École Normale Supérieure - PSL
75005, Paris

First-year thesis
L3 Mathematics and applications

Singular values of large random matrices

GRENIER Antoine and **KACHAIKIN** Marko

Supervisor: **CHAFAÏ** Djalil

Departement of
mathematics and applications

Second semester 2024

Contents

Introduction and historical context

Our thesis investigates next natural question:

Take matrix A with independent random entries, what is the probability that A is singular?

Naively, it seems intuitive to focus on the determinant of such a matrix and examine the quantity

$$\mathbb{P}(\det(A) = 0)$$

. While this approach exists, it is not the simplest one for addressing the question of invertibility (there are, in particular, estimates of $\log\det(A)$ when $n \to +\infty$).

One could propose then, to control invertibility by the smallest eigenvalue of A, since obviously $\det(A) = 0 \iff \lambda_{\min}(A) = 0$. The problem is that $\lambda_{\min}$ is a-priori a complex random variable, since A isn't supposed symmetric/hermitian and we can not study it's spectrum directly.

However, notice that $\det(A) = 0 \longleftrightarrow \det(AA^*) = 0$, so invertibility of A is equivalent to invertibility of hermitian matrix AA^*. From a spectral point of view, spectrum of AA^* is much more easier to understand it's a set of real non-negative numbers, they are called singular values, and it's our key to control the invertibility.

We've then decided firstly to revisit this key deterministic concept, defined as we will see for any rectangular matrix. We prove the singular value decomposition theorem and we mention the Courant-Fischer theorem.

We then employ these different deterministic notions in the context of large random matrices. We study the concentration of the largest and smallest singular values.

This study of the smallest singular value is actually fundamental to our initial problem: it characterizes the invertibility of a matrix. Therefore, we devote the final part of this thesis, in the much broader context of sub-Gaussian variables, to it.

Historically, singular values of matrices with random entries were firstly studied in two different contexts:

I) Empirical covariance matrices for random vectors

Classical statistical problem: given a finite sample $X_1, ..., X_n$ of observations, i.e. i.i.d random vectors of dimension d, what conclusions can we draw about the general population?
When the observations are Gaussian, some exact methods exist, for instance ANOVA or Student's test. However in most of applications, observations are non Gaussian and statistical methods are typically built using limit theorems.
Classical limit theorems such as Law of Large Numbers, Central limit theorem, etc. are considering that dimension of data d is fixed, while sample size tends to infinity. It implies particularly, if $\mathbb{E}||X_1||_2^2 < \infty$:

Vectorial central limit theory works:

$$\sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}X_1\Big) \xrightarrow{\mathbf{d}} \mathcal{N}\big(0, \mathbb{E}X_1X_1^*\big) \text{ when } d \text{ - fixed and } n \to \infty$$

This theory is widely accepted and works well in many applications, but it is not a panacea, especially in cases where the dimensionality of the data is huge and can be much larger than the sample size.

Large dimensional data appear in various fields, such as finances, medicine or genomics.

For instance in genomic cancer research, see [6], dimensionality (number of features) is of order 20000 comparing to number of patients of order 800.

However, the most fashionable example it's image processing, especially supervised training of image classifiers or image generators. Every image could be represented as a matrix of pixels, where index represent the location of a pixel and value represent the color (in RGB or White-Black scale). Depending on the image quality and size, that could be a matrix of a huge dimension. However, sample size of training could be relatively small comparing to dimension, for instance see famous Cats and Dogs dataset from Kaggle [3]. Size of dataset is ≈ 2000, but each photo is decoded by ≈ 10000 pixels!

Now we will show the example, see example 1.1 in [25], which is inspsired by work of Humpster, where the classical theory breaks down when the dimensionality of the data and sample size are both huge.

Let us take $X_1, ..., X_n$ - sample of i.i.d random vectors from $\mathcal{N}(0, Id_{\mathbb{R}^d})$.

Convergence a.s. of sample covariance matrix, implies particularly that eigenvalues $(\lambda_{n,j})_{j=1,...,d}$ of positive semi-define matrix $S_n = \frac{1}{n}\sum_{k=1}^{n} X_kX_k^*$ are converging to 1 and so log-det statistics converging to 0:

$$\log(\det(S_n)) = \sum_{j=1}^{d} \log(\lambda_{n,j}) \longrightarrow 0 \text{ **a.s.** when } d \text{ - fixed and } n \to \infty$$

Moreover using the scalar delta-method in vectorial CLT for log-det statistics and sequence of i.i.d random matrices $(X_iX_i^*)_{i\geq 1}$ (who are in fact random vectors in $\mathbb{R}^{d\times d}$ space), one could obtain that:

$$\sqrt{\frac{n}{2d}}\log\det(S_n) \xrightarrow{\mathbf{d}} \mathcal{N}(0,1) \text{ when } d \text{ - fixed and } n \to \infty$$

It may seem then logical to suppose that asymptotic normality of log-det statistics remains true when d is large, namely $d = \mathcal{O}(n)$.

However that's not the case! See the computational illustration of histogram of log-det statistics, for $n = d = 200$ and number of experiments of order 1000.

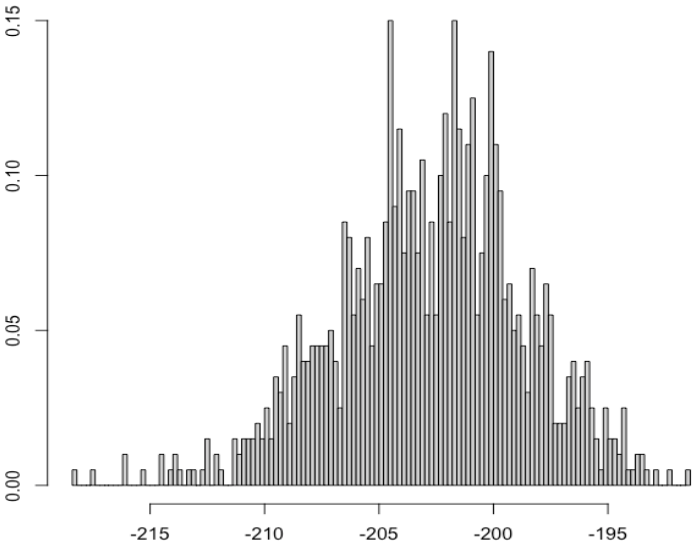


Figure 1: Histogram of log-det statistics

As we can see it's far away from density of standart normal distribution, it's also deviates near the smaller values when n increases, so one can suppose that it goes to $-\infty$ when $n \to \infty$.

This example show that the classical large sample limits are no longer valid for dealing with large dimensional data analysis.

To deal with it, a new area in asymptotic statistics has been developed where the data dimension d is no more fixed but tends to infinity together with the sample size n.

However, first breakthrough in this area was made not by statistians, but physicists from Ukraine - Marchenko and Pastur, who were studying originally the distribution of energy spectrum in disordered quantum systems, see paper [14]. Moreover, they essentially rediscovered the model of large sample covariance matrix.

It was proved by them, that for random matrix $\frac{1}{d}Y_nY_n^*$, where Y_n is a $n \times d$ matrix with i.i.d gaussian entries, which is in fact also sample covariance matrix $\frac{1}{d}\sum_{k=1}^{d} X_kX_k^*$ where $X_1, ..., X_d$ are vector columns of matrix Y_n, boundary of histogramm of eigenvalues of when $n, d \longrightarrow \infty$ and $n/d \to \alpha \in (0, 1]$ converges to deterministic shape given by explicit formula:

$$MP_\alpha(x) = \frac{\sqrt{((1+\sqrt{\alpha})^2 - x)(x - (1-\sqrt{\alpha})^2)}}{2\pi\alpha x}$$

That explains the phenomenon, that we've seen above, in fact when $n, d \longrightarrow \infty$ and $n/d \to \alpha \in (0, 1)$ (we choose $\alpha \neq 1$, so that $(1-\sqrt{\alpha})^2 \neq 0$) for continious bounded function log on interval $[(1-\sqrt{\alpha})^2, (1+\sqrt{\alpha})^2]$, next statistics (asymptotically !) holds:

$$\frac{1}{d}\log\det(S_n) \to \int_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} \log x \frac{\sqrt{((1+\sqrt{\alpha})^2 - x)(x - (1-\sqrt{\alpha})^2)}}{2\pi\alpha x} dx = \frac{\alpha - 1}{\alpha}\log(1-\alpha) - 1 < 0$$

And so, we conclude that as $n, d \to \infty$, and $n/d \to \alpha \in (0, 1)$ almost surely we have:

$$\sqrt{\frac{n}{d}}\log\det(S_n) \longrightarrow -\infty$$

Which is radically different from our assumption, that was written above, and explains figure 1 and so why log-det statistics goes to $-\infty$.

Moreover it gives us the clue about the concentration of the extremal eigenvalues, and so singular values near the edges. In fact we can prove that when when $n, d \longrightarrow \infty$ and $n/d \to \alpha \in (0, 1)$:

$$s_n(d^{-1/2}Y_n) \to (1-\sqrt{\alpha}) \text{ and } s_1(d^{-1/2}Y_n) \to (1+\sqrt{\alpha}) \textbf{ a.s.}$$

However, this result won't give us the information about the character of concentration (sub-gaussian, sub-exponential, etc.), we will see it later.

II) Condition number for random matrices.

In the end of 1940's, with the advent of the first computers, Team of John Von Neumann used random matrices as inputs for testing numerical algorithms for approximate solution of systems of linear equations.

The accuracy of the matrix algorithms, and sometimes their running time as well, as we will see, highly dependent on the extremal singular values and their proportion.

Based on heuristics, John von Neumann and Herman Holdstine conjectured in paper [7], that for matrix A of size $n \times n$ with independent random (gaussian) entries :

$$s_n(A) = \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big) \qquad \text{and} \qquad s_1(A) = \mathcal{O}\big(\sqrt{n}\big)$$

In next section we will see that the result for $s_{\max}$ could be obtained using standart ε-net argument for much more bigger class of random matrices.

However estimating s_n turned out to be very difficult. It was firsly proved by Alan Edelman in [4] that for matrices with i.i.d real/complex gaussian variables, it was done by the following procedure:

Let A_β be a $n \times n$ matrix with (real if $\beta = 1$, complex if $\beta = 2$) i.i.d standart gaussian variables.

For matrix $A_\beta A_\beta^*$, joint probability density functions f_β of ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$ are given by:

$$f_\beta\big(\lambda_1, ..., \lambda_n\big) = \frac{\mathbf{1}_{\lambda_1 \geq ... \geq \lambda_n \geq 0}}{\mathcal{Z}_{n,\beta}} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{k=1}^n \lambda_k^{\beta/2-1} \exp(-\beta\lambda_k/2)$$

Integrating over the all eigenvalues except the smallest one we could obtain, that:

For $\beta = 1$:

$$\mathbb{P}(s_n(A_\beta) \leq tn^{-1/2}) \to (1 - e^{-t^2/2-t})\mathbf{1}_{t\geq 0}, \quad n \to \infty$$

For $\beta = 2$:

$$\mathbb{P}(s_n(A_\beta) \leq tn^{-1/2}) = (1 - e^{-t^2})\mathbf{1}_{t\geq 0}, \quad \text{i.e.} \quad s_n^2(A_\beta) \sim \frac{1}{n}\mathcal{E}(1)$$

In principle it gives us that for small enough $t \geq 0$:

$$\mathbb{P}\big(s_n(A_\beta) \leq tn^{-1/2}\big) \leq t, \quad n \geq 1 \tag{1}$$

In fact result is true for all $t \geq 0$, it's trivial for $\beta = 2$ from numerical inequality $1 - \exp(-x^2) \leq x$ with $x \geq 0$, and for $\beta = 1$ see again [4] or for geometric proof [18].

It's evident that these methods do not work for general random matrices, especially those with discrete distributions, where rotation invariance and the joint density of eigenvalues are not available.

