q(\lambda, \mu)} \\ &= \exp(\lambda' n) (q(\lambda, \mu))^n. \end{aligned}$$

Since

$$\int_0^\infty 2s\lambda^2 \exp(-s^2\lambda^2) ds = 1$$

it follows that $q(\lambda, \mu) \in (0, 1)$ and the right choice of λ' proves the lemma. \square

Lemma 13 *Let $X = (X_1, \dots, X_n)$ be a sequence of independent, centered random variables with variances at least 1 and fourth moments bounded by B . We define for an arbitrary vector $a = (a_1, \dots, a_n) \in S^{n-1}$ the random sum $S = S(a) = \sum_{k \leq n} a_k X_k$. Then we have*

$$\mathbb{P}\left(|S| < \frac{1}{2}\right) \leq 1 - \frac{(\frac{3}{4})^2}{16\sqrt{12}B} \in (0, 1).$$

Proof: Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ be independent symmetric ± 1 -valued random variables independent to X . By the Payley-Zygmund inequality (Lemma 10) we can estimate

$$\mathbb{P}\left(|S| \geq \frac{1}{2}\right) \geq \frac{(\mathbb{E}|S|^2 - \frac{1}{4})^2}{\mathbb{E}|S|^4} = \frac{(\mathbb{E} \sum_{k \leq n} a_k^2 X_k^2 - \frac{1}{4})^2}{\mathbb{E} |\sum_{k \leq n} a_k X_k|^4} \geq \frac{(1 - \frac{1}{4})^2}{\mathbb{E} |\sum_{k \leq n} a_k X_k|^4}.$$

Now we use the standard-symmetrization inequality (Lemma 11) for the convex function $f : [0, \infty) \rightarrow [0, \infty)$, $x \mapsto x^4$

$$\mathbb{P}\left(|S| \geq \frac{1}{2}\right) \geq \frac{(\frac{3}{4})^2}{16\mathbb{E}_X \mathbb{E}_\varepsilon |\sum_{k \leq n} \varepsilon_k a_k X_k|^4}$$

(where $\mathbb{E}_X / \mathbb{E}_\varepsilon$ denotes the expectation with respect to $X_1, \dots, X_n / \varepsilon_1, \dots, \varepsilon_n$) and the Khintchine inequality (see [18] Lemma 4.1)

$$\mathbb{P}\left(|S| \geq \frac{1}{2}\right) \geq \frac{(\frac{3}{4})^2}{16\sqrt{12}\mathbb{E} |\sum_{k \leq n} a_k^2 X_k^2|^2} = \frac{(\frac{3}{4})^2}{16\sqrt{12}\mathbb{E} \sum_{j, k \leq n} a_j^2 a_k^2 X_j^2 X_k^2}.$$

By Jensen's inequality we have that

$$\mathbb{P}\left(|S| \geq \frac{1}{2}\right) \geq \frac{(\frac{3}{4})^2}{16\sqrt{12} \sum_{j, k \leq n} a_j^2 a_k^2 (\mathbb{E} X_j^4)^{\frac{1}{2}} (\mathbb{E} X_k^4)^{\frac{1}{2}}} = \frac{(\frac{3}{4})^2}{16\sqrt{12}B}$$

and therefore

$$\mathbb{P}\left(|S| < \frac{1}{2}\right) = 1 - \mathbb{P}\left(|S| > \frac{1}{2}\right) \leq 1 - \frac{\left(\frac{3}{4}\right)^2}{16\sqrt{12}B}$$

□

The following statement is essential to prove the bounds on the probability for rectangular random matrices.

Corollary 2 *Let G be a matrix as in Lemma 9. Then there exist constants $\nu, \mu \in (0, 1)$, depending only on B , with*

$$\mathbb{P}(\|Gx\|_2 < \nu\sqrt{n}) \leq \mu^n \text{ for all } x \in S^{k-1}.$$

Proof: Let $x \in S^{k-1}$ be an arbitrary unit vector. Let $G = (X_{i,j})_{i \leq n, j \leq k}$ be a random matrix that suffices the assumptions in Lemma 9. Then it follows that

$$\begin{aligned} \mathbb{P}(\|Gx\|_2 < \lambda\sqrt{n}) &= \mathbb{P}\left(\left(\sum_{i \leq n} \left| \langle (X_{i,j})_{j \leq k}, x \rangle \right|^2\right)^{\frac{1}{2}} < \lambda\sqrt{n}\right) \\ &= \mathbb{P}\left(\sum_{i \leq n} \left| \sum_{j \leq k} X_{i,j} x_j \right|^2 < \lambda^2 n\right) \end{aligned}$$

By Lemma 13 it follows for all $\lambda > 0$ that

$$\mathbb{P}\left(\left| \sum_{j \leq k} X_{i,j} x_j \right| < \frac{1}{2}\right) \leq 1 - \frac{\left(\frac{3}{4}\right)^2}{16\sqrt{12}B}$$

and by Lemma 12 the statement follows. □

Before we can start the proofs of Lemma 9 and Lemma 7 we need the following result for ε -nets.

Lemma 14 *For any $\varepsilon < 1$ there exists an ε -net $\mathfrak{N} \subset S^{n-1}$ such that*

$$|\mathfrak{N}| \leq \left(\frac{3}{\varepsilon}\right)^n.$$

Proof: Let \mathfrak{N} be a maximal ε -separated subset of S^{n-1} , i.e. for all $x, y \in \mathfrak{N}$ we have $\|x - y\|_2 \geq \varepsilon$ (which is the definition of an ε -separated subset) and for every ε -separated subset $\mathfrak{N}' \subset S^{n-1}$ with $\mathfrak{N} \subset \mathfrak{N}'$ we have $\mathfrak{N} = \mathfrak{N}'$. We show that \mathfrak{N} is also an ε -net of S^{n-1} . Therefore, assume that \mathfrak{N} is not an ε -net: Then there exists a $y \in S^{n-1}$ such that for all $x \in \mathfrak{N}$: $d(x, y) \geq \varepsilon$. But then $\mathfrak{N}' = \mathfrak{N} \cup \{y\}$ is a maximal ε -separated subset of S^{n-1} and $\mathfrak{N} \subset \mathfrak{N}'$ which contradicts the maximality of \mathfrak{N} . It follows that \mathfrak{N} is an ε -net of S^{n-1} .

For any distinct points $x, y \in \mathfrak{N}$ we have

$$\left(x + \frac{\varepsilon}{2}B_2^n\right) \cap \left(y + \frac{\varepsilon}{2}B_2^n\right) = \emptyset.$$

Therefore,

$$|\mathfrak{N}| \text{vol}\left(\frac{\varepsilon}{2}B_2^n\right) = \text{vol}\left(\bigcup_{x \in \mathfrak{N}} \left(x + \frac{\varepsilon}{2}B_2^n\right)\right) \leq \text{vol}\left(\left(1 + \frac{\varepsilon}{2}\right)B_2^n\right),$$

which leads to

$$|\mathfrak{N}| \leq \frac{\text{vol}\left(\left(1 + \frac{\varepsilon}{2}\right)B_2^n\right)}{\text{vol}\left(\frac{\varepsilon}{2}B_2^n\right)} = \frac{\left(1 + \frac{\varepsilon}{2}\right)^n}{\left(\frac{\varepsilon}{2}\right)^n} = \left(1 + \frac{2}{\varepsilon}\right)^n \leq \left(\frac{3}{\varepsilon}\right)^n.$$

Since $\mathfrak{N} \subset S^{n-1}$ is an ε -net the lemma is proven. \square

Proof (Lemma 9): In order to prove the statement we use an ε -net argument. Therefore, let $\varepsilon > 0$ to be chosen later. Then by Lemma 14 there exists an ε -net \mathfrak{N} in S^{k-1} with $|\mathfrak{N}| \leq \left(\frac{3}{\varepsilon}\right)^k$. Let $\mu, \nu \in (0, 1)$ denote the numbers in Corollary 2. Then

$$\mathbb{P}(\|Gx\|_2 < \nu\sqrt{n} \text{ for a } x \in \mathfrak{N}) \leq \sum_{x \in \mathfrak{N}} \mathbb{P}(\|Gx\|_2 < \nu\sqrt{n}) \leq \left(\frac{3}{\varepsilon}\right)^k \mu^n.$$

Now we use the properties of the ε -net to get an estimate on the whole sphere. Therefore, let V denote the event that $\|Gy\|_2 \leq \frac{1}{2}\nu\sqrt{n}$ for a vector $y \in S^{k-1}$ and $\|G\|_{2 \rightarrow 2} \leq K\sqrt{n}$. Consider V occurs then we choose a point $x \in \mathfrak{N}$ with $\|x - y\|_2 < \varepsilon$ and get that

$$\|Gx\|_2 \leq \|Gy\|_2 + \|G\|_{2 \rightarrow 2} \|x - y\|_2 \leq \frac{1}{2}\nu\sqrt{n} + K\sqrt{n}\varepsilon.$$

We choose $\varepsilon = \frac{\nu}{2K}$ and get

$$\|Gx\|_2 \leq \nu\sqrt{n}.$$

Then for proper chosen $\delta_0 < 1$ and $\frac{k}{n} < \delta_0$ we get that there exists $d > 0$ such that

$$\mathbb{P}(V) \leq \left(\mu \left(\frac{3}{\varepsilon}\right)^{\frac{k}{n}}\right)^n \leq \exp(-dn).$$

\square

Finally, we have all tools to prove Lemma 7:

Proof (Lemma 7): First we prove the statement for sparse vectors, where we use the estimate from Lemma 9. Then we need an ε -net argument similar to that one used in the proof of Lemma 9 to deduce an estimate for compressible vectors.

W.l.o.g. we can assume that $\delta_0 < \frac{1}{2}$ in Lemma 9. Furthermore we denote $k = \lfloor \delta n \rfloor$, then by Lemma 9

$$\begin{aligned} & \mathbb{P}\left(\inf_{\substack{x \in \text{Sparse}(\delta) \\ \|x\|_2=1}} \|Ax\|_2 \leq c\sqrt{n} \text{ and } \|A\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \\ &= \mathbb{P}\left(\exists \sigma \subset \{1, \dots, n\}, |\sigma| = k : \inf_{\substack{x \in \mathbb{R}^\sigma \\ \|x\|_2=1}} \|Ax\|_2 \leq c\sqrt{n} \text{ and } \|A\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \\ &\leq \sum_{\substack{\sigma \subset \{1, \dots, n\} \\ |\sigma|=k}} \mathbb{P}\left(\inf_{\substack{x \in \mathbb{R}^\sigma \\ \|x\|_2=1}} \|Ax\|_2 \leq c\sqrt{n} \text{ and } \|A\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \\ &\leq \binom{n}{k} \exp(-dn). \end{aligned}$$

By the binomial theorem we obtain

$$\binom{n}{k} \leq \sum_{m=0}^n \binom{n}{m} \left(\frac{k}{n}\right)^m \cdot \left(\frac{n}{k}\right)^k = \left(1 + \frac{k}{n}\right)^n \left(\frac{n}{k}\right)^k \leq \lim_{N \rightarrow \infty} \left(1 + \frac{k}{N}\right)^N \left(\frac{n}{k}\right)^k = \exp(k) \left(\frac{n}{k}\right)^k.$$

Therefore, by the definition of k we get that

$$\begin{aligned} \mathbb{P}\left(\inf_{\substack{x \in \text{Sparse}(\delta) \\ \|x\|_2=1}} \|Ax\|_2 \leq c\sqrt{n} \text{ and } \|A\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) &\leq \left(\frac{ne}{k}\right)^k \exp(-dn) \leq \left(\frac{e}{\frac{1}{2}\delta}\right)^{\delta n} \exp(-dn) \\ &= \exp\left(\delta n \ln\left(\frac{2e}{\delta}\right) - dn\right) \leq \exp(-d'n) \end{aligned}$$

for a well chosen $\delta < \delta_0$, depending on d . Now we want to transfer this estimate to compressible vectors. Let $C > 0$ and $\rho \in \left(0, \frac{1}{2}\right)$ be chosen later. We want to estimate $\mathbb{P}(V)$, where V denotes the event that $\|Ax\|_2 \leq C\sqrt{n}$ for a vector $x \in \text{Comp}(\delta, \rho)$ and $\|A\|_{2 \rightarrow 2} \leq K\sqrt{n}$. Assume that V occurs. For all $x \in \text{Comp}(\delta, \rho)$ we know that

$$d(x, \text{Sparse}(\delta)) = \inf_{y \in \text{Sparse}(\delta)} \|x - y\|_2 < \rho$$

by definition. Since

$$\text{Sparse}(\delta) = \bigcup_{\substack{\sigma \subset \{1, \dots, n\} \\ |\sigma|=k}} \mathbb{R}^\sigma$$

we know there exist a $y \in \text{Sparse}(\delta)$ and a $z \in \mathbb{R}^n$ with $\|z\|_2 \leq \rho$ such that $x = y + z$. It follows that $\|y\|_2 \geq 1 - \rho \geq \frac{1}{2}$ and

$$\|Ay\|_2 \leq \|Ax\|_2 + \|A\|_{2 \rightarrow 2} \|z\|_2 \leq C\sqrt{n} + \rho K\sqrt{n}.$$

We choose $C = \frac{c}{4}$ and $\rho = \frac{c}{4K}$ and get $\|Ay\|_2 \leq \frac{1}{2}c\sqrt{n}$. Since $\|y\|_2 \geq \frac{1}{2}$ we found in $u = \frac{y}{\|y\|_2}$ a sparse unit vector with $\|Au\|_2 \leq c\sqrt{n}$. It follows that $\mathbb{P}(V) \leq \exp(-d'n)$. \square

Chapter 4

Upper bound for the least singular value

In this chapter we prove our main result for the upper bound. More precisely in Theorem 11 we prove that the probability that an $n \times n$ random matrix $(a_i b_j X_{i,j})_{i,j \leq n}$ is bigger than

$$\frac{1}{\sqrt{n}} \min \left\{ \frac{\|(b_j)_{j \leq n}\|_2}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, \frac{\|(a_i)_{i \leq n}\|_2}{\|(\frac{1}{b_j})_{j \leq n}\|_2} \right\}$$

is small. In addition to that we show in Theorem 13 a similar bound for rectangular matrices. We want to remark that in the last setting we had to lower the assumptions on the distribution of the entries and that the bound we found on the probability is only useful for matrices that are almost square matrices. We also give an application of Theorem 13: We apply the theorem to the situation where we have to handle square random matrices which have rows whose entries are of unknown distribution.

An important tool are the incompressible vectors whose properties will be studied in the sequel. The next lemma, which can be found in [28] (see page 16) states that incompressible vectors can be seen as a generalization of spread vectors, i.e. vectors that suffice $\|x\|_2 = 1$ and $|x_k| \sim \frac{1}{\sqrt{n}}$ for all $k \in \{1, \dots, n\}$.

Lemma 15 *Let $x = (x_k)_{k \leq n} \in \text{Incomp}(\delta, \rho)$. Then there exists a set $\sigma(x) \subset \{1, \dots, n\}$ of cardinality $|\sigma(x)| \geq \frac{1}{2} \rho^2 \delta n$ and such that*

$$\frac{\rho}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\delta n}} \text{ for all } k \in \sigma(x).$$

Proof: Let $x \in \text{Incomp}(\delta, \rho)$. For a subset $\alpha \subset \{1, \dots, n\}$ let P_α denote the orthogonal projection onto $\mathbb{R}^\alpha \subset \mathbb{R}^n$. We define the sets $\sigma_1(x) := \left\{ k \in \{1, \dots, n\} \mid |x_k| \leq \frac{1}{\sqrt{\delta n}} \right\}$ and $\sigma_2(x) := \left\{ k \in \{1, \dots, n\} \mid |x_k| \geq \frac{\rho}{\sqrt{2n}} \right\}$. Then $\sigma(x) = \sigma_1(x) \cap \sigma_2(x)$.

First, we estimate $|\sigma_1^c(x)|$:

$$\begin{aligned}
1 &= \sum_{k \leq n} |x_k|^2 \\
&\geq \sum_{k \leq n} \frac{1}{\delta n} \mathbf{1}_{\{|x_k|^2 > \frac{1}{\delta n}\}} \\
&= \frac{1}{\delta n} \left| \left\{ k \in \{1, \dots, n\} \mid |x_k| > \frac{1}{\sqrt{\delta n}} \right\} \right| \\
&= \frac{1}{\delta n} |\sigma_1^c(x)|.
\end{aligned}$$

Therefore the cardinality of $\sigma_1^c(x)$ is less or equal δn . Then $y := P_{\sigma_1^c(x)}x \in \text{Sparse}(\delta)$, so from the incompressibility of the vector x it follows that $\|P_{\sigma_1(x)}x\|_2 = \|x - y\|_2 > \rho$. We use the definition of the set $\sigma_2(x)$ and get $\|P_{\sigma_2^c(x)}x\|_2^2 \leq n \cdot \frac{\rho^2}{2n} = \frac{\rho^2}{2}$. Since

$$\sigma_1(x) = \{\sigma_1(x) \cap \sigma_2(x)\} \cup \{\sigma_1(x) \cap \sigma_2^c(x)\} = \sigma(x) \cup \{\sigma_1(x) \cap \sigma_2^c(x)\} \subset \sigma(x) \cup \sigma_2^c(x)$$

it follows from the Pythagorean theorem that

$$\begin{aligned}
\|P_{\sigma_1(x)}x\|_2^2 &= \|P_{\sigma(x)}x + P_{\{\sigma_1(x) \cap \sigma_2^c(x)\}}x\|_2^2 \\
&= \|P_{\sigma(x)}x\|_2^2 + \|P_{\{\sigma_1(x) \cap \sigma_2^c(x)\}}x\|_2^2 \\
&\leq \|P_{\sigma(x)}x\|_2^2 + \|P_{\sigma_2^c(x)}x\|_2^2.
\end{aligned}$$

Therefore we have

$$\|P_{\sigma(x)}x\|_2^2 \geq \|P_{\sigma_1(x)}x\|_2^2 - \|P_{\sigma_2^c(x)}x\|_2^2 > \frac{\rho^2}{2}.$$

On the other hand, we have by the definition of $\sigma_1(x) \supset \sigma(x)$

$$\|P_{\sigma(x)}x\|_2^2 \leq \|P_{\sigma(x)}x\|_\infty^2 \cdot |\sigma(x)| \leq \frac{1}{\delta n} |\sigma(x)|.$$

It follows $|\sigma(x)| \geq \frac{1}{2}\rho^2\delta n$. □

Definition 6 Let $\delta, \rho \in (0, 1)$. We define $\text{Incomp}_1(\delta, \rho) := \{x \in \partial B_1^n \mid \exists \lambda \in \mathbb{R} : \lambda x \in \text{Incomp}(\delta, \rho)\}$, where ∂B_1^n is defined as $\partial B_1^n := \{x \in \mathbb{R}^n \mid \|x\|_1 = 1\}$.

The following lemma states that incompressible vectors are almost spherical in some sense.

Lemma 16 Let $\delta, \rho \in (0, 1)$. For all $x \in \text{Incomp}(\delta, \rho)$ we have

$$\frac{\rho^3 \delta}{\sqrt{2^3}} \sqrt{n} \|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2.$$

Proof: For all $x \in \text{Incomp}(\delta, \rho)$ it follows from Lemma 15 that there exists a set $\sigma(x) \subset \{1, \dots, n\}$ of cardinality $|\sigma(x)| \geq \frac{1}{2}\rho^2\delta n$ with the property

$$\frac{\rho}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\delta n}} \text{ for all } k \in \sigma(x).$$

Therefore we get that

$$\sqrt{n} \|x\|_2 \geq \|x\|_1 \geq \sum_{k \in \sigma(x)} |x_k| \geq \sum_{k \in \sigma(x)} \frac{\rho}{\sqrt{2n}} \geq \frac{\frac{1}{2}\rho^3\delta n}{\sqrt{2n}} = \frac{\rho^3\delta}{\sqrt{2^3}} \sqrt{n} \|x\|_2.$$

□

Let $1 \leq s \leq n$ be integers. Let $X = (X_{i,j})_{i,j \leq n}$ be a random matrix, whose entries are independent copies of a centered random variable with unit variance and subgaussian moment bounded by B . Then Lemma 17, which can be found in [43, page 15], ensures non-degeneracy of the s -dimensional subspace $V_{s,n}$, orthogonal to the columns X_{s+1}, \dots, X_n of X . This is useful, if one wants to apply Berry-Esséen's central limit theorem.

