

MATHEMATISCHES INSTITUT
UNIVERSITÄT HEIDELBERG

BACHELORARBEIT

ONLINE-SCHÄTZUNG DES
GEOMETRISCHEN MEDIANS IN
HILBERTRÄUMEN

Daniel Fridljand

betreut von
Prof. Dr. Jan JOHANNES

20. Oktober 2020

Abstract

In this thesis we consider a new algorithm estimating the geometric median. The geometric median is one possible generalization of the real median to Hilbert spaces. This new estimator is based on the stochastic gradient method and is calculated recursively. In comparison to the standard algorithm for this problem so far, this new approach is more efficient and faster, especially in case of high dimensional data. Furthermore the estimator can be simply updated when the data arrives sequentially. One most of the most important results of this thesis is the almost sure consistency of the estimator. We derive a rate of convergence in $\mathcal{L}_2$, show that our estimator is asymptotically normal and determine the exact asymptotic distribution. Lastly, we illustrate the asymptotic normality with a simulation.

Zusammenfassung

Gegenstand der vorliegenden Arbeit ist ein neuer Algorithmus zur Schätzung des geometrischen Medians. Der geometrische Median ist eine mögliche Verallgemeinerung des gewöhnlichen eindimensionalen Medians auf Hilberträumen. Dieser neue Schätzer basiert auf der stochastischen Gradientenmethode und wird rekursiv berechnet. Er zeichnet sich gegenüber dem bisherigen Standard-Algorithmus dadurch aus, dass er sich insbesondere bei Daten mit hoher Dimension viel schneller berechnen lässt und gut mit sequentiell eintreffenden Daten umgehen kann. Zu den wichtigsten gezeigten Resultaten dieser Arbeit gehört, dass der Schätzer $\mathbb{P}$ -fast sicher gegen den Median konvergiert. Wir leiten eine Konvergenzrate für die Konvergenz in $\mathcal{L}_2$ her. Wir zeigen, dass der Schätzer asymptotische Normalität aufweist und bestimmen die genaue Verteilung. Schließlich illustrieren wir die asymptotische Normalität mit einer Simulation.

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Bachelorarbeit unterstützt haben.

Zuerst gebührt mein Dank Prof. Jan Johannes, der meine Bachelorarbeit betreut und begutachtet hat. Für die hilfreichen Anregungen und die konstruktive Kritik bei der Erstellung dieser Arbeit möchte ich mich bedanken.

Ich bedanke mich bei der Gerhard C. Starck Stiftung und dem Stiftungsbüro, Frau Inge Frey. Die Gerhard C. Starck Stiftung hat mir mein bisheriges Studium an der Universität Heidelberg und der Hebräischen Universität in Jerusalem ermöglicht. Ohne die finanzielle Unterstützung wäre jenes in dieser Form nicht möglich gewesen. Bei Frau Frey möchte ich mich insbesondere für ihre Geduld und Nachsicht bedanken.

Inhaltsverzeichnis

1	Einleitung	9
2	Einführung	11
2.1	Funktionalanalysis	11
2.2	Definition geometrischer Median	15
2.3	Anwendungen	16
3	Geometrischer Median Eigenschaften	20
3.1	Annahmen	20
3.2	Ableitungen	21
3.3	Konvexität	29
3.4	Grundeigenschaften	34
4	Martingaldifferenz	38
4.1	Konvergenzarten	38
4.2	Bedingter Erwartungswert	39
4.3	Martingaldifferenz	46
4.4	Sätze	46
5	Algorithmus	49
5.1	Definition	49
5.2	$\mathbb{P}$ -fast sichere Konvergenz	55
5.3	$\mathcal{L}_2$ Konvergenzrate	58
5.4	Asymptotische Normalität	72
5.5	Nicht-asymptotische Konfidenzbereiche	78
6	Simulation	79

Kapitel 1

Einleitung

Ein klassisches Problem der Statistik ist es aus Beobachtungen eines Experimentes in Form von Messwerten Rückschlüsse über das Experiment selbst zu ziehen. Dadurch kann man z.B. einen Trend feststellen oder Prognosen über zukünftige Messwerte anstellen. Typischerweise geschieht das darüber, dass man einen Parameter des Experimentes schätzt, was bedeutet, dass man aus den Beobachtungen einen Parameter berechnet, der in einem gewissen Sinne nah an dem wahren Parameter liegt. Der wahre Parameter bleibt in einem realen Szenario immer unbekannt.

Der Median ist ein Beispiel für einen Parameter, an dem man häufig interessiert ist. Für Beobachtungen im eindimensionalen Fall, bei denen ein Messwert einfach eine reelle Zahl ist, ist der Median definiert als eine Zahl m , so dass die Wahrscheinlichkeit für eine Beobachtung, kleiner oder größer als m zu sein, gleich ist. Der geometrische Median ist eine mögliche Verallgemeinerung dieses Konzepts auf mehrdimensionale Räume (vgl. Gleichung 2.2). Der geometrische Median hat einige hilfreiche Eigenschaften. Zum Beispiel reagiert er tolerant auf Messfehler ohne zu sehr verfälscht zu werden (vgl. Unterabschnitt 3.4.2). Er hat viele Anwendung in Forschung und Praxis und kann auf sehr allgemein Räumen definiert werden (vgl. Abschnitt 2.3).

Lange Zeit war der verbreitetste Algorithmus, um aus Messwerten den geometrischen Median zu berechnen, eine Abwandlung vom Weiszfeld Algorithmus. Wir diskutieren in dieser Arbeit einen Algorithmus, der in [CCZ13] vorgestellt wurde und auf der stochastischen Gradienten-Methode basiert.

Dieser Algorithmus hat einige Vorzüge gegenüber dem von Weiszfeld. Zunächst schneidet er in Laufzeittests deutlich besser ab. Darüber hinaus kann er Daten online verarbeiten. Das bedeutet, dass er sequentiell eintreffende Daten aufnehmen kann, ohne die Berechnung ganz von vorne zu beginnen, sondern das Ergebnis aus der vorherigen Berechnung benutzen kann.

Das Hauptziel dieser Arbeit ist es, die Beweise aus unserer Hauptquelle [CCZ13] detailliert aufzuarbeiten. [CCZ13] betrachtet den neuen Algorithmus und zeigt die Konvergenz gegen den geometrischen Median. An einigen Stellen werden Beweise aus [CCG17] anstelle von [CCZ13] verwendet, da dort häufig stärkere Aussagen gezeigt werden. [CCG17] kann dabei als Fortsetzung bzw. Weiterentwicklung von [CCZ13] verstanden werden.

Zum Aufbau dieser Arbeit: Zunächst führen wir in Kapitel 2 den geometrischen Me-

dian ein und motivieren ihn. In Kapitel 3 beweisen wir einige nützliche Eigenschaften im Zusammenhang mit dem geometrischen Median, wie die Ableitungen und die Konvexität einer bestimmten Funktion. Auch wird gezeigt, dass er unter bestimmten Annahmen eindeutig existiert und robust ist. Dann geben wir in Kapitel 4 einen Überblick über den mehrdimensionalen bedingten Erwartungswert und benennen für uns wichtige Sätze. In Kapitel 5 wenden wir uns schließlich dem eigentlichen Algorithmus zu und beweise die Konvergenz. Schließlich geben in Kapitel 6 eine Simulation zur Illustration unserer Resultate.

Kapitel 2

Einführung

In diesem Kapitel führen wir den geometrischen Median ein noch ohne auf den entsprechenden Algorithmus zu sprechen zu kommen. Wir motivieren die Beschäftigung mit dem geometrischen Median und die Relevanz einer Berechnungsmethode.

2.1 Funktionalanalysis

In diesem Abschnitt führen wir notwendige Werkzeuge der Funktionalanalysis ein.

Definition 2.1.1. Ein **Banachraum** ist ein normierter Vektorraum, welcher vollständig unter der von der Norm induzierten Metrik ist.

Definition 2.1.2. Ein **komplexer/reeller Skalarraum** ist ein Vektorraum V über $\mathbb{R}/\mathbb{C}$ zusammen mit einer Abbildung $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}/\mathbb{C}$, welche als **Skalarprodukt** bezeichnet wird, sodass folgende Eigenschaften gelten:

1. Sie ist linear in der zweiten Variable: $\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle$
2. Sie ist streng positiv: für $x \neq 0$, $\langle x, x \rangle \in \mathbb{R}_+$.
3. Sie ist symmetrisch: $\langle x, y \rangle = \overline{\langle y, x \rangle}$

Bemerkung 2.1.3. Wegen der Linearität gilt $\langle x, 0 \rangle = 0$ für alle $x \in V$, insbesondere $\langle 0, 0 \rangle = 0$. Durch Symmetrie des Skalarprodukts ist antilinear in der ersten Variable, d.h. $\langle \alpha x + \beta y, z \rangle = \overline{\alpha} \langle x, z \rangle + \overline{\beta} \langle y, z \rangle$ (also ist das Skalarprodukt **sesquilinear**, d.h. linear in der einen und antilinear in der anderen Variable).

Definition 2.1.4. Wir setzen $\|x\| := \sqrt{\langle x, x \rangle}$ und nennen es **Norm** von x .

Definition 2.1.5. Ein **Hilbertraum** ist ein Skalarraum, der in der induzierten Norm ein Banachraum ist. Wir bezeichnen mit $\mathcal{H}$ immer einen Hilbertraum.

Definition 2.1.6. Sei (X, τ) ein topologischer Raum. (X, τ) heißt **separabel**, falls er eine abzählbare, dichte Teilmenge besitzt. Da eine Norm eine Topologie induziert, wird insbesondere ein Hilbertraum $\mathcal{H}$ separabel genannt, falls es eine dichte Teilmenge besitzt.

Beispiel 2.1.7. Das häufigste Beispiel ist $\mathcal{H} = \mathbb{R}^d$ mit dem Skalarprodukt $\langle (\alpha_1, \dots, \alpha_d), (\beta_1, \dots, \beta_d) \rangle = \sum_{j=1}^d \alpha_j \beta_j$. Die Vollständigkeit von $\mathcal{H}$ folgt direkt aus der Vollständigkeit von $\mathbb{R}$ und der endlichen Dimension. Offensichtlich ist $\mathbb{Q}^d$ eine dichte Menge

und da diese abzählbar ist, ist $\mathcal{H}$ separabel. Folglich ist $\mathcal{H}$ in der Tat ein separabler Hilbertraum.

Ein Beispiel dafür aus dem realen Leben, sind die Fernseherzuschauer der französischen Firma Médiamétrie. Das Beispiel befasst sich mit einer Stichprobe mit $n = 5423$ Datenpunkten bzw. Vektoren in $\mathbb{R}^d$ mit $d = 86400$, welche das Verhalten der einzelnen Zuschauer wiedergeben, gemessen jede Sekunde über einen Zeitraum von 24h. Man bemerke, dass $\mathcal{H}$ zwar endlichdimensional ist, d aber sehr groß ist. Siehe 2.3.1 für mehr Details.

Beispiel 2.1.8. Dann ist $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^d)$ ein separabler Hilbertraum mit dem Skalarprodukt $\langle f, g \rangle = \int_{\mathbb{R}^d} f(x)g(x) d\lambda^d(x)$. Die Vollständigkeit folgt aus dem Riesz-Fischer-Theorem [vgl. Sim15].

Lemma 2.1.9. $\mathcal{H}$ ist separabel.

Beweis. Wir benutzen den Beweis aus [Bre10], Theorem 4.13. Sei $\mathcal{R}$ die abzählbare Familie von Mengen in $\mathbb{R}^d$ der Form $R = \prod_{k=1}^d (a_k, b_k)$ mit $a_k, b_k \in \mathbb{Q}$. Sei $\mathcal{E}$ der über $\mathbb{Q}$ von den Funktionen $\{\mathbb{1}_R\}_{R \in \mathcal{R}}$ erzeugte Vektorraum, d.h. $\mathcal{E}$ besteht aus endlichen Linearkombinationen der $\mathbb{1}_R$ mit rationalen Koeffizienten, s.d. $\mathcal{E}$ abzählbar ist. Wir behaupten, dass $\mathcal{E}$ dicht in $\mathcal{H}$ ist. Sei $f \in \mathcal{H}$ gegeben und $\varepsilon > 0$. Dann existiert ein $f_1 \in C_c(\mathbb{R}^d)$, s.d. $\|f - f_1\|_2 < \varepsilon$. Sei $R \in \mathcal{R}$ ein beliebiger Würfel, der $\text{supp}(f_1)$ enthält (der Träger von f_1). Für $\delta > 0$ ist es leicht, eine Funktion $f_2 \in \mathcal{E}$ zu konstruieren, sodass $\|f_1 - f_2\|_\infty < \delta$ und f_2 außerhalb R verschwindet. Es genügt, R in kleine Würfel aus $\mathcal{R}$ aufzuteilen, in dem die Oszillation (d.h. $\sup - \inf$) von f_1 kleiner ist als δ . Daraus ergibt sich, dass $\|f_1 - f_2\|_2 \leq \|f_1 - f_2\|_\infty |R|^{\frac{1}{2}} < \delta |R|^{\frac{1}{2}}$. Wir folgern, dass $\|f - f_2\|_2 < 2\varepsilon$, wenn $\delta > 0$ so gewählt ist, dass $\delta |R|^{\frac{1}{2}} < \varepsilon$ $\square$

Auch dieser Hilbertraum ist von Relevanz. Heutzutage kommen häufig Daten vor, die als Zufallsvariablen mit Wertebereich in $\mathcal{H}$ modelliert werden können. Die EKG Kurve eines Patienten über einen gewissen Zeitraum, Spektroskopieanalyse von bestimmten Wellenlängen und auch die jährlichen Temperaturen an verschiedenen Orten sind Beispiele für solche Daten. Daten in $\mathcal{H}$ werden in functional data analysis studiert. Siehe 2.3.3 für mehr Beispiele.

Definition 2.1.10. Sei I eine Indexmenge und $\{e_i\}_{i \in I}$ ein Orthonormalsystem. Dann heißt $\{e_i\}_{i \in I}$ **Orthonormalbasis**, falls für alle $h \in \mathcal{H}$ aus $\forall i \in I : \langle h, e_i \rangle = 0$ folgt, dass $h = 0$.

Satz 2.1.11. Jeder Hilbertraum besitzt eine Orthonormalbasis.

Beweis. Nach dem Lemma von Zorn gibt es ein maximales Orthonormalsystem $\{e_i\}_{i \in I}$ in $\mathcal{H}$. Sei $h \in \mathcal{H}$ mit $\forall i \in I : \langle h, e_i \rangle = 0$. Wenn $h \neq 0$ wäre, wäre $\{e_i\}_{i \in I} \cup \{\frac{h}{\|h\|}\}$ ein Orthonormalsystem, was der Maximalität von $\{e_i\}_{i \in I}$ widerspräche. Also ist $\{e_i\}_{i \in I}$ schon eine Orthonormalbasis. $\square$

Bemerkung 2.1.12. Ein separabler Hilbertraum hat offensichtlich eine abzählbare Orthonormalbasis. Wir werden also im folgenden davon ausgehen, dass I abzählbar ist.

Lemma 2.1.13. Sei $\{e_i\}_{i \in I}$ eine Orthonormalbasis und $h \in \mathcal{H}$. Dann gilt

$$h = \sum_{i \in I} \langle h, e_i \rangle e_i$$

Beweis. Sei $y := h - \sum_{i \in I} \langle h, e_i \rangle e_i$. Dann gilt für $j \in I$

$$\langle y, e_j \rangle = \langle h, e_j \rangle - \sum_{i \in I} \langle h, e_i \rangle \langle e_i, e_j \rangle = \langle h, e_j \rangle - \langle h, e_j \rangle = 0$$

Und nach Definition der Orthonormalbasis somit $y = 0$. $\square$

Lemma 2.1.14. Sei $\{e_i\}_{i \in \mathbb{N}}$ eine Orthonormalbasis für $\mathcal{H}$ und seien $h, h' \in \mathcal{H}$ Vektoren. Dann gilt

$$\langle h, h' \rangle = \sum_{i \in I} \langle h, e_i \rangle \langle e_i, h' \rangle.$$

Beweis. Für $h = 0$ oder ein endliches I ist die Aussage klar. Sei also $h \neq 0$ und $I = \mathbb{N}$. Da nach Lemma 2.1.13 $h = \sum_{i \in I} \langle h, e_i \rangle e_i$ gilt, gibt es eine abzählbare Folge $\{i_k\}_{k \in \mathbb{N}} \subset \mathbb{N}$, sodass immer $\langle e_{i_k}, h \rangle \neq 0$ gilt und für andere Indizes $\langle e_i, h \rangle = 0 = \langle e_i, h' \rangle$. Dies ist möglich, da die Reihe sonst nicht konvergieren würde.

Sei $\varepsilon > 0$ vorgegeben. Nach Lemma 2.1.13 gibt es ein N , sodass für alle $n > N$ gilt: $\|h - \sum_{k=1}^n \langle e_{i_k}, h \rangle e_{i_k}\| \leq \sqrt{\varepsilon}$ und $\|h' - \sum_{k=1}^n \langle e_{i_k}, h' \rangle e_{i_k}\| \leq \sqrt{\varepsilon}$. Betrachte $|\langle h, h' \rangle - \sum_{k=1}^n \langle h, e_{i_k} \rangle \langle e_{i_k}, h' \rangle|$. Dann erhalten wir

$$\begin{aligned} \left| \langle h, h' \rangle - \sum_{k=1}^n \langle h, e_{i_k} \rangle \langle e_{i_k}, h' \rangle \right| &= \left| \left\langle h - \sum_{k=1}^n \langle e_{i_k}, h \rangle e_{i_k}, h' - \sum_{k=1}^n \langle e_{i_k}, h' \rangle e_{i_k} \right\rangle \right| \\ &\leq \left\| h - \sum_{k=1}^n \langle e_{i_k}, h \rangle e_{i_k} \right\| \left\| h' - \sum_{k=1}^n \langle e_{i_k}, h' \rangle e_{i_k} \right\| \leq \varepsilon, \end{aligned}$$

Wobei die Ungleichung Cauchy-Schwarz ist. Daraus folgt die Aussage. $\square$

Korollar 2.1.15. Sei $\{e_i\}_{i \in \mathbb{N}}$ eine Orthonormalbasis für $\mathcal{H}$ und seien $h \in \mathcal{H}$ Vektoren. Dann gilt

$$\|h\|^2 = \sum_{i \in I} |\langle h, e_i \rangle|^2$$

Beweis. Folgt direkt aus 2.1.14. $\square$

Definition 2.1.16. Sei $\mathcal{H}$ ein Hilbertraum. Die **schwache Topologie** auf $\mathcal{H}$ ist die schwächste Topologie auf $\mathcal{H}$, sodass, für alle $h \in \mathcal{H}$, die Abbildung $\langle h, \cdot \rangle : \mathcal{H} \rightarrow \mathbb{R}$ stetig unter dieser Topologie ist.

Sei $\{y_n\}_{n \in \mathbb{N}}$ eine Folge in $\mathcal{H}$ und $y \in \mathcal{H}$. Wir sagen y_n **konvergiert schwach** gegen y , falls y_n in der schwachen Topologie gegen y konvergiert, und wir schreiben $y_n \xrightarrow{w} y$.

Bemerkung 2.1.17. Aus der Definition folgt direkt, dass $y_n \xrightarrow{w} y$ genau dann, wenn für alle $h \in \mathcal{H}$ $\langle y_n, h \rangle \rightarrow \langle y, h \rangle$ gilt.

Lemma 2.1.18. Für $y \in \mathcal{H}$ gilt $\|y\| = \sup\{\langle y, u \rangle : u \in \mathcal{H}, \|u\| = 1\} =: \|\langle y, \cdot \rangle\|_L$.

Beweis. „ $\leq$ “:

$$\|y\|^2 = \langle y, y \rangle = \|y\| \cdot \left\langle y, \frac{y}{\|y\|} \right\rangle \leq \|y\| \cdot \|\langle y, \cdot \rangle\|_L \Rightarrow \|y\| \leq \|\langle y, \cdot \rangle\|_L$$

„ $\geq$ “: Die Cauchy-Schwarz-Ungleichung gibt uns für $u \in \mathcal{H}$ mit $\|u\| = 1$:

$$\langle y, u \rangle \leq \|y\| \cdot \|u\| = \|y\| \Rightarrow \|\langle y, \cdot \rangle\|_L \leq \|y\|$$

□

Lemma 2.1.19. *Die Normabbildung ist unterhalbstetig unter der schwachen Topologie. Das heißt: Ist $\{y_n\}_{n \in \mathbb{N}}$ eine Folge in $\mathcal{H}$, die schwach gegen $y \in \mathcal{H}$ konvergiert, dann gilt $\liminf_{n \rightarrow \infty} \|y_n\| \geq \|y\|$.*

Beweis. Nach Cauchy-Schwarz haben wir für $u \in \mathcal{H}$ $|\langle y_n, u \rangle| \leq \|y_n\| \|u\|$. Wegen Bemerkung 2.1.17 haben wir, wenn wir auf beiden Seiten den $\liminf$ nehmen, $|\langle y, u \rangle| = \liminf_{n \rightarrow \infty} |\langle y_n, u \rangle| \leq \|u\| \liminf_{n \rightarrow \infty} \|y_n\|$. Da u beliebig war erhalten wir $\|y\| = \|\langle y, \cdot \rangle\|_L \leq \liminf_{n \rightarrow \infty} \|y_n\|$. □

Satz 2.1.20. *Sei $\mathcal{H}$ ein separabler Hilbertraum und $\mathcal{B}_1$ der Einheitsball in $\mathcal{H}$. Dann ist die Beschränkung der schwachen Topologie von $\mathcal{H}$ auf $\mathcal{B}_1$ metrisierbar.*

Beweis. Sei $\{\phi_n\}_{n \in \mathbb{N}}$ eine dichte Teilmenge von $\mathcal{B}_1$. Definiere für $x, y \in \mathcal{H}$

$$\rho(x, y) := \sum_{j=1}^{\infty} \min \{2^{-j}, |\langle \phi_j, x - y \rangle|\}. \quad (2.1)$$

Dann ist es leicht zu zeigen, dass ρ tatsächlich eine Metrik ist, die die schwache Topologie induziert. □

Satz 2.1.21. *Sei $\mathcal{H}$ ein separabler Hilbertraum und $\mathcal{B}_1$ der Einheitsball in $\mathcal{H}$. Dann ist $\mathcal{B}_1$ kompakt unter der schwachen Topologie.*

Beweis. Für einen endlichdimensionalen Vektorraum ist die Behauptung trivial. Wir gehen deshalb im Folgenden von einem abzählbar unendlichdimensionalen Vektorraum aus. Sei $\{e_n\}_{n \in \mathbb{N}}$ eine Orthonormalbasis. Wegen Satz 2.1.20 genügt es, Folgenkompaktheit zu zeigen. Sei also $\{\eta_m\}_{m \in \mathbb{N}}$ eine Folge in $\mathcal{B}_1$. Wir müssen zeigen, dass es ein $\eta_\infty \in \mathcal{B}_1$ gibt, sodass eine Teilfolge von $\{\eta_m\}_{m \in \mathbb{N}}$ schwach dagegen konvergiert. Setze $a_j^{(m)} := \langle e_j, \eta_m \rangle$. Nach Cauchy-Schwarz muss $a_j \in [-1, 1]$.

Wegen der Kompaktheit von $[-1, 1]$ muss für alle $j \in \mathbb{N}$ die Folge $\{a_j^{(m)}\}_{m \in \mathbb{N}}$ eine konvergente Teilfolge haben. Nach Diagonalisierung können wir annehmen, dass für alle $j \in \mathbb{N}$ die Teilfolge $\{a_j^{(m_l)}\}_{l \in \mathbb{N}}$ gegen ein $a_j^{(\infty)} \in [-1, 1]$ konvergiert.

Für ein beliebiges $J \in \mathbb{N}$ gilt, dass

$$\sum_{j=1}^J \left| a_j^{(\infty)} \right|^2 = \sum_{j=1}^J \left| \lim_{l \rightarrow \infty} a_j^{(m_l)} \right|^2 = \lim_{l \rightarrow \infty} \sum_{j=1}^J \langle \varphi_j, \eta_{m_l} \rangle^2 \leq \limsup_{l \rightarrow \infty} \|\eta_{m_l}\|^2 \leq 1,$$

wobei wir bei der vorletzten Ungleichung Korollar 2.1.15 benutzen. Deshalb ist auch $\sum_{j=1}^{\infty} \left| a_j^{(\infty)} \right|^2 \leq 1$. Wieder nach Korollar 2.1.15 ist folglich $\eta_\infty := \sum_{j=1}^{\infty} a_j^{(\infty)} e_j$ wohldefiniert und in $\mathcal{B}_1$. Desweiteren gilt für alle $j \in \mathbb{N}$, dass $\langle e_j, \eta_{m_l} \rangle \rightarrow \langle e_j, \eta_\infty \rangle$. Dann gilt

aber für die Metrik aus Gleichung 2.1, dass $\rho(\eta_{m_l}, \eta_\infty) \rightarrow 0$. Nach Wahl der Metrik ρ folgt daraus $\eta_{m_l} \xrightarrow{w} \eta_\infty$. $\square$

Korollar 2.1.22. *Sei $\mathcal{H}$ ein separabler Hilbertraum. Für jedes $r \in \mathbb{R}^+$ ist der Ball $\mathcal{B}_r$ vom Radius r kompakt und metrisierbar unter der schwachen Topologie.*

Beweis. Nach Satz 2.1.20 und 2.1.21 ist der Einheitsball $\mathcal{B}_1$ kompakt und metrisierbar. Da die Abbildung $x \mapsto rx$ ein Homeomorphismus unter der schwachen Topologie ist, übertragen sich diese topologischen Invarianten auf $\mathcal{B}_r$. $\square$

Definition 2.1.23. Sei $A : \mathcal{H} \rightarrow \mathcal{H}$ ein beschränkter Operator. Dann heißt A **kompakter Operator**, falls für alle kompakten Teilmengen $K \subset \mathcal{H}$ gilt, dass $cls(A(K))$ auch wieder kompakt ist. Dabei bezeichnen wir mit $cls(\cdot)$ den Abschluss einer Menge.

Definition 2.1.24. Sei $A : \mathcal{H} \rightarrow \mathcal{H}$ ein beschränkter Operator. Dann heißt A **selbstadjungiert**, falls für alle $x, y \in \mathcal{H}$ gilt, dass $\langle Ax, y \rangle = \langle x, Ay \rangle$.

Lemma 2.1.25. *Sei $A : \mathcal{H} \rightarrow \mathcal{H}$ ein positiv semi-definiter, selbstadjungierter Operator. Sei eine Orthonormalbasis $\{e_i\}_{i \in I} \subset \mathcal{H}$ mit $\sum_{i \in I} \langle A(e_i), e_i \rangle < \infty$ gegeben. Dann ist A ein kompakter Operator.*

Beweis. Für den Beweis siehe z.B. Proposition 280 in [Gos11]. Er wird hier nicht ausgeführt, weil es den Rahmen dieser Arbeit überschreiten würde. $\square$

Satz 2.1.26. (Spektralsatz) *Sei $A : \mathcal{H} \rightarrow \mathcal{H}$ ein kompakter, selbstadjungierter Operator. Dann gibt es eine Orthonormalbasis $\{v_j\}_{j \in I} \subset \mathcal{H}$ und $\{\lambda_i\}_{i \in I} \subset \mathbb{R}$, sodass für alle $i \in \mathbb{N}$ $A(v_i) = \lambda_i v_i$ gilt. Falls $I = \mathbb{N}$, gilt $\lambda_i \rightarrow 0$.*

2.2 Definition geometrischer Median

In diesem Abschnitt führen wir den geometrischen Median mit Bezug auf die historische Motivierung ein. Weiterhin beweisen wir, dass er im reellen Fall mit klassischen Median übereinstimmt.

Der Median einer eindimensionalen Zufallsvariable X ist definiert als jede reelle Zahl m , welche $\mathbb{P}(X < m) \leq \frac{1}{2}$ und $\mathbb{P}(X > m) \leq \frac{1}{2}$ genügt. Eine natürliche Frage, welche schon am Anfang des 20. Jahrhunderts auftrat, ist, wie man diese Definition auf allgemeinere Räume, z.B. mehrdimensionale Vektorräume, erweitern kann. Es gibt einige Grundeigenschaften, die eine solche Definition erfüllen sollte. Erstens sollte sie offensichtlich im eindimensionalen Raum mit der Klassischen übereinstimmen. Darüber hinaus sollte bei einer symmetrischen Verteilung der Symmetriepunkt der Median sein (vgl.[Sma90]).

Eine mögliche Lösung hat Weber in [Web09] im Jahr 1909 vorgeschlagen. Er hat das idealisierte Problem betrachtet, den optimalen Standort für eine Firma auszusuchen, der minimalen euklidischen Abstand zu allen Kunden der Firma hat:

Definition 2.2.1. Seien $x_1, \dots, x_n \in \mathbb{R}^d$ Punkte. Dann ist m der geometrischer Median, falls

$$\sum_{i=1}^n \|x_i - m\|.$$

[RH13] erwähnt in 6.2 eine anschauliche geometrische Interpretation dieses Medians: Projiziere alle Beobachtungen auf eine Sphäre um m . Dann ist m genau dann der geometrische Median, wenn er der Mittelwert dieser Projektionen ist.

Das ist eine mögliche Verallgemeinerung des reellen Median. [Sma90] verglich sie mit einigen Anderen und stellte fest, dass sie einige nützliche Eigenschaften hat und deshalb in vielen wissenschaftlichen Abhandlungen bevorzugt wird. Der geometrische Median lässt sich z.B. einfach auf allgemeine, unendlichdimensionale Banachräume verallgemeinern, während das bei anderen, wie dem simplicial depth median (vgl. [Liu90]) oder dem simplicial volume median (vgl. [Oja83]), nicht sinnvoll funktioniert. Diese Verallgemeinerung ist in der Statistik relevanter und wird uns in dieser Arbeit ausschließlich beschäftigen wird.

Definition 2.2.2. Sei V ein separabler Banachraum. Sei X eine Zufallsvariable mit Wertebereich in V . Der **geometrische Median** (auch L^1 -Median oder **spatial median**) m von X ist definiert als

$$m := \arg \min_{h \in V} \mathbb{E}[\|X - h\| - \|X\|]. \quad (2.2)$$

Bemerkung 2.2.3. Man bemerke, dass $\mathbb{E}[\|X - h\| - \|X\|]$ selbst dann endlich ist, wenn $\mathbb{E}[\|X\|]$ es nicht ist, da $\mathbb{E}[\|X - h\| - \|X\|] \leq \mathbb{E}[\|X - h - X\|] = \mathbb{E}[\|h\|] = \|h\| < \infty$. Dies unterscheidet die Definition von 2.2.1. Wir werden uns ausschließlich mit dem Fall befassen, bei dem V ein Hilbertraum ist und diesen dann mit $\mathcal{H}$ bezeichnen. Wir werden später beweisen, dass $\arg \min_{h \in V} \mathbb{E}[\|X - h\| - \|X\|]$ unter bestimmten Annahmen existiert und eindeutig ist.

Wir möchten nun Nachweisen, dass der geometrische Median im eindimensionalen Fall mit dem klassischen Median übereinstimmt.

Lemma 2.2.4. *Sei X eine eindimensionale Zufallsvariable mit atomfreier Verteilung und m eine reelle Zahl. Dann ist m der geometrische Median von X gdw. m der Median von X ist:*

$$\mathbb{E}[|X - m| + |X|] = \min_{h \in \mathbb{R}} \mathbb{E}[|X - h| + |X|] \Leftrightarrow \mathbb{P}(X < m) \leq \frac{1}{2} \quad \text{und} \quad \mathbb{P}(X > m) \leq \frac{1}{2}$$

Beweis. Sei $G(h) := \mathbb{E}[|X - h| - |X|]$. Es ist leicht einzusehen, dass G konvex ist. Dass G bei m ein Minimum annimmt, ist deshalb äquivalent dazu, dass die Ableitung von G in m eine Nullstelle hat. Es gilt also $0 = G'(m) = \mathbb{E}[\text{sgn}(X - m)] = \mathbb{P}(X > m) - \mathbb{P}(X < m)$. Wenn man die letzte Gleichung umstellt, erhält man $\mathbb{P}(X > m) = \mathbb{P}(X < m)$. Dies ist im atomfreien Fall äquivalent dazu, dass $\mathbb{P}(X > m) = \mathbb{P}(X < m) = \frac{1}{2}$ bzw. $\mathbb{P}(X > m), \mathbb{P}(X < m) \leq \frac{1}{2}$. $\square$

2.3 Anwendungen

Um die Relevanz des geometrischen Medians etwas zu illustrieren, geben wir exemplarisch einige Anwendung wieder. Wir geben Beispiele aus der Praxis und der Forschung.