However, in parallel with studying of gaussian random matrices, theory for smallest singular value of random matrices with i.i.d Bernoulli entries were developed.

Let $M_n = (M_{i,j})_{i,j=1,...,n}$ - be $n \times n$ matrix with i.i.d entries distributed with respect to Rademacher distribution:

$$\mathbb{P}(M_{i,j} = \pm 1) = 1/2$$

It seems to be an old problem, to find asymptotics of $\mathbb{P}(M_n \text{ is singular})$.

It's almost obvious, that this probability is larger than proba that two columns/raws are equal:

$$\mathbb{P}(M_n \text{ is singular}) \geq 2\binom{n}{2}2^{-n} = \Big(\frac{1}{2} + o(1)\Big)^n$$

It was conjectured that inequality from above in fact is equality, see for instance talk of V.Vu on ICM 2014, in fact it was one of the main problem and oldest problem in combinatorial random matrix theory.

J. Komlós showed that $\mathbb{P}(M_n \text{ is singular}) = o_n(1)$, see paper [12]. Much later, the bound $\mathbb{P}(M_n \text{ is singular}) \leq (0.999)^n$ was obtained by J. Kahn, J. Komlós and E. Szemerédi in [11]. The upper bound was sequentially improved to $(3/4 + o_n(1))^n$ in paper by T. Tao and V.Vu [21], then to $(1/\sqrt{2} + o_n(1))^n$ in [2] and finally conjecture was proved by Tikhomirov in [23].

Spielman and Teng conjectured in their ICM 2002 talk [19], that estimate (1) should hold for the random Bernoulli matrices up to an exponentially small term that accounts for their singularity probability ν :

$$\mathbb{P}\big(s_n(M) \leq tn^{-1/2}\big) \leq Ct + \nu^n, \quad n \geq 1$$

In paper [16], Rudelson and Vershynin showed that for matrix X from much larger ensemble with i.i.d entries, this inequality holds:

$$\mathbb{P}\big(s_n(X) \leq tn^{-1/2}\big) \leq \underbrace{Ct}_{\text{optimal for gaussian}} + \underbrace{c^n}_{\text{optimal for discrete}}, \quad n \geq 1$$

Where constants $C > 0$ and $c \in (0, 1)$ depends only on distribution of entries, and not t or n.

Our thesis will be dedicated to demonstrate this result of Rudelson and Vershynin.

1 Singular value decomposition (SVD)

1.1 SVD, Courant-Fisher

In this first part, we introduce the main deterministic quantities associated with a matrix that will be useful to us later on.

Throughout, let A be an arbitrary $m \times n$ matrix with complex coefficients. We start with the fundamental concept of the singular value of a matrix, which we define through the singular value decomposition (SVD).

In this thesis, if $x \in \mathbb{R}^n$, $\|x\|_2$ will denote the euclidean norm of x and if $A \in M_{n,m}(\mathbb{R})$, $\|A\|$ will denote the norm of A subordinate to the euclidean norm on $\mathbb{R}^n$ and $\mathbb{R}^m$. The scalar product of two vectors x and $y \in \mathbb{R}^n$ will be denoted by $\langle x, y \rangle$.

Theorem 1.1 (Singular value decomposition)**.** *Let A be an $m \times n$ matrix with complex coefficients and let r be the rank of A. Thus there exists a strictly decreasing sequence $(s_1 \dots s_r)$ of positive real numbers and two unitary matrix $P \in U_m(\mathbb{C})$ and $Q \in U_n(\mathbb{C})$ such that*

$$M = PDQ$$

where $D = (d_{ij})$ is the matrix where

$$d_{ij} = \begin{cases} s_i & \text{if } 1 \leq i = j \leq r \\ 0 & \text{otherwise} \end{cases}$$

Proof. We consider the case where the matrix has real coefficients. The proof easily adapts in the general case. We adopt the equivalent geometric viewpoint by constructing two appropriate bases for an endomorphism $u : E \to F$ where E and F F are two Euclidean spaces of dimension n and m.

We denote as $\langle , \rangle_E$ and $\langle , \rangle_F$ the inner product on E and F and we denote u^* as the adjoint of u. For any real λ we define:

- $E_\lambda = \mathrm{Ker}(u^*u - \lambda id_E)$
- $F_\lambda = \mathrm{Ker}(uu^* - \lambda id_F)$

These are subspaces of vector spaces E and F. Firstly we denote that the endomorphism u^*u is self-adjoint and so is diagonalizable in an orthonormal basis. We denote as $\lambda_1 \dots \lambda_r$ the non-zero eigenvalues of u^*u counted with multiplicity.

$\forall i, \lambda_i > 0$ because if v is an eigenvector associated to λ_i we have $\lambda_i \|x\|^2 = \lambda_i \|u(x)\|^2$ using the definition of λ_i.

It is clear that if λ is a positive real, $u(E_\lambda) \subset F_\Lambda$ and $u^*(F_\lambda) \subset E_\lambda$. We denote $u_\lambda : E_\lambda \to F_\lambda$ the restriction and corestriction of u. Show that $\frac{1}{\sqrt{\lambda}} u_\lambda$ is an isometry.

The fact that $\|u_\lambda(x)\| = \|x\|$ is clear with the definitions: we must show that E_λ and F_λ have the same dimension. On E_λ, u_λ is isometric so injective and so $\mathrm{Dim}(E_\lambda) \leq \mathrm{Dim}(F_\lambda)$. y using u^* we prove the result.

Finally we use the spectral theorem: let $(e_1 \dots e_n)$ be an orthonormal basis of E which diagonalize u^*u. We suppose that $e_1 \dots e_r$ are associated to the non-zero eigenvalues. Because $\frac{1}{\sqrt{\lambda_i}} u_{\lambda_i}$ is isometric on E_{λ_i} for any λ_i the family $(f_1 = \frac{1}{\sqrt{\lambda_1}} u(e_1) \dots f_r \frac{1}{\sqrt{\lambda_r}} u(e_r))$ is linearly indepent and orthonormal. We complete this family with vectors $(f_{r+1} \dots f_n)$ such that $(f_1 \dots f_n)$ is an orthonormal basis of F.

Between the basis e and f, u is a diagonal endomorphism (because $\mathrm{Ker}(u^*u) = \mathrm{Ker}(u)$). It concludes the proof of the theorem.

□

Definition 1.1 (singular values)**.** *Real numbers $s_1 \geq \dots \geq s_n$ (and $s_k = 0$ for $k > r$) in Theorem 1.1 are called the singular values of A.*

Remark 1.1. *We have shown in the proof, that singular values of rectangular matrix A are in fact square roots of the eigenvalues of matrix A^*A or AA^* (up to multiplicity of 0 that changes for each matrix).*

We then remark that depending on our needs, we will speak about the eigenvalues of AA^* instead of singular values of A, which initially describe the same objects, up to an application of some function.

We mention here the Courant-Fischer theorem, which provides a variational characterization of the singular values of a matrix.

Theorem 1.2 (Courant-Fischer)**.** *Let A be an $m \times n$ with $n \leq m$ and $s_1 \geq \dots \geq s_n$ the singular values of A. One has*

$$s_k = \max_{V \in G_k} \min_{\substack{x \in V \\ \|x\|=1}} \|Ax\|_2 = \min_{W \in G_{n-k+1}} \max_{\substack{y \in W \\ \|y\|=1}} \|Ax\|_2$$

where G_k is the set of the subspace of $\mathbb{C}^n$ with dimension k

We will not prove this theorem in this paper because this famous formula is not very necessary for the following developments. Notice that for $k = 1$ or $k = n$ we find

$$s_1 = \max_{\|x\|=1} \|Ax\|_2 \quad \text{and} \quad s_n = \min_{\|x\|=1} \|Ax\|_2$$

.

This last formula is important: it shows that, if A is now an $n \times n$ matrix, A is singular only if $s_n = 0$. The problem of invertibility of a random matrix is so equivalent to the problem of the distribution of the smallest singular value. Moreover we have

$$|\det(A)| = \prod_{i=1}^{n} s_i.$$

We understand so why singular values are important: naively, one might have thought that the determinant would be the preferred tool for studying the invertibility of random matrices (thanks to its explicit expression). However, here we see that knowing the distributions of s_1 and s_n provides an bound on the determinant and vice versa.

We conclude this deterministic part with a purely computational lemma (see [1]), which provides an initial control of s_n and will later be used in the intuition to proof of the theorem by Rudelson and Vershynin.

Lemma 1.1 (Rows and operator norm of the inverse)**.** *Let A be a complex $n \times n$ matrix with rows $R_1, \dots, R_n$. Define the vector space $R_{-i} := \text{span}\{R_j : j \neq i\}$. We have then*

$$n^{-1/2} \min_{1 \leq i \leq n} \text{dist}(R_i, R_{-i}) \leq s_n(A) \leq \min_{1 \leq i \leq n} \text{dist}(R_i, R_{-i}).$$

1.2 Geometry and Computer Science

In this section, we provide some examples of applications of the concept of singular value.

We begin by providing a geometric interpretation. Let's consider a square matrix M in $M_n(\mathbb{R})$ and observe how it deforms the unit sphere in $\mathbb{R}^n$. The SVD decomposition provides two orthonormal bases of $\mathbb{R}^n$, given by columns of orthogonal (in complex case - unitary) matrices U, V, between which M is diagonal matrix Σ whose entries are exactly singular values of M, in sense that:

$$M = U\Sigma V^*$$

And so by changing the basis, one can assume M to be diagonal, with the coefficients being the singular values.

Let $x = (x_1, \dots, x_n) \in \mathbb{S}^{n-1}$ and $(x'_1, \dots, x'_n) = Mx$. We have

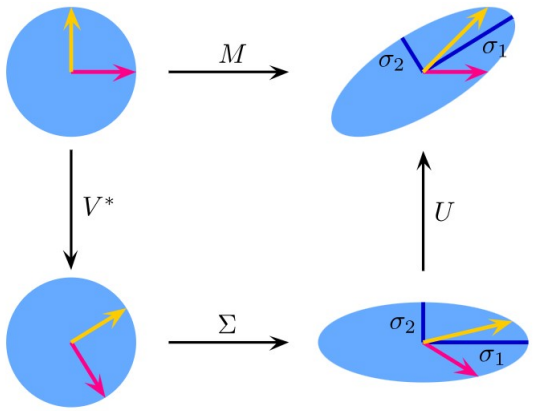
$$\frac{x_1'^2}{s_1^2} + \dots \frac{x_n'^2}{s_n^2} = 1$$

.

This proves that the matrix M deforms the sphere into an ellipse, and further into an ellipsoid, whose directions are given by the change of basis matrices and whose semi-axes are precisely the singular values of M.

Using the expressions of s_1 and s_n, we find that these quantities correspond respectively to the largest and smallest semi-axes of the ellipsoid.

Here is an illustration of this observation in the case where $n = 2$.



$$M = U \cdot \Sigma \cdot V^*$$

Let's now give a second concrete application: the pseudoinverse of a matrix $A \in M_{m,n}(\mathbb{R})$ (see [13]).

Definition 1.2. *Suppose A is an $m \times n$ matrix and the singular-value decomposition of A is given by PDQ. Pseudoinverse of A is a matrix A^+, which we define as:*

$$A^+ = Q^T D^+ P^T$$

where entries of diagonal matrix D^+ are given by $(D^+)_{i,i} = \frac{\mathbf{1}_{\{D_{i,i} \neq 0\}}}{D_{i,i}}$

We state without proof the following theorem, which illustrates the usefulness of the concept of pseudo-inverse.