Lemma 17 *Let $1 \leq s \leq n$ be integers and let $X_i \in \mathbb{R}^n$ ($i \in \{1, \dots, n\}$) be random vectors whose entries are independent copies of a centered subgaussian random variable with unit variance. Let E_q denote the event that there exists a unit vector $v = (v_1, \dots, v_n) \in V_{s,n} = \text{span}\{X_i \mid i = s+1, \dots, n\}^\perp$ such that $\max_{1 \leq i \leq n} |v_i| \geq n^{-q}$. If q is a sufficiently small absolute constant, then*

$$\mathbb{P}(E_q) \ll \exp(-n^Q) \text{ for some constant } Q > 0.$$

Remark: One can think of q as $q = \frac{1}{20}$ (see [43, page 15]).

The first one to mention that normal vectors on a random hyperplane (i.e. random unit vectors that are orthogonal to the random hyperplane) are non-degenerate were Tao and Vu in [42]. This idea has become a much used tool in the study of the least singular value of random matrices. However, the notion of non-degeneracy of random normals of Rudelson and Vershynin differs from the one in the lemma above: In [28] it was only required that random normals are incompressible vectors with high probability, which means that usually there is enough space between a random normal and the set of sparse vectors. But Lemma 17 sharpens this statement: no single coordinate of the random normal vector is able to have a significant portion of the mass of the vector. The following theorem deals with the invertibility of subgaussian random matrices and can be found in [28, page 3].

Theorem 10 (Rudelson, Vershynin) *Let $n \in \mathbb{N}$ be an integer. Let ξ_1, \dots, ξ_n be independent centered random variables with variances at least one and subgaussian moments bounded by an absolute constant $B > 0$. Define A to be the $n \times n$ matrix whose rows are independent copies of the random vector (ξ_1, \dots, ξ_n) . Then for every $\varepsilon \geq 0$ one has that*

$$\mathbb{P}\left(s_n(A) \leq \varepsilon \frac{1}{\sqrt{n}}\right) \leq C\varepsilon + c^n,$$

where $C > 0$ and $c \in (0, 1)$ are constants depending only on B .

If one thinks of a Gaussian matrix, i.e. matrices whose entries are i.i.d. normal distributed random variables, we know almost surely that the matrix is invertible. Therefore, one might expect that the property of being invertible holds for subgaussian matrices, too. This is not the case, since ± 1 -valued Bernoulli variables are subgaussian and random sign-matrices are singular with positive probability. Kahn, Komlós and Szemerédi showed in [13] that the singularity probability is exponentially small:

$$\mathbb{P}(n \times n \text{ random sign matrix is singular}) < c^n$$

for some universal constant $c \in (0, 1)$. Thus the summand c^n in the above inequality gives a bound on the event that the matrix $(X_{i,j})_{i,j \leq n}$ is not invertible.

Lemma 18 *For $k \in \{1, \dots, n\}$ let $Z_k : \Omega \rightarrow [0, \infty)$ be non-negative random variables and $a_k \geq 0$ be non-negative real numbers with $\sum_{k \leq n} a_k = 1$. Then for every $\varepsilon > 0$*

$$\mathbb{P}\left(\sum_{k \leq n} a_k Z_k \leq \varepsilon\right) \leq 2 \sum_{k \leq n} a_k \mathbb{P}(Z_k \leq 2\varepsilon).$$

Proof: It suffices to show that

$$\mathbb{1}_{\{\sum_{k \leq n} a_k Z_k \leq \varepsilon\}} \leq 2 \sum_{k \leq n} a_k \mathbb{1}_{\{Z_k \leq 2\varepsilon\}},$$

because then it follows that

$$\mathbb{P}\left(\sum_{k \leq n} a_k Z_k \leq \varepsilon\right) = \int \mathbb{1}_{\{\sum_{k \leq n} a_k Z_k \leq \varepsilon\}} d\mathbb{P} \leq \int 2 \sum_{k \leq n} a_k \mathbb{1}_{\{Z_k \leq 2\varepsilon\}} d\mathbb{P} = 2 \sum_{k \leq n} a_k \mathbb{P}(Z_k \leq 2\varepsilon).$$

For all $\omega \in \Omega$ such that $\sum_{k \leq n} a_k Z_k(\omega) > \varepsilon$ follows that

$$0 = \mathbb{1}_{\{\sum_{k \leq n} a_k Z_k(\omega) \leq \varepsilon\}} \leq 2 \sum_{k \leq n} a_k \mathbb{1}_{\{Z_k(\omega) \leq 2\varepsilon\}}.$$

So we only need to show the inequality in the case $\sum_{k \leq n} a_k Z_k(\omega) \leq \varepsilon$:

$$\begin{aligned} \mathbb{1}_{\{\sum_{k \leq n} a_k Z_k(\omega) \leq \varepsilon\}} &= \mathbb{1}_{\{2\varepsilon - \sum_{k \leq n} a_k Z_k(\omega) \geq \varepsilon\}} \\ &\leq \frac{2\varepsilon - \sum_{k \leq n} a_k Z_k(\omega)}{\varepsilon} \\ &= 2 - \sum_{k \leq n} \frac{a_k}{\varepsilon} Z_k(\omega) \\ &= 2 \sum_{k \leq n} a_k \left(1 - \frac{Z_k(\omega)}{2\varepsilon}\right) \\ &\leq 2 \sum_{k \leq n} a_k \mathbb{1}_{\{Z_k(\omega) \leq 2\varepsilon\}}. \end{aligned}$$

□

4.1 Upper bound for tensor matrices

We are now in the position to prove our main theorem.

Theorem 11 *Let $n \in \mathbb{N}$. Let $X = (X_{i,j})_{i,j \leq n}$ be a random matrix, whose entries are independent copies of a centered random variable with unit variance and subgaussian moment bounded by B . Let $a_1, \dots, a_n, b_1, \dots, b_n \in (0, 1]$ be finite sequences of reals. Then it exists a constant $c > 0$ depending only on B such that for every $c\sqrt{n} \geq \lambda \geq 2$ we have*

$$\mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > 2\lambda \frac{1}{\sqrt{n}} \min\left\{\frac{\|(b_i)_{i \leq n}\|_2}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, \frac{\|(a_i)_{i \leq n}\|_2}{\|(\frac{1}{b_i})_{i \leq n}\|_2}\right\}\right) \leq \left(\frac{6}{\lambda}\right) \ln(\lambda^2) + Qn^{-q}$$

for suitable constants $Q, q > 0$.

Proof: By Lemma 4 we know that

$$\begin{aligned} \mathbb{P}((a_i b_j X_{i,j})_{i,j \leq n} \text{ is singular}) &= \mathbb{P}((X_{i,j})_{i,j \leq n} \text{ is singular}) \\ &= \mathbb{P}(s_n((X_{i,j})_{i,j \leq n}) = 0) \\ &\leq c^n, \end{aligned}$$

for a $c \in (0, 1)$, where we used Theorem 10 in the inequality. It follows that

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \\
&= \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&\quad + \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is singular}\right) \\
&\leq \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) + c^n.
\end{aligned}$$

In the first step of our proof we want to get rid of the factors b_j in each coefficient of the random matrix to come into the position to deal with the matrix $(a_i X_{i,j})_{i,j \leq n}$. We use Lemma 2 to do so

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&= \mathbb{P}\left(\min_{\|x\|_2=1} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&= \mathbb{P}\left(\frac{1}{\max_{\|x\|_2=1} \|(a_i b_j X_{i,j})_{i,j \leq n}^{-1} x\|_2} > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right).
\end{aligned}$$

We denote the j -th entry of the i -th row of X^{-1} by $X_{i,j}^*$. Therefore we get by Lemma 4 that

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&= \mathbb{P}\left(\frac{1}{\max_{\|x\|_2=1} \|(\frac{1}{a_j} \frac{1}{b_i} X_{i,j}^*)_{i,j \leq n} x\|_2} > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&= \mathbb{P}\left(\frac{1}{\max_{\|x\|_2=1} (\sum_{i \leq n} |\frac{1}{b_i}|^2 |\sum_{j \leq n} \frac{1}{a_j} x_j X_{i,j}^*|^2)^{\frac{1}{2}} (\sum_{i \leq n} |b_i|^2)^{\frac{1}{2}}} > \frac{\lambda}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2},\right. \\
&\quad \left.(X_{i,j})_{i,j \leq n} \text{ is non-singular}\right).
\end{aligned}$$

By Hölder's inequality we know that

$$\sum_{i \leq n} \left| \sum_{j \leq n} \frac{1}{a_j} x_j X_{i,j}^* \right| \leq \left(\sum_{i \leq n} \left| \frac{1}{b_i} \right|^2 \left| \sum_{j \leq n} \frac{1}{a_j} x_j X_{i,j}^* \right|^2 \right)^{\frac{1}{2}} \left(\sum_{i \leq n} |b_i|^2 \right)^{\frac{1}{2}}$$

and obtain

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&\leq \mathbb{P}\left(\frac{1}{\max_{\|x\|_2=1} \sum_{i \leq n} |\sum_{j \leq n} \frac{1}{a_j} x_j X_{i,j}^*|} > \frac{\lambda}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&\stackrel{\text{Lemma 4}}{=} \mathbb{P}\left(\frac{1}{\max_{\|x\|_2=1} \|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_1} > \frac{\lambda}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
&\stackrel{\text{Lemma 2}}{=} \mathbb{P}\left(\min_{\|x\|_1=1} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\lambda}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right).
\end{aligned}$$

Choose $\delta, \rho \in (0, 1)$ such that they satisfy the estimate in Lemma 7. Since $\text{Incomp}_1(\delta, \rho) \subset \partial B_1^n$ it follows that

$$\begin{aligned}
& \mathbb{P} \left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right) \\
& \leq \mathbb{P} \left(\inf_{x \in \text{Incomp}_1(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\lambda}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right) \\
& = \mathbb{P} \left(\inf_{x \in \text{Incomp}_1(\delta, \rho)} \left(\left\| (a_i X_{i,j})_{i,j \leq n} \frac{x}{\|x\|_2} \right\|_2 \|x\|_2 \right) > \frac{\lambda}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right).
\end{aligned}$$

From the definition of $\text{Incomp}_1(\delta, \rho)$ it follows that $x \in \text{Incomp}_1(\delta, \rho)$ is equivalent to $\frac{x}{\|x\|_2} \in \text{Incomp}(\delta, \rho)$ and by Lemma 16 we conclude that for $x \in \text{Incomp}_1(\delta, \rho)$ we have

$$1 = \|x\|_1 \geq \frac{\sqrt{n} \delta \rho^3}{\sqrt{2^3}} \|x\|_2.$$

Therefore we get that

$$\begin{aligned}
& \mathbb{P} \left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right) \\
& \leq \mathbb{P} \left(\inf_{x \in \text{Incomp}_1(\delta, \rho)} \left\| (a_i X_{i,j})_{i,j \leq n} \frac{x}{\|x\|_2} \right\|_2 \frac{\sqrt{2^3}}{\sqrt{n} \delta \rho^3} > \frac{\lambda}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right) \\
& = \mathbb{P} \left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right) \\
& = \mathbb{P} \left(\left[\left\{ \inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right. \right. \\
& \quad \cap \left. \left\{ \inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right] \\
& \quad \cup \left[\left\{ \inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right. \\
& \quad \cap \left. \left\{ \inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right] \\
& \quad \cap \left. \left\{ (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right\} \right).
\end{aligned}$$

We use the union bound and get that

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& \leq \mathbb{P}\left(\left\{ \inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right. \\
& \quad \cap \left\{ \inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \\
& \quad \left. \cap \left\{ (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right\} \right) \\
& + \mathbb{P}\left(\left\{ \inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right. \\
& \quad \left. \cap \left\{ \inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right) \\
& \leq \mathbb{P}\left(\left\{ \inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right. \\
& \quad \cap \left\{ \inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \\
& \quad \left. \cap \left\{ (X_{i,j})_{i,j \leq n} \text{ is non-singular} \right\} \right) \\
& + \mathbb{P}\left(\left\{ \inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \right).
\end{aligned}$$

Since $S^{n-1} = \text{Comp}(\delta, \rho) \cup \text{Incomp}(\delta, \rho)$ we get that

$$\begin{aligned}
\left\{ \min_{\|x\|_2=1} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} &= \left\{ \inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\} \\
& \quad \cap \left\{ \inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \right\}.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \quad (4.1) \\
& \leq \mathbb{P}\left(\min_{\|x\|_2=1} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 > \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right).
\end{aligned}$$

These terms can be controlled separately. The second term in inequation (4.1) can be estimated by Lemma 7:

$$\begin{aligned}
& \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \\
&= \mathbb{P}\left(\left[\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right] \cap \left\{\|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right\}\right) \\
&\cup \left[\left[\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right] \cap \left\{\|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} > K\sqrt{n}\right\}\right] \\
&\leq \mathbb{P}\left(\left[\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right] \cap \left\{\|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right\}\right) \\
&\quad + \mathbb{P}\left(\left[\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right] \cap \left\{\|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} > K\sqrt{n}\right\}\right),
\end{aligned}$$

where we used the union bound and $K \geq 0$ is the constant from Lemma 8. It follows by Lemma 8 that

$$\begin{aligned}
& \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \\
&\leq \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, \|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \\
&\quad + \mathbb{P}\left(\|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} > K\sqrt{n}\right) \\
&\leq \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, \|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) + 2 \exp(-n).
\end{aligned}$$

Since $\|(\frac{1}{a_i})_{i \leq n}\|_2 \geq \|(\frac{1}{a_i})_{i \leq n}\|_\infty$ we obtain

$$\begin{aligned}
& \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \\
&\leq \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_\infty}, \|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) + 2 \exp(-n) \\
&= \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \lambda \min_{1 \leq i \leq n} |a_i|, \|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) + 2 \exp(-n) \\
&= \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \left(\sum_{i \leq n} |a_i|^2 < (X_{i,j})_{j \leq n}, x >^2\right)^{\frac{1}{2}} \leq \frac{\delta \rho^3}{\sqrt{2^3}} \lambda \min_{1 \leq i \leq n} |a_i|, \|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) \\
&\quad + 2 \exp(-n) \\
&\leq \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \left(\sum_{i \leq n} < (X_{i,j})_{j \leq n}, x >^2\right)^{\frac{1}{2}} \leq \frac{\delta \rho^3}{\sqrt{2^3}} \lambda, \|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) + 2 \exp(-n) \\
&= \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \lambda, \|(X_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} \leq K\sqrt{n}\right) + 2 \exp(-n).
\end{aligned}$$

We use Lemma 7 to obtain the existence of a constant $c > 0$ depending only on B such that for all $c\sqrt{n} \geq \lambda \geq 2$ we get a bound from above:

$$\mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\delta \rho^3}{\sqrt{2^3}} \frac{\lambda}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \leq \exp(-c'n) + 2 \exp(-n),$$

where c' only depends on B . Now we need to control the first term in inequality 4.1. We want to point out that we are in the situation that the event that $(X_{i,j})_{i,j \leq n}$ is non-singular occurs.

Therefore, we are able to compute the inverse. The following argument is orientated on [27]. In the sequel we denote by X_k the k -th column $(X_{i,j})_{i,j \leq n} e_k$ of $(X_{i,j})_{i,j \leq n}$. We want to remind the reader that (X_k, X_k^*) is a biorthogonal system, where X_k^* denotes $((X_{i,j})_{i,j \leq n}^{-1})^T e_k$. Let $u, v > 0$. Consider that there exists a vector $x \in \mathbb{R}^n$ such that

$$\|x\|_2 \leq u \text{ and } \|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2 \geq v \left\| \left(\frac{1}{a_i} \right)_{i \leq n} \right\|_2.$$

Then

$$s_n((a_i X_{i,j})_{i,j \leq n}) = \frac{1}{\|(a_i X_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2}} = \frac{1}{\max_{\|x\|_2 \neq 0} \frac{\|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2}{\|x\|_2}} \leq \frac{u}{v \left\| \left(\frac{1}{a_i} \right)_{i \leq n} \right\|_2}.$$

W.l.o.g. we can assume that $\max_{1 \leq i \leq n} |a_i| = a_1$ and that $a_1 = 1$. Define P_1 as the orthogonal projection onto $H_1 = \text{span}\{a_i X_i \mid i = 2, \dots, n\} = \text{span}\{X_i \mid i = 2, \dots, n\}$. We define

$$x = a_1 X_1 - P_1 a_1 X_1 = X_1 - P_1 X_1.$$

$X_1 - P_1 X_1$ is an element of H_1^\perp :

$$\begin{aligned} \langle X_1 - P_1 X_1, X_k \rangle &= \langle X_1, X_k \rangle - \langle P_1 X_1, X_k \rangle \\ &= \langle X_1, X_k \rangle - \langle X_1, P_1 X_k \rangle \\ &= \langle X_1, X_k \rangle - \langle X_1, X_k \rangle \\ &= 0 \text{ for all } k = 2, \dots, n. \end{aligned}$$

Therefore, we obtain by an analogue computation as in the proof of Lemma 3

$$\|x\|_2 = \|X_1 - P_1 X_1\|_2 = \text{dist}(X_1, H_1).$$

Analogue to the computation in the proof of Lemma 3 one can see that

$$\text{dist}(X_1, H_1) = |\langle N_1, X_1 \rangle|,$$

where N_1 is a random normal on H_1 , i.e. a vector of unit length that is orthonormal on the hyperplane H_1 . If one is only interested in the main idea of the estimation of $\mathbb{P}(\|x\|_2 > u)$ and want to omit the details, one can think of the entries of X as standard normal distributed random variables. In this particular case we have by Lemma 6 and the remark below that $\|x\|_2 = \text{dist}(X_1, H_1)$ is equal in distribution to $|g|$, where g is standard normal distributed and obtain

$$\mathbb{P}(\|x\|_2 > u) = \mathbb{P}(|g| > u) = \int_u^\infty \sqrt{\frac{2}{\pi}} \exp\left(-\frac{s^2}{2}\right) ds \leq 2 \exp\left(-\frac{u^2}{2}\right)$$

Since we want to handle a more general situation, we have to use Berry-Esséen's central limit theorem (Theorem 5) to get that $\|x\|_2$ is almost Gaussian distributed.

