Zur Spektralnorm hochdimensionaler Gaußscher Zufallsmatrizen

Bachelorarbeit

Max Justus Löchel

18. August 2020

Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg

Betreuer: Prof. Hajo Holzmann

Inhaltsverzeichnis

Kapitel 1

Einleitung

Die Spektralnorm von Zufallsmatrizen mit unabhängig normalverteilten Einträgen abzuschätzen und somit die Größenordnung der Zufallsmatrizen einzuschätzen, ist ein Standardproblem der Wahrscheinlichkeitstheorie, wie etwa in [Dav+01] und [Ver12] zu sehen ist, mit einigen Anwendungen. In dieser Bachelorarbeit „zur Spektralnorm hochdimensionaler Gaußscher Zufallsmatrizen“ steht die obere Abschätzung des Erwartungswerts der Spektralnorm Gaußscher Zufallsmatrizen im Mittelpunkt. Der Begriff *Gaußsche Zufallsmatrix* bezeichnet in unserem Fall eine symmetrische Matrix X, deren Einträge $X_{ij} = b_{ij}g_{ij}$ ergeben sich aus einem Skalar b_{ij} und einer unabhängig standardnormalverteilten Zufallsvariable g_{ij}. Das zentrale Ergebnis dieser Arbeit ist dabei der Satz 3.6, welchen wir im dritten Kapitel formulieren, diskutieren und beweisen werden. Dieser Satz entstammt der Arbeit *Sharp nonasymptotic bounds on the norm of random matrices with independent entries* von Afonso S. Bandeira und Ramon van Handel, [Ban+16, Theorem 1.1]. Dort wird auch noch eine Verallgemeinerung der Abschätzung auf nicht-symmetrische, rechteckige Matrizen gegeben, aber wir werden uns nur mit dem symmetrischen Fall beschäftigen. Insbesondere behandeln wir keinen rein asymptotischen Fall, indem wir uns nur die Grenzwerte ansehen, wie etwa in [Bai+88] oder [Gem+80], sondern das Resultat mit dem wir uns beschäftigen ist explizit auch im nich-asymptotischen Fall gültig, mit dem Ziel in einem möglichst allgemeinen Fall das Wachstum der Norm bis auf universelle Konstanten korrekt zu skalieren. Wir werden sehen, dass die Abschätzung, die wir betrachten, unter milden Annahmen scharf und besser ist, als einige bekannte Abschätzungen, etwa nach [Pis03] und [Dav+01]. Für den Beweis werden wir den Begriff des Gaußprozesses, sowie die Sudakov-Fernique Ungleichung und die Gauß-Konzentration von Wahrscheinlichkeitsmaßen benötigen. Deshalb werden wir uns im Kapitel 2 zunächst mit diesen Grundlagen befassen. Dabei gehen wir zuerst auf einige Eigenschaften der Spektralnorm ein und kommen dann über die multivariate Normalverteilung zu Gaußprozessen. Damit werden wir dann die Sudakov-Fernique Ungleichung beweisen. Außerdem werden wir uns mit der Gauß-Konzentration von Wahrscheinlichkeitsmaßen beschäftigen. Im vierten Kapitel betrachten wir weitere Eigenschaften der Spektralnorm hochdimensionaler Gaußscher Zufallsmatrizen, indem wir neben der oberen Abschätzung der Spektralnorm auch noch eine untere Abschätzung finden und eine dimensionsfreie Abschätzung beweisen. Zudem werden wir uns mit den Tails der Spektralnorm beschäftigen. Zum Schluss werden wir uns noch zwei Beispiele ansehen, wie sich die Spektralnorm und deren Abschätzung in Grenzfällen verhalten.

Kapitel 2

Grundlagen

In diesem Teil möchten wir auf einige wichtige Grundlagen aus der Linearen Algebra und Wahrscheinlichkeitstheorie eingehen. Dabei bewegen wir uns zumeist im reellen euklidischen Vektorraum $\mathbb{R}^n$ mit dem Standardskalarprodukt.

2.1 Symmetrische Matrizen und die Spektralnorm

Die Sätze zur Linearen Algebra in diesem Abschnitt orientieren sich an [Rol18]. Zunächst betrachten wir das Skalarprodukt:

Definition 2.1. Wir nennen die Abbildung

$$\langle -, - \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, \quad \langle v, w \rangle = v^\top w = \sum_{i=1}^{n} v_i w_i$$

das *Standardskalarprodukt im* $\mathbb{R}^n$.

Bemerkung 2.2. *Eigenschaften des Standardskalarprodukts.*
Bilinearität: Für alle $u, v, w \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$ gilt

$$\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \langle v, w \rangle$$

$$\langle u, v + \lambda w \rangle = \langle u, v \rangle + \lambda \langle u, w \rangle.$$

Symmetrie: Für alle $v, w \in \mathbb{R}^n$ gilt

$$\langle v, w \rangle = \langle w, v \rangle.$$

Positive Definitheit: Für alle $v \in \mathbb{R}^n$ gilt

$$\langle v, v \rangle \geq 0$$

und $\langle v, v \rangle = 0$ genau dann, wenn $v = 0$.
Verschiebungseigenschaft: Sei $A \in \mathbb{R}^{n \times n}$ eine Matrix, sowie $A^\top$ die transponierte von A, dann gilt

$$\langle Av, w \rangle = (Av)^\top w = v^\top A^\top w = \langle v, A^\top w \rangle.$$

Aus dem Skalarprodukt lässt sich eine Norm herleiten, da es auch um die Spektralnorm gehen soll, definieren wir zunächst den Begriff der Norm:

Definition 2.3. Sei V ein $\mathbb{K}$-Vektorraum, $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$. Wir nennen eine Abbildung

$$\|\cdot\| : V \to \mathbb{R}_0^+$$

eine Norm, falls
1. (Definitheit) für alle $v \in V$ gilt $\|v\| = 0 \Leftrightarrow v = 0$.
2. (Absolute Homogenität) für alle $v \in V$ und $\lambda \in \mathbb{K}$ gilt $\|\lambda v\| = |\lambda|\|v\|$.
3. (Dreiecksungleichung) für alle $v, w \in V$ gilt $\|v + w\| \leq \|v\| + \|w\|$.
Wir nennen das Paar $(V, \|\cdot\|)$ einen *normierten Raum.*

Beispiel 2.4. Wir betrachten den Vektorraum $\mathbb{R}^n$ mit dem Standardskalarprodukt, dann ist eine Norm gegeben durch

$$v = (v_1, \ldots, v_n) \to \|v\| = \sqrt{\langle v, v\rangle} = \left(\sum_{i=1}^{n} v_i^2\right)^{1/2}.$$

Wir nennen diese Norm auch *euklidische Norm* oder *2-Norm.*
Notation: $\|\cdot\|_2$.

Satz 2.5 (Cauchy-Schwarzsche Ungleichung)**.** *Wir betrachten die 2-Norm mit dem Standardskalarprodukt auf dem $\mathbb{R}^n$, dann gilt*

$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$

Weiter betrachten wir eine Verallgemeinerung des Kosinussatzes:

Satz 2.6. *Sei der $\mathbb{R}^n$ mit dem Standardskalarprodukt und der 2-Norm gegeben, dann gilt für Vektoren $v, w \in \mathbb{R}^n$*

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2\langle v, w\rangle.$$

Beweis. Wir rechnen elementar aus:

$$\begin{aligned} \|v - w\|^2 &= \langle v - w, v - w\rangle = \langle v, v\rangle - \langle v, w\rangle - \langle w, v\rangle + \langle w, w\rangle \\ &= \langle v, v\rangle + \langle w, w\rangle - 2\langle v, w\rangle = \|v\|^2 + \|w\|^2 - 2\langle v, w\rangle. \end{aligned}$$

□

Den Begriff der Norm auf allgemeinen Vektorräumen können wir auf Matrizen spezifizieren:

Definition 2.7. Seien $V = \mathbb{K}^n$ und $W = \mathbb{K}^m$ zwei normierte Vektorräume, $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, mit Normen $\|\cdot\|_V$ und $\|\cdot\|_W$, sowie $A \in \mathbb{K}^{m\times n}$ eine Matrix. Dann heißt

$$\|A\| = \max_{x\neq 0} \frac{\|Ax\|_W}{\|x\|_V} = \max_{\|x\|_V=1} \|Ax\|_W$$

die *natürliche Matrixnorm von A.*

Bemerkung 2.8. Die Matrixnorm erfüllt die Eigenschaften einer Norm:
Definitheit: Da $\|\cdot\|_V$ und $\|\cdot\|_W$ Normen sind folgt aus der Definition, dass $\|A\| \geq 0$. Weiter gilt

$$\|A\| = 0 \quad \Leftrightarrow \quad \|Ax\| = 0 \quad \forall x \in \mathbb{K}^n \quad \Leftrightarrow \quad Ax = 0 \quad \forall x \in \mathbb{K}^n \quad \Leftrightarrow \quad A = 0.$$

Absolute Homogenität: Wir rechnen für $\lambda \in \mathbb{K}$

$$\|\lambda A\| = \max_{x \neq 0} \frac{\|\lambda A x\|_W}{\|x\|_V} = \max_{x \neq 0} \frac{|\lambda| \|Ax\|_W}{\|x\|_V} = |\lambda| \|A\|.$$

Dreiecksungleichung: Für $A, B \in \mathbb{K}^{m \times n}$ gilt

$$\begin{aligned} \|A + B\| &= \max_{\|x\|_V = 1} \|Ax + Bx\|_W \leq \max_{\|x\|_V = 1} \|Ax\|_W + \|Bx\|_W \\ &\leq \max_{\|x\|_V = 1} \|Ax\|_W + \max_{\|y\|_V = 1} \|By\|_W = \|A\| + \|B\| \end{aligned}$$

Und schließlich folgt der Spezialfall der Spektralnorm, die im Mittelpunkt dieser Arbeit steht:

Definition 2.9. Sei $A \in \mathbb{K}^{m \times n}$, dann ist die *Spektralnorm von A* die von der 2-Norm abgeleitete natürliche Matrixnorm von A:

$$\|A\| = \|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2 = 1} \|Ax\|_2.$$

Aus Bemerkung 2.8 folgt dann schon, dass die Spektralnorm die Normeigenschaften erfüllt. Im weiteren Verlauf werden wir noch einige wichtige Eigenschaften symmetrischer Matrizen und der Spur benötigen auf welche wir nun eingehen.

Definition 2.10. **1.** Sei A eine $n \times n$-Matrix mit Einträgen a_{ij}, $1 \leq i, j \leq n$, dann nennen wir

$$\mathrm{Tr}(A) := \sum_{i=1}^{n} a_{ii}$$

die *Spur* der Matrix A.
2. Sei weiter $A^\top$ die transponierte Matrix von A, dann nennen wir A eine *symmetrische Matrix*, falls $A = A^\top$.

Ein entscheidendes Resultat ist folgender Satz:

Satz 2.11 (Spektralsatz für symmetrische Matrizen)**.** *Ist $A \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix, so ist A orthogonal diagonalisierbar, insbesondere hat A nur reelle Eigenwerte.*

Zum Beweis siehe [Jän08, Kapitel 10.3].

Bemerkung 2.12. Die Spur ist invariant unter Basistransformation, d.h. für eine $n \times n$-Matrix A und eine invertierbare $n \times n$-Matrix B ist

$$\mathrm{Tr}(A) = \mathrm{Tr}(B^{-1} A B).$$

Eine symmetrische Matrix A ist stets selbstadjungiert, d.h.

$$\langle Av, w \rangle = \langle v, A^\top w \rangle = \langle v, Aw \rangle.$$

Weiter ist dann
$$\mathrm{Tr}(A) = \lambda_1 + \ldots + \lambda_n.$$
Dies folgt direkt aus der Invarianz der Spur unter Basistransformation und dem Spektralsatz für symmetrische Matrizen.
Zudem gilt für eine symmetrische Matrix A, dass
$$(AA)^\top = A^\top A^\top = AA$$
und somit dass alle ganzzahligen Potenzen einer symmetrischen Matrix ebenfalls wieder symmetrisch sind. Zusammen mit den Potenzen der diagonalisierten Matrizen $B^{-1}AB$ ergibt sich für $p \geq 1$
$$\mathrm{Tr}(A^p) = \lambda_1^p + \ldots + \lambda_n^p.$$

Wir möchten noch auf den Zusammenhang zwischen Eigenwerten und der Spektralnorm symmetrischer Matrizen eingehen.

Proposition 2.13. *Sei $A \in \mathbb{R}^{n\times n}$ eine symmetrische Matrix mit Eigenwerten $\lambda_1, \ldots, \lambda_n$, dann gilt für die Spektralnorm von A*
$$\|A\|_2 = \max\{|\lambda_1|, \ldots, |\lambda_n|\}.$$

Beweis. Nach dem Spektralsatz für symmetrische Matrizen können wir eine Orthonormalbasis $(v_1, \ldots, v_n)$ des $\mathbb{R}^n$ zu den Eigenwerten $\lambda_1, \ldots, \lambda_n$ von A wählen.
Wir zeigen zuerst $\|A\|_2 \geq \max\{|\lambda_1|, \ldots, |\lambda_n|\}$: Sei v_i, $1 \leq i \leq n$ ein Eigenvektor zum Eigenwert λ_i aus der Orthonormalbasis. Dann ist
$$\frac{\|Av_i\|_2}{\|v_i\|_2} = \|\lambda_i v_i\|_2 = |\lambda_i|\|v_i\|_2 = |\lambda_i|.$$
Maximumsbildung liefert dann die gewünschte Ungleichung.
Andererseits gilt auch $\|A\|_2 \leq \max\{|\lambda_1|, \ldots, |\lambda_n|\} =: \lambda$. Sei $x = \sum_{i=1}^n a_i v_i \in \mathbb{R}^n$ mit $\|x\|_2 = 1$ beliebig. Dann gilt
$$1 = \|x\|_2^2 = \Big\| \sum_{i=1}^n a_i v_i \Big\|_2^2 = \Big\langle \sum_{i=1}^n a_i v_i, \sum_{i=1}^n a_i v_i \Big\rangle = \sum_{i,j=1}^n a_i a_j \langle v_i, v_j\rangle \overset{\text{ONB}}{=} \sum_{i=1}^n a_i^2.$$
Damit ergibt sich
$$\begin{aligned}
\|Ax\|_2 &= (\langle Ax, Ax\rangle)^{\frac{1}{2}} = \left(\left\langle A\sum_{i=1}^n a_i v_i, A\sum_{i=1}^n a_i v_i\right\rangle\right)^{\frac{1}{2}} \\
&= \left(\left\langle \sum_{i=1}^n \lambda_i a_i v_i, \sum_{i=1}^n \lambda_i a_i v_i\right\rangle\right)^{\frac{1}{2}} = \left(\sum_{i,j=1}^n a_i a_j \lambda_i \lambda_j \langle v_i, v_j\rangle\right)^{\frac{1}{2}} \\
&\overset{\text{ONB}}{=} \left(\sum_{i=1}^n a_i^2 \lambda_i^2\right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^n a_i^2 \lambda^2\right)^{\frac{1}{2}} = \left(\lambda^2\right)^{\frac{1}{2}} = \lambda.
\end{aligned}$$
Damit ist $\|A\|_2$ durch $\max\{|\lambda_1|, \ldots, |\lambda_n|\}$ beschränkt und damit folgt die Behauptung. $\square$

Lemma 2.14. *Sei $A \in \mathbb{R}^n$ eine symmetrische Matrix, dann gilt für*

$$\lambda_+ := \sup_{v \in S^{n-1}} \langle v, Av \rangle \quad \text{und} \quad \lambda_- := -\inf_{v \in S^{n-1}} \langle v, Av \rangle$$

dass

$$\|A\| = \max\{\lambda_+, \lambda_-\}.$$

Beweis. Analog zum Beweis von Proposition 2.13 können wir eine Orthonormalbasis $(v_1, \dots, v_n)$ des $\mathbb{R}^n$ zu den Eigenwerten $\lambda_1, \dots, \lambda_n$ von A wählen. Wir setzen $\lambda := \max\{|\lambda_1|, \dots, |\lambda_n|\}$.
Fall 1: A ist positiv definit. Dann können wir eine Matrix B so wählen, dass $B^2 = A$ und B ist symmetrisch mit Eigenwerten $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$, Details siehe Bemerkung 2.27. Wir erkennen, dass

$$\begin{aligned} \max\{\lambda_+, \lambda_-\} &= \sup_{v \in S^{n-1}} |\langle v, Av \rangle| = \sup_{v \in S^{n-1}} |\langle v, B^2 v \rangle| \\ &= \sup_{v \in S^{n-1}} \left(\sqrt{\langle Bv, Bv \rangle}\right)^2 = \|B\|_2^2 = (\max\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\})^2 \\ &= \lambda = \|A\|. \end{aligned}$$

Fall 2: A ist nicht positiv definit. Es nach Proposition 2.13

$$\lambda = \sup_{v \in S^{n-1}} |\langle v, Av \rangle|$$

zu zeigen. Für den Eigenvektor v, $\|v\| = 1$, zu λ gilt aber schon

$$\langle v, Av \rangle = \langle v, \lambda v \rangle = \lambda \langle v, v \rangle = \lambda_i$$

und somit ist

$$\lambda \leq \sup_{v \in S^{n-1}} |\langle v, Av \rangle|.$$

Andererseits gilt für $x \in S^{n-1}$

$$\begin{aligned} |\langle x, Ax \rangle| &= \left|\left\langle \sum_{i=1}^n a_i v_i, A \sum_{i=1}^n a_i v_i \right\rangle\right| = \left|\left\langle \sum_{i=1}^n a_i v_i, \sum_{i=1}^n \lambda_i a_i v_i \right\rangle\right| \\ &= \Big| \sum_{i,j=1}^n a_i a_j \lambda_j \langle v_i, v_j \rangle \Big| \overset{\text{ONB}}{=} \Big| \sum_{i=1}^n a_i^2 \lambda_i \Big| \leq \Big| \sum_{i=1}^n a_i^2 \lambda \Big| \leq \lambda. \end{aligned}$$

und somit

$$\lambda \geq \sup_{v \in S^{n-1}} |\langle v, Av \rangle|.$$

□

Weiter noch eine Definition von Graphen von der wir später noch Gebrauch machen werden.

Definition 2.15. Ein geordnetes Paar (V, E) der Mengen V und E nennen wir *Graph*, dabei ist V die Menge der Ecken und E die Mengen der Kanten. Wir nennen (V, E) einen *ungerichteten Graph ohne Mehrfachkanten*, falls E eine Teilmenge aller 2-elementigen Teilmengen von V ist.

Bemerkung 2.16. Topologisch können wir einen Graph auch als einen CW-Komplex der Dimension ≤ 1 auffassen. Dann sind die 0-Zellen die Ecken und die 1-Zellen die Kanten.

2.2 Normalverteilung und Gaußprozesse

Dieser Abschnitt orientiert sich an [Hol20]. Zunächst möchten wir wichtige Erkenntnisse aus der Wahrscheinlichkeitstheorie erwähnen, welche wir später noch gebrauchen werden. Dies ist insbesondere die Jensensche Ungleichung und die $L^p - Norm$.

Satz 2.17 (Jensensche Ungleichung)**.** *[Bib+19, Satz 1.15] Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $I \in \mathbb{R}$ ein Intervall und $X : \Omega \to I$ eine integrierbare Zufallsvariable, dann ist $\mathbb{E}[X] \in I$. Falls $\phi : I \to \mathbb{R}$ konvex ist, so ist $\phi(X)$ quasiintegrierbar und es gilt*

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)].$$

Weiter ist, falls ϕ strikt konvex und X fast sicher nicht konstant ist, sogar

$$\phi(\mathbb{E}[X]) < \mathbb{E}[\phi(X)].$$

Zum Beweis verweisen wir hier auf [Bib+19, Satz1.15].

Korollar 2.18. *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $I \in \mathbb{R}$ ein Intervall und $X : \Omega \to I$ eine integrierbare Zufallsvariable, dann ist $\mathbb{E}[X] \in I$. Falls $\phi : I \to \mathbb{R}$ konkav ist, so ist $\phi(X)$ quasiintegrierbar und es gilt*

$$\phi(\mathbb{E}[X]) \geq \mathbb{E}[\phi(X)].$$

Weiter ist, falls ϕ strikt konkav und X fast sicher nicht konstant ist, sogar

$$\phi(\mathbb{E}[X]) > \mathbb{E}[\phi(X)].$$

Beweis. Ist ϕ konkav, so gilt für alle $x, y \in I$, $\lambda \in [0,1]$

$$\begin{array}{rrcl} & \phi(\lambda x + (1-\lambda)y) & \geq & \lambda\phi(x) + (1-\lambda)\phi(y) \\ \Leftrightarrow & -\phi(\lambda x + (1-\lambda)y) & \leq & \lambda(-\phi(x)) + (1-\lambda)(-\phi(y)). \end{array}$$

Also ist $-\phi$ konvex und wir können die Jensensche Ungleichung anwenden:

$$\begin{array}{rrcl} & -\phi(\mathbb{E}[X]) & \leq & \mathbb{E}[-\phi(X)] \\ \Leftrightarrow & \phi(\mathbb{E}[X]) & \geq & \mathbb{E}[\phi(X)] \end{array}$$

Der strikt konkave Fall folgt analog. □

Bemerkung 2.19. Wir definieren für eine Zufallsvariable $X : \Omega \to \mathbb{R}$ und $p \geq 1$ mit $\mathbb{E}[|X|^p] < \infty$ die $\mathcal{L}^p$-Seminorm von X als

$$\|X\|_{\mathcal{L}^p} := (\mathbb{E}[|X|^p])^{1/p}.$$

Seminorm bedeutet in diesem Fall, dass die Normaxiome bis auf die Definitheit gegeben sind. Dies gilt im Übrigen auch für allgemeine Maßräume $(\Omega, \mathcal{A}, \mu)$. Setzt man $\mathcal{L}_0^p := \{f \in \mathcal{L}^p | f = 0 \text{ fast sicher}\}$, dann ist

$$L^p(\Omega, \mathcal{A}, \mu) := \mathcal{L}^p(\Omega, \mathcal{A}, \mu)/\mathcal{L}_0^p(\Omega, \mathcal{A}, \mu)$$

sogar für $\|[f]\|_p$ ein normierter Raum.
Aus der Jensenschen Ungleichung folgt weiter, falls $0 < r < s$ und $\mathbb{E}[|X|^s] < \infty$, dann gilt $\mathbb{E}[|X|^r] < \infty$ und

$$\|X\|_{\mathcal{L}^r} \leq \|X\|_{\mathcal{L}^s}.$$

Kommen wir nun zu normalverteilten Zufallsvariablen, welche im Mittelpunkt dieser Arbeit stehen. Wir beginnen mit dem univariaten Fall.

Definition 2.20. Wir nennen eine stetige Zufallsvariable X *normalverteilt*, $X \sim N(\mu, \sigma^2)$ falls X die Dichte- und Verteilungsfunktion

$$f(x;\mu,\sigma^2) = \tfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\tfrac{(x-\mu)^2}{2\sigma^2}\right), \quad F(x;\mu,\sigma^2) = \tfrac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

hat. Weiter nennen wir X *standardnormalverteilt*, falls $X \sim N(0,1)$. Dann schreiben wir für die Dichte- und Verteilungsfunktion von X

$$\varphi(x) = \tfrac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{x^2}{2}\right), \quad \Phi(x) = \tfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{\frac{1}{2}x^2} \mathrm{d}x.$$

Weiter betrachten wir den Fall der multivariaten Normalverteilung, welcher uns hauptsächlich beschäftigen wird. Dazu führen wir zunächst die charakteristische Funktion als Hilfsmittel ein.

Definition 2.21. Für einen d-variaten Zufallsvektor X heißt

$$\varphi_X(t) := \varphi_{\mathbb{P}_X}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \mathrm{d}\mathbb{P}_X(x) = \mathbb{E}\left[e^{i\langle t,X\rangle}\right]$$

die *charakteristische Funktion* von X.

Die letzte Gleichung in der Definition folgt aus der Transformationsformel.

Satz 2.22. *Sei X ein d-variater Zufallsvektor, dann ist durch die charakteristische Funktion φ_X die Verteilung von X schon eindeutig bestimmt.*

Zum Beweis siehe [Bib+19, Satz 1.37]. Und wir definieren die multivariate Normalverteilung.