Theorem 1.3 (Least Square Problem). *Suppose given A and b, where A is a $m \times n$ matrix and $b \in \mathbb{R}^m$, we want to find x such that Ax is closest to b. In order words, find x such that*

$$Ax = \min_{y \in \mathbb{R}^n} \|Ay - b\|_2$$

x is called the least square solution to the equation $Ax = b$.
The theorem then tells us that such an x exists and is explicitly given by

$$x = A^+ b$$

where A^+ is the pseudoinverse of A.

We thus see that the singular values of a matrix are important in the approximate solution of systems of the form $Ax = b$ We will deepen this idea with the notion of matrix conditioning (see [10]).
Let's start with an example. Consider solving the linear system $Ax = b$ with

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \text{ and } b = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

The solution to this system is $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$
Now, let's add a small perturbation δb to the right-hand side, resulting in the linear system

$$Ax_p = b + \delta b \text{ with } b = \begin{pmatrix} 0.01 \\ -0.01 \\ 0.01 \\ -0.01 \end{pmatrix}$$

The solution to this new system is $x_p = \begin{pmatrix} 1.82 \\ 0.36 \\ 1.35 \\ 0.79 \end{pmatrix}$ and is very different from x. Although the relative error on b which is $\frac{\|\delta b\|_2}{\|b\|_2} = 3.31910^{-4}$, is small, the relative error on x which is $\frac{\|x - x_p\|_2}{\|x\|_2} = 2460$!

Definition 1.3 (condition number of a matrix). *The condition number of an invertible matrix $A \in M_n(\mathbb{R})$ denoted by*

$$Cond(A) = \|A\|_s \|A^{-1}\|_s$$

where $\|\|_s$ is a subordinate matrix norm.

We immediately notice that Cond$(A) \geq 1$ and that if $s = 2$ in the definition we have

$$\text{Cond}(A) = \frac{s_1}{s_n}$$

We can measure the sensitivity of the linear system $Ax = b$ to perturbations in the data (i.e., perturbations in b or A) by the following theorem, that we will admit (see in the proof in [10])

Theorem 1.4. *Let $A \in M_n(\mathbb{R})$ be invertible and $b \in \mathbb{R}^n$, $b \neq 0$ given. Let $\delta b \in \mathbb{R}^n$ be a perturbation of b. If x and x_p are respectively solutions of $Ax = b$ and $Ax_p = b + \delta b$, then we have the inequality:*

$$\frac{\|x - x_p\|}{\|x\|} \leq cond(A) \frac{\|\delta b\|}{\|b\|}$$

and it is the best possible: for a given A, $b \neq 0$, and $\delta b \neq 0$ such that we have equality...

2 Singular values and invertibility of random matrices

After presenting various general properties of singular values in the deterministic case, our goal now is to use these concepts to prove the theorem of Rudelson and Vershynin mentioned in the introduction. We begin by summarizing the properties of sub-Gaussian variables, whose assumptions align with the theorem of Rudelson and Vershynin. Next, we show that the norm of a sub-Gaussian matrix can be controlled exponentially (a result that will be useful later), and finally, we prove the theorem of Rudelson and Vershynin.

2.1 Sub-Gaussian random variables

In this section we give the definition of sub-gaussian random variables in case when the mean of variable is zero and then provide their properties.

Definition 2.1 (centered sub-gaussian random variables)**.** *Let X be a real random variable with mean zero $\mathbb{E}X = 0$ We say that X is **sub-gaussian** if there exists $K > 0$ such that Laplace transform of X, for all $\lambda \in \mathbb{R}$ verifies:*

$$\mathbb{E}\exp\Big(\lambda X\Big) \leq \exp(\lambda^2 K^2)$$

Infimum over the all K so that inequality above holds is called the subgaussian moment of X and denoted by $\|X\|_{\Psi_2}$.

Remark 2.1. *Note that this definition encompasses both normal $\mathcal{N}(0,1)$ and discrete random variables with finite support and zero mean (Rademacher - $\frac{1}{2}(\delta_1 + \delta_{-1})$) , as well as many others, and so is very general.*

We subsequently provide various global properties of this type of random variables.

Theorem 2.1 (properties of centered subgaussian random variables)**.**

1. *Let X - be centered ($\mathbb{E}X = 0$) subgaussian r.v. with subgaussian moment $K := \|X\|_{\Psi_2}$, then for all $t \geq 0$:*

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{4K^2}\right) \qquad \textit{Gaussian tails}$$

2. *Let $X_1, ..., X_n$ - be centered **independent** subgaussian random variables, **not necessarily from the same law**, $a = (a_1, ..., a_n) \in \mathbb{R}^n$ - fixed deterministe vector, then for all $t \geq 0$:*

$$\mathbb{P}\left(\left|\sum_{k=1}^n a_i X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{4||a||_2^2 \max\limits_{1\leq k\leq n} ||X_i||_{\psi_2}^2}\right) \qquad \textit{Hoeffding's inequality}$$

Proof. *1.* For each $\alpha > 0, t \geq 0$ using Markov's inequality and monotonicity of $x \to e^{\alpha x}$:

$$\mathbb{P}(X \geq t) = \mathbb{P}\left(e^{\alpha Z} \geq e^{\alpha t}\right) \leq e^{-\alpha t}\mathbb{E}\left(e^{\alpha Z}\right) \leq \exp(\alpha^2 K^2 - \alpha t)$$

Optimizing for $\alpha > 0$ and t - fixed, we get that for each $t \geq 0$:

$$\mathbb{P}(Z \geq t) \leq \inf_{\alpha>0}\left(e^{\alpha^2 K^2 - \alpha t}\right) = e^{\alpha^2 K^2 - \alpha t}\Big|_{\alpha = t/2K^2} = \exp(-t^2/4K^2)$$

We can redo the same proof for the variable $-X$, so $\mathbb{P}(X \leq -t) \leq e^{-t^2/4K^2}$, from which we finally obtain:

$$\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/4K^2}$$

2. It's enough to prove that $\sum\limits_{k=1}^n a_i X_i$ - is a subgaussian random variable, with subgaussian moment that verifies:

$$\left|\left|\sum_{k=1}^n a_i X_i\right|\right|_{\psi_2} \leq ||a||_2^2 \max_{1\leq k\leq n} ||X_i||_{\psi_2}^2$$

Then first part of the theorem will imply the statement. By independence of the variables, we could write:

$$\mathbb{E}\exp\Big(\lambda\sum_{k=1}^n a_i X_i\Big) = \mathbb{E}\prod_{k=1}^n \exp\Big(\lambda a_i X_i\Big) = |\text{Independence}| = \prod_{k=1}^n \mathbb{E}\exp\Big(\lambda a_i X_i\Big)$$

$$\leq \prod_{k=1}^n \exp(c\lambda^2 a_i^2) = \exp\Big(c\lambda^2 \sum_{i=1}^n a_i^2 ||X_i||_{\psi_2}^2\Big) \leq \exp\Big(c\lambda^2 ||a||_2^2 \max_{1\leq k\leq n} ||X_i||_{\psi_2}^2\Big)$$

It's obvious that by linearity $\mathbb{E}\left[\sum\limits_{k=1}^n a_i X_i\right] = 0$, now definition of subgaussian r.v. and part 1 concludes the statement.

□

Remark 2.2. *One could prove that inverse statement is still true, namely - zero-mean r.v. with gaussian tails is always sub-gaussian. In general case with non-zero mean sub-gaussian random variable is defined as a r.v. with gaussian tails, see parts 2.5 and 2.6 of [24].*

2.2 Concentration of the largest singular value

Now, we will investigate the concentration of the largest singular value in general ensemble of square matrices with independent subgaussian entries.

This result is also utilized in proving the concentration of the smallest singular value.

Theorem 2.2. ***(Concentration of the largest singular value)***

Let $A = (\xi_{i,j})$ - be $n \times n$ matrix, whose entries $\xi_{i,j}$ are ***indepdendent*** *centered subgaussian random variables (not neccesarily from the same law).*

Then, with constant $C > 0$ that depend only on subgaussian norms of entries:

$$\mathbb{P}\big(s_1(A) \geq Cn^{1/2}\big) \leq 2e^{-n} \tag{2}$$

Some observations and preliminary lemmas before the proof are necessary.

We firstly note, that from the straighforward inequailities

$$s_1(A) = \max_{x,y \in \mathbb{S}^{n-1}} |\langle Ax, y\rangle| \quad \textit{where} \quad \mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : ||x||_2 = 1\}$$

For any fixed $x, y \in \mathbb{S}^{n-1}$: $\langle Ax, y\rangle = \sum\limits_{i,j=1}^{n} \xi_{i,j} x_i y_j$ - sum of independent subgaussian r.v., theorem 2.1 then implies:

$$\mathbb{P}\left(\Big|\langle Ax, y\rangle\Big| \geq n^{1/2}\right) \leq 2\exp\left(\frac{-n}{4||x||_2^2||y||_2^2 \max_{1\leq i,j\leq n} ||\xi_{i,j}||_{\psi_2}^2}\right) = 2\exp\left(\frac{-n}{4\max_{1\leq i,j\leq n} ||\xi_{i,j}||_{\psi_2}^2}\right) \tag{3}$$

Next:

$$\mathbb{P}\big(s_1(A) \geq tn^{1/2}\big) = \mathbb{P}\big(\exists x, y \in \mathbb{S}^{n-1} \ : |\langle Ax, y\rangle| \geq tn^{1/2}\big)$$

Obviously we cannot directly utilize the union bound over the infinite set. Instead we will discretize the sphere $\mathbb{S}^{n-1}$ (lemma 2.1), and show that norm on that discretization is near the real norm (lemma 2.2).

Lemma 2.1. ***(theorem-definition - ε-net on a high-dimensional sphere)***

For $\varepsilon \in (0,1)$, we define ε-net on the sphere $\mathbb{S}^{n-1}$, as a set of points $\mathcal{N}_\varepsilon \subset \mathbb{S}^{n-1}$, so that:

For any $x \in \mathbb{S}^{n-1}$, there is a point $y \in \mathcal{N}_\varepsilon$, such that:

$$||x - y||_2 \leq \varepsilon$$

Then, for every $\varepsilon \in (0,1)$ there exists a finite ε-net $\mathcal{N}_\varepsilon$ on sphere $\mathbb{S}^{n-1}$, so that:

$$\textbf{Card}(\mathcal{N}_\varepsilon) \leq \left(1 + \frac{2}{\varepsilon}\right)^n$$

Proof of lemma 2.1. We construct this ε-net explicitly, by the next alghoritm:

We take any $x_1 \in \mathbb{S}^{n-1}$, while possible we will continue taking $x_{i+1} \in \mathbb{S}^{n-1}$, such that: $\text{dist}(x_{i+1}, \{x_1, ..., x_i\}) > \varepsilon$.

Let $A = \{x_1, ..., x_N\}$ - be the set of N points obtained by algorithm, it's an ε-net by construction so we need to verify inequality on cardinality.

First of all, note that $\varepsilon/2$ closed balls centered at points x_j are disjoint.

By contradiction, if there exists $y \in \overline{B}(x_i, \varepsilon/2) \cap \overline{B}(x_j, \varepsilon/2)$ for $i \neq j$, then:

$$||y - x_i||_2 \leq \varepsilon/2 \quad \text{and} \quad ||y - x_j||_2 \leq \varepsilon/2$$

By triangle inequality:

$$||x_j - x_i||_2 \leq \varepsilon$$

What contradicts with the construction of points $\{x_1, ..., x_N\}$.