We want to remind the reader that the vector N_1 depends only on H_1 and thus can be determined by the vectors X_2, \dots, X_n . Therefore, N_1 is stochastically independent to X_1 . Since independence of two random variables X_1, N_1 is equivalent to the fact that the joint distribution of (X_1, N_1) namely $\mathbb{P}^{(X_1, N_1)}$ is equal to the product measure $\mathbb{P}^{X_1} \times \mathbb{P}^{N_1}$ we get

$$\begin{aligned} \mathbb{P}(\|x\|_2 > u) &= \mathbb{P}^{(X_1, N_1)}(|\langle N_1, X_1 \rangle| > u) \\ &= \mathbb{P}^{X_1} \times \mathbb{P}^{N_1}(|\langle N_1, X_1 \rangle| > u) \\ &= \int \mathbf{1}_{\{|\langle N_1, X_1 \rangle| > u\}} d\mathbb{P}^{X_1} \times \mathbb{P}^{N_1}. \end{aligned}$$

Therefore we get by Fubini's theorem

$$\begin{aligned}\mathbb{P}(\|x\|_2 > u) &= \int \int \mathbf{1}_{\{|\langle N_1, X_1 \rangle| > u\}} d\mathbb{P}^{X_1} d\mathbb{P}^{N_1} \\ &= \int \mathbb{P}^{X_1} \left(\left| \sum_{k \leq n} N_{1,k} X_{1,k} \right| > u \right) d\mathbb{P}^{N_1}.\end{aligned}$$

Now we use Berry-Esséen's central limit theorem (Theorem 5) to get (for suitable constants $c, C > 0$ that suffice the conditions in Lemma 17):

$$\begin{aligned}\mathbb{P}(\|x\|_2 > u) &\leq \int \mathbb{P}(|g| > u) + C \max_{1 \leq k \leq n} |N_{1,k}| d\mathbb{P}^{N_1} \\ &= \int_{\{\max_{1 \leq k \leq n} |N_{1,k}| \geq n^{-c}\}} \mathbb{P}(|g| > u) + C \max_{1 \leq k \leq n} |N_{1,k}| d\mathbb{P}^{N_1} \\ &\quad + \int_{\{\max_{1 \leq k \leq n} |N_{1,k}| \leq n^{-c}\}} \mathbb{P}(|g| > u) + C \max_{1 \leq k \leq n} |N_{1,k}| d\mathbb{P}^{N_1} \\ &\leq \int_{\{\max_{1 \leq k \leq n} |N_{1,k}| \geq n^{-c}\}} 1 + C d\mathbb{P}^{N_1} \\ &\quad + \int_{\{\max_{1 \leq k \leq n} |N_{1,k}| \leq n^{-c}\}} \mathbb{P}(|g| > u) + C \max_{1 \leq k \leq n} |N_{1,k}| d\mathbb{P}^{N_1} \\ &= (1 + C) \mathbb{P}^{N_1}(\max_{1 \leq k \leq n} |N_{1,k}| \geq n^{-c}) \\ &\quad + \int_{\{\max_{1 \leq k \leq n} |N_{1,k}| \leq n^{-c}\}} \mathbb{P}(|g| > u) + C \max_{1 \leq k \leq n} |N_{1,k}| d\mathbb{P}^{N_1}.\end{aligned}$$

By Lemma 17 we can estimate the first term by $(1 + C) \exp(-nc')$ for a suitable constant c' and we obtain

$$\begin{aligned}\mathbb{P}(\|x\|_2 > u) &\leq (1 + C) \exp(-nc') + \mathbb{P}(|g| > u) + Cn^{-c} \\ &\leq \mathbb{P}(|g| > u) + C'n^{-c} \\ &\leq 2 \exp\left(-\frac{u^2}{2}\right) + C'n^{-c},\end{aligned}$$

for a suitable constant $C' > 0$. Now we need to estimate $\|(a_i X_{i,j})_{i,j \leq n} x\|_2$ from below:

$$\|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2^2 = \|(a_i X_{i,j})_{i,j \leq n}^{-1} (X_1 - P_1 X_1)\|_2^2 = \|e_1 - (a_i X_{i,j})_{i,j \leq n}^{-1} P_1 X_1\|_2^2.$$

Since $P_1 X_1 \in H_1$ and $X_1^* \in H_1^\perp$ it follows that

$$\langle e_1, (a_i X_{i,j})_{i,j \leq n}^{-1} P_1 X_1 \rangle = \langle ((a_i X_{i,j})_{i,j \leq n}^{-1})^T e_1, P_1 X_1 \rangle = \langle X_1^*, P_1 X_1 \rangle = 0.$$

Consequently, we have $(a_i X_{i,j})_{i,j \leq n}^{-1} P_1 X_1 \perp e_1$ and it follows by the Pythagorean theorem that

$$\begin{aligned}
\|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2^2 &= \|e_1\|_2^2 + \|(a_i X_{i,j})_{i,j \leq n}^{-1} P_1 X_1\|_2^2 \\
&> \|(a_i X_{i,j})_{i,j \leq n}^{-1} P_1 X_1\|_2^2 \\
&= \sum_{k \leq n} \langle (a_i X_{i,j})_{i,j \leq n}^{-1} P_1 X_1, e_k \rangle^2 \\
&= \sum_{k \leq n} \langle P_1 ((a_i X_{i,j})_{i,j \leq n}^{-1})^T e_k, X_1 \rangle^2 \\
&= \sum_{k \leq n} \left\langle P_1 \left(\frac{1}{a_j} X_{i,j}^* \right)_{i,j \leq n}^T e_k, X_1 \right\rangle^2 \\
&= \sum_{k \leq n} \left\langle P_1 \frac{1}{a_k} X_k^*, X_1 \right\rangle^2,
\end{aligned} \tag{4.2}$$

where we have used Lemma 4. Since $X_1^* \in H_1^\perp$ it follows that $\ker P_1 = \text{span}\{X_1^*\}$ and we obtain

$$\|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2^2 > \sum_{k=2}^n \left| \frac{1}{a_k} \right|^2 \langle P_1 X_k^*, X_1 \rangle^2.$$

For $k = 2, \dots, n$ we denote $Y_k^* = P_1 X_k^* \in H_1$. We have

$$P_1(Y_k^* - X_k^*) = P_1 Y_k^* - P_1 X_k^* = P_1^2 X_k^* - P_1 X_k^* = 0$$

since P_1 is an orthogonal projection. Thus we get for all $k \in \{2, \dots, n\}$ that $Y_k^* - X_k^* \in \ker P_1 = \text{span}\{X_1^*\}$. Therefore, we get that $Y_k^* = X_k^* - \lambda_k X_1^*$ for some $\lambda_k \in \mathbb{R}$ and all $k = 2, \dots, n$. We use the fact that for all $k = 2, \dots, n$ we have $\langle X_1^*, X_k \rangle = 0$ to obtain

$$\langle Y_j^*, X_k \rangle = \langle X_j^* - \lambda_j X_1^*, X_k \rangle = \langle X_j^*, X_k \rangle - \lambda_j \langle X_1^*, X_k \rangle = \delta_{j,k}$$

for all $j, k = 2, \dots, n$. Therefore $(Y_k^*, X_k)_{k=2}^n$ is a biorthogonal system in $H_1 = \text{span}\{X_k \mid k = 2, \dots, n\}$. Furthermore by Lemma 3 this system is unique (we want to point out that $(X_i, X_i^*)_{i=2}^n$ is not contained in H_1) and the vectors $(Y_k^*)_{k=2}^n$ are uniquely determined by the system $(X_k)_{k=2}^n$. Thus the vectors $(Y_k^*)_{k=2}^n$ and X_1 are stochastically independent.

We use Lemma 3 to recognize that $\|Y_j^*\|_2 = \frac{1}{\text{dist}(X_j, H_{1,j})}$, where $H_{1,j} = \text{span}\{X_k \mid k \notin \{1, j\}\}$ and the related Hilbert space is now H_1 . So it follows that

$$\|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2^2 > \sum_{k=2}^n \left| \frac{1}{a_k} \right|^2 \left(\frac{s_k}{t_k} \right)^2,$$

where we define $s_k = \left| \left\langle \frac{Y_k^*}{\|Y_k^*\|_2}, X_1 \right\rangle \right|$ and $t_k = \text{dist}(X_k, H_{1,k})$, for $k = 2, \dots, n$.

W.l.o.g. we assume $k = 2$. For t_2 we get analogue to our computation for $\|x\|_2$ that for every $\nu > 0$

$$\mathbb{P}(t_2 > \nu) \leq 2 \exp\left(-\frac{\nu^2}{2}\right) + Cn^{-c},$$

for suitable constants c, C . Furthermore, since $\hat{Y}_2^* = \frac{Y_2^*}{\|Y_2^*\|_2} \in (H_{1,2})^\perp$ is stochastically independent to X_1 and $\sum_{k \leq n} |\hat{Y}_{2,i}^*|^2 = 1$ we get by Berry-Esséen's central limit theorem (Theorem 5) and Lemma 17 with similar computations as above that

$$\mathbb{P}(s_2 \leq \varepsilon) = \mathbb{P}\left(\left| \sum_{k \leq n} \hat{Y}_{2,k}^* X_{1,k} \right| \leq \varepsilon\right) \leq \mathbb{P}(|g| \leq \varepsilon) + Fn^{-f} \leq \varepsilon + Fn^{-f}$$

for suitable constants $f, F > 0$. It follows that

$$\begin{aligned}
\mathbb{P}\left(\frac{s_2}{t_2} \leq \frac{\varepsilon}{\nu}\right) &\leq \mathbb{P}(s_2 \leq \varepsilon \text{ or } t_2 \geq \nu) \\
&= \mathbb{P}(\{s_2 \leq \varepsilon\} \cup \{t_2 \geq \nu\}) \\
&\leq \mathbb{P}(s_2 \leq \varepsilon) + \mathbb{P}(t_2 \geq \nu) \\
&\leq \varepsilon + 2 \exp\left(-\frac{\nu^2}{2}\right) + Dn^{-d},
\end{aligned}$$

for suitable constants $d, D > 0$. By the use of Lemma 18 (applied on the convex combination

$\frac{|\frac{1}{a_2}|^2}{(\sum_{k=2}^n |\frac{1}{a_k}|^2)^{\frac{1}{2}}}, \dots, \frac{|\frac{1}{a_n}|^2}{(\sum_{k=2}^n |\frac{1}{a_k}|^2)^{\frac{1}{2}}}$) we get that

$$\begin{aligned}
\mathbb{P}\left(\sum_{k=2}^n \left|\frac{1}{a_k}\right|^2 \left|\frac{s_k}{t_k}\right|^2 \leq \varepsilon^2 \frac{\sum_{i=2}^n |\frac{1}{a_i}|^2}{\nu^2}\right) &\leq \frac{1}{\sum_{j=2}^n |\frac{1}{a_j}|^2} \sum_{k=2}^n \left|\frac{1}{a_k}\right|^2 \mathbb{P}\left(\left|\frac{s_k}{t_k}\right|^2 \leq \frac{2\varepsilon^2}{\nu^2}\right) \\
&= \mathbb{P}\left(\frac{s_2}{t_2} \leq \frac{2\varepsilon}{\nu}\right) \\
&\leq 2\varepsilon + 2 \exp\left(-\frac{\nu^2}{2}\right) + Dn^{-d}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbb{P}\left(\|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2 \leq \frac{\varepsilon}{\nu} \left(\sum_{k=2}^n \left|\frac{1}{a_k}\right|^2\right)^{\frac{1}{2}}\right) &\leq \mathbb{P}\left(\left(\sum_{m=2}^n \left|\frac{1}{a_m}\right|^2 \left|\frac{s_m}{t_m}\right|^2\right)^{\frac{1}{2}} \leq \frac{\varepsilon}{\nu} \left(\sum_{k=2}^n \left|\frac{1}{a_k}\right|^2\right)^{\frac{1}{2}}\right) \\
&= \mathbb{P}\left(\sum_{m=2}^n \left|\frac{1}{a_m}\right|^2 \left|\frac{s_m}{t_m}\right|^2 \leq \frac{\varepsilon^2}{\nu^2} \sum_{k=2}^n \left|\frac{1}{a_k}\right|^2\right) \\
&\leq 2\varepsilon + 2 \exp\left(-\frac{\nu^2}{2}\right) + Dn^{-d}
\end{aligned}$$

and finally,

$$\begin{aligned}
&\mathbb{P}\left(\min_{\|x\|_2=1} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \frac{u\nu}{\varepsilon} \left(\sum_{k=2}^n \left|\frac{1}{a_k}\right|^2\right)^{-\frac{1}{2}}\right) \\
&\geq \mathbb{P}\left(\|x\|_2 \leq u, \|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2 \geq \frac{\varepsilon}{\nu} \left(\sum_{k=2}^n \left|\frac{1}{a_k}\right|^2\right)^{\frac{1}{2}}\right) \\
&= 1 - \mathbb{P}\left(\left\{\|x\|_2 \leq u, \|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2 \geq \frac{\varepsilon}{\nu} \left(\sum_{k=2}^n \left|\frac{1}{a_k}\right|^2\right)^{\frac{1}{2}}\right\}^c\right) \\
&\geq 1 - 2 \exp\left(-\frac{u^2}{2}\right) - 2\varepsilon - 2 \exp\left(-\frac{\nu^2}{2}\right) - Hn^{-h},
\end{aligned}$$

for suitable constants $h, H > 0$. All in all we have shown that

$$\begin{aligned}
\mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > 2 \frac{u\nu}{\varepsilon} \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}\right) &\leq \exp(-c'n) + 2 \exp(-n) + 2 \exp\left(-\frac{u^2}{2}\right) \\
&\quad + 2 \exp\left(-\frac{\nu^2}{2}\right) + 2\varepsilon + Hn^{-h} + c^n,
\end{aligned}$$

for suitable constants $c', C, H, h > 0$, because $\|(\frac{1}{a_i})_{i \leq n}\|_2 \leq 2\|(\frac{1}{a_i})_{i=2}^n\|_2$. Choose $\varepsilon = \frac{1}{\lambda}$ and $u = t = \sqrt{\ln(\lambda^2)}$ to obtain

$$\begin{aligned} & \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) > 2 \ln(\lambda^2) \lambda \frac{\|(b_i)_{i \leq n}\|_2}{\sqrt{n} \|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \\ & \leq 2 \frac{1}{\lambda} + 2 \exp\left(-\frac{\ln(\lambda^2)}{2}\right) + 2 \exp\left(-\frac{\ln(\lambda^2)}{2}\right) + \exp(-c'n) + 2 \exp(-n) + Hn^{-h} + c^n \\ & \leq \frac{6}{\lambda} + Qn^{-q}, \end{aligned}$$

for suitable constants $Q, q > 0$. An analogue computation finishes the proof. \square

Remark:

1. Assume that $b_1 = \dots = b_n = 1$. Then $\frac{1}{\sqrt{n}} \min\left\{\frac{\|(1)_{i \leq n}\|_2}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, \frac{\|(a_i)_{i \leq n}\|_2}{\|(1)_{i \leq n}\|_2}\right\} = \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}$. That this equality is true follows directly from the following consideration:

$$\frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2} \leq \frac{\|(a_i)_{i \leq n}\|_2}{n} \iff n \leq \|(1)_{i \leq n}\|_2 \|(a_i)_{i \leq n}\|_2.$$

That this equality is always true is an easy conclusion of Hölder's inequality. Therefore, we have that

$$\mathbb{P}\left(s_n((a_i X_{i,j})_{i,j \leq n}) > 2\lambda \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \leq \left(\frac{6}{\lambda}\right) \ln(\lambda^2) + Qn^{-q}.$$

2. In the beginning of the proof of Theorem 11 we used the fact that we have an upper bound on the probability of the event that $(a_i b_j X_{i,j})_{i,j \leq n}$ is singular. We want to give an alternative consideration here which uses the following theorem of A. Sankar, D. A. Spielman and S.-H. Teng that can be found in [32, page 7]

Theorem 12 *Let $n \in \mathbb{N}$. Let A be an arbitrary $n \times n$ matrix, let $(g_{i,j})_{i,j \leq n}$ be an $n \times n$ random matrix whose entries are independent standard normal distributed random variables and let $0 < \sigma \leq 1$. Then*

$$\mathbb{P}\left(s_n(A + (\sigma g_{i,j})_{i,j \leq n}) \leq \frac{\varepsilon}{\sqrt{n}}\right) \leq 2.35 \frac{\varepsilon}{\sigma}.$$

We use Theorem 12 to show that without loss of generality we can think of $(a_i b_j X_{i,j})_{i,j \leq n}$ as an almost surely invertible matrix. Let $0 < \sigma \leq 1$. Let $(\sigma g_{i,j})_{i,j \leq n}$ be an $n \times n$ random matrix chosen independently of $(a_i b_j X_{i,j})_{i,j \leq n}$ whose entries are independent centered Gaussian random variables with variances σ^2 . By Weyl's inequality we know that

$$\begin{aligned} s_n(X) &= s_{(1+n-1)}(-(\sigma g_{i,j})_{i,j \leq n} + (X + (\sigma g_{i,j})_{i,j \leq n})) \\ &\leq \|(\sigma g_{i,j})_{i,j \leq n}\|_{2 \rightarrow 2} + s_n(X + (\sigma g_{i,j})_{i,j \leq n}). \end{aligned}$$

Since stochastic independence is equivalent to the fact that the joint distribution $\mathbb{P}((a_i b_j X_{i,j})_{i,j \leq n}, (\sigma g_{i,j})_{i,j \leq n})$ of $(a_i b_j X_{i,j})_{i,j \leq n}$ and $(\sigma g_{i,j})_{i,j \leq n}$ is equal to the product measure $\mathbb{P}^{(a_i b_j X_{i,j})_{i,j \leq n}} \times \mathbb{P}^{(\sigma g_{i,j})_{i,j \leq n}}$

we obtain by Fubini's theorem

$$\begin{aligned}
& \mathbb{P}(s_n((a_i b_j X_{i,j})_{i,j \leq n} + (\sigma g_{i,j})_{i,j \leq n}) = 0) \\
&= \mathbb{P}((a_i b_j X_{i,j})_{i,j \leq n}, (\sigma g_{i,j})_{i,j \leq n})(s_n((a_i b_j X_{i,j})_{i,j \leq n} + (\sigma g_{i,j})_{i,j \leq n}) = 0) \\
&= \mathbb{P}(a_i b_j X_{i,j})_{i,j \leq n} \times \mathbb{P}(\sigma g_{i,j})_{i,j \leq n}(s_n((a_i b_j X_{i,j})_{i,j \leq n} + (\sigma g_{i,j})_{i,j \leq n}) = 0) \\
&= \int \mathbb{1}_{\{s_n((a_i b_j X_{i,j})_{i,j \leq n} + (\sigma g_{i,j})_{i,j \leq n}) = 0\}} d\mathbb{P}^{(a_i b_j X_{i,j})_{i,j \leq n}} \times \mathbb{P}^{(\sigma g_{i,j})_{i,j \leq n}} \\
&= \int \underbrace{\mathbb{P}^{(\sigma g_{i,j})_{i,j \leq n}}(s_n((a_i b_j X_{i,j})_{i,j \leq n} + (\sigma g_{i,j})_{i,j \leq n}) = 0)}_{\text{Theorem 12}_0} d\mathbb{P}^{(a_i b_j X_{i,j})_{i,j \leq n}} \\
&= 0.
\end{aligned}$$

Thus the matrix $X + (\sigma g_{i,j})_{i,j \leq n}$ is almost surely invertible. Therefore, by choosing σ very small we can consider without loss of generality that X is invertible almost everywhere approximating a renormalization of the variance and a slightly larger bound for the fourth moments.

3. In the proof of Theorem 11 we used Berry-Esséen's theorem to bound the probability

$$\mathbb{P}(\|x\|_2 > u) = \int \mathbb{P}^{X_1} \left(\left| \sum_{k \leq n} N_{1,k} X_{1,k} \right| > u \right) d\mathbb{P}^{N_1}.$$

At this point one could have also used the approximate rotation invariance (see [46, page 12]) of subgaussian variables to obtain that the finite sum of subgaussian variables is a subgaussian variable. Than one can use the tail bound for subgaussian variables to come to similar results.

4. In [27] M. Rudelson and R. Vershynin used a least common denominator argument instead of our non-degeneracy argument (Lemma 17) to bound $\mathbb{P}(|\sum_{k \leq n} \hat{Y}_{2,k}^* X_{1,k}| \leq \varepsilon)$. Therefore they could improve the polynomial term Qn^{-q} to an exponential term c^n for $c \in (0, 1)$. We want to mark that the argumentation of [27] works in our situation, too. We just wanted to give an alternative, faster access to deal with this problem. The reader interested in the least common denominator technique is referred to [27] and [28].

4.2 Upper bound for rectangular matrices

The following crucial lemma can be found in [43] on page 30 (Lemma B.3). For the reader's convenience we give a short proof.

Lemma 19 (Projection lemma - Tao, Vu) *Let $1 \leq s \leq n$ be integers, let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ be the real or complex field, let A be an $n \times n$ \mathbb{K} -valued invertible matrix with columns X_1, \dots, X_n and let X_1^*, \dots, X_n^* denote the rows of A^{-1} . Let B be the $s \times n$ matrix with rows X_1^*, \dots, X_s^* . Let $V = \text{span}\{X_{s+1}, \dots, X_n\}^\perp$ be the orthogonal complement of the space spanned by X_{s+1}, \dots, X_n in \mathbb{K}^n . This space has dimension s and can be identified with \mathbb{K}^s via an orthonormal basis. Let*

$$P_V : \mathbb{K}^n \rightarrow \mathbb{K}^n \text{ be the orthogonal projection onto } V \equiv \mathbb{K}^s, j : \mathbb{K}^n \rightarrow V, j(x) = \begin{cases} x, & x \in V \\ 0, & x \in V^\perp. \end{cases}$$

Furthermore, define $\pi = j \circ P_V$. Let M be the $s \times s$ matrix with columns $\pi(X_1), \dots, \pi(X_s)$. Then M is invertible, and we have

$$BB^* = M^{-1}(M^{-1})^*.$$

In particular, we have

$$s_j(B) = (s_{s-j+1}(M))^{-1}$$

for all $1 \leq j \leq s$.