2.3.1 Median der Einschaltquoten

Dieses Beispiel wurde direkt der Hauptquelle [CCZ13, Abschnitt 4.2] übernommen.

Die Analyse des Nutzerverhaltens von Fernsehzuschauern, für verschiedene Kanäle oder verschiedenen Tagen im Jahr, ist ein essentielles Werkzeug um die Konsumgewohnheiten zu verstehen. Die französische Firma Médiamétrie bietet offizielle Einschaltquoten an. Médiamétrie arbeitet mit einem Panel von etwa 9000 Personen, deren Fernseher mit Sensoren ausgestattet sind, die jede Sekunde messen, ob der Fernseher eingeschaltet ist und, wenn ja, welcher Kanal.

Wegen der Weise, auf welche Médiamétrie die Stichprobe nimmt, ist leider aus theoretischer Sicht die Bedingung der stochastisch unabhängig, identisch verteilten Zufallsvariablen nicht erfüllt. Dennoch soll es hier zur Illustration genügen. Des Weiteren hat Médiamétrie selbst angemerkt, dass es teilweise atypische Verhalten in den Stichproben gibt, die Messfehler nahelegen. Deshalb kann hier die Robustheit des Medians seine Wirkung entfalten.

Wir haben als Beispiel die Daten des sechsten September 2010 gewählt. Nach Entfernen der Leute, die an dem Tag überhaupt nicht ferngesehen haben, bleibt man bei $n = 5423$. Für jedes Element i in der Stichprobe haben wir einen Vektor $X_i \in \{0, 1\}^{86400}$, wobei 86400 für die Anzahl der Sekunden an einem Tag steht und eine Null bedeutet, dass zu der entsprechenden Sekunde Person i nicht Fernseher geschaut hat.

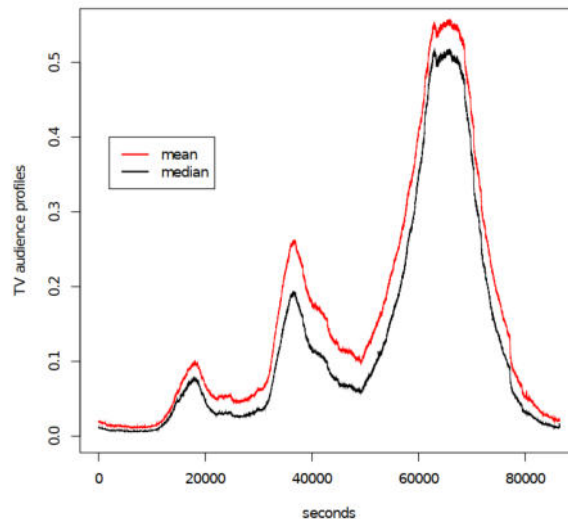


Abbildung 2.1: Mittelwert und geometrischer Median der Einschaltquoten, jede Sekunde am 6ten September 2010 [aus CCZ13, Abschnitt 5].

Wichtig zu bemerken ist, dass aus theoretischer Sicht, wegen Messfehlern und der begrenzten Stichprobengröße, beide Werte jeweils als Schätzungen des entsprechenden wahren Parameters der gesamten französischen Fernsehzuschauer zu verstehen sind. Wir werden später darauf zurückkommen, mit welchem Algorithmus der Median berechnet wurde.

In Abbildung 2.1 ist der klassische Indikator, der Durchschnitt, neben den Median eingezeichnet. Wir wollen die Beiden nun vergleichen. Der Durchschnitt gibt einfach an,

welcher Anteil der Personen aus der Stichprobe zu einer bestimmten Sekunde Fernseher geschaut haben. Wir können zunächst beobachten, dass beide Werte dem gleichen Trend folgen. Es gibt drei Hochpunkte während des Tages mit den höchsten Einschaltquoten zwischen 20 Uhr und 22 Uhr. Der geschätzte Median, welcher weniger sensibel auf kleine Störungen und Ausreißer reagiert, ist allgemein kleiner. Das impliziert auch, dass die Verteilungen der einzelnen Einschaltkurven nicht symmetrisch um den Mittelwert sind, da sonst Mittelwert und Median übereinstimmen würden.

2.3.2 Steigern der Konfidenz von Schätzern

Theorem 3.1. aus [Min+15] zeigt, dass speziell der geometrische Median eine Eigenschaft hat, die ihn in unterschiedlichsten statistischen Problemen extrem nützlich macht. Sei $\mathcal{H}$ ein Hilbertraum, $\mu \in \mathcal{H}$ ein interessierender Parameter und eine Folge von unabhängigen Schätzern $\hat{\mu}_1, \dots, \hat{\mu}_k \in \mathcal{H}$ für μ , von denen man annimmt, dass sie schwach um μ konzentriert sind. Dann ist der geometrische Median $\hat{\mu}$ dieser Schätzer ein neuer, eigener Schätzer, der einen wesentlich engeren und somit besseren Konfidenzbereich um μ hat. In anderen Worten hat $\hat{\mu}$ eine kleinere Standardabweichung. Damit gibt [Min+15] eine relativ allgemeine Methode, in Hilberträumen Schätzer zu optimieren. Es folgt die präzise mathematische Formulierung des Theorems.

Satz 2.3.1. *Sei $\mu \in \mathcal{H}$ der interessierender Parameter und seien $\hat{\mu}_1, \dots, \hat{\mu}_k \in \mathcal{H}$ eine endliche Folge von unabhängigen Schätzern für μ . Fixiere $\alpha \in (0, \frac{1}{2})$. Wähle $0 < p < \alpha$ und $\varepsilon > 0$, sodass für alle $j, 1 \leq j \leq k$*

$$\mathbb{P}(\|\hat{\mu}_j - \mu\| > \varepsilon) \leq p.$$

Definiere

$$C_\alpha := (1 - \alpha) \sqrt{\frac{1}{1 - 2\alpha}},$$

$$\psi(\alpha; p) := (1 - \alpha) \log \frac{1 - \alpha}{1 - p} + \alpha \log \frac{\alpha}{p}.$$

Setze

$$\hat{\mu} := \text{med}(\hat{\mu}_1, \dots, \hat{\mu}_k).$$

Dann

$$\mathbb{P}(\|\hat{\mu} - \mu\| > C_\alpha \varepsilon) \leq e^{-k\psi(\alpha; p)}.$$

Der Beweis findet sich in [Min+15] Theorem 3.1.

Bemerkung 2.3.2. Das Ergebnis kann auch benutzt werden, wenn man mehrere autonome Recheneinheiten (z.B. Prozessoren) zur Verfügung hat. Die einzelnen Schätzwerte $\hat{\mu}_j, j = 1, \dots, k$ können parallel berechnet werden um sie dann zum Abschluss über den geometrischen Median zusammenführt. In vielen Situationen führt das zu einer signifikanten Verbesserung des Rechenaufwands.

2.3.3 Weitere Anwendungen

[CCG17] und [CC+14] erwähnen jeweils noch einige andere Anwendungen in der Einleitung. Der geometrische Median wurde ähnlich zum Beispiel 2.3.1 in [CG12] benutzt, um ein median profile für für den elektrischen Verbrauch in Frankreich zu erstellen. Der Autor von [Ger08] schlug eine funktionale Hauptkomponentenanalyse, mit dem geometrischen Median als zentralen Lageparameter, vor und benutzte diese, um Daten mit Lippenbewegungen zu analysieren (vergleiche weiter [Loc+99] und [Bal+11]). In functional data analysis hat sich [KP12] resistente Schätzer für den Kovarianzoperator angeschaut, die auf dem geometrischen Median basieren. Der geometrischer Median von Schätzern unter der a-posteriori Verteilung wurde auch schon im Zusammenhang mit robuster Bayesscher Statistik in [Min+14] genutzt. Schließlich wurde eine allgemeine Definition des geometrischen Medians auf Mannigfaltigkeiten in [FVJ09] für ein Problem in der Signalverarbeitung gegeben.

Kapitel 3

Geometrischer Median Eigenschaften

Nachdem wir im vorherigen Kapitel gezeigt haben, dass der geometrische Median Relevanz hat, möchten wir in diesem einige interessante Eigenschaften zeigen. Insbesondere bestimmen wir die Ableitungen der Funktion G , die der geometrische Median minimiert. Des Weiteren zeigen wir die Existenz, Eindeutigkeit und Robustheit des geometrischen Median.

3.1 Annahmen

Zunächst gehen wir auf einige Annahmen ein, die wir treffen müssen, damit der geometrische Median existiert und der Algorithmus funktioniert. Wir orientieren uns an [CCZ13] Abschnitt 2.1 und Abschnitt 2 in [CCG17].

Sei $\mathcal{H}$ ein Hilbertraum und X eine Zufallsvariable mit Wertebereich in $\mathcal{H}$.

Annahme 1. Die Zufallsvariable X ist nicht auf einer Geraden konzentriert: für alle $h \in \mathcal{H}$, existiert $h' \in \mathcal{H}$, so dass $\langle h, h' \rangle = 0$ und

$$\text{Var}(\langle h', X \rangle) > 0.$$

Annahme 2. Die Verteilung $\mathbb{P}$ von X ist eine konvexe Kombination von zwei Verteilungen $\mathbb{P}_c$ und $\mathbb{P}_d$ (d.h. $\mathbb{P} = \lambda\mathbb{P}_c + (1 - \lambda)\mathbb{P}_d$), so dass gilt:

- $\mathbb{P}_c$ ist atomfrei und nicht um Punkte konzentriert: Sei Y eine Zufallsvariable mit Verteilung $\mathbb{P}_c$. Dann gibt es eine Konstante $C > 0$, so dass für alle $h \in \mathcal{B}_r$

$$\mathbb{E}_c \left[\frac{1}{\|X - h\|} \right] \leq C, \quad \mathbb{P}_c(\{h\}) = 0.$$

- $\mathbb{P}_d$ ist diskret: $\mathbb{P}_d = \sum_i p_i \delta_{x_i}$. Wir bezeichnen mit D den Träger von $\mathbb{P}_d^X$ und nehmen an, dass $m \notin D$.

Annahme 3. Es gibt eine positive Konstante C , so dass für alle $h \in \mathcal{H}$

$$\mathbb{E}_c \left[\frac{1}{\|X - h\|^2} \right] \leq C$$

gilt

Annahme 4. Wir werden zusätzlich zu den Annahmen aus [CCZ13] immer von $\dim(\mathcal{H}) \geq 3$ ausgehen.

Bemerkung 3.1.1. Aus Annahmen 2 und 3 folgt direkt, dass für $h \in \mathcal{H} \setminus D$ für ein C

$$\mathbb{E} \left[\frac{1}{\|X - h\|} \right] \leq C, \quad \mathbb{E} \left[\frac{1}{\|X - h\|^2} \right] \leq C,$$

gilt.

Bemerkung 3.1.2. Wie in Abschnitt 3.4.1 noch gezeigt wird, stellt Annahme 1 sicher, dass der geometrische Median m eindeutig ist. Da der in dieser Arbeit studierte Algorithmus in kleinen Dimensionen nicht sinnvoll ist, ist Annahme 4 keine bedeutende Einschränkung. Des Weiteren werden wir davon ausgehen, dass $m \notin D$.

Beispiel 3.1.3. Sei $\mathcal{H} = \mathbb{R}^d$ (vgl. Beispiel 2.1.7) mit $d \geq 3$ nach Annahme 4. Angenommen $\mathbb{P}_c^X$ hat eine Wahrscheinlichkeitsdichte f bezüglich des Lebesgue-Maßes, die auf jeder kompakten Teilmenge beschränkt ist (d.h. $\forall K \subseteq \mathbb{R}^d$ kompakt: $\sup_{x \in K} f(x) < \infty$). Dann kann gezeigt werden, dass Annahmen 2 und 3 erfüllt sind. Dies wird hier nicht bewiesen, da es den Rahmen dieser Arbeit überschreiten würde.

Beispiel 3.1.4. Sei $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^d)$ (vgl. 2.1.8). Dann ist die Dimension nicht endlich und die Wahrscheinlichkeit von kleinen Bällen wurden nur für einige bestimmte Fälle von Gauß-Prozessen hergeleitet. Wegen der Symmetrie stimmen Median und Mittelwert dann überein und in vielen Fällen gilt nach Theorem 1.2 aus [Naz07], dass es positive Konstanten C_1, C_2, C_3, C_4 gibt, so dass für $\varepsilon > 0$ und eine Zufallsvariable Y mit entsprechender Verteilung

$$\mathbb{P}[\|Y - m\| \leq \varepsilon] \leq C_1 \varepsilon^{C_4} \exp(-C_2 \varepsilon^{-C_3}).$$

gilt. Daraus folgt, dass für alle positiven β , $\mathbb{E}[\|Y - m\|^{-\beta}] < \infty$ gilt. Dies wird hier nicht bewiesen, da es den Rahmen dieser Arbeit überschreiten würde.

3.2 Ableitungen

In diesem Abschnitt schauen wir uns einige für die Beweise nützliche Funktionen an und rechnen einige Eigenschaften nach. Wir orientieren uns an 2.2 aus [CCZ13].

3.2.1 zu minimierende Funktion G

In Anlehnung an die Definition des geometrischen Medians in 2.2, definieren wir die Funktion $G : \mathcal{H} \rightarrow \mathbb{R}$, die wir zu minimieren versuchen

$$G(\alpha) := \mathbb{E}[\|X - \alpha\| - \|X\|]. \quad (3.1)$$

Als Verkettung konvexer Funktionen ist G auch konvex. Wegen Annahme 2 lässt sich diese Funktion in zwei Teile aufteilen:

$$G(\alpha) = \mathbb{E}[\|X - \alpha\| - \|X\|] = \lambda \mathbb{E}_c[\|X - \alpha\| - \|X\|] + (1 - \lambda) \mathbb{E}_d[\|X - \alpha\| - \|X\|]$$

$$= \lambda G_c(\alpha) + (1 - \lambda) G_d(\alpha),$$

wobei $\mathbb{E}_d$ und $\mathbb{E}_c$ jeweils für die Erwartung unter μ_c und μ_d stehen.

Nach Definition von μ_d ist $G_d(\alpha) = \sum_i p_i (\|x_i - \alpha\| - \|x_i\|)$. Es wird jedoch für die Ableitung nützlich sein, G auf D , dem Träger von μ_d und somit Unstetigkeitsstellen, etwas umzudefinieren. Für $\alpha \in D$ definieren wir G_d durch „Vergessen“ von α :

$$\tilde{G}_d(\alpha) = \begin{cases} \sum_{i, x_i \neq \alpha} p_i (\|x_i - \alpha\| - \|x_i\|) & \alpha \in D \\ \sum_i p_i (\|x_i - \alpha\| - \|x_i\|) & \text{sonst.} \end{cases} \quad (3.2)$$

3.2.2 Differential Φ

Wir möchten uns im Folgenden Differentiale von G anschauen. Zunächst müssen wir jene für Hilberträume definieren.

Definition 3.2.1. Sei $\mathcal{H}$ ein Hilbertraum über $\mathbb{R}$. Ein Funktional $G : \mathcal{H} \rightarrow \mathbb{R}$ ist **Fréchet differenzierbar** an der Stelle $y \in \mathcal{H}$ mit Ableitung $G'(y)$, falls $G'(y) : \mathcal{H} \rightarrow \mathbb{R}$ ein beschränkter, linearer Operator ist mit

$$\lim_{\|h\| \rightarrow 0} \frac{1}{\|h\|} |G(y+h) - G(y) - G'(y)h| = 0.$$

Mit beschränkt meinen wir, dass die Operatornorm $\|G'\|_L$ endlich ist. Die Fréchet-Ableitung ist die direkte Verallgemeinerung des totalen Differentials von endlichdimensionalen Vektorräumen.

Definition 3.2.2. Sei $\mathcal{H}$ ein Hilbertraum über $\mathbb{R}$ und $G : \mathcal{H} \rightarrow \mathbb{R}$ ein überall fréchet-differenzierbares Funktional. $\Phi : \mathcal{H} \rightarrow \mathcal{H}$ heißt **Gradient** von G , falls für alle $y \in \mathcal{H}$ $\langle \Phi(y), \cdot \rangle$ die Fréchet-Ableitung an der Stelle y ist.

Definition 3.2.3. Sei $\mathcal{H}$ ein Hilbertraum über $\mathbb{R}$ und $G : \mathcal{H} \rightarrow \mathbb{R}$ ein überall Fréchet-differenzierbares Funktional mit Fréchet-Ableitung G' . Dann heißt G **zweifach Fréchet differenzierbar** an der Stelle $y \in \mathcal{H}$ mit zweiter Ableitung $G''(y)$, falls $G''(y) : \mathcal{H} \rightarrow L(\mathcal{H}, \mathbb{R})$ ein beschränkter, linearer Operator ist mit

$$\lim_{\|h\| \rightarrow 0} \frac{1}{\|h\|} \|G'(y+h) - G'(y) - G''(y)h\|_L = 0.$$

Mit $L(X, Y)$ wird der Raum der beschränkten, linearen Abbildungen von X nach Y bezeichnet.

Definition 3.2.4. Sei $\mathcal{H}$ ein Hilbertraum über $\mathbb{R}$ und $G : \mathcal{H} \rightarrow \mathbb{R}$ ein überall zweifach fréchet-differenzierbares Funktional. Dann heißt $\Gamma_y : \mathcal{H} \rightarrow \mathcal{H}$ **Hesse-Matrix** an der Stelle $y \in \mathcal{H}$, falls $\mathcal{H} \rightarrow L(\mathcal{H}, \mathbb{R}), h \mapsto \langle \Gamma_y(h), \cdot \rangle$ die zweite Fréchet Ableitung ist.

Bemerkung 3.2.5. Übliche Eigenschaften der Ableitung, wie die Taylor Approximation, sind immer noch gültig, werden aber im Rahmen dieser Arbeit nicht bewiesen, obwohl sie benutzt werden.

Lemma 3.2.6. *Sei*

$$\Phi_c(y) := -\mathbb{E}_c \left[\frac{X - y}{\|X - y\|} \right].$$

Dann ist Φ_c der Gradient von G_c .

Für den Beweis orientieren wir uns an Lemma 1 aus dem Appendix für [Ger08].

Beweis. Für $y, h \in \mathcal{H}$ gilt

$$\begin{aligned} G_c(y+h) - G_c(y) - \langle \Phi_c(y), h \rangle &= \mathbb{E}_c \left[\|X - y - h\| - \|X - y\| - \left\langle -\frac{X - y}{\|X - y\|}, h \right\rangle \right] \\ &= \int \|x - y - h\| - \|x - y\| + \frac{\langle x - y, h \rangle}{\|x - y\|} d\mathbb{P}_c(x) \\ &= \int_{x \neq y} \|x - y - h\| - \|x - y\| + \frac{\langle x - y, h \rangle}{\|x - y\|} d\mathbb{P}_c(x). \end{aligned}$$

Bei der letzten Gleichung benutzen wir, dass wir $\mathbb{P}_c(\{y\}) = 0$ annehmen. Eine Taylor-Entwicklung zweiten Grades der Funktion $f(t) := \sqrt{t}$ um $t_0 := \|x - y\|^2$ für $x \neq y$ gibt uns für ein η zwischen $\|x - y - h\|^2$ und $\|x - y\|^2$

$$\begin{aligned} \|x - y - h\| &= f(\|x - y - h\|^2) \\ &= f(t_0) + f'(t_0)(\|x - y - h\|^2 - \|x - y\|^2) \\ &\quad + \frac{f''(\eta)}{2}(\|x - y - h\|^2 - \|x - y\|^2)^2 \\ &= f(t_0) + f'(t_0)(\|h\|^2 - 2\langle x - y, h \rangle) \\ &\quad + \frac{f''(\eta)}{2}(\|h\|^2 - 2\langle x - y, h \rangle)^2 \\ &= \|x - y\| + \frac{1}{2\|x - y\|}(\|h\|^2 - 2\langle x - y, h \rangle) \\ &\quad - \frac{1}{8\eta^{\frac{3}{2}}}(\|h\|^2 - 2\langle x - y, h \rangle)^2. \end{aligned} \tag{3.3}$$

Dabei benutzen wir bei der dritten Gleichung, dass $\|a - b\|^2 = \|a\|^2 + \|b\|^2 - 2\langle a, b \rangle$. Somit gilt für $x \neq y$

$$\begin{aligned} &\|x - y - h\| - \|x - y\| + \frac{\langle x - y, h \rangle}{\|x - y\|} \\ &= \frac{1}{2\|x - y\|}(\|h\|^2 - 2\langle x - y, h \rangle) - \frac{1}{8\eta^{\frac{3}{2}}}(\|h\|^2 - 2\langle x - y, h \rangle)^2 + \frac{\langle x - y, h \rangle}{\|x - y\|} \\ &= \frac{\|h\|^2}{2\|x - y\|} - \frac{(\|h\|^2 - 2\langle x - y, h \rangle)^2}{8\eta^{\frac{3}{2}}}. \end{aligned}$$

Daraus folgt

$$\begin{aligned} \left| \|x - y - h\| - \|x - y\| + \frac{\langle x - y, h \rangle}{\|x - y\|} \right| &= \left| \frac{\|h\|^2}{2\|x - y\|} - \frac{(\|h\|^2 - 2\langle x - y, h \rangle)^2}{8\eta^{\frac{3}{2}}} \right| \\ &\leq \left| \frac{\|h\|^2}{2\|x - y\|} \right| + \left| \frac{(\|h\|^2 - 2\langle x - y, h \rangle)^2}{8\eta^{\frac{3}{2}}} \right| \\ &\leq \frac{\|h\|^2}{2\|x - y\|} + \frac{(\|h\|^2 + 2\|x - y\|\|h\|)^2}{8\eta^{\frac{3}{2}}} \\ &\leq \frac{\|h\|^2}{2\|x - y\|} + \frac{(\|h\|^2 + 2\|x - y\|\|h\|)^2}{8 \min\{\|x - y - h\|^2, \|x - y\|^2\}^{\frac{3}{2}}} \\ &\leq \frac{\|h\|^2}{2\|x - y\|} \left(\frac{1}{2} + \frac{(\|h\| + 2\|x - y\|)^2}{8 \min\{\|x - y - h\|, \|x - y\|\}^3} \right). \end{aligned}$$

Folglich gilt

$$\begin{aligned}
 & \lim_{\|h\| \rightarrow 0} \frac{|G_c(y+h) - G_c(y) - \langle \Phi_c(y), h \rangle|}{\|h\|} \\
 &= \lim_{\|h\| \rightarrow 0} \frac{1}{\|h\|} \left| \int_{x \neq y} \|x-y-h\| - \|x-y\| + \frac{\langle x-y, h \rangle}{\|x-y\|} d\mathbb{P}_c(x) \right| \\
 &\leq \lim_{\|h\| \rightarrow 0} \frac{1}{\|h\|} \int_{x \neq y} \left| \|x-y-h\| - \|x-y\| + \frac{\langle x-y, h \rangle}{\|x-y\|} \right| d\mathbb{P}_c(x) \\
 &\leq \lim_{\|h\| \rightarrow 0} \|h\| \int_{x \neq y} \frac{1}{2\|x-y\|} + \frac{(\|h\|+2\|x-y\|)^2}{8\|x-y\|\min\{\|x-y-h\|, \|x-y\|\}^3} d\mathbb{P}_c(x).
 \end{aligned}$$

Da $\lim_{\|h\| \rightarrow 0} \frac{(\|h\|+2\|x-y\|)^2}{8\|x-y\|\min\{\|x-y-h\|, \|x-y\|\}^3} = \frac{2}{\|x-y\|^2}$ und $\int_{x \neq y} \frac{1}{2\|x-y\|} + \frac{2}{\|x-y\|^2} d\mathbb{P}_c(x) = \mathbb{E}_c \left[\frac{1}{2\|X-y\|} + \frac{2}{\|X-y\|^2} \right] < \infty$ nach Annahme 2 und 3, ist der umklammerte Teil mit dem Satz der dominierten Konvergenz endlich. Deshalb konvergiert der gesamte Ausdruck gegen 0. $\square$

Definiere weiterhin

$$\Phi_d(x) := \begin{cases} \sum_{i, x_i \neq x} p_i \frac{x-x_i}{\|x-x_i\|} & x \in D \\ \sum_i p_i \frac{x-x_i}{\|x-x_i\|} & \text{sonst,} \end{cases}$$

$$\Phi := \lambda \Phi_c + (1-\lambda) \Phi_d.$$

Bemerkung 3.2.7. Es gilt somit für $y \notin D$

$$\begin{aligned}
 \Phi(y) &= \lambda \Phi_c(y) + (1-\lambda) \Phi_d(y) = -\lambda \mathbb{E}_c \left[\frac{X-y}{\|X-y\|} \right] + (1-\lambda) \sum_i p_i \frac{y-x_i}{\|y-x_i\|} \\
 &= -\lambda \mathbb{E}_c \left[\frac{X-y}{\|X-y\|} \right] - (1-\lambda) \mathbb{E}_d \left[\frac{X-y}{\|X-y\|} \right] = -\mathbb{E} \left[\frac{X-y}{\|X-y\|} \right].
 \end{aligned}$$

Offensichtlich ist dann Φ_d der Gradient von $\tilde{G}_d$ (vgl 3.2). Desweiteren gilt für $x \notin D$: $\tilde{G}_d(x) = G_d(x)$ und somit $G(x) = G_c(x) + \tilde{G}_d(x)$. Daraus folgt, dass dann auch $\Phi(x)$ der Gradient von G an der Stelle x ist.

Für $x \in D$ gilt dies jedoch nicht allgemein. Wir können aber eine eine andere, etwas schwächere Aussage treffen.

Definition 3.2.8. Sei $g : \mathcal{H} \rightarrow \mathbb{R}$ eine konvexe Abbildung, $f : \mathcal{H} \rightarrow \mathcal{H}$ eine Abbildung und $x_0 \in \mathcal{H}$. Dann ist $f(x_0)$ ein **Subdifferential** von g an der Stelle x_0 , falls

$$\forall y \in \mathcal{H} : \quad g(y) - g(x_0) \geq \langle f(x_0), y - x_0 \rangle.$$

f heißt Subdifferential von g , falls für alle $x_0 \in \mathcal{H}$ $f(x_0)$ ein Subdifferential von g an der Stelle x_0 ist.

Lemma 3.2.9. Die Fréchet-Ableitung f einer eindimensionalen, konvexen Funktion g ist immer ein Subdifferential.

Beweis. Seien $x_0, y \in \mathcal{H}$ gegeben. Setze $y_t := x_0 + t(y - x_0) = (1 - t)x_0 + ty$. Dann gilt

$$g(y) = g\left(y_t + \frac{1}{t}(1 - t)(y_t - x_0)\right) \geq g(y_t) + \frac{1 - t}{t}(g(y_t) - g(x_0)).$$

Wobei die Ungleichung einfach die Konvexität von g ist. Sei $h := y - x_0$, so dass $y_t = x_0 + th$. Es gilt

$$\begin{aligned} \frac{1 - t}{t}(g(y_t) - g(x_0)) &= \frac{1 - t}{t}(g(x_0 + th) - g(x_0)) \\ &= \frac{1}{t}(g(x_0 + th) - g(x_0)) - (g(x_0 + th) - g(x_0)) \\ &\rightarrow \langle f(x_0), h \rangle + 0 = \langle f(x_0), y - x_0 \rangle, \end{aligned}$$

für $t \rightarrow 0$, woraus die Aussage direkt folgt. $\square$

Lemma 3.2.10. *Die Funktion Φ_d ist ein Subdifferential von G_d .*

Beweis. Für den Fall, dass $x_0 \notin D$, ist die Aussage durch Lemma 3.2.9 klar, da dann Φ_d die Fréchet-Ableitung an der Stelle x_0 ist. Sei also $x_0 \in D$

$$\begin{aligned} \langle \Phi_d(x_0), y - x_0 \rangle &= \left\langle \sum_{i, x_i \neq x_0} p_i \frac{x_0 - x_i}{\|x_0 - x_i\|}, y - x_0 \right\rangle \\ &= \sum_{i, x_i \neq x_0} p_i \frac{\langle x_0 - x_i, y - x_0 \rangle}{\|x_0 - x_i\|} = \sum_{i, x_i \neq x_0} p_i \frac{\langle x_0 - x_i, y - x_i + x_i - x_0 \rangle}{\|x_0 - x_i\|} \\ &= \sum_{i, x_i \neq x_0} p_i \frac{\langle x_0 - x_i, y - x_i \rangle - \langle x_0 - x_i, x_0 - x_i \rangle}{\|x_0 - x_i\|} \\ &= \sum_{i, x_i \neq x_0} p_i \frac{\langle x_0 - x_i, y - x_i \rangle - \|x_0 - x_i\|^2}{\|x_0 - x_i\|} \\ &= \sum_{i, x_i \neq x_0} p_i \frac{\langle x_0 - x_i, y - x_i \rangle}{\|x_0 - x_i\|} - \sum_{i, x_i \neq x_0} p_i \|x_0 - x_i\| \\ &\leq \sum_{i, x_i \neq x_0} p_i \frac{\|y - x_i\| \cdot \|x_0 - x_i\|}{\|x_0 - x_i\|} - \sum_{i, x_i \neq x_0} p_i \|x_0 - x_i\| \\ &= \sum_{i, x_i \neq x_0} p_i \|y - x_i\| - \sum_{i, x_i \neq x_0} p_i \|x_0 - x_i\| \\ &= \sum_{i, x_i \neq x_0} p_i \|y - x_i\| - \sum_i p_i \|x_0 - x_i\| \\ &\leq \sum_i p_i \|y - x_i\| - \sum_i p_i \|x_0 - x_i\| = G_d(y) - G_d(x_0), \end{aligned}$$

wobei die erste Ungleichung aus der Cauchy-Schwarz-Ungleichung folgt. Die Zweite folgt direkt aus der Definition von G_d . $\square$

Satz 3.2.11. *Die Funktion Φ ist ein Subdifferential von G .*

Beweis. Nach Lemma 3.2.10 ist Φ_d ein Subdifferential von G_d . Nach Lemma 3.2.6 ist Φ_c die Fréchet-Ableitung von G_c . Aus Lemma 3.2.9 folgt, dass Φ_c auch ein Subdifferential von

G_c . Entsprechend folgt

$$\begin{aligned}
 \langle \Phi(x_0), y - x \rangle &= \lambda \langle \Phi_c(x_0), y - x \rangle + (1 - \lambda) \langle \Phi_d(x_0), y - x \rangle \\
 &\leq \lambda (G_c(y) - G_c(x_0)) + (1 - \lambda) (G_d(y) - G_d(x_0)) \\
 &= (\lambda G_c(y) + (1 - \lambda) G_d(y)) - (\lambda G_c(x_0) + (1 - \lambda) G_d(x_0)) \\
 &= G(y) - G(x).
 \end{aligned}$$

Dabei benutzt die zweite Zeile, dass Φ_d bzw. Φ_c das Subdifferential von G_d bzw. G_c ist. $\square$

Lemma 3.2.12. *Seien g, x_0 wie in 3.2.8 gegeben. Dann ist x_0 ein globales Minimum von g , genau dann wenn es ein Subdifferential f von g an der Stelle x_0 gibt mit $f(x_0) = 0$*

Beweis. „ $\Rightarrow$ “. x_0 ein globales Minimum von $g \Rightarrow \forall h \in \mathcal{H} : g(x_0 + h) \geq g(x_0) \Rightarrow \langle f(x_0), h \rangle \leq g(x_0 + h) - g(x_0) \leq 0 \Rightarrow f(x_0) \in \mathcal{H}^\perp = \{0\} \Rightarrow f(x_0) = 0$

„ $\Leftarrow$ “. $f(x_0) = 0 \Rightarrow \forall y \in \mathcal{H} : g(y) - g(x_0) \geq \langle f(x_0), y - x_0 \rangle = 0 \Rightarrow g(y) \geq g(x_0) \Rightarrow x_0$ ist ein globales Minimum von g $\square$

Bemerkung 3.2.13. Somit ist nach Lemma 3.2.11 und Lemma 3.2.12 der Median m eine Lösung der nicht-linearen Gleichung

$$\Phi(x) = 0.$$

3.2.3 Hesse-Matrix Γ_α

Wir führen auch die Hesse-Matrix von G ein.