Definition 2.23. Wir nennen einen Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_d)^\top$ *multivariat standardnormalverteilt*, $\mathbf{X} \sim N(0, I_d)$, falls die Koordinaten $X_1, \ldots, X_d$ unabhängig standardnormalverteilt sind. Dann hat $\mathbf{X}$ die Lebesgue-Dichte

$$f(\mathbf{x}) = \tfrac{1}{(2\pi)^{d/2}} e^{-\mathbf{x}^\top \mathbf{x}/2}, \quad \mathbf{x} = (x_1, \ldots, x_d)^\top$$

und die charakteristische Funktion

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{-\mathbf{t}^\top \mathbf{t}/2}, \quad \mathbf{t} = (t_1, \ldots, t_d)^\top.$$

Weiter nennen wir einen Zufallsvektor $\mathbf{X} \in \mathbb{R}^d$ *multivariat normalverteilt*, falls $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d\times d}$ existieren, wobei Σ symmetrisch und positiv definit ist, so dass $\mathbf{X}$ die charakteristische Funktion

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mu^\top \mathbf{t}} e^{-\mathbf{t}^\top \Sigma \mathbf{t}/2}, \quad \mathbf{t} \in \mathbb{R}^d$$

hat. Wir notieren $\mathbf{X} \sim N(\mu, \Sigma)$.

Einen multivariat normalverteilten Zufallsvektor bezeichnen wir auch als *gaußschen Zufallsvektor*. Nun folgen noch einige wichtige Eigenschaften der multivariaten Normalverteilung.

Satz 2.24. *Wir betrachten den multivariat normalverteilten Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim N(\mu, \Sigma)$ mit $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d\times d}$, dann ist die Lebesguedichte von $\boldsymbol{X}$ für $\boldsymbol{x} \in \mathbb{R}^d$*

$$f(\boldsymbol{x}, \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^\top \Sigma^{-1}(\boldsymbol{x}-\mu)\right).$$

Satz 2.25. *Wir betrachten wieder den multivariat normalverteilten Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim N(\mu, \Sigma)$ mit $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d\times d}$, dann gilt für $A \in \mathbb{R}^{n\times d}$ und $a \in \mathbb{R}^n$, dass*

$$A\boldsymbol{X} + a \sim N(A\mu + a, A\Sigma A^\top).$$

Beweis. Wir betrachten die charakteristische Funktion von $A\mathbf{X} + a$ für $t \in \mathbb{R}^n$:

$$\begin{aligned}\mathbb{E}\left[e^{i\langle t, A\mathbf{X}+a\rangle}\right] &= \mathbb{E}\left[e^{it^\top(A\mathbf{X}+a)}\right] = e^{it^\top a}\mathbb{E}\left[e^{i(t^\top A)\mathbf{X}}\right] \\ &= e^{it^\top a}e^{i(t^\top A)\mu}e^{-(t^\top A)\Sigma(A^\top t)/2}\end{aligned}$$

Dabei verwenden wir in der letzten Gleichheit, dass $\mathbf{X} \sim N(\mu, \Sigma)$. Die rechte Seite entspricht aber schon der charakteristischen Funktion von $N(A\mu+a, A\Sigma A^\top)$, damit folgt die Aussage. □

Korollar 2.26. *Ist $\boldsymbol{X} \sim N(0, I_d)$, falls $A\boldsymbol{X} + \mu$ für $A \in \mathbb{R}^{d\times d}$ mit $AA^\top = \Sigma$ und $\mu \in \mathbb{R}^d$ die charakteristische Funktion*

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = e^{i\mu^\top \boldsymbol{t}}e^{-\boldsymbol{t}^\top\Sigma\boldsymbol{t}/2}, \quad \boldsymbol{t} \in \mathbb{R}^d$$

hat, so gilt $A\boldsymbol{X} + \mu \sim N(\mu, \Sigma)$.

Bemerkung 2.27. Sei die selbe Situation wie im obigen Korollar gegeben. Da Σ symmetrisch ist, existiert nach Satz 2.11 eine Spektralzerlegung $\Sigma = Q \operatorname{diag}(\lambda_1, \ldots, \lambda_d)\, Q^\top$, wobei Q eine orthogonale $d \times d$-Matrix ist und $\operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ die Diagonalmatrix mit den Eigenwerten von Σ auf der diagonalen. Dann ist $A = Q \operatorname{diag}(\lambda_1^{1/2}, \ldots, \lambda_d^{1/2})$. Wir bezeichnen ein solche Matrix A im folgenden auch als $\Sigma^{1/2}$.

Proposition 2.28. *Sei $X_v = \langle v, g\rangle$ für einen ein standardnormalverteilten Zufallsvektor $g \in \mathbb{R}^d$, $g \sim N(0, I_d)$, und $v \in \mathbb{R}^d$. Dann gilt*

$$\mathit{Var}(X_v) = \|v\|^2$$

und insbesondere $X_v \sim N(0, \|v\|_2^2)$.
Ist weiter $X_w = \langle w, g\rangle$ mit $w \in \mathbb{R}^d$ eine weiter Zufallsvariable, dann gilt

$$\mathit{Var}(X_v - X_w) = \|v - w\|^2$$

und insbesondere $(X_v - X_w) \sim N(0, \|v - w\|_2^2)$.

Bemerkung 2.29. Aus Proposition 2.28 folgt für $X_v - X_w$ direkt

$$\|X_v - X_w\|_{\mathcal{L}^2} = \operatorname{Var}(X_v - X_w)^{1/2} = \|v - w\|_2.$$

Beweis von Proposition 2.28. Nach Satz 2.25 folgt direkt, dass $X_v = v^\top g \sim N(v^\top 0, v^\top I_d v)$, wobei $v^\top I_d v = \langle v, v\rangle = \|v\|_2^2$, also $X_v \sim N(0, \|v\|_2^2)$. Für $(X_v - X_w)$ gilt

$$X_v - X_w = \langle g, v\rangle - \langle g, w\rangle = \langle g, v - w\rangle =: X_{v-w}$$

wegen der Bilinearität des Standardskalarprodukts. Die zweite Aussage folgt dann mit der ersten Teilaussage. □

Auf Basis der multivariaten Normalverteilung können wir nun den Begriff des Gaußprozesses definieren:

Definition 2.30. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(E, \mathcal{E})$ ein messbarer Raum. Dann nennen wir eine Familie von E-wertigen Zufallsvariablen $\mathcal{X} = (X_t)_{t\in T}$ mit $X_t : \Omega \to E$ einen *stochastischen Prozess mit Zustandsraum* $(E, \mathcal{E})$. Dabei ist T eine beliebige Indexmenge, falls $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ ist, nennen wir den Prozess *reellwertig*.

Definition 2.31. Sei $\mathcal{X} = (X_t)_{t\in T}$ ein stochastischer Prozess, dann nennen wir für jede endliche Teilmenge $\{t_1, \dots, t_n\} \subset T$ die Wahrscheinlichkeitsverteilung $\mathbb{P}_{(X_{t_1},\dots,X_{t_n})}$ *endlichdimensionale Randverteilung von* $\mathcal{X}$.

Definition 2.32. (nach [Hol20]) Ein (reellwertiger) stochastischer Prozess $\mathcal{X} = (X_t)_{t\in T}$ heißt *Gaußprozess*, falls alle endlichdimensionalen Randverteilungen von $\mathcal{X}$ multivariat normalverteilt sind.

Im folgenden betrachten wir bestimmte Eigenschaften von Gaußprozessen noch etwas genauer:

Definition 2.33. Wir betrachten einen reellwertigen stochastischen Prozess $\mathcal{X} = (X_t)_{t\in T}$ mit Indexmenge T. Falls $\mathbb{E}\left[|X_t|\right] < \infty$ ist, so nennen wir $m(t) = \mathbb{E}\left[X_t\right]$, $t \in T$, die *Erwartungswertfunktion* von $\mathcal{X}$, sowie falls $\mathbb{E}\left[X_t^2\right] < \infty$ ist, so heißt $\gamma(s,t) = \mathrm{Cov}(X_s, X_t)$, $s, t \in T$, die *Kovarianzfunktion* von $\mathcal{X}$.

Satz 2.34. *Durch die Erwartungswertfunktion und Kovarianzfunktion eines Gaußprozesses sind die endlichdimensionalen Randverteilungen eindeutig bestimmt.*

Beweis. Wenn wir die Erwartungswert- und Kovarianzfunktion wie in der vorherigen Definition mit $m(t) = \mathbb{E}\left[X_t\right]$ und $\gamma(s,t) = \mathrm{Cov}(X_s, X_t)$ notieren, dann ergibt sich für die multivariate Normalverteilung $(X_1, \dots, X_n)^\top$ mit $n \in \mathbb{N}$, $t_1 < \ldots < t_n$ und $t_1, \dots, t_n \in T$, dass der Erwartungswertvektor

$$\left(\mathbb{E}\left[X_{t_1}\right], \dots, \mathbb{E}\left[X_{t_n}\right]\right)^\top = (m(t_1), \dots, m(t_n))^\top$$

und die Kovarianzmatrix

$$(\mathrm{Cov}(X_{t_i}, X_{t_j}))_{i,j=1,\dots,n} = (\gamma(t_i, t_j))_{i,j=1,\dots,n}$$

ist. □

Wir betrachten nun eine stochastischen Prozess $\mathcal{X} = (X_t)_{t\in T}$ mit $\mathbb{E}\left[X_t\right] = 0$ für alle $t \in T$. Dann ist die Kovarianzfunktion

$$\gamma(t,s) = \mathrm{Cov}(X_s, X_t) = \mathbb{E}\left[X_t X_s\right]$$

für $t, s \in T$ und wir definieren die $\mathcal{L}^2$-*Norm der Inkremente* des stochastischen Prozesses als

$$d(t,s) = \|X_t - X_s\|$$

für $t, s \in T$.

Beispiel 2.35. Wir betrachten wieder, wie in Proposition 2.28, den Zufallsvektor $X_t = \langle g, t\rangle$ für einen multivariat standardnormalverteilten Zufallsvektor $g \sim N(0, I_d)$, $g \in \mathbb{R}^d$, und $t \in T \subseteq \mathbb{R}^d$.
Dann nennen wir den $\mathcal{X} = (X_t)_{t\in T}$ den *kanonischen Gaußprozess*.
Für alle $t \in T$ gilt dann $\mathbb{E}[X_t] = 0$. Die Inkremente des kanonischen Gaußprozesses sind nach Bemerkung 2.29

$$d(t,s) = \|X_t - X_s\|_{\mathcal{L}^2} = \|t - s\|_2.$$

Weiter gilt für die Kovarianzfunktion

$$\gamma(t,s) = \mathbb{E}[X_t X_s] = \mathbb{E}\left[t^\top g g^\top s\right] = t^\top I_d s = \langle t, s\rangle.$$

2.3 Sudakov-Fernique Ungleichung

Der Abschnitt orientiert sich an [Ver20]. Ein wichtiger Baustein im Beweis des zentralen Satzes dieser Arbeit ist folgende Ungleichung:

Satz 2.36 (Sudakov-Fernique Ungleichung)**.** *[Ver20, Theorem 7.2.11] Seien $X = (X_1, \dots, X_n)^\top$ und $Y = (Y_1, \dots, Y_n)^\top$, $X, Y \in \mathbb{R}^n$, zwei gaußsche Zufallsvektoren mit Erwartungswert 0. Weiter sei für alle $s, t \in \{1, \dots, n\}$*

$$\mathbb{E}\left[(X_t - X_s)^2\right] \leq \mathbb{E}\left[(Y_t - Y_s)^2\right],$$

dann gilt

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] \leq \mathbb{E}\left[\max_{1\leq i\leq n} Y_i\right].$$

Die Sudakov-Fernique Ungleichung bildet somit ein wichtiges Hilfsmittel um die Maxima von gaußschen Zufallvektoren gegeneinander abzuschätzen. Dieses Resultat werden wir später noch zur Abschätzung der Spektralnorm gaußscher Zufallsmatrizen benötigen. Über die endlichdimensionalen Randverteilungen eines Gaußprozesses können wir diese Aussage später auch in eine Aussage über Gaußprozesse umwandeln. Eine ähnliche Aussage bildet die Slepian Ungleichung, die allerdings noch eine weitere Voraussetzung hat:

Satz 2.37 (Slepian Ungleichung)**.** *[Ver20, Theorem 7.2.1] Seien $(X_t)_{t\in T}$, sowie $(Y_t)_{t\in T}$ zwei Gaußprozesse mit Erwartungswert 0. Weiter sei für alle $s, t \in T$*

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[Y_t^2\right] \quad \text{und} \quad \mathbb{E}\left[(X_t - X_s)^2\right] \leq \mathbb{E}\left[(Y_t - Y_s)^2\right],$$

dann gilt für jedes $r \in \mathbb{R}$

$$\mathbb{P}\Big\{\sup_{t\in T} X_t \geq r\Big\} \leq \mathbb{P}\Big\{\sup_{t\in T} Y_t \geq r\Big\}$$

und insbesonders

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq \mathbb{E}\left[\sup_{t\in T} Y_t\right].$$

Da wir nur die etwas allgemeinere Aussage der Sudakov-Fernique Ungleichung benötigen, werden wir nur diese beweisen, der Beweis der Slepian Ungleichung ist dem sehr ähnlich. Für den Beweis der Sudakov-Fernique Ungleichung benötigen wir den Begriff der Gaußinterpolation. Dafür betrachten wir zunächst folgendes Lemma zur partiellen Gaußintegration:

Lemma 2.38. *[Ver20, Lemma 7.2.5] Sei $X \sim N(0, \Sigma)$, $X \in \mathbb{R}^d$, ein multivariat normalverteilter Zufallsvektor mit Erwartungswert 0. Dann gilt für jede differenzierbare Funktion $f : \mathbb{R}^d \to \mathbb{R}$, welche maximal, wie auch ihre partiellen Ableitungen, exponentiell wächst, dass*

$$\mathbb{E}\left[Xf(X)\right] = \Sigma \cdot \mathbb{E}\left[\nabla f(X)\right].$$

Das maximal exponentielle Wachstum von f bedeutet, dass für $c_1, c_2 \in \mathbb{R}$, $c_1, c_2 > 0$, $x \in \mathbb{R}^d$ gilt:

$$|f(x)| \leq c_1 e^{c_2 \|x\|} \quad \text{und für } 1 \leq i \leq d \quad \left|\frac{\partial f}{\partial x_i}(x)\right| \leq c_1 e^{c_2\|x\|}.$$

Wir betrachten für den Beweis zunächst den eindimensionalen Fall:

Lemma 2.39. *Sei $X \sim N(0, \sigma^2)$ normalverteilt, dann gilt für eine differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $\lim_{x\to\pm\infty} g(x)f(x) = 0$, wobei $f(x)$ die Lebesguedichte von X ist, dass*

$$\mathbb{E}\left[Xg(X)\right] = \sigma^2 \mathbb{E}\left[g'(X)\right].$$

Beweis von Lemma 2.39. Wir betrachten die Dichte von X

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

sowie die Ableitung

$$f'(x) = -\frac{x}{\sigma^2}\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

und erhalten

$$\sigma^2 f'(x) = -xf(x).$$

Nun wenden wir dies und partielle Integration an und erhalten:

$$\begin{aligned} \sigma^2\mathbb{E}\left[g'(X)\right] &= \sigma^2 \int_{\mathbb{R}} g'(x)f(x)\mathrm{d}x \\ &= \sigma^2 \left([g(x)f(x)]_{-\infty}^{+\infty} - \int_{\mathbb{R}} g(x)f'(x)\mathrm{d}x \right) \\ &= -\int_{\mathbb{R}} g(x)\sigma^2 f'(x)\mathrm{d}x \\ &= \int_{\mathbb{R}} g(x)xf(x)\mathrm{d}x = \mathbb{E}\left[Xg(X)\right]. \end{aligned}$$

Dabei ist $[g(x)f(x)]_{-\infty}^{+\infty} = 0$, wegen der oben genannten asymptotischen Voraussetzung an g. □

Damit können wir nun zum Beweis der partiellen Gaußintegration im multivariaten Fall kommen:

Beweis von Lemma 2.38. Sei $X \sim N(0, \Sigma)$, $X \in \mathbb{R}^d$ gegeben. Wir notieren $X = (X_1, \ldots, X_d)^\top$, dann ist die Aussage aus dem Lemma für $1 \leq i \leq d$ äquivalent zu

$$\mathbb{E}\left[X_i f(X)\right] = \sum_{j=1}^n \mathrm{Cov}(X_i, X_j)\mathbb{E}\left[\frac{\partial f}{\partial x_j}(X)\right] = \sum_{j=1}^n \mathbb{E}\left[X_i, X_j\right]\mathbb{E}\left[\frac{\partial f}{\partial x_j}(X)\right],$$

welches die komponentenweise Betrachtung der Matrixmultiplikation aus dem Lemma ist. Die Kovarianz können wir wegen $\mu = 0$ umschreiben in $\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j]$. Wir zeigen nun diese äquivalente Aussage. Wenn $\sigma_i^2 := \mathrm{Var}(X_i) > 0$ ist, dann definieren wir den gaußschen Zufallsvektor $X^{(i)} = (X_1^{(i)}, \dots, X_d^{(i)})^\top$ für $1 \leq j \leq d$ als

$$X_j^{(i)} := X_j - \lambda_j X_i \quad \text{mit} \quad \lambda_j^{(i)} = \tfrac{\mathbb{E}[X_i X_j]}{\sigma_i^2}.$$

Dadurch sind X_i und $X_j^{(i)}$ unabhängig, denn

$$\mathrm{Cov}(X_i, X_j^{(i)}) = \mathbb{E}\left[X_i X_j^{(i)}\right] = \mathbb{E}[X_i X_j] - \lambda_j^{(i)} \sigma_i^2 = 0.$$

Weiter notieren wir

$$\lambda^{(i)} = (\lambda_1^{(i)}, \dots, \lambda_d^{(i)})^\top \quad \text{und erhalten} \quad X = X^{(i)} + X_i \lambda^{(i)}.$$

Wir betrachten zunächst den Erwartungswert $\mathbb{E}_i$ in X_i für festes $X^{(i)}$. Nun können wir die univariate partielle Gaußintegration anwenden, welches wegen der Wachstumsbedingungen an f möglich ist, und erhalten

$$\mathbb{E}_i[X_i f(X)] = \mathbb{E}_i\left[X_i f(X^{(i)} + X_i \lambda^{(i)})\right] = \sigma_i^2 \mathbb{E}_i\left[\frac{\partial f}{\partial t}(X^{(i)} + t\lambda^{(i)})\Big|_{t=X_i}\right].$$

Wir nehmen an, dass

$$|X_1 f(X^{(i)} + X_i \lambda^{(i)})| \quad \text{und} \quad \left|\tfrac{\partial f}{\partial t}(X^{(i)} + t\lambda^{(i)})\Big|_{t=X_i}\right|$$

über $X^{(i)}$ integrierbar sind. Dann können wir über $X^{(i)}$ integrieren, indem wir den Satz von Fubini anwenden und erhalten

$$\mathbb{E}[X_i f(X)] = \mathbb{E}\left[X_i f(X^{(i)} + X_i \lambda^{(i)})\right] = \sigma_i^2 \mathbb{E}\left[\frac{\partial f}{\partial t}(X^{(i)} + t\lambda^{(i)})\Big|_{t=X_i}\right].$$

Wir rechnen nun noch die partiellen Ableitungen auf der rechten Seite aus

$$\frac{\partial f}{\partial t}(X^{(i)} + t\lambda^{(i)})\Big|_{t=X_i} = \sum_{i=1}^{d} \lambda_j^{(i)} \frac{\partial f}{\partial x_j}(X^{(i)} + X_i \lambda^{(i)}) = \sum_{i=1}^{d} \lambda_j^{(i)} \frac{\partial f}{\partial x_j}(X^{(i)} + X_i \lambda^{(i)})$$

und setzen diese ein

$$\mathbb{E}[X_i f(X)] = \sigma_i^2 \mathbb{E}\left[\sum_{i=1}^{d} \lambda_j^{(i)} \frac{\partial f}{\partial j}(X^{(i)} + X_i \lambda^{(i)})\right] = \sum_{j=1}^{n} \mathbb{E}[X_i, X_j]\, \mathbb{E}\left[\frac{\partial f}{\partial x_j}(X)\right]$$

und erhalten, was zu zeigen war. $\square$

Mit der partiellen Gaußintegration können wir die Gaußinterpolation formulieren, welche wir im Beweis der Sudakov-Fernique Ungleichung verwenden werden. Dazu betrachten wir für eine (endliche) Indexmenge T, $|T| = d$, die gaußschen Zufallsvektoren $X = (X_t)_{t \in T}$, sowie $Y = (Y_t)_{t \in T}$, $X, Y \in \mathbb{R}^d$, welche unabhängig sind. Dann definieren wir den gaußschen Vektor $Z(u) \in \mathbb{R}^d$ als jenen Vektor, der mit

$$Z(u) := \sqrt{u}X + \sqrt{1-u}Y, \quad u \in [0,1]$$

stetig zwischen X und Y interpoliert. Aus der Unabhängigkeit von X und Y folgt für die Kovarianzen die Interpolation mit $1 \leq i, j \leq d$

$$\begin{aligned}
\mathrm{Cov}(Z_i(u), Z_j(u)) \quad &= \mathbb{E}\left[Z_i(u)Z_j(u)\right] \\
&= \mathbb{E}\left[(\sqrt{u}X_i + \sqrt{1-u}Y_i)(\sqrt{u}X_j + \sqrt{1-u}Y_j)\right] \\
&= \mathbb{E}\left[uX_iX_j + (1-u)Y_iY_j\right] \\
&= u\mathrm{Cov}(X_i, X_j) + (1-u)\mathrm{Cov}(Y_i, Y_j).
\end{aligned}$$

Weiter gilt $Z(0) = Y$ und $Z(1) = X$.