Proof: By construction, we have $\langle X_i^*, X_j \rangle = \delta_{ij}$ for all $1 \leq i, j \leq n$. In particular, since $\langle X_k^*, X_l \rangle = 0$ for $l \in \{s+1, \dots, n\}$ and $k \in \{1, \dots, s\}$ we know that X_1^*, \dots, X_s^* lie in V , and thus,

$$\begin{aligned} \langle \pi(X_i^*), \pi(X_j) \rangle &= \langle jP_V(X_i^*), jP_V(X_j) \rangle \\ &= \langle (jP_V)^T jP_V(X_i^*), X_j \rangle \\ &= \langle P_V^T j^T jP_V(X_i^*), X_j \rangle \\ &= \langle P_V^2(X_i^*), X_j \rangle \\ &= \langle X_i^*, X_j \rangle \\ &= \delta_{ij} \end{aligned}$$

for $1 \leq i, j \leq s$. That means $(\pi(X_i^*), \pi(X_j))_{i,j=1}^s$ is a biorthogonal system. Thus by Lemma 3 M is invertible, and the rows of M^{-1} are given by $\pi(X_1^*), \dots, \pi(X_s^*)$. Thus the ij -th entry of $(M^{-1})(M^{-1})^*$ is given by $\langle \pi(X_i^*), \pi(X_j^*) \rangle$, while the ij -th entry of BB^* is given by $\langle X_i^*, X_j^* \rangle$. Since X_1^*, \dots, X_s^* lie in V , and π is an isometry on V the claim follows. \square

Theorem 13 *Let $n \in \mathbb{N}$. Let $1 \leq k \leq n$. Let $b_1, \dots, b_{n-k} \in (0, 1]$ be a sequence of real numbers and $g_{i,j}$ ($i = 1, \dots, n$ and $j = 1, \dots, n-k$) be i.i.d. standard normal distributed random variables. Define G as the $n \times (n-k)$ random matrix $G = (b_j g_{i,j})_{i \leq n, j \leq n-k}$. Then*

$$\mathbb{P}\left(s_{n-k}(G) > 2\lambda \frac{1}{\|(\frac{1}{b_j})_{j \leq n-k}\|_2}\right) \leq \frac{2(1 + 2^{\frac{k+1}{2}})}{\lambda} \ln(\lambda^4).$$

Proof: W.l.o.g. we assume that $1 = b_1 \geq \dots \geq b_{n-k}$. Let $A = (\bar{b}_j g_{i,j})_{i,j \leq n}$, where $\bar{b}_j := \begin{cases} b_j & 1 \leq j \leq n-k \\ 1 & n-k+1 \leq j \leq n \end{cases}$ and $g_{i,j}$ ($i, j \leq n$) are independent standard normal distributed random variables. We denote by $\bar{b}_1 X_1, \dots, \bar{b}_n X_n$ the columns of A , respectively the rows of A^T . We define G as the $n \times (n-k)$ random matrix that we obtain by taking the first $(n-k)$ rows $b_1 X_1, \dots, b_{n-k} X_{n-k}$ of A^T . By Lemma 4 and Theorem 8 we know that A^T is almost surely invertible as a tensor matrix of an square Gaussian matrix. Therefore, w.l.o.g. we can assume that A is an invertible matrix. Denote the columns of $(A^T)^{-1}$ by $\frac{1}{b_1} X_1^*, \dots, \frac{1}{b_n} X_n^*$ and apply the projection lemma on $(A^T)^{-1}$ to obtain

$$s_1(M)^{-1} = s_{n-k}(G) = s_{n-k}(G^T),$$

where we used equality 2.1 and M is the invertible $(n-k) \times (n-k)$ matrix from Lemma 19 with columns $\pi\left(\frac{1}{b_1} X_1^*\right), \dots, \pi\left(\frac{1}{b_{n-k}} X_{n-k}^*\right)$ where $\pi = j \circ P_V : \mathbb{R}^n \rightarrow V$ denotes the map from Lemma 19 onto $V = \text{span}\{X_{n-k+1}^*, \dots, X_n^*\}^\perp$. Let $u, v > 0$. Analogue to the proof of Theorem 11 we are looking for a suitable $x \in \mathbb{R}^n$ such that $\|x\|_2 \leq u$ and $\|Mx\|_2 \geq \frac{v}{\|(\frac{1}{b_j})_{j \leq n-k}\|_2}$. Then

$$s_{n-k}(B) = \frac{1}{s_1(M)} = \frac{1}{\|M\|_{2 \rightarrow 2}} = \frac{1}{\|M^T\|_{2 \rightarrow 2}} \leq \frac{u}{v \|(\frac{1}{b_j})_{j \leq n-k}\|_2}.$$

Define P_1^{n-k} as the orthogonal projection onto $H_1^{n-k} = \text{span}\{b_i X_i \mid i \in \{2, \dots, n-k\}\} = \text{span}\{X_i \mid i \in \{2, \dots, n-k\}\}$. We set $x = \pi(b_1 X_1 - P_1^{n-k} b_1 X_1) = \pi(X_1 - P_1^{n-k} X_1)$ and obtain since π is an

isometry on V and by the use of Lemma 6

$$\begin{aligned}
\|x\|_2 &= \|\pi(X_1 - P_1^{n-k} X_1)\|_2 \\
&\leq \|\pi\|_{2 \rightarrow 2} \|X_1 - P_1^{n-k} X_1\|_2 \\
&= \|X_1 - P_1^{n-k} X_1\|_2 \\
&= \inf_{Y \in H_1^{n-k}} \|X_1 - Y\|_2 \\
&= \text{dist}(X_1, H_1^{n-k}) \\
&\stackrel{\cong}{=} \|(g_i)_{i \leq k+1}\|_2,
\end{aligned}$$

where the last equality is again an equality in distribution and the g_1, \dots, g_{k+1} are independent standard normal distributed Gaussian random variables. Therefore we get for all $u > 0$

$$\begin{aligned}
\mathbb{P}(\|x\|_2 > u) &\leq \mathbb{P}(\|(g_i)_{i \leq k+1}\|_2 > u) \\
&= (2\pi)^{-\frac{k+1}{2}} \int_{\{\sum_{i=1}^{k+1} x_i^2 > u^2\}} \exp\left(-\frac{\sum_{i=1}^{k+1} x_i^2}{2}\right) dx_1 \dots dx_{k+1} \\
&\leq (2\pi)^{-\frac{k+1}{2}} \exp\left(-\frac{u^2}{4}\right) \int \exp\left(-\frac{\sum_{i=1}^{k+1} x_i^2}{4}\right) dx_1 \dots dx_{k+1} \\
&= (2\pi)^{-\frac{k+1}{2}} \exp\left(-\frac{u^2}{4}\right) \left(\int \exp\left(-\frac{x^2}{4}\right) dx\right)^{k+1} \\
&= (2\pi)^{-\frac{k+1}{2}} \exp\left(-\frac{u^2}{4}\right) (2\pi)^{\frac{k+1}{2}} 2^{\frac{k+1}{2}} \\
&= 2^{\frac{k+1}{2}} \exp\left(-\frac{u^2}{4}\right).
\end{aligned}$$

Analogue to the proof of Theorem 11 we need to estimate $\|M^T x\|_2$ from below:

Since $\langle X_i^*, X_j \rangle = \delta_{i,j}$ for all $i, j \in \{1, \dots, n\}$ we know that $(X_j)_{j \leq n}$ is a linearly independent system of vectors and thus it follows that $V = \text{span}\{X_1, \dots, X_{n-k}\}$. Therefore we obtain

$$\begin{aligned}
\langle M e_i, \pi(X_1) \rangle &\stackrel{\text{Lemma 4}}{=} \left\langle \pi\left(\frac{1}{b_i} X_i^*\right), \pi(X_1) \right\rangle \\
&= \frac{1}{b_i} \langle \pi(X_i^*), \pi(X_1) \rangle \\
&= \frac{1}{b_i} \langle X_i^*, \pi^T \pi X_1 \rangle \\
&= \frac{1}{b_i} \langle X_i^*, P_V^T j^T j P_V X_1 \rangle \\
&= \frac{1}{b_i} \langle X_i^*, P_V^2 X_1 \rangle \\
&= \frac{1}{b_i} \langle X_i^*, X_1 \rangle \\
&= \delta_{i,1},
\end{aligned}$$

for all $i \in \{1, \dots, n-k\}$. Consequently,

$$\begin{aligned}
\|M^T x\|_2^2 &= \|M^T \pi(X_1 - P_1^{n-k} X_1)\|_2^2 = \|M^T \pi(X_1) - M^T \pi(P_1^{n-k} X_1)\|_2^2 \\
&= \|e_1 - M^T \underbrace{\pi(P_1^{n-k} X_1)}_{\in \mathbb{R}^n}\|_2^2.
\end{aligned}$$

As $\pi(\underbrace{P_1^{n-k} X_1}_{\in H_1^{n-k}}) \in \text{span}\{\pi(X_i) \mid i \in \{2, \dots, n-k\}\}$ we get that

$$\begin{aligned}
\langle e_1, M^T \pi(P_1^{n-k} X_1) \rangle &= \langle M e_1, \pi(P_1^{n-k} X_1) \rangle \\
&= \left\langle \pi(X_1^*), \pi(P_1^{n-k} X_1) \right\rangle \\
&= \langle \pi(X_1^*), \sum_{j=2}^{n-k} \lambda_j \pi(X_j) \rangle \\
&= \sum_{j=2}^{n-k} \lambda_j \langle \pi(X_1^*), \pi(X_j) \rangle \\
&= \sum_{j=2}^{n-k} \lambda_j \langle X_1^*, \pi^T \pi(X_j) \rangle
\end{aligned}$$

and with analogue computations as above we have that

$$\begin{aligned}
\langle e_1, M^T \pi(P_1^{n-k} X_1) \rangle &= \sum_{j=2}^{n-k} \langle X_1^*, X_j \rangle \\
&= 0,
\end{aligned}$$

where $\sum_{j=2}^{n-k} \lambda_j \pi(X_j) = \pi(P_1^{n-k} X_1)$ where $\lambda_j \in \mathbb{R}$. By the Pythagorean theorem we get that

$$\begin{aligned}
\|M^T x\|_2^2 &= \|e_1\|_2^2 + \|M^T \pi(P_1^{n-k} X_1)\|_2^2 \\
&> \|M^T \pi(P_1^{n-k} X_1)\|_2^2 \\
&= \sum_{i \leq n-k} |\langle M^T \pi(P_1^{n-k} X_1), e_i \rangle|^2 \\
&= \sum_{i \leq n-k} |\langle P_1^{n-k} X_1, \pi^T M e_i \rangle|^2.
\end{aligned}$$

Since $M e_i = \pi\left(\frac{1}{b_i} X_i^*\right)$ we obtain

$$\begin{aligned}
\|M^T x\|_2^2 &> \sum_{i \leq n-k} \left| \left\langle P_1^{n-k} X_1, \pi^T \pi\left(\frac{1}{b_i} X_i^*\right) \right\rangle \right|^2 \\
&= \sum_{i \leq n-k} \left| \left\langle \pi^T \pi \underbrace{P_1^{n-k} X_1}_{\in H_1^{n-k}}, \frac{1}{b_i} X_i^* \right\rangle \right|^2.
\end{aligned}$$

Since $H_1^{n-k} = \text{span}\{X_j \mid j \in \{2, \dots, n-k\}\}$ we can write

$$P_1^{n-k} X_1 = \sum_{j=2}^{n-k} \lambda_j X_j,$$

where $\lambda_j \in \mathbb{R}$. It follows that

$$\begin{aligned}
\|M^T x\|_2^2 &> \sum_{i \leq n-k} \left| \left\langle \pi^T \pi \underbrace{\sum_{j=2}^{n-k} \lambda_j X_j, \frac{1}{b_i} X_i^*}_{\in V} \right\rangle \right|^2 \\
&= \sum_{i \leq n-k} \left| \left\langle \sum_{j=2}^{n-k} \lambda_j X_j, \frac{1}{b_i} X_i^* \right\rangle \right|^2 \\
&= \sum_{i \leq n-k} \left| \frac{1}{b_i} \right|^2 | \langle P_1^{n-k} X_1, X_i^* \rangle |^2 \\
&= \sum_{i \leq n-k} \left| \frac{1}{b_i} \right|^2 | \langle X_1, P_1^{n-k} X_i^* \rangle |^2,
\end{aligned}$$

where we used the fact that $V = \text{span}\{X_1, \dots, X_{n-k}\}$. Since $X_1^* \in (H_1^{n-k})^\perp$ it follows that $X_1^* \in \ker(P_1^{n-k})$ and therefore

$$\|M^T x\|_2^2 > \sum_{i=2}^{n-k} \left| \frac{1}{b_i} \right|^2 | \langle X_1, P_1^{n-k} X_i^* \rangle |^2$$

For $k = 2, \dots, n-k$ we denote $Y_k^* = P_1^{n-k} X_i^* \in H_1^{n-k}$. Now, similar to the proof above we have to check whether the system $(Y_i^*, X_i)_{i=2}^{n-k}$ is a biorthogonal system:

As

$$P_1^{n-k}(Y_i^* - X_i^*) = P_1^{n-k}(Y_i^*) - P_1^{n-k}(X_i^*) = P_1^{n-k}(P_1^{n-k} X_i^*) - P_1^{n-k}(X_i^*) = 0$$

we get that

$$Y_i^* - X_i^* \in \ker(P_1^{n-k}) = \text{span}\{X_i^* \mid i = \underbrace{\{1, n-k+1, \dots, n\}}_{=Q}\}.$$

It follows that $Y_i^* = X_i^* - \sum_{j \in Q} \lambda_j X_j^*$ where $\lambda_j \in \mathbb{R}$ for every $j \in Q$. We use the fact $\langle X_j^*, X_i \rangle = 0$ for all $j \in Q$, $i \in \{2, \dots, n-k\}$ to obtain

$$\langle Y_m^*, X_l \rangle = \langle X_m^* - \sum_{q \in Q} \lambda_q X_q^*, X_l \rangle = \langle X_m^*, X_l \rangle - \sum_{q \in Q} \lambda_q \langle X_q^*, X_l \rangle = \langle X_m^*, X_l \rangle = \delta_{m,l}$$

for all $m, l \in \{2, \dots, n-k\}$. Therefore $(Y_i^*, X_i)_{i=2}^{n-k}$ is a biorthogonal system in $H_1^{n-k} = \text{span}\{X_l \mid l \in \{2, \dots, n-k\}\}$. Furthermore by Lemma 3 this system is unique, where we once again want to point out that $(X_i^*, X_i)_{i=2}^{n-k}$ is not contained in H_1^{n-k} , and the vectors $(Y_i^*)_{i=2}^{n-k}$ are uniquely determined by the system $(X_i)_{i=2}^{n-k}$. In particular, since every Y_j^* ($j \in \{2, \dots, n-k\}$) is a measurable function of the vectors X_2, \dots, X_{n-k} we obtain that the system $(Y_j^*)_{j=2}^{n-k}$ is stochastically independent to the vectors $X_1, X_{n-k+1}, \dots, X_n$. We define $H_{1,j}^{n-k} = \text{span}\{X_i \mid i \notin \{1, j, n-k+1, \dots, n\}\}$. Then by Lemma 3 we get that $\|Y_j^*\|_2 = \frac{1}{\text{dist}(X_j, H_{1,j}^{n-k})}$ for $j \in \{2, \dots, n-k\}$, where the related Hilbert space is now H_1^{n-k} .

For $j \in \{2, \dots, n-k\}$ set $q_j = \left| \left\langle \frac{Y_j^*}{\|Y_j^*\|_2}, X_1 \right\rangle \right|$ and $p_j = \|Y_j^*\|_2 = \frac{1}{\text{dist}(X_j, H_{1,j}^{n-k})}$. Then we can write

$$\|M^T x\|_2^2 > \sum_{j=2}^{n-k} \left| \frac{1}{b_j} \right|^2 \left| \frac{q_j}{p_j} \right|^2.$$

W.l.o.g. we assume $j = 2$. Then by an analogue computation as for $\|x\|_2$ we get

$$\mathbb{P}(p_2 > t) = \mathbb{P}(\text{dist}(X_2, H_{1,2}^{n-k}) > t) \leq 2^{\frac{k+1}{2}} \exp\left(-\frac{t^2}{4}\right).$$

Furthermore, since $\hat{Y}_2^* = \frac{Y_2^*}{\|Y_2^*\|_2} \in (H_{1,2}^{n-k})^\perp$ is stochastically independent to X_1 we have that

$$\begin{aligned} \mathbb{P}(q_2 \leq \varepsilon) &= \mathbb{P}^{(\hat{Y}_2^*, X_1)}\left(\left|\sum_{i=1}^n \hat{Y}_2^*(i) X_1(i)\right| \leq \varepsilon\right) \\ &= \mathbb{P}^{\hat{Y}_2^*} \times \mathbb{P}^{X_1}\left(\left|\sum_{i=1}^n \hat{Y}_2^*(i) X_1(i)\right| \leq \varepsilon\right) \\ &= \int \mathbf{1}_{\{|\sum_{i=1}^n \hat{Y}_2^*(i) X_1(i)| \leq \varepsilon\}} d\mathbb{P}^{\hat{Y}_2^*} \times \mathbb{P}^{X_1} \end{aligned}$$

Thus we get by Fubini's theorem and the rotational invariance of the Gaussian distribution

$$\begin{aligned} \mathbb{P}(q_2 \leq \varepsilon) &= \int \int \mathbf{1}_{\{|\sum_{i=1}^n \hat{Y}_2^*(i) X_1(i)| \leq \varepsilon\}} d\mathbb{P}^{X_1} d\mathbb{P}^{\hat{Y}_2^*} \\ &= \int \mathbb{P}^{X_1}\left(\left|\sum_{i=1}^n \hat{Y}_2^*(i) X_1(i)\right| \leq \varepsilon\right) d\mathbb{P}^{\hat{Y}_2^*} \\ &= \int \mathbb{P}(|g| \leq \varepsilon) d\mathbb{P}^{\hat{Y}_2^*} \\ &\leq \mathbb{P}(|g| \leq \varepsilon) \\ &\leq \varepsilon. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{P}\left(\frac{q_j}{p_j} \leq \frac{\varepsilon}{t}\right) &\leq \mathbb{P}(q_j \leq \varepsilon \text{ or } p_j \geq t) \\ &= \mathbb{P}(\{q_j \leq \varepsilon\} \cup \{p_j \geq t\}) \\ &\leq \mathbb{P}(q_j \leq \varepsilon) + \mathbb{P}(p_j \geq t) \\ &\leq \varepsilon + 2^{\frac{k+1}{2}} \exp\left(-\frac{t^2}{4}\right). \end{aligned}$$

We use Lemma 18 for the convex combination $\frac{|\frac{1}{b_2}|^2}{\sum_{j=2}^{n-k} |\frac{1}{b_j}|^2}, \dots, \frac{|\frac{1}{b_{n-k}}|^2}{\sum_{j=2}^{n-k} |\frac{1}{b_j}|^2}$ and obtain

$$\begin{aligned} \mathbb{P}\left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2 \left(\frac{q_j}{p_j}\right)^2 \leq \frac{\varepsilon \sum_{j=2}^{n-k} |\frac{1}{b_j}|^2}{t}\right) &\leq \frac{1}{\sum_{j=2}^{n-k} |\frac{1}{b_j}|^2} \sum_{i=2}^{n-k} \left|\frac{1}{b_i}\right|^2 \mathbb{P}\left(\frac{q_i}{p_i} \leq \frac{2\varepsilon}{t}\right) \\ &= \mathbb{P}\left(\frac{q_2}{p_2} \leq \frac{2\varepsilon}{t}\right) \\ &\leq 2\varepsilon + 2^{\frac{k+1}{2}} \exp\left(-\frac{t^2}{4}\right). \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{P}\left(\|M^{-1}x\|_2 \leq \frac{\varepsilon}{t} \left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2\right)^{\frac{1}{2}}\right) &\leq \mathbb{P}\left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2 \left(\frac{q_j}{p_j}\right)^2 \leq \frac{\varepsilon^2 \sum_{j=2}^{n-k} |\frac{1}{b_j}|^2}{t^2}\right) \\ &\leq 2\varepsilon + 2^{\frac{k+1}{2}} \exp\left(-\frac{t^2}{4}\right). \end{aligned}$$

Finally,

$$\begin{aligned}
\mathbb{P}\left(s_{n-k}(B) \leq \frac{ut}{\varepsilon} \left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2\right)^{\frac{1}{2}}\right) &\geq \mathbb{P}\left(\|x\|_2 \leq u, \|M^{-1}x\|_2 \geq \frac{\varepsilon}{t} \left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2\right)^{\frac{1}{2}}\right) \\
&= 1 - \mathbb{P}\left(\left\{\|x\|_2 \leq u, \|M^{-1}x\|_2 \geq \frac{\varepsilon}{t} \left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2\right)^{\frac{1}{2}}\right\}^c\right) \\
&\geq 1 - \mathbb{P}\left(\left\{\|x\|_2 > u\right\} \cup \left\{\|M^{-1}x\|_2 \leq \frac{\varepsilon}{t} \left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2\right)^{\frac{1}{2}}\right\}\right) \\
&\geq 1 - 2^{\frac{k+1}{2}} \exp\left(-\frac{u^2}{4}\right) - 2^{\frac{k+1}{2}} \exp\left(-\frac{t^2}{4}\right) - 2\varepsilon.
\end{aligned}$$

Choosing $\varepsilon = \frac{1}{\lambda}$, $u = t = \sqrt{\ln(\lambda^4)}$ we get by an analogue computation as in the proof of Theorem 11:

$$\begin{aligned}
\mathbb{P}\left(s_{n-k}(B) > 2\lambda \ln(\lambda^4) \left(\sum_{j=1}^{n-k} \left|\frac{1}{b_j}\right|^2\right)^{\frac{1}{2}}\right) &\leq \mathbb{P}\left(s_{n-k}(B) > \lambda \ln(\lambda^4) \left(\sum_{j=2}^{n-k} \left|\frac{1}{b_j}\right|^2\right)^{\frac{1}{2}}\right) \\
&\leq \frac{2(1 + 2^{\frac{k+1}{2}})}{\lambda}.
\end{aligned}$$