Lemma 3.2.14. *Sei $y \in \mathcal{H} \setminus D$. Dann ist*

$$\Gamma_y = \mathbb{E} \left[\frac{1}{\|X - y\|} \left(\mathbf{I}_{\mathcal{H}} - \frac{(X - y) \otimes (X - y)}{\|X - y\|^2} \right) \right]$$

die Hesse Matrix von G an der Stelle y . Dabei ist $\mathbf{I}_{\mathcal{H}}$ die Identität und $u \otimes v(h) = \langle u, h \rangle v$ für $u, v, h \in \mathcal{H}$

Für den Beweis orientieren wir uns an Lemma 1 (i) aus dem Appendix für [Ger08].

Beweis. Wir wissen nach Bemerkung 3.2.7, dass Φ der Gradient von G auf $\mathcal{H} \setminus D$ ist. Seien $f, y, h \in \mathcal{H}$ mit $\|f\| = 1$. Dann gilt

$$\begin{aligned}
 &\langle \Phi(y + h), f \rangle - \langle \Phi(y), f \rangle - \langle \Gamma_y(h), f \rangle \\
 &= \mathbb{E} \left[\left\langle \frac{X - y}{\|X - y\|} - \frac{X - y - h}{\|X - y - h\|} - \frac{1}{\|X - y\|} \left(\mathbf{I}_{\mathcal{H}} - \frac{(X - y) \otimes (X - y)}{\|X - y\|^2} \right) (h), f \right\rangle \right] \\
 &= \mathbb{E} \left[\left\langle \frac{X - y}{\|X - y\|} - \frac{X - y - h}{\|X - y - h\|} + \frac{1}{\|X - y\|} \left(h - \frac{\langle X - y, h \rangle (X - y)}{\|X - y\|^2} \right), f \right\rangle \right] \\
 &= \mathbb{E} \left[\frac{\langle X - y, f \rangle}{\|X - y\|} - \frac{\langle X - y - h, f \rangle}{\|X - y - h\|} - \frac{1}{\|X - y\|} \left(\langle h, f \rangle - \frac{\langle X - y, h \rangle \langle X - y, f \rangle}{\|X - y\|^2} \right) \right] \\
 &= \mathbb{E} \left[\frac{\langle X - y, f \rangle}{\|X - y\|} - \frac{\langle X - y - h, f \rangle}{\|X - y - h\|} - \frac{\langle h, f \rangle}{\|X - y\|} + \frac{\langle X - y, h \rangle \langle X - y, f \rangle}{\|X - y\|^3} \right] = A + B,
 \end{aligned}$$

mit

$$\begin{aligned} A &:= -\mathbb{E} \left[\langle X - y, f \rangle \left(\frac{1}{\|X - y - h\|} - \frac{1}{\|X - y\|} - \frac{\langle X - y, h \rangle}{\|X - y\|^3} \right) \right], \\ B &:= -\mathbb{E} \left[\frac{\langle h, f \rangle}{\|X - y\|} - \frac{\langle h, f \rangle}{\|X - y - h\|} \right], \end{aligned}$$

Wir möchten nun für beide Terme zeigen, dass sie schneller als h gegen 0 konvergieren. Ähnlich zur Gleichung 3.3 aus dem Beweis zu Lemma 3.2.6 führen wir eine Taylor-Entwicklung durch. Setze $f(t) := \frac{1}{\sqrt{t}}$ und $t_0 := \|x - y\|^2$. Dann gilt für ein η zwischen $\|x - y - h\|^2$ und $\|x - y\|^2$

$$\begin{aligned} \frac{1}{\|x - y - h\|} &= f(\|x - y - h\|^2) \\ &= f(t_0) + f'(t_0) (\|h\|^2 - 2\langle x - y, h \rangle) + \frac{f''(\eta)}{2} (\|h\|^2 - 2\langle x - y, h \rangle)^2 \\ &= \frac{1}{\sqrt{t_0}} - \frac{1}{2t_0^{\frac{3}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle) + \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle)^2 \\ &= \frac{1}{\|x - y\|} - \frac{1}{2\|x - y\|^3} (\|h\|^2 - 2\langle x - y, h \rangle) + \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle)^2 \\ &= \frac{1}{\|x - y\|} + \frac{\langle x - y, h \rangle}{\|x - y\|^3} - \frac{\|h\|^2}{2\|x - y\|^3} + \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle)^2. \end{aligned} \tag{3.4}$$

Und deshalb gilt

$$\begin{aligned} &\left| \frac{1}{\|x - y\|} - \frac{1}{\|x - y - h\|} \right| \\ &= \left| \frac{1}{2\|x - y\|^3} (\|h\|^2 - 2\langle x - y, h \rangle) - \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle)^2 \right| \\ &= \|h\|^2 - 2\langle x - y, h \rangle \left| \frac{1}{2\|x - y\|^3} - \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle) \right| \\ &\leq \|h\| (\|h\| + 2\|x - y\|) \left| \frac{1}{2\|x - y\|^3} - \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle) \right|. \end{aligned} \tag{3.5}$$

Bei der Ungleichung benutzen wir die Dreiecksungleichung und Cauchy-Schwarz. Bei A gilt deshalb

$$\begin{aligned} &\left| \langle x - y, f \rangle \left(\frac{1}{\|x - y - h\|} - \frac{1}{\|x - y\|} - \frac{\langle x - y, h \rangle}{\|x - y\|^3} \right) \right| \\ &\leq \|x - y\| \left| \frac{1}{\|x - y - h\|} - \frac{1}{\|x - y\|} - \frac{\langle x - y, h \rangle}{\|x - y\|^3} \right| \\ &= \|x - y\| \left| \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x - y, h \rangle)^2 - \frac{\|h\|^2}{2\|x - y\|^3} \right| \\ &\leq \|x - y\| \left(\frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 + 2|\langle x - y, h \rangle|)^2 + \frac{\|h\|^2}{2\|x - y\|^3} \right) \\ &\leq \|x - y\| \left(\frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 + 2\|x - y\| \|h\|)^2 + \frac{\|h\|^2}{2\|x - y\|^3} \right) \\ &= \|h\|^2 \|x - y\| \left(\frac{3}{8\eta^{\frac{5}{2}}} (\|h\| + 2\|x - y\|)^2 + \frac{1}{2\|x - y\|^3} \right) \\ &= \|h\|^2 \left(\frac{3\|x - y\|}{8\eta^{\frac{5}{2}}} (\|h\| + 2\|x - y\|)^2 + \frac{1}{2\|x - y\|^2} \right). \end{aligned}$$

Dabei benutzen wir in der dritten Zeile Gleichung 3.4, bei der ersten und letzten Ungleichung Cauchy-Schwarz und bei der zweiten Ungleichung die Dreiecksungleichung. Da $\lim_{h \rightarrow 0} \left(\frac{3\|x-y\|}{8\eta^{\frac{5}{2}}} (\|h\| + 2\|x-y\|)^2 + \frac{1}{2\|x-y\|^2} \right) = \frac{3}{2\|x-y\|^2} + \frac{1}{2\|x-y\|^2} = \frac{2}{\|x-y\|^2}$ gilt und nach Bemerkung 3.1.1 $\mathbb{E} \left[\frac{1}{\|X-y\|^2} \right] < \infty$, da wir $y \notin D$ voraussetzen, folgt insgesamt

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{|A|}{\|h\|} &\leq \lim_{h \rightarrow 0} \frac{1}{\|h\|} \mathbb{E} \left[\left| \langle X-y, f \rangle \left(\frac{1}{\|X-y-h\|} - \frac{1}{\|X-y\|} - \frac{\langle X-y, h \rangle}{\|X-y\|^3} \right) \right| \right] \\ &\leq \lim_{h \rightarrow 0} \|h\| \mathbb{E} \left[\frac{3\|X-y\|}{8\eta^{\frac{5}{2}}} (\|h\| + 2\|X-y\|)^2 + \frac{1}{2\|X-y\|^2} \right] = 0. \end{aligned}$$

Nach ähnlichen Argumenten können wir auch B abschätzen.

$$\begin{aligned} &\left| \frac{\langle h, f \rangle}{\|x-y\|} - \frac{\langle h, f \rangle}{\|x-y-h\|} \right| \\ &= |\langle h, f \rangle| \left| \frac{1}{\|x-y\|} - \frac{1}{\|x-y-h\|} \right| \leq \|h\| \left| \frac{1}{\|x-y\|} - \frac{1}{\|x-y-h\|} \right| \\ &\leq \|h\|^2 (\|h\| + 2\|x-y\|) \left| \frac{1}{2\|x-y\|^3} - \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x-y, h \rangle) \right|. \end{aligned}$$

Bei der Ungleichung benutzen wir Gleichung 3.5. Da $\lim_{h \rightarrow 0} (\|h\| + 2\|x-y\|) \left| \frac{1}{2\|x-y\|^3} - \frac{3}{8\eta^{\frac{5}{2}}} (\|h\|^2 - 2\langle x-y, h \rangle) \right| = \frac{1}{\|x-y\|^2}$ und nach Bemerkung 3.1.1 $\mathbb{E} \left[\frac{1}{\|X-y\|^2} \right] < \infty$ gilt, da wir $y \notin D$ voraussetzen, folgt insgesamt

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{|B|}{\|h\|} &\leq \lim_{h \rightarrow 0} \frac{1}{\|h\|} \mathbb{E} \left[\left| \frac{\langle h, f \rangle}{\|X-y\|} - \frac{\langle h, f \rangle}{\|X-y-h\|} \right| \right] \\ &\leq \lim_{h \rightarrow 0} \|h\| \mathbb{E} \left[(\|h\| + 2\|X-y\|) \left| \frac{1}{2\|X-y\|^3} - \frac{3(\|h\|^2 - 2\langle X-y, h \rangle)}{8\eta^{\frac{5}{2}}} \right| \right] \\ &= 0. \end{aligned}$$

Insgesamt haben wir

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\|\Phi(y+h) - \Phi(y) - \Gamma_y(h)\|_L}{\|h\|} &= \sup_{\|f\|=1} \lim_{h \rightarrow 0} \frac{|\langle \Phi(y+h) - \Phi(y) - \Gamma_y(h), f \rangle|}{\|h\|} \\ &\leq \sup_{\|f\|=1} \lim_{h \rightarrow 0} \frac{|A|}{\|h\|} + \frac{|B|}{\|h\|} = 0. \end{aligned}$$

□

Lemma 3.2.15. Für $\alpha \in \mathcal{H} \setminus D$ ist Γ_α unter Annahme 2 ein beschränkter Operator.

Beweis. Sei $h \in \mathcal{H}$ mit $\|h\| = 1$. Dann gilt

$$\begin{aligned}
 \|\Gamma_\alpha h\| &= \left\| \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(\mathbf{I}_{\mathcal{H}} - \frac{(X - \alpha) \otimes (X - \alpha)}{\|X - \alpha\|^2} \right) \right] (h) \right\| \\
 &= \left\| \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(h - \frac{\langle X - \alpha, h \rangle (X - \alpha)}{\|X - \alpha\|^2} \right) \right] \right\| \\
 &\leq \mathbb{E} \left[\left\| \frac{1}{\|X - \alpha\|} \left(h - \frac{\langle X - \alpha, h \rangle (X - \alpha)}{\|X - \alpha\|^2} \right) \right\| \right] \\
 &\leq \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(\|h\| + \frac{\|\langle X - \alpha, h \rangle (X - \alpha)\|}{\|X - \alpha\|^2} \right) \right] \\
 &\leq \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(1 + \frac{\|X - \alpha\| \cdot \|h\| \|X - \alpha\|}{\|X - \alpha\|^2} \right) \right] \\
 &= \mathbb{E} \left[\frac{1}{\|X - \alpha\|} (1 + 1) \right] = 2 \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \right] < \infty.
 \end{aligned}$$

Dabei sind die ersten beiden Ungleichungen Dreiecksungleichung und die Dritte Cauchy-Schwarz-Ungleichung. Dass das Ergebnis endlich ist, war genau Bemerkung 3.1.1. $\square$

Bemerkung 3.2.16. Γ_α ist kein kompakter Operator.

3.3 Konvexität

In diesem Abschnitt führen Konvexität allgemein ein und zeigen mit Hilfe der Ableitungen, dass G konvex ist.

3.3.1 Allgemeine Einführung

Definition 3.3.1. Eine Funktion f heißt **streng konvex**, falls für $x, y \in \mathcal{H}$ mit $x \neq y$ und alle $t \in (0, 1)$

$$f(tx + (1 - t)y) < tf(x) + (1 - t)f(y)$$

gilt.

Lemma 3.3.2. Sei f eine zweifach differenzierbare Funktion von $\mathcal{H}$ nach $\mathbb{R}$ und Γ_α die Hesse-Matrix von f an der Stelle α . Wenn Γ_α überall positiv definit ist, das heißt für alle $h, \alpha \in \mathcal{H}$ gilt $\langle h, \Gamma_\alpha h \rangle > 0$, dann ist f streng konvex.

Beweis. Seien $x, y \in \mathcal{H}$ und $t \in (0, 1)$. Setze $z := tx + (1 - t)y$. Nach Taylor Approximation gilt für gewisse $\eta, \eta' \in \mathcal{H}$

$$f(x) = f(z) + \nabla f(z)^T(x - z) + \frac{1}{2} \langle x - z, \Gamma_\eta(x - z) \rangle > f(z) + \nabla f(z)^T(x - z),$$

$$f(y) = f(z) + \nabla f(z)^T(y - z) + \frac{1}{2} \langle y - z, \Gamma_{\eta'}(y - z) \rangle > f(z) + \nabla f(z)^T(y - z).$$

Dabei benutzen wir bei den Ungleichungen, dass die Hesse Matrix streng positiv definit ist. Daraus folgt

$$\begin{aligned}
 tf(x) + (1 - t)f(y) &> t(f(z) + \nabla f(z)^T(x - z)) + (1 - t)(f(z) + \nabla f(z)^T(y - z)) \\
 &= f(z) + \nabla f(z)^T(t(x - z) + (1 - t)(y - z)) \\
 &= f(z) + \nabla f(z)^T(tx + (1 - t)y - z) = f(z) + \nabla f(z)^T(0) \\
 &= f(tx + (1 - t)y).
 \end{aligned}$$

□

Lemma 3.3.3. *Sei f eine streng konvexe Funktion von $\mathcal{H}$ nach $\mathbb{R}$. Dann besitzt f höchstens ein globales Minimum.*

Beweis. Seien $x, y \in \mathcal{H}$ globale Minima von f . Insbesondere ist dann $f(x) = f(y)$. Weil f streng konvex ist, gilt dann

$$f\left(\frac{1}{2}x + \frac{1}{2}y\right) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = f(x).$$

Somit wäre $f(\frac{x+y}{2}) < f(x)$, was der Minimalität widersprechen würde. □

3.3.2 Anwendung

Wir möchten nun die Konvexität von G nachweisen und für uns hilfreiche Rückschlüsse ziehen.

Im Folgenden bezeichne Γ_α stets die Hesse-Matrix von G (vgl. Lemma 3.2.14). Seien $\alpha, h \in \mathcal{H}$. Setze $\bar{h} = h/\|h\|$.

Lemma 3.3.4. *Es gilt*

$$\langle h, \Gamma_\alpha h \rangle = \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \left(1 - \frac{\langle \bar{h}, \alpha - X \rangle^2}{\|\alpha - X\|^2} \right) \right].$$

Beweis. Zur Erinnerung: $u \otimes v(h) = \langle u, h \rangle v$ für $u, v, h \in \mathcal{H}$.

$$\begin{aligned} \langle h, \Gamma_\alpha h \rangle &= \left\langle h, \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(\mathbf{I}_{\mathcal{H}} - \frac{(X - \alpha) \otimes (X - \alpha)}{\|X - \alpha\|^2} \right) \right] (h) \right\rangle \\ &= \left\langle h, \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(h - \frac{\langle X - \alpha, h \rangle (X - \alpha)}{\|X - \alpha\|^2} \right) \right] \right\rangle \\ &= \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(\langle h, h \rangle - \frac{\langle h, \langle X - \alpha, h \rangle (X - \alpha) \rangle}{\|X - \alpha\|^2} \right) \right] \\ &= \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(\|h\|^2 - \frac{\langle X - \alpha, h \rangle^2}{\|X - \alpha\|^2} \right) \right] \\ &= \mathbb{E} \left[\frac{1}{\|X - \alpha\|} \left(\|h\|^2 - \frac{\|h\|^2 \cdot \langle \bar{h}, X - \alpha \rangle^2}{\|X - \alpha\|^2} \right) \right] \\ &= \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \left(1 - \frac{\langle \bar{h}, \alpha - X \rangle^2}{\|\alpha - X\|^2} \right) \right] \end{aligned}$$

In der ersten Zeile benutzen wir Lemma 3.2.14. Wobei die dritte Gleichung aus Linearität des Skalarprodukts und des Erwartungswerts folgt. □

Lemma 3.3.5. *Sei P_h die Projektion auf das orthogonale Komplement von h . Dann gilt*

$$\langle h, \Gamma_\alpha h \rangle = \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \frac{\|P_h(\alpha - X)\|^2}{\|\alpha - X\|^2} \right].$$

.

Beweis. Offensichtlich ist $\{\bar{h}\}$ eine Orthonormalbasis für $\text{span}(h)$. Deshalb gilt $P_{\bar{h}} = \mathbf{I}_{\mathcal{H}} - \langle \bar{h}, \cdot \rangle \bar{h}$. Aus Lemma 3.3.4 folgt

$$\begin{aligned} \langle h, \Gamma_{\alpha} h \rangle &= \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \left(1 - \frac{\langle \bar{h}, \alpha - X \rangle^2}{\|\alpha - X\|^2} \right) \right] \\ &= \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \left(\frac{\|\alpha - X\|^2 - \langle \bar{h}, \alpha - X \rangle^2}{\|\alpha - X\|^2} \right) \right] \\ &= \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \frac{\|(\alpha - X) - \langle \bar{h}, \alpha - X \rangle \bar{h}\|^2}{\|\alpha - X\|^2} \right] \\ &= \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \frac{\|P_{\bar{h}}(\alpha - X)\|^2}{\|\alpha - X\|^2} \right]. \end{aligned}$$

Wobei die dritte Gleichung aus dem Satz Pythagoras folgt

$$\begin{aligned} \|(\alpha - X) - \langle \bar{h}, \alpha - X \rangle \bar{h}\|^2 &= \|\alpha - X\|^2 + \|\langle \bar{h}, \alpha - X \rangle \bar{h}\|^2 \\ &\quad - 2\langle \alpha - X, \langle \bar{h}, \alpha - X \rangle \bar{h} \rangle \\ &= \|\alpha - X\|^2 + \langle \bar{h}, \alpha - X \rangle^2 \|\bar{h}\|^2 \\ &\quad - 2\langle \bar{h}, \alpha - X \rangle \langle \bar{h}, \alpha - X \rangle \\ &= \|\alpha - X\|^2 + \langle \bar{h}, \alpha - X \rangle^2 \cdot 1 - 2\langle \bar{h}, \alpha - X \rangle^2 \\ &= \|\alpha - X\|^2 - \langle \bar{h}, \alpha - X \rangle^2. \end{aligned}$$

□

Lemma 3.3.6. Sei A eine positive Konstante und $\mathcal{B}_A$ der Ball von Radius A . Unter Annahmen 1 und 2 existieren positive Konstanten c_A, C_A , sodass

$$\forall \alpha \in \mathcal{B}_A \setminus D, \forall h \in \mathcal{H}, \quad c_A \|h\|^2 \leq \langle h, \Gamma_{\alpha} h \rangle \leq C_A \|h\|^2,$$

wobei D aus Annahmen 2 stammt.

Beweis. Abschätzung von oben:

$$\begin{aligned} \langle h, \Gamma_{\alpha} h \rangle &= \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \frac{\|P_{\bar{h}}(\alpha - X)\|^2}{\|\alpha - X\|^2} \right] \leq \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \frac{\|\alpha - X\|^2}{\|\alpha - X\|^2} \right] \\ &= \|h\|^2 \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \right] := C_A \|h\|^2 \end{aligned}$$

Wobei die erste Gleichung aus Lemma 3.3.5 folgt. Die erste Ungleichung rührt daher, dass $P_{\bar{h}}$ eine orthogonale Projektion ist. $C_A := \mathbb{E} \left[\frac{1}{\|\alpha - X\|} \right]$ ist wegen Annahme 2 endlich, da $\alpha \notin D$. Für die Abschätzung von unten genügt es wegen Lemma 3.3.5

$$\forall \alpha \in \mathcal{H}, \forall u \in \mathcal{H}, \|u\| = 1, \quad \langle u, \Gamma_{\alpha} u \rangle = \mathbb{E} \left[\frac{\|P_u(X - \alpha)\|^2}{\|X - \alpha\|^3} \right] \geq c_A \quad (3.6)$$

zu zeigen. Sei $\mathcal{U} := \{U \subseteq \mathcal{H} \mid \text{Untervektorraum} \mid \forall x \in U : \text{Var}(\langle x, X \rangle) = 0\}$. Da $\{0\} \in \mathcal{U}$, ist $\mathcal{U}$ nicht leer. Offensichtlich ist das eine halbgeordnete Menge, in der jede Kette eine obere Schranke hat. Nach dem Lemma von Zorn hat $\mathcal{U}$ hinsichtlich der Mengeninklusion ein maximales Element U' .

Zwischenbehauptung. Es gibt zwei orthogonale Vektoren $0 \neq v_1, v_2$ in $U'^\perp$

Beweis Zur Erinnerung: Annahme 1 besagt, dass für $\forall h \in \mathcal{H} \exists h' \in \mathcal{H} : \langle h, h' \rangle = 0$ und $\text{Var}(\langle h', X \rangle) > 0$. Daraus folgt direkt, dass $U' \neq \mathcal{H}$. Da nach Annahme 4 $\dim(\mathcal{H}) \geq 3$, gibt es ein $0 \neq v_1 \in U'^\perp$. Wieder nach Annahme 1 gibt es ein $\tilde{v}_2 \in \mathcal{H}$ für das $\langle v_1, \tilde{v}_2 \rangle = 0$ und $\text{Var}(\langle \tilde{v}_2, X \rangle) > 0$ gilt. Letzteres zeigt uns, dass $\tilde{v}_2 \notin U'$ und somit $\text{codim}(U' \oplus \text{span}(v_1)) \geq 1$. Aus Dimensionsgründen können wir ein $0 \neq v_2 \in U'^\perp$ wählen, das immer noch orthogonal zu v_1 ist (d.h. $v_2 \in (U' \oplus \text{span}(v_1))^\perp$).

#

Seien nun solche v_1, v_2 gegeben. Definiere für $t \in [0, 2\pi]$ $v_t = \cos(t)v_1 + \sin(t)v_2$ und die Abbildung $f : [0, 2\pi] \rightarrow \mathbb{R}$, $t \mapsto \text{Var}(\langle v_t, X \rangle)$. f ist als Verkettung stetiger Funktionen stetig und da es auf einer kompakten Menge definiert ist, nimmt es ein globales Minimum c an.

Zwischenbehauptung. $c > 0$

Beweis Angenommen $c = 0$. Dann gibt es ein $t_0 \in [0, 2\pi]$ mit $f(t_0) = 0$. Betrachte v_{t_0} . Nach Definition gilt $\text{Var}(\langle v_{t_0}, X \rangle) = 0$, weshalb $\text{span}(v_{t_0}) \in \mathcal{U}$. Da U' maximal in $\mathcal{U}$ ist, ist $\text{span}(v_{t_0}) \subseteq U'$ also $\cos(t_0)v_1 + \sin(t_0)v_2 = v_{t_0} \in U'$. Da $v_1, v_2 \in U'^\perp$ und $|\sin(t_0)| + |\cos(t_0)| \neq 0$ kann das aber nicht sein.

#

Das bedeutet, dass für alle $v \in V$ für $V := \text{span}(v_1, v_2)$ mit $\|v\| = 1$ $\text{Var}(\langle X, v \rangle) \geq c > 0$ gilt. Kommen wir zurück zu u aus der eigentlichen Behauptung 3.6. Aus Dimensionsgründen müssen $\text{span}(u)^\perp$ und V einen nicht-trivialen Schnitt haben. Sei nun $v \in V$ mit $\langle v, u \rangle = 0$ und $\|v\| = 1$ gegeben. Da also v ein Einheitsvektor in $\text{span}(u)^\perp$ ist, gilt für die orthogonale Projektion P_u auf $\text{span}(u)^\perp$ für alle $y \in \mathcal{H}$: $\|P_u y\|^2 \geq \langle y, v \rangle^2$. Insbesondere gilt $\|P_u(X - \alpha)\|^2 \geq \langle X - \alpha, v \rangle^2$. Sei $K \in \mathbb{R}$ und betrachte $\text{Var}(\langle v, X \rangle) \mathbb{1}_{\{\|X\| \leq K\}}$. Wir möchten zeigen, dass Letzteres, wenn man K groß genug wählt, auch immer echt positiv ist:

$$\begin{aligned} \mathbb{1}_{\{\|X\| \leq K\}} &\xrightarrow{\mathbb{P}} 1 \text{ für } K \rightarrow \infty \text{ und die Abbildung } x \mapsto \text{Var}(\langle v, X \rangle) x \text{ ist stetig} \\ &\Rightarrow \text{Var}(\langle v, X \rangle) \mathbb{1}_{\{\|X\| \leq K\}} \xrightarrow{\mathbb{P}} \text{Var}(\langle v, X \rangle) \cdot 1 \geq c > 0 \text{ für } K \rightarrow \infty \\ &\Rightarrow \exists K, c' > 0 : \text{Var}(\langle v, X \rangle) \mathbb{1}_{\{\|X\| \leq K\}} \geq c' \mathbb{P} - f.s. \end{aligned}$$

Seien nun solche $K, c' > 0$ gegeben. Wenden wir uns schließlich der Ungleichung aus 3.6 zu. Wir erinnern uns, dass nach Voraussetzung $\alpha \in \mathcal{B}_A$. Sei $M := \frac{1}{(A+K)^3}$

$$\begin{aligned} \mathbb{E} \left[\frac{\|P_u(X - \alpha)\|^2}{\|X - \alpha\|^3} \right] &\geq \mathbb{E} \left[\frac{\langle v, X - \alpha \rangle^2}{\|X - \alpha\|^3} \right] \geq \mathbb{E} \left[\frac{\langle v, X - \alpha \rangle^2}{\|X - \alpha\|^3} \mathbb{1}_{\{\|X\| \leq K\}} \right] \\ &\geq \frac{1}{(A+K)^3} \mathbb{E} [\langle v, X - \alpha \rangle^2 \mathbb{1}_{\{\|X\| \leq K\}}] \\ &= M \cdot (\text{Var}(\langle v, X - \alpha \rangle \mathbb{1}_{\{\|X\| \leq K\}}) + \mathbb{E} [\langle v, X - \alpha \rangle \mathbb{1}_{\{\|X\| \leq K\}}]^2) \\ &= M \cdot (\text{Var}(\langle v, X \rangle \mathbb{1}_{\{\|X\| \leq K\}}) + \mathbb{E} [\langle v, X - \alpha \rangle \mathbb{1}_{\{\|X\| \leq K\}}]^2) \\ &\geq M \cdot (c' + 0) = Mc' =: c_A > 0 \end{aligned}$$

Dabei ist die erste Gleichung der Verschiebungssatz und die Zweite die Translationsinvarianz der Varianz. Die letzten beiden Ungleichungen folgen aus oben bewiesenen Eigenschaften von c' . $\square$

Bemerkung 3.3.7. Als Spezialfall erhalten wir, dass es für den Median m zwei positive Konstanten $0 < c_m \leq C_m \leq \mathbb{E} [\|X - m\|^{-1}] < \infty$ gibt, sodass

$$c_m \|h\|^2 \leq \langle h, \Gamma_m h \rangle \leq C_m \|h\|^2.$$

Folglich ist Γ_m positiv definit und entsprechend invertierbar.

Korollar 3.3.8. G ist streng konvex in $\mathcal{H}$ (vgl. 3.3.1).

Beweis. Aus Stetigkeitsgründen ist die Hesse-Matrix von G überall streng positiv. Nach Lemma 3.3.2 ist also G streng konvex in $\mathcal{B}_A$. Da A beliebig groß gewählt werden kann, ist G auf ganz $\mathcal{H}$ streng konvex. $\square$

Korollar 3.3.9. G ist stark konvex auf jeder beschränkten Menge: Für $A > 0$ gibt es $\tilde{c}_A > 0$, sodass

$$\forall \alpha_1, \alpha_2 \in \mathcal{B}_A : \quad \langle \Phi(\alpha_2) - \Phi(\alpha_1), \alpha_2 - \alpha_1 \rangle \geq \tilde{c}_A \|\alpha_2 - \alpha_1\|^2.$$

Beweis. Sei $A > 0$ und $\alpha_1, \alpha_2 \in \mathcal{B}$.

$$\begin{aligned} \langle \Phi(\alpha_2) - \Phi(\alpha_1), \alpha_2 - \alpha_1 \rangle &= \lambda \langle \Phi_c(\alpha_2) - \Phi_c(\alpha_1), \alpha_2 - \alpha_1 \rangle \\ &\quad + (1 - \lambda) \langle \Phi_d(\alpha_2) - \Phi_d(\alpha_1), \alpha_2 - \alpha_1 \rangle \end{aligned}$$

Aus Lemma 3.2.11 folgt

$$\begin{aligned} \langle \Phi_d(\alpha_2) - \Phi_d(\alpha_1), \alpha_2 - \alpha_1 \rangle &= \langle \Phi_d(\alpha_2), \alpha_2 - \alpha_1 \rangle + \langle \Phi_d(\alpha_1), \alpha_1 - \alpha_2 \rangle \\ &\geq G(\alpha_2) - G(\alpha_1) + G(\alpha_1) - G(\alpha_2) \\ &= 0. \end{aligned}$$

Weshalb also $\langle \Phi(\alpha_2) - \Phi(\alpha_1), \alpha_2 - \alpha_1 \rangle \geq \lambda \langle \Phi_c(\alpha_2) - \Phi_c(\alpha_1), \alpha_2 - \alpha_1 \rangle + 0$ gilt. Definiert man $\alpha_t = (1 - t)\alpha_1 + t\alpha_2$ und $f(t) = \Phi(\alpha_t)$, so kann man $\Phi(\alpha_2) - \Phi(\alpha_1) = \int_0^1 f'(t) dt$ umschreiben. Setze weiterhin $u := \frac{\alpha_2 - \alpha_1}{\|\alpha_2 - \alpha_1\|}$.

Zwischenbehauptung.

$$f'(t) = \|\alpha_2 - \alpha_1\| \Gamma_{\alpha_t}(u)$$

Beweis

$$\begin{aligned} f'(t) &= \frac{d}{dt} f(t) = \frac{d}{dt} \Phi(\alpha_t) = \Phi'(\alpha_t) \cdot \left(\frac{d}{dt} \alpha_t \right) = \Gamma_{\alpha_t} \left(\frac{d}{dt} (1 - t)\alpha_1 + t\alpha_2 \right) \\ &= \Gamma_{\alpha_t}(\alpha_2 - \alpha_1) = \|\alpha_2 - \alpha_1\| \Gamma_{\alpha_t}(u) \end{aligned}$$

#

Deshalb

$$\begin{aligned}
 \langle \Phi_c(\alpha_2) - \Phi_c(\alpha_1), \alpha_2 - \alpha_1 \rangle &= \left\langle \int_0^1 f'(t) dt, \alpha_2 - \alpha_1 \right\rangle = \|\alpha_2 - \alpha_1\| \int_0^1 \langle f'(t), u \rangle dt \\
 &= \|\alpha_2 - \alpha_1\|^2 \int_0^1 \langle \Gamma_{\alpha_t}(u), u \rangle dt \geq \|\alpha_2 - \alpha_1\|^2 \int_0^1 c_A dt \\
 &= \|\alpha_2 - \alpha_1\|^2 c_A.
 \end{aligned}$$

Wobei c_A und die entsprechende Ungleichung aus Lemma 3.3.6 stammen. Insgesamt folgt

$$\begin{aligned}
 \langle \Phi(\alpha_2) - \Phi(\alpha_1), \alpha_2 - \alpha_1 \rangle &\geq \lambda \langle \Phi_c(\alpha_2) - \Phi_c(\alpha_1), \alpha_2 - \alpha_1 \rangle \\
 &\geq \lambda \|\alpha_2 - \alpha_1\|^2 c_A =: \tilde{c}_A \|\alpha_2 - \alpha_1\|^2.
 \end{aligned}$$

□

3.4 Grundeigenschaften

3.4.1 Existenz, Eindeutigkeit

[LM92] zeigte, dass die Existenz des geometrischen Medians ohne unsere Annahmen im Allgemeinen nicht gesichert ist, indem er ein Gegenbeispiel in einem Banachraum konstruiert hat.