Lemma 2.40 (Gaußinterpolation)**.** *[Ver20, Lemma 7.2.7] Wir betrachten zwei unabhängige multivariat normalverteilte Zufallsvektoren $X \sim N(0, \Sigma_X)$ und $Y \sim N(0, \Sigma_Y)$, $X, Y \in \mathbb{R}^d$, sowie die Gaußinterpolation*

$$Z(u) := \sqrt{u}X + \sqrt{1-u}Y, \quad u \in [0,1]$$

Dann gilt für jede zweimal differenzierbare Funktion $f : \mathbb{R}^d \to \mathbb{R}$ mit höchstens exponentiellen Wachstum ihrer partiellen Ableitungen, wie oben beschrieben, dass

$$\frac{d}{du}\mathbb{E}\left[f(Z(u))\right] = \frac{1}{2}\sum_{i,j=1}^{d}\big(Cov(X_i, X_j) - Cov(Y_i, Y_j)\big)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u))\right].$$

Beweis von Lemma 2.40. Wir berechnen mit der Ketttenregel komponentenweise die Ableitung

$$\begin{aligned}
\tfrac{d}{du}\mathbb{E}\left[f(Z(u))\right] \quad &= \sum_{i=1}^{d}\mathbb{E}\left[\tfrac{\partial f}{\partial x_i}(Z(u))\tfrac{dZ_i}{du}\right] \\
&= \sum_{i=1}^{d}\mathbb{E}\left[\tfrac{\partial f}{\partial x_i}(Z(u))\tfrac{d}{du}(\sqrt{u}X_i + \sqrt{1-u}Y_i)\right] \\
&= \tfrac{1}{2}\sum_{i=1}^{d}\mathbb{E}\left[\tfrac{\partial f}{\partial x_i}(Z(u)) - \left(\tfrac{X_i}{\sqrt{u}} - \tfrac{Y_i}{\sqrt{1-u}}\right)\right] \\
&= \tfrac{1}{2}\left(\sum_{i=1}^{d}\tfrac{1}{\sqrt{u}}\mathbb{E}\left[X_i\tfrac{\partial f}{\partial x_i}(Z(u))\right] - \sum_{i=1}^{d}\tfrac{1}{\sqrt{1-u}}\mathbb{E}\left[Y_i\tfrac{\partial f}{\partial x_i}(Z(u))\right]\right).
\end{aligned}$$

Nun betrachten wir die beiden Summen einzeln. Zuerst

$$\sum_{i=1}^{d}\frac{1}{\sqrt{u}}\mathbb{E}\left[X_i\frac{\partial f}{\partial x_i}(Z(u))\right] = \sum_{i=1}^{d}\frac{1}{\sqrt{u}}\mathbb{E}\left[X_i(g_i(X))\right],$$

wobei

$$g_i(X) = \frac{\partial f}{\partial x_i}\left(\sqrt{u}X + \sqrt{1-u}Y\right)$$

ist. Dabei betrachten wir zunächst den Term nur bezüglich X und bedingen auf (festes) Y. Dann können wir auf $\mathbb{E}\left[X_i(g_i(X))\right]$ die partielle gaußsche Integration anwenden und erhalten

$$\begin{aligned}
\mathbb{E}\left[X_i(g_i(X))\right] &= \sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j)\mathbb{E}\left[\frac{\partial g_i}{\partial x_j}(X)\right] \\
&= \sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\left(\sqrt{u}X + \sqrt{1-u}Y\right) \cdot \sqrt{u}\right] \\
&= \sqrt{u} \sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\left(Z(u)\right)\right],
\end{aligned}$$

wobei der Faktor $\sqrt{u}$ durch das Ableiten hinzukommt. Dieses Ergebnis setzen wir nun wieder ein:

$$\begin{aligned}
\sum_{i=1}^{d} \frac{1}{\sqrt{u}}\mathbb{E}\left[X_i \frac{\partial f}{\partial x_i}(Z(u))\right] &= \sum_{i=1}^{d} \frac{1}{\sqrt{u}}\mathbb{E}\left[X_i(g_i(X))\right] \\
&= \sum_{i=1}^{d} \frac{1}{\sqrt{u}}\left(\sqrt{u} \sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\left(Z(u)\right)\right]\right) \\
&= \sum_{i,j=1}^{d} \operatorname{Cov}(X_i, X_j)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\left(Z(u)\right)\right]
\end{aligned}$$

Wir haben zunächst auf (festes) Y bedingt. Das lösen wir nun wieder auf, indem wir auf beiden Seiten den Erwartungswert über Y bilden, das Ergebnis bleibt das Gleiche. Für die Y_i ergibt sich analog

$$\sum_{i=1}^{d} \frac{1}{\sqrt{1-u}}\mathbb{E}\left[Y_i \frac{\partial f}{\partial x_i}(Z(u))\right] = \sum_{i,j=1}^{d} \operatorname{Cov}(Y_i, Y_j)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\left(Z(u)\right)\right].$$

Nun können wir die beiden Summanden wieder zusammenführen:

$$\begin{aligned}
\frac{d}{du}\mathbb{E}\left[f(Z(u))\right] &= \frac{1}{2}\left(\sum_{i=1}^{d} \frac{1}{\sqrt{u}}\mathbb{E}\left[X_i \frac{\partial f}{\partial x_i}(Z(u))\right] - \sum_{i=1}^{d} \frac{1}{\sqrt{1-u}}\mathbb{E}\left[Y_i \frac{\partial f}{\partial x_i}(Z(u))\right]\right) \\
&= \frac{1}{2}\left(\sum_{i,j=1}^{d} \operatorname{Cov}(X_i, X_j)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\left(Z(u)\right)\right]\right. \\
&\left. - \sum_{i,j=1}^{d} \operatorname{Cov}(Y_i, Y_j)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\left(Z(u)\right)\right]\right) \\
&= \frac{1}{2}\sum_{i,j=1}^{d} \left(\operatorname{Cov}(X_i, X_j) - \operatorname{Cov}(Y_i, Y_j)\right)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u))\right]
\end{aligned}$$

□

Mit der Gaußinterpolation können wir nun die Sudakov-Fernique Ungleichung beweisen. Die zu beweisende Ungleichung ist

$$\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] \leq \mathbb{E}\left[\max_{1 \leq i \leq n} Y_i\right]$$

unter der Voraussetzung, dass $X = (X_1, \ldots, X_n)^\top$ und $Y = (Y_1, \ldots, Y_n)^\top$, $X, Y \in \mathbb{R}^n$, zwei gaußsche Zufallsvektoren mit Erwartungswert 0 sind und für alle $s, t \in \{1, \ldots, n\}$

$$\mathbb{E}\left[(X_t - X_s)^2\right] \leq \mathbb{E}\left[(Y_t - Y_s)^2\right]$$

gilt. Dazu werden wir eine Funktion f betrachten, die das $\max_{1\leq i\leq n} x_i$ approximiert und zweimal differenzierbar ist, um dann

$$\mathbb{E}\left[f(X)\right] \leq \mathbb{E}\left[f(Y)\right]$$

mit der Gaußinterpolation abschätzen.

Beweis von Satz 2.36. Es reicht aus die Aussage für zwei gaußsche Zufallsvektoren $X = (X_1, \dots, X_n)^\top$ und $Y = (Y_1, \dots, Y_n)^\top$, $X, Y \in \mathbb{R}^n$, zu beweisen. Zur Approximation des Maximums $\max_{1\leq i\leq n} x_i$ der Vektoren betrachten wir die Funktion

$$f_\beta(x) := \frac{1}{\beta} \log \sum_{i=1}^n e^{\beta x_i}$$

für ein $\beta > 0$ und $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$. Es gilt

$$f_\beta(x) \to \max_{1\leq i\leq n} x_i \qquad \text{für } \beta \to \infty,$$

denn es gilt

$$\max_{1\leq i\leq n} x_i \leq f_\beta(x) = \tfrac{1}{\beta} \log \sum_{i=1}^n e^{\beta x_i} \leq \max_{1\leq i\leq n} x_i + \tfrac{\log n}{\beta} \to \max_{1\leq i\leq n} x_i \qquad \text{für } \beta \to \infty.$$

Die linke Ungleichung folgt aus

$$\max_{1\leq i\leq n} x_i = \frac{1}{\beta} \log e^{\beta \max_{1\leq i\leq n} x_i} \leq \frac{1}{\beta} \log \sum_{i=1}^n e^{\beta x_i}$$

und der Monotonie der Logarithmusfunktion. Die rechte Ungleichung ergibt sich durch das Einsetzen von $\max_{1\leq i\leq n} x_i$ für alle x_i:

$$\begin{aligned} f_\beta(x) = \tfrac{1}{\beta} \log \sum_{i=1}^n e^{\beta x_i} \quad &\leq \tfrac{1}{\beta} \log \sum_{i=1}^n e^{\beta \max_{1\leq i\leq n} x_i} = \tfrac{1}{\beta} \log \left(n e^{\beta \max_{1\leq i\leq n} x_i}\right) \\ &= \tfrac{\log n}{\beta} + \tfrac{1}{\beta} \log \left(\exp\left(\beta \max_{1\leq i\leq n} x_i\right)\right) = \max_{1\leq i\leq n} x_i + \tfrac{\log n}{\beta}, \end{aligned}$$

sowie aus der Monotonie der Exponentialfunktion. Als nächstes setzen wir in die Gaußinterpolation ein und zeigen

$$\frac{d}{du} \mathbb{E}\left[f_\beta(Z(u))\right] \leq 0$$

für $Z(u) = \sqrt{u} X + \sqrt{1-u} Y$. Daraus folgt dann, dass

$$\mathbb{E}\left[f_\beta(Z(0))\right] = \mathbb{E}\left[f_\beta(Y)\right] \geq \mathbb{E}\left[f_\beta(X)\right] = \mathbb{E}\left[f_\beta(Z(1))\right]$$

und somit gilt für $\beta \to \infty$ mit majorisierter Konvergenz

$$\mathbb{E}\left[\max_{1\leq i\leq n} Y_i\right] \geq \mathbb{E}\left[\max_{1\leq i\leq n} X_i\right],$$

was zu zeigen war. Die majorisierte Konvergenz dürfen wir hier anwenden, denn aus den obigen Ungleichungen ergibt sich mit $\max\left(0, \max_{i\leq i\leq n} x_i\right) + \log n$ eine Majorante für alle f_β.

Zeigen wir nun noch $\frac{d}{du}\mathbb{E}\left[f(Z(u))\right] \leq 0$: Für die partielle Ableitung von f in x_i definieren wir

$$p_i(x) := \frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{1}{\beta}\log\sum_{i=1}^n e^{\beta x_i}\right) = \frac{e^{\beta x_i}}{\sum_{j=1}^n e^{\beta x_j}},$$

was wir mit der Kettenregel erhalten. Die zweite partielle Ableitung erhalten wir mit der Quotientenregel

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_i^2} &= \frac{\partial f}{\partial x_i}\left(\frac{e^{\beta x_i}}{\sum_{j=1}^n e^{\beta x_j}}\right) = \frac{\beta e^{\beta x_i}\cdot\sum_{j=1}^n e^{\beta x_j} - e^{\beta x_i}\cdot\beta e^{\beta x_i}}{\left(\sum_{j=1}^n e^{\beta x_j}\right)^2} \\
&= \beta\left(\frac{e^{\beta x_i}\sum_{j=i}^n e^{\beta x_j}}{\left(\sum_{j=1}^n e^{\beta x_j}\right)^2} - \frac{\left(e^{\beta x_i}\right)^2}{\left(\sum_{j=1}^n e^{\beta x_j}\right)^2}\right) = \beta\left(p_i(x) - p_i(x)^2\right)
\end{aligned}$$

und falls $i \neq j$ ist, gilt

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial f}{\partial x_j}\left(\frac{e^{\beta x_i}}{\sum_{k=1}^n e^{\beta x_k}}\right) = \frac{-e^{\beta x_i}\cdot\beta e^{\beta x_j}}{\left(\sum_{k=1}^n e^{\beta x_k}\right)^2} = -\beta p_i(x)p_j(x).$$

Und wir beobachten, dass

$$\sum_{i=1}^n p_i(x) = \sum_{i=1}^n \frac{e^{\beta x_i}}{\sum\limits_{j=1}^n e^{\beta x_j}} = 1$$

ist. Wir multiplizieren beide Seiten zunächst mit $p_i(x)$

$$\sum_{j=1}^n p_i(x)p_j(x) = p_i(x)$$

und subtrahieren $p_i(x)^2$ und erhalten

$$\sum_{j\neq i} p_i(x)p_j(x) = p_i(x) - p_i(x)^2.$$

Damit können wir nun $\frac{d}{du}\mathbb{E}\left[f(Z(u))\right]$ mit der Gaußinterpolation berechnen:

$$\begin{aligned}
\frac{d}{du}\mathbb{E}\left[f(Z(u))\right] &= \frac{1}{2}\sum_{i,j=1}^d \big(\mathrm{Cov}(X_i,X_j) - \mathrm{Cov}(Y_i,Y_j)\big)\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i\partial x_j}(Z(u))\right] \\
&= \frac{1}{2}\sum_{i=1}^n \big(\mathrm{Cov}(X_i,X_i) - \mathrm{Cov}(Y_i,Y_i)\big)\mathbb{E}\left[\beta\left(p_i - p_i^2\right)\right] \\
&+ \frac{1}{2}\sum_{i\neq j}\big(\mathrm{Cov}(X_i,X_j) - \mathrm{Cov}(Y_i,Y_j)\big)\mathbb{E}\left[-\beta p_i p_j\right] \\
&= \frac{\beta}{2}\sum_{i=1}^n \big(\mathrm{Var}(X_i) - \mathrm{Var}(Y_i)\big)\mathbb{E}\big[\big(\sum_{j\neq i} p_i p_j\big)\big] \\
&- \frac{\beta}{2}\sum_{i\neq j}\big(\mathrm{Cov}(X_i,X_j) - \mathrm{Cov}(Y_i,Y_j)\big)\mathbb{E}\left[p_i p_j\right] \\
&= \frac{\beta}{2}\sum_{i\neq j}\big(\mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[Y_i^2\right]\big)\mathbb{E}\left[p_i p_j\right] \\
&- \frac{\beta}{2}\sum_{i\neq j}\big(\mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[Y_i Y_j\right]\big)\mathbb{E}\left[p_i p_j\right]
\end{aligned}$$

Wir können $\sum_{i\neq j}\left(\mathbb{E}\left[X_i^2\right]-\mathbb{E}\left[Y_i^2\right]\right)\mathbb{E}\left[p_ip_j\right]$ auch mit getauschten Indizes als $\sum_{i\neq j}\left(\mathbb{E}\left[X_j^2\right]-\mathbb{E}\left[Y_j^2\right]\right)\mathbb{E}\left[p_ip_j\right]$ schreiben. Wir nehmen dann den Mittelwert der beiden:

$$\sum_{i\neq j}\left(\mathbb{E}\left[X_i^2\right]-\mathbb{E}\left[Y_i^2\right]\right)\mathbb{E}\left[p_ip_j\right]=\frac{1}{2}\sum_{i\neq j}\left(\mathbb{E}\left[X_i^2\right]+\mathbb{E}\left[X_j^2\right]-\mathbb{E}\left[Y_i^2\right]-\mathbb{E}\left[Y_j^2\right]\right)\mathbb{E}\left[p_ip_j\right]$$

Das setzen wir nun in die Gaußinterpolation ein und erhalten

$$\begin{aligned}
\tfrac{d}{du}\mathbb{E}\left[f(Z(u))\right] &= \tfrac{\beta}{4}\sum_{i\neq j}\left(\mathbb{E}\left[X_i^2\right]+\mathbb{E}\left[X_j^2\right]-\mathbb{E}\left[Y_i^2\right]-\mathbb{E}\left[Y_j^2\right]\right)\mathbb{E}\left[p_ip_j\right]\\
&\quad -\tfrac{\beta}{4}\sum_{i\neq j}\left(2\mathbb{E}\left[X_iX_j\right]-2\mathbb{E}\left[Y_iY_j\right]\right)\mathbb{E}\left[p_ip_j\right]\\
&= \tfrac{\beta}{4}\sum_{i\neq j}\left(\mathbb{E}\left[X_i^2\right]-2\mathbb{E}\left[X_iX_j\right]+\mathbb{E}\left[X_j^2\right]\right.\\
&\quad \left.-\mathbb{E}\left[Y_i^2\right]+2\mathbb{E}\left[Y_iY_j\right]-\mathbb{E}\left[Y_j^2\right]\right)\mathbb{E}\left[p_ip_j\right]\\
&= \tfrac{\beta}{4}\sum_{i\neq j}\left(\mathbb{E}\left[(X_i-X_j)^2\right]-\mathbb{E}\left[(Y_i-Y_j)^2\right]\right)\mathbb{E}\left[p_ip_j\right].
\end{aligned}$$

Nun müssen wir noch zeigen, dass für dieses Ergebnis $\frac{d}{du}\mathbb{E}\left[f(Z(u))\right]\leq 0$ ist. Zunächst ist nach Wahl von $\beta>0$ auch $\frac{\beta}{4}>0$. Weiter ist

$$\mathbb{E}\left[p_ip_j\right]=\mathbb{E}\left[\frac{e^{\beta x_i}e^{\beta x_j}}{\left(\sum_{j=1}^n e^{\beta x_j}\right)^2}\right]>0,$$

weil alle Terme im Erwartungswert positiv sind. Nun gilt nach Voraussetzung, dass

$$\mathbb{E}\left[(X_i-X_j)^2\right]\leq\mathbb{E}\left[(Y_i-Y_j)^2\right],$$

also ist

$$\mathbb{E}\left[(X_i-X_j)^2\right]-\mathbb{E}\left[(Y_i-Y_j)^2\right]\leq 0$$

und somit auch $\frac{d}{du}\mathbb{E}\left[f(Z(u))\right]\leq 0$. □

2.4 Gauß-Konzentration

Dieser Abschnitt orientiert sich im Allgemeinen an [Bou+12]. In diesem Abschnitt möchten wir auf die Gauß-Konzentration und deren Auswirkung auf die Abschätzung der Tails von Gaußprozessen eingehen. Darauf aufbauend werden wir einen Abschätzung des $2q$-ten Moments sub-gaußscher Zufallsvariablen kennen lernen. Bei der Abschätzung der Spektralnorm gaußscher Zufallsmatrizen werden wir auf diese Ergebnisse zurückkommen.

Definition 2.41. Eine Teilmenge T eines metrischer Raum mit Metrik (M,d) heißt *totalbeschränkt* oder auch *präkompakt*, falls es für jedes $\varepsilon>0$ eine endliche Menge $\{t_1,t_2,\ldots,t_n\}$ gibt, so dass

$$T\subseteq\bigcup_{i=1}^n\{t\in M|d(t,t_i)<\varepsilon\}.$$

Satz 2.42. *[Bou+12, Theorem 5.8] Sei $\mathcal{X} = (X_t)_{t\in T}$ ein fast sicher stetiger Gaußprozess mit Erwartungswert 0 und einer totalbeschränkten Menge T. Weiter sei $Z = \sup_{t\in T} X_t$ und $\mathbb{E}[Z] < \infty$. Falls*

$$\sigma^2 = \sup_{t\in T} \mathbb{E}\left[X_t^2\right] < \infty$$

ist, dann gilt für alle $u > 0$

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq u) \leq e^{-u^2/(2\sigma^2)}$$

und

$$\mathbb{P}(\mathbb{E}[Z] - Z \geq u) \leq e^{-u^2/(2\sigma^2)}.$$

Für den Beweis dieses Satzes benötigen wir die Ungleichung über die Gauß-Konzentration für Lipschitz-Funktionen:

Satz 2.43 (Gauß-Konzentration für Lipschitz-Funktionen)**.** *[Bou+12, Theorem 5.6] Sei $X = (X_1, \ldots, X_n)^\top$ ein multivariat standardnormalverteilter Zufallsvektor $X \sim N(0, I_n)$. Weiter sei $f : \mathbb{R}^n \to \mathbb{R}$ eine lipschitzstetige Funktion mit Lipschitzkonstante L, d.h. $\|f(x) - f(y)\| \leq L \cdot \|x - y\|$ für $x, y \in \mathbb{R}^n$. Dann gilt für alle $u > 0$*

$$\mathbb{P}(f(X) - \mathbb{E}[f(X)] \geq u) \leq e^{-u^2/(2L^2)}.$$

Der Gauß-Konzentration für Lipschitz-Funktionen liegen einige Erkenntnisse, wie etwa die Azuma-Hoeffding Ungleichung zugrunde. Deshalb möchten wir hier auf einen Beweis der Ungleichung verzichten, Details siehe [Bou+12, Theorem 5.6].

Beweis von Satz 2.42. Zunächst betrachten wir denn Fall, dass T endlich ist. Zur Einfachheit nehmen wir an, dass $T = \{1, \ldots, n\}$ ist, dann ist $X = (X_1, \ldots, X_n)^\top$ und $X \sim N(0, \Sigma)$ mit Kovarianzmatrix Σ. Wir notieren $\Sigma^{1/2}$ als die Quadrtawurzel von Σ, wie in Bemerkung 2.27 beschrieben. Sein nun $Y = (Y_1, \ldots, Y_n)^\top$ ein standardnormalverteilter Zufallsvektor $Y \sim N(0, I_n)$ im $\mathbb{R}^n$, dann gilt nach Korollar 2.26, dass $\Sigma^{1/2}Y$ und X identisch verteilt sind. Wir betrachten

$$f(Y) := \max_{1\leq i\leq n} (\Sigma^{1/2}Y)_i$$

und erkennen, dass $f(Y)$ identisch verteilt ist zu $\max_{1\leq i\leq n} X_i$. Falls wir zeigen können, dass die Lipschitzkonstante L von f durch σ begrenzt ist können wir die Gauß-Konzentration anwenden. Dazu betrachten wir für alle $u, v \in \mathbb{R}^n$ und $1 \leq i \leq n$

$$\left|(\Sigma^{1/2}u)_i - (\Sigma^{1/2}v)_i\right| = \left|\sum_{j=1}^n \left(\Sigma^{1/2}\right)_{ij} (u_j - v_j)\right| \leq \left(\sum_{j=1}^n \left(\Sigma^{1/2}\right)_{ij}^2\right)^{1/2} \|u - v\|,$$

wobei wir rechts die Cauchy-Schwarzsche Ungleichung benutzt haben. Wir beobachten, dass $\left(\Sigma^{1/2}\right)_{ij}$ für $1 \leq i \leq n$ die i-te Zeile der Matrix $\Sigma^{1/2}$ ist und wegen der Symmetrie von $\Sigma^{1/2}$ auch die i-te Spalte ist. Aus den Regeln der Matrixmultiplikation und der Definition von $\Sigma^{1/2}$ folgt

$$\sum_{j=1}^n \left(\Sigma^{1/2}\right)_{ij}^2 = \Sigma_{ii} = \mathrm{Var}(X_i) \leq \max_{1\leq i\leq n} \mathrm{Var}(X_i) \leq \sigma^2,$$

wobei die letzte Ungleichung nach der an X gestellten Voraussetzung gilt. Das setzen wir ein und erhalten

$$\begin{aligned}|f(u)-f(v)| &= \big|\max_{1\le i\le n}(\Sigma^{1/2}u)_i - \max_{1\le i\le n}(\Sigma^{1/2}v)_i\big| \\ &\le \max_{1\le i\le n}\big|(\Sigma^{1/2}u)_i - (\Sigma^{1/2}v)_i\big| \le \sigma\|u-v\|.\end{aligned}$$

Damit ist f lipschitzstetig mit Lipschitzkonstante $L=\sigma$. Dann ergibt sich aus der Gauß-Konzentration für Lipschitz-Funktionen für $Z=\max_{1\le i\le n} X_i$, dass

$$\mathbb{P}(Z-\mathbb{E}[Z]\ge u) = \mathbb{P}(f(x)-\mathbb{E}[f(X)]\ge u) \le e^{-u^2/(2\sigma^2)}$$

ist. Die Ungleichung

$$\mathbb{P}(\mathbb{E}[Z]-Z\ge u)\le e^{-u^2/(2\sigma^2)}$$

folgt aus Symmetriegründen. Es bleibt noch die Ungleichung für eine beliebige total beschränkte Menge T zu zeigen. Dazu betrachten wir eine in T dicht liegende abzählbare Teilmenge D. Dann gilt wegen der Pfadstetigkeit von $\sup X_t$, dass

$$Z=\sup_{t\in T} X_t = \sup_{t\in D} X_t.$$

Also ergibt sich die Ungleichung aus dem Grenzwert

$$\max_{1\le i\le n} X_i \to \sup_{t\in D} X_t = Z \qquad \text{für } n\to\infty.$$

Dazu bleibt noch zu zeigen, dass $\mathbb{E}[Z]$ auch existiert. Da $\max_{1\le i\le k} X_i \le \max_{1\le i\le l} X_i$ für $k\le l$ gilt mit monotoner Konvergenz

$$\mathbb{E}\left[\max_{1\le i\le n} X_i\right] \to \mathbb{E}\left[\sup_{t\in D} X_t\right] = \mathbb{E}[Z] \qquad \text{für } n\to\infty.$$

Aus der totalen Beschränktheit von T folgt aber, dass auch $\mathbb{E}\left[\max_{1\le i\le n} X_i\right]$ für alle n beschränkt ist und somit die Existenz des Erwartungswerts, welcher nach Voraussetzung endlich ist. □

Damit haben wir ein erstes Resultat aus der Gauß-Konzentration abgeleitet. Wir möchten noch auf einen weiteren Satz eingehen, bei dem direkt auffällt, dass die gestellte Voraussetzung ähnlich dem gerade bewiesenen Resultat ist. Genau mit diesem Zusammenhang werden wir später die beiden Sätze verwenden.

Definition 2.44. Eine Zufallsvariable X deren Tails für eine Konstante $C>0$ für alle $x\ge 0$ die Ungleichung

$$\mathbb{P}(|X|\ge x)\le e^{-x^2/C^2}$$

erfüllen, nennen wir *sub-gaußsche Zufallsvariable.*

Satz 2.45. *[Bou+12, Theorem 2.1] Sei X eine sub-gaußsche Zufallsvariable mit Erwartungswert $\mathbb{E}[X]=0$. Falls für ein $v>0$ gilt*

$$\max(\mathbb{P}(X>x),\mathbb{P}(-X>x))\le e^{-x^2/(2v)}$$

für alle $x>0$, dann gilt für jede ganze Zahl $q\ge 1$, dass

$$\mathbb{E}\left[X^{2q}\right]\le 2q!(2v)^q\le q!(4v)^q.$$

Beweis. Wir beweisen den Satz zuerst für den Fall $v = 1$ und schließen dann auf ein allgemeines $v > 0$. Da $X^{2q} > 0$ ist und x^{2q} monoton wachsend, sowie für $x = 0$ ebenfalls $x^{2q} = 0$ ist, gilt mit dem Tail-Integral des Erwartungswerts

$$\begin{aligned} \mathbb{E}\left[X^{2q}\right] &= \mathbb{E}\left[|X|^{2q}\right] \\ &= \int_0^\infty 2q \cdot x^{2q-1}\mathbb{P}\{|X| > x\}\mathrm{d}x \\ &\leq \int_0^\infty 2q \cdot x^{2q-1} 2e^{-x^2/2}\mathrm{d}x = 4q\int_0^\infty x^{2q-1}e^{-x^2/2}\mathrm{d}x. \end{aligned}$$

Die Ungleichung folgt dabei aus der an X gestellten Voraussetzung. Zur Berechnung des Integrals setzen wir $x = \sqrt{2t}$ und erhalten

$$\mathbb{E}\left[X^{2q}\right] = 4q\int_0^\infty (2t)^{q-1}e^{-t}\mathrm{d}x = 4q2^{q-1}\int_0^\infty t^{q-1}e^{-t}\mathrm{d}t = 4q2^{q-1}\Gamma(q).$$

Nun nutzen wir, dass für eine ganze positive Zahl q gilt $\Gamma(q) = (q-1)!$ und erhalten

$$\mathbb{E}\left[X^{2q}\right] = 4q2^{q-1}\Gamma(q) = 4q2^{q-1}(q-1)! = 2q!2^q \leq q!4^q.$$

Damit ist der Satz für $v = 1$ gezeigt. Falls $v \neq 1$ normieren wir $\frac{X}{\sqrt{v}}$ und erhalten

$$\mathbb{E}\left[X^{2q}\right] = \mathbb{E}\left[\left(\sqrt{v}\frac{X}{\sqrt{v}}\right)^{2q}\right] = v^q\mathbb{E}\left[\left(\frac{X}{\sqrt{v}}\right)^{2q}\right] \leq v^q 2q!2^q \leq 2q!(2v)^q \leq q!(4v)^q.$$

□

Kapitel 3

Abschätzung der Spektralnorm Gaußscher Zufallsmatrizen

In diesem Abschnitt möchten wir uns mit der Abschätzung des Erwartungswerts der Spektralnorm von Gaußschen Zufallsmatrizen befassen. Zentral ist dabei der Satz 3.6, den wir im Folgenden zunächst formulieren und dann beweisen werden. Der zentrale Satz und der Beweis entstammen der Arbeit von Bandeira und van Handel [Ban+16].