□

An easy application of Theorem 13 is the following corollary:

Corollary 3 *Let $n \geq k \geq 0$ be natural numbers. Let $b_1, \dots, b_{n-k} \in (0, 1]$ and $a_{1,n-k+1}, \dots, a_{1,n}, \dots, a_{n,n-k+1}, \dots, a_{n,n}$ be sequences of real numbers. We can assume w.l.o.g. that $0 < b_1 \leq \dots \leq b_{n-k}$. Let $g_{i,j}$ ($1 \leq i \leq n$, $1 \leq j \leq n-k$) be i.i.d. standard normal distributed random variables. Then for arbitrary random variables $X_{i,j}$ ($1 \leq i \leq n$, $n-k+1 \leq j \leq n$) we define the $n \times n$ random matrix*

$$X = \begin{pmatrix} b_1 g_{1,1} & \dots & b_{n-k} g_{1,n-k} & a_{1,n-k+1} X_{1,n-k+1} & \dots & a_{1,n} X_{1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ b_1 g_{n,1} & \dots & b_{n-k} g_{n,n-k} & a_{n,n-k+1} X_{n,n-k+1} & \dots & a_{n,n} X_{n,n} \end{pmatrix}.$$

Then

$$\mathbb{P}\left(s_n(X) > 2\lambda \frac{1}{\|(\frac{1}{b_i})_{i \leq n-k}\|_2}\right) \leq \frac{2(1 + 2^{\frac{k+1}{2}})}{\lambda} \ln(\lambda^4).$$

Proof: We define the $(n-k) \times n$ matrix

$$G^T = \begin{pmatrix} b_1 g_{1,1} & \dots & b_1 g_{n,1} \\ \dots & \dots & \dots \\ b_{n-k} g_{1,n-k} & \dots & b_{n-k} g_{n,n-k} \end{pmatrix}$$

By the Cauchy interlacing theorem (Theorem 3) we know that

$$s_n(X) = s_n(X^T) \leq s_{n-k}(G^T) = s_{n-k}(G).$$

It follows that

$$\begin{aligned}\mathbb{P}\left(s_n(X) > \lambda \frac{1}{\|(\frac{1}{b_i})_{i \leq n-k}\|_2}\right) &\leq \mathbb{P}\left(s_{n-k}(G) > \lambda \frac{1}{\|(\frac{1}{b_i})_{i \leq n-k}\|_2}\right) \\ &\leq \frac{2(1 + 2^{\frac{k+1}{2}})}{\lambda} \ln(\lambda^4).\end{aligned}$$

□

Chapter 5

Lower bound for the least singular value

In this chapter we want to study probabilistic estimates to bound the smallest singular value from below. We start this in section 5.1 by looking at random matrices with normal distributed entries, more precisely we investigate matrices $(a_j g_{i,j})_{i,j \leq n}$, where $a_1, \dots, a_n \in \mathbb{R} \setminus \{0\}$ and the $g_{i,j}$ ($i, j \in \{1, \dots, n\}$) are i.i.d. standard normal distributed random variables. We will show in the sequel that the probability that

$$s_n((a_j g_{i,j})_{i,j \leq n}) < \frac{\varepsilon}{\|(\frac{1}{a_j})_{j \leq n}\|_2}$$

is small. This bound coincides with the bound we found in chapter 3. In section 5.2 we generalize this estimate to the class of subgaussian random variables. In Theorem 17 we show that under some assumptions specified later the probability of

$$s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\| \|(\frac{1}{b_j})_{j \leq n}\|_2}$$

is smaller or equal than $C\varepsilon + c^n$ for a constant $C > 0$ and $c \in (0, 1)$ plus a term depending on the $\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2$ and $\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n}^T x\|_2$. This leads to our investigations in section 5.3, where we will show that although compressible vectors do not have a great impact on the smallest singular value in the classical setting (see section 3.2), they can play a major role for tensor matrices. The main ingredient here is the so called *invertibility via distance* lemma which was taught to us by M. Rudelson and R. Vershynin in [28, page 16 f]. We will illustrate this by a simple example.

5.1 Estimate for Gaussian random matrices

As mentioned above, in this section we will study the smallest singular value of $n \times n$ random matrices $X = (a_j g_{i,j})_{i,j \leq n}$, where the $g_{i,j}$ are i.i.d. standard normal distributed random variables and a_1, \dots, a_n are reals that do not equal zero. We will need the fact that every entry of the i -th column vector of X has the same variance so we can use the rotational invariance of the Gaussian distribution. This is the main tool in our proof of Theorem 14. But before we can start, we need the following lemma which can be found in [32, page 34, Lemma B1]

Lemma 20 *Let $n \geq 2$ be an integer and let $\xi = (\xi_1, \dots, \xi_n)$ be a random unit vector, which is uniformly distributed on the Euclidean sphere. Then for $0 < c \leq 1$,*

$$\mathbb{P}\left(|\xi_1| \geq \sqrt{\frac{c}{n}}\right) \geq \mathbb{P}(|g| \geq \sqrt{c}),$$

where g is a standard normal distributed random variable.

Proof: From Lemma 5 we know that we can get a random unit vector $\xi = (\xi_1, \dots, \xi_n)$ which is uniformly distributed on the n -dimensional Euclidean sphere by choosing n i.i.d. standard normal distributed random variables g_1, \dots, g_n and setting $\xi_i = \frac{g_i}{\|(g_i)_{i \leq n}\|_2}$, $i = 1, \dots, n$. It follows

$$\begin{aligned} \mathbb{P}\left(|\xi_1|^2 \geq \frac{c}{n}\right) &= \mathbb{P}\left(\frac{|g_1|^2}{\|(g_i)_{i \leq n}\|_2^2} \geq \frac{c}{n}\right) \\ &= \mathbb{P}\left(\frac{(n-1)|g_1|^2}{\|(g_i)_{i \leq n}\|_2^2} \geq \frac{(n-1)c}{n}\right) \\ &\geq \mathbb{P}\left(\frac{(n-1)|g_1|^2}{\|(g_i)_{i \leq n}\|_2^2} \geq \frac{(n-1)c}{n-1}\right) \\ &= \mathbb{P}\left(\frac{(n-1)|g_1|^2}{\|(g_i)_{i \leq n}\|_2^2} \geq c\right). \end{aligned}$$

We want point out that

$$t_n = \frac{\sqrt{(n-1)}g_1}{\|(g_i)_{i \leq n}\|_2}$$

is a random variable distributed to the t -distribution with $n-1$ degrees of freedom. Let g be a standard normal distributed random variable. Since t_n and g are symmetric about the origin and by the use of the inequality

$$\mathbb{P}(t_n > \sqrt{c}) \geq \mathbb{P}(g > \sqrt{c}), \text{ for all } 0 < c \leq 1$$

which can be found in [12, Chapter 28, section 2] we get the stated inequality. \square

Theorem 14 Let $n \in \mathbb{N}$. Let $X = (a_j g_{i,j})_{i,j \leq n}$ be an $n \times n$ random matrix, where $a_1, \dots, a_n \in \mathbb{R} \setminus \{0\}$ and the $g_{i,j}$ are i.i.d. standard normal distributed random variables. Then for all $\varepsilon > 0$

$$\mathbb{P}\left(s_n(X) \leq \frac{\varepsilon}{\|(\frac{1}{a_j})_{j \leq n}\|_2}\right) \leq 2.35\varepsilon.$$

Proof: As in the proof of Theorem 9 we know by Theorem 8 that the probability that $s_n(X) = 0$ is equal to zero and therefore, we get by analogue considerations and Lemma 4 that X is almost surely invertible. In the sequel we will always assume that X is invertible. The proof consists of two steps. In step one we estimate the probability $\mathbb{P}\left(\|X^{-1}x\|_2 > \frac{1}{\varepsilon}\right)$ for an arbitrary fixed vector in the Euclidean sphere S^{n-1} . In the second step we show that the estimate from step one also holds for a uniform distributed random unit vector and use this result to derive an estimate for

$$\begin{aligned} \mathbb{P}\left(s_n(X) \leq \frac{\varepsilon}{\|(\frac{1}{a_j})_{j \leq n}\|_2}\right) &= \mathbb{P}\left(\min_{x \in S^{n-1}} \|Xx\|_2 \leq \frac{\varepsilon}{\|(\frac{1}{a_j})_{j \leq n}\|_2}\right) \\ &= \mathbb{P}\left(\frac{1}{\max_{x \in S^{n-1}} \|X^{-1}x\|_2} \leq \frac{\varepsilon}{\|(\frac{1}{a_j})_{j \leq n}\|_2}\right). \end{aligned}$$

Step 1: Let $x \in S^{n-1}$ be an arbitrary vector in the sphere and let Q be the orthonormal $n \times n$ matrix, i.e. $Q^T Q = \mathbf{1}_{n \times n}$, such that $Q^T e_1 = x$. Then we get

$$\begin{aligned} \|(a_j g_{i,j \leq n})_{i,j}^{-1} x\|_2 &= \|(a_j g_{i,j})_{i,j \leq n}^{-1} Q^T e_1\|_2 \\ &= \|((g_{i,j})_{i,j \leq n} \text{diag}(a_1, \dots, a_n))^{-1} Q^T e_1\|_2 \\ &= \|(Q(g_{i,j})_{i,j \leq n} \text{diag}(a_1, \dots, a_n))^{-1} e_1\|_2 \\ &\stackrel{\mathfrak{D}}{=} \|(a_j g_{i,j})_{i,j \leq n}^{-1} e_1\|_2, \end{aligned}$$

where we used the rotational invariance of the Gaussian distribution to obtain that $Q(g_{i,j})_{i,j \leq n}$ and $(g_{i,j})_{i,j \leq n}$ have the same distribution. In the sequel X_i^* denotes the i -th row and $X_{i,j}^*$ the j -th entry in the i -th row of $(g_{i,j})_{i,j \leq n}^{-1}$. Consequently, by Lemma 4 we have the following equality in distribution:

$$\|(a_j g_{i,j})_{i,j}^{-1} x\|_2 \stackrel{\mathcal{D}}{=} \left\| \left(\frac{1}{a_i} X_{i,j}^* \right)_{i,j} e_1 \right\|_2 = \left\| \left(\frac{1}{a_i} X_{i,1}^* \right)_{i \leq n} \right\|_2.$$

We apply Lemma 3 to the biorthogonal system $((X_{i,j})_{j \leq n}, (X_{j,k}^*)_{j \leq n})$ to get

$$\|(a_j g_{i,j})_{i,j}^{-1} x\|_2 \stackrel{\mathcal{D}}{=} \frac{1}{\text{dist}((a_j g_{1,j})_{j \leq n}, \text{span}\{(a_j g_{i,j})_{j \leq n} \mid i = 2, \dots, n\})},$$

Analogue to the computations in the proof of Lemma 3 we obtain

$$\|(a_j g_{i,j})_{i,j \leq n}^{-1} x\|_2 \stackrel{\mathcal{D}}{=} \frac{1}{|\langle \bar{N}_1, (a_j g_{1,j})_{j \leq n} \rangle|},$$

where \bar{N}_1 is a random normal on the (random) hyperplane $\text{span}\{(a_j g_{i,j})_{j \leq n} \mid i = 2, \dots, n\}$, i.e. $\bar{N}_1 \in \text{span}\{(a_j g_{i,j})_{j \leq n} \mid i = 2, \dots, n\}^\perp$ and $\|\bar{N}_1\|_2 = 1$. Let N_1 be a random normal on $H_1 = \text{span}\{(g_{i,j})_{j \leq n} \mid i = 2, \dots, n\}$. Since

$$\left\langle \text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1, \left(a_j g_{k,j}\right)_{j \leq n} \right\rangle = \langle N_1, (g_{k,j})_{j \leq n} \rangle = 0$$

for all $k = 2, \dots, n$ it follows that $\text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1 \in \text{span}\{(a_j g_{i,j})_{j \leq n} \mid i = 2, \dots, n\}^\perp$. Therefore, since $\text{span}\{(a_j g_{i,j})_{j \leq n} \mid i = 2, \dots, n\}$ is almost surely a hyperplane, we get that

$$\begin{aligned} \frac{1}{|\langle \bar{N}_1, (a_j g_{1,j})_{j \leq n} \rangle|} &= \frac{\|\text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1\|_2}{|\langle \text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1, (a_j g_{1,j})_{j \leq n} \rangle|} \\ &= \frac{\|\text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1\|_2}{|\langle N_1, (g_{1,j})_{j \leq n} \rangle|} \end{aligned}$$

almost surely.

We want to use the fact that N_1 is uniformly distributed on the Euclidean sphere. To see that this is true, it is enough to show that the distribution of X_1^* is rotational invariant, because $N_1 \in \text{span}\{X_1^*\}$ and the rotational invariant measure is unique on the Euclidean sphere. We know that for any orthonormal matrix \hat{Q} we have

$$\mathbb{1}_{n \times n} = (X_{i,j}^*)_{i,j \leq n} (g_{i,j})_{i,j \leq n} = (X_{i,j}^*)_{i,j \leq n} \underbrace{\hat{Q} \hat{Q}^T}_{=\mathbb{1}_{n \times n}} (g_{i,j})_{i,j \leq n}.$$

By the rotational invariance of the Gaussian distribution we get that

$$\mathbb{1}_{n \times n} \stackrel{\mathcal{D}}{=} (X_{i,j}^*)_{i,j \leq n} \hat{Q} (g_{i,j})_{i,j \leq n}.$$

It follows that

$$(X_{i,j}^*)_{i,j \leq n} \stackrel{\mathcal{D}}{=} (X_{i,j}^*)_{i,j \leq n} \hat{Q}.$$

Therefore, X_1^* is rotationally invariant and thus N_1 is uniformly distributed on the Euclidean sphere. We obtain

$$\begin{aligned} \mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} x\|_2 > \varepsilon) &= \mathbb{P}\left(\frac{\|\text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1\|_2}{|\langle N_1, (g_{1,j})_{j \leq n} \rangle|} > \varepsilon\right) \\ &= \mathbb{P}\left(|\langle N_1, (g_{1,j})_{j \leq n} \rangle| < \frac{\|\text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1\|_2}{\varepsilon}\right) \\ &= \mathbb{P}^{(N_1, (g_{1,j})_{j \leq n})} \left(|\langle N_1, (g_{1,j})_{j \leq n} \rangle| < \frac{\|\text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) N_1\|_2}{\varepsilon}\right). \end{aligned}$$

Since N_1 is determined by the hyperplane spanned by $(g_{2,j})_{j \leq n}, \dots, (g_{n,j})_{j \leq n}$ we get that N_1 is stochastically independent to $(g_{1,j})_{j \leq n}$ as a measurable function of $(g_{2,j})_{j \leq n}, \dots, (g_{n,j})_{j \leq n}$. Again, we want to remark that stochastic independence is equivalent to the fact that the joint distribution $\mathbb{P}^{(N_1, (g_{1,j})_{j \leq n})}$ of N_1 and $(g_{1,j})_{j \leq n}$ is the product measure $\mathbb{P}^{N_1} \times \mathbb{P}^{(g_{1,j})_{j \leq n}}$. Thus, we get by Fubini's theorem

$$\begin{aligned}
\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} x\|_2 > \varepsilon) &= \mathbb{P}^{N_1} \times \mathbb{P}^{(g_{1,j})_{j \leq n}} \left(\left| \langle N_1, (g_{1,j})_{j \leq n} \rangle \right| < \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon} \right) \\
&= \int \mathbf{1}_{\{|\langle N_1, (g_{1,j})_{j \leq n} \rangle| < \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon}\}} d\mathbb{P}^{N_1} \times \mathbb{P}^{(g_{1,j})_{j \leq n}} \\
&= \int \int \mathbf{1}_{\{|\langle N_1, (g_{1,j})_{j \leq n} \rangle| < \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon}\}} d\mathbb{P}^{(g_{1,j})_{j \leq n}} d\mathbb{P}^{N_1} \\
&= \int \mathbb{P}^{(g_{1,j})_{j \leq n}} \left(\left| \langle N_1, (g_{1,j})_{j \leq n} \rangle \right| < \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon} \right) d\mathbb{P}^{N_1} \\
&= \int \mathbb{P}^{(g_{1,j})_{j \leq n}} \left(\left| \sum_{j \leq n} N_1(j) g_{1,j} \right| < \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon} \right) d\mathbb{P}^{N_1}.
\end{aligned}$$

We use the rotation invariance of the Gaussian distribution to obtain

$$\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} x\|_2 > \varepsilon) = \int \mathbb{P} \left(|g| < \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon} \right) d\mathbb{P}^{N_1},$$

where g is a standard normal distributed random variable. We know that

$$\begin{aligned}
\mathbb{P} \left(|g| < \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon} \right) &= \frac{2}{\pi} \int_0^{\frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon}} \exp \left(-\frac{x^2}{2} \right) dx \\
&\leq \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon}
\end{aligned} \tag{5.1}$$

and obtain

$$\begin{aligned}
\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} x\|_2 > \varepsilon) &\leq \int \frac{\|\text{diag}(\frac{1}{a_1}, \dots, \frac{1}{a_n}) N_1\|_2}{\varepsilon} d\mathbb{P}^{N_1} \\
&= \frac{1}{\varepsilon} \int_{S^{n-1}} \left(\sum_{j=1}^n \left| \frac{1}{a_j} \right|^2 |N_1(j)|^2 \right)^{\frac{1}{2}} d\mathbb{P}^{N_1},
\end{aligned}$$

where \mathbb{P}^{N_1} is the unique rotationally invariant, normalized measure on the Euclidean sphere. We use Jensen's inequality and obtain

$$\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} x\|_2 > \varepsilon) \leq \frac{1}{\varepsilon} \left(\sum_{j=1}^n \left| \frac{1}{a_j} \right|^2 \int_{S^{n-1}} |N_1(j)|^2 d\sigma \right)^{\frac{1}{2}},$$

We use Lemma 5 to express $|N_1(j)|^2$ by $\frac{|g_j|^2}{\|(g_i)_{i \leq n}\|_2^2}$ where the g_j ($j = 1, \dots, n$) are i.i.d. standard normal distributed random variables and obtain

$$\begin{aligned}
\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} x\|_2 > \varepsilon) &\leq \frac{1}{\varepsilon} \left(\sum_{j=1}^n \left| \frac{1}{a_j} \right|^2 \int \frac{|g_j|^2}{\|(g_i)_{i \leq n}\|_2^2} d\mathbb{P} \right)^{\frac{1}{2}} \\
&= \frac{1}{\varepsilon} \frac{1}{\sqrt{n}} \left(\sum_{j=1}^n \left| \frac{1}{a_j} \right|^2 \int \frac{\sum_{i=1}^n |g_i|^2}{\|(g_i)_{i \leq n}\|_2^2} d\mathbb{P} \right)^{\frac{1}{2}} \\
&= \frac{1}{\varepsilon} \frac{1}{\sqrt{n}} \left(\sum_{j=1}^n \left| \frac{1}{a_j} \right|^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

Step 2: Let $v \in S^{n-1}$ be uniformly distributed on the Euclidean sphere and independent to $(a_j g_{i,j})_{i,j \leq n}$. Then it follows by Fubini's theorem and step 1 that

$$\begin{aligned}
\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon) &= \mathbb{P}^{((a_j g_{i,j})_{i,j \leq n}^{-1}, v)}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon) \\
&= \mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} \times \mathbb{P}^v(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon) \\
&= \int \mathbf{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon\}} d\mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} \times \mathbb{P}^v \\
&= \int \int \mathbf{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon\}} d\mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} d\mathbb{P}^v \\
&= \int \mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon) d\mathbb{P}^v \\
&\leq \int \frac{1}{\varepsilon} \frac{1}{\sqrt{n}} \left(\sum_{j=1}^n \left| \frac{1}{a_j} \right|^2 \right)^{\frac{1}{2}} d\mathbb{P}^v \\
&= \frac{1}{\varepsilon} \frac{1}{\sqrt{n}} \left(\sum_{j=1}^n \left| \frac{1}{a_j} \right|^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

Furthermore, since $(a_j g_{i,j})_{i,j \leq n}$ is invertible almost surely we can use the singular value decomposition for $(a_j g_{i,j})_{i,j \leq n}^{-1}$ to obtain the existence of random vectors $u_i \in S^{n-1}$ ($i \leq n$) such that

$$s_i((a_j g_{i,j})_{i,j \leq n}^{-1}) = \|(a_j g_{i,j})_{i,j \leq n}^{-1} u_i\|_2.$$

Especially, we have

$$\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} = \|(a_j g_{i,j})_{i,j \leq n}^{-1} u_1\|_2.$$