Für den Beweis der Existenz mit unseren Annahmen orientieren wir uns an [Kem87] Lemma 2.3 und Theorem 3.6. Wir vereinfachen den dortigen Beweis, da wir es mit einem Hilbertraum statt eines Banachraums zu tun haben. Für die Eindeutigkeit benutzen wir bereits gezeigte Resultate.

Lemma 3.4.1. *Der geometrische Median m ist eindeutig.*

Beweis. Nach Lemma 3.3.8 ist G streng konvex. Nach Lemma 3.3.3 hat G dann höchstens ein Minimum. □

Man könnte die Existenz des Minimums von G auch aus Konvexität von G folgern. Wir geben hier einen anderen Beweis um ein besseres Verständnis für G zu erhalten und, weil wir Teile der Beweisschritte später nochmal benötigen.

Lemma 3.4.2. $\lim_{\|y\| \rightarrow \infty} \frac{G(y)}{\|y\|} = 1$

Beweis. Für alle $y \in \mathcal{H}$ gilt

$$\left| \frac{G(y)}{\|y\|} \right| = \left| \frac{\mathbb{E}[\|X - y\| - \|X\|]}{\|y\|} \right| \leq \frac{\mathbb{E}[\|X - y\| - \|X\|]}{\|y\|} \leq \frac{\mathbb{E}[\|y\|]}{\|y\|} = 1. \quad (3.7)$$

Setze für die andere Ungleichung $H(u) := \mathbb{P}(\{x \in \mathcal{H} : \|x\| \leq u\})$, $\tilde{H}(u) := \mathbb{P}(\{x \in \mathcal{H} : \|x\| > u\}) = 1 - H(u) \leq 1$. Dann ist $H(u)$ die Dichte der Abbildung $\|\cdot\|$ bezüglich $\mathbb{P}^X$. Setze $r := \|y\|$. Dann gilt

$$\begin{aligned}
 G(y) &= \mathbb{E}[\|X - y\| - \|X\|] = \int_{\mathcal{H}} \|x - y\| - \|x\| d\mathbb{P}^X(x) \\
 &\geq \int_{\mathcal{H}} \max\{\|y\| - 2\|x\|, -\|y\|\} d\mathbb{P}^X(x) = \int_0^\infty \max\{r - 2u, -r\} dH(u) \\
 &= \int_0^r \max\{r - 2u, -r\} dH(u) \geq \int_0^r (r - 2u) dH(u) = \int_0^r (r - 2u) H'(u) du \\
 &= [(r - 2u)H(u)]_0^r - \int_0^r -2H(u) du = -rH(r) + \int_0^r 2H(u) du \\
 &= \int_0^r 2H(u) - H(r) du \geq \int_0^r 2H(u) - 1 du = \int_0^r 1 - 2(1 - H(u)) du \\
 &= \int_0^r 1 - 2\tilde{H}(u) du.
 \end{aligned}$$

Da $\tilde{H}(u) \downarrow 0$ für $u \rightarrow \infty$, gibt es für jedes ε ein u_0 , sodass für alle $u \geq u_0$ gilt: $\tilde{H}(u) \leq \frac{\varepsilon}{2}$. Somit folgt

$$\begin{aligned}
 \frac{G(y)}{\|y\|} &\geq \frac{1}{r} \int_0^r 1 - 2\tilde{H}(u) du \geq \frac{1}{r} \int_{u_0}^r 1 - 2\tilde{H}(u) du \\
 &\geq \frac{1}{r} \int_{u_0}^r (1 - \varepsilon) du = \frac{r - u_0}{r} (1 - \varepsilon) = \left(1 - \frac{u_0}{r}\right) (1 - \varepsilon).
 \end{aligned}$$

Deshalb gilt

$$\lim_{\|y\| \rightarrow \infty} \frac{G(y)}{\|y\|} \geq \lim_{r \rightarrow \infty} \left(1 - \frac{u_0}{r}\right) (1 - \varepsilon) = 1 - \varepsilon.$$

Und weil ε beliebig klein gewählt werden kann, folgt $\lim_{\|y\| \rightarrow \infty} \frac{G(y)}{\|y\|} \geq 1$. $\square$

Lemma 3.4.3. *Es gilt: $G_* := \inf_{\alpha \in \mathcal{H}} G(\alpha) > -\infty$.*

Beweis. Sei $\{\alpha_n\}_{n \in \mathbb{N}}$ eine Folge in $\mathcal{H}$, sodass $\lim_{n \rightarrow \infty} G(\alpha_n) = G_*$ gilt. Ist $\{\alpha_n\}_{n \in \mathbb{N}}$ beschränkt, so ist es nach Gleichung 3.7 auch $\{G(\alpha_n)\}_{n \in \mathbb{N}}$ und somit auch G_* . Ist $\{\alpha_n\}_{n \in \mathbb{N}}$ unbeschränkt, so strebt $G(\alpha_n)$ nach Lemma 3.4.2 gegen ∞ . $\square$

Lemma 3.4.4. *Es gibt $r \in \mathbb{R}_{>0}^+$, sodass $G_* := \inf_{\alpha \in \mathcal{H}} G(\alpha) = \inf_{\alpha \in \mathcal{B}_r} G(\alpha)$. Somit kann sich der geometrischer Median wegen seiner Eindeutigkeit nur in $\mathcal{B}_r$ befinden.*

Beweis. Es gilt nach Lemma 3.4.3 $-\infty < G_* \leq G(0) = 0$. Setze $r := \sup\{\|y\| : G(y) \leq 0\}$. Da nach Lemma 3.4.2 $\lim_{\|y\| \rightarrow \infty} G(y) = \infty$, ist r endlich. Desweiteren hat r die Eigenschaft, dass aus $\|y\| > r$ folgt, dass $G(y) > 0$. Daraus folgt die Aussage. $\square$

Satz 3.4.5. *Sei $\mathcal{H}$ ein separabler Hilbertraum und X eine Zufallsvariable mit Werten in $\mathcal{H}$. Dann existiert ein geometrischer Median.*

Beweis. Sei r aus Lemma 3.4.4 gegeben. Nach Wahl von r genügt es zu zeigen, dass G aus 3.1 auf $\mathcal{B}_r$ ein Minimum annimmt. Man bemerke, dass im unendlichdimensionalen Fall $\mathcal{B}_r$ nicht kompakt in der Normtopologie ist. Nach Korollar 2.1.22 ist jedoch die schwache Topologie (vgl. 2.1.16) auf $\mathcal{B}_r$ kompakt und metrisierbar.

Sei $\{\alpha_n\}_{n \in \mathbb{N}}$ eine Folge in $\mathcal{B}_r$, sodass $\lim_{n \rightarrow \infty} G(\alpha_n) = G_*$ gilt. Da $\mathcal{B}_r$ kompakt und metrisierbar bezüglich der schwachen Topologie ist, können wir über eine Teilfolge ohne

Beschränkung der Allgemeinheit annehmen, dass $\{\alpha_n\}_{n \in \mathbb{N}}$ schwach konvergiert mit $\alpha_n \xrightarrow{w} m \in \mathcal{B}_r$. Nach Lemma 2.1.19 gilt für alle $x \in \mathcal{H}$, dass

$$\|m - x\| - \|x\| \leq \liminf_{n \rightarrow \infty} \|\alpha_n - x\| - \|x\|.$$

Mit Monotonie des Erwartungswerts und dem Lemma von Fatou folgt

$$\begin{aligned} G(m) &= \mathbb{E}[\|X - m\| - \|X\|] \leq \mathbb{E}[\liminf_{n \rightarrow \infty} \|X - \alpha_n\| - \|X\|] \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{E}[\|X - \alpha_n\| - \|X\|] = \liminf_{n \rightarrow \infty} G(\alpha_n) = G_*. \end{aligned}$$

Aber da $m \in \mathcal{B}_r$ gilt auch $G(m) \geq G_*$ und somit $G(m) = G_*$. Deshalb ist m ein geometrischer Median. $\square$

3.4.2 Robustheit

An dieser Stelle könnte man sich fragen, warum wir uns entschlossen haben uns auf den Median statt auf den Mittelwert zu konzentrieren. Letzterer scheint in sehr relevanten Kontexten die intuitive Wahl zu sein, da er z.B. der optimale Schätzer für die Lageparameter der Normalverteilung ist. Das liegt an folgendem entscheidenden Unterschied:

In vielen Anwendungen aus der realen Welt, sind gesammelte Daten mit Störungen verunreinigt und können z.B. wegen grober Messfehler große Ausreißer beinhalten. Diese Ausreißer automatisiert zu erkennen ist nicht einfach in Stichproben mit hoher Dimension und Stichprobengröße, wie z.B. in 2.3.1.

Der Durchschnitt $\bar{y}$ kann durch eine einzige Störung im Datensatz vollkommen durcheinander gebracht werden: Wenn irgendein Datenwert $y_i \rightarrow \pm\infty$, dann $\bar{y} \rightarrow \pm\infty$. Andererseits wird der Median durch einen einzelnen sich gegen $\pm\infty$ bewegendem Wert kaum beeinflusst. Der Median ist also resistent gegenüber groben Fehlern, während der Mittelwert es nicht ist. Tatsächlich kann der Median bis zu 50% Fehlerwerte tolerieren bevor er beliebig groß gemacht werden kann. Wir sagen, dass der **breakdown point** bei 50% liegt, während der des Mittelwerts bei 0% ist. Bemerkenswerterweise hat der Median somit den höchsten breakdown point, den ein Indikator haben kann. Denn intuitiv ist klar, dass wenn mehr als 50% der Beobachtungen verfälscht sind, der Schätzer nicht mehr zwischen der eigentlichen Verteilung und der verfälschten Verteilung unterscheiden kann. Diese Eigenschaft des Medians nennt sich **Robustheit** und macht es zum bevorzugten Indikator in realer Anwendung.

Wir möchten nun beweisen, dass der breakdown point bei 50% liegt. Wir orientieren uns dafür an [Ger08] Abschnitt 5. Zunächst müssen wir ihn formal definieren. Sei $\varepsilon \in [0, 1]$ vorgegeben. Betrachte die **Kontaminationsumgebung** von $\mathbb{P}$

$$\mathcal{N}_\varepsilon(\mathbb{P}) = \{(1 - \varepsilon)\mathbb{P} + \varepsilon\mathbb{Q} : \mathbb{Q} \text{ ist ein Wahrscheinlichkeitsmaß auf } \mathcal{H}\}.$$

Wir bezeichnen mit $M(\tilde{\mathbb{P}})$ immer den Median eines Wahrscheinlichkeitsmaßes $\tilde{\mathbb{P}}$. Der **maximale Bias** des Medians über $\mathcal{N}_\varepsilon(\mathbb{P})$ ist definiert als

$$B_M(\varepsilon) = \sup \left\{ \|M(\tilde{\mathbb{P}}) - M(\mathbb{P})\| : \tilde{\mathbb{P}} \in \mathcal{N}_\varepsilon(\mathbb{P}) \right\}.$$

Bemerke, dass $B_M(\varepsilon)$ nicht endlich sein muss. Tatsächlich wird es unendlich, wenn ε groß genug ist. Für $\varepsilon = 1$ wäre zum Beispiel $\tilde{\mathbb{P}}$ ein beliebiges Wahrscheinlichkeitsmaß und entsprechend $M(\tilde{\mathbb{P}})$ beliebig weit weg von $M(\mathbb{P})$. Definiere nun also den **breakdown point**

$$\varepsilon_M^* = \inf \{ \varepsilon : B_M(\varepsilon) = \infty \}$$

Satz 3.4.6. $\varepsilon_M^* = 0,5$

Beweis. Wenn $B_M(\varepsilon) = \infty$, dann gibt es eine Folge von Wahrscheinlichkeitsmaßen $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$, sodass für $\tilde{\mathbb{P}}_n := (1 - \varepsilon)\mathbb{P} + \varepsilon\mathbb{Q}_n$, $\|M(\tilde{\mathbb{P}}_n)\| \rightarrow \infty$ gilt. Setze $G_{\tilde{\mathbb{P}}}(y) := \int (\|x - y\| - \|x\|) \tilde{\mathbb{P}}(x)$ für ein Wahrscheinlichkeitsmaß $\tilde{\mathbb{P}}$ in Anlehnung an die zu minimierende Funktion G aus 3.1. Offenkundig gilt $G_{\mathbb{Q}_n}(y) \geq -\|y\|$ und nach 3.4.2 $\lim_{\|y\| \rightarrow \infty} \frac{G_{\mathbb{P}}(y)}{\|y\|} = 1$. Weiterhin gilt

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{G_{\tilde{\mathbb{P}}_n}(M(\tilde{\mathbb{P}}_n))}{\|M(\tilde{\mathbb{P}}_n)\|} &= \liminf_{n \rightarrow \infty} \frac{(1 - \varepsilon)G_{\mathbb{P}}(M(\tilde{\mathbb{P}}_n)) + \varepsilon G_{\mathbb{Q}_n}(M(\tilde{\mathbb{P}}_n))}{\|M(\tilde{\mathbb{P}}_n)\|} \\ &\geq (1 - \varepsilon) \liminf_{n \rightarrow \infty} \frac{G_{\mathbb{P}}(M(\tilde{\mathbb{P}}_n))}{\|M(\tilde{\mathbb{P}}_n)\|} - \varepsilon \\ &= 1 - 2\varepsilon. \end{aligned}$$

Aber, da $M(\tilde{\mathbb{P}}_n)$ das Minimum von $G_{\tilde{\mathbb{P}}_n}$ ist, gilt $G_{\tilde{\mathbb{P}}_n}(M(\tilde{\mathbb{P}}_n)) \leq G_{\tilde{\mathbb{P}}_n}(0) = 0$, weshalb $\varepsilon \geq \frac{1}{2}$. Somit ist $\varepsilon_M^* \geq \frac{1}{2}$. Wie oben ausgeführt, kann aber ε_M^* nicht größer als $\frac{1}{2}$ sein. $\square$

Diese Robustheit lässt sich auch mit der geometrischen Vorstellung 2.2 veranschaulichen: Wenn ein Median und die entsprechende Sphäre gegeben ist, hat ein einzelner sich entfernender Punkt keine Auswirkung auf seine Projektion.

Kapitel 4

Martingaldifferenz

Wir möchten in diesem Kapitel etwas Theorie im Zusammenhang mit dem bedingten Erwartungswerts und Martingalen, die wir für spätere Beweise benötigen, einführen. Der Großteil des in diesem Abschnitt eingeführten Materie ist Gegenstand einer weiterführenden Vorlesung in Wahrscheinlichkeitstheorie.

4.1 Konvergenzarten

Seien im Folgenden Y und Y_n für $n \in \mathbb{N}$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Wertebereich in einem metrischen Raum $(\mathcal{Y}, d)$.

Definition 4.1.1. Die Folge $\{Y_n\}_{n \in \mathbb{N}}$ konvergiert **$\mathbb{P}$ -fast sicher** gegen Y , falls $\mathbb{P}(\lim_{n \rightarrow \infty} d(Y_n, Y) = 0) = 1$ gilt. Wir schreiben

$$\lim_{n \rightarrow \infty} Y_n = Y \quad \mathbb{P}\text{-f.s. oder } Y_n \xrightarrow{\mathbb{P}\text{-f.s.}} Y.$$

Bemerkung 4.1.2. Angenommen $\lim_{n \rightarrow \infty} Y_n = Y$ $\mathbb{P}$ -f.s.. Dann gibt es also für jedes $\varepsilon > 0$, eine $\mathbb{P}^Y$ -Nullmenge N und eine natürliche Zahl $n_\varepsilon \in \mathbb{N}$, sodass

$$\forall n \geq n_\varepsilon : \{Y(\omega) : \omega \in \Omega\} \subseteq N \cup \{y \in \mathcal{Y} : \|Y_n(\omega) - y\| < \varepsilon\}.$$

Definition 4.1.3. Sei $p \geq 1$. Die Folge $\{Y_n\}_{n \in \mathbb{N}}$ konvergiert in $\mathcal{L}_p$ gegen Y , falls $\lim_{n \rightarrow \infty} \mathbb{E}[d(Y_n, Y)^p] = 0$. Wir schreiben

$$\lim_{n \rightarrow \infty} Y_n = Y \text{ in } \mathcal{L}_p \text{ oder } Y_n \xrightarrow{\mathcal{L}_p} Y.$$

Definition 4.1.4. Die Folge $\{Y_n\}_{n \in \mathbb{N}}$ konvergiert in **Verteilung** gegen Y , falls für alle reellen, stetigen, beschränkten Funktionen f $\lim_{n \rightarrow \infty} \mathbb{P}^{Y_n} f = \mathbb{P}^Y f$ gilt. Wir schreiben

$$\lim_{n \rightarrow \infty} Y_n = Y \text{ in Verteilung oder } Y_n \xrightarrow{d} Y.$$

Lemma 4.1.5. Aus Konvergenz in $\mathcal{L}_p$ folgt Konvergenz in Verteilung, d.h. aus $Y_n \xrightarrow{\mathcal{L}_p} Y$ folgt $Y_n \xrightarrow{d} Y$.

Beweis. Die Aussage sollte in der Einführung in die Wahrscheinlichkeitstheorie bewiesen worden sein. □

Definition 4.1.6. Die Folge $\{Y_n\}_{n \in \mathbb{N}}$ heißt **asymptotisch normalverteilt**, falls es eine normalverteilte Zufallsvariable $W \sim \mathcal{N}(\mu, \Sigma)$ und Folgen $\{a_n\}_{n \in \mathbb{N}}$ und $\{b_n\}_{n \in \mathbb{N}}$ gibt, sodass

$$\frac{Y_n - a_n}{b_n} \xrightarrow{d} W.$$

Wir schreiben kurz

$$\frac{Y_n - a_n}{b_n} \xrightarrow{d} \mathcal{N}(\mu, \Sigma).$$

4.2 Bedingter Erwartungswert

Wir führen den bedingten Erwartungswert ein und zeigen Rechenregeln, die wir benötigen. Der bedingte Erwartungswert ist Gegenstand einer weiterführenden Vorlesung in Wahrscheinlichkeitstheorie. Wir orientieren uns an Abschnitt 8 in [Kle06].

4.2.1 Definition

Sei im Folgenden immer $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{F} \subset \mathcal{A}$ eine Unter- σ -Algebra und $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ ein separabler Hilbertraum (vgl. 2.1.5). Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathcal{H}$ eine Zufallsvariable mit endlichem Erwartungswert.

Definition 4.2.1. Eine Zufallsvariable Y heißt **bedingte Erwartung** von X gegeben $\mathcal{F}$, in Zeichen $\mathbb{E}[X|\mathcal{F}] := Y$, falls

- i Y $\mathcal{F}$ -messbar ist,
- ii Für jedes $A \in \mathcal{F}$ gilt $\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A]$

Bemerkung 4.2.2. Aus der Definition folgt direkt, dass, falls X bezüglich $\mathcal{F}$ -messbar ist, $\mathbb{E}[X|\mathcal{F}] = X$ gilt. Zum Beispiel folgt aus $\sigma(X) \subseteq \mathcal{F}$, dass $\mathbb{E}[X|\mathcal{F}] = X$ gilt.

Satz 4.2.3. Der bedingte Erwartungswert $\mathbb{E}[X|\mathcal{F}]$ existiert.

Beweis. Setze $\mu^X : A \mapsto \mathbb{E}[X \mathbb{1}_A]$ für $A \in \mathcal{A}$. Dann ist μ^X offenkundig ein Maß. Da X nach Annahme eine endliche Erwartung hat, ist $\mu^X(\Omega) = \mathbb{E}[X \mathbb{1}_\Omega] = \mathbb{E}[X] < \infty$ und μ^X somit endlich. Es ist leicht einzusehen, dass $\mu^X \ll \mathbb{P}$ gilt: Sei N eine $\mathbb{P}$ -Nullmenge. Dann ist $\mu^X(N) = \mathbb{E}[X \mathbb{1}_N] = 0$. Sei h die natürliche Einbettung von $\mathcal{F}$ in $\mathcal{A}$. Dann sind $\mu^X \circ h = \mu^X|_{\mathcal{F}}$ und $\mathbb{P} \circ h = \mathbb{P}|_{\mathcal{F}}$ die natürlichen Restriktionen. Dann wird $\mu^X|_{\mathcal{F}}$ immer noch von $\mathbb{P}|_{\mathcal{F}}$ dominiert. Wir können also die Radon–Nikodym-Dichte nehmen

$$\mathbb{E}[X|\mathcal{F}] = \frac{d\mu^X|_{\mathcal{F}}}{d\mathbb{P}|_{\mathcal{F}}} = \frac{d(\mu^X \circ h)}{d(\mathbb{P} \circ h)}.$$

Dann ist $\mathbb{E}[X|\mathcal{F}]$ $\mathcal{F}$ -messbar und für $F \in \mathcal{F}$ gilt

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}] \mathbb{1}_F] = \int_F \mathbb{E}[X|\mathcal{F}] d\mathbb{P} = \int_F \frac{d\mu^X|_{\mathcal{F}}}{d\mathbb{P}|_{\mathcal{F}}} d\mathbb{P} = \int_F \mu^X = \mu^X(F) = \mathbb{E}[X \mathbb{1}_F].$$

□

Sei ferner auch $Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathcal{H}$ eine Zufallsvariable mit endlichem Erwartungswert.

Satz 4.2.4. (Linearität) Sei $\lambda \in \mathbb{R}$ ein Skalar. Der bedingte Erwartungswert ist linear, also $\mathbb{E}[\lambda X + Y | \mathcal{F}] = \lambda \mathbb{E}[X | \mathcal{F}] + \mathbb{E}[Y | \mathcal{F}]$

Beweis. Die rechte Seite ist $\mathcal{F}$ -messbar, und für $A \in \mathcal{F}$ ist

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A(\lambda \mathbb{E}[X | \mathcal{F}] + \mathbb{E}[Y | \mathcal{F}])] &= \lambda \mathbb{E}[\mathbb{1}_A \mathbb{E}[X | \mathcal{F}]] + \mathbb{E}[\mathbb{1}_A \mathbb{E}[Y | \mathcal{F}]] \\ &= \lambda \mathbb{E}[\mathbb{1}_A X] + \mathbb{E}[\mathbb{1}_A Y] \\ &= \mathbb{E}[\mathbb{1}_A(\lambda X + Y)]. \end{aligned}$$

□

Lemma 4.2.5. Sei X stochastisch unabhängig von $\mathcal{F}$ (d.h. $\sigma(X)$ und $\mathcal{F}$ sind unabhängig). Dann ist $\mathbb{E}[X | \mathcal{F}] = \mathbb{E}[X]$.

Beweis. Wir müssen die beiden Bedingungen aus der Definition des bedingten Erwartungswerts in 4.2.1 nachrechnen.

i Trivialerweise ist $\mathbb{E}[X]$ als Konstante messbar bezüglich $\mathcal{F}$.

ii Sei $A \in \mathcal{F}$. Dann sind X und $\mathbb{1}_A$ unabhängig. Es folgt

$$\mathbb{E}[\mathbb{E}[X] \mathbb{1}_A] = \mathbb{E}[X] \mathbb{E}[\mathbb{1}_A] = \mathbb{E}[X \mathbb{1}_A].$$

□

4.2.2 Eigenschaften reellwertiger Fall

In diesem Unterabschnitt rechnen wir einige Eigenschaften des bedingten Erwartungswerts für Zufallsvariablen mit Wertebereich in $\mathbb{R}$ nach. Wir werden sie später erweitern.

Lemma 4.2.6. Sei $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Dann ist $\mathbb{E}[X | \mathcal{F}]$ $\mathbb{P}$ -fast sicher eindeutig.

Beweis. Seien Y und Y' Zufallsvariablen, die Bedingungen i und ii aus der Definition des bedingten Erwartungswerts in 4.2.1 erfüllen. Setze $A := \{\|Y - Y'\| > 0\}$. Da Y und Y' nach i bezüglich $\mathcal{F}$ -messbar sind, ist $A \in \mathcal{F}$. Dann ist nach Bedingung ii aus Definition 4.2.1

$$0 = \mathbb{E}[X \mathbb{1}_A] - \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A] - \mathbb{E}[Y' \mathbb{1}_A] = \mathbb{E}[(Y - Y') \mathbb{1}_A].$$

Wegen $(Y - Y') \mathbb{1}_A \geq 0$, ist dann $\mathbb{P}(A) = 0$, also $Y \leq Y'$ $\mathbb{P}$ -fast sicher. Aus Symmetrie folgt $Y \geq Y'$ und somit $Y = Y'$. □

Lemma 4.2.7. (Monotonie) Seien X und Y reellwertige Zufallsvariablen. Ist $X \geq Y$ $\mathbb{P}$ -fast sicher, so ist $\mathbb{E}[X | \mathcal{F}] \geq \mathbb{E}[Y | \mathcal{F}]$ $\mathbb{P}$ -fast sicher.

Beweis. Sei $A := \{\mathbb{E}[X | \mathcal{F}] < \mathbb{E}[Y | \mathcal{F}]\} \in \mathcal{F}$. Wegen $X \geq Y$ ist $\mathbb{E}[\mathbb{1}_A(X - Y)] \geq 0$, also $\mathbb{P}(A) = 0$. □

Lemma 4.2.8. (Dreiecksungleichung) Sei X eine reellwertige Zufallsvariable. Dann gilt $\mathbb{E}[|X| | \mathcal{F}] \geq |\mathbb{E}[X | \mathcal{F}]|$.

Beweis. Seien $X^+, X^- \geq 0$ Zufallsvariablen mit $X = X^+ - X^-$ gegeben. Dann folgt aus der Monotonie und Linearität

$$\begin{aligned}\mathbb{E}[|X||\mathcal{F}] &= \mathbb{E}[X^+ + X^-|\mathcal{F}] = \mathbb{E}[X^+|\mathcal{F}] + \mathbb{E}[X^-|\mathcal{F}] \\ &\geq \max\{\mathbb{E}[X|\mathcal{F}], 0\} + \max\{-\mathbb{E}[X|\mathcal{F}], 0\} = |\mathbb{E}[X|\mathcal{F}]|.\end{aligned}$$

□

Satz 4.2.9. (Majorisierte Konvergenz) Sei $Y \in \mathcal{L}^1$ reellwertige Zufallsvariable mit $Y \geq 0$ und ist $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von reellwertigen Zufallsvariablen mit $|X_n| \leq Y$ für $n \in \mathbb{N}$. Gilt zusätzlich $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$ (vgl. 4.1.1), so gilt:

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}] \quad \mathbb{P}\text{-f.s.}.$$

Beweis. Seien alle Voraussetzungen aus dem Satz gegeben. Setze $Z_n := \sup_{k \geq n} |X_k - X|$. Dann ist $0 \leq Z_n \leq 2Y$ und $Z_n \xrightarrow{\text{f.s.}} 0$. Nach der majorisierten Konvergenz für den klassischen Erwartungswert gilt $\mathbb{E}[Z_n] \xrightarrow{n \rightarrow \infty} 0$, also nach der Dreiecksungleichung

$$\mathbb{E}[|\mathbb{E}[X_n|\mathcal{F}] - \mathbb{E}[X|\mathcal{F}]|] \leq \mathbb{E}[\mathbb{E}[|X_n - X||\mathcal{F}]] = \mathbb{E}[|X_n - X|] \leq \mathbb{E}[Z_n] \xrightarrow{n \rightarrow \infty} 0.$$

Dies entspricht aber der $\mathcal{L}_1$ -Konvergenz (vgl. 4.1.3). Da $\{Z_n\}_{n \in \mathbb{N}}$ monoton fallend ist, gilt dies nach der Monotonie auch für $\{\mathbb{E}[Z_n|\mathcal{F}]\}_{n \in \mathbb{N}}$. Da Z_n nichtnegativ ist, existiert $Z := \lim_{n \rightarrow \infty} \mathbb{E}[Z_n|\mathcal{F}]$. Nach dem Lemma von Fatou ist

$$\begin{aligned}\mathbb{E}[Z] &\leq \lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{E}[Z_n|\mathcal{F}]] = \lim_{n \rightarrow \infty} \mathbb{E}[Z_n] \\ &= \mathbb{E}\left[\lim_{n \rightarrow \infty} \sup_{k \geq n} |X_k - X|\right] = \mathbb{E}\left[\lim_{n \rightarrow \infty} |X_n - X|\right] = 0.\end{aligned}$$

Da Z nicht negativ ist, ist also $Z = 0$ und damit $\mathbb{E}[Z_n|\mathcal{F}] \xrightarrow{n \rightarrow \infty} 0$ $\mathbb{P}$ -fast sicher. Weiter gilt

$$|\mathbb{E}[X_n|\mathcal{F}] - \mathbb{E}[X|\mathcal{F}]| \leq \mathbb{E}[|X_n - X||\mathcal{F}] \leq \mathbb{E}\left[\sup_{k \geq n} |X_k - X||\mathcal{F}\right] = \mathbb{E}[Z_n|\mathcal{F}] \xrightarrow{n \rightarrow \infty} 0$$

Dabei ist die erste Ungleichung die Dreiecksungleichung aus 4.2.8 und die Zweite die Monotonie aus 4.2.7. □

Lemma 4.2.10. Sei $\{X_i\}_{i \in \mathbb{N}}$ eine Folge von reellwertigen Zufallsvariablen. Dann gilt

$$\mathbb{E}\left[\sum_{i \in \mathbb{N}} X_i|\mathcal{F}\right] = \sum_{i \in \mathbb{N}} \mathbb{E}[X_i|\mathcal{F}],$$

falls $\sum_{i \in \mathbb{N}} |X_i| \in \mathcal{L}^1$.

Beweis. Setze $Z_n := \sum_{i=1}^n X_i$ und $X := \sum_{i \in \mathbb{N}} X_i$. Dann $Z_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$ und $|Z_n| \leq \sum_{i \in \mathbb{N}} |X_i|$. Nach der majorisierten Konvergenz aus 4.2.9 gilt

$$\begin{aligned}\mathbb{E}\left[\sum_{i \in \mathbb{N}} X_i|\mathcal{F}\right] &= \mathbb{E}[X|\mathcal{F}] = \lim_{n \rightarrow \infty} \mathbb{E}[Z_n|\mathcal{F}] = \lim_{n \rightarrow \infty} \mathbb{E}\left[\sum_{i=1}^n X_i|\mathcal{F}\right] \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}[X_i|\mathcal{F}] = \sum_{i=1}^{\infty} \mathbb{E}[X_i|\mathcal{F}].\end{aligned}$$

□

Lemma 4.2.11. *Seien X und Y reellwertige Zufallsvariablen. Ist $\mathbb{E}[|XY|] < \infty$ und ist Y messbar bezüglich $\mathcal{F}$, dann ist*

$$\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}].$$

Beweis. Sei zunächst $X \geq 0$ und $Y \geq 0$. Für $n \in \mathbb{N}$ setze $Y_n = 2^{-n} \lfloor 2^n Y \rfloor$. Dann ist $Y_n \uparrow Y$ sowie $Y_n \mathbb{E}[X|\mathcal{F}] \uparrow Y \mathbb{E}[X|\mathcal{F}]$, da wegen der Monotonie aus Lemma 4.2.7 $\mathbb{E}[X|\mathcal{F}] \geq 0$ gilt. Es gilt nach dem Satz von der monotonen Konvergenz

$$\mathbb{E}[\mathbb{1}_A Y_n \mathbb{E}[X|\mathcal{F}]] \xrightarrow{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_A Y \mathbb{E}[X|\mathcal{F}]].$$

Andererseits ist

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A Y_n \mathbb{E}[X|\mathcal{F}]] &= \sum_{k=1}^{\infty} \mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{Y_n = k2^{-n}\}} k 2^{-n} \mathbb{E}[X|\mathcal{F}]] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{Y_n = k2^{-n}\}} k 2^{-n} X] \\ &= \mathbb{E}[\mathbb{1}_A Y_n X] \xrightarrow{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_A Y X]. \end{aligned}$$

Also gilt $\mathbb{E}[\mathbb{1}_A Y \mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[\mathbb{1}_A Y X]$. Im allgemeinen Fall schreiben wir $X = X^+ - X^-$ und $Y = Y^+ - Y^-$ und nutzen die Linearität der bedingten Erwartung aus. $\square$

Lemma 4.2.12. (Turmeigenschaft) *Sei X eine reellwertige Zufallsvariable und $\mathcal{G} \subset \mathcal{F} \subset \mathcal{A}$ σ -Algebren. Dann gilt $\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{G}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{F}] = \mathbb{E}[X|\mathcal{G}]$.*

Beweis. Die rechte Gleichung folgt aus Bemerkung 4.2.2, da $\mathbb{E}[X|\mathcal{G}]$ messbar bezüglich $\mathcal{F}$ ist. Sei nun $A \in \mathcal{G} \subseteq \mathcal{F}$, das heißt A ist insbesondere in $\mathcal{F}$.

$$\mathbb{E}[\mathbb{1}_A \mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{G}]] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{G}]].$$

$\square$

4.2.3 Eigenschaften mehrdimensionaler Fall

In diesem Abschnitt erweitern einige der Resultate aus dem letzten Abschnitt. Im folgenden ist der Wertebereich immer ein allgemeiner, separabler Hilbertraum $\mathcal{H}$.