Definition 3.1. Sei X eine symmetrische $n \times n$ -Matrix mit Einträgen $X_{ij} = g_{ij} b_{ij}$, wobei die $g_{ij} \overset{u.i.v}{\sim} N(0,1)$ mit $i \leq j$ unabhängig, identisch standardnormalverteilte Zufallsvariablen und die b_{ij} gegebene Skalare mit $i \leq j$ sind.
Dafür definieren wir

$$\sigma := \max_i \sqrt{\sum_j b_{ij}^2} \tag{3.1}$$

und

$$\sigma_* := \max_{ij} |b_{ij}|. \tag{3.2}$$

3.1 Abschätzungen der Spektralnorm

Zunächst möchten wir auf einige bekannte Abschätzungen der Spektralnorm gaußscher Zufallsmatrizen ohne diese zu beweisen eingehen. Später werden wir dann feststellen, dass die Abschätzung nach Bandeira und van Handel, auf die wir im nächsten Abschnitt genauer eingehen werden, im Vergleich zu diesen scharf ist. Als erstes betrachten wir aber das asymptotische Verhalten der Spektralnorm von gaußschen Wigner Matrizen.

Definition 3.2. Wir nennen eine Matrix W ein *gaußsche Wigner Matrix*, falls in der Situation von Definitition 3.1 alle $b_{ij} = 1$ sind.

Dann ist ein klassisches Resultat für gaußsche Wigner Matrizen, dass unter milden Annahmen bezüglich der Momente

$$\frac{\|W\|}{\sqrt{n}} \to 2$$

für $n \to \infty$ gilt, siehe [Bai+88].

Notation 3.3. Wir schreiben für zwei reelle Zahlen(folgen)

$$x \lesssim y,$$

wenn

$$x \leq Cy$$

für eine universelle Konstante $C \in \mathbb{R}$ gilt. Weiter notieren wir

$$x \asymp y,$$

falls $x \lesssim y$ und $x \gtrsim y$.

Als nächstes geben wir eine nichtasymptotische Beschränkung der Spektralnorm von $\|X\|$, wie oben definiert, welche als Konsequenz aus der nichtkommutativen Khintchine Ungleichung von Lust-Piquard und Pisier [Pis03]. In unserer Situation gilt:

Bemerkung 3.4. Sei X wie zuvor definiert, dann gilt

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma\sqrt{\log n}.$$

Diese Abschätzung ist aber schon im simplen Fall einer gaußschen Wigner Matrix nicht mehr scharf, denn dann ist $\sigma = \sqrt{n}$ und somit $\mathbb{E}\left[\|X\|\right] \lesssim \sqrt{n \log n}$, wobei die korrekte Skalierung $\mathbb{E}\left[\|X\|\right] \lesssim \sqrt{n}$ ist. Eine weitere Begrenzung von $\|X\|$ folgt aus einer Methode von Gordon, welche die Slepian Ungleichung für Gaußprozesse erweitert, siehe [Dav+01].

Bemerkung 3.5. Sei X wie zuvor, dann ist

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma_* \sqrt{n}.$$

Hier erscheint der Parameter σ_* natürlich viel kleiner als σ, während die dimensionale Skalierung deutlich schlechter ist. Das fällt inbesonders bei Diagonalmatrizen ins Gewicht.

3.2 Abschätzung der Spektralnorm nach Bandeira und van Handel

Wir betrachten die Abschätzung der Spektralnorm von gaußschen Zufallsmatrizen nach Bandeira und van Handel:

Satz 3.6. *[Ban+16, Theorem 1.1] Seien X wie in Definition 3.1 und σ wie in (3.1), sowie σ_* wie in (3.2) definiert. Dann gilt für $0 < \varepsilon \leq \frac{1}{2}$, dass*

$$\mathbb{E}\left[\|X\|\right] \leq (1+\varepsilon)\left(2\sigma + \frac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_* \sqrt{\log n}\right).$$

Wenn wir die Notation nach 3.3 benutzen können wir den Satz auch wie folgt formulieren:

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma + \sigma_* \sqrt{\log n}$$

Diese Abschätzung ist deutlich besser als die Ergebnisse 3.4 und 3.5, denn diese sind die Grenzfälle der Abschätzung nach Bandeira und van Handel: So gilt nach den Eigenschaften von σ und σ_*, dass

$$1 \leq \frac{\sigma}{\sigma_*} \leq \sqrt{n}.$$

Daraus schließen wir, dass der Fall $\frac{\sigma}{\sigma_*} \lesssim 1$ schon minimal ist. In diesem Fall gilt asymptotisch

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma + \sigma \sqrt{\log n} \asymp \sigma \sqrt{\log n}.$$

Die Abschätzung der Spektralnorm nach Bandeira und van Handel entspricht dann also jener der nichtkommutativen Khintchine Ungleichung, Bemerkung 3.4. Im nicht minimalen Fall ist die Abschätzung nach Bandeira und van Handel jedoch besser. Betrachten wir den maximalen Fall $\frac{\sigma}{\sigma_*} \gtrsim \sqrt{n}$, so entspricht die Abschätzung nach Bandeira und van Handel jener die nach Gordon, Bemerkung 3.5, folgt und ist im nicht maximalen Fall besser:

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma_* \sqrt{n} + \sigma_* \sqrt{\log n}$$

Somit spiegelt die Abschätzung in diesem Sinne eine Interpolation zwischen Bemerkung 3.4 und Bemerkung 3.5 wider.

Bemerkung 3.7. Wir werden in Kapitel 4 noch sehen, dass

$$\mathbb{E}\left[\|X\|\right] \gtrsim \sigma + \sigma_* \sqrt{\log n}$$

ist, sofern wir bestimmte Bedingungen an die Koeffizienten b_{ij} stellen. In diesem Fall ist also

$$\mathbb{E}\left[\|X\|\right] \asymp \sigma + \sigma_* \sqrt{\log n}$$

und somit ist die Abschätzung scharf bis auf Konstanten.

Aus der Bemerkung und dem Vergleich zu den Abschätzungen nach Khintchine und Gordon schließen wir, dass die Abschätzung nach Bandeira und van Handel das optimale Ergebnis ihrer Art ist. Allerdings müssen wir feststellen, dass Satz 3.6 nicht unbedingt scharf in Bezug auf die Konstanten ist, hier sind bessere Ergebnisse denkbar.

Bemerkung 3.8. Die Abschätzung der Spektralnorm nach Satz 3.6 lässt sich auch auf nichtsymmetrische rechteckige Matrizen verallgemeinern. Des Weiteren sind auch Abschätzungen ohne Einbezug der Dimension n möglich. Dies werden wir noch in Kapitel 4 betrachten.

Wir möchten noch betrachten, wie sich Satz 3.6 im Falle einer gaußschen Wigner Matrix verhält.

Beispiel 3.9. Sei W eine gaußsche Wigner Matrix, dann gilt in der Situation von Satz 3.6, dass $\sigma_* = 1$ und $\sigma = \sqrt{n}$. Es folgt

$$\mathbb{E}\left[\|W\|\right] \leq (1+\varepsilon)2\sqrt{n} + o(\sqrt{n})$$

für hinreichend kleines ε. Somit erhalten wir aus Satz 3.6 nicht nur eine scharfe Abschätzung, sondern dieser gibt sogar das korrekte asymptotische Verhalten $\|W\|/\sqrt{n} \to 2$ wider.

Kommen wir zum Beweis des Satzes 3.6, dazu benötigen wir noch die folgenden drei Hilfsaussagen: Diese Aussagen benutzen eine r-dimensionale gaußsche Wigner Matrix Y_r, also eine symmetrische $r \times r$-Matrix mit unabhängigen $N(0,1)$ verteilten Zufallsvariablen als Einträgen $Y_{ij} = g_{ij}$, wobei $g_{ij} \sim N(0,1)$ für alle $1 \leq i \leq j \leq r$.

Lemma 3.10. *[Ban+16, Lemma 2.2] Sei Y_r eine r-dimensionale gaußsche Wigner Matrix, dann gilt für jedes $p \geq 2$, dass*

$$\mathbb{E}\left[\|Y_r\|^{2p}\right]^{\frac{1}{2p}} \leq 2\sqrt{r} + 2\sqrt{2p}.$$

Proposition 3.11. *[Ban+16, Proposition 2.1] Sei Y_r eine r-dimensionale gaußsche Wigner Matrix und X wie in Satz 3.6, dann gilt für $\sigma_* \leq 1$ und $p \in \mathbb{N}$*

$$\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] \leq \frac{n}{\lceil \sigma^2 \rceil + p} \mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil \sigma^2 \rceil + p}\right)\right].$$

Dabei ist für $x \in \mathbb{R}$ die obere Gaußklammer $\lceil x \rceil$ definiert als

$$\lceil x \rceil := \min\{k \in \mathbb{Z} | k \geq x\}.$$

Lemma 3.12. *Sei X eine symmetrische $n \times n$-Matrix mit reellen Einträgen, dann gilt für $p \geq 1$*

$$\|X\| \leq \left(\operatorname{Tr}\left(X^{2p}\right)\right)^{\frac{1}{2p}} \leq n^{\frac{1}{2p}} \|X\|$$

bzw.

$$\|X\|^{2p} \leq \operatorname{Tr}\left(X^{2p}\right) \leq n\|X\|^{2p}.$$

Der Beweis des Satzes 3.6 ist nun mit dem Lemma 3.10 und der Proposition 3.11 recht direkt.

Beweis von Satz 3.6. Wir betrachten zunächst den Fall, dass $\sigma_* = 1$ ist und schließen dann auf ein allgemeines σ_*:
Wir wenden Lemma 3.12 an und erhalten für $p \geq 2$

$$\mathbb{E}\left[\|X\|\right] \leq \mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)^{\frac{1}{2p}}\right] \leq \mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right]^{\frac{1}{2p}}.$$

Die zweite Ungleichung ergibt sich aus der Jensenschen Ungleichung, genauer Korollar 2.18, denn für $p \geq 1$ ist $x \mapsto x^{\frac{1}{2p}}$ konkav. Da $\sigma_* = 1$ können wir nun Proposition 3.11 benutzen:

$$\begin{aligned} \mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right]^{\frac{1}{2p}} &\leq \left(\tfrac{n}{\lceil \sigma^2 \rceil + p} \mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil \sigma^2 \rceil + p}\right)\right]\right)^{\frac{1}{2p}} \\ &= \tfrac{n^{1/(2p)}}{(\lceil \sigma^2 \rceil + p)^{1/(2p)}} \mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil \sigma^2 \rceil + p}\right)\right]^{\frac{1}{2p}} \end{aligned} \tag{3.3}$$

Wir wenden nun nochmal Lemma 3.12, sowie die Linearität des Erwartungswerts an:

$$\begin{aligned} \tfrac{n^{1/(2p)}}{(\lceil \sigma^2 \rceil + p)^{1/(2p)}} \mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil \sigma^2 \rceil + p}\right)\right]^{\frac{1}{2p}} &\leq \tfrac{n^{1/(2p)}}{(\lceil \sigma^2 \rceil + p)^{1/(2p)}} \mathbb{E}\left[(\lceil \sigma^2 \rceil + p)\|Y_{\lceil \sigma^2 \rceil + p}\|^{2p}\right]^{\frac{1}{2p}} \\ &= n^{1/(2p)} \mathbb{E}\left[\|Y_{\lceil \sigma^2 \rceil + p}\|^{2p}\right]^{\frac{1}{2p}} \end{aligned}$$

Weiter können wir nun Lemma 3.10 für $r = \lceil \sigma^2 \rceil + p$ anwenden (für $p \geq 2$):

$$n^{1/(2p)} \mathbb{E}\left[\|Y_{\lceil \sigma^2 \rceil + p}\|^{2p}\right]^{\frac{1}{2p}} \leq n^{1/(2p)} \left(2\sqrt{\lceil \sigma^2 \rceil + p} + 2\sqrt{2p}\right).$$

Als nächstes wählen wir $p = \lceil a \log n \rceil$, mit $\lceil x \rceil \leq x + 1$, $x \in \mathbb{R}$, folgt

$$\begin{aligned}
\mathbb{E}\left[\|X\|\right] &\leq n^{1/(2p)} \left(2\sqrt{\lceil \sigma^2 \rceil + p} + 2\sqrt{2p}\right) \\
&= n^{1/(2\lceil a \log n \rceil)} \left(2\left(\lceil \sigma^2 \rceil + \lceil a \log n \rceil\right)^{\frac{1}{2}} + 2\left(2\lceil a \log n \rceil\right)^{\frac{1}{2}}\right) \\
&\leq e^{1/(2a)} \left(2\left(\sigma^2 + a \log n + 2\right)^{\frac{1}{2}} + 2\left(2a \log n + 2\right)^{\frac{1}{2}}\right) \\
&\leq e^{1/(2a)} \left(2\sigma + 2\left(a \log n + 2\right)^{\frac{1}{2}} + 2\left(2a \log n + 2\right)^{\frac{1}{2}}\right).
\end{aligned}$$

Dabei gilt mit der Monotonie der Exponentialfunktion

$$n^{1/(2\lceil a \log n \rceil)} = \exp(\log n)^{1/(2\lceil a \log n \rceil)} \leq e^{\log n/(2a \log n)} = e^{1/(2a)}.$$

Wir setzen $e^{1/(2a)} = 1 + \varepsilon$. Nach Voraussetzung ist $\varepsilon \leq \frac{1}{2}$ also muss $a \geq 1$ sein.
Fall 1: $n \geq 2$ und $p \geq 2$. Wir sehen $2 \leq 3 \log 2 \leq 3a \log n$ und rechnen

$$\begin{aligned}
\mathbb{E}\left[\|X\|\right] &\leq e^{1/(2a)} \left(2\sigma + 2\left(a \log n + 2\right)^{\frac{1}{2}} + 2\left(2a \log n + 2\right)^{\frac{1}{2}}\right) \\
&\leq e^{1/(2a)} \left(2\sigma + 2\left(a \log n + 3a \log n\right)^{\frac{1}{2}} + 2\left(2a \log n + 3a \log n\right)^{\frac{1}{2}}\right) \\
&= e^{1/(2a)} \left(2\sigma + 2\left(4a \log n\right)^{\frac{1}{2}} + 2\left(5a \log n\right)^{\frac{1}{2}}\right) \\
&= e^{1/(2a)} \left(2\sigma + 2\sqrt{2}\left(2a \log n\right)^{\frac{1}{2}} + 2\sqrt{\tfrac{5}{2}}\left(2a \log n\right)^{\frac{1}{2}}\right) \\
&= e^{1/(2a)} \left(2\sigma + 2(\sqrt{2} + \sqrt{\tfrac{5}{2}})\left(2a \log n\right)^{\frac{1}{2}}\right) \\
&\leq e^{1/(2a)} \left(2\sigma + 2 \cdot 3\left(2a \log n\right)^{\frac{1}{2}}\right) \\
&= e^{1/(2a)} \left(2\sigma + 6\left(2a \log n\right)^{\frac{1}{2}}\right) \\
&= (1 + \varepsilon) \left(2\sigma + 6\left(\frac{1}{\log(1+\varepsilon)} \log n\right)^{\frac{1}{2}}\right) \\
&= (1 + \varepsilon) \left(2\sigma + \frac{6}{\sqrt{\log(1+\varepsilon)}} \sigma_* \sqrt{\log n}\right).
\end{aligned}$$

Dabei benutzen wir, dass $\sigma_* = 1$ ist, sowie die Äquivalenz

$$e^{1/(2a)} = 1 + \varepsilon \quad \Leftrightarrow \quad 2a = \tfrac{1}{\log(1+\varepsilon)}.$$

Fall 2: n=1. Dann ist $X = (bg)$, wobei $g \sim N(0,1)$ und $|b| = \sigma_* = \sigma$. Daraus ergibt sich

$$\begin{aligned}
&\mathbb{E}\left[\|X\|\right] = \mathbb{E}\left[|b||g|\right] = \sigma \mathbb{E}\left[|g|\right] \leq \sigma(1 + \mathbb{E}\left[g^2\right]) = 2\sigma \\
&\leq (1 + \varepsilon) \left(2\sigma + \frac{6}{\sqrt{\log(1+\varepsilon)}} \sigma_* \sqrt{\log n}\right).
\end{aligned}$$

Dabei ist $\log 1 = 0$.
Fall 3: p=1. Mit der Wahl $p = \lceil a \log n \rceil$ folgt, da $a \geq 1$, schon dass $n \leq 2$. Den Fall $n = 1$ haben wir bereits abgehandelt. Somit ist $n = 2$ und $\lceil \sigma^2 \rceil = 1$ oder

$\lceil\sigma^2\rceil = 2$. Das können wir nun in (3.3) einsetzen. Zunächst für $\lceil\sigma^2\rceil = 1$:

$$\begin{aligned}
\mathbb{E}\left[\|X\|\right] &\leq \left(\tfrac{n}{\lceil\sigma^2\rceil+p}\mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil\sigma^2\rceil+p}\right)\right]\right)^{\frac{1}{2p}} = \sqrt{\mathbb{E}\left[\operatorname{Tr}\left(Y_2^2\right)\right]} \\
&= \sqrt{\mathbb{E}\left[g_{11}^2+g_{12}^2+g_{21}^2+g_{22}^2\right]} \\
&= \sqrt{\mathbb{E}\left[g_{11}^2\right]+\mathbb{E}\left[g_{12}^2\right]+\mathbb{E}\left[g_{21}^2\right]+\mathbb{E}\left[g_{22}^2\right]} \\
&= \sqrt{4} = 2 = 2\sigma \leq (1+\varepsilon)\left(2\sigma + \tfrac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right).
\end{aligned}$$

Denn mit $\sigma_* = 1$ gilt schon $\sigma \geq 1$. Ist nun $\lceil\sigma^2\rceil = 2$ gilt

$$\begin{aligned}
\mathbb{E}\left[\|X\|\right] &\leq \left(\tfrac{n}{\lceil\sigma^2\rceil+p}\mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil\sigma^2\rceil+p}\right)\right]\right)^{\frac{1}{2p}} = \left(\tfrac{2}{3}\mathbb{E}\left[\operatorname{Tr}\left(Y_3^2\right)\right]\right)^{\frac{1}{2}} \\
&= \left(\tfrac{2}{3}\mathbb{E}\left[g_{11}^2+g_{12}^2+g_{13}^2+g_{21}^2+g_{22}^2+g_{23}^2+g_{31}^2+g_{32}^2+g_{33}^2\right]\right)^{\frac{1}{2}} \\
&= \sqrt{\tfrac{2}{3}\cdot 9} = \sqrt{6} \leq 2\sigma + \tfrac{6}{\sqrt{\log(\frac{3}{2})}}\sigma_*\sqrt{\log 2} \\
&\leq (1+\varepsilon)\left(2\sigma + \tfrac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right).
\end{aligned}$$

Damit ist der Satz für $\sigma_* = 1$ bewiesen.
Für ein allgemeines $\sigma_* > 0$ gilt, falls $\sigma_* \neq 1$ normieren wir X mit $\frac{1}{\sigma_*}$. Dann gilt für $X = X' \cdot \sigma_*$, dass $\sigma'_* = 1$ von X' ist. Wir rechnen

$$\begin{aligned}
\mathbb{E}\left[\|X\|\right] &= \mathbb{E}\left[\|X'\sigma_*\|\right] \\
&= \sigma_*\mathbb{E}\left[\|X'\|\right] \\
&\leq \sigma_*(1+\varepsilon)\left(2\sigma' + \tfrac{6}{\sqrt{\log(1+\varepsilon)}}1\sqrt{\log n}\right) \\
&= (1+\varepsilon)\left(2\sigma'\sigma_* + \tfrac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right) \\
&= (1+\varepsilon)\left(2\sigma + \tfrac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right).
\end{aligned}$$

Dabei ist $\sigma' = \max_i \sqrt{\left(\frac{b_{ij}}{\sigma_*}\right)^2}$ und somit

$$\sigma_* \cdot \sigma' = \max_i \sqrt{\sigma_*^2\left(\frac{b_{ij}}{\sigma_*}\right)^2} = \sigma.$$

Somit folgt die Behauptung auch für $\sigma_* > 0$. Für $\sigma_* = 0$ gilt

$$\mathbb{E}\left[\|X\|\right] = 0 = (1+\varepsilon)\left(2\sigma + \frac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right).$$

Aufgrund der Definition von σ_* über den Betrag ist $\sigma_* < 0$ nicht möglich. □

3.3 Beweis der Hilfsaussagen zur Abschätzung der Spektralnorm

Nun möchten wir die Lemmata 3.10 und 3.12, sowie die Proposition 3.11 beweisen, welche dem Beweis von Satz 3.6 vorangegangen sind. Wir beginnen mit

Lemma 3.10, in dessen Beweis eine ganze Reihe der zu Anfang besprochenen Grundlagen einfließt:

Um die Aussage des Lemma 3.10 zu beweisen gehen wir in drei Schritten vor: Um die Ungleichung

$$\mathbb{E}\left[\|Y_r\|^{2p}\right]^{\frac{1}{2p}} \leq 2\sqrt{r} + 2\sqrt{2p}$$

zu zeigen überlegen wir uns zunächst im ersten Schritt, dass auf der linken Seite die L^{2p}-Norm von der Spektralnorm von Y_r steht. Diese entspricht jedoch dem betragsmäßig größten Eigenwert von Y_r, welcher der maximale Eigenwert λ_+ oder das negative des minimalen Eigenwertes λ_- ist. Somit können wir diesen Ausdruck mit der Dreiecksungleichung in zwei Teile aufspalten. Zur Abschätzung fassen wir dann im zweiten Schritt den einen Teil $\mathbb{E}[\lambda_+]$ als das Supremum eines Gaußprozess auf und schätzen diesen mit der Sudakov-Fernique Ungleichung gegen $2\sqrt{r}$ ab. Den anderen Teil schätzen wir dann im dritten Schritt mit Hilfe der Gauß-Konzentration ab.