Note, that all this $\varepsilon/2$ closed balls centered at points x_j are lies in the bigger $(1 + \varepsilon/2)$-ball centered at 0:

$$\bigcup_{k=1}^{N} \overline{B}(x_k, \varepsilon/2) \subset \overline{B}(0, 1 + \varepsilon/2)$$

But then:

$$\text{Vol}\Big(B(0, 1 + \varepsilon/2)\Big) \geq N \times \text{Vol}\Big(B(0, \varepsilon/2)\Big)$$

We then conclude:

$$N \leq \frac{\text{Vol}\Big(B(0, 1 + \varepsilon/2)\Big)}{\text{Vol}\Big(B(0, \varepsilon/2)\Big)} = \left(\frac{1 + \varepsilon/2}{\varepsilon/2}\right)^n = \left(1 + \frac{2}{\varepsilon}\right)^n$$

□

Lemma 2.2. ***(computation of operator norm on a net)***
Let $\mathcal{N}_\varepsilon$ - be an ε-net on the sphere $\mathbb{S}^{n-1}$, with fixed $\varepsilon \in (0,1)$, and $A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ - linear operator (matrix), then:

$$||A||_{\mathbb{C}^n \to \mathbb{C}^n} \leq \frac{1}{(1-\varepsilon)^2} \max_{x,y \in \mathcal{N}_\varepsilon} |\langle Ax, y\rangle|$$

Proof of lemma 2.2. By the definition of ε-net, for every $x \in \mathbb{S}^{n-1}_{\mathbb{C}}$: $x = y + \nu$, with $\nu \in \mathcal{N}_\varepsilon$ and $||y||_2 \leq \varepsilon$.
So, by triangle inequality:

$$||Ax||_2 \leq ||A\nu||_2 + ||Ay||_2 \leq \max_{\nu \in \mathcal{N}_\varepsilon} ||A\nu||_2 + \max_{||y||_2 \leq \varepsilon} ||Ay||_2 \leq \max_{\nu \in \mathcal{N}_\varepsilon} ||A\nu||_2 + \varepsilon ||A||$$

Taking the supremum on $x \in \mathbb{S}^{n-1}_{\mathbb{C}}$:

$$(1-\varepsilon)||A|| \leq \max_{\nu \in \mathcal{N}_\varepsilon} ||A\nu||_2$$

Notice now that $||A\nu||_2 = \max_{\eta \in \mathbb{S}^{n-1}_{\mathbb{C}}} \langle A\nu, \eta\rangle$, redoing the same proof, one could obtain:

$$(1-\varepsilon)||A\nu||_2 \leq \max_{\eta \in \mathcal{N}_\varepsilon} \langle A\nu, \eta\rangle$$

So we obtained the statement of the lemma.

□

We now return to the proof of theorem, using discretization technique.

Proof of theorem 2.2 .

Let $\mathcal{N}$ - be 1/2-net on $\mathbb{S}^{n-1}$. Lemma 2.2 and monotonnicity of measure then implies for all $u \geq 0$:

$$\mathbb{P}\big(s_1(A) \geq u\big) \leq \mathbb{P}\left(\max_{x,y \in \mathcal{N}} |\langle Ax, y\rangle| \geq \frac{u}{4}\right)$$

Denote by $K := \max_{1 \leq i,j \leq n} ||\xi_{i,j}||_{\psi_2}$, union bound and lemma 2.1 now implies, that:

$$\mathbb{P}\left(\max_{x,y \in \mathcal{N}} |\langle Ax, y\rangle| \geq \frac{u}{4}\right) = \mathbb{P}\left(\exists x, y \in \mathcal{N} : |\langle Ax, y\rangle| \geq \frac{u}{4}\right) \leq$$

$$\mathbf{Card}(\mathcal{N}) \times \mathbb{P}\left(|\langle Ax, y\rangle| \geq \frac{u}{4}\right) \leq 5^n \times \mathbb{P}\left(|\langle Ax, y\rangle| \geq \frac{u}{4}\right) \leq |\text{ formula (2)}| \leq 2\exp\left(\frac{-u^2}{4K^2} + n\log 5\right)$$

Taking now $u = 2K\sqrt{2n\log 5}$, will imply the statement of the theorem with $C_1 = 2K\sqrt{2\log 5}$

□

2.3 Rudelson and Vershynin: Invertibility of random matrices

In this section, we use the previous parts to demonstrate the theorems of Rudelson and Vershynin [16], which we recall as follows:

Theorem 2.3 (Rudelson and Vershynin, general version)**.** *Let $A = (\xi_{i,j})$ - be $n \times n$ matrix, whose entries $\xi_{i,j}$ are independent random variables with zero mean and variance at least one, then:*

1. *If $(\xi_{i,j})$ - are r.v. with uniformly bounded fourth moments by B: $\sup_{i,j} \mathbb{E}|\xi_{i,j}|^4 \leq B$.*
   Then, for all $\varepsilon \geq 0$:

$$\mathbb{P}\big(s_n(A) \leq \varepsilon n^{-1/2}\big) \leq C\varepsilon + c^n + \mathbb{P}\big(||A|| \geq Kn^{1/2}\big)$$

2. *If $(\xi_{i,j})$ are sub-gaussian r.v. with subgaussian moments bounded by B.*
   Then, for all $\varepsilon \geq 0$:

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

Where $K, C > 0, c \in (0,1)$ - constants, that depends only on B, and ***not*** *of n or ε.*

2.3.1 General strategy of the proof

Before going to the proof of the theorem, we need to discuss the strategy of the proof in several parts.

I) Sphere decomposition in high dimensions

As we will see, problem of bounding from below value $||Ax||_2$, could be somehow reduced to bounding from below a linear form $\langle a, \xi \rangle$ with a respect to a vector a on a different subsets of unit sphere $\mathbb{S}^{n-1}$, where ξ - random coefficients vector.

However, behaviour of this linear form strongly dependent of the direction taken on a unit sphere, and so from internal geometry of high-dimensional sphere.

For example, taking $\xi = (\xi_1, ...\xi_n)$ with ξ_i - i.i.d from Rademacher distribution $\mathbb{P}(\xi_i = \pm 1) = \frac{1}{2}$.

For $a = (1, \dots, 0) \in \mathbb{S}^{n-1}$: $\langle a, \xi \rangle = \xi_1 -$ discret supported on $\{\pm 1\}$

For $a = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}) \in \mathbb{S}^{n-1}$: $\langle a, \xi \rangle = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k -$ Almost $\mathcal{N}(0,1)$ for large n by CLT!

One could expect that if a big number (say, $\mathcal{O}(n)$) of coordinates of vector a are non-zero (and so of order $\mathcal{O}(\frac{1}{\sqrt{n}})$) , we would investigate the phenomenons similar to central limit theorem: spreading of the values and concentration around zero, as opposed to the case when only small number of coordinates are non zero. We then have to distinguish this cases, as we expect the different regimes of behaviour.

We now introuduce the sphere decomposition in the next form:

Definition 2.2. *Let $\delta, \rho \in (0,1)$. A vector $x \in \mathbb{R}^n$ is called sparse if $|\operatorname{supp}(x)| \leq \delta n$. A vector $x \in S^{n-1}$ is called compressible if x is within Euclidean distance ρ from the set of all sparse vectors. A vector $x \in S^{n-1}$ is called incompressible if it is not compressible. The sets of sparse, compressible, and incompressible vectors will be denoted by Sparse = Sparse(δ), Comp = Comp(δ, ρ), and Incomp = Incomp(δ, ρ) respectively.*

In our argument, the parameter δ and ρ will be chosen as constant that will depend only on B.

We can write now

$$\mathbb{P}(s_n(A) \leq \varepsilon n^{-1/2}) \leq \mathbb{P}\Big(\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \leq \varepsilon n^{-1/2}\Big) + \mathbb{P}\Big(\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 \leq \varepsilon n^{-1/2}\Big)$$

And after all:

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

$$\mathbb{P}\big(\|A\| \geq K n^{1/2}\big) +$$

$$+\mathbb{P}\left(\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \leq \varepsilon n^{-1/2} \text{ and } \|A\| \leq K n^{1/2}\right) \tag{4}$$

$$+\mathbb{P}\left(\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 \leq \varepsilon n^{-1/2} \text{ and } \|A\| \leq K n^{1/2}\right) \tag{5}$$

As we've seen for subgaussian random variables first term is exponentially small, for well chosen K that depends on B. For general version first term will remain untouched.

For compressible and incompressible vectors, two independent separate arguments will be provided.

II) Invertibility on compressible vectors

For compressible vectors, again, most of the coordinates are almost zero, and only few of them are relatively big, and so the contribution for $||Ax||_2$ comes from the columns that corresponds to the biggest coordinates of x. We could replace then our matrix A, with a submatrix of A of size $n \times [\delta n]$.

Observation, that the smallest singular value of rectangular matrix where number of columns is small relatively to the number of rows is big enough and well exponentially controled, up to a probability that controls largest singular value, see Proposition 2.3. It was well known before the Rudelson-Vershynin and could be again obtained by bounding the linear form from below, see Lemma 2.4, and then by tenzorization techniques, see Lemma 2.3.

Choosing δ - sufficiently small, almost immediatly proves the invertibility on set of sparse vectors, via union bound, condition $\|A\| \leq K n^{1/2}$ helps to deduce the invertibility on full set of compressible vectors.

III) Invertibility on incompressible vectors

Recalling that Lemma 1.1. shows, that s_n is bounded from both sides by minimum of distances from rows X_j vectors to linear subspaces H_n spanned by rows $(X_1, ..., X_{j-1}, X_{j+1}, ..., X_n)$, up to a factor of $\sqrt{n}$.

One could see, that for set of incompressible vectors, providing the careful union argument, we could replace the minimum by average to bound above the probability (6), namely:

$$\mathbb{P}\left(\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 < \varepsilon \rho n^{-1/2}\right) \leq \frac{1}{\delta n} \sum_{k=1}^{n} \mathbb{P}\left(\operatorname{dist}(X_k, H_k) < \varepsilon\right)$$

Obtained distance, could be easily bounded from below by $\text{dist}(X_k, H_k) > |\langle X_k, X^* \rangle|$, where X^* is a unit normal vector to a hyperplane H_n, that we choose so that X^* and X_k are independent. As it turnes out, in general setting, X^* is in $Incomp(\delta, \rho)$ with large probability $1 - e^{-cn}$, see proposition 2.5.

Conditioning upon a realization of $X_1, ...X_{n-1}$, will finally reduces our problem to bounding small probability function, which we define as follows:

Definition 2.3. *Let $\xi_1, \dots, \xi_n$ be independent random variables and $a = (a_1, \dots a_n)$ be a vector of real coefficients. Let*

$$S := \sum_{k=1}^{n} a_k \xi_k$$

.

The small ball probability is defined, for $\varepsilon > 0$ by

$$p_\varepsilon(a) = \sup_{v \in \mathbb{R}} \mathbb{P}(|S - v| \leq \varepsilon)$$

Our final goal will be find a sharp estimation for $p_\varepsilon(a)$, for $a \in Incomp(\delta, \rho))$.

Estimation of $p_\varepsilon(a)$ is a classical problem in probability theory, called a problem of Littlewood and Offord, and was popularized of course by Erdös, see for instance papers [5], [22].

One could atack estimation of $p_\varepsilon(a)$ almost directly. For incompressible vectors, most of the coordinates are at fixed distance from the zero. One could see, that it implies that those vectors have a lot of coordinates of same order $\mathcal{O}(\frac{1}{\sqrt{n}})$, more precisely, next lemma holds

Lemma 2.3 (Incompressible vectors have a big number of coordinates of order $\mathcal{O}(n^{-1/2})$)**.** *Let $x \in Incomp(\delta, \rho)$. Then there exists a set $\sigma \subseteq \{1, \dots, n\}$ of cardinality $|\sigma| \geq \frac{1}{2}\rho^2\delta n$ such that for all $k \in \sigma$:*

$$\frac{\rho}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\delta n}}$$

We then expect for linear forms $\langle a, \xi \rangle$, with ξ - random vector $a \in Incomp(\delta, \rho))$ to have almost gaussian behaviour in large dimensions from CLT, we could use the famous Berry-Essen theorem to deduce the weaker form of concentration result, namely:

Theorem 2.4. *Let $\xi_1, \dots \xi_n$ be random variables (with variance at least 1 and fourth moment bounded by B), $\delta, \rho \in]0\,;1[$ and $a \in Incomp(\delta, \rho)$, then for every $\epsilon > 0$ we have*

$$p_\epsilon(a) \leq C_5(\epsilon + Bn^{-1/2})$$

where C_5 and c depend only on ρ and δ.