Da nach Annahme der Hilbertraum $\mathcal{H}$ separabel ist, haben wir nach 2.1.11 eine Orthonormalbasis $\{e_i\}_{i \in \mathbb{N}}$ (vgl. 2.1.10). Sei im Folgenden $\{e_i\}_{i \in \mathbb{N}}$ immer diese Orthonormalbasis.

Bemerkung 4.2.13. Sei $h \in \mathcal{H}$ ein Vektor im Hilbertraum. Sei X eine Zufallsvariable mit Wertebereich in $\mathcal{H}$. Dann folgt aus der Linearität des bedingten Erwartungswerts (vgl. 4.2.4)

$$\mathbb{E}[\langle X, h \rangle | \mathcal{F}] = \langle \mathbb{E}[X|\mathcal{F}], h \rangle.$$

Lemma 4.2.14. *Sei X eine Zufallsvariable mit Wertebereich in $\mathcal{H}$. Dann ist $\mathbb{E}[X|\mathcal{F}]$ $\mathbb{P}$ -fast sicher eindeutig.*

Beweis. Seien Y und Y' Zufallsvariablen, die Bedingungen i und ii aus der Definition des bedingten Erwartungswerts in 4.2.1 erfüllen. Setze $X_k := \langle X, e_k \rangle$, $Y_k := \langle Y, e_k \rangle$ und $Y'_k := \langle Y', e_k \rangle$. Dann sind Y_k und Y'_k beide bedingte Erwartungswerte bezüglich $\mathcal{F}$ für X_k , denn beide sind $\mathcal{F}$ -messbar und für jedes $A \in \mathcal{F}$ gilt

$$\mathbb{E}[X_k \mathbb{1}_A] = \langle \mathbb{E}[X \mathbb{1}_A], e_k \rangle = \langle \mathbb{E}[Y \mathbb{1}_A], e_k \rangle = \mathbb{E}[Y_k \mathbb{1}_A].$$

Dasselbe gilt für Y'_k . Da Y_k und Y'_k reellwertige Zufallsvariablen sind, gilt nach Lemma 4.2.6 $Y_k = Y'_k$ $\mathbb{P}$ -fast sicher. Da k beliebig war, muss nach Definition der Orthonormalbasis $Y = Y'$ $\mathbb{P}$ -fast sicher gelten. $\square$

Folgender Satz erweitert die Aussage aus 4.2.10.

Satz 4.2.15. *Sei $\{X_i\}_{i \in \mathbb{N}}$ eine Folge von Zufallsvariablen. Dann gilt*

$$\mathbb{E} \left[\sum_{i \in \mathbb{N}} X_i | \mathcal{F} \right] = \sum_{i \in \mathbb{N}} \mathbb{E}[X_i | \mathcal{F}]$$

falls für alle $k \in \mathbb{N}$ $\sum_{i \in \mathbb{N}} |\langle e_k, X_i \rangle| \in \mathcal{L}^1$ gilt.

Beweis. Definiere für $i, j \in \mathbb{N}$ $X_i^j := \langle e_j, X_i \rangle$. Dann ist X_i^j eine reellwertige Zufallsvariable. Sei $k \in \mathbb{N}$. Dann gilt

$$\begin{aligned} \left\langle e_k, \sum_{i \in \mathbb{N}} \mathbb{E}[X_i | \mathcal{F}] \right\rangle &= \sum_{i \in \mathbb{N}} \langle e_k, \mathbb{E}[X_i | \mathcal{F}] \rangle = \sum_{i \in \mathbb{N}} \mathbb{E}[\langle e_k, X_i \rangle | \mathcal{F}] = \sum_{i \in \mathbb{N}} \mathbb{E}[X_i^k | \mathcal{F}] \\ &= \mathbb{E} \left[\sum_{i \in \mathbb{N}} X_i^k | \mathcal{F} \right] = \mathbb{E} \left[\sum_{i \in \mathbb{N}} \langle e_k, X_i \rangle | \mathcal{F} \right] = \mathbb{E} \left[\left\langle e_k, \sum_{i \in \mathbb{N}} X_i \right\rangle | \mathcal{F} \right] \\ &= \left\langle e_k, \mathbb{E} \left[\sum_{i \in \mathbb{N}} X_i | \mathcal{F} \right] \right\rangle. \end{aligned}$$

Dabei benutzen wir in der zweiten Zeile Lemma 4.2.10. Alle anderen Gleichungen nutzen die Linearität des Skalarprodukts und des bedingten Erwartungswerts. Somit gilt für ein beliebiges $k \in \mathbb{N}$

$$\left\langle e_k, \sum_{i \in \mathbb{N}} \mathbb{E}[X_i | \mathcal{F}] - \mathbb{E} \left[\sum_{i \in \mathbb{N}} X_i | \mathcal{F} \right] \right\rangle = 0.$$

Nach der Definition der Orthonormalbasis (vgl. 2.1.10) folgt die Aussage. $\square$

Lemma 4.2.16. *Sei λ eine Zufallsvariable mit Werten in $\mathbb{R}$ und X eine Zufallsvariable mit Werten in $\mathcal{H}$. Sei $\mathbb{E}[|\lambda X|] < \infty$. Ist λ messbar bezüglich $\mathcal{F}$, dann ist $\mathbb{E}[\lambda X | \mathcal{F}] = \lambda \mathbb{E}[X | \mathcal{F}]$. Ist X messbar bezüglich $\mathcal{F}$, dann ist $\mathbb{E}[\lambda X | \mathcal{F}] = \mathbb{E}[\lambda | \mathcal{F}] X$.*

Beweis. Nach Lemma 2.1.13 können wir $X = \sum_{j \in \mathbb{N}} \langle X, e_j \rangle e_j =: \sum_{j \in \mathbb{N}} X^j e_j$ mit X_j , wie in Satz 4.2.15, schreiben. Somit gilt durch Satz 4.2.15

$$\mathbb{E}[\lambda X | \mathcal{F}] = \mathbb{E} \left[\lambda \sum_{j \in \mathbb{N}} X^j e_j | \mathcal{F} \right] = \sum_{j \in \mathbb{N}} \mathbb{E}[\lambda X^j | \mathcal{F}] e_j.$$

Fall 1: Ist λ messbar bezüglich $\mathcal{F}$, so folgt durch 4.2.11, X_j reellwertig ist

$$\sum_{j \in \mathbb{N}} \mathbb{E} [\lambda X^j | \mathcal{F}] e_j = \lambda \sum_{j \in \mathbb{N}} \mathbb{E} [X^j | \mathcal{F}] e_j = \lambda \mathbb{E} \left[\sum_{j \in \mathbb{N}} X^j e_j | \mathcal{F} \right] = \lambda \mathbb{E} [X | \mathcal{F}].$$

Fall 2: Ist X messbar bezüglich $\mathcal{F}$, so folgt durch 4.2.11

$$\sum_{j \in \mathbb{N}} \mathbb{E} [\lambda X^j | \mathcal{F}] e_j = \sum_{j \in \mathbb{N}} \mathbb{E} [\lambda | \mathcal{F}] X^j e_j = \mathbb{E} [\lambda | \mathcal{F}] X.$$

□

Lemma 4.2.17. Sei $g : \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}$ eine messbare Abbildung. Sei X unabhängig von $\mathcal{F}$ und Y messbar bezüglich $\mathcal{F}$. Wenn zusätzlich gilt, dass $\mathbb{E} [\mathbb{E} [\|g(X, y)\|] |_{y=Y}] < \infty$. Dann gilt für $\omega \in \Omega$

$$\mathbb{E} [g(X, Y) | \mathcal{F}] = \mathbb{E} [g(X, y)] |_{y=Y}.$$

Beweis.

$$\begin{aligned} \mathbb{E} [\|g(X, Y)\|] &= \int \|g(x, y)\| d\mathbb{P}^{(X, Y)}(x, y) \\ &= \iint \|g(x, y)\| d\mathbb{P}^X(x) d\mathbb{P}^Y(y) \\ &= \int \mathbb{E} [\|g(X, y)\|] d\mathbb{P}^Y(y) \\ &= \mathbb{E} [\mathbb{E} [\|g(X, y)\|] |_{y=Y}] < \infty. \end{aligned}$$

Die zweite Zeile benutzt, dass $X \perp\!\!\!\perp Y$ und den Satz von Fubini. Setze

$Z := \mathbb{E} [g(X, y)] |_{y=Y}$. Wir rechnen die Eigenschaften des bedingten Erwartungswerts nach:

1. Y $\mathcal{F}$ -messbar, g messbar $\Rightarrow \mathbb{E} [g(X, y)] |_{y=Y}$ $\mathcal{F}$ -messbar
2. Sei $F \in \mathcal{F}$ und $B := Y(F)$

$$\begin{aligned} \int_F \mathbb{E} [g(X, y)] |_{y=Y} d\mathbb{P} &= \int_B \mathbb{E} [g(X, y)] d\mathbb{P}^Y(y) \\ &= \int_B \int_{\mathcal{H}} g(x, y) d\mathbb{P}^X(x) d\mathbb{P}^Y(y) \\ &= \int_{\mathcal{X} \times B} g(x, y) d\mathbb{P}^{(X, Y)}(x, y) \\ &= \int_F g(X, Y) d\mathbb{P}. \end{aligned}$$

Dabei benutzen wir den Satz von Fubini in der dritten Zeile.

□

Wir möchten im Folgenden die Aussage 4.2.11 des klassischen Produkts auf Skalarprodukte in Hilberträumen erweitern.

Satz 4.2.18. Ist für alle $i \in \mathbb{N}$ $\mathbb{E} [\|\langle X, e_i \rangle \langle e_i, Y \rangle\|] < \infty$ und ist Y messbar bezüglich $\mathcal{F}$, dann ist

$$\mathbb{E} [\langle X, Y \rangle | \mathcal{F}] = \langle \mathbb{E} [X | \mathcal{F}], Y \rangle.$$

Beweis. Nach 2.1.14 gilt $\langle X, Y \rangle = \sum_{i \in \mathbb{N}} \langle X, e_i \rangle \langle e_i, Y \rangle$. Aus Satz 4.2.15 folgt

$$\mathbb{E}[\langle X, Y \rangle | \mathcal{F}] = \sum_{i \in \mathbb{N}} \mathbb{E}[\langle X, e_i \rangle \langle e_i, Y \rangle | \mathcal{F}].$$

Setzen wir $X_i := \langle X, e_i \rangle$ und $Y_i := \langle e_i, Y \rangle$, so sind X_i und Y_i reellwertige Zufallsvariablen. Nach Annahme gilt $\mathbb{E}[|X_i Y_i|] < \infty$. Weil Y bezüglich $\mathcal{F}$ -messbar ist, ist Y_i bezüglich $\mathcal{F}$ -messbar und wir können Lemma 4.2.11 anwenden

$$\begin{aligned} \mathbb{E}[\langle X, Y \rangle | \mathcal{F}] &= \sum_{i \in \mathbb{N}} \mathbb{E}[X_i Y_i | \mathcal{F}] = \sum_{i \in \mathbb{N}} Y_i \mathbb{E}[X_i | \mathcal{F}] = \sum_{i \in \mathbb{N}} Y_i \mathbb{E}[\langle X, e_i \rangle | \mathcal{F}] \\ &= \sum_{i \in \mathbb{N}} Y_i \langle \mathbb{E}[X | \mathcal{F}], e_i \rangle = \sum_{i \in \mathbb{N}} \langle \mathbb{E}[X | \mathcal{F}], e_i \rangle \langle e_i, Y \rangle = \langle \mathbb{E}[X | \mathcal{F}], Y \rangle. \end{aligned}$$

Wobei wir bei der vierten Gleichung Bemerkung 4.2.13 benutzen. Die letzte Gleichung ist wieder Satz 2.1.14. $\square$

Wir definieren für $u, v, h \in \mathcal{H}$ $u \otimes v$ durch $u \otimes v(h) := \langle u, h \rangle v$.

Lemma 4.2.19. Für $u, v \in \mathcal{H}$ gilt $\|u \otimes v\|_L \leq \|u\| \cdot \|v\|$.

Beweis.

$$\begin{aligned} \|u \otimes v\|_L &= \sup_{h \in \mathcal{H}} \frac{\|u \otimes v(h)\|}{\|h\|} = \sup_{h \in \mathcal{H}} \frac{\|\langle u, h \rangle v\|}{\|h\|} = \sup_{h \in \mathcal{H}} \frac{|\langle u, h \rangle| \|v\|}{\|h\|} \\ &\leq \sup_{h \in \mathcal{H}} \frac{\|u\| \|h\| \|v\|}{\|h\|} = \|u\| \cdot \|v\| \end{aligned}$$

$\square$

Lemma 4.2.20. Der Operator $\otimes$ genügt dem Distributivgesetz: Für $u, v, x \in \mathcal{H}$ gilt: $(u + v) \otimes x = u \otimes x + v \otimes x$ und $x \otimes (u + v) = x \otimes u + x \otimes v$.

Beweis. Sei $h \in \mathcal{H}$. Dann gilt

$$\begin{aligned} ((u + v) \otimes x)(h) &= \langle u + v, h \rangle x = \langle u, h \rangle x + \langle v, h \rangle x = (u \otimes x)(h) + (v \otimes x)(h), \\ (x \otimes (u + v))(h) &= \langle x, h \rangle (u + v) = \langle x, h \rangle u + \langle x, h \rangle v = x \otimes u(h) + x \otimes v(h). \end{aligned}$$

$\square$

Lemma 4.2.21. Sei Y messbar bezüglich $\mathcal{F}$. Dann gilt

$$\mathbb{E}[X \otimes Y | \mathcal{F}] = \mathbb{E}[X | \mathcal{F}] \otimes Y \quad \text{und} \quad \mathbb{E}[Y \otimes X | \mathcal{F}] = Y \otimes \mathbb{E}[X | \mathcal{F}].$$

Beweis.

$$\mathbb{E}[X \otimes Y | \mathcal{F}] = \mathbb{E}[\langle X, \cdot \rangle Y | \mathcal{F}] = \mathbb{E}[\langle X, \cdot \rangle | \mathcal{F}] Y = \langle \mathbb{E}[X | \mathcal{F}], \cdot \rangle Y = \mathbb{E}[X | \mathcal{F}] \otimes Y$$

Die zweite Gleichung folgt aus Lemma 4.2.16. Die dritte Gleichung folgt aus Linearität.

$$\mathbb{E}[Y \otimes X | \mathcal{F}] = \mathbb{E}[\langle Y, \cdot \rangle X | \mathcal{F}] = \langle Y, \cdot \rangle \mathbb{E}[X | \mathcal{F}] = Y \otimes \mathbb{E}[X | \mathcal{F}]$$

Die zweite Gleichung folgt aus Lemma 4.2.16. $\square$

4.3 Martingaldifferenz

Sei im Folgenden $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $I \subseteq \mathbb{R}$.

Definition 4.3.1. Eine Familie $\mathbb{F} = \{\mathcal{F}_t\}_{t \in I}$ von σ -Algebren mit $\mathcal{F}_t \subseteq \mathcal{F}$ für jedes $t \in I$, heißt **Filtrierung** falls $\mathcal{F}_s \subset \mathcal{F}_t$ für alle $s, t \in I$ mit $s \leq t$.

Sei im Folgenden $\mathbb{F}$ eine Filtrierung.

Definition 4.3.2. Eine Folge von Zufallsvariablen $X = \{X_t\}_{t \in I}$ heißt **adaptiert** an die Filtration $\mathbb{F}$, falls X_t für jedes $t \in I$ bezüglich $\mathcal{F}_t$ messbar ist.

Sei im Folgenden $X = \{X_t\}_{t \in I}$ eine adaptierte Folge von Zufallsvariablen.

Definition 4.3.3. Sei X reellwertig und für $t \in I$ gelte $\mathbb{E}[|X_t^+|] < \infty$. Dann heißt X **Supermartingal**, falls für alle $s, t \in I$ mit $t > s$ $\mathbb{E}[X_t | \mathcal{F}_s] \leq X_s$ gilt.

Definition 4.3.4. Sei $\{X_n\}_{n \in \mathbb{N}}$ eine an $\mathbb{F}$ adaptierte Folge von Zufallsvariablen. Dann heißt die Folge $\{X_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$ **Martingaldifferenz**, falls es die folgenden beiden Bedingungen erfüllt:

1. $\mathbb{E}[|X_n|] < \infty$,
2. $\mathbb{E}[X_n | \mathcal{F}_{n-1}] = 0$, f.s.

Definition 4.3.5. Sei $X = \{X_{n,k} : k, n \in \mathbb{N}\}$ ein Array von Zufallsvariablen mit Werten in $\mathcal{H}$ und $\mathbb{F} = \{\mathcal{F}_{n,k} : k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}, n \in \mathbb{N}\}$ ein Array von Unter- σ -Algebren von $\mathcal{A}$, sodass die Reihen jeweils eine Filtrierung bilden (d.h. für $k \in \mathbb{N}$ gilt $\mathcal{F}_{n,k-1} \subseteq \mathcal{F}_{n,k}$). Wir sagen X ist **adaptiert** an $\mathbb{F}$, falls für alle $n, k \in \mathbb{N}$ gilt, $X_{n,k}$ messbar bezüglich $\mathcal{F}_{n,k}$ ist.

Definition 4.3.6. Sei $(X, \mathbb{F})$ ein adaptierter Array. $(X, \mathbb{F})$ heißt **Martingaldifferenzarray**, falls $\mathbb{E}[|X_{n,k}|] < \infty$ und $\mathbb{E}[X_{n,k} | \mathcal{F}_{n,k-1}] = 0$

4.4 Sätze

Wir geben einige größere Sätze der Wahrscheinlichkeitstheorie an, die wir später benutzen werden. Die meisten werden hier nicht bewiesen, da es den Rahmen dieser Arbeit überschreiten würde.

Satz 4.4.1. (Martingalkonvergenzatz) *Ist Y ein von unten beschränktes Supermartingal, so gibt es eine $\mathcal{F}$ -messbare Zufallsvariable Y_∞ mit $\mathbb{E}[Y_\infty] \leq \mathbb{E}[Y_0]$ und $Y_n \xrightarrow{\mathbb{P}\text{-f.s.}} Y_\infty$.*

Beweis. Ein Beweis findet sich bei 11.5 in [Kle06]. □

Wir geben im Folgenden eine schwächere Form vom Robbins und Siemund Theorem an, weil es für unsere Zwecke genügt und wir nicht durch überflüssige Notation irritieren wollen. Dieses Theorem ist allgemein sehr wichtig im Zusammenhang mit stochastischer Approximation. Wir halten uns bei der Formulierung und dem Beweis an [Duf13, Theorem 1.3.12].

Satz 4.4.2. (Robbins und Siemund Theorem) Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathbb{F} := \{\mathcal{F}_n\}_{n \in \mathbb{N}}$ eine Filtrierung. Seien $\{V_n\}_{n \in \mathbb{N}}$ und $\{\chi_n\}_{n \in \mathbb{N}}$ Folgen von reellwertigen, nicht-negativen und an $\mathbb{F}$ adaptierte Zufallsvariablen, sodass $\sum_{k=1}^{\infty} \chi_k < \infty$ $\mathbb{P}$ -f.s. und

$$\mathbb{E}[V_{n+1} | \mathcal{F}_n] \leq V_n + \chi_n.$$

gilt. Dann konvergiert V_n $\mathbb{P}$ -f.s. gegen eine endliche Zufallsvariable V .

Beweis. Setze $Y_n := V_n - \sum_{k=1}^{n-1} \chi_k$. Dann ist Y_n ein Supermartingal, denn für $n, m \in \mathbb{N}$ mit $n > m$ gilt

$$\begin{aligned} \mathbb{E}[Y_n | \mathcal{F}_m] &= \mathbb{E}\left[V_n - \sum_{k=1}^{n-1} \chi_k | \mathcal{F}_m\right] = \mathbb{E}[V_n | \mathcal{F}_m] - \sum_{k=1}^{n-1} \mathbb{E}[\chi_k | \mathcal{F}_m] \\ &= \mathbb{E}[\mathbb{E}[V_n | \mathcal{F}_{n-1}] | \mathcal{F}_m] - \sum_{k=1}^{n-1} \mathbb{E}[\chi_k | \mathcal{F}_m] \\ &\leq \mathbb{E}[V_{n-1} + \chi_{n-1} | \mathcal{F}_m] - \sum_{k=1}^{n-1} \mathbb{E}[\chi_k | \mathcal{F}_m] \\ &= \mathbb{E}[V_{n-1} | \mathcal{F}_m] - \sum_{k=1}^{n-2} \mathbb{E}[\chi_k | \mathcal{F}_m] \leq \dots \\ &\leq \mathbb{E}[V_{m+1} | \mathcal{F}_m] - \sum_{k=1}^m \mathbb{E}[\chi_k | \mathcal{F}_m] \leq V_m - \sum_{k=1}^{m-1} \mathbb{E}[\chi_k | \mathcal{F}_m]. \end{aligned}$$

Dabei benutzen wir bei der ersten Gleichung die Voraussetzung und die Monotonie des bedingten Erwartungswerts. Bei der dritten Gleichung benutzen wir die Turmeigenschaft (vgl. 4.2.12). Danach folgt eine offensichtliche Induktion.

Nun gilt für $k \in \{1, \dots, m-1\}$ $\mathcal{F}_k \subseteq \mathcal{F}_m$ und somit ist χ_k bezüglich $\mathcal{F}_m$ messbar. Somit folgt mit Bemerkung 4.2.2 $\chi_k = \mathbb{E}[\chi_k | \mathcal{F}_m]$ und deshalb

$$\mathbb{E}[Y_n | \mathcal{F}_m] \leq V_m - \sum_{k=1}^{m-1} \mathbb{E}[\chi_k | \mathcal{F}_m] = V_m - \sum_{k=1}^{m-1} \chi_k = Y_m.$$

Wähle nun $\sum_{k=1}^{\infty} \chi_k < a < \infty$ (möglich nach Annahme). Dann ist Y_n von unten durch a beschränkt, weil alle V_n, χ_n nicht-negativ sind. Nach dem Martingalkonvergenzsatz (4.4.1) konvergiert deshalb Y_n gegen eine endliche Zufallsvariable Y_{∞} . Somit $V_n = Y_n + \sum_{k=1}^{n-1} \chi_k \xrightarrow{\mathbb{P}\text{-f.s.}} Y_{\infty} + \sum_{k=1}^{\infty} \chi_k =: V$. Damit konvergiert V_n gegen die endliche Zufallsvariable V . $\square$

Um die asymptotische Normalität unseres Algorithmus zu zeigen, ist Theorem 5.1 aus [Jak88] essentiell:

Satz 4.4.3. (Zentraler Grenzwertsatz für Martingaldifferenzarrays in Hilberträumen) Sei $\mathcal{H}$ ein Hilbertraum mit Orthonormalbasis $\{e_i\}_{i \in \mathbb{N}}$. Sei $(X, \mathbb{F})$ ein Martingaldifferenzarray (vgl. 4.3.6) mit Wertebereich in $\mathcal{H}$. Sei $\Sigma : \mathcal{H} \rightarrow \mathcal{H}$ ein linearer Operator und seien folgende Bedingungen erfüllt:

1.

$$\mathbb{E} \left(\sup_{1 \leq k \leq n} \|X_{n,k}\| \right) \rightarrow 0,$$

2. Für alle $i, j \in \mathbb{N}$:

$$\sum_{k=1}^n \langle X_{n,k}, e_i \rangle \langle X_{n,k}, e_j \rangle \xrightarrow{d} \langle \Sigma e_i, e_j \rangle,$$

3. Setze $r_N^2(x) = \sum_{i=N}^{\infty} \langle x, e_i \rangle^2$. Für alle $\varepsilon > 0$:

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{k=1}^n r_N^2(X_{nk}) > \varepsilon \right) = 0.$$

Dann gilt $\sum_{k=1}^n X_{n,k} \xrightarrow{d} \mathcal{N}(0, \Sigma)$.

Um die zweite Bedingung aus Satz 4.4.3 zu prüfen, wird Proposition 1.3.19 (i) aus [Duf13] nützlich. Die Proposition ist eine direkte Folgerung vom Chows Lemma (vgl. Theorem 1.3.18 aus [Duf13]). Den entsprechenden führen wir hier jedoch nicht aus, da es den Rahmen dieser Arbeit überschreiten würde.

Satz 4.4.4. *Sei $\mathbb{F}$ eine Filtrierung und $\{\eta_n\}_{n \in \mathbb{N}}$ eine adaptierte Folge von Zufallsvariablen. Angenommen:*

1. Für alle $n \in \mathbb{N}$ ist $\mathbb{E}[\eta_{n+1} | \mathcal{F}_n] = 0$,
2. $\sup_{n \in \mathbb{N}} \mathbb{E}[|\eta_{n+1}|^a | \mathcal{F}_n] < \infty$ für ein $1 < a \leq 2$

Dann gilt:

$$\frac{1}{n-1} \sum_{k=0}^{n-1} \eta_{k+1} \xrightarrow{\mathbb{P}\text{-f.s.}} 0$$

Kapitel 5

Algorithmus

In diesem Kapitel betrachten wir den Algorithmus und zeigen weisen verschiedene Konvergenzarten gegen den geometrischen Median.

5.1 Definition

In diesem Abschnitt definieren wir den Algorithmus, auf den bislang hingearbeitet wurde. Wir orientieren uns an Abschnitt 2.3 aus [CCZ13]. Wir geben auch verschiedene Darstellungen des Algorithmus mit Hilfe neuer Funktionen. Für diese Funktionen zeigen wir für uns nützliche Eigenschaften.

5.1.1 Algorithmus

Sei $\gamma_n := c_\gamma n^{-\alpha}$, wobei c_γ eine positive Konstante ist und $\alpha \in (\frac{1}{2}, 1)$. Im Folgenden werden wir γ_n als Schrittgröße oder Tuningparameter bezeichnen. Der Tuningparameter kontrolliert die Konvergenzgeschwindigkeit des Algorithmus. Die beste Wahl ist in der Praxis nicht bekannt.

Lemma 5.1.1.

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty$$

Beweis. Da c_γ , gilt

$$\begin{aligned} \sum_{n=1}^{\infty} \gamma_n &= \sum_{n=1}^{\infty} c_\gamma n^{-\alpha} = c_\gamma \sum_{n=1}^{\infty} n^{-\alpha} = \infty \text{ (da } \alpha < 1), \\ \sum_{n=1}^{\infty} \gamma_n^2 &= \sum_{n=1}^{\infty} (c_\gamma n^{-\alpha})^2 = c_\gamma^2 \sum_{n=1}^{\infty} n^{-2\alpha} < \infty \text{ (da } 2\alpha > 1). \end{aligned}$$

□

Seien nun $X_0, X_2, \dots, X_n$ unsere n Beobachtungen. Fixiere $M' \in \mathbb{R}$. Wir definieren den Algorithmus als eine rekursive Folge von Schätzern:

$$Z_0 = X_0 \mathbb{1}_{\{\|X_0\| \leq M'\}}, \tag{5.1}$$

$$Z_{n+1} = Z_n - \gamma_n U_{n+1}, \quad (5.2)$$

$$U_{n+1} = \begin{cases} 0 & X_{n+1} = Z_n \\ -\frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|} & X_{n+1} \neq Z_n. \end{cases} \quad (5.3)$$

Dieser Algorithmus kann als „stochastische Gradienten-Methode“ betrachtet werden. Die Schrittgröße γ_n reguliert die Konvergenzrate. Die Richtung U_{n+1} kann als Schätzung für den Gradienten Φ von G bei Z_n verstanden werden, wie in den folgenden beiden Lemmata gezeigt wird.

Wir werden auch im Folgenden den Fall $X_{n+1} \neq Z_n$ als „typischen Fall“ bezeichnen und häufiger annehmen. Im typischen Fall gilt

$$Z_{n+1} = Z_n + \gamma_n \frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|}. \quad (5.4)$$

Bemerkung 5.1.2. Da wir Z_1 beschränkt gewählt haben, gibt es eine Konstante $M > 0$, sodass $\mathbb{E} \left[\|Z_1 - m\|^2 \right] \leq M$ gilt.

Lemma 5.1.3. *Es gilt*

$$\mathcal{F}_n := \sigma(Z_0, \dots, Z_n) \subseteq \sigma(X_0, \dots, X_n).$$

Beweis. Wir geben einen Beweis durch vollständige Induktion.

„ $n = 0$ “:

Definiere in Anlehnung an Gleichung 5.1 $f : \mathcal{H} \rightarrow \mathcal{H}, x \mapsto x \mathbb{1}_{\{\|x\| \leq M'\}}$. Dann ist f eine offene Abbildung und es gilt $Z_0 = X_0 \mathbb{1}_{\{\|X_0\| \leq M'\}} = f(X_0)$. Daraus folgt $\sigma(Z_0) = \sigma(f(X_0)) = \sigma(X_0)$

„ $n \rightsquigarrow n + 1$ “:

Setze $f_n : \mathcal{H}^2 \rightarrow \mathcal{H}, (x, z) \mapsto z + \mathbb{1}_{\{x \neq z\}} \gamma_n \frac{x - z}{\|x - z\|}$, sodass $Z_{n+1} = f_n(X_{n+1}, Z_n)$ gilt. Offensichtlich ist f_n als Komposition messbarer Funktionen messbar und deshalb gilt $\sigma(Z_{n+1}) = \sigma(f_n(X_{n+1}, Z_n)) \subseteq \sigma(X_{n+1}, Z_n)$. Wegen der Induktionsvoraussetzung folgt

$$\sigma(Z_0, \dots, Z_{n+1}) \subseteq \sigma(Z_0, \dots, Z_n, X_{n+1}) = \sigma(X_0, \dots, X_{n+1}).$$

□

Lemma 5.1.4. *Im „typischen Fall“, d.h. $X_{n+1} \neq Z_n$ gilt*

$$\mathbb{E}[U_{n+1} | \mathcal{F}_n] = \Phi(Z_n) = \mathbb{E}[\Phi(Z_n) | \mathcal{F}_n] \quad \mathbb{P} - f.s..$$

Beweis. Offensichtlich ist Z_n messbar bezüglich $\mathcal{F}_n$. Da nach Lemma 5.1.3 $\mathcal{F}_n \subseteq \sigma(X_0, \dots, X_n)$ und X_{n+1} unabhängig von $\sigma(X_0, \dots, X_n)$ ist, ist X_{n+1} unabhängig von $\mathcal{F}_n$. Die Abbildung $g : \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}, (x, z) \mapsto \frac{x - z}{\|x - z\|}$ ist offensichtlich stetig. Für $\omega \in \Omega$

gilt

$$\begin{aligned}
 \mathbb{E}[U_{n+1}|\mathcal{F}_n](\omega) &= \mathbb{E}\left[-\frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|} \middle| \mathcal{F}_n\right](\omega) \\
 &= \mathbb{E}[-g(X_{n+1}, Z_n)|\mathcal{F}_n](\omega) \\
 &= \mathbb{E}[-g(X_{n+1}, Z_n(\omega))] \\
 &= -\mathbb{E}\left[\frac{X - Z_n(\omega)}{\|X - Z_n(\omega)\|}\right] \\
 &= \Phi(Z_n(\omega)) = \Phi(Z_n)(\omega).
 \end{aligned}$$

Die zweite Gleichung ist klar, da Φ stetig ist und Z_n messbar bezüglich $\mathcal{F}_n$ ist. In der dritten Gleichung benutzen wir Lemma 4.2.17, da die Bedingung des Lemmas erfüllt ist, da $\|g(x, y)\| = 1$. $\square$

Definieren wir nun ξ als den Fehler der Abschätzung von U_{n+1} bezüglich Φ ,

$$\xi_{n+1} = \Phi(Z_n) - U_{n+1}. \quad (5.5)$$

Bemerkung 5.1.5. Dank Lemma 5.1.4 wissen wir, dass

$$\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] = \mathbb{E}[\Phi(Z_n) - U_{n+1}|\mathcal{F}_n] = \Phi(Z_n) - \mathbb{E}[U_{n+1}|\mathcal{F}_n] = 0.$$

Daher ist die Folge $\{\xi_n\}_{n \in \mathbb{N}}$ eine Martingaldifferenz (vgl. Definition 4.3.4).