Beweis von Lemma 3.10. Schritt 1: Zunächst notieren wir

$$\lambda_+ := \sup_{v\in S^{r-1}} \langle v, Y_r v\rangle \quad , \quad \lambda_- := -\inf_{v\in S^{r-1}} \langle v, Y_r v\rangle$$

Dabei ist S^{r-1} die Einheitsphäre des $\mathbb{R}^r$. Dann können wir nach Lemma 2.14 schreiben

$$\|Y_r\| = \max(\lambda_+, \lambda_-).$$

Da die $Y_{ij} = g_{ij}$ als standardnormalverteilte Zufallsvariablen symmetrisch verteilt sind, haben dann Y_r und $-Y_r$ die gleiche Verteilung. Somit gilt

$$\lambda_+ = \sup_{v\in S}\langle v, Y_r v\rangle = -\inf_{v\in S}\langle v, -Y_r v\rangle \overset{D}{=} -\inf_{v\in S}\langle v, Y_r v\rangle = \lambda_-.$$

Also sind λ_+ und λ_- identisch verteilt und somit ist auch $\mathbb{E}[\lambda_+] = \mathbb{E}[\lambda_-]$. Dann können wir mit der L^{2p}-Norm, vgl. Bemerkung 2.19, in (a) $(p \geq 2)$ und der Dreiecksungleichung für die L^{2p}-Norm in (b) schreiben:

$$\begin{aligned}
\mathbb{E}\left[\|Y_r\|^{2p}\right]^{\frac{1}{2p}} &= \mathbb{E}\left[\max(\lambda_+,\lambda_-)^{2p}\right]^{\frac{1}{2}p} \\
&\overset{(a)}{=} \|\max(\lambda_+,\lambda_-)\|_{2p} \\
&= \|\mathbb{E}[\lambda_+] - \mathbb{E}[\lambda_-] + \max(\lambda_+,\lambda_-)\|_{2p} \\
&\overset{(b)}{\leq} \|\mathbb{E}[\lambda_+]\| + \|\max(\lambda_+,\lambda_-) - \mathbb{E}[\lambda_-]\|_{2p} \\
&\leq \mathbb{E}[\lambda_+] + \|\max(\lambda_+,\lambda_-) - \mathbb{E}[\lambda_-]\|_{2p} \\
&= \mathbb{E}[\lambda_+] + \|\max(\lambda_+ - \mathbb{E}[\lambda_+], \lambda_- - \mathbb{E}[\lambda_-])\|_{2p}
\end{aligned}$$

Schritt 2: Wir befassen uns mit $\mathbb{E}[\lambda_+]$:
Wir wissen, dass

$$\mathbb{E}[\lambda_+] = \mathbb{E}\left[\sup_{v\in S^{r-1}} \langle v, Y_r v\rangle\right]$$

ist. Das möchten wir nutzen um $\mathbb{E}[\lambda_+]$ mit der Sudakov-Fernique Ungleichung, Satz 2.36, abzuschätzen. Dazu fassen wir $\{\langle v, Y_r v\rangle\}_{v\in S^{r-1}}$ als Gaußprozess auf und schätzen dessen Supremum gegen jenes des Gaußprozess $\{2\langle v, g\rangle\}_{v\in S^{r-1}}$

ab, wobei g ein multivariat standardnormalverteilter Vektor im $\mathbb{R}^r$ ist. Dieser Gaußprozess entspricht dem kanonischen Gaußprozess, Beispiel 2.35, mit Vorfaktor 2. Um die Sudakov-Fernique Ungleichung anwenden zu können müssen wir die Voraussetzungen überprüfen und zeigen zunächst für $v, w \in S^{r-1}$, dass

$$\mathbb{E}\left[(\langle v, Y_r v\rangle - \langle w, Y_r w\rangle)^2\right] \leq \mathbb{E}\left[(2\langle v, g\rangle - 2\langle w, g\rangle)^2\right]$$

gilt. Wir schreiben $v := (v_1, \ldots, v_r)^\top$, $w := (w_1, \ldots, w_r)^\top$ und rechnen

$$\begin{aligned} & \mathbb{E}\left[(\langle v, Y_r v\rangle - \langle w, Y_r w\rangle)^2\right] \\ = \quad & \mathbb{E}\left[\left(\left(\sum_{ij} v_i Y_{ij} v_j\right) - \left(\sum_{ij} w_i Y_{ij} w_j\right)\right)^2\right] \\ \overset{Y_{ij} \text{ ausklammern}}{=} \quad & \mathbb{E}\left[\left(\sum_{ij} Y_{ij}(v_i v_j - w_i w_j)\right)^2\right] \end{aligned}$$

Durch das Ausmultiplizieren dieses letzten Ausdrucks erhalten wir mit der Linearität des Erwartungswerts und da Y_r symmetrisch ist

$$\begin{aligned} & \mathbb{E}\Bigg[\sum_{ij} Y_{ij}^2 (v_i v_j - w_i w_j)^2 + \sum_{i \neq j} Y_{ij} Y_{ji} (v_i v_j - w_i w_j)(v_j v_i - w_j w_i) \\ + \quad & \sum_{i \neq k \text{ oder } j \neq l} Y_{ij} Y_{kl} (v_i v_j - w_i w_j)(v_k v_l - w_k w_l)\Bigg] \\ = \quad & \mathbb{E}\left[\sum_{ij} Y_{ij}^2 (v_i v_j - w_i w_j)^2\right] + \mathbb{E}\left[\sum_{i \neq j} Y_{ij}^2 (v_i v_j - w_i w_j)^2\right] \\ + \quad & \mathbb{E}\left[\sum_{i \neq k \text{ oder } j \neq l} Y_{ij} Y_{kl} (v_i v_j - w_i w_j)(v_k v_l - w_k w_l)\right] \\ = \quad & \sum_{ij} (v_i v_j - w_i w_j)^2 \mathbb{E}\left[Y_{ij}^2\right] + \sum_{i \neq j} (v_i v_j - w_i w_j)^2 \mathbb{E}\left[Y_{ij}^2\right] \\ + \quad & \sum_{i \neq k \text{ oder } j \neq l} (v_i v_j - w_i w_j)(v_k v_l - w_k w_l) \mathbb{E}\left[Y_{ij} Y_{kl}\right] \\ = \quad & \sum_{ij} (v_i v_j - w_i w_j)^2 + \sum_{i \neq j} (v_i v_j - w_i w_j)^2 \leq 2 \sum_{ij} (v_i v_j - w_i w_j)^2 \end{aligned}$$

Dabei ist $\mathbb{E}\left[Y_{ij}^2\right] = 1$, da $Y_{ij} \sim N(0,1)$, und $\mathbb{E}\left[Y_{ij} Y_{kl}\right] = \mathbb{E}\left[Y_{ij}\right] \cdot \mathbb{E}\left[Y_{kl}\right] = 0$, weil die Y_{ij} unabhängig sind.

Wir betrachten unter Ausnutzung der Eigenschaften der Norm von v und w

$$\begin{aligned} & \sum_{ij} (v_i v_j - w_i w_j)^2 = \sum_{ij} v_i^2 v_j^2 - 2 v_i v_j w_i w_j + w_i^2 w_j^2 \\ & = \sum_i \left[v_i^2 \left(\sum_j v_j^2\right) - 2 v_i w_i \left(\sum_j v_j w_j\right) + w_i^2 \left(\sum_j w_j^2\right)\right] \\ & = \sum_i \left(v_i^2 - 2 v_i w_i \langle v, w\rangle + w_i^2\right) \\ & = \|v\| + \|w\| - 2\langle v, w\rangle\langle v, w\rangle = 2 - 2\langle v, w\rangle^2. \end{aligned}$$

und mit dem Kosinussatz, Satz 2.6, gilt

$$2\|v - w\|^2 = 2\left(\|v\|^2 + \|w\|^2 - 2\langle v, w\rangle\right) = 2(2 - 2\langle v, w\rangle) = 4 - 4\langle v, w\rangle$$

Nun erhalten wir die Äquivalenzen

$$\begin{aligned} & 2\|v - w\|^2 & \geq \quad & \sum_{ij} (v_i v_j - w_i w_j)^2 \\ \Leftrightarrow \quad & 4 - 4\langle v, w\rangle & \geq \quad & \|v\| + \|w\| - 2\langle v, w\rangle^2 \\ \Leftrightarrow \quad & 2\langle v, w\rangle^2 - 4\langle v, w\rangle + 2 & \geq \quad & 0 \end{aligned}$$

Die letzte Ungleichung ist jedoch schon erfüllt, da das Polynom $2x^2 - 4x + x = 2(x-1)^2$ nur eine (doppelte) Nullstelle bei $x = 1$ besitzt und sonst größer 0 ist, also gilt

$$2\sum_{ij}(v_i v_j - w_i w_j)^2 \leq 4\|v - w\|.$$

Weiter gilt nach Proposition 2.28, dass

$$\mathbb{E}\left[(2\langle v, g\rangle - 2\langle w, g\rangle)^2\right] = 4\mathbb{E}\left[(\langle v, g\rangle - \langle w, g\rangle)^2\right] = 4\|v - w\|^2.$$

Zusammen erhalten wir

$$\begin{aligned}&\mathbb{E}\left[(\langle v, Y_r v\rangle - \langle w, Y_r w\rangle)^2\right] \leq 2\sum_{ij}(v_i v_j - w_i w_j)^2 \leq 4\|v - w\|^2\\&= \mathbb{E}\left[(2\langle v, g\rangle - 2\langle w, g\rangle)^2\right]\end{aligned} \tag{3.4}$$

Damit ist die oben genannte Voraussetzung geprüft. Es bleibt zu untersuchen, ob beide Prozesse Erwartungswert 0 haben. Es ist

$$\mathbb{E}\left[\langle v, Y_r v\rangle\right] = \mathbb{E}\left[\sum_{ij} v_i Y_{ij} v_j\right] = \sum_{ij} v_i v_j \mathbb{E}\left[Y_{ij}\right] = 0$$

denn $\mathbb{E}\left[Y_{ij}\right] = 0$. Mit $\mathbb{E}\left[g_i\right] = 0$ gilt weiter

$$\mathbb{E}\left[2\langle v, g\rangle\right] = 2\sum_{i=1}^{r} v_i \mathbb{E}\left[g_i\right] = 0$$

Also haben beide Prozesse Erwartungswert 0. Damit sind die Voraussetzungen für die Sudakov-Fernique Ungleichung, Satz 2.36, erfüllt und wir können diese auf die endlichdimensionalen Randverteilungen der Gaußprozesse anwenden. Es gilt also für $k \in \mathbb{N}$ und $v_1, \ldots, v_k \in S^{r-1}$

$$\mathbb{E}\left[\max_{1\leq i\leq k}\langle v_i, Y_r v_i\rangle\right] \leq \mathbb{E}\left[\max_{1\leq i\leq k} 2\langle v_i, g\rangle\right].$$

Da ein Gaußprozess über seine endlichdimensionalen Randverteilungen eindeutig bestimmt ist und hier beide Gaußprozesse sowohl stetige Pfade, als auch eine total beschränkte Indexmenge S^{r-1} besitzen, können wir zum Supremum übergehen (wie im Beweis von Satz 2.42) und erhalten

$$\mathbb{E}\left[\sup_{v\in S^{r-1}}\langle v, Y_r v\rangle\right] \leq \mathbb{E}\left[\sup_{v\in S^{r-1}} 2\langle v, g\rangle\right].$$

Weiter ist nach der Cauchy-Schwarz Ungleichung $\langle v, g\rangle \leq \|v\|\|g\|$, weshalb gilt

$$\sup_{v\in S^{r-1}} 2\langle v, g\rangle \leq \sup_{v\in S^{r-1}} 2\|v\|\|g\| = 2\|g\|.$$

Zusammen erhalten wir

$$\mathbb{E}\left[\lambda_+\right] = \mathbb{E}\left[\sup_{v\in S^{r-1}}\langle v, Y_r v\rangle\right] \leq \mathbb{E}\left[\sup_{v\in S^{r-1}} 2\langle v, g\rangle\right] = \mathbb{E}\left[2\|g\|\right] \leq 2\sqrt{r}$$

Dabei gilt die letzte Ungleichung wegen der Jensenschen Ungleichung und der Linearität des Erwartungswerts:

$$\begin{aligned} & \mathbb{E}\left[2\|g\|\right] = 2\mathbb{E}\left[\sqrt{g_1^2+\ldots+g_r^2}\right] \overset{\text{Jensen}}{\leq} 2\sqrt{\mathbb{E}\left[g_1^2+\ldots+g_r^2\right]} \\ = \; & 2\sqrt{\mathbb{E}\left[g_1^2\right]+\ldots+\mathbb{E}\left[g_r^2\right]} = 2\underbrace{\sqrt{1+\ldots+1}}_{r-\text{mal}} = 2\sqrt{r} \end{aligned}$$

Die Jensensche Ungleichung dürfen wir hier so anwenden, genauer Korollar 2.18, weil $x \mapsto \sqrt{x}$ konkav ist.
Schritt 3: Es bleibt noch $\|\max(\lambda_+ - \mathbb{E}\left[\lambda_+\right], \lambda_- - \mathbb{E}\left[\lambda_-\right])\|_{2p}$ abzuschätzen. Hierfür verwenden wir die Gauß-Konzentration. Dafür benutzen wir zunächst Satz 2.42. Wir betrachten wieder den Gaußprozess $\{\langle v, Y_r v\rangle\}_{v\in S^{r-1}}$ und berechnen für $w=0$ aus der Ungleichung (3.4)

$$\mathbb{E}\left[\langle v, Y_r v\rangle^2\right] \leq 4\|v\| = 4,$$

also ist

$$\sigma^2 = \sup_{v\in S^{r-1}} \mathbb{E}\left[\langle v, Y_r v\rangle^2\right] \leq 4.$$

Zudem ist $\mathbb{E}\left[\langle v, Y_r v\rangle^2\right] = 0$, $\{\langle v, Y_r v\rangle\}_{v\in S^{r-1}}$ fast sicher stetig, S^{r-1} total beschränkt und $\mathbb{E}\left[\lambda_+\right] < \infty$ nach Schritt 2. Damit folgt dann aus Satz 2.42 für $x > 0$, dass

$$\mathbb{P}(\lambda_+ - \mathbb{E}\left[\lambda_+\right] \geq x) \leq e^{-x^2/(2\cdot 4)}$$

und

$$\mathbb{P}(\mathbb{E}\left[\lambda_+\right] - \lambda_+ \geq x) \leq e^{-x^2/(2\cdot 4)}.$$

Zusammen erhalten wir

$$\max(\mathbb{P}(\lambda_+ - \mathbb{E}\left[\lambda_+\right] > x), \mathbb{P}(-(\lambda_+ - \mathbb{E}\left[\lambda_+\right]) > x)) \leq e^{-x^2/(2\cdot 4)}.$$

Weiter gilt mit $\lambda_+ - \mathbb{E}\left[\lambda_+\right] \overset{D}{=} \lambda_- - \mathbb{E}\left[\lambda_-\right]$ die analoge Aussage für $\lambda_- - \mathbb{E}\left[\lambda_-\right]$. Dann folgt mit der Tatsache, dass $\mathbb{E}\left[\lambda_+ - \mathbb{E}\left[\lambda_+\right]\right] = 0$, sowie $\mathbb{E}\left[\lambda_- - \mathbb{E}\left[\lambda_-\right]\right] = 0$ aus Satz 2.45 direkt schon ($p \geq 2$)

$$\mathbb{E}\left[\max((\lambda_+ - \mathbb{E}\left[\lambda_+\right])^{2p}, (\lambda_- - \mathbb{E}\left[\lambda_-\right])^{2p})\right] \leq p!(4\cdot 4)^p \leq \left(2\sqrt{2p}\right)^{2p}.$$

Dabei gilt die letzte Ungleichung, denn mit $p \geq 2$ und somit $p! \leq p^p$ folgt

$$p!(4\cdot 4)^p \leq p^p 8^p = (4\cdot 2p)^p = \left(2\sqrt{2p}\right)^{2p}.$$

Es folgt

$$\begin{aligned} & \|\max(\lambda_+ - \mathbb{E}\left[\lambda_+\right], \lambda_- - \mathbb{E}\left[\lambda_-\right])\|_{2p} \\ & = \left(\mathbb{E}\left[\max((\lambda_+ - \mathbb{E}\left[\lambda_+\right])^{2p}, (\lambda_- - \mathbb{E}\left[\lambda_-\right])^{2p})\right]\right)^{\frac{1}{2p}} \leq 2\sqrt{2p}. \end{aligned}$$

Insgesamt folgt somit die Behauptung:

$$\begin{aligned} \mathbb{E}\left[\|Y_r\|^{2p}\right]^{\frac{1}{2p}} \; & = \; \mathbb{E}\left[\lambda_+\right] + \|\max(\lambda_+ - \mathbb{E}\left[\lambda_-+\right], \lambda_- - \mathbb{E}\left[\lambda_-\right])\|_{2p} \\ & \leq \; 2\sqrt{r} + 2\sqrt{2p}. \end{aligned}$$

□

Damit haben wir die erste wichtige Hilfsaussage zum Beweis des Satzes 3.6 bewiesen. Nun folgt noch eine zweite Aussage mit der wir die Matrix X gegen die Matrix Y, wie oben beschrieben, abschätzen können:

Im Folgenden wird der Beweis von Proposition 3.11 gegeben. Dazu schreiben wir in einem ersten Schritt zunächst $\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right]$ als ein Summe aus und schätzen diese dann im zweiten Schritt dadurch ab, indem wir sie über bestimmte Folgen von Kanten eines Graphen summieren. Diese Betrachtung über die Kanten eines Graphen ist der entscheidende Schritt, der Rest geht dann recht zügig. Analog machen wir dies im dritten Schritt mit Y_r. Zuletzt führen wir im vierten Schritt dann beide Ergebnisse zusammen.

Schritt 1: Der Beweis beginnt mit folgendem Lemma:

Lemma 3.13. *Sei X wie in Satz 3.6 definiert, dann gilt für $p \in \mathbb{N}$, dass*

$$\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] = \sum_{u_1,\ldots,u_{2p}\in\{1,\ldots,n\}} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1}\mathbb{E}\left[g_{u_1u_2}g_{u_2u_3}\cdots g_{u_{2p}u_1}\right].$$

Beweis von Lemma 3.13. Wir betrachten zunächst die $n\times n$-Matrix C mit den Einträgen c_{ij}, wobei $1\leq i,j\leq n$. Dann schreiben wir für $t\geq 2$

$$C^t := (c_{ij}^{(t)})_{i,j\in\{1,\ldots,n\}}.$$

Dann sind die $c_{ij}^{(t)}$ die Einträge der Matrix C^t. Diese können wir dann nach den Regeln der Matrixmultiplikation jeweils schreiben als:

$$\begin{aligned} c_{ij}^{(t)} &= \sum_{u_2=1}^{n} c_{iu_2}c_{u_2j}^{(t-1)} = \sum_{u_2=1}^{n} c_{iu_2}\sum_{u_3=1}^{n} c_{u_2u_3}c_{u_3j}^{(t-2)} \\ &= \sum_{u_2=1}^{n} c_{iu_2}\sum_{u_3=1}^{n} c_{u_2u_3}\cdots\sum_{u_t=1}^{n} c_{u_{t-1}u_t}c_{u_tj} \\ &= \sum_{u_2,\ldots,u_t\in\{1,\ldots,n\}} c_{iu_2}c_{u_2u_3}\cdots c_{u_tj} \end{aligned}$$

Damit ergibt sich für die Spur von C^t

$$\begin{aligned} \operatorname{Tr}\left(C^t\right) &= \sum_{i=1}^{n}\sum_{u_2,\ldots,u_t\in\{1,\ldots,n\}} c_{iu_2}c_{u_2u_3}\cdots c_{u_ti} \\ &= \sum_{u_1,\ldots,u_t\in\{1,\ldots,n\}} c_{u_1u_2}c_{u_2u_3}\cdots c_{u_tu_1} \end{aligned}$$

Wir setzen nun $C = X$ mit $c_{ij} = b_{ij}g_{ij}$, sowie $t = 2p$ und erhalten mit der Linearität des Erwartungswerts

$$\begin{aligned} &\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] = \mathbb{E}\left[\sum_{u_1,\ldots,u_{2p}\in\{1,\ldots,n\}} b_{u_1u_2}g_{u_1u_2}b_{u_2u_3}g_{u_2u_3}\cdots b_{u_{2p}u_1}g_{u_{2p}u_1}\right] \\ &= \sum_{u_1,\ldots,u_{2p}\in\{1,\ldots,n\}} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1}\mathbb{E}\left[g_{u_1u_2}g_{u_2u_3}\cdots g_{u_{2p}u_1}\right]. \end{aligned}$$

□

Damit haben wir nun $\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right]$ aus der Proposition 3.11 in eine Form gebracht, welche wir gut bearbeiten können. Um dies zu tun, betrachten wir die rechte Summe aus Lemma 3.13 als Summe von geschlossenen Kantenfolgen im (ungerichteten) Graph, vgl. Definition 2.15,

$$G_n := (V_n, E_n)\,, \qquad V_n := \{1, \dots, n\}, \qquad E_n := \{\{u_i, u_j\} | u_i, u_j \in V_n\}$$

Um die geschlossenen Kantenfolgen in der Summe verwenden zu können definieren wir diese als

$$\mathbf{u} := (u_1, \dots, u_{2p}) \in V_n^{2p}.$$

Dann betrachten wir $\mathbf{u}$ als die geschlossene Kantenfolge

$$u_1 \to u_2 \to \dots \to u_{2p} \to u_1$$

der Länge $2p$. Diese hat dann die Kanten $u_1u_2, u_2u_3, \dots, u_{2p}u_1$. Somit können wir die Indizes der b_{ij} und g_{ij} auffassen als Kanten von $\mathbf{u}$ und schreiben mit dem Ergebnis aus Lemma 3.13:

$$\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] = \sum_{\mathbf{u} \in V_n^{2p}} b_{u_1u_2} b_{u_2u_3} \cdots b_{u_{2p}u_1} \mathbb{E}\left[g_{u_1u_2} g_{u_2u_3} \cdots g_{u_{2p}u_1}\right].$$

Man beachte dabei, dass bei den Kanten selbst keine Richtung vorgegeben ist, es sind also die Kanten u_iu_j und u_ju_i gleich für $1 \leq i, j \leq n$. Dies ist hier nützlich, da die Matrix X (und auch später Y_r) symmetrisch ist, also $g_{ij} = g_{ji}$, bzw. für die b_{ij} ebenso. Das benutzen wir nämlich jetzt, wenn wir die Anzahl der verschiedenen Kanten, die i-mal durchlaufen werden, definieren, wie auch danach, wenn wir über eine gerade Anzahl Kanten sprechen.
Wir definieren also $n_i(\mathbf{u})$ als die Anzahl der verschiedenen Kanten u_ku_l, die genau i-mal von $\mathbf{u}$ durchlaufen werden. Dabei ist egal ob eine Kante als u_ku_l oder u_lu_k durchlaufen wird. Das nutzen wir jetzt, indem wir für jeden Summanden zunächst $\mathbb{E}\left[g_{u_1u_2} g_{u_2u_3} \cdots g_{u_{2p}u_1}\right]$ betrachten. Wir stellen fest, dass die $g_{ij} \overset{u.i.v}{\sim} N(0,1)$, $i \leq j$, unabhängig sind, wir diese also als das Produkt der Erwartungswerte schreiben können. Dabei ist zu beachten, dass, falls die $g_{ij} = g_{ji}$ identisch sind, diese nicht unabhängig sind, also nicht als Produkt der Erwartungswerte geschrieben werden dürfen. Insbesondere ist X symmetrisch. Deshalb spielt die Richtung einer Kante auch keine Rolle. Wir erhalten somit

$$\begin{aligned}
&\sum_{\mathbf{u} \in V_n^{2p}} b_{u_1u_2} b_{u_2u_3} \cdots b_{u_{2p}u_1} \mathbb{E}\left[g_{u_1u_2} g_{u_2u_3} \cdots g_{u_{2p}u_1}\right] \\
&= \sum_{\mathbf{u} \in V_n^{2p}} b_{u_1u_2} b_{u_2u_3} \cdots b_{u_{2p}u_1} \prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})}.
\end{aligned}$$

Dabei misst i, wie oft ein g_{kl} vorkommt, und $n_i(\mathbf{u})$ gibt, wie zuvor definiert, an, für wie viele g_{kl} das gilt. Da die g_{ij} identisch verteilt sind, können wir bei der Erwartungswertbildung für diese ein $g \sim N(0,1)$ verwenden.
Um weiter fortzufahren betrachten wir nun $\mathbf{u}$ genauer und definieren:
Wir sagen, dass $\mathbf{u}$ gerade ist, falls jede Kante genau in einer geraden Anzahl von $\mathbf{u}$ durchlaufen wird, d.h. falls i ungerade ist, muss $n_i(\mathbf{u}) = 0$ sein.
Ist nun $\mathbf{u}$ nicht gerade, so gilt mindestens für ein i, dass $n_i(\mathbf{u}) \neq 0$ und somit

$$\prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})} = \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})} \prod_{j \neq i} \left(\mathbb{E}\left[g^j\right]\right)^{n_j(\mathbf{u})}.$$

Da nun aber $g \sim N(0,1)$ ist, und die Normalverteilung achsensymmetrisch an der y-Achse ist, ist $\left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})} = 0$. Es folgt somit, dass nur die geraden $\mathbf{u}$ berücksichtigt werden müssen:

$$\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] = \sum_{\mathbf{u} \in V_n^{2p}, \mathbf{u} \text{ gerade}} b_{u_1 u_2} b_{u_2 u_3} \cdots b_{u_{2p} u_1} \prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})}.$$

Wir fassen nun die verschiedenen Kantenfolgen gleicher Form zusammen: Sei $\mathbf{u}$ eine geschlossene Folge von Kanten, dann definieren wir $\mathbf{s(u)}$ als sie Form von $\mathbf{u}$, indem wir $\mathbf{u}$ nach dem Zeitpunkt des Auftretens der einzelnen Ecken reskalieren.

Beispiel 3.14.

- Wir betrachten die geschlossene Kantenfolge $\mathbf{u}_1 = (5,6,8,9,7,9) \in V_n^6$ mit $n \geq 9$. Dann können wir $\mathbf{u}_1$ schreiben als

  $$5 \to 6 \to 8 \to 9 \to 7 \to 9 \to 5.$$

  Das ergibt die Form $\mathbf{s(u_2)}$ von $\mathbf{u}_2$

  $$1 \to 2 \to 3 \to 4 \to 5 \to 4 \to 1$$

  Es gilt weiter $n_1(\mathbf{u}_1) = 4$ und $n_2(\mathbf{u}_1) = 1$, somit ist $\mathbf{u}_1$ nicht gerade.