By the singular value decomposition we also get

$$\begin{aligned}
\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2^2 &= \left| \langle ((a_j g_{i,j})_{i,j \leq n}^{-1})^T (a_j g_{i,j})_{i,j \leq n}^{-1} v, v \rangle \right|^2 \\
&= \left| \left\langle \sum_{i=1}^n s_i^2 ((a_j g_{i,j})_{i,j \leq n}^{-1}) \langle u_i, v \rangle u_i, v \right\rangle \right|^2 \\
&= \sum_{i=1}^n s_i^2 ((a_j g_{i,j})_{i,j \leq n}^{-1}) |\langle u_i, v \rangle|^2 \\
&\geq s_1^2 ((a_j g_{i,j})_{i,j \leq n}^{-1}) |\langle u_1, v \rangle|^2 \\
&= \|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2}^2 |\langle u_1, v \rangle|^2.
\end{aligned}$$

It follows for all $c > 0$ and for all uniformly distributed $v \in S^{n-1}$

$$\begin{aligned}
\mathbb{P}\left(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon \sqrt{\frac{c}{n}}\right) &= \mathbb{P}^{((a_j g_{i,j})_{i,j \leq n}^{-1}, v)}\left(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon \sqrt{\frac{c}{n}}\right) \\
&= \mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} \times \mathbb{P}^v\left(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon \sqrt{\frac{c}{n}}\right) \\
&\geq \mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} \times \mathbb{P}^v\left(\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} |\langle u_1, v \rangle| \geq \varepsilon \sqrt{\frac{c}{n}}\right) \\
&\geq \mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} \times \mathbb{P}^v\left(\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon, |\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\right) \\
&= \int \mathbf{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon, |\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\}} d\mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} \times \mathbb{P}^v.
\end{aligned}$$

We use Fubini's theorem and

$$\mathbb{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon, |\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\}} = \mathbb{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon\}} \mathbb{1}_{\{|\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\}}$$

to obtain

$$\begin{aligned} \mathbb{P}\left(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon \sqrt{\frac{c}{n}}\right) &= \int \int \mathbb{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon, |\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\}} d\mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} d\mathbb{P}^v \\ &= \int \int \mathbb{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon\}} \mathbb{1}_{\{|\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\}} d\mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} d\mathbb{P}^v \\ &= \int \mathbb{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon\}} \int \mathbb{1}_{\{|\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\}} d\mathbb{P}^v d\mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}} \\ &= \int \mathbb{1}_{\{\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon\}} \mathbb{P}^v\left(|\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\right) d\mathbb{P}^{(a_j g_{i,j})_{i,j \leq n}^{-1}}. \end{aligned}$$

We know that every realization of $\{u_i\}_{i \leq n}$ is an orthonormal basis of \mathbb{R}^n . With respect to this basis v is again a uniform distributed random unit vector, since the distribution of v is rotationally invariant and the first coordinate of v regarding this basis is $\langle u_1, v \rangle$. Therefore, we can apply Lemma 20 and obtain

$$\mathbb{P}^v\left(|\langle u_1, v \rangle| \geq \sqrt{\frac{c}{n}}\right) \geq \mathbb{P}(|g| \geq \sqrt{c}),$$

where g is a standard normal distributed random variable. It follows that

$$\mathbb{P}\left(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon \sqrt{\frac{c}{n}}\right) \geq \mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon) \mathbb{P}(|g| \geq \sqrt{c}).$$

We obtain

$$\begin{aligned} \mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon) &\leq \frac{\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1} v\|_2 \geq \varepsilon \sqrt{\frac{c}{n}})}{\mathbb{P}(|g| \geq \sqrt{c})} \\ &\leq \sqrt{\frac{2}{\pi}} \frac{\|(1/a_i)_{i \leq n}\|_2}{\sqrt{c} \varepsilon \mathbb{P}(|g| \geq \sqrt{c})}. \end{aligned}$$

Since this inequality is true for every $c > 0$, we can choose a positive c to get rid of the $\sqrt{c} \mathbb{P}(|g| \geq \sqrt{c})$ term. Since Sankar, Spielman and Teng found out, that $c = 0.57$ almost minimizes the term on the right hand side and evaluated the error function numerically (see [32, page 8]) we use this result to obtain

$$\mathbb{P}(\|(a_j g_{i,j})_{i,j \leq n}^{-1}\|_{2 \rightarrow 2} \geq \varepsilon) \leq_{c=0.57} \frac{2.35 \|(1/a_i)_{i \leq n}\|_2}{\varepsilon}.$$

□

As an application of Theorem 14 we want to estimate the expect value of random Block matrices. Let m and n_1, \dots, n_m be natural numbers and define $n = n_1 + \dots + n_m$. Let $a_1^{(1)}, \dots, a_{n_1}^{(1)}, \dots, a_1^{(m)}, \dots, a_{n_m}^{(m)} \in \mathbb{R} \setminus \{0\}$. Furthermore, for every $i \in \{1, \dots, m\}$ let $g_{k,l}^{(i)}$ ($k, l \leq n_i$) be independent standard normal distributed random variables. We want to study the $n \times n$ random block matrix

$$X = \begin{pmatrix} X_1 & & 0 \\ & \dots & \\ 0 & & X_m \end{pmatrix},$$

where for all $i = 1, \dots, m$ $X_i = (a_l^{(i)} g_{k,l}^{(i)})_{k,l \leq n_i}$. Since

$$\mathbb{E} s_n(X) = \mathbb{E} \min_{1 \leq i \leq m} s_{n_i}(X_i)$$

and

$$\mathbb{P}\left(\left\|\left(\frac{1}{a_j}\right)_{j \leq n_i}\right\|_2 s_{n_i}(X_i) \leq \varepsilon\right) \leq 2.35\varepsilon$$

by Theorem 14, we can use the following result from [11, page 7]:

Theorem 15 *Let $\alpha > 0$. Let $p > 0$. Let $x_1, \dots, x_n \in \mathbb{R} \setminus \{0\}$ be real numbers and X_1, \dots, X_n be random variables satisfying*

$$\mathbb{P}(|X_i| \leq t) \leq \alpha t \text{ for all } i = 1, \dots, m \text{ and } t \geq 0.$$

Then

$$\frac{1}{(1+p)\alpha^p} \frac{1}{\left(\sum_{i=1}^n \frac{1}{|x_i|}\right)^p} \leq \mathbb{E} \min_{1 \leq i \leq n} |x_i X_i|^p.$$

Moreover, if the X_i 's are independent and satisfy the condition

$$\mathbb{P}(|X_i| > t) \leq \exp(-\beta t) \text{ for every } t > 0,$$

where $\beta > 0$ is a constant, then

$$\mathbb{E} \min_{1 \leq i \leq n} |x_i X_i|^p \leq \beta^{-p} \Gamma(1+p) \frac{1}{\left(\sum_{i=1}^n \frac{1}{|x_i|}\right)^p}.$$

From the first statement in Theorem 15 it follows that for the expected value of the smallest singular value of the Block matrix X we have

$$\mathbb{E} s_n(X) \geq \frac{1}{2} 2.35^{-1} \left(\sum_{i=1}^m \frac{1}{\left\|\left(\frac{1}{a_j^{(i)}}\right)_{j \leq n_i}\right\|_2} \right)^{-1}.$$

We want to point out that we don't need the independence of the matrices X_i . For $p = 2$ Theorem 15 states that for a class of random variables which satisfy both conditions, one knows the order of the expected value of the smallest singular value of the $n \times n$ diagonal matrix $\text{diag}(x_1 X_1, \dots, x_n X_n)$. As it is discussed in [11] standard normal distributed and exponentially distributed random variables do satisfy this conditions.

Let us look at complex Gaussian matrices $G_{\mathbb{C}}$, i.e. complex-valued $n \times n$ random matrices whose entries are independent copies of a complex Gaussian random ξ with $\mathbb{E}\xi = 0$ and $\mathbb{E}\text{Re}(\xi)^2 = \frac{1}{2} = \mathbb{E}\text{Im}(\xi)^2$, where $\text{Re}(\xi)$ denotes the real part and $\text{Im}(\xi)$ denotes the imaginary part of ξ . A model example of such a random variable is $\frac{1}{\sqrt{2}}g_1 + \frac{i}{\sqrt{2}}g_2$, where g_1 and g_2 are independent standard normal distributed random variables. For $G_{\mathbb{C}}$ we have the following theorem due to A. Edelman which can be found in [43, page 3].

Theorem 16 *Let $n \in \mathbb{N}$. Let $G_{\mathbb{C}}$ be an $n \times n$ complex-valued Gaussian matrix. Then for any fixed $t > 0$ we have the exact formula*

$$\mathbb{P}(n s_n(G_{\mathbb{C}})^2 \leq t) = \int_0^t \exp(-x) dx.$$

An easy consequence of the theorem above is the following:

$$\mathbb{P}(s_n(G_{\mathbb{C}})^2 > t) = \exp(-t^2/n).$$

Let $n, m \in \mathbb{N}$. It follows that for integers $n_1, \dots, n_m \in \mathbb{N}$ with $\sum_{j \leq m} n_j = n$, real numbers $a_1, \dots, a_m \in \mathbb{R} \setminus \{0\}$ and

$$X_{\mathbb{C}} = \begin{pmatrix} a_1 X_{1, \mathbb{C}} & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & a_m X_{m, \mathbb{C}} \end{pmatrix},$$

where $X_{j, \mathbb{C}}$ are independent $n_j \times n_j$ complex-valued Gaussian matrices we have the following equality for the smallest singular value of $X_{\mathbb{C}}$:

$$\begin{aligned} \mathbb{E} s_n(X_{\mathbb{C}}) &= \mathbb{E} \min_{1 \leq j \leq m} s_{n_j}(a_j X_{j, \mathbb{C}}) \\ &= \int_0^\infty \mathbb{P}(\min_{1 \leq j \leq m} s_{n_j}(a_j X_{j, \mathbb{C}}) > t) dt \\ &= \int_0^\infty \mathbb{P}(\bigcap_{1 \leq j \leq m} \{s_{n_j}(a_j X_{j, \mathbb{C}}) > t\}) dt. \end{aligned}$$

Using the independence of the $X_{j, \mathbb{C}}$'s we obtain

$$\begin{aligned} \mathbb{E} s_n(X_{\mathbb{C}}) &= \int_0^\infty \prod_{1 \leq j \leq m} \mathbb{P}(s_{n_j}(a_j X_{j, \mathbb{C}}) > t) dt \\ &= \int_0^\infty \exp\left(-t^2 \sum_{j \leq m} \frac{n_j}{a_j}\right) dt \\ &= \frac{\sqrt{\pi}}{2} \left(\sum_{j \leq m} \frac{n_j}{a_j}\right)^{-\frac{1}{2}}. \end{aligned}$$

We want to point out that for $a_1 = \dots = a_m = 1$ we have

$$\mathbb{E} s_n(X_{\mathbb{C}}) = \frac{\sqrt{\pi}}{2} n^{-\frac{1}{2}}.$$

5.2 General estimate

In this section we will derive an estimate for more general random tensor matrices. Again the decomposition of the sphere is the crucial tool in our proof.

Theorem 17 *Let $n \in \mathbb{N}$ be an integer. Let ξ_1, \dots, ξ_n be independent centered random variables with variances at least one and subgaussian moments bounded by an absolute constant $B > 0$. Define $X = (X_{i,j})_{i,j \leq n}$ to be the random matrix whose rows are independent copies of the random vector (ξ_1, \dots, ξ_n) . Then for every $\lambda > 0$ and all sequences $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R} \setminus \{0\}$ and all $\delta, \rho \in (0, 1)$ we have*

$$\begin{aligned} &\mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\ &\leq C\lambda + 2c^n + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\ &\quad + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n}^T x\|_2 \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right), \end{aligned}$$

where $C > 0$ and $c \in (0, 1)$ are constants depending on B .

Proof: Let $\delta, \rho \in (0, 1)$. Then we have

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
= & \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is singular}\right) \\
& + \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
\leq & c^n + \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right),
\end{aligned}$$

where $c \in (0, 1)$. We used Theorem 10 for $\varepsilon = 0$ to obtain the inequality. We decompose the euclidean sphere by $S^{n-1} = \text{Comp}(\delta, \rho) \cup \text{Incom}(\delta, \rho)$. It follows

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
= & c^n + \mathbb{P}\left(\min_{\|x\|_2=1} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
= & c^n + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
& \text{or } \inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
\leq & c^n + \mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right).
\end{aligned}$$

By Lemma 16 we know that for all $x \in \text{Incomp}(\delta, \rho)$ we have $\|x\|_1 \geq \frac{\rho^3 \delta}{\sqrt{2^3}} \sqrt{n} \|x\|_2$. It follows

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
= & c^n + \mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \left(\|x\|_1 \left\| (a_i b_j X_{i,j})_{i,j \leq n} \frac{x}{\|x\|_1} \right\|_2\right) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2},\right. \\
& \left. (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
\leq & c^n + \mathbb{P}\left(\frac{\rho^3 \delta}{\sqrt{2^3}} \sqrt{n} \inf_{x \in \text{Incomp}_1(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2},\right. \\
& \left. (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right).
\end{aligned}$$

At this point we want to remind on the definition of the set

$$\text{Incomp}_1(\delta, \rho) = \{x \in \partial B_1^n \mid \exists \nu \in \mathbb{R} : \nu x \in \text{Incom}(\delta, \rho)\}.$$

Since $\text{Incomp}_1 \subset \partial B_1^n$ we obtain

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
\leq & c^n + \mathbb{P}\left(\min_{\|x\|_1=1} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2},\right. \\
& \left.(X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
\leq & c^n + \mathbb{P}\left(\frac{1}{\max_{\|x\|_2=1} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_1} \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2},\right. \\
& \left.(X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right).
\end{aligned}$$

By Lemma 4 we get that

$$(a_i b_j X_{i,j})_{i,j \leq n}^{-1} = \left(\frac{1}{b_i} \frac{1}{a_j} X_{i,j}^*\right)_{i,j \leq n},$$

where $(X_{i,j}^*)_{i,j \leq n} = (X_{i,j})_{i,j \leq n}^{-1}$. It follows by Lemma 2

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
\leq & c^n + \mathbb{P}\left(\frac{1}{\max_{\|x\|_2=1} \|(\frac{1}{b_i} \frac{1}{a_j} X_{i,j}^*)_{i,j \leq n} x\|_1} \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2},\right. \\
& \left.(X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
\leq & c^n + \mathbb{P}\left(\max_{\|x\|_2=1} \sum_{i \leq n} \left|\frac{1}{b_i}\right| \left\|\sum_{j \leq n} \frac{1}{a_j} X_{i,j}^* x\right\| \geq \frac{1}{\lambda} \frac{\rho^3 \delta}{\sqrt{2^3}} \left\|\left(\frac{1}{a_i}\right)_{i \leq n}\right\|_2 \left\|\left(\frac{1}{b_j}\right)_{j \leq n}\right\|_2,\right. \\
& \left.(X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right).
\end{aligned}$$

By Hölder's inequality we get that

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
\leq & c^n + \mathbb{P}\left(\max_{\|x\|_2=1} \left(\sum_{i \leq n} \left|\frac{1}{b_i}\right|^2\right)^{\frac{1}{2}} \left(\sum_{i \leq n} \left|\sum_{j \leq n} \frac{1}{a_j} X_{i,j}^* x_j\right|^2\right)^{\frac{1}{2}} \geq \frac{1}{\lambda} \frac{\rho^3 \delta}{\sqrt{2^3}} \left\| \left(\frac{1}{a_i}\right)_{i \leq n} \right\|_2 \left\| \left(\frac{1}{b_j}\right)_{j \leq n} \right\|_2,\right. \\
& \left. (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
= & c^n + \mathbb{P}\left(\max_{\|x\|_2=1} \|(a_i X_{i,j})_{i,j \leq n}^{-1} x\|_2 \geq \frac{1}{\lambda} \frac{\rho^3 \delta}{\sqrt{2^3}} \left\| \left(\frac{1}{a_i}\right)_{i \leq n} \right\|_2, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
= & c^n + \mathbb{P}\left(\min_{\|x\|_2=1} \|(a_i X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right),
\end{aligned}$$

where we used Lemma 4 and Lemma 2, again. By the equality

$$s_n((a_i X_{i,j})_{i,j \leq n}) = s_n((a_i X_{i,j})_{i,j \leq n}^T)$$

we get

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
\leq & c^n + \mathbb{P}\left(s_n((a_i X_{i,j})_{i,j \leq n}^T) \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right).
\end{aligned}$$

By repeating the argument from above for

$$\mathbb{P}\left(s_n((a_i X_{i,j})_{i,j \leq n}^T) \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right)$$

we get that

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
& \leq c^n + \mathbb{P}\left(s_n((X_{i,j})_{i,j \leq n}^T \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-2} n^{-\frac{1}{2}}, (X_{i,j})_{i,j \leq n} \text{ is non-singular}\right) \\
& \quad + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
& \quad + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n}^T x\|_2 \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right) \\
& \leq c^n + \mathbb{P}\left(s_n((X_{i,j})_{i,j \leq n}) \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-2} n^{-\frac{1}{2}}\right) \\
& \quad + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
& \quad + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n}^T x\|_2 \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right).
\end{aligned}$$

By Theorem 10 it follows for a suitable constant $C > 0$

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j X_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
& \leq 2c^n + C\lambda + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j X_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
& \quad + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i X_{i,j})_{i,j \leq n}^T x\|_2 \leq \lambda \left(\frac{\rho^3 \delta}{\sqrt{2^3}}\right)^{-1} \frac{1}{\|(\frac{1}{a_i})_{i \leq n}\|_2}\right).
\end{aligned}$$

□

Corollary 4 *Let $X = (a_i b_j g_{i,j})_{i,j \leq n}$ be an $n \times n$ random matrix, where $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R} \setminus \{0\}$ and the $g_{i,j}$ are i.i.d. standard normal distributed random variables. Then for all $\varepsilon > 0$ and all $\delta, \rho \in (0, 1)$ we have that*

$$\begin{aligned}
& \mathbb{P}\left(s_n((a_i b_j g_{i,j})_{i,j \leq n}) \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right) \\
& \leq 2.35 \frac{\rho^3 \delta}{\sqrt{2^3}} \lambda + \mathbb{P}\left(\inf_{x \in \text{Comp}(\delta, \rho)} \|(a_i b_j g_{i,j})_{i,j \leq n} x\|_2 \leq \lambda \frac{\sqrt{n}}{\|(\frac{1}{a_i})_{i \leq n}\|_2 \|(\frac{1}{b_j})_{j \leq n}\|_2}\right).
\end{aligned}$$

Proof: As discussed several times before the matrix $(g_{i,j})_{i,j \leq n}$ is almost surely invertible and we can use the argument from the proof of Theorem 17, but instead of repeating it, we continue analog to our computations in the proof of Theorem 14. This completes the proof. □

5.3 A new look at compressible vectors

We start this section with an lemma which connects the invertibility problem for incompressible vectors of an matrix A with the distances of its column vectors X_k to the hyperplane H_k spanned by the other $n - 1$ column vectors. This lemma is a crucial and often used tool when dealing with invertibility problems on the incompressible vectors. It can be found in [28, page 16 f].