Wir können nun eine andere Darstellung des Algorithmus geben

$$Z_{n+1} = Z_n + \gamma_n(-\Phi(Z_n) + \xi_{n+1}). \quad (5.6)$$

Lemma 5.1.6. *Sei $x \in \mathcal{H}$ ein Vektor. Für $k' \neq k$ gilt dann*

$$\mathbb{E}[\langle \xi_{k'+1}, x \rangle \langle \xi_{k+1}, x \rangle] = 0.$$

Beweis. Ohne Beschränkung der Allgemeinheit nehmen wir $k' < k$ an.

$$\begin{aligned}
 \mathbb{E}[\langle \xi_{k'+1}, x \rangle \langle \xi_{k+1}, x \rangle] &= \mathbb{E}[\mathbb{E}[\langle \xi_{k'+1}, x \rangle \langle \xi_{k+1}, x \rangle | \mathcal{F}_k]] \\
 &= \langle \xi_{k'+1}, x \rangle \mathbb{E}[\mathbb{E}[\langle \xi_{k+1}, x \rangle | \mathcal{F}_k]] \\
 &= \langle \xi_{k'+1}, x \rangle \mathbb{E}[\langle \mathbb{E}[\xi_{k+1} | \mathcal{F}_k], x \rangle] \\
 &= \langle \xi_{k'+1}, x \rangle \mathbb{E}[\langle 0, x \rangle] \\
 &= 0
 \end{aligned}$$

Die zweite Gleichung benutzt, dass $\xi_{k'}$ messbar bezüglich $\mathcal{F}_k$ ist und wir deshalb Lemma 4.2.11 anwenden können. Die dritte Zeile ist Bemerkung 4.2.13. Die vierte Gleichung verwendet das Resultat aus Bemerkung 5.1.5. $\square$

Lemma 5.1.7.

$$\|\xi_{n+1}\| \leq 2 \quad \mathbb{P} - f.s.$$

Beweis.

$$\begin{aligned}
 \|\xi_{n+1}\| &= \|\Phi(Z_n) - U_{n+1}\| \leq \|\Phi(Z_n)\| + \|U_{n+1}\| \\
 &= \left\| \mathbb{E} \left[\frac{X - Z_n}{\|X - Z_n\|} \right] \right\| + \|U_{n+1}\| \leq \mathbb{E} \left[\left\| \frac{X - Z_n}{\|X - Z_n\|} \right\| \right] + \|U_{n+1}\| \\
 &= \mathbb{E}[1] + \|U_{n+1}\| = 1 + \|U_{n+1}\|
 \end{aligned}$$

Nach Definition gilt

$$\|U_{n+1}\| = \begin{cases} 0 & X_{n+1} = Z_n \\ \left\| \frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|} \right\| & X_{n+1} \neq Z_n \end{cases} = \begin{cases} 0 & X_{n+1} = Z_n \\ 1 & X_{n+1} \neq Z_n \end{cases} \leq 1. \quad (5.7)$$

Daraus folgt die Behauptung direkt. $\square$

Lemma 5.1.8. *Es gilt*

$$\mathbb{E} \left[\|\xi_{n+1}\|^2 | \mathcal{F}_n \right] \leq 1 - \|\Phi(Z_n)\|^2.$$

Und im typischen Fall (d.h. $X_{n+1} \neq Z_n$) sogar Gleichheit. Des weiteren gilt

$$\mathbb{E} \left[\|\xi_{n+1}\|^2 \right] \leq 1.$$

Beweis.

$$\begin{aligned}
 \mathbb{E} \left[\|\xi_{n+1}\|^2 | \mathcal{F}_n \right] &= \mathbb{E} \left[\|U_{n+1}\|^2 + \|\Phi(Z_n)\|^2 - 2\langle \Phi(Z_n), U_{n+1} \rangle | \mathcal{F}_n \right] \\
 &= \mathbb{E} \left[\|U_{n+1}\|^2 | \mathcal{F}_n \right] + \|\Phi(Z_n)\|^2 - 2\langle \Phi(Z_n), \mathbb{E}[U_{n+1} | \mathcal{F}_n] \rangle \\
 &= \mathbb{E} \left[\|U_{n+1}\|^2 | \mathcal{F}_n \right] + \|\Phi(Z_n)\|^2 - 2\langle \Phi(Z_n), \Phi(Z_n) \rangle \\
 &= \mathbb{E} \left[\|U_{n+1}\|^2 | \mathcal{F}_n \right] + \|\Phi(Z_n)\|^2 - 2\|\Phi(Z_n)\|^2 \\
 &= \mathbb{E} \left[\|U_{n+1}\|^2 | \mathcal{F}_n \right] - \|\Phi(Z_n)\|^2
 \end{aligned}$$

Wobei wir bei der zweiten Gleichung 4.2.18 benutzen. Die dritte Zeile ist Lemma 5.1.4.

Nach dem selbigen Argument, wie in Gleichung 5.7, gilt $\|U_{n+1}\|^2 \leq 1$.

$$\begin{aligned}
 \mathbb{E} \left[\|\xi_{n+1}\|^2 | \mathcal{F}_n \right] &= \mathbb{E} \left[\|U_{n+1}\|^2 | \mathcal{F}_n \right] - \|\Phi(Z_n)\|^2 \\
 &\leq \mathbb{E}[1 | \mathcal{F}_n] - \|\Phi(Z_n)\|^2 = 1 - \|\Phi(Z_n)\|^2 \leq 1
 \end{aligned}$$

In der zweiten Zeile benutzen wir die Monotonie des bedingten Erwartungswerts aus Lemma 4.2.7. Im typischen Fall kann auf die Fallunterscheidung verzichtet werden und es gilt automatisch Gleichheit. Für die zweite Aussage bemerken wir, dass direkt aus der Definition vom bedingten Erwartungswert (vgl. 4.2.1) folgt

$$\begin{aligned}
 \mathbb{E} \left[\|\xi_{n+1}\|^2 \right] &= \mathbb{E} \left[\mathbb{E} \left[\|\xi_{n+1}\|^2 | \mathcal{F}_n \right] \right] \\
 &\leq \mathbb{E} \left[1 - \|\Phi(Z_n)\|^2 \right] \\
 &\leq \mathbb{E}[1] \leq 1.
 \end{aligned}$$

$\square$

5.1.2 Andere Darstellungen

Wir möchten noch andere Darstellungsformen des Algorithmus einführen. Wir werden diese neue Darstellungen in den späteren beweisen benutzen. Setze

$$\begin{aligned}\alpha_k &:= \mathbf{I}_H - \gamma_k \Gamma_m, & \beta_n &:= \alpha_n \alpha_{n-1} \cdots \alpha_1, \\ \delta_n &:= \Phi(Z_n) - \Gamma_m(Z_n - m),\end{aligned}\tag{5.8}$$

wobei Γ_m die Hesse-Matrix an der Stelle m aus Abschnitt 3.2.3 ist und $\mathbf{I}_H$ wieder die Identität. Wegen Lemma 3.2.15 ist Γ_m beschränkt und deshalb sind die Operatoren wohldefiniert ($0 \cdot \infty$ kann nicht vorkommen).

Bemerkung 5.1.9. Weil $\mathbf{I}_H$ und Γ_m jeweils linear sind, sind es auch alle α_k . Entsprechend ist auch β_n linear.

Dann lässt sich die Darstellung aus Gleichung 5.6 mit Hilfe von Gleichung 5.8 noch weiter umschreiben

$$\begin{aligned}Z_{n+1} &= Z_n + \gamma_n \xi_{n+1} - \gamma_n \Phi(Z_n) \\ &= Z_n + \gamma_n \xi_{n+1} - \gamma_n (\Gamma_m(Z_n - m) + \delta_n).\end{aligned}\tag{5.9}$$

Und entsprechend auch

$$\begin{aligned}Z_{n+1} - m &= (Z_n - m) - \gamma_n \Gamma_m(Z_n - m) + \gamma_n \xi_{n+1} - \gamma_n \delta_n \\ &= \alpha_n (Z_n - m) + \gamma_n \xi_{n+1} - \gamma_n \delta_n.\end{aligned}\tag{5.10}$$

Da $\alpha_{n-1} \alpha_{n-2} \cdots \alpha_{k+1} = \beta_{n-1} \beta_k^{-1}$ gilt, folgt daher induktiv

$$\begin{aligned}Z_n - m &= \alpha_{n-1} (Z_{n-1} - m) + \gamma_{n-1} \xi_n - \gamma_{n-1} \delta_{n-1} \\ &= \alpha_{n-1} (\alpha_{n-2} (Z_{n-2} - m) + \gamma_{n-2} \xi_{n-1} - \gamma_{n-2} \delta_{n-2}) + \gamma_{n-1} \xi_n - \gamma_{n-1} \delta_{n-1} \\ &= \alpha_{n-1} \alpha_{n-2} (Z_{n-2} - m) + (\alpha_{n-1} \gamma_{n-2} \xi_{n-1} + \gamma_{n-1} \xi_n) \\ &\quad - (\alpha_{n-1} \gamma_{n-2} \delta_{n-2} + \gamma_{n-1} \delta_{n-1}) \\ &= \beta_{n-1} \beta_{n-3}^{-1} (Z_{n-2} - m) + \beta_{n-1} (\beta_{n-2}^{-1} \gamma_{n-2} \xi_{n-1} + \beta_{n-1}^{-1} \gamma_{n-1} \xi_n) \\ &\quad - \beta_{n-1} (\beta_{n-2}^{-1} \gamma_{n-2} \delta_{n-2} + \beta_{n-1}^{-1} \gamma_{n-1} \delta_{n-1}) \\ &= \dots \\ &= \beta_{n-1} (Z_1 - m) + \beta_{n-1} \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \xi_{k+1} - \beta_{n-1} \sum_{k=1}^{n-2} \gamma_k \beta_k^{-1} \delta_k \\ &=: \beta_{n-1} (Z_1 - m) + \beta_{n-1} M_n - \beta_{n-1} R_{n-1},\end{aligned}\tag{5.11}$$

mit

$$R_n := \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \delta_k \quad M_n := \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \xi_{k+1}.\tag{5.12}$$

Lemma 5.1.10. *Es gibt eine Konstante $\tilde{C}_m$, sodass für alle $k \in \mathbb{N}$*

$$\|\delta_k\| \leq \tilde{C}_m \|Z_k - m\|^2$$

gilt.

Beweis. Die Restgliedformel der 0-ten Taylor-Approximation von Φ an der Stelle Z_k um den Entwicklungspunkt m , gibt uns

$$\Phi(Z_k) = \Phi(m) + \int_0^1 \Gamma_{m+t(Z_k-m)}(Z_k - m) dt = \int_0^1 \Gamma_{m+t(Z_k-m)}(Z_k - m) dt.$$

Dabei verwenden wir, dass Γ die Ableitung von Φ ist und nach 3.2.13 $\Phi(m) = 0$ gilt. Deshalb kann man δ_k folgendermaßen darstellen:

$$\begin{aligned} \delta_k = \Phi(Z_k) - \Gamma_m(Z_k - m) &= \int_0^1 \Gamma_{m+t(Z_k-m)}(Z_k - m) dt - \int_0^1 \Gamma_m(Z_k - m) dt \\ &= \int_0^1 (\Gamma_{m+t(Z_k-m)} - \Gamma_m)(Z_k - m) dt \end{aligned}$$

Für $h, h' \in \mathcal{H}$ definieren wir die Abbildung

$$\begin{aligned} \varphi_{h,h'} : [0, 1] &\longrightarrow \mathcal{H} \\ t &\longmapsto \varphi_{h,h'}(t) := \Gamma_{m+th}(h') \\ &= \mathbb{E} \left[\frac{1}{\|X - (m + th)\|} \left(h' - \frac{\langle X - (m + th), h' \rangle (X - (m + th))}{\|X - (m + th)\|^2} \right) \right]. \end{aligned}$$

Setze weiterhin die Zufallsvariablen $U_h : [0, 1] \rightarrow \mathbb{R}_+$ und $V_{h,h'} : [0, 1] \rightarrow \mathcal{H}$ für $t \in [0, 1]$ mit

$$\begin{aligned} U_h(t) &:= \frac{1}{\|X - m - th\|} \\ V_{h,h'}(t) &:= h' - \frac{\langle X - m - th, h' \rangle (X - m - th)}{\|X - m - th\|^2}, \end{aligned}$$

sodass $\varphi_{h,h'}(t) = \mathbb{E} [U_h(t)V_{h,h'}(t)]$. Weil U_h und $V_{h,h'}$ jeweils differenzierbar sind, ist $\varphi_{h,h'}$ nach dem Satz von der majorisierten Konvergenz auch differenzierbar. Es gilt nach Produktregel und Dreiecksungleichung

$$\begin{aligned} \|\varphi'_{h,h'}(t)\| &= \|\mathbb{E} [U'_h(t)V_{h,h'}(t) + U_h(t)V'_{h,h'}(t)]\| \\ &\leq \mathbb{E} [|U'_h(t)| \|V_{h,h'}(t)\| + |U_h(t)| \|V'_{h,h'}(t)\|]. \end{aligned}$$

Wir können die einzelnen Edukte abschätzen:

$$\begin{aligned} U_h(t) &:= \frac{1}{\|X - m - th\|}, \\ |U'_h(t)| &\leq \frac{\|h\|}{\|X - m - th\|^2}, \\ \|V_{h,h'}(t)\| &\leq \|h'\| + \frac{|\langle X - m - th, h' \rangle| \|X - m - th\|}{\|X - m - th\|^2} \leq 2 \|h'\|. \end{aligned}$$

$$\begin{aligned} \|V'_{h,h'}(t)\| &= \left\| \frac{d}{dt} \frac{\langle X - m - th, h' \rangle (X - m - th)}{\|X - m - th\|^2} \right\| \\ &\leq \left\| \frac{\langle X - m - th, h' \rangle (-h)}{\|X - m - th\|^2} \right\| + \left\| \frac{\langle -h, h' \rangle (X - m - th)}{\|X - m - th\|^2} \right\| \\ &\quad + \left\| \frac{\langle X - m - th, h' \rangle (X - m - th) \cdot (-2\|h\|)}{\|X - m - th\|^3} \right\| \\ &= \frac{4\|h\|\|h'\|}{\|X - m - th\|} \end{aligned}$$

Dabei ist die erste Ungleichung die Dreiecksungleichung und die Zweite Cauchy-Schwarz. Setze $N_h := \{t \in [0, 1] : m + th \in D\}$ mit D aus Annahme 2. Offensichtlich ist N_h eine Nullmenge bezüglich des Lebesgue-Maßes. Wenn man alle Abschätzungen zusammensetzt, folgt für $t \in N_h^c$

$$\begin{aligned} \|\varphi_{h,h'}(t)\| &\leq \mathbb{E} [|U'_h(t)| \|V_{h,h'}(t)\| + |U_h(t)| \|V'_{h,h'}(t)\|] \\ &= \mathbb{E} \left[\frac{2 \|h'\| \|h\|}{\|X - m - th\|^2} + \frac{4 \|h\| \|h'\|}{\|X - m - th\|^2} \right] \\ &= 6 \|h\| \|h'\| \mathbb{E} \left[\frac{1}{\|X - m - th\|^2} \right] \leq 6 \|h\| \|h'\| C. \end{aligned}$$

Bei der letzten Ungleichung benutzen wir, dass nach Bemerkung 3.1.1 für ein $C < \infty$ uniform $\mathbb{E} \left[\frac{1}{\|X - m - th\|^2} \right] \leq C$ gilt, da $t \in N_h^c$ und somit $m + th \notin D$. Somit erhalten wir für ein beliebiges $h \in \mathcal{H}$

$$\begin{aligned} \|\Phi(m + h) - \Gamma_m(h)\| &\leq \int_0^1 \|\Gamma_{m+th}(h) - \Gamma_m(h)\| dt = \int_0^1 \|\varphi_{h,h}(t) - \varphi_{h,h}(0)\| dt \\ &= \int_0^1 \left\| \int_0^t \varphi'_{h,h}(t') dt' \right\| dt \leq \int_0^1 \int_0^t \|\varphi'_{h,h}(t')\| dt' dt \\ &= \int_0^1 \int_0^t \mathbb{1}_{N_h^c}(t') \|\varphi'_{h,h}(t')\| dt' dt \\ &\leq \int_0^1 \int_0^t \sup_{t' \in [0,t] \cap N_h^c} \|\varphi'_{h,h}(t')\| dt' dt \leq \int_0^1 \sup_{t \in [0,1] \cap N_h^c} \|\varphi'_{h,h}(t)\| dt \\ &\leq \int_0^1 6C \|h\|^2 dt = 6C \|h\|^2. \end{aligned}$$

Setzen wir $h = Z_k - m$, so folgt für $\tilde{C}_m := 6C$

$$\|\delta_k\| = \|\Phi(Z_k) - \Gamma_m(Z_k - m)\| \leq \tilde{C}_m \|Z_k - m\|^2.$$

□

5.1.3 Averaged Version

Unser zweiter Algorithmus ist eine Abwandlung des ersten, der den Durchschnitt aller Schätzungen bildet

$$\bar{Z}_{n+1} = \bar{Z}_n + \frac{1}{n+1} (Z_{n+1} - \bar{Z}_n), \quad (5.13)$$

mit $\bar{Z}_0 = 0$, sodass $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$.

Wir werden in den folgenden Abschnitten unter ausgewählten Konvergenzbegriffen der Statistik, die in Abschnitt 4.1 eingeführt wurden, den Algorithmus analysieren. Offenkundig sollte die Folge von Zufallsvariablen Z_n gegen den geometrischen Median m konvergieren. Wir werden präzisere bzw. stärkere Aussagen formulieren und beweisen.

5.2 $\mathbb{P}$ -fast sichere Konvergenz

In diesem Abschnitt möchten wir die $\mathbb{P}$ -fast sichere Konvergenz (vgl. 4.1.1) von Z_n gegen m herleiten. Wir orientieren uns an Abschnitt 6.2. aus [CCZ13].

Lemma 5.2.1. *Es gibt eine endliche Zufallsvariable $V \geq 0$, sodass*

$$\lim_{n \rightarrow \infty} \|Z_n - m\|^2 = V, \quad \mathbb{P}\text{-f.s.}$$

Beweis. Betrachten wir $V_n := \|Z_n - m\|^2$. Zur Erinnerung: $Z_{n+1} = Z_n - \gamma_n \Phi(Z_n) + \gamma_n \xi_{n+1}$ (vgl. 5.6). Nach dem Satz des Pythagoras folgt

$$V_{n+1} = \|Z_n - m - \gamma_n \Phi(Z_n)\|^2 + \gamma_n^2 \|\xi_{n+1}\|^2 + 2\gamma_n \langle \xi_{n+1}, Z_n - \gamma_n \Phi(Z_n) \rangle. \quad (5.14)$$

Wegen Satz 4.2.18 und, weil ξ_{n+1} eine Martingaldifferenz ist, gilt

$$\mathbb{E}[\langle \xi_{n+1}, Z_n - \gamma_n \Phi(Z_n) \rangle | \mathcal{F}_n] = 0.$$

Also

$$\begin{aligned} \mathbb{E}[V_{n+1} | \mathcal{F}_n] &= \|Z_n - m - \gamma_n \Phi(Z_n)\|^2 + \gamma_n^2 \mathbb{E}[\|\xi_{n+1}\|^2 | \mathcal{F}_n] \\ &\quad + \gamma_n \mathbb{E}[\langle \xi_{n+1}, Z_n - \gamma_n \Phi(Z_n) \rangle | \mathcal{F}_n] \\ &= \|Z_n - m - \gamma_n \Phi(Z_n)\|^2 + \gamma_n^2 \mathbb{E}[\|\xi_{n+1}\|^2 | \mathcal{F}_n] \\ &= \|Z_n - m\|^2 - 2\gamma_n \langle Z_n - m, \Phi(Z_n) \rangle \\ &\quad + \gamma_n^2 (\|\Phi(Z_n)\|^2 + \mathbb{E}[\|\xi_{n+1}\|^2 | \mathcal{F}_n]) \\ &= V_n - 2\gamma_n \langle Z_n - m, \Phi(Z_n) \rangle + \gamma_n^2. \end{aligned} \quad (5.15)$$

Wobei bei der letzten Gleichung benutzt wird, dass wir vom typischen Fall ausgehen können und somit nach Lemma 5.1.8 $\|\Phi(Z_n)\|^2 + \mathbb{E}[\|\xi_{n+1}\|^2 | \mathcal{F}_n] = 1$ gilt. Nach Korollar 3.3.9 ist der mittlere Summand nicht positiv, da

$$\langle Z_n - m, \Phi(Z_n) \rangle = \langle Z_n - m, \Phi(Z_n) - \Phi(m) \rangle \geq 0.$$

Insgesamt folgt also für alle n mit dem typischen Fall, dass $\mathbb{E}[V_{n+1} | \mathcal{F}_n] \leq V_n + \gamma_n^2$. Für $\chi_n := \gamma_n^2$ folgt also, da $\sum_n \chi_n = \sum_n \gamma_n^2 < \infty$ nach 5.1.1 gilt, aus dem Robbins und Siegmund Theorem (siehe 4.4.2), dass V_n $\mathbb{P}$ -f.s. gegen eine Zufallsvariable V konvergiert. Des Weiteren bemerken wir, dass

$$\begin{aligned} \mathbb{E}[V_{n+1}] &= \mathbb{E}[\mathbb{E}[V_{n+1} | \mathcal{F}_n]] \\ &= \mathbb{E}[V_n - 2\gamma_n \langle Z_n - m, \Phi(Z_n) \rangle + \gamma_n^2] \\ &= \mathbb{E}[V_n] + \gamma_n^2 - 2\gamma_n \mathbb{E}[\langle \Phi(Z_n), Z_n - m \rangle] \\ &= \dots \\ &= \mathbb{E}[V_0] + \sum_{k=1}^n \gamma_k^2 - 2 \sum_{k=1}^n \gamma_k \mathbb{E}[\langle \Phi(Z_k), Z_k - m \rangle]. \end{aligned} \quad (5.16)$$

gilt. Die erste Gleichung ist das Gesetz der totalen Wahrscheinlichkeit bzw. ein direktes Resultat von der Definition des bedingten Erwartungswertes. Die Zweite ist obige Gleichung

5.15 und die Letzte eine offensichtliche Induktion. Insgesamt folgt

$$\begin{aligned}
 \mathbb{E}[V_{n+1}] &= \mathbb{E}[V_0] + \sum_{k=1}^n \gamma_k^2 - 2 \sum_{k=1}^n \gamma_k \mathbb{E}[\langle \Phi(Z_k), Z_k - m \rangle] \\
 &\leq \mathbb{E}[V_0] + \sum_{k=1}^n \gamma_k^2 \leq \mathbb{E}[\|Z_0 - m\|^2] + \sum_{k=1}^{\infty} \gamma_k^2 \\
 &= \mathbb{E}[\|X_0 \mathbb{1}_{\{\|X_0\| \leq M'\}} - m\|^2] + \sum_{k=1}^{\infty} \gamma_k^2 < \infty.
 \end{aligned}$$

Weshalb auch $\mathbb{E}[V] < \infty$ erfüllt ist. $V \geq 0$ ist klar. $\square$

Satz 5.2.2. *Es gilt*

$$\lim_{n \rightarrow \infty} Z_n = m, \quad \mathbb{P}\text{-f.s.}$$

Beweis. Nach Definition 4.1.1 ist die Aussage äquivalent zu

$$\lim_{n \rightarrow \infty} \|Z_n - m\| = 0, \quad \mathbb{P}\text{-f.s.}$$

Wegen Stetigkeit genügt es zu zeigen, dass V aus obigem Lemma $\mathbb{P}$ -f.s. 0 entspricht. Wenn man 5.16 umstellt, erhält man

$$2 \sum_{k=1}^n \gamma_k \mathbb{E}[\langle \Phi(Z_k), Z_k - m \rangle] = \mathbb{E}[V_0] + \sum_{k=1}^n \gamma_k^2 - \mathbb{E}[V_{n+1}] \leq \mathbb{E}[V_0] + \sum_{k=1}^{\infty} \gamma_k^2 < \infty.$$

Da $\gamma_k \mathbb{E}[\langle \Phi(Z_k), Z_k - m \rangle]$ immer nicht negativ ist, konvergiert die Reihe auf der linken Seite. Insbesondere folgt

$$\sum_{n=1}^{\infty} \gamma_n \langle \Phi(Z_n), Z_n - m \rangle < +\infty \quad \mathbb{P}\text{-f.s.} \quad (5.17)$$

Wir wollen nun zeigen, dass das nur sein kann, wenn Z_n gegen m konvergiert. Sei $\varepsilon \in (0, 1)$ und definiere die Menge

$$\Omega_\varepsilon := \{\omega \in \Omega : \exists n_\varepsilon(\omega) \geq 1 : \forall n \geq n_\varepsilon(\omega) : \varepsilon^2 < V_n(\omega) < \varepsilon^{-2}\}.$$

Für $\omega \in \Omega_\varepsilon$ gilt per Definition

$$\begin{aligned}
 \forall n \geq n_\varepsilon(\omega) : \varepsilon^2 < V_n(\omega) = \|Z_n(\omega) - m\|^2 < \varepsilon^{-2} &\Rightarrow \varepsilon < \|Z_n(\omega) - m\| < \varepsilon \\
 &\Rightarrow \forall n \geq n_\varepsilon(\omega) : Z_n(\omega) \in \{\alpha \in \mathcal{H} : \varepsilon < \|\alpha - m\| < \varepsilon\}.
 \end{aligned} \quad (5.18)$$

Diese Relation können wir nun nutzen um die Reihe aus Gleichung 5.17 auf ω abzuschätzen

$$\begin{aligned}
 \sum_{n=1}^{\infty} \gamma_n \langle \Phi(Z_n(\omega)), Z_n(\omega) - m \rangle &\geq \sum_{n=n_\varepsilon}^{\infty} \gamma_n \langle \Phi(Z_n(\omega)), Z_n(\omega) - m \rangle \\
 &\geq \inf_{n \geq n_\varepsilon} \langle \Phi(Z_n(\omega)), Z_n(\omega) - m \rangle \left(\sum_{n=n_\varepsilon}^{\infty} \gamma_n \right) \\
 &\geq \inf_{\varepsilon < \|\alpha - m\| < \varepsilon^{-1}} \langle \Phi(\alpha), \alpha - m \rangle \left(\sum_{n=n_\varepsilon}^{\infty} \gamma_n \right) \\
 &= \inf_{\varepsilon < \|\alpha - m\| < \varepsilon^{-1}} \langle \Phi(\alpha) - \Phi(m), \alpha - m \rangle \left(\sum_{n=n_\varepsilon}^{\infty} \gamma_n \right) \\
 &\geq \inf_{\varepsilon < \|\alpha - m\| < \varepsilon^{-1}} c_m \|\alpha - m\|^2 \left(\sum_{n=n_\varepsilon}^{\infty} \gamma_n \right) \\
 &\geq c_m \varepsilon^2 \left(\sum_{n=n_\varepsilon}^{\infty} \gamma_n \right) = \infty.
 \end{aligned}$$

Wobei wir bei der ersten Ungleichung wieder benutzen, dass die Summanden nicht negativ sind. Die Dritte ist eben die obige Relation 5.18. $c_m > 0$ stammt aus Korollar 3.3.9. Das Ergebnis ist wegen Lemma 5.1.1 unendlich. Da dieser Term also für alle $\omega \in \Omega_\varepsilon$ unendlich ist, der Term auf der linken Seite in Gleichung 5.17 aber $\mathbb{P}$ -f.s. endlich ist, muss Ω_ε eine $\mathbb{P}$ -Nullmenge sein.

Führen wir nun alles zusammen um zu zeigen, dass $V = 0$ $\mathbb{P}$ -f.s. gilt. Wir wissen schon aus Lemma 5.2.1, dass $V \geq 0$ und es ein $\tilde{c} \geq 0$ gibt mit $V < \tilde{c}$ $\mathbb{P}$ -f.s.. Sei $\varepsilon > 0$ vorgegeben. Wir wollen zeigen, dass $V > \varepsilon$ eine $\mathbb{P}$ -Nullmenge ist. Sei $c := \min\{\varepsilon, \tilde{c}^{-1}\} > 0$.

$$\mathbb{P}(\varepsilon < V) = \mathbb{P}(\varepsilon < V < \tilde{c}) + \mathbb{P}(\tilde{c} \leq V) = \mathbb{P}(V \in (\varepsilon, \tilde{c})) + 0 \leq \mathbb{P}(V \in (c, c^{-1}))$$

Da $\lim_{n \rightarrow \infty} V_n = V$ $\mathbb{P}$ -f.s. gibt es nach Gleichung 4.1.2 eine $\mathbb{P}$ -Nullmenge N , sodass gilt

$$\begin{aligned}
 \{V \in (c, c^{-1})\} &= \{\omega \in \Omega : c < V(\omega) < c^{-1}\} \\
 &\subseteq \left\{ \omega \in \Omega : \exists n_\varepsilon \geq 1 : \forall n \geq n_\varepsilon(\omega) : \left(\frac{c}{2}\right) < V_n(\omega) < \left(\frac{c}{2}\right)^{-1} \right\} \cup N \\
 &= \Omega_{\sqrt{\frac{c}{2}}} \cup N,
 \end{aligned}$$

und deshalb $\mathbb{P}(\varepsilon < V) \leq \mathbb{P}(V \in (c, c^{-1})) \leq \mathbb{P}(\Omega_{\sqrt{\frac{c}{2}}}) + \mathbb{P}(N) = 0$. $\square$

5.3 $\mathcal{L}_2$ Konvergenzrate

In diesem Abschnitt möchten wir eine Konvergenzrate in Landau Notation für die Konvergenz von Z_n gegen m in $\mathcal{L}_2$ (vgl. 4.1.3) herleiten. Wir orientieren uns an Abschnitt 6.3 in [CCZ13].