- Wir betrachten die geschlossene Kantenfolge $\mathbf{u}_2 = (4,6,8,1,8,6) \in V_n^6$ mit $n \geq 9$. Dann können wir $\mathbf{u}_2$ schreiben als

  $$4 \to 6 \to 8 \to 1 \to 8 \to 6 \to 4.$$

  Das ergibt die Form $\mathbf{s(u_1)}$ von $\mathbf{u}_1$

  $$1 \to 2 \to 3 \to 4 \to 3 \to 2 \to 1$$

  Es gilt weiter $n_1(\mathbf{u}_2) = 0$ und $n_2(\mathbf{u}_2) = 3$, somit ist $\mathbf{u}_2$ gerade.

- Sei $\mathbf{u}_3 = (9,3,5,2,5,3) \in V_n^6$ eine weiter Kantenfolge

  $$9 \to 3 \to 5 \to 2 \to 5 \to 3 \to 9,$$

  dann ist $\mathbf{s(u_3)}$

  $$1 \to 2 \to 3 \to 4 \to 3 \to 2 \to 1$$

  und somit habe $\mathbf{u}_2$ und $\mathbf{u}_3$ die gleiche Form $\mathbf{s(u_2)} = \mathbf{s(u_3)}$.

Wir beobachten, dass

$$\prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})}$$

nicht mehr von $\mathbf{u}$ selbst abhängt, sondern nur noch von der Form $\mathbf{s(u)}$ von $\mathbf{u}$. Dies ist so, da g unabhängig von den einzelnen $u_i u_j$ ist und die Anzahl, wie oft bestimmte Kanten durchlaufen werden, nur von der Form abhängig ist, wie in folgendem Lemma gezeigt wird:

Lemma 3.15. *Sei $\boldsymbol{u} \in V_n^{2p}$ mit Form $\boldsymbol{s(u)} = \boldsymbol{s}$, wobei $p, n \in \mathbb{N}$, dann gilt*

$$n_i(\boldsymbol{u}) = n_i(\boldsymbol{s})$$

für alle $i \geq 0$.

Beweis von Lemma 3.15. Sei $\mathbf{u} \in V_n^{2p}$ mit

$$\mathbf{u} := (u_1, \dots, u_{2p}) \in V_n^{2p}$$

und sei $\mathbf{s(u)} = \mathbf{s}$ die Form von $\mathbf{u}$ mit

$$\mathbf{s} := (1, s_2, \dots, s_{2p}),$$

wobei die $s_2, \dots, s_{2p}$ die nach dem Zeitpunkt des ersten Auftretens reskalierten Ecken sind. Wir schreiben

$$u_1 \to u_2 \to \cdots \to u_{2p} \to u_1,$$

sowie

$$1 \to s_2 \to \cdots \to s_{2p} \to 1$$

und erkennen, dass in der reskalierten Formschreibweise die Kante u_1u_2, genau der Kante $1s_2$, und $u_2u_3$der Kante s_2s_3, usw. entspricht. Wir schließen, wenn die Kante u_1u_2 genau i-mal vorkommt, dass dann auch die Kante $1s_2$ genau i-mal vorkommt, wenn die Kante u_2u_3 genau i-mal vorkommt, dass dann auch die Kante s_2s_3 genau i-mal vorkommt, für die weiteren Kanten analog. D.h. es gibt jeweils genau gleich viele Kanten, die i-mal besucht werden für jedes i. Somit ergibt sich $n_i(\mathbf{u}) = n_i(\mathbf{s})$ für alle i. □

Nach diesen Überlegungen lässt sich die Aussage auf die Formen gerader Kantenfolgen der Länge $2p$ zurückführen, diese definieren wir als:

$$S_{2p} = \{\mathbf{s(u)} | \mathbf{u} \text{ ist gerade Kantenfolge der Länge } 2p\}$$

Damit können wir jetzt von $n_i(\mathbf{s})$ sprechen für $\mathbf{s} \in S_{2p}$. Es ist also

$$\prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})} = \prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})}$$

für alle $\mathbf{u}$ mit Form $\mathbf{s}$.
Um dies nun in der Summe über alle $\mathbf{u}$ schreiben zu können führen wir noch den Begriff der Menge aller geraden Kantenfolgen mit der Form $\mathbf{s} \in S_{2p}$ ein, die an einer bestimmten Ecke $u \in V_n$ starten:

$$\Gamma_{\mathbf{s},u} = \{\mathbf{u} \in V_n | \mathbf{s(u)} = \mathbf{s}, u_1 = u\}.$$

Damit können wir schreiben

$$\begin{aligned} \mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] &= \sum_{\mathbf{u} \in V_n^{2p}, \mathbf{u} \text{ gerade}} b_{u_1u_2} b_{u_2u_3} \cdots b_{u_{2p}u_1} \prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{u})} \\ &= \sum_{u \in \{1,\dots,n\}} \sum_{\mathbf{s} \in S_{2p}} \prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \sum_{\mathbf{u} \in \Gamma_{\mathbf{s},u}} b_{uu_2} b_{u_2u_3} \cdots b_{u_{2p}u}. \end{aligned} \tag{3.5}$$

Damit haben wir die linke Seite in eine Form gebracht, die wir gut abschätzen können. Das tun wir in den nächsten beiden Lemmata mit denen der Beweis der Proposition 3.11 dann sehr zügig gehen wird. Davor führen wir noch einen Begriff ein, den wir im nächsten Lemma benötigen:
Sei $\mathbf{s} = (s_1, \dots, s_{2p}) \in S_{2p}$ eine Form von Kanten, dann definieren wir

$$m(\mathbf{s}) = \max_{1 \leq i \leq 2p} s_i$$

als die Anzahl der verschiedenen von $\mathbf{s}$ durchlaufenen Ecken. Man beachte, dass bei obiger Notation eine Kante, die bereits durchlaufen wurde den ersten Index bei dem sie Auftritt auch für ihr weiteres Auftreten behält.
Weiter definieren wir für eine Ecke u und eine Form $\mathbf{s} = (s_1, \dots, s_{2p}) \in S_{2p}$ den Zeitpunkt

$$i(k) := \inf\{j | s_j = k\}$$

an dem die k-te unterschiedliche Ecke das erste mal erscheint. Aus der vorherigen Definition von $m(\mathbf{s})$ resultiert, dass $1 \leq k \leq m(\mathbf{s})$. Zudem ergibt sich, dass $i(1) = 1$ per Definition.

Beispiel 3.16. Wir betrachten wieder die geschlossene Kantenfolge $\mathbf{u}_1 = (5, 6, 8, 9, 7, 9) \in V_n^6$ mit $n \geq 9$ mit der Form

$$1 \to 2 \to 3 \to 4 \to 5 \to 4 \to 1$$

und erkennen, dass $m(\mathbf{s(u_1)}) = 5$, sowie $i(k) = k$ für alle $1 \leq k \leq 5$.

Schritt 2: Nun schätzen wir die für die Spur beschriebene Summe ab:

Lemma 3.17. *Sei $\sigma_* \leq 1$, dann gilt für jedes $u \in \{1, \dots, n\}$ und $\boldsymbol{s} \in S_{2p}$, dass*

$$\sum_{\boldsymbol{u} \in \Gamma_{s,u}} b_{uu_2} b_{u_2u_3} \cdots b_{u_{2p}u} \leq \sigma^{2(m(\boldsymbol{s})-1)}.$$

Insbesondere gilt dann auch

$$\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] \leq n \sum_{\boldsymbol{s} \in S_{2p}} \sigma^{2(m(\boldsymbol{s})-1)} \prod_{i \geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\boldsymbol{s})}.$$

Beweis von Lemma 3.17. Fall 1: $m(\boldsymbol{s}) \geq 2$. Wir betrachten eine Kantenfolge $\mathbf{u} \in \Gamma_{\mathbf{s},u}$. Da $\mathbf{u}$ die Form $\mathbf{s} \in S_{2p}$ hat und $\mathbf{s}$ gerade ist (wie zuvor überlegt), muss jede Kante, die durchlaufen wird, mindestens zweimal durchlaufen werden. Insbesondere gilt das dann auch für jede Kante $\{u_{i(k)-1}, u_{i(k)}\}$ für ein $2 \leq k \leq m(\mathbf{s})$. Zudem muss die Kante $\{u_{i(k)-1}, u_{i(k)}\}$ sich von allen vorherigen Kanten $\{u_{i(l)-1}, u_{i(l)}\}$ mit $l < k$ unterscheiden, da wie oben definiert, die Ecke $u_{i(k)}$ zu diesem Zeitpunkt das erste Mal auftritt, weil $i(k)$ ja genau der Index des ersten Auftretens der k-ten unterschiedlichen Ecke ist. Mit $\max_{ij} |b_{ij}| = \sigma_* \leq 1$ können wir abschätzen

$$\begin{aligned}
&\sum_{\mathbf{u} \in \Gamma_{\mathbf{s},u}} b_{uu_2} b_{u_2u_3} \cdots b_{u_{2p}u} \leq \sum_{\mathbf{u} \in \Gamma_{\mathbf{s},u}} b^2_{uu_{i(2)}} b^2_{u_{i(3)-1}u_{i(3)}} \cdots b^2_{u_{i(m(\mathbf{s}))-1}u_{i(m(\mathbf{s}))}} \\
&= \sum_{v_2,\dots,v_{m(\mathbf{s})}, u \neq v_i \neq v_j \text{ für } i \neq j} b^2_{uv_2} b^2_{v_{s_{i(3)-1}} v_{i(3)}} \cdots b^2_{v_{s_{i(m(\mathbf{s}))-1}} v_{m(\mathbf{s})}}
\end{aligned}$$

denn mit $\sigma_* \leq 1$ sind alle höheren Potenzen ebenfalls kleiner/gleich. Die zweite Gleichheit gilt, da die obere Summe nur noch von der Form und den Kombinationsmöglichkeiten abhängt. Weiter ist $s_{i(k)-1} < k$ per Konstruktion und somit auch nur von den $u, v_2, \dots, v_{k-1}$ abhängig. Außerdem ist durch die feste Form auch schon festgelegt, welches $s_{i(k)-1} \in u, v_2, \dots, v_{k-1}$ gesetzt werden muss. Das führt zu folgendem Ergebnis:

$$\begin{aligned}
&\sum_{v_2,\dots,v_{m(\mathbf{s})},u\neq v_i\neq v_j \text{ für } i\neq j} b_{uv_2}^2 b_{v_{s_{i(3)-1}}v_{i(3)}}^2 \cdots b_{v_{s_{i(m(\mathbf{s}))-1}}v_{m(\mathbf{s})}}^2 \\
&= \sum_{u\neq v_2} b_{uv_2}^2 \Bigg(\sum_{v_3,u\neq v_3\neq v_2} b_{v_{s_{i(3)-1}}v_{i(3)}}^2 \\
&\cdots \Bigg(\sum_{v_{m(\mathbf{s})},u\neq v_{m(\mathbf{s})}\neq v_j \text{ für } j<m(\mathbf{s})} b_{v_{s_{i(m(\mathbf{s}))-1}}v_{m(\mathbf{s})}}^2 \Bigg)\Bigg) \\
&\leq \underbrace{\sigma^2(\sigma^2\cdots(\sigma^2))}_{m(\mathbf{s})-1\text{-mal}} = \sigma^{2(m(\mathbf{s})-1)}
\end{aligned}$$

Dabei gilt die Ungleichung, weil jedes $\sum_{v_k,u\neq v_k\neq v_i \text{ für } i<k} b_{v_{s_{i(k)-1}}v_k}^2 \leq \sigma^2$ per Definition von σ^2 und weil, wie oben beschrieben, $s_{i(k)-1}$ schon durch die vorherigen $u, v_2, \dots, v_{k-1}$ festgelegt ist.

Fall 2: $m(\boldsymbol{s}) = 1$. Dann gilt schon, da nur eine Ecke/Kante immer wieder durchlaufen wird,

$$\sum_{\mathbf{u}\in\Gamma_{\mathbf{s},u}} b_{uu_2}b_{u_2u_3}\cdots b_{u_{2p}u} = b_{uu}^{2p} \leq 1 = \sigma^0 = \sigma^{2(m(\mathbf{s})-1)},$$

denn $|b_{uu}| \leq \sigma_* = 1$.

Damit ist der erste Teil des Lemmas gezeigt, der zweite folgt durch Einsetzen in die Gleichung (3.5):

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] &= \sum_{u\in\{1,\dots,n\}} \sum_{\mathbf{s}\in S_{2p}} \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \sum_{\mathbf{u}\in\Gamma_{\mathbf{s},u}} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1} \\
&\leq \sum_{u\in\{1,\dots,n\}} \underbrace{\sum_{\mathbf{s}\in S_{2p}} \sigma^{2(m(\mathbf{s})-1)} \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})}}_{\text{unabhängig von } u} \\
&= n \sum_{\mathbf{s}\in S_{2p}} \sigma^{2(m(\mathbf{s})-1)} \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})}
\end{aligned}$$

□

Schritt 3: Um die Proposition 3.11 beweisen zu können benötigen wir noch eine analoge Abschätzung für die Matrix Y_r, welche in der Proposition 3.11 ebenfalls auftritt.

Lemma 3.18. *Sei Y_r wie in Proposition 3.11 definiert, sowie $r > p$, dann gilt*

$$\mathbb{E}\left[\operatorname{Tr}\left(Y_r^{2p}\right)\right] = r \sum_{\boldsymbol{s}\in S_{2p}} (r-1)(r-2)\cdots(r-m(\boldsymbol{s})+1) \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\boldsymbol{s})}.$$

Beweis von Lemma 3.18. Wir erkennen zunächst, dass Y_r ein Spezialfall von X aus den vorherigen Überlegungen mit $n = r$ und $b_{ij} = 1$ für alle $1 \leq i, j \leq r$ ist. Also gilt Gleichung (3.5) für Y_r:

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Tr}\left(Y_r^{2p}\right)\right] &= \sum_{u\in\{1,\dots,n\}} \sum_{\mathbf{s}\in S_{2p}} \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \sum_{\mathbf{u}\in\Gamma_{\mathbf{s},u}} \underbrace{b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1}}_{=1} \\
&= \sum_{\mathbf{s}\in S_{2p}} \sum_{u\in\{1,\dots,n\}} \sum_{\mathbf{u}\in\Gamma_{\mathbf{s},u}} \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \\
&= \sum_{\mathbf{s}\in S_{2p}} \left|\{\mathbf{u}\in\{1,\dots,r\}^{2p} : \mathbf{s}(\mathbf{u}) = \mathbf{s}\}\right| \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})}
\end{aligned}$$

Als nächstes bestimmen wir nun $|\{\mathbf{u} \in \{1,\dots,r\}^{2p} : \mathbf{s(u)} = \mathbf{s}\}|$. Dazu überlegen wir uns, dass jede geschlossene Folge von Kanten $\{\mathbf{u} \in \{1,\dots,r\}^{2p}$ der Länge $2p$ und der Form $\mathbf{s(u)} = \mathbf{s}$ schon durch seine $m(\mathbf{s})$ verschiedenen Ecken in der Reihenfolge des ersten Auftretens hinreichend bestimmt ist.
Um $|\{\mathbf{u} \in \{1,\dots,r\}^{2p} : \mathbf{s(u)} = \mathbf{s}\}|$ zu bestimmen, genügt es also alle möglichen Kombinationen des ersten Auftretens einer Ecke aufzuschreiben. Dabei beachten wir, dass, wenn eine Ecke das erste Mal auftritt, diese nicht noch einmal zum ersten Mal auftreten kann. Falls nun $m(\mathbf{s}) \leq r$ ist, gilt für ein $\mathbf{s} \in S_{2p}$

$$|\{\mathbf{u} \in \{1,\dots,r\}^{2p} : \mathbf{s(u)} = \mathbf{s}\}| = r(r-1)\cdots(r-m(\mathbf{s})+1)$$

Die Bedingung $m(\mathbf{s}) \leq r$ ist erfüllt, denn jede gerade Kantenfolge der Länge $2p$ kann maximal $m(\mathbf{s}) \leq p+1$ verschiedene Ecken durchlaufen. Mit der Voraussetzung $p < r$ gilt dann $m(\mathbf{s}) \leq p+1 \leq r$. Wir setzen ein:

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Tr}\left(Y_r^{2p}\right)\right] &= \sum_{\mathbf{s}\in S_{2p}} |\{\mathbf{u} \in \{1,\dots,r\}^{2p} : \mathbf{s(u)} = \mathbf{s}\}| \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \\
&= \sum_{\mathbf{s}\in S_{2p}} r(r-1)\cdots(r-m(\mathbf{s})+1) \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \\
&= r \sum_{\mathbf{s}\in S_{2p}} (r-1)(r-2)\cdots(r-m(\mathbf{s})+1) \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})}
\end{aligned}$$

Das r können wir ausklammern, da $m(\mathbf{s}) \geq 1$ für alle $\mathbf{s} \in S_{2p}$. □

Schritt 4: Nun kommen wir mit den Ergebnissen der Lemmata 3.17 und 3.18 zum Beweis von Proposition 3.11:

Beweis von Proposition 3.11. Zur Erinnerung: für $\sigma_* \leq 1$ möchten wir zeigen

$$\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] \leq \frac{n}{\lceil\sigma^2\rceil + p}\mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil\sigma^2\rceil+p}\right)\right].$$

Dazu fixieren wir ein $p \in \mathbb{N}$ und setzen $r = \lceil\sigma^2\rceil + p$. Folglich gilt für $l \leq m(\mathbf{s})-1$, dass

$$r - l \geq r - m(\mathbf{s}) + 1 \geq \sigma^2 + p - m(\mathbf{s}) + 1,$$

denn $\sigma^2 + p \leq \lceil\sigma^2\rceil + p = r$. Es folgt

$$(r-1)(r-2)\cdots(r-m(\mathbf{s})+1) \geq (\sigma^2 + p - m(\mathbf{s}) + 1)^{m(\mathbf{s})-1} \geq \sigma^{2m(\mathbf{s})-1}.$$

Die zweite Ungleichung gilt, denn $p - m(\mathbf{s}) + 1 \geq 0$, weil jede gerade Kantenfolge der Länge $2p$ maximal $m(\mathbf{s}) \leq p+1$ verschiedene Ecken durchlaufen kann. Mit diesem Ergebnis und mit Lemma 3.17, welches wir anwenden können, da $\sigma_* \leq 1$, folgt

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Tr}\left(X^{2p}\right)\right] &\leq n \sum_{\mathbf{s}\in S_{2p}} \sigma^{2(m(\mathbf{s})-1)} \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \\
&\leq n \sum_{\mathbf{s}\in S_{2p}} (r-1)(r-2)\cdots(r-m(\mathbf{s})+1) \prod_{i\geq 1} \left(\mathbb{E}\left[g^i\right]\right)^{n_i(\mathbf{s})} \\
&= n\tfrac{1}{r}\mathbb{E}\left[\operatorname{Tr}\left(Y_r^{2p}\right)\right] \\
&= \tfrac{n}{\lceil\sigma^2\rceil+p}\mathbb{E}\left[\operatorname{Tr}\left(Y^{2p}_{\lceil\sigma^2\rceil+p}\right)\right].
\end{aligned}$$

Zudem haben wir Lemma 3.18 benutzt, dieses können wir benutzen, weil $r = \lceil\sigma^2\rceil + p \geq p$. Im letzten Schritt haben wir noch das r, wie am Anfang gewählt, ersetzt. □

Damit haben wir zwei wichtige Hilfsaussagen für den Beweis des Satz 3.6 bewiesen. Zu Beginn des Beweises von Satz 3.6 benutzen wir noch, dass ebenfalls zuvor formulierte Lemma 3.12. Dessen Aussage

$$\|X\| \leq \left(\operatorname{Tr}\left(X^{2p}\right)\right)^{\frac{1}{2p}} \leq n^{\frac{1}{2p}} \|X\|$$

für X als symmetrische $n \times n$-Matrix möchten wir nun beweisen:

Beweis von Lemma 3.12. Da X symmetrisch ist, ist X nach dem Spektralsatz für symmetrisch Matrizen orthogonal diagonalisierbar mit Eigenwerten $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Somit gilt nach Proposition 2.13, dass

$$\|X\| = \max_{1 \leq i \leq n} |\lambda_i|.$$

Weiter gilt nach Bemerkung 2.12, da X symmetrisch ist, auch, dass X^{2p} symmetrisch mit Eigenwerten $\lambda_1^{2p}, \dots, \lambda_n^{2p}$ ist. Deswegen gilt auch

$$\operatorname{Tr}(X^{2p}) = \sum_{i=1}^{n} \lambda_i^{2p}.$$

Mit der Ungleichung

$$\max_{1 \leq i \leq n} |\lambda_i|^{2p} \leq \sum_{i=1}^{n} \lambda_i^{2p} \leq n \max_{1 \leq i \leq n} |\lambda_i|^{2p}$$

folgt schließlich

$$\|X\| = \max_{1 \leq i \leq n} |\lambda_i| \leq \underbrace{\left(\sum_{i=1}^{n} \lambda_i^{2p}\right)^{\frac{1}{2p}}}_{=(\operatorname{Tr}(X^{2p}))^{\frac{1}{2p}}} \leq n^{\frac{1}{2p}} \max_{1 \leq i \leq n} |\lambda_i| = n^{\frac{1}{2p}} \|X\|$$

und damit die Behauptung. $\square$

Kapitel 4

Ergänzungen

Dieses Kapitel orientiert sich an [Ban+16, Kapitel 3].

4.1 Weitere Eigenschaften der Spektralnorm Gaußscher Zufallsmatrizen

Zunächst möchten wir, nachdem wir uns in Kapitel 3 ausführlich damit beschäftigt haben, wie wir die Spektralnorm gaußscher Zufallsmatrizen nach oben abschätzen können, uns überlegen, wie dies auch nach unten geht. Dazu betrachten wir die gleiche Situation wie in der Abschätzung der Spektralnorm nach oben, Satz 3.6. Es ist also X symmetrisch mit Einträgen $b_{ij}g_{ij}$.