Lemma 21 *Let A be an arbitrary random matrix. Let X_1, \dots, X_n denote the column vectors of A , and let $H_k = \text{span}\{X_i \mid i \neq k\}$ denote the span of all column vectors except the k -th. Then for every $\delta, \rho \in (0, 1)$ and every $\varepsilon > 0$, one has*

$$\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Ax\|_2 < \varepsilon \rho n^{-\frac{1}{2}}\right) \leq \frac{1}{\delta n} \sum_{k \leq n} \mathbb{P}(\text{dist}(X_k, H_k) < \varepsilon).$$

Proof: Let $x \in \text{Incomp}(\delta, \rho)$. Then we have

$$\begin{aligned} \|Ax\|_2 &= \|Ax - 0\|_2 \\ &\geq \max_{1 \leq k \leq n} \inf_{y \in H_k} \|Ax - y\|_2 \\ &= \max_{1 \leq k \leq n} \text{dist}(Ax, H_k) \\ &= \max_{1 \leq k \leq n} \text{dist}\left(\sum_{j \leq n} x_j X_j, H_k\right) \\ &= \max_{1 \leq k \leq n} \text{dist}(x_k X_k, H_k) \\ &= \max_{1 \leq k \leq n} |x_k| \text{dist}(X_k, H_k). \end{aligned}$$

Furthermore, we get

$$\begin{aligned} \mathbb{E}|\{k \in \{1, \dots, n\} \mid \text{dist}(X_k, H_k) < \varepsilon\}| &= \mathbb{E} \sum_{k \leq n} \mathbf{1}_{\{\text{dist}(X_k, H_k) < \varepsilon\}} \\ &= \sum_{k \leq n} \mathbb{P}(\text{dist}(X_k, H_k) < \varepsilon). \end{aligned}$$

We define U to be the event that the set $\sigma_1 = \{k \in \{1, \dots, n\} \mid \text{dist}(X_k, H_k) \geq \varepsilon\}$ contains more than $(1 - \delta)n$ elements. Then it follows by Markov's inequality that

$$\begin{aligned} \mathbb{P}(U^c) &= \mathbb{P}(|\{k \in \{1, \dots, n\} \mid \text{dist}(X_k, H_k) \geq \varepsilon\}| < (1 - \delta)n) \\ &= \mathbb{P}(|\{k \in \{1, \dots, n\} \mid \text{dist}(X_k, H_k) < \varepsilon\}| \geq \delta n) \\ &\leq \frac{\mathbb{E}|\{k \in \{1, \dots, n\} \mid \text{dist}(X_k, H_k) < \varepsilon\}|}{\delta n} \\ &= \frac{1}{\delta n} \sum_{k \leq n} \mathbb{E} \mathbf{1}_{\{\text{dist}(X_k, H_k) < \varepsilon\}} \\ &= \frac{1}{\delta n} \sum_{k \leq n} \mathbb{P}(\text{dist}(X_k, H_k) < \varepsilon). \end{aligned}$$

For a subset $\sigma \subset \{1, \dots, n\}$ we denote the orthogonal projection onto \mathbb{R}^σ in \mathbb{R}^n by P_σ . We know that at least δn elements are contained in the set $\sigma_2(x) = \{k \in \{1, \dots, n\} \mid |x_k| \geq \rho n^{-\frac{1}{2}}\}$, since otherwise we would contradict the incompressibility of x :

Assume that $\sigma_2(x)$ contains less than δn elements. Then it follows that $\|x - y\|_2 \leq \rho$ for the sparse vector $y = P_{\sigma_2(x)} x$, since $\|P_{\sigma_2(x)} x\|_2 \leq \rho$. This can not be true as x is incompressible.

Assume that the event U occurs. Let $x \in \text{Incomp}(\delta, \rho)$ be an arbitrary but fixed vector. Then we have

$$|\sigma_1| + |\sigma_2(x)| > (1 - \delta)n + \delta n = n.$$

Therefore, it follows that the intersection $\sigma_1 \cap \sigma_2(x)$ is nonempty. Let $k \in \sigma_1 \cap \sigma_2(x)$. Then by the definitions of the sets σ_1 and $\sigma_2(x)$, we have that

$$\|Ax\|_2 \geq |x_k| \text{dist}(X_k, H_k) \geq \rho n^{-\frac{1}{2}} \varepsilon.$$

It follows that the occurrence of U implies $\|Ax\|_2 \geq \rho \varepsilon n^{-\frac{1}{2}}$ for all $x \in \text{Incomp}(\delta, \rho)$ and we obtain

$$\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Ax\|_2 < \varepsilon \rho n^{-\frac{1}{2}}\right) \leq \mathbb{P}(U^c) \leq \frac{1}{\delta n} \sum_{k \leq n} \mathbb{P}(\text{dist}(X_k, H_k) < \varepsilon).$$

□

We can use Lemma 21 to show that for matrices $(b_j g_{i,j})_{i,j \leq n}$ the compressible vectors play an important role. This is surprising, as in the cases studied before compressible vectors have little impact on the least singular value (see for example [28]). To proof this, we need the following lemma which gives an probabilistic estimate for a lower bound of

$$\inf_{x \in \text{Incomp}(\delta, \rho)} \|Xx\|_2.$$

A main ingredient of this lemma is Lemma 21.

Lemma 22 *Let $X = (a_j g_{i,j})_{i,j \leq n}$ be an $n \times n$ random matrix, where $a_1, \dots, a_n \in \mathbb{R} \setminus \{0\}$ and the $g_{i,j}$ are i.i.d. standard normal distributed random variables. Then for every $\delta, \rho \in (0, 1)$ and every $\varepsilon > 0$ we have*

$$\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Xx\|_2 \leq \frac{\varepsilon \rho \sqrt{n}}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \leq \frac{\varepsilon}{\delta}.$$

Proof: Define $H_k = \text{span}\{(a_j g_{i,j})_{i \leq n} \mid j \neq k\} = \text{span}\{(g_{i,j})_{i \leq n} \mid j \neq k\}$ as the span of the columns of X except the k -th. We use Lemma 21 and obtain

$$\begin{aligned} \mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Xx\|_2 \leq \frac{\varepsilon \rho \sqrt{n}}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) &= \mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|(a_j g_{i,j})_{i,j \leq n} x\|_2 \leq \frac{\varepsilon \rho}{\sqrt{n}} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \\ &\leq \frac{1}{\delta n} \sum_{k \leq n} \mathbb{P}\left(\text{dist}((a_k g_{i,k})_{i \leq n}, H_k) < \varepsilon \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \\ &= \frac{1}{\delta n} \sum_{k \leq n} \mathbb{P}\left(\text{dist}((g_{i,k})_{i \leq n}, H_k) < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right). \end{aligned}$$

Like before we write the distance from the random vectors $(g_{i,k})_{i \leq n}$ to the subspaces H_k (which are almost surely hyperplanes as the matrix $(g_{i,j})_{i,j \leq n}$ is almost surely invertible by Theorem 8) as $|\langle (g_{i,k})_{i \leq n}, N_k \rangle|$ where N_k is a random normal on H_k . Again, N_k is stochastically independent to $(g_{i,k})_{i \leq n}$ as a measurable function of the random vectors $(g_{i,j})_{i \leq n}$ ($j \neq k$). We compute

$$\begin{aligned} &\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Xx\|_2 \leq \frac{\varepsilon \rho \sqrt{n}}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \\ &\leq \frac{1}{\delta n} \sum_{k \leq n} \mathbb{P}^{((g_{i,k})_{i \leq n}, N_k)}\left(|\langle (g_{i,k})_{i \leq n}, N_k \rangle| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \\ &= \frac{1}{\delta n} \sum_{k \leq n} \mathbb{P}^{((g_{i,k})_{i \leq n})} \times \mathbb{P}^{N_k}\left(|\langle (g_{i,k})_{i \leq n}, N_k \rangle| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \\ &= \frac{1}{\delta n} \sum_{k \leq n} \int \mathbb{1}_{\{|\langle (g_{i,k})_{i \leq n}, N_k \rangle| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\}} d\mathbb{P}^{((g_{i,k})_{i \leq n})} \times \mathbb{P}^{N_k}. \end{aligned}$$

Therefore, we obtain by Fubini's theorem

$$\begin{aligned}
& \mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Xx\|_2 \leq \frac{\varepsilon \rho \sqrt{n}}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \\
& \leq \frac{1}{\delta n} \sum_{k \leq n} \int \int \mathbf{1}_{\{|\langle (g_{i,k})_{i \leq n}, N_k \rangle| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\}} d\mathbb{P}^{(g_{i,k})_{i \leq n}} d\mathbb{P}^{N_k} \\
& = \frac{1}{\delta n} \sum_{k \leq n} \int \mathbb{P}^{(g_{i,k})_{i \leq n}} \left(|\langle (g_{i,k})_{i \leq n}, N_k \rangle| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1} \right) d\mathbb{P}^{N_k} \\
& = \frac{1}{\delta n} \sum_{k \leq n} \int \mathbb{P}^{(g_{i,k})_{i \leq n}} \left(\left| \sum_{i \leq n} g_{i,k} N_k(i) \right| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1} \right) d\mathbb{P}^{N_k}.
\end{aligned}$$

Now we use the rotational invariance of the Gaussian distribution to get that

$$\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Xx\|_2 \leq \frac{\varepsilon \rho \sqrt{n}}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) \leq \frac{1}{\delta n} \sum_{k \leq n} \int \mathbb{P}\left(|g| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) d\mathbb{P}^{N_k}$$

where g is a standard normal distributed random variable. Since

$$\begin{aligned}
\mathbb{P}\left(|g| < \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) & = \sqrt{\frac{2}{\pi}} \int_0^{\frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}} \exp\left(-\frac{x^2}{2}\right) dx \\
& \leq \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1}
\end{aligned}$$

we obtain

$$\begin{aligned}
\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|Xx\|_2 \leq \frac{\varepsilon \rho \sqrt{n}}{\|(\frac{1}{a_j})_{j \leq n}\|_1}\right) & \leq \frac{1}{\delta n} \sum_{k \leq n} \int \frac{\varepsilon}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1} d\mathbb{P}^{N_k} \\
& = \frac{\varepsilon}{\delta n} \sum_{k \leq n} \frac{1}{|a_k|} \frac{n}{\|(\frac{1}{a_j})_{j \leq n}\|_1} \\
& = \frac{\varepsilon}{\delta}.
\end{aligned}$$

□

Now we want to use Lemma 22 to conclude that there exist random tensor matrices X such that the compressible vectors have a big impact on the smallest singular value of the matrix X . Therefore, we look at the following example:

For the first example we want to remind of the situation in Theorem 9: For any $\varepsilon > 0$ we want to study the least singular value of the matrix

$$G_\varepsilon = \begin{pmatrix} g_{1,1} & \cdots & g_{1,n-1} & \varepsilon g_{1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ g_{n,1} & \cdots & g_{n,n-1} & \varepsilon g_{n,n} \end{pmatrix},$$

where for $i, j \in \{1, \dots, n\}$ the $g_{i,j}$ are i.i.d. standard normal distributed random variables. Furthermore, let $\delta, \rho \in (0, 1)$. It follows by Lemma 22 that

$$\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|G_\varepsilon x\|_2 \leq \frac{\lambda \rho \sqrt{n}}{\|(\underbrace{1, \dots, 1}_{n-1 \text{ times}}, \frac{1}{\varepsilon})\|_1}\right) \leq \frac{\lambda}{\delta}.$$

Setting $\lambda = \frac{\delta}{2}$ we get that

$$\mathbb{P}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|G_\varepsilon x\|_2 \geq \frac{\frac{\delta}{2} \rho \sqrt{n}}{\|(1, \dots, 1, \frac{1}{\varepsilon})\|_1}\right) \geq \frac{1}{2}.$$

For a random variable X let MX denote the median of X . Then we have the following inequality due to Markov's inequality:

$$\frac{1}{2} \leq \mathbb{P}(X \geq \text{MX}) \leq \frac{\mathbb{E}X}{\text{MX}}.$$

By this considerations we obtain

$$\frac{\delta \rho \sqrt{n}}{2\|(1, \dots, 1, \frac{1}{\varepsilon})\|_1} \leq \mathbb{M}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|G_\varepsilon x\|_2\right) \leq 2\mathbb{E}\left(\inf_{x \in \text{Incomp}(\delta, \rho)} \|G_\varepsilon x\|_2\right).$$

But by Theorem 9 it follows that

$$\mathbb{E}s_n(G_\varepsilon) \sim \min\left\{\varepsilon, \frac{1}{\sqrt{n}}\right\} \sim \frac{1}{\|(1, \dots, 1, \varepsilon)\|_2}.$$

Since $\frac{\delta \rho \sqrt{n}}{2\|(1, \dots, 1, \frac{1}{\varepsilon})\|_1} \sim \min\left\{\varepsilon \sqrt{n}, \frac{1}{\sqrt{n}}\right\}$ we obtain that $\inf_{x \in \text{Incomp}(\delta, \rho)} \|Gx\|_2$ does not have a big impact on $s_n(G_\varepsilon)$ for very small $\varepsilon > 0$.

Chapter 6

Random simplices and connections to singular values of random matrices

We tried to find a geometric representation of the smallest singular value in terms of random simplices to understand what causes the effect that the smallest singular value of an $n \times n$ Gaussian random matrix is of order $\frac{1}{\sqrt{n}}$ whereas the largest singular value is about factor n greater. Even though we did not manage to find such an geometric representation, we learned important techniques that inspired us in our main theorems. In addition to that we made some observations on random simplices which we will present in the sequel.

As we have seen in Lemma 5 one can write a random vector $(g_i)_{i \leq n}$ whose coordinates g_i are independent standard normal distributed random variables as

$$(g_i)_{i \leq n} = \|(g_i)_{i \leq n}\|_2 \frac{(g_i)_{i \leq n}}{\|(g_i)_{i \leq n}\|_2},$$

where $\frac{(g_i)_{i \leq n}}{\|(g_i)_{i \leq n}\|_2}$ is a random unit vector uniformly distributed on the Euclidean sphere which is stochastically independent to $\|(g_i)_{i \leq n}\|_2$. Since, many important geometric properties do not depend on the length of the vectors rather than their position to each other we will study normalized Gaussian vectors most of the time.

Let $n, m \in \mathbb{N}$ with $m \leq n + 1$. Let X_1, \dots, X_m be i.i.d. random vectors chosen uniformly at random from the Euclidean sphere. Then we have seen in the proof of Lemma 6 that X_1, \dots, X_m are linearly independent almost surely and thus the convex hull of X_1, \dots, X_m is almost surely a $m - 1$ -dimensional simplex to which we will refer as random simplex.

Lemma 23 *Let $n \in \mathbb{N}$. Let X_1, \dots, X_n be i.i.d. random vectors chosen uniformly at random from the Euclidean sphere. For all $i, j \in \{1, \dots, n\}$ with $i \neq j$ we have that*

$$\sqrt{2}(\sqrt{2} - 1) \leq \mathbb{E}\|X_i - X_j\|_2 \leq \sqrt{2}.$$

Proof: We apply Jensen's inequality on the concave function $f : [0, \infty) \rightarrow [0, \infty)$ $x \mapsto \sqrt{x}$ and

obtain

$$\begin{aligned}
\mathbb{E}\|X_i - X_j\|_2 &= \mathbb{E}\left(\sum_{k \leq n} |X_{i,k} - X_{j,k}|^2\right)^{\frac{1}{2}} \\
&= \mathbb{E}\left(\sum_{k \leq n} X_{i,k}^2 - 2X_{i,k}X_{j,k} + X_{j,k}^2\right)^{\frac{1}{2}} \\
&\leq \left(\mathbb{E}\sum_{k \leq n} X_{i,k}^2 - 2X_{i,k}X_{j,k} + X_{j,k}^2\right)^{\frac{1}{2}} \\
&= \left(\mathbb{E}\|X_i\|_2^2 + \mathbb{E}\|X_j\|_2^2 - \underbrace{2\sum_{k \leq n} \mathbb{E}X_{i,k}\mathbb{E}X_{j,k}}_{=0}\right)^{\frac{1}{2}} \\
&= \sqrt{2}.
\end{aligned}$$

The last inequality is due to the fact that for all $i \in \{1, \dots, n\}$ each coordinate of X_i has mean zero. On the other hand we get by the parallelogram law

$$\begin{aligned}
\mathbb{E}(\|X_i + X_j\|_2 + \|X_i - X_j\|_2) &\geq \mathbb{E}(\|X_i + X_j\|_2^2 + \|X_i - X_j\|_2^2)^{\frac{1}{2}} \\
&= \sqrt{2}\mathbb{E}(\|X_i\|_2^2 + \|X_j\|_2^2)^{\frac{1}{2}}.
\end{aligned}$$

Again by Jensen's inequality it follows that

$$\begin{aligned}
\mathbb{E}\|X_i - X_j\|_2 &\geq \sqrt{2}\mathbb{E}(\|X_i\|_2^2 + \|X_j\|_2^2)^{\frac{1}{2}} - \mathbb{E}\|X_i + X_j\|_2 \\
&= 2 - \mathbb{E}\|X_i + X_j\|_2 \\
&\geq 2 - (\mathbb{E}\|X_i + X_j\|_2^2)^{\frac{1}{2}} \\
&= 2 - (\mathbb{E}\langle X_i + X_j, X_i + X_j \rangle)^{\frac{1}{2}} \\
&= 2 - (\mathbb{E}\|X_i\|_2^2 + \mathbb{E}\|X_j\|_2^2)^{\frac{1}{2}}
\end{aligned}$$

where the last equality follows by an analogue consideration as above. Finally, we get that

$$\mathbb{E}\|X_i - X_j\|_2^2 \geq \sqrt{2}(\sqrt{2} - 1)$$

which ends the proof. \square

We calculated that the expected (Euclidean) distance between two vertices of the random simplex is of order $\sqrt{2}$. The next lemma sharpens this statement.

Lemma 24 *Let $n \in \mathbb{N}$. Let X_1, \dots, X_n be independent unit random vectors chosen uniformly at random from the Euclidean sphere. Then for all $i \in \{1, \dots, n\}$ we have*

$$\mathbb{P}(\{\exists j \in \{1, \dots, n\} \setminus \{i\} : d(X_j, X_i) < \varepsilon\}) \leq (n-1) \exp\left(-\frac{(n-2)(\frac{\pi}{2} - \varepsilon)^2}{2}\right),$$

where $d(X_j, X_i)$ denotes the geodesic distance between X_j and X_i .

The following proposition states that most of the mass of the Euclidean sphere lies concentrated around any set which carries at least half of the mass of the sphere. It can be found in [17] on page 1.

Proposition 1 *Denote by σ_{n-1} the unique normalized rotational invariant measure on S^{n-1} . Let $A \subset S^{n-1}$ be any measurable set with $\sigma_{n-1}(A) \geq \frac{1}{2}$. Then for every $\varepsilon > 0$ we have for the ε -neighbourhood $A_\varepsilon = \{x \in S^{n-1} \mid d(x, A) < \varepsilon\}$ of A , where the distance is measured in the geodesic distance, that*

$$\sigma_{n-1}(A_\varepsilon) \geq 1 - \exp\left(-\frac{(n-2)\varepsilon^2}{2}\right).$$

Proof (Lemma 24): Let $\varepsilon \in (0, \frac{\pi}{2})$. We have for all $i \in \{1, \dots, n\}$ that

$$\begin{aligned} \mathbb{P}(\{\exists j \in \{1, \dots, n\} \setminus \{i\} : d(X_j, X_i) < \varepsilon\}) &= \mathbb{P}\left(\bigcup_{\substack{j \leq n \\ j \neq i}} \{d(X_j, X_i) < \varepsilon\}\right) \\ &\leq \sum_{\substack{j \leq n \\ j \neq i}} \mathbb{P}(d(X_j, X_i) < \varepsilon) \\ &= (n-1) \mathbb{P}(d(X_1, X_i) < \varepsilon) \\ &= (n-1) \int \mathbf{1}_{\{d(X_1, X_i) < \varepsilon\}} d\mathbb{P}^{X_1} \times \mathbb{P}^{X_i}, \end{aligned}$$

where we used the independence of X_1 and X_i ($i = 2, \dots, n$) in the last equality. We use Fubini's theorem and obtain

$$\mathbb{P}(\{\exists j \in \{1, \dots, n\} \setminus \{i\} : d(X_j, X_i) < \varepsilon\}) \leq (n-1) \int \mathbb{P}^{X_1}(d(X_1, X_i) < \varepsilon) d\mathbb{P}^{X_i}.$$

By Proposition 1 we get that

$$\mathbb{P}(\{\exists j \in \{1, \dots, n\} \setminus \{i\} : d(X_j, X_i) < \varepsilon\}) \leq (n-1) \exp\left(-\frac{(n-2)(\frac{\pi}{2} - \varepsilon)^2}{2}\right),$$

where we used that every hemisphere H has $\sigma_{n-1}(H) = \frac{1}{2}$. □

Another important quantity of the (random) simplex is its center of gravity.

Lemma 25 *Let X_1, \dots, X_n be independent random variables that are uniformly distributed on the sphere. Denote by $\Delta_{n-1} = \text{conv}\{X_1, \dots, X_n\}$ the random simplex spanned by X_1, \dots, X_n . Furthermore, let*

$$G_{\Delta_{n-1}} = \frac{1}{n} \sum_{i \leq n} X_i$$

denote the center of gravity of Δ_{n-1} . Then we have

$$\mathbb{E}\|G_{\Delta_{n-1}}\|_2 \sim \frac{1}{\sqrt{n}}.$$

Proof: Since l_2^n is of type and cotype 2 we have

$$\begin{aligned} C_2(l_2^n)^{-1} \frac{1}{\sqrt{n}} &= C_2(l_2^n)^{-1} \frac{1}{n} \mathbb{E}\left(\sum_{i \leq n} \|X_i\|_2^2\right)^{\frac{1}{2}} \\ &\leq \frac{1}{n} \mathbb{E} \int_0^1 \left\| \sum_{i \leq n} r_i(t) X_i \right\|_2 dt \\ &\leq T_2(l_2^n) \frac{1}{n} \mathbb{E}\left(\sum_{i \leq n} \|X_i\|_2^2\right)^{\frac{1}{2}} \\ &= T_2(l_2^n) \frac{1}{\sqrt{n}}, \end{aligned}$$

where $(r_i)_{i \leq n}$ is a Rademacher sequence, $C_2(l_2^n)$ and $T_2(l_2^n)$ denote the cotype and type constant of l_2^n . Since for all $i \in \{1, \dots, n\}$ the X_i are symmetric distributed it follows that

$$\mathbb{E}\|G_{\Delta_{n-1}}\|_2 = \frac{1}{n} \mathbb{E}\left\| \sum_{i \leq n} X_i \right\|_2 = \frac{1}{n} \int_0^1 \mathbb{E}\left\| \sum_{i \leq n} r_i(t) X_i \right\|_2 dt.$$

Thus it follows by Fubini's theorem that

$$\mathbb{E}\|G_{\Delta_{n-1}}\|_2 = \frac{1}{n} \mathbb{E} \int_0^1 \left\| \sum_{i \leq n} r_i(t) X_i \right\|_2 dt \sim \frac{1}{\sqrt{n}}.$$

□

In [20] R. Miles gave an answer to the question about the volume of a random simplex: Consider we choose $i + j$ independent random points in the Euclidean ball B_2^n , where i are chosen uniformly at random in the unit ball B_2^n and j are chosen uniformly at random on the sphere S^{n-1} . In the case $2 \leq i + j \leq n + 1$ the convex hull of the $i + j$ random points is almost surely an m -dimensional simplex Δ_m ($m = i + j - 1$) and the $i + j$ random points are the vertices of Δ_m .