However, this results utilise only that incompressible vectors are well spread, and not more thin information. We won't prove it to concentrate ourselves on exponential bound.

To obtain an exponential in dimension result, Rudelson and Vershynin developped a new much sharper tool, inspired by resolution of inverse Littlewood-Offord problem for random-sign integer sums $\sum_k \pm a_k$ in paper of Tao and Vu, see [20]. They investigated the following arithemitic phenomenon:

If the small ball probability $p_0(a)$ is large, then vector a has rich additive structure

One can suppose that more general phenomenon holds for more general real linear forms rather that integer plus-minus sums:

The smaller the value of $p_\varepsilon(a)$, the more non-comparable are coordinates of a in arithmetical sense

and

The bigger the value of $p_\varepsilon(a)$, the less the distance from a to some artihmetic progression

To investigate those phenomenon, next features was introduced:

Definition 2.4 (Recurrence set)**.** *Let $\alpha \in (0, 1)$ and $\kappa \geq 0$. The* recurrence set $I(a) = I_{\alpha,\kappa}(a)$ *of a vector $a \in \mathbb{R}^n$ is defined as the set of all $t \in \mathbb{R}$ such that all except κ coordinates of the vector ta are of distance at most α from $\mathbb{Z}$.*

Regarding t as time, we can think of the recurrence set as the moments when most of the particles moving along the unit torus with speeds $a_1, \dots, a_n$ return close to their initial positions, this describes ergodic nature of the recurrence set.

Key-role in the estimate of $p_\varepsilon(a)$ will play the infimum of recurence set, that we will call essential LCD:

Definition 2.5 (Essential LCD)**.** *Let $\alpha \in (0, 1)$ and $\kappa \geq 0$. The* essential least common denominator $D(a) = D_{\alpha,\kappa}(a)$ *of a vector $a \in \mathbb{R}^n$ is defined as the infimum of $t > 0$ such that all except κ coordinates of the vector ta are of distance at most α from nonzero integers.*

We then reformulate our guess in the following form:

$p_\varepsilon(a)$ shall be controlled by the inverse of essential LCD, because the more coordinates of a are arithmetically incomparable, the bigger LCD is and the smaller $p_\varepsilon(a)$ is!

This intuition is confirmed in the following theorem (all parameters are appropriately chosen, and a_k are of the same order of magnitude) :

$$p_\varepsilon(a) \leq \frac{CBK^3}{\sqrt{\kappa}}\Big(\varepsilon + \frac{1}{D_{2\alpha,2\kappa}(a)}\Big) + C\exp\Big(-\frac{c\alpha^2\kappa}{B^2}\Big)$$

Heart of the proof of Rudelson-Vershynin theorem, is lies in the fact that essential LCD could be of any order up to an exponential.

One could show that it will imply that essential LCD of random normal X^* - is of exponential order, with probability $1 - e^{-cn}$.

It follows from the qualitative observation, let $\widetilde{A}$ be a matrix with rows $X_1, ..., X_n$ and so $\widetilde{A}X^* = 0$, then:

$$\text{If } ||\widetilde{A}x||_2 > 0 \text{ for all } x \in S \text{ some subset, then } X^* \notin S$$

Again, intuitively, the more arithemitic noncomparable coefficients of X^* are, the more likely that $\widetilde{A}X^* > 0$

And so matrix $\widetilde{A}$ is more likely invertible on the subsets S of the unit sphere where the essential LCD is of order below an exponential. Then, by observation above, the random normal X^* will not lie in S Therefore, the essential LCD of X^* will be at least of exponential order!

Summing up the observations, one could show that:

$$\mathbb{P}\big(D_{\alpha,\beta n}(\widehat{X^*}) < e^{cn} \text{ and } \|A\| \leq Kn^{1/2}\big) \leq e^{-c'n}.$$

Conditioning upon a realization of $X_1, ..., X_{n-1}$ and small ball probability estimate via LCD will implies the statement!

We now are going to give the precise details of these differents parts. However, we will not provide the details of the final part that allows for the exponential control of $p_\epsilon(a)$ using the tools developed by Rudelson and Vershynin (thus, we will only prove the weak version with a decay of $\frac{1}{\sqrt{n}}$).

2.3.2 Invertibility on compressible vectors

Next lemma shows, how to get from concentration of linear forms $(Ax)_1$ to concentration of $||Ax||_2$.

Lemma 2.4 (Tensorization)**.** *Let $\xi_1, \ldots, \xi_n$ be independent non-negative random variables and let $K, \varepsilon_0 \geq 0$*

- *Assume that for each k*
$$\mathbb{P}(\xi_k < \varepsilon) \leq K\varepsilon \text{ for all } \varepsilon \geq \varepsilon_0$$
Then
$$\mathbb{P}(\sum_{k=1}^{n} \xi_k^2 < \varepsilon^2 n) \leq (CK\varepsilon)^n$$
where C is an absolute constant.

- *Assume that there exist $\lambda > 0$ and $\mu \in]0\,;1[$ such that for each k*
$$\mathbb{P}(\xi_k < \lambda) \leq \mu$$
Then there exists $\lambda_1 > 0$ and $\mu_1 \in]0\,;1[$ that depend only on λ and μ such that
$$\mathbb{P}(\sum_{k=1}^{n} \xi_k^2 < \lambda_1 n) \leq \mu_1^n$$

We now estimating the smallest singular value in the case of rectangular matrices. The first important proposition is the following:

Proposition 2.1 (Smallest singular values of rectangular matrices)**.** *Let G be an $n \times k$ matrix whose entries are independent centered random variables with variance at least 1 and fourth moment bounded by B. Let $K \geq 1$. Then there exist $c_1, c_2 > 0$ and δ_0 that depend only on B and K such that if $k < \delta_0 n$ then*

$$\mathbb{P}(\inf_{x \in \mathbb{S}^{k-1}} \|Gx\|_2 \leq c_1 n^{\frac{1}{2}} \text{ and } \|G\| \leq Kn^{\frac{1}{2}}) \leq e^{-c_2 n}$$

Proof. We use the following lemma, whose proof we omit.

Lemma 2.5. *Let $\xi_1 \dots \xi_n$ be independent cenrered random variables with variance at least 1 and fourth moment bounded by B. Then there exists $\mu \in]0\,;1[$ depending only on B, such that for every coefficient vector $a = (a_1 \dots a_n) \in \mathbb{S}^{n-1}$ the random sum $S = \sum_{k=1}^{n} a_k \xi_k$ satisfies*

$$\mathbb{P}(|S| < \frac{1}{2}) \leq \mu$$

(This lemma is an application of Paley-Sigmund inequality, we refer to original paper [16] for proof.)

Combining this lemma with the tensorization lemma, we obtain the following estimate:

Let G be a matrix as in the proposition 2.3. Then there exist constants $\eta, \nu \in]0\,;1[$ depending only on B such that for every $x \in \mathbb{S}^{k-1}$

$$\mathbb{P}(\|Gx\|_2 < \eta n^{\frac{1}{2}}) \leq \nu^n$$

We prove now the proposition

Let ε to be chosen later ($\varepsilon > 1$). By the lemma 2.1 there exists an ε-net N in $\mathbb{S}^{k-1}$ of cardinality $|N| \leq (\frac{3}{\varepsilon})^n$. Let η and ν be the numbers present in the previous lemma be. Then by union bound,

$$\mathbb{P}(\exists x \in N, \|Gx\|_2 < \eta n^{\frac{1}{2}}) \leq \nu^n$$

Let V be the event that $\|G\| \leq Kn^{\frac{1}{2}}$ and $\|Gy\|_2 \leq \frac{1}{2}\eta n^{\frac{1}{2}}$ for some point $y \in \mathbb{S}^{k-1}$. Assume that V occurs, and choose a point $x \in N$ such that $\|y - x\|_2 < \varepsilon$. Then

$$\|Gx\|_2 \leq \|Gy\|_2 + \|G\|\varepsilon \leq \eta n^{\frac{1}{2}}$$

,

if we set $\varepsilon = \frac{\eta}{2K}$.

So we have

$$\mathbb{P}(V) \leq (\nu(\frac{3}{\varepsilon})^{\frac{k}{n}})^n \leq e^{-c_2 n}$$

if we assume that $\frac{k}{n} \leq \delta_0 < 1$

□

We now can deduce invertibility on a set of compressible vectors.

Lemma 2.6 (invertibility for compressible vectors)**.** *Let A be a random matrix as in Theorem 2.3, and let $K \geq 1$. Then there exist $\delta, \rho, c_3, c_4 > 0$ that depend only on B and K, and such that*

$$\mathbb{P}\left(\inf_{x \in Comp(\delta,\rho)} \|Ax\|_2 \leq c_3 n^{1/2} \text{ and } \|A\| \leq Kn^{1/2}\right) \leq e^{-c_4 n}$$

Remark 2.3. *The bound in this lemma is much stronger that we need for prove the Rudelson and Vershynin's theorem. Indeed by choosing the constant C in proposition 3.3 large enough, we can assume that $n > \frac{1}{c_3}$ and $\varepsilon < 1$. Then the value $c_3 n^{1/2}$ is bigger than $\varepsilon n^{-1/2}$*

Proof of Lemma 2.6. We start with an estimate on set of sparse vectors.

We note firstly that for $x \in \text{Sparse}(\delta)$, we have $\|Ax\|_2 = \|A'x'\|_2$, where $A' = (\xi_{i,j})_{1 \leq i \leq n, j \in \text{supp}(x)}$ - submatrix of A obtained by removing from A columns that corresponds to coordinates of x that equals to zero, and $x' = (x_i)_{i \in \text{supp}(x)}$ - non zero coordinates of x, there are exactly $[\delta n]$ of them. Moreover:

$$\inf_{x \in \text{Sparse}(\delta)} \|Ax\|_2 = \inf_{x' \in \mathbb{S}^{[\delta n]-1}} \|A'x'\|_2$$

Then, using an union bound and proposition 2.3, with $k = [\delta n]$, with $\delta \leq \delta_0 \leq 1/2$ - appropriately chosen :

$$\mathbb{P}\left(\inf_{x \in \text{Sparse}(\delta)} \|Ax\|_2 \leq c_3 n^{1/2} \text{ and } \|A\| \leq Kn^{1/2}\right) =$$

$$\mathbb{P}\left(\exists \Delta = \text{supp}(x) \subset \{1, \dots, n\} : \inf_{x' \in \mathbb{S}^{[\delta n]-1}} \|A'x'\|_2 \leq c_3 n^{1/2} \text{ and } \|A'\| \leq Kn^{1/2}\right)$$

$$\binom{n}{[\delta n]} \times \mathbb{P}\left(\inf_{x' \in \mathbb{S}^{[\delta n]-1}} \|A'x'\|_2 \leq c_3 n^{1/2} \text{ and } \|A'\| \leq Kn^{1/2}\right) \leq \binom{n}{[\delta n]} e^{-c_2 n} \leq e^{-c_2 n/2}$$

Since $\binom{n}{[\delta n]}$ is a number of possible choices of $[\delta n]$ non-zero coordinates from n possible.

We now deduce invertibility on a set of compressible vectors, which is simple technical statement. □

2.3.3 Invertibility on incompressible vectors via distance

We now need to control invertibility on incompressible vectors. One of the most impressive results is the following, which allows reducing the problem to calculating a lower bound on the distance between a random vector and a random hyperplane, which was discussed above.