Proposition 4.1. *[Ban+16, Lemma 3.14] In der Situation wie in Satz 3.6 gilt*

$$\mathbb{E}\left[\|X\|\right] \gtrsim \sigma + \mathbb{E}\left[\max_{1\le i,j\le n} |b_{ij}g_{ij}|\right].$$

Beweis. Schritt 1: Wenn wir die kanonische Basis $\{e_i\}$ des $\mathbb{R}^n$ betrachten, können wir für $1 \le i \le n$ rechnen

$$\begin{aligned}\max_i \|Xe_i\| &= \max_i \|(X_{i1},\dots,X_{in})^\top\| = \max_i \left(X_{i1}^2+\dots+X_{in}^2\right)^{\frac{1}{2}} \\ &\ge \max_i(\max_j |X_{ij}|) = \max_{1\le i,j\le n} |X_{ij}|\end{aligned}$$

Da $\|X\| = \max\limits_{\|v\|=1} \|Xv\|$ und $\|e_i\| = 1$ ist, folgt schon, dass

$$\|X\| \ge \max_{1\le i,j\le n} |X_{ij}| = \max_{1\le i,j\le n} |b_{ij}g_{ij}|$$

gilt. Mit Erwartungswertbildung ergibt sich dann

$$\mathbb{E}\left[\|X\|\right] \ge \mathbb{E}\left[\max_{1\le i,j\le n} |X_{ij}|\right] = \mathbb{E}\left[\max_{1\le i,j\le n} |b_{ij}g_{ij}|\right]. \tag{4.1}$$

Schritt 2: Wir betrachten wieder die kanonische Basis $\{e_i\}$ des $\mathbb{R}^n$ und stellen fest, dass $\|X\| \ge \max_i \|Xe_i\|$ gilt. Zum Abschätzen des Erwartungswerts der Norm benötigen wir zunächst folgende Überlegung: Nach den Standardregeln für die Varianz einer Zufallsvariablen Y gilt $\mathrm{Var}(Y) = \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2$, es ergibt sich also für $Y \sim \|Xe_i\|$

$$\mathbb{E}\left[\|Xe_i\|^2\right] = (\mathbb{E}[\|Xe_i\|])^2 + \mathrm{Var}(\|Xe_i\|). \tag{4.2}$$

Wir betrachten die Abschätzung

$$\mathrm{Var}(\|Xe_i\|) \leq \max_{1\leq j\leq n} b_{ij}^2 \lesssim \max_{1\leq j\leq n} (\mathbb{E}[b_{ij}|g_{ij}|])^2 \leq (\mathbb{E}[\|Xe_i\|])^2. \tag{4.3}$$

Die erste Ungleichung ergibt sich dabei aus der Gaußschen Poincaré Ungleichung [Bou+12, Theorem 3.20]: Für einen multivariat standardnormalverteilten Zufallsvektor $Y = (Y_1, \ldots, Y_n)^\top$ und eine stetig differenzierbare Funktion $f : \mathbb{R}^n \to \mathbb{R}$ gilt

$$\mathrm{Var}(f(Y)) \leq \mathbb{E}\left[\|\nabla f(Y)\|^2\right].$$

In unserem Fall gilt für festes i, dass $Y_j = g_{ij}$ und

$$f(Y) = \|Xe_i\| = \left(b_{i1}^2 g_{i1}^2 + \ldots + b_{in}^2 g_{in}^2\right)^{\frac{1}{2}}.$$

Mit der Kettenregel folgt

$$\nabla f(Y) = \frac{\left(b_{i1}^2 g_{i1}, \ldots, b_{in}^2 g_{in}\right)^\top}{f(Y)}.$$

Somit erhalten wir

$$\begin{aligned}
\mathrm{Var}(f(Y)) &\leq \mathbb{E}\left[\|\nabla f(Y)\|^2\right] = \mathbb{E}\left[\left\|\frac{\left(b_{i1}^2 g_{i1}, \ldots, b_{in}^2 g_{in}\right)^\top}{f(Y)}\right\|^2\right] \\
&= \mathbb{E}\left[\frac{\left(b_{i1}^4 g_{i1}^2 + \ldots + b_{in}^4 g_{in}^2\right)}{\|Xe_i\|}\right] \\
&\leq \mathbb{E}\left[\max_{1\leq j\leq n} b_{ij}^2 \cdot \frac{\left(b_{i1}^2 g_{i1}^2 + \ldots + b_{in}^2 g_{in}^2\right)}{\|Xe_i\|}\right] \\
&= \mathbb{E}\left[\max_{1\leq j\leq n} b_{ij}^2 \cdot \frac{\|Xe_i\|}{\|Xe_i\|}\right] = \max_{1\leq j\leq n} b_{ij}^2,
\end{aligned}$$

wobei $\|Xe_i\| \neq 0$ fast sicher.
Die zweite Ungleichung gilt, da der Erwartungswert der Halbnormaverteilung endlich und größer 0 ist:

$$\max_{1\leq j\leq n} b_{ij}^2 \lesssim \max_{1\leq j\leq n} b_{ij}^2 (\mathbb{E}[|g_{ij}|])^2 = \max_{1\leq j\leq n} (\mathbb{E}[b_{ij}|g_{ij}|])^2.$$

Die dritte Ungleichung gilt, weil $b_{ij}|g_{ij}| \leq \|Xe_i\|$ ist. Wenn wir (4.3) in (4.2) einsetzen folgt

$$\mathbb{E}\left[\|Xe_i\|^2\right] = (\mathbb{E}[\|Xe_i\|])^2 + \mathrm{Var}(\|Xe_i\|) \lesssim (\mathbb{E}[\|Xe_i\|])^2.$$

Damit und mit der Beobachtung zu Beginn von Schritt 2 können wir schreiben

$$\mathbb{E}[\|X\|] \geq \max_{1\leq i\leq n} \mathbb{E}[\|Xe_i\|] \gtrsim \max_{1\leq i\leq n} \left(\mathbb{E}\left[\|Xe_i\|^2\right]\right)^{\frac{1}{2}} = \sigma. \tag{4.4}$$

Schritt 3: Wir fassen die Ergebnisse (4.1) und (4.4) aus Schritt 1 und 2 zusammen und bilden den Mittelwert:

$$\mathbb{E}[\|X\|] \gtrsim \sigma + \mathbb{E}\left[\max_{1\leq i,j\leq n} |b_{ij} g_{ij}|\right]$$

Dabei müssen wir aufgrund der $\gtrsim$ Notation keinen Vorfaktor hinzufügen, was den Beweis beendet. □

Damit haben wir eine adäquate untere Abschätzung der Spektralnorm von X gefunden. Wenn man genug große Koeffizienten b_{ij} in X hat, dann zeigt diese untere Abschätzung einerseits, dass die obere Abschätzung nach Bandeira und van Handel, Satz 3.6, scharf ist. Andererseits folgt aber auch aus Satz 3.6, dass in diesem Fall diese untere Abschätzung scharf ist. Die genaue Situation in der dies auftritt und den Beweis dazu betrachten wir in dem folgenden Korollar:

Korollar 4.2. *[Ban+16, Corollary 3.15] Wir betrachten die Situation von Satz 3.6 und nehmen für zwei Konstanten $c, \alpha > 0$ an, dass*

$$|\{ij : |b_{ij}| \geq c\sigma_*\}| \geq n^\alpha.$$

Dann gilt

$$\mathbb{E}[\|X\|] \asymp \sigma + \sigma_* \sqrt{\log n},$$

dabei hängt die universelle Konstante für die untere Grenze nur von c und α ab.

Damit gilt die Schärfe der Abschätzung für alle Fälle sobald in jeder n^α-ten Reihe, z.B. $\alpha = 1$, ein großer Koeffizient ist. Das ist somit in vielen Beispielen der Fall. Für den Beweis des Korollars benötigen wir noch folgendes Lemma:

Lemma 4.3. *Für $n \in \mathbb{N}$ unabhängig standardnormalverteilte Zufallsvariablen, $X_1, \ldots, X_n \overset{u.i.v}{\sim} N(0,1)$, gilt*

$$\mathbb{E}\left[\max_{1 \leq i \leq n} |X_i|\right] \gtrsim \sqrt{\log n}.$$

Für den Beweis des Lemmas verweisen wir hier auf [Ora+15].

Beweis von Korollar 4.2. Wir notieren zunächst

$$I := \{ij : |b_{ij}| \geq c\sigma_*\},$$

also ist $n^\alpha \leq |I| \leq n^2$. Damit können wir schreiben

$$\mathbb{E}\left[\max_{ij} |b_{ij} g_{ij}|\right] \geq \mathbb{E}\left[\max_{ij \in I} |b_{ij} g_{ij}|\right] \geq c\sigma_* \mathbb{E}\left[\max_{ij \in I} |g_{ij}|\right].$$

Die rechte Ungleichung folgt dabei aus den Voraussetzungen an I und der Linearität des Erwartungswerts. Mit Lemma 4.3 können wir nun die rechte Seite nach unten abschätzen:

$$c\sigma_* \mathbb{E}\left[\max_{ij \in I} |g_{ij}|\right] \gtrsim \sigma_* \sqrt{\log |I|}$$

Dabei ist die universelle Konstante von dem gewählten c abhängig. Aus der Voraussetzung $n^\alpha \leq |I|$und der Monotonie des Logarithmus folgt dann

$$\sigma_* \sqrt{\log |I|} \gtrsim \sigma_* \sqrt{\log(n^\alpha)} = \sigma_* \sqrt{\alpha} \sqrt{\log n} \gtrsim \sigma_* \sqrt{\log n}.$$

Insgesamt gilt also

$$\mathbb{E}\left[\max_{ij} |b_{ij} g_{ij}|\right] \gtrsim \sigma_* \sqrt{\log n},$$

wobei die universellen Konstanten von der Wahl von c und α abhängen. Mit Proposition 4.1 folgt somit

$$\mathbb{E}\left[\|X\|\right] \gtrsim \sigma + \mathbb{E}\left[\max_{1\leq i,j\leq n} |b_{ij}g_{ij}|\right] \gtrsim \sigma + \sigma_* \sqrt{\log n}.$$

Die Abschätzung nach oben folgt direkt aus Satz 3.6:

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma + \sigma_* \sqrt{\log n}$$

Zusammen folgt die Behauptung. □

Nachdem wir jetzt den Erwartungswert der Spektralnom nach unten und oben abgeschätzt haben, möchten wir uns weiter mit der Wahrscheinlichkeitsverteilung der Spektralnorm der Matrix X, wie in Satz 3.6, beschäftigen, genauer gesagt mit deren Tails. Dazu werden wir wieder die Gauß-Konzentration benutzen.

Proposition 4.4. *[Ban+16, Corollary 3.9] Wir betrachten die Situation von Satz 3.6 für $0 < \varepsilon \leq \frac{1}{2}$. Dann gilt für $t \geq 0$, dass*

$$\mathbb{P}\left(\|X\| \geq (1+\varepsilon)\left(2\sigma + \frac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right) + t\right) \leq e^{-t^2/(4\sigma_*^2)}.$$

Damit haben wir ohne großen Aufwand eine nützliche Abschätzung der Tails gewonnen. Des Weiteren erkennen wir, dass die Spektralnorm von X, wie in Satz 3.6, eine sub-gaußsche Zufallsvariable ist.

Beweis. Nach Lemma 2.14 können wir schreiben

$$\|X\| = \sup_{v\in S^{n-1}} |\langle v, Xv\rangle|.$$

Wir möchten die Gauß-Konzentration, genauer Satz 2.42 anwenden. Dazu fassen wir $\|X\|$ als das Supremum des Gaußprozesses $\{|\langle v, Xv\rangle|\}_{v\in S^{n-1}}$ auf. Mit der Symmetrie von X und Unabhängigkeit der Einträge gilt

$$\begin{aligned}
\mathbb{E}\left[|\langle v, Xv\rangle|^2\right] &= \mathbb{E}\left[\left(\sum_{ij} v_i X_{ij} v_j\right)^2\right] \\
&= \mathbb{E}\left[\sum_{ij} X_{ij}^2 v_i^2 v_j^2 + \sum_{i\neq j} X_{ij}^2 v_i^2 v_j^2 + \sum_{i\neq k \text{ oder } j\neq l} X_{ij}X_{kl}v_iv_jv_kv_l\right] \\
&= \mathbb{E}\left[\sum_{ij} b_{ij}^2 g_{ij}^2 v_i^2 v_j^2 + \sum_{i\neq j} b_{ij}^2 g_{ij}^2 v_i^2 v_j^2\right] \\
&= \sum_{ij} b_{ij}^2 v_i^2 v_j^2 + \sum_{i\neq j} b_{ij}^2 v_i^2 v_j^2 \\
&\leq 2\sigma_*^2,
\end{aligned}$$

wobei wir uns die Linearität des Erwartungswerts, die Varianz der Standardnormalverteilung und die Eigenschaften der v_i zu Nutze machen. Nach Satz 3.6 gilt $\mathbb{E}\left[\|X\|\right] < \infty$. Mit Satz 2.42 folgt also für $t \geq 0$

$$\begin{aligned}
&\mathbb{P}\left(\|X\| - \mathbb{E}\left[\|X\|\right] \geq t\right) \leq e^{-t^2/(2\cdot 2\sigma_*^2)} \\
\Leftrightarrow\ &\mathbb{P}\left(\|X\| \geq \mathbb{E}\left[\|X\|\right] + t\right) \leq e^{-t^2/(4\sigma_*^2)} \\
\Rightarrow\ &\mathbb{P}\left(\|X\| \geq (1+\varepsilon)\left(2\sigma + \tfrac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right) + t\right) \leq e^{-t^2/(4\sigma_*^2)}
\end{aligned}$$

Die letzte Gleichung gilt dabei wegen der Monotonie der Wahrscheinlichkeitsverteilung und Satz 3.6. □

4.2 Dimensionsfreie Abschätzung

Das Hauptresultat, Satz 3.6,

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma + \sigma_* \sqrt{\log n}$$

ist zur Abschätzung des Erwartungswerts der Spektralnorm von der Dimension n abhängig. Da

$$\sqrt{\log n} \to \infty \qquad \text{für } n \to \infty$$

bringt das einige Nachteile für große Dimensionen mit sich, insbesondere, wenn eine Matrix mit niedriger Dimension eingebettet wird. In diesem Fall ist auch die untere Abschätzung der Spektralnorm, Korollar 4.2, nicht mehr gültig und die obere Abschätzung ist nicht mehr scharf. Somit ergibt sich die Idee die Dimension n durch eine „effektive Dimension“ zu ersetzen. Dafür definieren wir

$$|(b_{ij})|_p := \left[\sum_{1 \leq i,j \leq n} |b_{ij}|^p \right]^{\frac{1}{p}}, \tag{4.5}$$

welches unsere „effektive Dimension“ sein wird. Damit können wir folgende Abschätzung formulieren:

Korollar 4.5. *[Ban+16, Corollary 3.7] Sei X wie in der Situation von Satz 3.6. Dann gilt für $1 \leq p < 2$, dass*

$$\mathbb{E}\left[\|X\|\right] \lesssim \sigma + \sigma_* \sqrt{\log \frac{|(b_{ij})|_p}{\sigma_*}}.$$

Dabei ist die universelle Konstante nur von der Wahl von p abhängig.

Die Idee des Beweises des Korollars ist es die Matrix X in einem ersten Schritt nach der Größe ihrer Koeffizienten b_{ij} aufzuspalten und diese dann in einem zweiten und dritten Schritt mit Hilfe von Satz 3.6 abzuschätzen. Im vierten Schritt bringen wir dann wieder beide Abschätzungen zusammen.

Beweis. Fall 1: $\sigma_* = 1$: Schritt 1: Um, wie zuvor beschrieben, die Matrix X in Teilmatrizen aufzuteilen, definieren wir für $k \in \mathbb{N}$ die Matrix $X^{(k)}$ über ihre Einträge:

$$X^{(k)}_{ij} = g_{ij} b_{ij} \mathbf{1}_{2^{-k} < |b_{ij}| \leq 2^{-k+1}}$$

Das heißt, dass die Matrix $X^{(k)}$ dann an den Stellen die Einträge von X hat, für die gilt $2^{-k} < |b_{ij}| \leq 2^{-k+1}$, und sonst 0 ist. Wenn wir nun alle Matrizen $X^{(k)}$ nun wieder zusammenfügen, erhalten wir wieder die Matrix X, somit gilt mit $k_0 \in \mathbb{N}$

$$\mathbb{E}\left[\|X\|\right] = \mathbb{E}\left[\left\| \sum_{k=1}^{\infty} X^{(k)} \right\| \right] \leq \mathbb{E}\left[\left\| \sum_{k=1}^{k_0 - 1} X^{(k)} \right\| \right] + \sum_{k=k_0}^{\infty} \mathbb{E}\left[\|X^{(k)}\| \right].$$

Dabei wenden wir die Dreiecksungleichung und monotone Konvergenz an. Die Konstante k_0 werden wir im Folgenden noch wählen.
Schritt 2: Wir schätzen

$$\sum_{k=k_0}^{\infty} \mathbb{E}\left[\|X^{(k)}\|\right]$$

ab. Dazu betrachten wir die Anzahl der Einträge von $X^{(k)}$, die ungleich 0 sind, welche wir als $c(k)$ schreiben. Mit obiger Überlegung über die Einträge von $X^{(k)}$ können wir schreiben

$$c(k) := |\{ij : 2^{-k} < |b_{ij}| \leq 2^{-k+1}\}|.$$

Mit $|(b_{ij})|_p$, wie in (4.5) definiert, können wir nun abschätzen:

$$2^{-kp}c(k) \leq \sum_{1\leq i,j\leq n} b_{ij}^p \mathbf{1}_{2^{-k}<|b_{ij}|\leq 2^{-k+1}} \leq \sum_{1\leq i,j\leq n} |b_{ij}|^p = |(b_{ij})|_p^p$$

Somit folgt

$$c(k) \leq 2^{kp}|(b_{ij})|_p^p.$$

Da $X^{(k)}$ nun $c(k)$ Einträge hat die nicht 0 sind, müssen diese folglich in einer $c(k) \times c(k)$ Submatrix enthalten sein. Für diese können wir nun σ abschätzen mit

$$\sigma = \max_{1\leq i\leq c(k)} \sqrt{\sum_{j=1}^{c(k)} (b_{ij}^{(k)})^2} \leq \sqrt{c(k)(2^{-k+1})^2} \leq 2^{-k+1}\sqrt{c(k)},$$

sowie $\sigma_* = \max_{ij} |b_{ij}^{(k)}| \leq 2^{-k+1}$, wobei $b_{ij}^{(k)}$ die Einträge von $X^{(k)}$ sind. Mit obigen Abschätzungen und Satz 3.6 folgt nun

$$\begin{aligned}\mathbb{E}\left[\|X^{(k)}\|\right] &\lesssim \sigma + \sigma_*\sqrt{\log c(k)} \lesssim 2^{-k+1}\left(\sqrt{c(k)} + \sqrt{\log c(k)}\right) \\ &\leq 2^{-k+1}\left(\sqrt{2^{kp}|(b_{ij})|_p^p} + \sqrt{\log(2^{kp}|(b_{ij})|_p^p)}\right) \\ &\lesssim 2^{-k(1-p/2)}|(b_{ij})|_p^{p/2}.\end{aligned}$$

Da $p < 2$ ist, fällt die rechte Seite für $k \to \infty$ exponentiell ab. Deshalb existiert ein $k \in \mathbb{N}$ für das gilt

$$2^{-k(1-p/2)}|(b_{ij})|_p^{p/2} \leq 1.$$

Sei nun k_0 die kleinste ganze Zahl für die dies erfüllt ist. Dann können wir wie folgt abschätzen:

$$\sum_{k=k_0}^{\infty} \mathbb{E}\left[\|X^{(k)}\|\right] \lesssim \sum_{k=k_0}^{\infty} 2^{-k(1-p/2)}|(b_{ij})|_p^{p/2} \leq \sum_{k=0}^{\infty} 2^{-k(1-p/2)} \lesssim \sigma$$

Dabei nutzen wir in der ersten Ungleichung die Abschätzung von vorhin, in der zweiten Ungleichung die Wahl von k_0, so dass $2^{-k(1-2/p)}|(b_{ij})|_p^{p/2} \leq 1$ ist, und in der dritten Ungleichung das exponentielle Abfallen der Summanden und somit die Konvergenz der Reihe. Dazu nutzen wir, dass $\sigma \geq \sigma_*$ ist. Die universelle Konstante hängt dabei von p ab.
Schritt 3: Es bleibt noch

$$\mathbb{E}\left[\left\|\sum_{k=1}^{k_0-1} X^{(k)}\right\|\right]$$

abzuschätzen. Wir möchten dafür wieder Satz 3.6 benutzen. Dazu benötigen wir, dass die Matrix $\sum_{k=1}^{k_0-1} X^{(k)}$ höchstens

$$\sum_{k=1}^{k_0-1} c(k) \leq \sum_{k=1}^{k_0-1} 2^{kp} |(b_{ij})|_p^p \lesssim 2^{k_0 p} |(b_{ij})|_p^p \lesssim |(b_{ij})|_p^{p+p^2/(2-p)}$$

Einträge hat. Die Anzahl der Einträge folgt dabei aus der Definition von $c(k)$ und die erste Ungleichung aus der Abschätzung von $c(k)$ aus Schritt 2. Die zweite Ungleichung folgt, da alle $k \leq k_0$ in der Summe sind. Die universelle Konstante ist hier wieder von p abhängig. Die letzte Ungleichung folgt aus der Definition von k_0. Dies wird klar, wenn wir

$$2^{k_0 p} \lesssim |(b_{ij})|_p^{p^2/(2-p)} \tag{4.6}$$

zeigen. Aus der Wahl von k_0 in Schritt 2 wissen wir, dass k_0 die kleinste ganze Zahl ist, die

$$2^{-k_0(1-p/2)} |(b_{ij})|_p^{p/2} \leq 1$$

erfüllt. Da nun aber $p \in [1, 2)$ gilt $k_0 < k_0(1 - p/2)$ und somit ist

$$1 \lesssim 2^{-k_0} |(b_{ij})|_p^{p/2} \leq 2^{-k_0} |(b_{ij})|_p^{p^2/(2-p)},$$

wobei die universelle Konstante von p abhängt. Wenn man nun die linke und rechte Seite mit 2^{k_0} multipliziert erhält man (4.6). Nun können wir Satz 3.6 anwenden und schätzen

$$\mathbb{E}\left[\left\| \sum_{k=1}^{k_0-1} X^{(k)} \right\|\right] \lesssim \sigma + \sigma_* \sqrt{\log |(b_{ij})|_p^{p+p^2/(2-p)}} \lesssim \sigma + \sqrt{\log |(b_{ij})|_p},$$

wobei wir die Eigenschaften des Logarithmus nutzen.
Schritt 4: Wir müsse nun noch die Schritte 2 und 3 wie im ersten Schritt aufgeteilt zusammenfügen:

$$\begin{aligned}
\mathbb{E}\left[\|X\|\right] &\leq \mathbb{E}\left[\left\| \sum_{k=1}^{k_0-1} X^{(k)} \right\|\right] + \sum_{k=k_0}^{\infty} \mathbb{E}\left[\|X^{(k)}\|\right] \\
&\lesssim \sigma + \sqrt{\log |(b_{ij})|_p} + \sigma \\
&\lesssim \sigma + \sqrt{\log |(b_{ij})|_p}
\end{aligned}$$

Fall 2: $\sigma_* \neq 1$. Wir gehen analog zum Beweis von Satz 3.6 vor: Für ein allgemeines $\sigma_* > 0$ gilt, falls $\sigma_* \neq 1$ normieren wir X mit $\frac{1}{\sigma_*}$. Dann gilt für $X = X' \cdot \sigma_*$, dass $\sigma'_* = 1$ von X' ist. Wir rechnen

$$\begin{aligned}
\mathbb{E}\left[\|X\|\right] &= \mathbb{E}\left[\|X'\sigma_*\|\right] = \sigma_* \mathbb{E}\left[\|X'\|\right] \\
&\lesssim \sigma_* \left(\sigma' + \sqrt{\log |(b_{ij})|'_p}\right) = \left(\sigma' \sigma_* + \sigma_* \sqrt{\log |(b_{ij})|'_p}\right) \\
&= \left(\sigma + \sigma_* \sqrt{\log \frac{|(b_{ij})|_p}{\sigma_*}}\right).
\end{aligned}$$

Dabei ist $\sigma' = \max_i \sqrt{\left(\frac{b_{ij}}{\sigma_*}\right)^2}$ und somit ist

$$\sigma_* \cdot \sigma' = \max_i \sqrt{\sigma_*^2 \left(\frac{b_{ij}}{\sigma_*}\right)^2} = \sigma.$$

Weiter ist

$$|(b_{ij})|'_p = \left[\sum_{1 \le i,j \le n} \left| \frac{b_{ij}}{\sigma_*} \right|^p \right]^{\frac{1}{p}} = \frac{\left[\sum_{1 \le i,j \le n} |b_{ij}|^p \right]^{\frac{1}{p}}}{\sigma_*} = \frac{|(b_{ij})|_p}{\sigma_*}.$$

Somit folgt die Behauptung auch für $\sigma_* > 0$.
Für $\sigma_* = 0$ ist die Behauptung trivial. Aufgrund der Definition von σ_* über den Betrag ist $\sigma_* < 0$ nicht möglich. $\square$

Bemerkung 4.6. Aus dem Beweis geht hervor, weshalb wir $p < 2$ wählen müssen. Denn sonst würde $2^{-k(1-p/2)}$ nicht mehr exponentiell abfallen und die zur Abschätzung benötigte Konvergenz, die wir in den Schritte 2 und 3 benutzen, wäre nicht mehr gegeben. Insbesondere wird auch die universelle Konstante, die von p abhängt für $p \to 2$ unendlich groß.

Mit dem Korollar 4.5 haben wir nun eine dimensionfreie Abschätzung. Diese hat zwei große Vorteile gegenüber Satz 3.6. Einerseits ist dies, dass wir nun relativ problemfrei niedrigdimensionale Matrizen in hohe Dimensionen einbetten können, ohne dass unsere Abschätzung durch $\sqrt{\log n}$ schlechter wird. Andererseits ergibt sich so auch die Möglichkeit unendlichdimensionale Räume zu betrachten, da für $n \to \infty$ unsere Abschätzung nicht auch gegen unendlich geht.

Kapitel 5

Beispiele

Im Folgenden werden wir uns noch das Verhalten der Spektralnorm an bestimmten Beispielen ansehen. Dazu werden wir Zufallsmatrizen simulieren und deren empirischen Erwartungswert betrachten, welchen wir mit unseren Abschätzungen vergleichen werden.