Theorem 18 (Miles) *We have for all $k = 1, 2, \dots$ that*

$$\begin{aligned} \mathbb{E}|vol_m(\Delta_m)|^k &= \left(\frac{1}{m!}\right)^k \left(\frac{n}{n+k}\right)^i \frac{\Gamma(\frac{1}{2}(m+1)(n+k) - j + 1)}{\Gamma(\frac{1}{2}[(m+1)n + mk] - j + 1)} \left(\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}(n+k))}\right)^m \\ &\quad \times \prod_{l=1}^{m-1} \frac{\Gamma(\frac{1}{2}(n-m+k+l))}{\Gamma(\frac{1}{2}(n-m+l))}, \end{aligned}$$

where $vol_m(\Delta_m)$ denotes the m -dimensional volume of Δ_m .

This result is a consequence of the well known Blaschke-Petkantschin formula which can be found in [33, page 201] or [37, page 77]: Let $n, k \in \mathbb{N}$, $k \leq n$. Let dL_k^n denote the measure on all k -dimensional affine subspaces of \mathbb{R}^n

$$\wedge_{i=0}^k dx_i^n = (k! vol_k(\text{conv}\{x_0, \dots, x_k\}))^{n-k} \wedge_{i=0}^k dx_i^k dL_k^n.$$

Furthermore, for the case $k = n - 1$ one can decompose dL_{n-1}^n as follows (see [37, page 77])

$$dL_{n-1}^n = dp d\mu_{S^{n-1}}(\xi),$$

where p denotes the distance of the hyperplane to the origin and ξ is the normal vector of the hyperplane.

In [20] Miles showed that choosing n points independently and uniformly at random from the sphere S^{n-1} is stochastically equivalent to the following (see also [4]) First, one chooses a random hyperplane, determined by its random normal vector which is uniformly distributed on the sphere, and its distance p to the origin which has the probability density function

$$p : \mathbb{R} \rightarrow [0, \infty), \begin{cases} \frac{2}{B(\frac{1}{2}, \frac{1}{2}(n-1)^2)} (1-x^2)^{\frac{n^2-2n-1}{2}}, & x \in [0, 1] \\ 0, & \text{else} \end{cases},$$

where

$$B : (0, \infty) \times (0, \infty) \rightarrow \mathbb{R}, \quad B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

denotes beta function. Then one chooses n random unit vectors x_1, \dots, x_n from the intersection of the hyperplane with the sphere by weighting the jointly uniform distribution by

$$(n-1)! \frac{B(\frac{1}{2}, \frac{1}{2}(n-1)^2)}{2} \frac{vol_{n-2}(S^{n-2})^n}{vol_{n-1}(S^{n-1})^{n-1}} \frac{1}{(1-p^2)^{\frac{n-1}{2}}} vol_{n-1}(\text{conv}\{x_1, \dots, x_n\}).$$

Lemma 26 Let $n \in \mathbb{N}$. Let X_1, \dots, X_n be independent random vectors chosen uniformly at random from S^{n-1} . Furthermore, let p denote the distance of the (almost surely affine) hyperplane spanned by X_1, \dots, X_n to the origin. Then

$$\mathbb{E}p \sim \sqrt{\frac{2}{\pi}} \frac{1}{n}.$$

Proof: By the sequential construction of Miles we know that

$$\mathbb{E}p = \int_0^1 x \frac{2}{B(\frac{1}{2}, \frac{1}{2}(n-1)^2)} (1-x^2)^{\frac{n^2-2n-1}{2}} dx.$$

Since we have the identity

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

for the beta function we get that

$$\begin{aligned} \mathbb{E}p &= \frac{2\Gamma(\frac{1}{2}(n^2-2n)+1)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(n-1)^2)} \int_0^1 x(1-x^2)^{\frac{n^2-2n-1}{2}} dx \\ &= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n^2-2n}{2})\frac{n^2-2n}{2}}{\Gamma(\frac{1}{2}(n-1)^2)} \int_0^1 2x(1-x^2)^{\frac{n^2-2n-1}{2}} dx, \end{aligned}$$

where we used that $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. By a change of variables we obtain

$$\begin{aligned} \mathbb{E}p &= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n^2-2n}{2})\frac{n^2-2n}{2}}{\Gamma(\frac{1}{2}(n-1)^2)} \int_0^1 y^{\frac{n^2-2n-1}{2}} dy \\ &= \frac{1}{\sqrt{\pi}} \frac{n^2-2n}{(n-1)^2} \frac{\Gamma(\frac{n^2-2n}{2})}{\Gamma(\frac{(n-1)^2}{2})}. \end{aligned}$$

It follows by Stirling's formula that

$$\begin{aligned} \mathbb{E}p &\sim \frac{1}{\sqrt{\pi}} \frac{n^2-2n}{(n-1)^2} \frac{\sqrt{\frac{4\pi}{n^2-2n}} \left(\frac{n^2-2n}{2e}\right)^{\frac{n^2-2n}{2}}}{\sqrt{\frac{4\pi}{(n-1)^2}} \left(\frac{(n-1)^2}{2e}\right)^{\frac{(n-1)^2}{2}}} \\ &= \frac{1}{\sqrt{\pi}} \frac{n^2-2n}{(n-1)^2} \sqrt{\frac{(n-1)^2}{n^2-2n}} \left(\frac{n^2-2n}{(n-1)^2}\right)^{\frac{n^2-2n}{2}} \left(\frac{2e}{(n-1)^2}\right)^{\frac{1}{2}} \\ &\sim \sqrt{\frac{2}{\pi}} \frac{1}{n-1} \\ &\sim \sqrt{\frac{2}{\pi}} \frac{1}{n}. \end{aligned}$$

□

In [23, page 11 ff] H. Ruben and R. E. Miles showed that if $X_i = (a_i g_{i,j})_{j \leq n}$ where $a_1, \dots, a_n \in \mathbb{R} \setminus \{0\}$ is a sequence of reals and $g_{i,j}$ ($i, j \leq n$) are independent standard normal distributed random variables we have the following equality in distribution

$$|\text{vol}_{n-1}(\text{conv}\{X_1, \dots, X_n\})| \stackrel{\mathfrak{D}}{=} \frac{1}{(n-1)!} \left(\prod_{i \leq n} a_i^2 \sum_{i \leq n} \frac{1}{a_i^2} \right)^{\frac{1}{2}} \prod_{i=1}^{n-1} \chi_{n-i+1},$$

where $\chi_{n-i+1} \stackrel{\mathfrak{D}}{=} \sqrt{\sum_{j \leq n-i+1} g_j^2}$ (the g_j denote i.i.d. standard normal distributed random variables) denotes the distribution and the $n-1$ χ -variates are independent. In particular, since

$$\prod_{i \leq n} s_i(X) = \det(X) \stackrel{\mathfrak{D}}{=} (n-1)! |\text{vol}_{n-1}(\text{conv}\{X_1, \dots, X_n\})|,$$

where X denotes the $n \times n$ random matrix whose i -th column is given by X_i , we have a direct connection to the singular values. It would be nice, to find a geometric meaning of the singular values of X , especially of the smallest singular, as a quantity in the random simplex $\text{conv}\{X_1, \dots, X_n\}$.

The following theorem is stated for Gaussian random vectors since this formulation relieves the computation. But one sees by Lemma 5 that there is an analogue formulation for unit vectors chosen uniformly at random.

We are interested in an estimate for the inradius $r_{\Delta_{n-1}}$ (i.e. $r_{\Delta_{n-1}}$ is the radius of the biggest ball completely contained in Δ_{n-1}) of an random simplex Δ_{n-1} .

Theorem 19 *Let X_1, \dots, X_n be independent random vectors whose entries are i.i.d. standard normal distributed random variables. We denote the almost surely existing $(n-1)$ -dimensional random simplex with vertices X_i by $\Delta_{n-1} = \text{conv}\{X_1, \dots, X_n\}$. Then we get for the inradius $r_{\Delta_{n-1}}$ of Δ_{n-1} that*

$$\mathbb{P}(r_{\Delta_{n-1}} < \varepsilon) \geq 1 - 2\sqrt{\frac{2}{\pi}} \exp\left(-\frac{\varepsilon^2 n^2}{8}\right).$$

Proof: We denote the faces of Δ_{n-1} by F_1, \dots, F_n , where for all $i \in \{1, \dots, n\}$ F_i denotes the face containing all but the i -th vertex X_i . Since we know that the volume of an n -dimensional simplex with base B and height h is equal to $\frac{1}{n} \text{vol}_{n-1}(B)h$ it follows that

$$\text{vol}_{n-1}(\Delta_{n-1}) = \frac{r_{\Delta_{n-1}}}{n-1} \sum_{i \leq n} \text{vol}_{n-2}(F_i)$$

since one can decompose Δ_{n-1} into n random simplices with bases F_1, \dots, F_n and height $r_{\Delta_{n-1}}$. Therefore, we get that

$$\begin{aligned} \mathbb{P}(r_{\Delta_{n-1}} < \varepsilon) &= 1 - \mathbb{P}(r_{\Delta_{n-1}} \geq \varepsilon) \\ &= 1 - \mathbb{P}\left(\frac{1}{r_{\Delta_{n-1}}} \leq \frac{1}{\varepsilon}\right) \\ &= 1 - \mathbb{P}\left(\frac{1}{n-1} \sum_{i \leq n} \frac{\text{vol}_{n-2}(F_i)}{\text{vol}_{n-1}(\Delta_{n-1})} \leq \frac{1}{\varepsilon}\right) \\ &= 1 - \mathbb{P}\left(\frac{1}{n-1} \sum_{i \leq n} \frac{\text{vol}_{n-2}(F_i)}{\text{vol}_{n-2}(F_i) \frac{1}{n-1} \text{dist}(X_i, H_i)} \leq \frac{1}{\varepsilon}\right), \end{aligned}$$

where $H_i = \text{span}\{X_j \mid j \neq i\}$. We obtain

$$\begin{aligned} \mathbb{P}(r_{\Delta_{n-1}} < \varepsilon) &= 1 - \mathbb{P}\left(\sum_{i \leq n} \frac{1}{\text{dist}(X_i, H_i)} \leq \frac{1}{\varepsilon}\right) \\ &= 1 - \mathbb{P}\left(\frac{1}{n} \sum_{i \leq n} \frac{1}{\text{dist}(X_i, H_i)} \leq \frac{1}{\varepsilon n}\right) \\ &\geq 1 - \frac{2}{n} \sum_{i \leq n} \mathbb{P}\left(\frac{1}{\text{dist}(X_i, H_i)} \leq \frac{2}{\varepsilon n}\right), \end{aligned}$$

where we used Lemma 18. By Lemma 6 we get

$$\begin{aligned} \mathbb{P}\left(r_{\Delta_{n-1}} < \varepsilon\right) &\geq 1 - 2\mathbb{P}(\text{dist}(X_1, H_1) \geq \frac{\varepsilon n}{2}) \\ &= 1 - 2\mathbb{P}\left(|g| \geq \frac{\varepsilon n}{2}\right), \end{aligned}$$

where g is a standard normal distributed random variable. It follows that

$$\begin{aligned}\mathbb{P}(r_{\Delta_{n-1}} < \varepsilon) &\geq 1 - 2 \int_{\frac{\varepsilon n}{2}}^{\infty} \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &\geq 1 - 2\sqrt{\frac{2}{\pi}} \exp\left(-\frac{\varepsilon^2 n^2}{8}\right).\end{aligned}$$

□

Chapter 7

Bibliography

- [1] Z. D. Bai, Jack W. Silverstein and Y. Q. Yin. *A note on the largest eigenvalue of a large-dimensional sample covariance matrix*. J. Multivariate Anal., 26(2), pages 166-168, 1988.
- [2] A. Bandeira and R. van Handel. *Sharp nonasymptotic bounds on the norm of random matrices with independent entries*. Annals of Probability, to appear, 2014.
- [3] R. H. Berk. *Biorthogonal and Dual Configurations and the Reciprocal Normal Distribution*. Ann. Math. Statist. 40 (1969), no. 2, pages 393-398, 1969.
- [4] C. Buchta, J. Müller and R. F. Tichy. *Stochastic approximation of convex bodies*. Math. Ann. 271, pages 225-235, 1985.
- [5] D. Chafaï. *Singular values of random matrices*. djalil.chafai.net/docs/sing.pdf, 22.10.2015.
- [6] S. Chevet. *Séries de variables aléatoires gaussiennes a valeurs dans $E \hat{\otimes} \varepsilon F$ Application aux produits d'espaces de Wiener abstraits*. In Séminaire sur la Géométrie des Espaces de Banach (1977-1978), pages Exp. No. 19, 15. École Polytech., Palaiseau, 1978.
- [7] A. Edelman. *Eigenvalues and condition numbers of random matrices*. SIAM Journal on Matrix Analysis and Applications, 9(4), 543-560, 1988.
- [8] A. Edelman. *Eigenvalues and Condition Numbers of Random Matrices*. MIT PhD Dissertation(4), 1989.
- [9] B. Farrell, R. Vershynin. *Smoothed analysis of symmetric random matrices with continuous distributions*. Proceedings of the AMS 144 (2016), pages 2257-2261, 2016.
- [10] O. Friedland, O. Giladi. *A simple observation on random matrices with continuous diagonal entries*. Electronic Communications in Probability, 18(53), pages 1-7, 2013.
- [11] Y. Gordon, A. Litvak, C. Schütt and E. Werner. *On the minimum of several random variables*. Proceedings of the AMS, 134 (2006), 3665-3675, 2006.
- [12] N. Johnson, S. Kotz and N. Balakrishnan. *Continuous Univariate Distributions*. Vol. 2, John Wiley & Sons, New York, 1995.
- [13] J. Kahn, J. Komlós, E. Szemerédi. *On the probability that a random ± 1 -matrix is singular*. J. Amer. Math. Soc. 8 (1995), no. 1, 223-240, 1995.
- [14] J. Kallsen *Lecture Notes: Risk Management*. www.math.uni-kiel.de/finmath/de/material/dateien/riskmanagement, 26.07.2016.

- [15] J. Komlós. *On the determinant of $(0,1)$ matrices*. Studia Sci Math. Hungar. 2, pp. 7-21, 1967.
- [16] R. Latała. *Some estimates of norms of random matrices*. Proc. Amer. Math. Soc. 133, 1273-1282, 2005.
- [17] M. Ledoux. *The concentration of measure phenomenon*. Mathematical Surveys and Monographs 89, American Mathematical Society, 2001.
- [18] M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Volume 23 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], Springer-Verlag, Berlin, 1991. Isoperimetry and processes.
- [19] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann. *Smallest singular value of random matrices and geometry of random polytopes*. Adv. Math. 195, pages 491-523, 2005.
- [20] R. E. Miles. *Isotropic random simplices*. Advances in Appl. Probability 3, pp. 353-382, 1971.
- [21] G. Pisier. *The volume of convex bodies and Banach space geometry*. volume 94 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1989.
- [22] S. Riemer, C. Schütt. *On the expectation of the norm of random matrices with non-identically distributed entries*. Electron. J. Probab., 18:no. 29, 13, 2013.
- [23] H. Ruben and R. E. Miles. *A canonical decomposition of the probability measure of sets of isotropic random points in \mathbb{R}^n* . J. Multivariate Anal. 10(1), pages 1-18, 1980.
- [24] J. S. Müller. *Approximation of the ball by random polytopes*. Journal of Approximation Theory 63, pages 198-209, 1990.
- [25] S. Riemer. *Größte singuläre Zahl von Zufallsmatrizen: Beiträge zur Abschätzung der Größenordnung durch Orlicz-Normen*. Kiel, Univ., Diss., 2011.
- [26] M. Rudelson. *Recent developments in non-asymptotic theory of random matrices*. arXiv:1301.2382[math.PR], 2013.
- [27] M. Rudelson and R. Vershynin. *The least singular value of a random square matrix is $O(n^{-1/2})$* . C. R. Acad. Sci. Paris, Ser. I 346 (2008).
- [28] M. Rudelson and R. Vershynin. *The Littlewood-Offord problem and invertibility of random matrices*. Adv. Math., 218(2): 600-633, 2008.
- [29] M. Rudelson, R. Vershynin. *The smallest singular value of a rectangular random matrix*. Comm. Pure Appl. Math. 62 (2009), no 12, pages 1707-1739, 2009.
- [30] M. Rudelson and R. Vershynin. *The smallest singular value of a rectangular random matrix*. Comm. Pure Appl. Math. 62, no. 12, 1707-1739, 2009.
- [31] M. Rudelson and O. Zeitouni. *Singular values of Gaussian matrices and permanent estimators*. to appear in Random Structures and Algorithms.
- [32] A. Sankar, D. A. Spielman and S.-H. Teng. *Smoothed Analysis of the Condition Numbers and Growth Factors of Matrices*. SIMAX Volume 28, Issue 2, 446-476, 2006.
- [33] L. A. Santaló. *Integral Geometry and Geometric Probability*. Encyclopedia of Mathematics and its Applications, Addison-Wesley, 1976.
- [34] G. Schechtman, J. Zinn. *On the volume of the intersection of two L_p^n balls*. Proc. Amer. Math. Soc. 110, pages 217-224, 1990.

- [35] R. Schneider, W. Weil. *Stochastic and integral geometry*. Springer, Berlin, 2008.
- [36] C. Schütt *Lecture Notes: Funktionalanalysis*. 20.07.2010
- [37] C. Schütt and E. Werner. *Random polytopes of points chosen from the boundary of a convex body*. Gafa Seminar Notes, Lecture Notes in Mathematics 1807, Springer, pages 241-422, 2002.
- [38] D. Spielman and S.-H. Teng. *Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time*. J. ACM, 51 (2004), pp. 385-463, 2004.
- [39] D. W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, 1993.
- [40] S. J. Szarek. *Condition numbers of random matrices*. J. Complexity, 7(2), 131-149, 1991.
- [41] T. Tao. *Topics in random matrix theory*. American Mathematical Society, 132 (2012), 2012.
- [42] T. Tao, V. Vu. *On random ± 1 matrices: Singularity and Determinant*. Random Structures and Algorithms 28, pages 1-23, 2006.
- [43] T. Tao and V. Vu. *Random matrices: the distribution of the smallest singular values*. Geom. Funct. Anal. 20 (2010), no. 1, 260-297, 2010.
- [44] T. Tao and V. Vu. *Smoothed analysis of the condition number and the least singular value*. Math. Comp. 79, 2333-2352, 2010.
- [45] R. Vershynin. *Lecture Notes: Non-asymptotic Random Matrix Theory*. www-personal.umich.edu/~romanv/teaching/2006-07/280/course.html, 05.04.2014.
- [46] R. Vershynin. *Introduction to the non-asymptotic analysis of random matrices*. Compressed sensing, 210-268, Cambridge Univ. Press, Cambridge, 2012.
- [47] J. von Neumann. *Collected works Vol. V: Design of computers, theory of automata and numerical analysis*. General editor: A. H. Taub. A Pergamon Press Book The Macmillan Co., New York 1963.
- [48] 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), pages 509-51, 1988.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit abgesehen von der Beratung durch den Betreuer meiner Promotion, unter Einhaltung guter wissenschaftlicher Praxis selbstständig angefertigt und keine anderen als die angegebenen Hilfsmittel verwendet habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Diese Arbeit hat weder ganz noch zum Teil an anderer Stelle im Rahmen eines Prüfungsverfahrens vorgelegen. Des Weiteren habe ich noch keinen Promotionsversuch unternommen.

Kiel, 27. Juli 2016

(Colin Kleinschmidt)