Lemma 2.7 (invertibility via distance)**.** *Let A be any random matrix. Let $X_1, \dots, X_n$ denote the column vectors of A, and let H_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*

$$\mathbb{P}\left(\inf_{x \in Incomp(\delta,\rho)} \|Ax\|_2 < \varepsilon \rho n^{-1/2} \right) \le \frac{1}{\delta n} \sum_{k=1}^{n} \mathbb{P}\left(\operatorname{dist}(X_k, H_k) < \varepsilon\right)$$

Proof. Let $x \in \operatorname{Incomp}(\delta, \rho)$. Writing $Ax = \sum_{k=1}^{n} x_k X_k$, we have

$$\|Ax\|_2 \ge \max_{k=1,\dots,n} \operatorname{dist}(Ax, H_k) = \max_{k=1,\dots,n} \operatorname{dist}(x_k X_k, H_k) = \max_{k=1,\dots,n} |x_k| \operatorname{dist}(X_k, H_k).$$

Denote

$$p_k := P(\operatorname{dist}(X_k, H_k) < \varepsilon).$$

Then

$$\mathbb{E}\,|\{k : \operatorname{dist}(X_k, H_k) < \varepsilon\}| = \sum_{k=1}^{n} p_k.$$

Denote by U the event that the set $\sigma_1 := \{k : \operatorname{dist}(X_k, H_k) \ge \varepsilon\}$ contains more than $(1-\delta)n$ elements. Then by Chebyshev's inequality,

$$\mathbb{P}(U^c) \le \frac{1}{\delta n} \sum_{k=1}^{n} p_k.$$

On the other hand, for every incompressible vector x, the set $\sigma_2(x) := \{k : |x_k| \ge \rho n^{-1/2}\}$ contains at least δn elements.

Assume that the event U occurs. Fix any incompressible vector x. Then $|\sigma_1| + |\sigma_2(x)| > (1-\delta)n + \delta n > n$, so the sets σ_1 and $\sigma_2(x)$ have a nonempty intersection. Let $k \in \sigma_1 \cap \sigma_2(x)$. Then by (3.7) and by the definitions of the sets σ_1 and $\sigma_2(x)$, we have

$$\|Ax\|_2 \ge |x_k| \operatorname{dist}(X_k, H_k) \ge \rho n^{-1/2} \cdot \varepsilon.$$

Summarizing, we have shown that

$$\mathbb{P}\left(\inf_{x \in \operatorname{Incomp}(\delta,\rho)} \|Ax\|_2 < \varepsilon \rho n^{-1/2} \right) \le \mathbb{P}(U^c) \le \frac{1}{\delta n} \sum_{k=1}^{n} p_k.$$

This completes the proof. □

The last part of the paper of Rudelson and Vershynin aims at prove the following proposition which permits to easily conclude about the initial theorem:

Proposition 2.2 (Strong Distance bound)**.** *Let A be a random matrix as in Theorem 2.3. Let $X_1, \dots, X_n$ denote its column vectors, and consider the subspace $H_n = \operatorname{span}(X_1, \dots, X_{n-1})$. Let $K \ge 1$. Then for every $\varepsilon \ge 0$, one has*

$$\mathbb{P}\left(\operatorname{dist}(X_n, H_n) < \varepsilon \text{ and } \|A\| \le K n^{1/2}\right) \le C(\varepsilon + c^n),$$

where C and $c \in (0,1)$ depend only on B and K.

For the end of this paper we give an idea of the proof of this proposition. We begin with a replacement of $\operatorname{dist}(X_n, H_n)$.

To this end let X^* be any unit vector orthogonal to $X_1 \dots X_{n-1}$. We call it a random normal. We can choose X^* so that it is a random vector that depends only on $X_1 \dots X_{n-1}$ and is independent of X_n.

We have easily that

$$\operatorname{dist}(X_n, H_n) \ge |\langle X^*, X_n \rangle|$$

And so:

$$\mathbb{P}(\operatorname{dist}(X_n, H_n) \le \varepsilon \text{ and } \|A\| \le K n^{1/2}) \le \mathbb{P}(|\langle X^*, X_n \rangle| \le \varepsilon \text{ and } \|A\| \le K n^{1/2}) \tag{6}$$

It is easy to show that a random normal is often incompressible, more precisely

Proposition 2.3. *Let δ, ρ, c_4 be as in lemma 2.5. Then*

$$\mathbb{P}(X^* \in Comp(\delta, \rho) \text{ and } \|A\| \le K n^{1/2}) \le e^{-c_4 n}$$

Proof of proposition 2.3 . Let $\widetilde{A}$ - be a matrix with rows $X_1, ..., X_{n-1}$. By definition of random normal:

$$\widetilde{A}X^* = 0$$

And so, if $X^* \in Comp(\delta, \rho)$, then $\inf_{x \in \mathrm{Comp}(\delta,\rho)} \|\widetilde{A}x\|_2 = 0$, applying now invertibility on set of compressible vectors for $(n-1) \times n$ matrix, one can deduce the lemma. Notice that lemma 2.5 is valid also for $(n-1) \times n$, we omit to prove it here, since it's almost the same as for $n \times n$. □

We could write:

$$\mathbb{P}(|\langle X^*, X_n \rangle| \le \varepsilon \text{ and } \|A\| \le Kn^{1/2}) \le$$

$$\mathbb{P}(|\langle X^*, X_n \rangle| \le \varepsilon \text{ and } X^* \in Incomp(\delta, \rho)) + \mathbb{P}(X^* \in Comp(\delta, \rho) \text{ and } \|A\| \le Kn^{1/2})$$

Second term is exponentially small, by a previous lemma, so we need to bound first term. Notice that fixing $X_1, ..., X_{n-1}$ uniquely defines and fixes random normal X^* that independent from X_n, and so we will do a conditioning on a realisation of $X_1, ..., X_{n-1}$. By law of total expectation:

$$\mathbb{P}(|\langle X^*, X_n \rangle| \le \varepsilon \text{ and } X^* \in Incomp(\delta, \rho)) =$$

$$\mathbb{E}_{X_1,\dots,X_{n-1}} \left[\mathbb{P}\left(|\langle X^*, X_n \rangle| \le \varepsilon \text{ and } X^* \in Incomp(\delta, \rho) \middle| X_1, \dots, X_{n-1} \right) \right]$$

Now, probability under the expectation we could bound in the following way:

$$\mathbb{P}\left(|\langle X^*, X_n \rangle| \le \varepsilon \text{ and } X^* \in Incomp(\delta, \rho) \middle| X_1, \dots, X_{n-1} \right) \le \sup_{v \in \mathbb{R}} \mathbb{P}\Big(|\sum_{k=1}^{n} a_k \xi_{k,n} - v| \le \varepsilon \Big) \le p_\varepsilon(a)$$

Where $a = (a_1, ..., a_n) = X^*$ - which is fixed vector from $Incomp(\delta, \rho)$, upon conditioning on $X_1, ..., X_{n-1}$.

To the end of this thesis, we will estimate $p_\varepsilon(a)$, for $a \in Incomp(\delta, \rho)$.

Remark 2.4. *In fact, we will use other decomposition to deduce the statement, however it shows why we are interested in solving Littlewood-Offord problem with incompressible coefficients. One can use this decomposition to deduce weaker result on concentration with order $n^{-1/2}$ as it was told in theorem 2.5. We will go straightforward to Vershynin-Rudelson proof instead.*

2.3.4 Small ball probability via Arithmetico-Erdogic features

In this section, we show how to bound small ball probability by inverse of essential LCD, that was discussed above. More precisely, we will show next result:

Theorem 2.5 (Small Ball Probability)**.** *Let ξ be a centered random variable with variance at least 1 and with the third moment bounded by B. Consider independent copies $\xi_1, \dots, \xi_n$ of ξ. Let $a = (a_1, \dots, a_n)$ be a coefficient vector and let $K \ge 1$ be such that*

$$1 \le |a_k| \le K \quad \text{for all } k. \tag{7}$$

Let $0 < \alpha < 1/6K$ and $0 < \kappa < n$. Then for every $\varepsilon \ge 0$ one has

$$p_\varepsilon(a) \le \frac{CBK^3}{\sqrt{\kappa}} \Big(\varepsilon + \frac{1}{D_{2\alpha,2\kappa}(a)} \Big) + C \exp\Big(- \frac{c\alpha^2\kappa}{B^2} \Big),$$

where $C, c > 0$ are absolute constants.

Proof of theorem 2.5. We will not give the proof in full of details, however we will illustrate main techniques.

We start from the Esseen inequality, see paper , on a small ball probability, via characteristic function. Let $S = \sum_{k=1}^{n} a_k \xi_k$ and $\phi(t) = \mathbb{E}e^{iSt}$, then:

$$p_\varepsilon(a) = \sup_{v \in \mathbb{R}} \mathbb{P}(|S - v| \le \varepsilon) \le C \int_{-\pi/2}^{\pi/2} |\phi(t/\varepsilon)| \, dt,$$

□

Now, by independence ξ_k, we obviously have:

$$\phi(t) = \mathbb{E}e^{iSt} = \prod_{k=1}^{n} \mathbb{E}e^{i\xi_k t}$$

To estimate the integral in Esseen inequality , we notice now that

$$|\mathbb{E}e^{i\xi_k t}|^2 = \mathbb{E}\cos(a_k|\xi_k - \xi_k^{'}|t).$$

Where $\xi_k^{'}$ - is an independent copie of ξ_k.

Using the inequality $|x| \le \exp(-\frac{1}{2}(1-x^2))$ valid for all x, we then obtain

$$|\phi(t)| \le \prod_{k=1}^{n} \exp\Big(-\frac{1}{2}(1-|\phi_k(t)|^2)\Big)$$
$$= \exp\Big(-\mathbb{E}\sum_{k=1}^{n}\frac{1}{2}(1-\cos(a_k|\xi_k-\xi_k^{'}|t))\Big) = \exp\big(-\mathbb{E}f(|\xi_1-\xi_1^{'}|t)\big),$$

where

$$f(t) := \sum_{k=1}^{n} \sin^2\big(\frac{1}{2}a_k t\big).$$

Now, we notice that from Paley-Zigmund inequality:

$$\mathbb{P}(|\xi_1 - \xi_1^{'}|) > 1) := \beta > 0$$

We then replace random variable $|\xi_1 - \xi_1^{'}|$, by itself conditioned on event that $|\xi_1 - \xi_1^{'}| > 1$, which we denote by ζ. That will change nothing but absolute constant in an exponent meaning that:

$$|\phi(t)| \le \exp\big(-\beta\,\mathbb{E}f(\zeta t)\big).$$

Then by Esseen's and using Jensen's inequalities, we estimate the small ball probability as

$$p_\varepsilon(a) \le C\int_{-\pi/2}^{\pi/2} |\phi(t/\varepsilon)|\,dt \le C\sup_{z\ge 1}\int_{-\pi/2}^{\pi/2}\exp\big(-\beta f(zt/\varepsilon)\big)\,dt. \tag{8}$$

Fix $z \ge 1$. It's easy to see that for $M := \max\limits_{|t|\le\pi/2} f(zt/\varepsilon)$:

$$\frac{n}{4} \le M \le n$$

Now we consider the level sets of f, defined for $m, r \ge 0$ as

$$T(m,r) := \{t:\ |t| \le r,\ f(zt/\varepsilon) \le m\}.$$

We will use now the statement called Halász lemma, see [9]:

Lemma 2.8. *Let $l \in \mathbb{N}$ be such that $l^2 m \le M$. Then*

$$|T(m, \frac{\pi}{2})| \le \frac{2}{l}\cdot|T(l^2 m, \pi)|.$$

Hence, for every $\eta \in (0,1)$ such that $m \le \eta M$, one has:

$$|T(m,\frac{\pi}{2})| \le 4\sqrt{\frac{m}{\eta M}}\cdot|T(\eta M,\pi)|. \tag{9}$$

(Apply previous $l = \lfloor\sqrt{\frac{\eta M}{m}}\rfloor$).