Lemma 5.3.1. *Es gibt eine Orthonormalbasis $\{v_j\}_{j \in I}$ und eine Folge von positiven Eigenwerten $\{\lambda_j\}_{j \in I} \subset (0, \mathbb{E}[\|X - m\|^{-1}]]$ mit*

$$\forall j \in \mathbb{N} : \quad \Gamma_m v_j = \lambda_j v_j.$$

Und falls I unendlich ist

$$\lambda_j \xrightarrow{j \rightarrow \infty} \mathbb{E} [\|X - m\|^{-1}].$$

Beweis.

$$\begin{aligned} \Gamma_m &= \mathbb{E} \left[\frac{1}{\|X - m\|} \left(\mathbf{I}_H - \frac{(X - m) \otimes (X - m)}{\|X - m\|^2} \right) \right] \\ &= \mathbb{E} [\|X - m\|^{-1}] \mathbf{I}_H - \mathbb{E} \left[\frac{1}{\|X - m\|} \left(\frac{(X - m) \otimes (X - m)}{\|X - m\|^2} \right) \right] \\ &=: \mathbb{E} [\|X - m\|^{-1}] \mathbf{I}_H - \Delta_m. \end{aligned}$$

Nach Lemma 3.2.15 und 3.3.7 ist Γ_m ein beschränkter und positiv definiter Operator. Da $\mathbf{I}_H$ offensichtlich beschränkt ist, ist dann auch Δ_m beschränkt. Für $x, y \in \mathcal{H}$ gilt

$$\begin{aligned} \langle \Delta_m(x), y \rangle &= \left\langle \mathbb{E} \left[\left(\frac{(X - m) \otimes (X - m)}{\|X - m\|^3} \right) \right] (x), y \right\rangle \\ &= \mathbb{E} \left[\left(\frac{\langle X - m, x \rangle \langle X - m, y \rangle}{\|X - m\|^3} \right) \right] \\ &= \langle x, \Delta_m(y) \rangle, \end{aligned}$$

$$\langle \Delta_m(x), x \rangle = \mathbb{E} \left[\left(\frac{\langle X - m, x \rangle^2}{\|X - m\|^3} \right) \right] \geq 0.$$

Deshalb ist Δ_m selbstadjungiert und positiv semi-definit. Des Weiteren gilt

$$\begin{aligned} \sum_{j \in I} \langle e_j, \Delta_m e_j \rangle &= \sum_{j \in I} \mathbb{E} \left[\frac{\langle X - m, e_j \rangle^2}{\|X - m\|^3} \right] \\ &\leq \mathbb{E} \left[\frac{1}{\|X - m\|} \right] < \infty. \end{aligned}$$

Wobei wir in der zweiten Zeile Korollar 2.1.15 benutzen. Nach 2.1.25 ist Δ_m dann ein kompakter Operator. Nach dem Spektralsatz (vgl. 2.1.26) gibt es dann eine Orthonormalbasis $\{v_j\}_{j \in I}$ und eine Folge von positiven Eigenwerten $\{\lambda'_j\}_{j \in I}$ mit $\Delta_m(v_j) = \lambda'_j v_j$. Entsprechend gilt für $\lambda_j := \mathbb{E} [\|X - m\|^{-1}] - \lambda'_j$, dass $\Gamma_m(v_j) = \mathbb{E} [\|X - m\|^{-1}] v_j - \Delta_m(v_j) = (\mathbb{E} [\|X - m\|^{-1}] - \lambda'_j) v_j = \lambda_j v_j$. Weil Γ_m positiv definit und Δ_m positiv semi-definit ist, sind $0 \leq \lambda'_j, 0 < \lambda_j$ und somit $\lambda_j \in (0, \mathbb{E} [\|X - m\|^{-1}])$.

Falls I unendlich ist, folgt aus Satz 2.1.26 auch, dass $\lambda'_j \rightarrow 0$ und somit $\lambda_j \rightarrow \mathbb{E} [\|X - m\|^{-1}]$. $\square$

Aus Lemma 2.1.13 folgt dann, dass für alle $x \in \mathcal{H}$: $\Gamma_m x = \sum_{j \in I} \lambda_j \langle v_j, x \rangle v_j$. Um die Notation zu vereinfachen schreiben wir

$$\Gamma_m x = \sum_{\lambda \in \Lambda} \lambda \langle v_\lambda, x \rangle v_\lambda. \quad (5.19)$$

mit $\Lambda := \{\lambda_j\}_{j \in I}$ mit Mehrfachnennung. Nach Gleichung 5.11 gilt

$$Z_n - m = \beta_{n-1} (Z_1 - m) + \beta_{n-1} M_n - \beta_{n-1} R_{n-1}.$$

Wir möchten nun das zweite Moment der drei einzelnen Summanden abschätzen. Dazu wird die Folge von Funktion

$$f_n(x) = \prod_{k=1}^n (1 - \gamma_k x)$$

nützlich sein. Dann gilt

$$\begin{aligned} \beta_n x &= \prod_{k=1}^n \alpha_k(x) = \prod_{k=1}^n (x - \gamma_k \Gamma_m x) \\ &= \prod_{k=1}^n \left(\sum_{\lambda \in \Lambda} \langle e_\lambda, x \rangle v_\lambda - \gamma_k \left(\sum_{\lambda \in \Lambda} \lambda \langle e_\lambda, x \rangle v_\lambda \right) \right) \\ &= \sum_{\lambda \in \Lambda} \prod_{k=1}^n (1 - \gamma_k \lambda) \langle v_\lambda, x \rangle v_\lambda \\ &= \sum_{\lambda \in \Lambda} f_n(\lambda) \langle v_\lambda, x \rangle v_\lambda. \end{aligned} \tag{5.20}$$

Dabei benutzen wir in der zweiten Zeile Gleichung 5.19 und Lemma 2.1.13. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass n groß genug ist, sodass $\gamma_n \mathbb{E} [\|X - m\|^{-1}] < 1$ gilt, da $\{\gamma_n\}_{n \in \mathbb{N}}$ eine Nullfolge und $\mathbb{E} [\|X - m\|^{-1}]$ endlich ist. Dann ist f_n auf $(0, \mathbb{E} [\|X - m\|^{-1}])$ streng monoton und deshalb injektiv. Entsprechend ist f_n^{-1} wohldefiniert. Da alle einzelnen α_n invertierbar sind, sind auch alle β_n invertierbar und es gilt nach der selben Argumentation, wie in Gleichung 5.20 $\beta_n^{-1} x = \sum_{\lambda \in \Lambda} f_n^{-1}(\lambda) \langle v_\lambda, x \rangle v_\lambda$.

Lemma 5.3.2. *Es gibt eine Konstante c_1 , sodass*

$$\|\beta_{n-1} \beta_k^{-1}\|_L \leq c_1 \exp \left(-2 \frac{c_m c_\gamma}{1 - \alpha} (n^{1-\alpha} - (k+1)^{1-\alpha}) \right), \quad \|\beta_{n-1} \beta_k^{-1}\|_L \leq c_1.$$

Wobei $c_m > 0$ aus Bemerkung 3.3.7 stammt. Insbesondere gilt

$$\|\beta_{n-1}\|_L = \|\beta_{n-1} \beta_0^{-1}\|_L \leq c_1 \exp \left(-2 \frac{c_m c_\gamma}{1 - \alpha} (n^{1-\alpha} - 1) \right).$$

Beweis. Sei $\lambda_{\min} := \inf_{\lambda \in \Lambda} \lambda$ das Infimum der Eigenwerte von Γ_m . Wir wissen nach Bemerkung 3.3.7, dass $\lambda_{\min} \geq c_m > 0$. Nach Gleichung 5.20 gilt $\beta_n(x) = \sum_{\lambda \in \Lambda} f_n(\lambda) \langle v_\lambda, x \rangle v_\lambda$. Es gilt

$$\begin{aligned} \|\beta_{n-1} \beta_k^{-1}\|_L &= \sup_{\|x\|=1} \|\beta_{n-1} \beta_k^{-1}(x)\| \\ &= \sup_{\|x\|=1} \sum_{\lambda \in \Lambda} f_{n-1}(\lambda) \langle v_\lambda, \beta_k^{-1}(x) \rangle v_\lambda \\ &= \sup_{\|x\|=1} \sum_{\lambda \in \Lambda} f_{n-1}(\lambda) \left\langle v_\lambda, \sum_{\lambda' \in \Lambda} f_k^{-1}(\lambda') \langle v_{\lambda'}, x \rangle v_{\lambda'} \right\rangle v_\lambda \\ &= \sup_{\|x\|=1} \sum_{\lambda \in \Lambda} \sum_{\lambda' \in \Lambda} \frac{f_{n-1}(\lambda)}{f_k(\lambda')} \langle v_\lambda, x \rangle \langle v_{\lambda'}, v_{\lambda'} \rangle v_\lambda \\ &= \sup_{\|x\|=1} \sum_{\lambda \in \Lambda} \frac{f_{n-1}(\lambda)}{f_k(\lambda)} \langle v_\lambda, x \rangle v_\lambda. \end{aligned}$$

Dabei ist die zweite Zeile Lemma 2.1.13. Es folgt

$$\begin{aligned}
 \|\beta_{n-1}\beta_k^{-1}\|_L &= \sup_{\|x\|=1} \sum_{\lambda \in \Lambda} \sum_{\lambda' \in \Lambda} \frac{f_{n-1}(\lambda)}{f_k(\lambda')} \langle v_\lambda, x \rangle \langle v_\lambda, v_{\lambda'} \rangle v_\lambda \\
 &= \sup_{\|x\|=1} \sum_{\lambda \in \Lambda} \frac{f_{n-1}(\lambda)}{f_k(\lambda)} \langle v_\lambda, x \rangle v_\lambda \\
 &= \sup_{\lambda \in \Lambda} \frac{f_{n-1}(\lambda)}{f_k(\lambda)} \\
 &= \sup_{\lambda \in \Lambda} \frac{\prod_{j=1}^{n-1} (1 - \gamma_j \lambda)}{\prod_{j=1}^k (1 - \gamma_j \lambda)} \\
 &= \sup_{\lambda \in \Lambda} \prod_{j=k+1}^{n-1} (1 - \gamma_j \lambda) \\
 &\leq \prod_{j=k+1}^{n-1} (1 - \gamma_j \lambda_{inf}) \\
 &\leq \prod_{j=k+1}^{n-1} (1 - \gamma_j c_m).
 \end{aligned}$$

Dabei benutzen wir in den letzten beiden Zeilen, dass alle $\gamma_j > 0$ sind und entsprechend $x \mapsto 1 - \gamma_j x$ sowie $x \mapsto \prod_{j=k+1}^{n-1} (1 - \gamma_j x)$ monoton fallend sind. Da $\{\gamma_j\}_{j \in \mathbb{N}}$ eine Nullfolge ist, gibt es ein $k_0 \in \mathbb{N}$, sodass $k \geq k_0 \Rightarrow \gamma_j c_m \leq 1$ gilt. Ohne Beschränkung der Allgemeinheit dürfen wir $n > k_0$ annehmen.

Setze $c_1 := \max_{k \in \{1, \dots, k_0\}} \left\{ \prod_{j=k+1}^{k_0-1} (1 - \gamma_j c_m) \exp \left(c_m \sum_{j=k+1}^{k_0-1} \gamma_j \right) \right\}$. Wir nutzen im Folgenden die Notation $\prod_{j=a}^b a_j = \prod_{j=b}^a a_j^{-1}$ und $\sum_{j=a}^b a_j = - \sum_{j=b}^a a_j$. Dann gilt

$$\begin{aligned}
 \|\beta_{n-1}\beta_k^{-1}\|_L &\leq \prod_{j=k+1}^{n-1} (1 - \gamma_j c_m) \\
 &= \prod_{j=k+1}^{k_0-1} (1 - \gamma_j c_m) \prod_{j=k_0}^{n-1} (1 - \gamma_j c_m) \\
 &\leq c_1 \exp \left(-c_m \sum_{j=k+1}^{k_0-1} \gamma_j \right) \prod_{j=k_0}^{n-1} (1 - \gamma_j c_m) \\
 &= c_1 \exp \left(-c_m \sum_{j=k+1}^{k_0-1} \gamma_j \right) \exp \left(\log \left(\prod_{j=k_0}^{n-1} (1 - \gamma_j c_m) \right) \right) \\
 &= c_1 \exp \left(-c_m \sum_{j=k+1}^{k_0-1} \gamma_j \right) \exp \left(\sum_{j=k_0}^{n-1} \log (1 - \gamma_j c_m) \right).
 \end{aligned}$$

Die dritte Zeile folgt direkt aus der Wahl von c_1 .

$$\begin{aligned}
 \|\beta_{n-1}\beta_k^{-1}\|_L &\leq c_1 \exp\left(-c_m \sum_{j=k+1}^{k_0-1} \gamma_j\right) \exp\left(\sum_{j=k_0}^{n-1} \log(1 - \gamma_j c_m)\right) \\
 &\leq c_1 \exp\left(-c_m \sum_{j=k+1}^{k_0-1} \gamma_j\right) \exp\left(\sum_{j=k_0}^{n-1} -\gamma_j c_m\right) \\
 &= c_1 \exp\left(-c_m \sum_{j=k+1}^{k_0-1} \gamma_j\right) \exp\left(-c_m \sum_{j=k_0}^{n-1} \gamma_j\right) \\
 &= c_1 \exp\left(-c_m \sum_{j=k+1}^{n-1} \gamma_j\right).
 \end{aligned}$$

Die letzte Ungleichung benutzt, dass allgemein $\log(1+x) \leq x$ für $x \geq -1$ gilt und wir $\gamma_j c_m \leq 1$ für $k \geq k_0$ annehmen. Da immer $c_m, \gamma_j > 0$ gilt, sieht man an dieser Stelle direkt ein, dass $\|\beta_{n-1}\beta_k^{-1}\|_L \leq c_1 \exp(0) = c_1$ gilt.

$$\begin{aligned}
 \|\beta_{n-1}\beta_k^{-1}\|_L &\leq c_1 \exp\left(-c_m \sum_{j=k+1}^{n-1} \gamma_j\right) \\
 &= c_1 \exp\left(-c_m c_\gamma \sum_{j=k+1}^{n-1} j^{-\alpha}\right) \\
 &\leq c_1 \exp\left(-c_m c_\gamma \int_{k+1}^n t^{-\alpha} dt\right) \\
 &= c_1 \exp\left(-2 \frac{c_m c_\gamma}{1-\alpha} (n^{1-\alpha} - (k+1)^{1-\alpha})\right)
 \end{aligned}$$

Die zweite Zeile benutzt, dass allgemein $\sum_{k=1}^{n-1} k^{-\alpha} \geq \int_1^n t^{-\alpha} dt$ gilt, da die linke Seite eine Obersumme für das Integral auf der rechten Seite ist. $\square$

Korollar 5.3.3. *Es gibt eine Konstante C_2 , sodass*

$$\mathbb{E} \left[\|\beta_{n-1}(Z_1 - m)\|^2 \right] \leq C_2 \exp\left(-2 \frac{c_m c_\gamma}{1-\alpha} n^{1-\alpha}\right).$$

Insbesondere gilt $\mathbb{E} \left[\|\beta_{n-1}(Z_1 - m)\|^2 \right] = \mathcal{O} \left(\exp\left(-2 \frac{c_m c_\gamma}{1-\alpha} n^{1-\alpha}\right) \right) = \mathcal{O} \left(\frac{\log(n)}{n^\alpha} \right)$.

Beweis. Es gilt

$$\begin{aligned}
 \mathbb{E} \left[\|\beta_{n-1}(Z_1 - m)\|^2 \right] &\leq c_1 \|\beta_{n-1}\|_L^2 \mathbb{E} \left[\|Z_1 - m\|^2 \right] \\
 &\leq c_1 \exp\left(-2 \frac{c_m c_\gamma}{1-\alpha} (n^{1-\alpha} - 1)\right) \mathbb{E} \left[\|Z_1 - m\|^2 \right] \\
 &\leq c_1 M \exp\left(2 \frac{c_m c_\gamma}{1-\alpha}\right) \exp\left(-2 \frac{c_m c_\gamma}{1-\alpha} n^{1-\alpha}\right) \\
 &=: C_2 \exp\left(-2 \frac{c_m c_\gamma}{1-\alpha} n^{1-\alpha}\right).
 \end{aligned}$$

Dabei ist die zweite Zeile Lemma 5.3.2. Die vorletzte Zeile ist Bemerkung 5.1.2. $\square$

Sei nun $c_\alpha > 0$ beliebig und $l(n)$ so, dass

$$(l(n) + 1)^{1-\alpha} = n^{1-\alpha} - c_\alpha \log(n). \quad (5.21)$$

$l(n)$ wird im Folgenden eine nützliche Funktion sein um Summen aufzuspalten. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass c_α so gewählt ist, dass $l(n)$ eine natürliche Zahl ist, da es am Ende nur darum gehen wird, dass c_α groß genug für unsere Zwecke ist. Zunächst möchten wir eine wichtige Eigenschaft zeigen.

Lemma 5.3.4.

$$n - l(n) \sim \frac{c_\alpha}{1-\alpha} \log(n) n^\alpha, \quad \gamma_{l(n)} \sim c_\gamma n^{-\alpha}.$$

Und somit $(n - l(n))\gamma_{l(n)}^2 \sim \frac{c_\alpha c_\gamma^2}{1-\alpha} \log(n) n^{-\alpha}$. Insbesondere ist $(n - l(n))\gamma_{l(n)}^2$ beschränkt.

Beweis. Wähle $\tilde{l}(n)$ so, dass

$$\tilde{l}(n)^{1-\alpha} = n^{1-\alpha} - c_\alpha \log(n). \quad (5.22)$$

Dann gilt

$$\left(\lim_{n \rightarrow \infty} \frac{\tilde{l}(n)}{l(n)} \right)^{1-\alpha} = \lim_{n \rightarrow \infty} \frac{\tilde{l}(n)^{1-\alpha}}{l(n)^{1-\alpha}} = \lim_{n \rightarrow \infty} \frac{\tilde{l}(n)^{1-\alpha}}{(l(n) + 1)^{1-\alpha}} = \frac{n^{1-\alpha} - c_\alpha \log(n)}{n^{1-\alpha} - c_\alpha \log(n)} = 1.$$

und deshalb $\lim_{n \rightarrow \infty} \frac{\tilde{l}(n)}{l(n)} = 1$ womit $\tilde{l}(n) \sim l(n)$ folgt. Deshalb genügt es die beiden Behauptungen für $\tilde{l}(n)$ zu zeigen. Wenn man Gleichung 5.21 umstellt, erhält man $\tilde{l}(n) = {}^{1-\alpha}\sqrt{n^{1-\alpha} - c_\alpha \log(n)} = n {}^{1-\alpha}\sqrt{1 - c_\alpha \frac{\log(n)}{n^{1-\alpha}}}$. Setze nun $g(x) := c_\alpha \frac{\log(x)}{x^{1-\alpha}}$. Da $\alpha < 1$, gilt offensichtlich $\lim_{x \rightarrow \infty} g(x) = 0$. Es folgt

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n - \tilde{l}(n)}{\frac{c_\alpha}{1-\alpha} \log(n) n^\alpha} &= \lim_{n \rightarrow \infty} \frac{n - n {}^{1-\alpha}\sqrt{1 - c_\alpha \frac{\log(n)}{n^{1-\alpha}}}}{\frac{c_\alpha}{1-\alpha} \log(n) n^\alpha} \\ &= \lim_{n \rightarrow \infty} \frac{(1-\alpha)n^{1-\alpha}}{c_\alpha \log(n)} \left(1 - {}^{1-\alpha}\sqrt{1 - c_\alpha \frac{\log(n)}{n^{1-\alpha}}} \right) \\ &= \lim_{n \rightarrow \infty} (1-\alpha) \frac{1 - {}^{1-\alpha}\sqrt{1 - g(n)}}{g(n)} \\ &= \lim_{n \rightarrow \infty} \frac{(1 - g(n))^{\frac{1}{1-\alpha}-1} g'(n)}{g'(n)} \\ &= \lim_{n \rightarrow \infty} (1 - g(n))^{\frac{1}{1-\alpha}-1} = 1. \end{aligned}$$

Dabei benutzen wir in der dritten Zeile die Regel von de L'Hospital. Wenden wir uns nun

der zweiten Behauptung zu

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \frac{c_\gamma n^{-\alpha}}{\gamma_{\tilde{l}(n)}} &= \frac{n^{-\alpha}}{\tilde{l}(n)^{-\alpha}} \\
 &= \lim_{n \rightarrow \infty} \left(\frac{\tilde{l}(n)}{n} \right)^\alpha \\
 &= \lim_{n \rightarrow \infty} \left(1 - \frac{n - \tilde{l}(n)}{n} \right)^\alpha \\
 &= \lim_{n \rightarrow \infty} \left(1 - \frac{\frac{c_\alpha}{1-\alpha} \log(n) n^\alpha}{n} \frac{n - \tilde{l}(n)}{\frac{c_\alpha}{1-\alpha} \log(n) n^\alpha} \right)^\alpha \\
 &= (1 - 0 \cdot 1)^\alpha = 1.
 \end{aligned}$$

□

Lemma 5.3.5.

$$\mathbb{E} \left[\|\beta_{n-1} M_n\|^2 \right] \leq C \frac{\log(n)}{n^\alpha}$$

Beweis. Zur Erinnerung: $M_n = \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \xi_{k+1}$ (vgl. 5.12). Setze $M_n^\lambda := \langle v_\lambda, M_n \rangle$, sodass $M_n = \sum_{\lambda \in \Lambda} \langle v_\lambda, M_n \rangle v_\lambda = \sum_{\lambda} M_n^\lambda v_\lambda$.

$$\begin{aligned}
 \mathbb{E} \left[\left(M_n^\lambda \right)^2 \right] &= \mathbb{E} \left[\left\langle v_\lambda, \sum_{k=1}^{n-1} \gamma_k \beta_k^{-1} \xi_{k+1} \right\rangle^2 \right] \\
 &= \mathbb{E} \left[\left\langle v_\lambda, \sum_{k=1}^{n-1} \gamma_k \sum_{\lambda' \in \Lambda} f_k^{-1}(\lambda') \langle v_{\lambda'}, \xi_{k+1} \rangle v_{\lambda'} \right\rangle^2 \right] \\
 &= \mathbb{E} \left[\left(\sum_{k=1}^{n-1} \gamma_k f_k^{-1}(\lambda) \langle v_\lambda, \xi_{k+1} \rangle \right)^2 \right] \tag{5.23} \\
 &= \sum_{k, k'=1}^{n-1} \gamma_k \gamma_{k'} f_k^{-1}(\lambda) f_{k'}^{-1}(\lambda) \mathbb{E} [\langle v_\lambda, \xi_{k+1} \rangle \langle v_\lambda, \xi_{k'+1} \rangle] \\
 &= \sum_{k=1}^{n-1} \gamma_k^2 f_k^{-2}(\lambda) \mathbb{E} [\langle \xi_{k+1}, v_\lambda \rangle^2] \\
 &= \sum_{k=1}^{n-1} \gamma_k^2 \mathbb{E} [\langle f_k^{-1}(\lambda) \xi_{k+1}, v_\lambda \rangle^2]
 \end{aligned}$$

Dabei benutzen wir bei der dritten Gleichung, dass nach Lemma 5.1.6 für $k' \neq k$ $\mathbb{E} [\langle \xi_{k'+1}, v_\lambda \rangle \langle \xi_{k+1}, v_\lambda \rangle] = 0$ gilt. Somit erhalten wir

$$\begin{aligned}
 \mathbb{E} \left[\|\beta_{n-1} M_n\|^2 \right] &= \mathbb{E} \left[\left\| \sum_{\lambda \in \Lambda} f_{n-1}(\lambda) \langle v_\lambda, M_n \rangle v_\lambda \right\|^2 \right] \\
 &= \mathbb{E} \left[\sum_{\lambda \in \Lambda} |f_{n-1}(\lambda) \langle v_\lambda, M_n \rangle|^2 \right] \\
 &= \sum_{\lambda \in \Lambda} f_{n-1}^2(\lambda) \mathbb{E} \left[\left(M_n^\lambda \right)^2 \right] \\
 &= \sum_{\lambda \in \Lambda} f_{n-1}^2(\lambda) \sum_{k=1}^{n-1} \gamma_k^2 \mathbb{E} \left[\langle f_k^{-1}(\lambda) \xi_{k+1}, v_\lambda \rangle^2 \right].
 \end{aligned}$$

Die zweite Zeile ist Korollar 2.1.15. Die Vierte ist Gleichung 5.23 von oben.

$$\begin{aligned}
 \mathbb{E} \left[\|\beta_{n-1} M_n\|^2 \right] &= \sum_{\lambda \in \Lambda} f_{n-1}^2(\lambda) \sum_{k=1}^{n-1} \gamma_k^2 \mathbb{E} \left[\langle f_k^{-1}(\lambda) \xi_{k+1}, v_\lambda \rangle^2 \right] \\
 &= \sum_{k=1}^{n-1} \gamma_k^2 \mathbb{E} \left[\sum_{\lambda \in \Lambda} \langle f_{n-1}(\lambda) f_k^{-1}(\lambda) \xi_{k+1}, v_\lambda \rangle^2 \right] \\
 &= \sum_{k=1}^{n-1} \gamma_k^2 \mathbb{E} \left[\sum_{\lambda \in \Lambda} \langle \beta_{n-1} \beta_k^{-1} \xi_{k+1}, v_\lambda \rangle^2 \right] \\
 &= \sum_{k=1}^{n-1} \gamma_k^2 \mathbb{E} \left[\|\beta_{n-1} \beta_k^{-1} \xi_{k+1}\|^2 \right] \\
 &\leq \sum_{k=1}^{n-1} \gamma_k^2 \|\beta_{n-1} \beta_k^{-1}\|_L^2 \mathbb{E} \left[\|\xi_{k+1}\|^2 \right] \\
 &\leq \sum_{k=1}^{n-1} \gamma_k^2 \|\beta_{n-1} \beta_k^{-1}\|_L^2 \\
 &\leq c_1 \sum_{k=1}^{n-1} \gamma_k^2 \exp \left(-2 \frac{c_m c_\gamma}{1-\alpha} (n^{1-\alpha} - (k+1)^{1-\alpha}) \right).
 \end{aligned}$$

Die dritte Gleichheit folgt direkt aus Gleichung 5.20 und Lemma 2.1.13. Die Vierte ist wieder Lemma 2.1.15. Die zweite Ungleichung benutzt, dass nach Lemma 5.1.8 $\mathbb{E} \left[\|\xi_{k+1}\|^2 \right] \leq 1$. Die Letzte ist Lemma 5.3.2.

Wir möchten $c_1 \sum_{k=1}^{n-1} \gamma_k^2 \exp \left(-\frac{c_\gamma}{1-\alpha} (n^{1-\alpha} - (k+1)^{1-\alpha}) \right)$ an $l(n)$ (vgl. 5.21) in zwei Teilsommen aufteilen und jeweils abschätzen.

$$\begin{aligned}
 &c_1 \sum_{k=1}^{l(n)} \gamma_k^2 \exp \left(-\frac{c_\gamma}{1-\alpha} (n^{1-\alpha} - (k+1)^{1-\alpha}) \right) \\
 &\leq c_1 \sum_{k=1}^{l(n)} \gamma_k^2 \exp \left(-\frac{c_\gamma}{1-\alpha} (n^{1-\alpha} - (l(n)+1)^{1-\alpha}) \right) \\
 &= c_1 \sum_{k=1}^{l(n)} \gamma_k^2 \exp \left(-\frac{c_\gamma c_\alpha}{1-\alpha} \log(n) \right) \\
 &= c_1 n^{-\frac{c_\gamma c_\alpha}{1-\alpha}} \sum_{k=1}^{l(n)} \gamma_k^2 \leq c_1 n^{-\frac{c_\gamma c_\alpha}{1-\alpha}} \sum_{k=1}^{\infty} \gamma_k^2 =: C' n^{-\frac{c_\gamma c_\alpha}{1-\alpha}}.
 \end{aligned}$$

Dabei benutzen wir in der dritten Zeile Gleichung 5.21. Nach Lemma 5.1.1 wissen wir, dass $C' := \sum_{k=1}^{\infty} \gamma_k^2 < \infty$. Wir können c_α groß genug wählen, sodass $C' n^{-\frac{c_\gamma c_\alpha}{1-\alpha}} < n^{-42}$ für alle $n \in \mathbb{N}$.

Nun haben wir erfolgreich den linken Teil der Summe $\mathbb{E} \left[\|\beta_{n-1} M_n\|^2 \right]$ abgeschätzt. Wenden wir uns dem rechten Teil zu

$$\begin{aligned} c_1 \sum_{k=l(n)+1}^{n-1} \gamma_k^2 \exp \left(-\frac{c_\gamma}{1-\alpha} (n^{1-\alpha} - (k+1)^{1-\alpha}) \right) \\ \leq \sum_{k=l(n)+1}^{n-1} \gamma_k^2 \exp \left(-\frac{c_\gamma}{1-\alpha} (n^{1-\alpha} - n^{1-\alpha}) \right) \\ = c_1 \sum_{k=l(n)+1}^{n-1} \gamma_k^2 \exp(0) \leq c_1 \sum_{k=l(n)+1}^{n-1} \gamma_{l(n)}^2 \\ \leq c_1 (n - l(n) - 2) \gamma_{l(n)}^2 \leq c_1 (n - l(n)) \gamma_{l(n)}^2. \end{aligned}$$

In der zweiten Ungleichung benutzen wir, dass $k \mapsto \gamma_k = \frac{c_\gamma}{k^\alpha}$ monoton fallend ist. Um alles zusammenzuführen gilt, wenn c_α groß genug ist

$$\begin{aligned} \mathbb{E} \left[\|\beta_{n-1} M_n\|^2 \right] &\leq n^{-42} + c_1 (n - l(n)) \gamma_{l(n)}^2 \\ &\sim n^{-42} + \frac{c_1 c_\alpha c_\gamma^2}{1-\alpha} \log(n) n^{-\alpha} \\ &\leq C \frac{\log(n)}{n^\alpha}. \end{aligned}$$

Dabei benutzen wir in der zweiten Zeile Lemma 5.3.4. $\square$

Zur Erinnerung: $\delta_k := \Phi(Z_k) - \Gamma_m(Z_k - m)$. Sei $K > 0$ und $l(n)$ wie in Gleichung 5.21. Definiere die Folge von Ereignissen

$$\Omega_N = \left\{ \begin{array}{l} \forall n \geq N, \forall k \geq n - l(n) : \|Z_k(\omega) - m\| \leq 1/K, \\ \forall k : \|\delta_k(\omega)\| \leq N \end{array} \right\}. \quad (5.24)$$

Lemma 5.3.6. Ω_N ist monoton wachsend und es gibt eine Nullmenge M_0 , sodass $\Omega = \bigcup_{N \in \mathbb{N}} \Omega_N \cup M_0$.

Beweis. Dass Ω_N ist monoton wachsend, ist klar. Aus Gleichung 5.3.4 folgt, dass $\lim_{n \rightarrow \infty} n - l(n) = \infty$. k wird deshalb wegen $k \geq n - l(n)$ und $n \geq N$ beliebig groß, wenn man $N \in \mathbb{N}$ groß genug wählt. Da Z_k nach Satz 5.2.2 fast sicher gegen m konvergiert, gibt es also ein von K abhängiges $N \in \mathbb{N}$ und eine Nullmenge M_N , sodass $\|Z_k(\omega) - m\| \leq 1/K$ für $\omega \in M_N^c$ gilt. Da dann $\|Z_k(\omega) - m\| \leq 1/K$ erfüllt ist, gilt mit Korollar 5.1.10 $\|\delta_k\| \leq \tilde{C}_m \|Z_k - m\|^2 \leq \frac{\tilde{C}_m}{K^2}$ auf M_N^c . Offensichtlich gilt irgendwann $\frac{\tilde{C}_m}{K^2} \leq N$. Insgesamt gilt also für die Nullmenge $M_0 := \bigcup_{N \in \mathbb{N}} M_N$, dass $\Omega = \bigcup_{N \in \mathbb{N}} \Omega_N \cup M_0$. $\square$

Lemma 5.3.7. Wir können in Gleichung 5.21 ein c_α groß genug wählen, sodass für alle $n \geq 2$ (was wir ohne Beschränkung der Allgemeinheit annehmen werden) gilt

$$\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \leq n^{-21}.$$

Beweis. Es gilt

$$\begin{aligned}
 \sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L &\leq c_1 \sum_{k=1}^{l(n)} \gamma_k \exp \left(-2 \frac{c_m c_\gamma}{1-\alpha} (n^{1-\alpha} - (k+1)^{1-\alpha}) \right) \\
 &\leq c_1 \sum_{k=1}^{l(n)} \gamma_k \exp \left(-2 \frac{c_m c_\gamma}{1-\alpha} (n^{1-\alpha} - (l(n)+1)^{1-\alpha}) \right) \\
 &= c_1 \sum_{k=1}^{l(n)} \gamma_k \exp \left(-2 \frac{c_m c_\gamma}{1-\alpha} (c_\alpha \log(n)) \right) \\
 &= c_1 c_\gamma \sum_{k=1}^{l(n)} k^{-\alpha} n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \\
 &= c_1 c_\gamma n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma \sum_{k=2}^{l(n)} k^{-\alpha} n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \\
 &\leq c_1 c_\gamma n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma \sum_{k=2}^{l(n)} k^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \\
 &\leq c_1 c_\gamma n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma \sum_{k=2}^n k^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}}.
 \end{aligned}$$

Wir wollen über vollständige Induktion zeigen, dass mit der richtigen Wahl von c_α $c_1 c_\gamma n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma \sum_{k=2}^n k^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \leq n^{-21}$ gilt.