5.1 Gaußsche Wigner Matrizen

Zunächst schauen wir uns gaußsche Wigner Matrizen an, wobei die Koeffizienten b_{ij} alle gleich 1 sind. Das heißt wir betrachten symmetrische $n \times n$-Zufallsmatrizen der Form W mit unabhängig, identisch verteilten Einträgen $W_{ij} \sim N(0,1)$. Dazu simulieren wir jeweils $N = 100$ gaußsche Wigner Matrizen der Dimension $n = 1, \dots, 100$ und berechnen deren Spektralnorm. Dann bilden wir für jedes n den empirischen Erwartungswert. Außerdem berechnen wir für jedes n die Abschätzung des Erwartungswerts der Spektralnorm nach Bandeira und van Handel, Satz 3.6, wobei wir $\varepsilon = \frac{1}{2}$ wählen, welches in diesem Beispiel gut geeignet scheint. Da alle $b_{ij} = 1$ sind ergibt sich $\sigma_* = 1$ und $\sigma = \max_i \sqrt{\sum_{j=1}^n b_{ij}^2} = \sqrt{n}$

$$\begin{aligned}\mathbb{E}\left[\|X\|\right] &\leq (1+\varepsilon)\left(2\sigma + \frac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_*\sqrt{\log n}\right)\\ &= 3\left(\sqrt{n} + \frac{3}{\sqrt{\log\frac{3}{2}}}\sqrt{\log n}\right).\end{aligned}$$

Weiter wissen wir aus Kapitel 3, dass wir das asymptotische Verhalten

$$\frac{\|W\|}{\sqrt{n}} \to 2$$

für $n \to \infty$ erwarten. Wir entnehmen dem die Approximation

$$\|W\| \approx 2\sqrt{n}$$

für großes n. Diese Approximation werden wir ebenfalls mit den Ergebnissen vergleichen. Wir simulieren die Zufallsmatrizen in R:

```
# Simulation des Erwartungswerts der Norm
# der Wigner Matrizen mit Dimension n

wigner=function(n,N){
  if(n==1){return(mean(rnorm(N)))} # Fall n=1
  s=c()
  for (k in 1:N) {
    W=matrix(rnorm(n^2),nrow=n,ncol = n)
    # symmterische Wigner Matrizen simulieren
    for (i in 2:n) {
      for (j in 1:(i-1)) {
      W[i,j]=W[j,i]
    }
  }
  s=c(s,norm(W,"2"))
  }
  mean(s) # empirischen Erwartungswert bilden
}

# Durchfuehrung der Simulation bis Dimension n=100
Wsim=c()
for (n in 1:100) {
  Wsim=c(Wsim,wigner(n,100))
}
plot(Wsim, ylim=c(0,70), xlab = "Dimension n",
ylab = "E[||X||]",
main = "Simulation Gaussscher Wigner Matrizen", pch=16)
# Plotten der Ergebnisse

# asymptotisch zu erwartender Grenzwert
curve(sqrt(x)*2,col=3, add = T, lwd=3)

# Abschaetzung nach Bandeira und van Handel
eps=0.5
Ban=function(s,sstar,n,e){
  (1+e)*(2*s+6/(sqrt(log(1+e)))*sstar*sqrt(log(n)))
}

ab=Ban(sqrt(1:100),1,1:100,eps)
points(ab, col=2, pch=16) # zum Plot hinzufuegen
legend("topleft", c("simulierte Norm",
"Normabschaetzung", "asymptotisches Verhalten"),
col = c(1,2,3), pch=c(16,16,16))
```

Dabei haben wir zunächst die Zufallsmatrizen, wie oben beschrieben, simuliert und deren Spektralnorm berechnet, weiter haben wir die Ergebnisse, wie auch den asymptotisch zu erwartenden Grenzwert geplottet, siehe Abbildung 5.1, die Simulationen in schwarz und die Approximation in grün. Des Weiteren haben wir die Normabschätzung, Satz 3.6, durchgeführt und die Ergebnisse in rot dem Plot hinzugefügt. Dabei können wir gut sehen, wie die asymptotisch

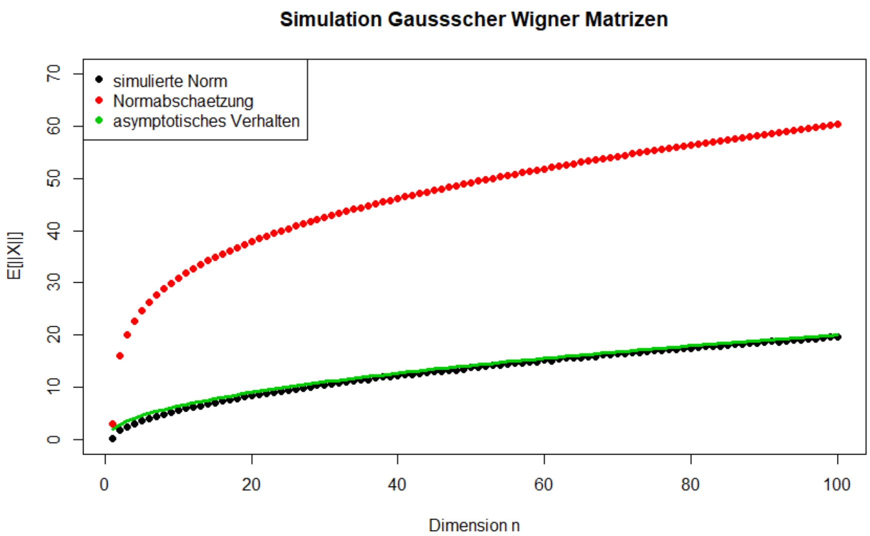

Abbildung 5.1: Simulation des Erwartungswerts der Spektralnorm Gaußscher Wigner Matrizen und Vergleich mit Normabschätzung

Approximation zu den simulierten Ergebnissen passt. Ebenfalls erfüllt die Normabschätzung die erwarteten Eigenschaften, nämlich dass sie die Norm auch wirklich abschätzt und weiter, dass das Wachstum der Abschätzung nur bis auf universelle Konstanten dem der simulierten Norm entspricht. Wenn wir den Quotienten aus der Abschätzung der Spektralnorm und deren Simulation betrachten, fällt uns auf, dass der Faktor bei $n = 100$ zwischen 3 und 4 liegt. Bei einem Vergleich mit der Abschätzung, bei der wir vor $\sqrt{n}$ den Faktor 3 haben, und des asymptotischen Verhaltens von $2\sqrt{n}$ erscheint dieses Verhältnis doch recht hoch. Dieser größerer Faktor lässt sich durch die Komponente, die durch $\sqrt{\log n}$ entsteht, erklären, für $n = 100$ rechnen wir:

$$\frac{3\left(\sqrt{100} + \frac{3}{\sqrt{\log \frac{3}{2}}}\sqrt{\log 100}\right)}{2\sqrt{100}} \approx 3,017.$$

Der zweite Term der Abschätzung fällt also auch bei der Dimension $n = 100$ noch deutlich ins Gewicht. Der Quotient fällt aber mit zunehmender Dimension, wie sich in Abbildung 5.2 erkennen lässt, was man auf das langsamere Wachstum von $\sqrt{\log n}$ gegenüber $\sqrt{n}$ zurückführen kann.
Die Abschätzung bis auf universelle Konstanten möchten wir uns genauer ansehen: Wir betrachten die Skalierung verschiedener Abschätzungen bis auf universelle Konstanten C_i:
Korrekte Skalierung: Wie auch schon in dem Plot zu sehen können wir von dem Wachstum der Spektralnorm der Matrizen von

$$\mathbb{E}\left[\|W\|\right] \approx C_1\sqrt{n},$$

wobei hier $C_1 = 2$ ist.

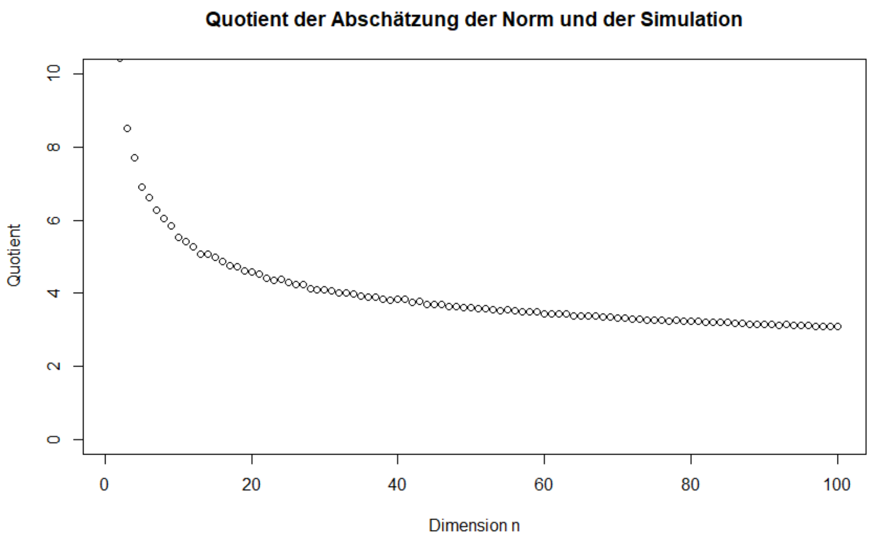

Abbildung 5.2: Quotient der Abschätzung der Spektralnorm und der simulierten Spektralnorm

Satz 3.6: Nach der Abschätzung von Bandeira und van Handel können wir schreiben:

$$\mathbb{E}\left[\|W\|\right] \lesssim \sqrt{n} + \sqrt{\log n}.$$

Es ergibt sich also

$$\mathbb{E}\left[\|W\|\right] \leq C_3\sqrt{n},$$

was bis auf einer universellen Konstanten dem Wachstum der korrekten Skalierung entspricht. Dies können wir uns auch an der unteren Abschätzung, Korollar 4.2, erkennen. Die Voraussetzung von Korollar 4.2 sind hier erfüllt, denn es gibt n^2 große Koeffizienten und somit gilt

$$\mathbb{E}\left[\|W\|\right] \gtrsim \sqrt{n} + \sqrt{\log n}$$

und daher

$$\mathbb{E}\left[\|W\|\right] \geq C_3\sqrt{n}.$$

Die Abschätzung ist also bis auf universelle Konstanten in diesem Fall scharf.
Bemerkung 3.4: Wenn wir die Abschätzung nach Khintchine

$$\mathbb{E}\left[\|W\|\right] \lesssim \sigma\sqrt{\log n},$$

hier betrachten und $\sigma = \sqrt{n}$ einsetzten, erhalten wir

$$\mathbb{E}\left[\|W\|\right] \leq C_4\sqrt{n\log n}.$$

Wir hatten diesen Fall auch schon in Kapitel 3 gesehen. Hier ist die Skalierung also schlechter und die Abschätzung ist nicht mehr scharf.
Bemerkung 3.5: Nach Gordon können wir abschätzen

$$\mathbb{E}\left[\|W\|\right] \lesssim \sigma_*\sqrt{n}.$$

Da hier $\sigma_* = 1$ ist erhalten wir

$$\mathbb{E}\left[\|W\|\right] \leq C_5\sqrt{n},$$

was die korrekte Skalierung ist.
Korollar 4.5: Für $p = 1$ erhalten wir

$$|(b_{ij})|_1 = \sum_{1\leq i,j\leq n} b_{ij} = n^2$$

mit σ und σ_* ergibt sich

$$\mathbb{E}\left[\|W\|\right] \lesssim \sqrt{n} + \sqrt{\log(n^2)} \lesssim \sqrt{n},$$

was der korrekten Skalierung

$$\mathbb{E}\left[\|W\|\right] \leq C_6\sqrt{n}$$

entspricht.
Insgesamt können wir also festhalten, dass Die Abschätzungen nach Bandeira und van Handel, sowie nach Gordon in diesem Besipiel korrekt skalieren, während die aus der nichtkommutativen Khintchine Ungleichung resultierende Abschätzung hier schlechter ist. Die Unterschiede in der Güte der Abschätzungen waren jedoch zu erwarten, denn in Kapitel 3 haben wir uns auch Gedanken dazu gemacht, wann die Abschätzungen nach Khintchine und Gordon der Güte derer von Bandeira und van Handel entsprechen. Dies war zum Einen der Fall $\frac{\sigma}{\sigma_*} \lesssim 1$, wenn die nach Khintchine genauso gut ist, und zum Anderen der Fall $\frac{\sigma}{\sigma_*} \gtrsim \sqrt{n}$, wenn die Abschätzung nach Gordon der Güte derer nach Bandeira und van Handel genügt. Da in diesem Beispiel aber $\frac{\sigma}{\sigma_*} = \frac{\sqrt{n}}{1} = \sqrt{n}$ gilt, liegt genau der zweite Grenzfall vor.
In einem weiteren Beispiel möchten wir uns noch mit dem anderen Grenzfall beschäftigen.

5.2 Schwach besetzte Matrizen

Wir betrachten jetzt den Fall, dass die Matrix X eine Diagonalmatrix ist. Dazu setzen wir die $b_{ii} = 1$ für $1 \leq i \leq n$ und alle $b_{ij} = 0$ für $i \neq j$. Es ergibt sich direkt, dass $\sigma = \sigma_* = 1$ gilt. Analog zu dem Beispiel Gaußscher Wigner Matrizen simulieren wir wieder je $N = 100$ Zufallsmatrizen der Dimension $n = 1, \ldots, 100$ und berechnen den Erwartungswert der Spektralnorm. Wir führen die Simulation wieder mit R aus.

```
# Simulation des Erwartungswerts der Norm normalverteilter
# Diagonalmatrizen mit Dimension n, N ist Anzahl der Matrizen
diagonalGauss=function(n,N){
  if(n==1){return(mean(rnorm(N)))} # Fall n=1
  s=c()
  for (k in 1:N) {
    x=matrix(0,nrow=n,ncol=n)
    diag(x)=rnorm(n)
    s=c(s,norm(x,"2"))
```

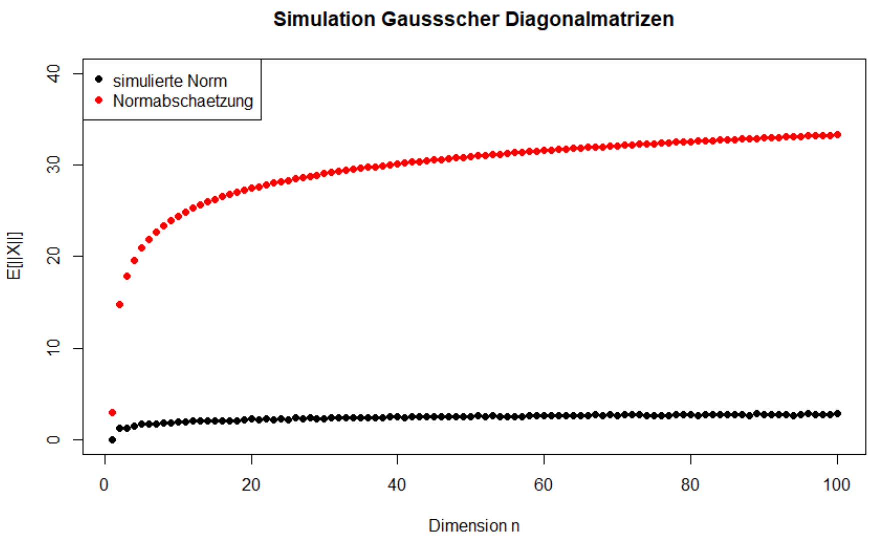

Abbildung 5.3: Simulation des Erwartungswerts der Spektralnorm Gaußscher Diagonalmatrizen und Vergleich mit Normabschätzung

```
  }
  mean(s) # empirischen Erwartungswert bilden
}

Dsim=c()
for (n in 1:100) {
  # Durchfuehrung der Simulation bis Dimension n=100
  Dsim=c(Dsim, diagonalGauss(n,100))
}
# Plotten der Ergebnisse
plot(Dsim, ylim=c(0,40), xlab = "Dimension n", ylab = "E[||X||]",
main = "Simulation Gaussscher Diagonalmatrizen", pch=16)

# Abschaetzung nach Bandeira und van Handel
eps=0.5
Ban=function(s,sstar,n,e){
  (1+e)*(2*s+6/(sqrt(log(1+e)))*sstar*sqrt(log(n)))
}

ab=Ban(1,1,1:100,eps)
points(ab, col=2, pch=16) # zum Plot hinzufuegen
legend("topleft", c("simulierte Norm",
"Normabschaetzung"), col = c(1,2), pch=c(16,16))
```

Außerdem berechnen wir wieder die Abschätzung der Spektralnorm nach Bandeira und van Handel, Satz 3.6, und wählen dafür wieder $\varepsilon = \frac{1}{2}$. Somit ergibt

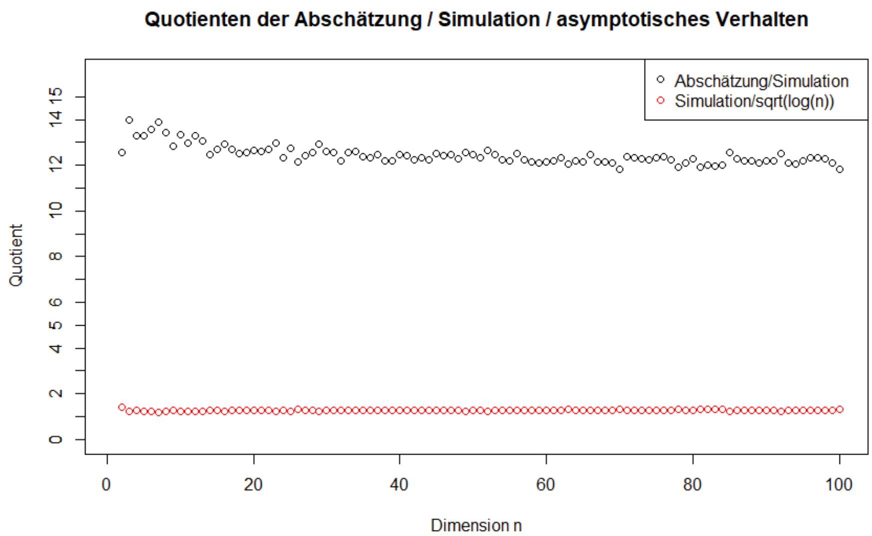

Abbildung 5.4: Quotienten aus Abschätzung der Spektralnorm, deren Simulation und $\sqrt{\log n}$

sich

$$\begin{aligned}\mathbb{E}\left[\|X\|\right] &\leq (1+\varepsilon)\left(2\sigma + \frac{6}{\sqrt{\log{(1+\varepsilon)}}}\sigma_*\sqrt{\log n}\right)\\ &= 3\left(1+\frac{3}{\sqrt{\log\frac{3}{2}}}\sqrt{\log n}\right)\\ &\lesssim \sqrt{\log n}\end{aligned}$$

Das entspricht

$$\mathbb{E}\left[\|X\|\right] \leq C_0\sqrt{\log n}$$

für eine universelle Konstante C_0 und $n \geq 2$. In Abbildung 5.3 erkennen wir, dass trotz eines asymptotisch ähnlichen Wachstums die Abschätzung im Absoluten nicht besonders genau scheint. Den Quotienten aus der Abschätzung und der Simulation sehen wir in Abbildung 5.4. Der relativ große Quotient lässt sich auf den Vorfaktor $\frac{9}{\sqrt{\log 3/2}} \approx 14,134$ zurückführen, während die Spektralnorm wie $\sqrt{\log n}$ mit einem Vorfaktor zwischen 1 und 2 wächst, was ebenfalls aus Abbildung 5.4 ersichtlich wird.
Ob die Abschätzung bis auf universelle Konstanten trotzdem scharf ist, können wir mit Korollar 4.2 prüfen. Da es genau n Einträge $b_{ii} = 1$ gibt, ist die Voraussetzung für Korollar 4.2 erfüllt und wir schätzen nach unten ab:

$$\mathbb{E}\left[\|X\|\right] \gtrsim \sigma + \sigma_*\sqrt{\log n} \gtrsim \sqrt{\log n},$$

falls $n \geq 2$ ist. Somit gilt in diesem Fall

$$\mathbb{E}\left[\|X\|\right] \geq C_1\sqrt{\log n}$$

und die Abschätzung ist bis auf universelle Konstanten scharf. Weiter möchten wir die Abschätzung den anderen Abschätzungen, die wir kennen gelernt haben,

vergleichen:
Bemerkung 3.4: Mit der Abschätzung nach Khintchine,

$$\mathbb{E}[\|W\|] \lesssim \sigma\sqrt{\log n},$$

erhalten wir für $\sigma = 1$

$$\mathbb{E}[\|W\|] \leq C_2\sqrt{\log n},$$

und somit eine bis auf universelle Konstanten scharfe Abschätzung.
Bemerkung 3.5: Nach Gordon können wir abschätzen

$$\mathbb{E}[\|W\|] \lesssim \sigma_*\sqrt{n}.$$

Da hier $\sigma_* = 1$ ist erhalten wir

$$\mathbb{E}[\|W\|] \leq C_3\sqrt{n}$$

für eine universelle Konstante C_3. Das entspricht jedoch keiner scharfen Abschätzung.
Korollar 4.5: Für $p = 1$ erhalten wir mit

$$|(b_{ij})|_1 = \sum_{1\leq i,j\leq n} b_{ij} = n$$

und $\sigma = \sigma_* = 1$ für $n \geq 2$

$$\mathbb{E}[\|W\|] \lesssim 1 + \sqrt{\log n} \lesssim \sqrt{\log n},$$

was der korrekten Skalierung

$$\mathbb{E}[\|W\|] \leq C_4\sqrt{\log n}$$

für eine universelle Konstante C_4 entspricht.
Wenn wir die Abschätzungen vergleichen stellen wir fest, dass diese bis auf jene nach Gordon, Bemerkung 3.5, bis auf universelle Konstanten scharf sind. Im Fall von Bemerkung 3.5 ist das Wachstum der Abschätzung mit $\sqrt{n}$ im Vergleich zu $\sqrt{\log n}$ deutlich größer. Analog zu dem Beispiel 5.1 erkennen wir wieder, dass hier ein Grenzfall vorliegt. Denn es gilt $\frac{\sigma}{\sigma_*} = 1$ und somit liegt der minimale mögliche Fall $\frac{\sigma}{\sigma_*} \lesssim 1$. Wir hatten uns bereits überlegt, dass in diesem Fall die Abschätzung aus Bemerkung 3.4 bis auf universelle Konstanten genauso gut ist, wie die nach Bandeira und van Handel. Diese Abschätzung hatte jedoch in dem anderen extremen Fall im Beispiel 5.1 schlechter abgeschnitten. Die Abschätzung aus Bemerkung 3.5 hatte hingegen in dem maximalen Fall im Beispiel 5.1 gut abgeschnitten, verfehlt aber in dem hier vorliegenden minimalen Fall das korrekte Wachstum deutlich.
Bemerkenswert ist zudem, dass auch unsere dimensionsfreie Abschätzung, Korollar 4.2, in beiden Beispielen eine bis auf universelle Konstanten scharfe Abschätzung liefert. Da sie aber, im Gegensatz zu Satz 3.6, auch für niedrigdimensionale Matrizen, die in höhere Dimensionen eingebettet sind, sowie für bestimmte unendlichdimensionale Matrizen geeignet ist, scheint sie eine gute alternative Abschätzung zu bieten.

Literaturverzeichnis

[Ban+16] *A. S. Bandeira* und *R. van Handel*, Sharp nonasymptotic bounds on the norm of random matrices with independent entries. Ann. Probab. 44, No. 4, 2479–2506 (2016; Zbl 1372.60004)

[Jän08] *K. Jänich*, Lineare Algebra. Berlin: Springer (2008; Zbl 1136.15001)

[Bib+19] *M. Bibinger* und *H. Holzmann*, Wahrscheinlichkeitstheorie. Skript zur Vorlesung, 13. März 2019

[Bou+12] *S. Boucheron*, *G. Lugosi* und *P. Massart*, Concentration inequalities. A nonasymptotic theory of independence. Oxford: Oxford University Press (2013; Zbl 1279.60005)

[Ver20] *R. Vershynin*, High-dimensional probability. An introduction with applications in data science. Cambridge: Cambridge University Press (2018; Zbl 1430.60005)

[Hol20] *H. Holzmann*, Wahrscheinlichkeitstheorie. Handouts zur Vorlesung, 2020

[Rol18] *S. Rollenske*, Lineare Algebra I und Grundlagen der Mathematik. Skript zur Vorlesung, 30. Januar 2018

[Bai+88] *Z. D. Bai* und *Y. Q. Yin*, Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. Ann. Probab. 16, No. 4, 1729–1741 (1988; Zbl 0677.60038)

[Pis03] *G. Pisier*, Introduction to operator space theory. London Mathematical Society Lecture Note Series, Volume 294, Cambridge University Press, Cambridge, 2003

[Ora+15] *F. Orabona* und *D. Pal*, Optimal Non-Asymptotic Lower Bound on the Minimax Regret of Learning with Expert Advice, arXiv: 1511.02176, 2015

[Dav+01] *K. R. Davidson* und *S. J. Szarek* Local operator theory, random matrices and Banach spaces. In Handbook of the Geometry of Banach Spaces, Vol. I 317–366. North- Holland, Amsterdam. (2001)

[Ver12] *R. Vershynin* Introduction to the non-asymptotic analysis of random matrices. In Compressed Sensing 210–268. Cambridge Univ. Press, Cambridge. (2012)

[Gem+80] *S. Geman* A limit theorem for the norm of random matrices. Ann. Probab. 8 252–261.(1980)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig, ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Marburg, 18. August 2020

Ort, Datum	Unterschrift