Now we can estimate the integral in (8) by the integral distribution formula. Using (9) for small m and the trivial bound $|T(m,\pi/2)| \le \pi$ for large m, we get

$$p_\varepsilon(a) \le C\sup_{z\ge1}\int_{-\pi/2}^{\pi/2}\exp\big(-\beta f(zt/\varepsilon)\big)\,dt \le C\int_0^\infty |T(m,\frac{\pi}{2})|\ \beta e^{-\beta m}\,dm$$
$$\le C\int_0^{\eta M} 4\sqrt{\frac{m}{\eta M}}\cdot|T(\eta M,\pi)|\ \beta e^{-\beta m}\,dm + C\int_{\eta M}^{\infty}\pi\ \beta e^{-\beta m}\,dm$$
$$\le \frac{C_1}{\sqrt{\beta\eta M}}\cdot|T(\eta M,\pi)| + C\pi e^{-\beta\eta M} \le \frac{C_2 B}{\sqrt{\eta n}}\cdot|T(\eta n,\pi)| + C\pi e^{-c_2\eta n/B^2}.$$

We shall now bound the measure of the level set $|T(\eta n, \pi)|$ by the density of the *recurrence set* of a.

Consider any $t \in T(\eta n, \pi)$ and set $y := z/2\varepsilon$. Then $y \geq 1/2\varepsilon$, and

$$f(zt/\varepsilon) = \sum_{k=1}^{n} \sin^2(a_k y t) \leq \eta n. \tag{10}$$

Let us fix

$$\eta := \frac{\alpha^2 \kappa}{4n}. \tag{11}$$

Then at least $n - \kappa$ terms in the sum in (10) satisfy

$$\sin^2(a_k y t) \leq \frac{\eta n}{\kappa} = \frac{\alpha^2}{4} < \frac{1}{144},$$

which implies for those terms that $\operatorname{dist}(a_k y t, \pi\mathbb{Z}) \leq \alpha$. Thus yt/π belongs to the recurrence set of a, which we defined above.

Our argument thus shows that $T(\eta n, \pi) \subseteq \frac{\pi}{y} I_{\alpha,\kappa}(a)$. Thus

$$|T(\eta n, \pi)| \leq \Big|\frac{\pi}{y} I_{\alpha,\kappa}(a) \cap [-\pi, \pi]\Big| = \frac{\pi}{y} \cdot |I_{\alpha,\kappa}(a) \cap [-y, y]|.$$

The quantity

$$\operatorname{dens}(I, y) := \frac{1}{2y} \cdot |I \cap [-y, y]|$$

can be interpreted as the *density* of the set I. We have thus shown that

$$|T(\eta n, \pi)| \leq 2\pi \operatorname{dens}(I_{\alpha,\kappa}(a), y).$$

Using this bound and our choice (11) of η in (**??**), we conclude that:

$$p_\varepsilon(a) \leq \frac{C_3 B}{\alpha\sqrt{\kappa}} \cdot \sup_{y \geq 1/2\varepsilon} \operatorname{dens}(I_{\alpha,\kappa}(a), y) + C\pi e^{-c_3 \alpha^2 \kappa / B^2}. \tag{12}$$

It remains to bound the density of the recurrence set $I(a)$ by the reciprocal of the essential LCD $D(a)$. We will derive this from the following structural lemma, which shows that: (1) the recurrence set has lots of gaps; (2) each gap bounds below the essential LCD of a.

Lemma 2.9 (Gaps in the recurrence set)**.** *Under the assumptions of Theorem 2.5, let $t_0 \in I_{\alpha,\kappa}(a)$. Then:*

1. *$t_0 + 3\alpha \notin I_{\alpha,\kappa}(a)$.*
2. *Let $t_1 \in I_{\alpha,\kappa}(a)$ be such that $t_1 > t_0 + 3\alpha$. Then $t_1 - t_0 \geq D_{2\alpha,2\kappa}(a)$.*

Since $D_{2\alpha,2\kappa}(a) \geq (1-2\alpha)/K > 4\alpha$, this lemma implies that the recurrence set I has gaps of size at least $D_{2\alpha,2\kappa}(a) - 4\alpha$.

We won't prove this lemma since it rather technical than ideological. One can use it to bound the density of the recurrence set via the reciprocal of the essential LCD.

Lemma 2.10 (Recurrence set via essential LCD)**.** *Under the assumptions of Theorem 2.5, we have for every $y > 0$:*

$$\operatorname{dens}(I_{\alpha,\kappa}(a), y) \leq 3\alpha\Big(\frac{1}{2y} + \frac{2}{D_{2\alpha,2\kappa}(a)}\Big). \tag{13}$$

The contribution of the first term in (13) comes from the $O(\alpha)$-neighborhood of zero, which is contained in the recurrence set. This is the initial time when all of the moving particles are still close to 0.

Proof. Denote $I := I_{\alpha,\kappa}(a) \cap [-y, y]$. This set is closed and nonempty (it contains 0). Set $t_0 := \min\{t : \ t \in I\}$. If $I \subseteq [t_0, t_0 + 3\alpha]$, then

$$\operatorname{dens}(I, y) = \frac{|I|}{2y} \leq \frac{3\alpha}{2y}, \tag{14}$$

which completes the proof in this case.

Assume then that $I \not\subseteq [t_0, t_0 + 3\alpha]$. Then we can define inductively the maximal sequence of points $t_1, t_2, \ldots, t_L \in I$ by

$$t_l := \min\{t : \ t \in I; \ t > t_{l-1} + 3\alpha\}.$$

Note that by Lemma 2.9, $t_{l-1} + 3\alpha \notin I$. Thus the strict inequality in the definition of t_l can be replaced by the non-strict inequality, so the minimum makes sense.

Part 1 of Lemma 2.9 yields

$$I \subseteq \bigcup_{l=0}^{L} [t_l, t_l + 3\alpha),$$

while part 2 implies

$$t_L - t_0 \geq \sum_{l=1}^{L} (t_l - t_{l-1}) \geq L \cdot D_{2\alpha,2\kappa}(a).$$

On the other hand, since $t_0, t_L \in I \subseteq [-y, y]$, we have $t_L - t_0 \leq 2y$. We conclude that

$$\mathrm{dens}(I, y) \leq \frac{|\bigcup_{l=0}^{L} [t_l, t_l + 3\alpha)|}{t_L - t_0} \leq \frac{(L+1) \cdot 3\alpha}{L \cdot D_{2\alpha,2\kappa}(a)} \leq \frac{6\alpha}{D_{2\alpha,2\kappa}(a)}.$$

This completes the proof. □

We now conclude the theorem 2.5.

For a general coefficient vector a, not necessarily with well coefficients of the same order, we will rewrite the theorem restricting a onto its spread part, which we define as follows:

Definition 2.6 (Spread part)**.** *Let $0 < K_1 < K_2$ be fixed. For a vector $x \in \mathbb{R}^n$, we consider the subset $\sigma(x) \subseteq \{1, \dots, n\}$ defined as*

$$k \in \sigma(x) \quad \text{if} \quad K_1 \leq |n^{1/2} x_k| \leq K_2,$$

and, if $\sigma(x) \neq \emptyset$, we define the spread part of x *as*

$$\hat{x} := (n^{1/2} x_k)_{k \in \sigma(x)}.$$

If $\sigma(x) = \emptyset$, the spread part of x is not defined.

We notice also that:

Lemma 2.11 (restriction)**.** *For any $a \in \mathbb{R}^n$, any $\sigma \subseteq \{1, \dots, n\}$, orthogonal projection P_σ on space R^σ and any $\epsilon \geq 0$, we have*

$$p_\epsilon(a) \leq p_\epsilon(P_\sigma a).$$

One can deduce general theorem for small ball probability:

Corollary 2.1 (Small ball probability for general vectors)**.** *Let $\xi_1, \dots, \xi_n$ be random variables as in Theorem 2.6. Let $a \in \mathbb{R}^n$ be a vector of real coefficients whose spread part $\hat{a}$ is well defined (for some fixed truncation levels $K_1, K_2 > 0$). Let $\alpha \in (0, 1)$ and $\beta \in (0, 1/2)$. Then for for every $\epsilon \geq 0$ one has*

$$p_\epsilon(a) \leq \frac{C}{\sqrt{\beta}} \Big(\epsilon + \frac{1}{\sqrt{n}\, D_{\alpha,\beta n}(\hat{a})} \Big) + C e^{-c\alpha^2 \beta n},$$

where $C, c > 0$ depend (polynomially) only on B, K_1, K_2.

Remark 2.5. *We set $D_{\alpha,\kappa}(\hat{a}) = 0$ if $\hat{a}$ is not defined.*

2.3.5 Final step

The final step in the paper, is the following theorem:

Theorem 2.6 (Random normal)**.** *Let $X_1, \dots, X_{n-1}$ be random vectors as in proposition 3.4 Consider a unit vector X^* orthogonal to all these vectors. Let $K \geq 1$. Then there exist constants $K_1, K_2, \alpha, \beta, c, c' > 0$ that depend only on B and K, and such that*

$$\mathbb{P}\big(D_{\alpha,\beta n}(\widehat{X^*}) < e^{cn} \text{ and } \|A\| \leq K n^{1/2} \big) \leq e^{-c' n}.$$

Firstly, let us show, that theorem 2.6 will implies theorem 2.3.

Proof of theorem 2.3. It's enough to prove that theorem 2.6 implies the proposition 2.2. , using the formula (7), we will bound the quantity $\mathbb{P}(|\langle X^*, X_n \rangle| \leq \epsilon \text{ and } \|A\| \leq K n^{1/2})$. Let $K_1, K_2, \alpha, \beta, c, c' > 0$ constants as in theorem 2.6.

We write:

$$\mathbb{P}(|\langle X^*, X_n \rangle| \leq \epsilon \text{ and } \|A\| \leq K n^{1/2}) \leq$$

$$\mathbb{P}\big(D_{\alpha,\beta n}(\widehat{X^*}) < e^{cn} \text{ and } \|A\| \leq K n^{1/2} \big) + \mathbb{P}\big(D_{\alpha,\beta n}(\widehat{X^*}) \geq e^{cn} \text{ and } |\langle X^*, X_n \rangle| \leq \epsilon \big)$$

First term is less than $e^{-c'n}$, by the theorem 3.11. We need then to bound now the second term. Fixing $X_1, ..., X_{n-1}$ uniquely defines and fixes random normal X^* that independent from X_n, and so we will do a conditioning on a realisation of $X_1, ..., X_{n-1}$. By law of total expectation:

$$\mathbb{P}\big(D_{\alpha,\beta n}(\widehat{X^*}) \geq e^{cn} \text{ and } |\langle X^*, X_n\rangle| \leq \epsilon\big) =$$

$$\mathbb{E}_{X_1,\dots,X_{n-1}}\left[\mathbb{P}\left(\big(D_{\alpha,\beta n}(\widehat{X^*}) \geq e^{cn} \text{ and } |\langle X^*, X_n\rangle| \leq \epsilon\big)\middle| X_1, \dots, X_{n-1}\right)\right]$$

For internal probability, X^* - is fixed and $D_{\alpha,\beta n}(\widehat{X^*})$ as well. But then, we could apply the small ball probability estimate obtained in theorem 2.5 and corollary 2.1!

That implies particularly that:

$$\mathbb{P}\big(D_{\alpha,\beta n}(\widehat{X^*}) \geq e^{cn} \text{ and } |\langle X^*, X_n\rangle| \leq \epsilon\big) \leq C_1\varepsilon + C_2 e^{-c_3 n}$$

With constants $C_1, C_2, c_3 > 0$ depends only on K, B. That ends the proof of theorem 2.3!

□

We now return to the proof of theorem 2.6.

As it was described in a strategy, matrix $\widetilde{A}$ is more likely invertible on the subsets S of the unit sphere where the essential LCD is of order below an exponential. We then need to define preferable subsets on the sphere, we will omit all of the details about the choosing the parameters, and will provide the general ideas instead.

Definition 2.7 (Level sets of LCD)**.** *We define the level set $S_D \subseteq S^{n-1}$ as*

$$S_D := \{x \in \mathit{Incomp} :\ D \leq D_{\alpha,n_0/2}(\hat{x}) < 2D\}.$$

We want to show the invertibility of the random matrix $\widetilde{A}$ on the level sets S_D for all D up to an exponential order. This will be done using again $\varepsilon-$net argument. We will first show the invertibility on a single vector $\in S_D$. Next, we will find a small (α/D)-net in S_D. Then, by a union bound, the invertibility will hold for each point in this net. By approximation, we will extend the invertibility to the whole S_D.

The invertibility on a single vector $x \in S_D$ will easily follow from our general small ball probability estimates and the tensorization.

Lemma 2.12 (Invertibility on a single vector)**.** *There exist $c, C_>0$ that depend only on B and K, and such that the following holds. Let $\alpha \in (0,1)$ and $D_0 \leq D < \frac{1}{\sqrt{n}}\, e^{c\alpha^2 n}$. Then for every vector $x \in S_D$ and for every $t \geq 0$, one has*

$$\mathbb{P}\big(\|A'x\|_2 < tn^{1/2}\big) \leq \Big(Ct + \frac{C}{\sqrt{n}\,D}\Big)^{n-1}.$$

Careful argument on cardinality of ε-net on S_D, which we will not discuss here, implies that:

Lemma 2.13 (Invertibility on a level set)**.** *There exist $\alpha, c, C > 0$ that depend only on B and K, such that for D, which verifies $D_0 \leq D < e^{cn}$. Then*

$$\mathbb{P}\big(\inf_{x\in S_D} \|A'x\|_2 < \frac{c_{10}}{D}\, n^{1/2} \text{ and } \|A\| \leq Kn^{1/2}\big) \leq e^{-n}.$$

Proof of theorem 2.6. Let α and c be as in Lemma 2.13.

If $x \in S^{n-1}$ is such that $D(\hat{x}) < e^{cn}$ then, by the definition of the level sets S_D, either x is compressible or $x \in S_D$ for some $D \in \mathcal{D}$, where

$$\mathcal{D} = \{D :\ D_0/2 \leq D < e^{cn},\ D = 2^k,\ k \in \mathbb{Z}\}.$$

We have then:

$$\mathbb{P}\big(D(\widehat{X^*}) < e^{cn} \text{ and } \|A\| \leq Kn^{1/2}\big) \leq$$

$$\mathbb{P}\big(X^* \in \mathit{Comp} \text{ and } \|A\| \leq Kn^{1/2}\big) + \sum_{D\in\mathcal{D}} \mathbb{P}(X^* \in S_D \text{ and } \|A\| \leq Kn^{1/2}).$$

By proposition 2.3 , $\mathbb{P}(X^* \in \mathit{Comp} \text{ and } \|A\| \leq Kn^{1/2}) \leq e^{-c_4 n}$. By lemma 2.13, for every $D \in \mathcal{D}$ we have

$$\mathbb{P}(X^* \in S_D \text{ and } \|A\| \leq Kn^{1/2}) \leq \mathbb{P}\big(\inf_{x\in S_D} \|A'x\|_2 = 0 \text{ and } \|A\| \leq Kn^{1/2}\big) \leq e^{-n}.$$

Since $|\mathcal{D}| \leq C'n$, we conclude that

$$\mathbb{P}\big(D(\widehat{X^*}) < e^{cn} \text{ and } \|A\| \leq Kn^{1/2}\big) \leq e^{-c_4 n} + C'n \cdot e^{-n} \leq e^{-c'n}.$$

This completes the proof of Theorem 2.6.

□

3 Conclusion

Main goal of this thesis was studying the invertibility of large random matrices via developing the estimation on smallest singular value.

Despite the fact that the formulation seems very natural:

Take matrix A with random entries, what is the probability that A is singular?

and is understandable for any student in their first to third year, solving this problem combines various aspects of modern mathematics, from studying the geometry of a sphere in huge dimensions and its coverings to arithmetic progressions and sums with random coefficients.

To conclude, the main result described in our thesis is:

For square matrix A with general independent entries, probability of singularity of A is exponentially small.

We would like to thank prof. Djalil Chafaï for his valuable advices and collaboration, as well as support during the exploration of the topic and for the opportunity to consistently discuss the concepts at weekly meetings at 6 PM on Wednesdays.

4 Further developpements

I Tikhomirov's theorem

In 2019, Tikhomirov proved the Spielman-Teng conjecture for Bernoulli matrices, which we recall here:

Theorem 4.1 (Tikhomirov (2019), see [23])**.** *For every $p \in]0\,;\frac{1}{2}]$ and $\varepsilon > 0$ there are $n_{p,\varepsilon}, C_{p,\varepsilon} > 0$ depending only on p and ε with the following property. Let $n > n_{p,\varepsilon}$ and let $B_n(p)$ be the $n \times n$ random matrix with independent entries b_{ij} such that $\mathbb{P}(b_{ij}) = 1) = p$ and $\mathbb{P}(b_{ij} = 0) = 1 - p$. Then for any $s \in [-1\,;0]$ and any $t > 0$,*

$$\mathbb{P}(s_n(B_n(p) + s1_n 1_n^T) < tn^{-1/2}) \leq (1 - p + \varepsilon)^n + C_{p,\varepsilon} t$$

where 1_n is the column vector consisting solely of 1

II Rebrova-Tikhomirov on relaxation of fourth moment condition

One could ask a question:

Is fourth moment condition crucial in theorem of Rudelson and Vershynin or it could be relaxed?

In 2017, Elisaveta Rebrova - PhD student of Roman Vershynin in colaboration with Konstantin Tikhomirov in paper [15], proved that for case when entries are i.i.d fourth moment condition isn't necessary and probability that operator norm is big is absent in estimate, more precisely they proved:

Theorem 4.2 (Rebrova-Tikhomirov)**.** *Let A be an $n \times n$ random matrix, whose elements are independent copies of a mean zero subgaussian random variable with unit variance. Then, for every $\varepsilon > 0$, we have*

$$\mathbb{P}\Big(s_n(A) \leq \varepsilon n^{1/2}\Big) \leq C\varepsilon + c^n \tag{15}$$

where $C, c > 0$ depend (polynomially) only on the subgaussian moment B.

III Recent(very) work on a precision of a constant in Spielman-Teng conjecture

In the work of 30 May 2024, Ashwin Sah, Julian Sahasrabudhe and Mehtaab Sawhney, published thinner version of Vershynin-Rudelson-Rebrova-Tikhomirov's theorem, where they precise the constant C which stands in front of ε.

More precisely they proved in paper [17]:

Theorem 4.3. *Let A be an $n \times n$ random matrix with iid subgaussian entries, with mean 0 and variance 1. Then, for all $\varepsilon \geq 0$,*

$$\mathcal{P}\big(s_n(A) \leq \varepsilon n^{-1/2}\big) \leq (1 + o(1))\varepsilon + e^{-cn}, \tag{16}$$

where the $o(1)$ term decays as $C(\log n)^{-1/16}$, where $C, c > 0$ depends only on distribution of entries.

IV Open questions on Heavy-Tailed matrices: We could pose the question about the invertibility of random matrices, when entries do not have second or even first moment, and so CLT or LLN are not longer available. It's well-known that those matrices behaves very different from those whose entries have second moments, see for instance paper of Alice Guionnet [8]. It would be then interesting to understand the concentration of the smallest singular value in this ensembles.

References

[1] C. Bordenave and D. Chafaï. Around the circular law, 2012. arXiv: 1109.3343 [math.PR].

[2] J. Bourgain, V. Vu, and P. Wood. On the singularity probability of discrete random matrices. *Journal of Functional Analysis*, 258:559–603, Jan. 2010. DOI: 10.1016/j.jfa.2009.04.016.

[3] W. Cukierski. Dogs vs. cats, 2013. URL: https://kaggle.com/competitions/dogs-vs-cats.

[4] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM Journal on Matrix Analysis and Applications*, 9(4):543–560, 1988. DOI: 10.1137/0609045. eprint: https://doi.org/10.1137/0609045. URL: https://doi.org/10.1137/0609045.

[5] P. Erdös. On a lemma of Littlewood and Offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.

[6] S. Fiorini. gene expression cancer RNA-Seq. UCI Machine Learning Repository, 2016. DOI: https://doi.org/10.24432/C5R88

[7] H. H. Goldstine and J. V. Neumann. Numerical inverting of matrices of high order. ii. *Proceedings of the American Mathematical Society*, 2(2):188–202, 1951. ISSN: 00029939, 10886826. URL: http://www.jstor.org/stable/2032484 (visited on 05/07/2024).

[8] A. Guionnet. *Heavy tailed random matrices: how they differ from the goe, and open problems: the abel symposium, rosendal, norway, august 2016*. In Jan. 2018, pages 415–427. ISBN: 978-3-030-01592-3. DOI: 10.1007/978-3-030-01593-0_15.

[9] G. Halász. On the distribution of additive arithmetic functions. *Acta Arithmetica*, 27(1):143–152, 1975.

[10] C. Japhet. Conditionnement d'un système linéaire. URL: https://www.math.univ-paris13.fr/~japhet/L2/2020-2021/Conditionnement.pdf (visited on 11/15/2020).

[11] J. Kahn, J. Komlos, and E. Szemeredi. On the probability that a random ± 1matrix is singular. *Journal of The American Mathematical Society*, 8, Jan. 1995. DOI: 10.2307/2152887.

[12] J. Komlós. On the determinant of (0-1) matrices. *Studia Scientiarium Mathematicarum Hungarica*, 2, 1967. DOI: https://doi.org/10.7282/t3-jxdr-am37.

[13] R. Macausland. The moore-penrose inverse and least squares. URL: http://buzzard.ups.edu/courses/2014spring/420projects/math420-UPS-spring-2014-macausland-pseudo-inverse.pdf.

[14] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. English. *Math. USSR, Sb.*, 1:457–483, 1968. ISSN: 0025-5734. DOI: 10.1070/SM1967v001n04ABEH001994.

[15] E. Rebrova and K. Tikhomirov. Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries, 2017. arXiv: 1508.06690 [math.PR].

[16] M. Rudelson and R. Vershynin. The littlewood-offord problem and invertibility of random matrices, 2008. arXiv: math/0703503 [math.PR].

[17] A. Sah, J. Sahasrabudhe, and M. Sawhney. On the spielman-teng conjecture, 2024. arXiv: 2405.20308 [math.PR].

[18] A. Sankar, D. A. Spielman, and S.-H. Teng. Smoothed analysis of the condition numbers and growth factors of matrices, 2005. arXiv: cs/0310022 [cs.NA].

[19] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms, 2002. arXiv: math/0212413 [math.OC].

[20] T. Tao and V. Vu. Inverse littlewood-offord theorems and the condition number of random discrete matrices, 2007. arXiv: math/0511215 [math.PR].

[21] T. Tao and V. Vu. On the singularity probability of random bernoulli matrices, 2008. arXiv: math/0501313 [math.CO].

[22] T. Tao and V. Vu. The littlewood-offord problem in high dimensions and a conjecture of frankl and füredi, 2011. arXiv: 1002.5028 [math.CO].

[23] K. Tikhomirov. Singularity of random bernoulli matrices, 2019. arXiv: 1812.09016 [math.PR].

[24] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, number 47 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018. ISBN: 978-1-108-41519-4.

[25] J. Yao, S. Zheng, and Z. Bai. *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Mar. 2015. ISBN: 9781107065178. DOI: 10.1017/CBO9781107588080.