„ $n = 2$ “:

Offensichtlich kann man c_α in Abhängigkeit von c_γ, c_m und α groß genug wählen, sodass

$$c_1 c_\gamma n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma \sum_{k=2}^n k^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} = c_1 c_\gamma 2^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma 2^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \leq 2^{-21}$$

gilt.

„ $n \rightsquigarrow n+1$ “:

Wähle $c_\alpha \geq \frac{1-\alpha}{2c_m c_\gamma} \left(\frac{\log(c_1 c_\gamma)}{\log(2)} - \alpha \right)$. Dann gilt

$$\begin{aligned}
 c_\alpha &\geq \frac{1-\alpha}{2c_m c_\gamma} \left(\frac{\log(c_1 c_\gamma)}{\log(2)} - \alpha \right) \\
 &\geq \frac{1-\alpha}{2c_m c_\gamma} \left(\frac{\log(c_1 c_\gamma)}{\log(n+1)} - \alpha \right) \\
 &= \frac{1-\alpha}{-2c_m c_\gamma} \left(\alpha - \frac{\log(c_1 c_\gamma)}{\log(n+1)} \right) \\
 &\geq \frac{1-\alpha}{-2c_m c_\gamma} \left(\alpha + \frac{\log((n+1)^{-21} - n^{-21})}{\log(n+1)} - \frac{\log(c_1 c_\gamma)}{\log(n+1)} \right) \\
 &= \frac{1-\alpha}{-2c_m c_\gamma} \left(\alpha + \frac{\log((n+1)^{-21} - n^{-21}) - \log(c_1 c_\gamma)}{\log(n+1)} \right).
 \end{aligned}$$

Daraus folgt

$$\begin{aligned}
 c_\alpha &\geq \frac{1-\alpha}{-2c_m c_\gamma} \left(\alpha + \frac{\log((n+1)^{-21} - n^{-21}) - \log(c_1 c_\gamma)}{\log(n+1)} \right) \\
 &\Leftrightarrow -2 \frac{c_m c_\gamma c_\alpha}{1-\alpha} \leq \alpha + \frac{\log((n+1)^{-21} - n^{-21}) - \log(c_1 c_\gamma)}{\log(n+1)} \\
 &\Leftrightarrow -\alpha - 2 \frac{c_m c_\gamma c_\alpha}{1-\alpha} \leq \frac{\log((n+1)^{-21} - n^{-21}) - \log(c_1 c_\gamma)}{\log(n+1)} \\
 &\Leftrightarrow (n+1)^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \leq \frac{(n+1)^{-21} - n^{-21}}{c_1 c_\gamma} \\
 &\Leftrightarrow n^{-21} + c_1 c_\gamma (n+1)^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \leq (n+1)^{-21}.
 \end{aligned}$$

Aus der Induktionsvoraussetzung folgt also mit einer solchen Wahl von c_α

$$\begin{aligned}
 (n+1)^{-21} &\geq n^{-21} + c_1 c_\gamma (n+1)^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \\
 &\geq c_1 c_\gamma n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma \sum_{k=2}^n k^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma (n+1)^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} \\
 &= c_1 c_\gamma n^{-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}} + c_1 c_\gamma \sum_{k=2}^{n+1} k^{-\alpha-2 \frac{c_m c_\gamma c_\alpha}{1-\alpha}}.
 \end{aligned}$$

Daraus folgt die Behauptung direkt. □

Lemma 5.3.8. *Es gibt eine Folge von Konstanten $\{C_N\}_{N \in \mathbb{N}}$ mit*

$$\forall N \in \mathbb{N} : \quad \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|\beta_{n-1} R_n\|^2 \right] \leq \frac{C_N}{n^{42}} + \frac{C_N}{K^2} \sum_{k=l(n)+1}^n \gamma_k \mathbb{E} \left[\|Z_k - m\|^2 \right].$$

Beweis. Sei $N \in \mathbb{N}$ und $\omega \in \Omega_N$. Ohne Beschränkung der Allgemeinheit sei $n \geq N$. Dann können wir die Summe wieder an der Stelle $l(n)$ in Zwei aufteilen

$$\begin{aligned}
 \|\beta_{n-1} R_n(\omega)\|^2 &= \left\| \beta_{n-1} \sum_{k=1}^n \gamma_k \beta_k^{-1} \delta_k(\omega) \right\|^2 \\
 &\leq \left(\sum_{k=1}^n \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \|\delta_k(\omega)\| \right)^2 \\
 &= \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \|\delta_k(\omega)\| + \sum_{k=l(n)+1}^n \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \|\delta_k(\omega)\| \right)^2 \\
 &\leq \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L N + \sum_{k=l(n)+1}^n \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \tilde{C}_m \|Z_k(\omega) - m\|^2 \right)^2,
 \end{aligned}$$

Dabei benutzen wir in der vierten Zeile, dass nach Lemma 5.1.10 und, weil $\omega \in \Omega_N$,

$\|\delta_k\| \leq N, \tilde{C}_m \|Z_k - m\|^2$ gilt.

$$\begin{aligned} \|\beta_{n-1} R_n(\omega)\|^2 &\leq \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L N + \sum_{k=l(n)+1}^n \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \tilde{C}_m \|Z_k(\omega) - m\|^2 \right)^2 \\ &\leq 2N^2 \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \right)^2 \\ &\quad + 2\tilde{C}_m^2 \left(\sum_{k=l(n)+1}^n \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \|Z_k(\omega) - m\|^2 \right)^2 \\ &\leq 2N^2 \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \right)^2 + 2c_1^2 \tilde{C}_m^2 \left(\sum_{k=l(n)+1}^n \gamma_k \|Z_k(\omega) - m\|^2 \right)^2 \end{aligned}$$

Die zweite Zeile ist $(a+b)^2 \leq 2a^2 + 2b^2$. Die Dritte ist $\|\beta_{n-1} \beta_k^{-1}\|_L \leq c_1$ nach 5.3.2.

$$\begin{aligned} \|\beta_{n-1} R_n(\omega)\|^2 &\leq 2N^2 \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \right)^2 \\ &\quad + 2c_1^2 \tilde{C}_m^2 \left(\sum_{k=l(n)+1}^n \gamma_k \|Z_k(\omega) - m\|^2 \right) \left(\sum_{k=l(n)+1}^n \gamma_k \|Z_k(\omega) - m\|^2 \right) \\ &\leq 2N^2 \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \right)^2 \\ &\quad + 2c_1^2 \tilde{C}_m^2 \left(\sum_{k=l(n)+1}^n \gamma_k \frac{1}{K^2} \right) \left(\sum_{k=l(n)+1}^n \gamma_k \|Z_k(\omega) - m\|^2 \right) \\ &\leq 2N^2 \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \right)^2 \\ &\quad + 2 \frac{c_1^2 \tilde{C}_m^2}{K^2} (n - l(n)) \gamma_{l(n)} \sum_{k=l(n)+1}^n \gamma_k \|Z_k(\omega) - m\|^2. \end{aligned}$$

Die zweite Ungleichung folgt auch wieder daraus, dass $\omega \in \Omega_N$. Nach Lemma 5.3.7 gilt $\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \leq n^{-21}$. Deshalb $2N^2 \left(\sum_{k=1}^{l(n)} \gamma_k \|\beta_{n-1} \beta_k^{-1}\|_L \right)^2 \leq 2 \frac{N^2}{n^{42}}$. Wir haben schon in Lemma 5.3.4 gesehen, dass $(n - l(n)) \gamma_{l(n)}$ beschränkt ist. Entsprechend kann man $2\tilde{C}_m^2 (n - l(n)) \gamma_{l(n)}$ durch eine Konstante beschränken. Insgesamt gibt es eine Konstante C_N , sodass

$$\|\beta_{n-1} R_n(\omega)\|^2 \leq \frac{C_N}{n^{42}} + \frac{C_N}{K^2} \sum_{k=l(n)+1}^n \gamma_k \|Z_k(\omega) - m\|^2,$$

und entsprechend gilt

$$\mathbb{E} \left[\mathbb{1}_{\Omega_N} \|\beta_{n-1} R_n\|^2 \right] \leq \frac{C_N}{n^{42}} + \frac{C_N}{K^2} \sum_{k=l(n)+1}^n \gamma_k \mathbb{E} \left[\|Z_k - m\|^2 \right].$$

□

Satz 5.3.9. Für die Folge von Ereignissen $\{\Omega_N\}_{N \in \mathbb{N}}$ aus Gleichung 5.24 gibt es eine Folge Konstanten $\{C_N\}_{N \in \mathbb{N}}$ mit

$$\forall N \in \mathbb{N} : \quad \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_n - m\|^2 \right] \leq C_N \gamma_n \log \left(\sum_{k=1}^n \gamma_k \right) \leq C_N \frac{\log(n)}{n^\alpha}.$$

Beweis. Führen wir erst einmal alles zusammen, was wir bisher gesehen haben. Nach Gleichung 5.11 gilt

$$Z_n - m = \beta_{n-1} (Z_1 - m) + \beta_{n-1} M_n - \beta_{n-1} R_{n-1}.$$

Nach Lemma 5.3.3 gilt

$$\mathbb{E} \left[\|\beta_{n-1} (Z_1 - m)\|^2 \right] = \mathcal{O} \left(\frac{\log(n)}{n^\alpha} \right).$$

Nach Lemma 5.3.5 gilt

$$\mathbb{E} \left[\|\beta_{n-1} M_n\|^2 \right] \leq C \frac{\log(n)}{n^\alpha}.$$

Nach Lemma 5.3.8 gilt

$$\forall N \in \mathbb{N} : \quad \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|\beta_{n-1} R_n\|^2 \right] \leq \frac{C_N}{n^{42}} + \frac{C_N}{K^2} \sum_{k=l(n)+1}^n \gamma_k \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_k - m\|^2 \right].$$

Entsprechend folgt für alle $N \in \mathbb{N}$ und $n \geq N$ für neue Konstanten C, C'

$$\begin{aligned} \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_n - m\|^2 \right] &\leq 3 \left(\mathbb{E} \left[\|\beta_{n-1} (Z_1 - m)\|^2 \right] + \mathbb{E} \left[\|\beta_{n-1} M_n\|^2 \right] \right. \\ &\quad \left. + \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|\beta_{n-1} R_n\|^2 \right] \right) \\ &\leq \frac{C \log(n)}{n^\alpha} + \frac{C}{K^2} \sum_{k=l(n)+1}^n \gamma_k \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_k - m\|^2 \right] \\ &\leq \frac{C \log(n)}{n^\alpha} + \frac{C}{K^2} (n - l(n) - 1) \sup_{l(n) < k \leq n} \gamma_k \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_k - m\|^2 \right] \\ &\leq \frac{C \log(n)}{n^\alpha} + \frac{C}{K^2} (n - l(n)) \gamma_{l(n)} \sup_{l(n) < k \leq n} \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_k - m\|^2 \right] \\ &\leq \frac{C \log(n)}{n^\alpha} + \frac{C'}{K^2} \sup_{l(n) < k \leq n} \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_k - m\|^2 \right]. \end{aligned}$$

In der letzten Zeile benutzen wir wieder, dass $(n - l(n)) \gamma_{l(n)}$ nach Gleichung 5.3.4 beschränkt ist. Wenn man $u_n := \mathbb{E} \left[\mathbb{1}_{\Omega_N} \|Z_n - m\|^2 \right]$ setzt, ließt sich das als

$$u_n \leq \frac{C \log(n)}{n^\alpha} + \frac{C'}{K^2} \max_{l(n) < k \leq n} u_n.$$

Wähle $K^2 \geq 2C'$. Dann gilt

$$u_n \leq \frac{C \log(n)}{n^\alpha} + \frac{1}{2} \max_{l(n) < k \leq n} u_n. \tag{5.25}$$

Wir möchten nun mit vollständiger Induktion zeigen, dass, wenn wir N' groß genug und $C'' = 4C$ wählen,

$$\forall n \geq N' : \quad u_n \leq \frac{C'' \log(n)}{n^\alpha}$$

gilt. Der Induktionsanfang ist klar. Angenommen $u_k \leq \frac{C'' \log(k)}{k^\alpha}$ gilt für alle $k \leq n$. Wir müssen nun also $u_{n+1} \leq \frac{C'' \log(n+1)}{(n+1)^\alpha}$ zeigen. Mit Gleichung 5.25 folgt

$$\begin{aligned}
 u_{n+1} &\leq \frac{C \log(n+1)}{(n+1)^\alpha} + \frac{1}{2} \max_{l(n+1) < k \leq n+1} u_k \\
 &= \frac{C \log(n+1)}{(n+1)^\alpha} + \frac{1}{2} \max \left(u_{n+1}, \max_{l(n+1) < k \leq n} u_k \right) \\
 &\leq \frac{C \log(n+1)}{(n+1)^\alpha} + \frac{1}{2} \max \left(u_{n+1}, \max_{l(n+1) < k \leq n} \frac{C'' \log(k)}{k^\alpha} \right) \\
 &= \frac{C \log(n+1)}{(n+1)^\alpha} + \frac{1}{2} \max \left(u_{n+1}, \frac{C'' \log(l(n+1))}{l(n+1)^\alpha} \right).
 \end{aligned} \tag{5.26}$$

Dabei benutzen wir in der dritten die Induktionsvoraussetzung und in der letzten Zeile, dass $k \mapsto \frac{C'' \log(k)}{k^\alpha}$ monoton fallend ist. Nun führen wir eine Fallunterscheidung durch:

1. Fall: $u_{n+1} \geq \frac{C'' \log(l(n+1))}{l(n+1)^\alpha}$:

Dann erhalten wir direkt aus Gleichung 5.26

$$\begin{aligned}
 u_{n+1} &\leq \frac{C \log(n+1)}{(n+1)^\alpha} + \frac{1}{2} u_{n+1} \\
 \Leftrightarrow \frac{1}{2} u_{n+1} &\leq \frac{C \log(n+1)}{(n+1)^\alpha} \\
 \Leftrightarrow u_{n+1} &\leq \frac{2C \log(n+1)}{(n+1)^\alpha} \leq \frac{4C \log(n+1)}{(n+1)^\alpha} = \frac{C'' \log(n+1)}{(n+1)^\alpha}.
 \end{aligned}$$

2. Fall: $u_{n+1} < \frac{C'' \log(l(n+1))}{l(n+1)^\alpha}$:

Dann erhalten wir direkt aus Gleichung 5.26

$$u_{n+1} \leq \frac{C \log(n+1)}{(n+1)^\alpha} + \frac{1}{2} \frac{C'' \log(l(n+1))}{l(n+1)^\alpha}.$$

Nun gilt $l(n+1) \sim (n+1)$ (folgt leicht aus Lemma 5.3.4). Wenn wir also N' groß genug wählen, gilt $\frac{\log(l(n+1))}{l(n+1)^\alpha} \leq \frac{3 \log(n+1)}{2(n+1)^\alpha}$, da wir von $n \geq N'$ ausgehen. Somit erhalten wir

$$\begin{aligned}
 u_{n+1} &\leq \frac{C \log(n+1)}{(n+1)^\alpha} + \frac{3}{4} \frac{C'' \log(n+1)}{(n+1)^\alpha} \\
 &\leq 4C \frac{\log(n+1)}{(n+1)^\alpha} = C'' \frac{\log(n+1)}{(n+1)^\alpha}.
 \end{aligned}$$

Dies schließt die Induktion und somit den gesamten Beweis ab. □

Korollar 5.3.10.

$$\|Z_n - m\|^2 = O_{\mathbb{P}} \left(\frac{\log(n)}{n^\alpha} \right)$$

Beweis. Folgt direkt aus 5.3.9 □

Bemerkung 5.3.11. Man kann sogar eine bessere Konvergenzrate finden. Für $\alpha \in (\frac{1}{2}, 1)$ und $\alpha < \beta < 3\alpha - 1$ gilt

$$\begin{aligned}
 \mathbb{E} \left[\|Z_n - m\|^2 \right] &= O \left(\frac{1}{n^\alpha} \right), \\
 \mathbb{E} \left[\|Z_n - m\|^4 \right] &= O \left(\frac{1}{n^\beta} \right).
 \end{aligned}$$

Der Beweis findet sich in [CCG17] in Abschnitt 5.1.

5.4 Asymptotische Normalität

In diesem Abschnitt möchten wir die asymptotische Normalität (vgl. 4.1.6) von $\{Z_n\}_{n \in \mathbb{N}}$ nachweisen. Wir orientieren uns an Abschnitt 6.4 aus [CCZ13].

Zur Erinnerung: $\xi_{n+1} := \Phi(Z_n) - U_{n+1}$ ist eine Martingaldifferenz und $\delta_n := \Phi(Z_n) - \Gamma_m(Z_n - m)$. Setzen wir

$$T_n := Z_n - m, \quad \bar{T}_n := \bar{Z}_n - m \quad \text{und} \quad \tilde{M}_{n+1} := \sum_{k=1}^n \xi_{k+1},$$

kommen wir durch Äquivalenzumformungen der Gleichung aus 5.9 zu

$$\begin{aligned} Z_{n+1} &= Z_n + \gamma_n \xi_{n+1} - \gamma_n (\Gamma_m(Z_n - m) + \delta_n) \\ \Leftrightarrow \gamma_n \Gamma_m(Z_n - m) &= Z_n - Z_{n+1} + \gamma_n \xi_{n+1} - \gamma_n \delta_n \\ \Leftrightarrow \Gamma_m(Z_n - m) &= \frac{1}{\gamma_n} (Z_n - Z_{n+1}) + \xi_{n+1} - \delta_n \\ \Leftrightarrow \Gamma_m T_n &= \xi_{n+1} - \delta_n + \frac{1}{\gamma_n} (T_n - T_{n+1}). \end{aligned}$$

Summerien wir diese Gleichungen, kommen wir auf

$$\begin{aligned} n \Gamma_m \bar{T}_n &= \Gamma_m(n(\bar{Z}_n - m)) = \Gamma_m\left(\sum_{k=1}^n Z_k - \sum_{k=1}^n m\right) = \Gamma_m\left(\sum_{k=1}^n Z_k - m\right) \\ &= \Gamma_m\left(\sum_{k=1}^n T_k\right) = \sum_{k=1}^n \Gamma_m(T_k) = \sum_{k=1}^n \left(\xi_{k+1} - \delta_k + \frac{1}{\gamma_k} (T_k - T_{k+1})\right) \\ &= \tilde{M}_{n+1} - \sum_{k=1}^n \delta_k + \sum_{k=1}^n \frac{1}{\gamma_k} (T_k - T_{k+1}) \\ &= \tilde{M}_{n+1} - \sum_{k=1}^n \delta_k + \sum_{k=1}^n \frac{T_k}{\gamma_k} - \sum_{k=1}^n \frac{T_{k+1}}{\gamma_k} \\ &= \tilde{M}_{n+1} - \sum_{k=1}^n \delta_k + \frac{T_1}{\gamma_1} - \frac{T_{n+1}}{\gamma_n} + \sum_{k=2}^n \frac{T_k}{\gamma_k} - \sum_{k=2}^n \frac{T_k}{\gamma_{k-1}} \\ &= \tilde{M}_{n+1} - \sum_{k=1}^n \delta_k + \frac{T_1}{\gamma_1} - \frac{T_{n+1}}{\gamma_n} + \sum_{k=2}^n T_k \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}}\right). \end{aligned}$$

Wenn man beide Seiten durch $\sqrt{n}$ teilt, erhalten wir

$$\sqrt{n} \Gamma_m \bar{T}_n = \frac{1}{\sqrt{n}} \tilde{M}_{n+1} + \frac{1}{\sqrt{n}} \left(\frac{T_1}{\gamma_1} - \frac{T_{n+1}}{\gamma_n} + \sum_{k=2}^n T_k \left[\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right] - \sum_{k=1}^n \delta_k \right). \quad (5.27)$$

Wir geben nun ein Lemma, was uns ermöglicht für die Konvergenz den rechten Summanden zu ignorieren und uns ausschließlich auf $\frac{\tilde{M}_{n+1}}{\sqrt{n}}$, zu konzentrieren.

Lemma 5.4.1.

$$\frac{1}{\sqrt{n}} \left(\frac{T_1}{\gamma_1} - \frac{T_{n+1}}{\gamma_n} + \sum_{k=2}^n T_k \left[\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right] - \sum_{k=1}^n \delta_k \right) \xrightarrow{d} 0$$

Beweis. Setze

$$A_n := \frac{T_1}{\sqrt{n}\gamma_1}, \quad A'_n := \frac{T_{n+1}}{\sqrt{n}\gamma_n}, \quad A''_n := \frac{1}{\sqrt{n}} \sum_{k=2}^n T_k \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right), \quad A'''_n := \frac{1}{\sqrt{n}} \sum_{k=1}^n \delta_k.$$

Wir werden die Konvergenz für jeden einzelnen Summanden zeigen. $A_n \xrightarrow{d} 0$ ist klar, da $\frac{T_1}{\gamma_1}$ nicht von n abhängt. Zur Erinnerung: Nach Satz 5.3.10 gilt $\mathbb{E} [\mathbb{1}_{\Omega_N} \|T_n\|^2] \leq C_N \frac{\log(n)}{n^\alpha}$. Für A'_n haben wir somit

$$\begin{aligned} \mathbb{E} [\mathbb{1}_{\Omega_N} \|A'_n\|^2] &= \frac{1}{n\gamma_n^2} \mathbb{E} [\mathbb{1}_{\Omega_N} \|T_{n+1}\|^2] \leq \frac{1}{n\gamma_n^2} \cdot C_N \frac{\log(n+1)}{(n+1)^\alpha} \\ &\leq \frac{C_N \log(n+1)}{\gamma_n^2 n^{1+\alpha}} = \frac{C_N \log(n+1)}{c_\gamma^2 n^{-2\alpha} n^{1+\alpha}} = \frac{C_N \log(n+1)}{c_\gamma^2 n^{1-\alpha}}. \end{aligned}$$

Da $\alpha < 1$, gilt $\frac{C_N \log(n+1)}{c_\gamma^2 n^{1-\alpha}} \xrightarrow{n \rightarrow \infty} 0$. Da für eine Nullmenge M_0 $\Omega = \bigcup_{N \in \mathbb{N}} \Omega_N \cup M_0$ gilt, gilt $\mathbb{E} [\|A'_n\|^2] = \sup_{N \in \mathbb{N}} \mathbb{E} [\mathbb{1}_{\Omega_N} \|A'_n\|^2]$. Insgesamt folgt also $A'_n \xrightarrow{\mathcal{L}_2} 0$. Aus Lemma 4.1.5 folgt $A'_n \xrightarrow{d} 0$. Wenden wir uns A''_n zu. Es gilt

$$\begin{aligned} \left| \frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right| &= \frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} = \frac{k^\alpha}{c_\gamma} - \frac{(k-1)^\alpha}{c_\gamma} = \frac{k^\alpha}{c_\gamma} \left(1 - \left(1 - \frac{1}{k} \right)^\alpha \right) \\ &\leq \frac{k^\alpha}{c_\gamma} \left(1 - \left(1 - \frac{\alpha}{k} \right) \right) = \alpha c_\gamma^{-1} k^{\alpha-1}. \end{aligned} \tag{5.28}$$

Dabei benutzen wir, dass wir die Bernoullische Ungleichung anwenden können, weil $\alpha \leq 1$ und $-\frac{1}{k} \geq -1$. Entsprechend folgt für $C'_N := \frac{\alpha}{c_\gamma} \sqrt{C_N}$

$$\begin{aligned} \mathbb{E} [\mathbb{1}_{\Omega_N} \|A''_n\|] &= \mathbb{E} \left[\left\| \frac{1}{\sqrt{n}} \sum_{k=2}^n \mathbb{1}_{\Omega_N} T_k \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right) \right\| \right] \\ &\leq \frac{1}{\sqrt{n}} \sum_{k=2}^n \mathbb{E} [\|\mathbb{1}_{\Omega_N} T_k\|] \left| \frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right| \\ &\leq \frac{\alpha}{c_\gamma \sqrt{n}} \sum_{k=2}^n \mathbb{E} [\|\mathbb{1}_{\Omega_N} T_k\|] k^{\alpha-1} \\ &\leq \frac{\alpha}{c_\gamma \sqrt{n}} \sum_{k=2}^n \sqrt{\mathbb{E} [\|\mathbb{1}_{\Omega_N} T_k\|^2]} k^{\alpha-1} \\ &\leq \frac{\alpha}{c_\gamma \sqrt{n}} \sum_{k=2}^n \sqrt{C_N \frac{\log(k)}{k^\alpha}} k^{\alpha-1} \\ &= \frac{C'_N}{\sqrt{n}} \sum_{k=2}^n \sqrt{\log(k)} k^{\frac{\alpha}{2}-1} \\ &\leq \frac{C'_N}{\sqrt{n}} \sum_{k=2}^n \sqrt{\log(n)} n^{\frac{\alpha}{2}-1} \\ &= C'_N \frac{n}{\sqrt{n}} \sqrt{\log(n)} n^{\frac{\alpha}{2}-1} \\ &= C'_N \sqrt{\log(n)} n^{\frac{1}{2}(\alpha-1)} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Dabei ist die erste Ungleichung die Dreiecksungleichung und die zweite folgt aus Gleichung 5.28. Die Dritte ist die Hölder-Ungleichung. Die Vierte nutzt wieder Satz 5.3.10.

Die letzte Zeile konvergiert gegen 0, weil $\alpha > \frac{1}{2}$. Wie oben folgt daraus $A_n'' \xrightarrow{\mathcal{L}_1} 0$ und $A_n'' \xrightarrow{d} 0$.

Wenden wir uns schließlich A_n''' zu. Wegen Lemma 5.1.10 gibt es eine Konstante $\tilde{C}_m$ mit $\|\delta_k\| \leq \tilde{C}_m \|Z_k - m\|^2 = \tilde{C}_m \|T_k\|^2$.

$$\begin{aligned} \mathbb{E} [\mathbb{1}_{\Omega_N} \|A_n'''\|] &= \frac{1}{\sqrt{n}} \mathbb{E} \left[\left\| \sum_{k=1}^n \mathbb{1}_{\Omega_N} \delta_k \right\| \right] \leq \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbb{E} [\|\mathbb{1}_{\Omega_N} \delta_k\|] \\ &\leq \frac{\tilde{C}_m}{\sqrt{n}} \sum_{k=1}^n \mathbb{E} [\|\mathbb{1}_{\Omega_N} T_k\|^2] \leq \frac{\tilde{C}_m C_N}{\sqrt{n}} \sum_{k=1}^n \frac{\log(k)}{k^\alpha} \end{aligned}$$

Da $\alpha > \frac{1}{2}$, gilt $\sum_{k=1}^{\infty} \log(k) k^{-\alpha-\frac{1}{2}} < \infty$. Nach dem Kroneckerschen Lemma folgt

$$\mathbb{E} [\mathbb{1}_{\Omega_N} \|A_n'''\|] \leq \frac{\tilde{C}_m C_N}{\sqrt{n}} \sum_{k=1}^n \frac{\log(k)}{k^\alpha} \xrightarrow{n \rightarrow \infty} 0.$$

Somit folgt nach derselben Logik wie oben, dass $A_n''' \xrightarrow{d} 0$ aus $A_n''' \xrightarrow{\mathcal{L}_1} 0$ und Lemma 4.1.5. $\square$

Zur Erinnerung: für $u, v, h \in \mathcal{H}$ gilt $u \otimes v(h) = \langle u, h \rangle v$. Definiere im Folgenden

$$\Sigma := \mathbb{E} \left[\frac{(X - m)}{\|X - m\|} \otimes \frac{(X - m)}{\|X - m\|} \right], \quad \Sigma_n := \mathbb{E} [\xi_{n+1} \otimes \xi_{n+1} | \mathcal{F}_n].$$

Lemma 5.4.2. *Es gilt $\lim_{n \rightarrow \infty} \Sigma_n = \Sigma$ $\mathbb{P}$ -f.s. in der von $\|\cdot\|_L$ induzierten Metrik, d.h. $\|\Sigma_n - \Sigma\|_L \rightarrow 0$ $\mathbb{P}$ -f.s..*

Beweis. Wir können Σ_n folgendermaßen zerlegen

$$\begin{aligned} \Sigma_n &= \mathbb{E} [\xi_{n+1} \otimes \xi_{n+1} | \mathcal{F}_n] = \mathbb{E} [(\Phi(Z_n) - U_{n+1}) \otimes (\Phi(Z_n) - U_{n+1}) | \mathcal{F}_n] \\ &= \mathbb{E} [\Phi(Z_n) \otimes \Phi(Z_n) | \mathcal{F}_n] + \mathbb{E} [U_{n+1} \otimes U_{n+1} | \mathcal{F}_n] \\ &\quad - \mathbb{E} [U_{n+1} \otimes \Phi(Z_n) | \mathcal{F}_n] - \mathbb{E} [\Phi(Z_n) \otimes U_{n+1} | \mathcal{F}_n] \\ &= \mathbb{E} [U_{n+1} \otimes U_{n+1} | \mathcal{F}_n] + \Phi(Z_n) \otimes \Phi(Z_n) \\ &\quad - \mathbb{E} [U_{n+1} | \mathcal{F}_n] \otimes \Phi(Z_n) - \Phi(Z_n) \otimes \mathbb{E} [U_{n+1} | \mathcal{F}_n] \\ &= \mathbb{E} [U_{n+1} \otimes U_{n+1} | \mathcal{F}_n] - \Phi(Z_n) \otimes \Phi(Z_n). \end{aligned}$$

Dabei benutzen wir bei der dritten Gleichheit die Distributivität von $\otimes$ (vgl. 4.2.20) und die Linearität des bedingten Erwartungswerts. Bei der vorletzten Gleichung benutzen wir die Eigenschaft des bedingten Erwartungswerts und des Tensorprodukts aus Lemma 4.2.21. Die letzte Gleichung benutzt, dass nach Lemma 5.1.4 $\mathbb{E} [U_{n+1} | \mathcal{F}_n] = \Phi(Z_n)$ gilt.

$$\begin{aligned} \|\Phi(Z_n)\| &= \|\delta_n + \Gamma_m(Z_n - m)\| \leq \|\delta_n\| + \|\Gamma_m(Z_n - m)\| \\ &\leq \tilde{C}_m \|Z_n - m\|^2 + \|\Gamma_m\|_L \|Z_n - m\|. \end{aligned}$$

In der zweiten Zeile benutzen wir Lemma 5.1.10. Nach Lemma 3.2.15 ist $\|\Gamma_m\|_L$ endlich. Da nach Satz 5.2.2 Z_n fast sicher gegen m konvergiert, konvergiert $\Phi(Z_n)$ fast sicher gegen 0. Mit Lemma 4.2.19 folgt, dass auch $\Phi(Z_n) \otimes \Phi(Z_n)$ in $\|\cdot\|_L$ fast sicher gegen 0 konvergiert. Wir können uns für die Aussage auf Folgendes beschränken

$$\begin{aligned}
 & \|\Sigma - \mathbb{E}[U_{n+1} \otimes U_{n+1} | \mathcal{F}_n]\|_L \\
 &= \left\| \mathbb{E} \left[\frac{(X-m)}{\|X-m\|} \otimes \frac{(X-m)}{\|X-m\|} \right] - \mathbb{E} \left[\frac{(X-Z_n)}{\|X-Z_n\|} \otimes \frac{(X-Z_n)}{\|X-Z_n\|} \middle| \mathcal{F}_n \right] \right\|_L \\
 &= \left\| \mathbb{E} \left[\frac{(X-m)}{\|X-m\|} \otimes \frac{(X-m)}{\|X-m\|} \middle| \mathcal{F}_n \right] - \mathbb{E} \left[\frac{(X-Z_n)}{\|X-Z_n\|} \otimes \frac{(X-Z_n)}{\|X-Z_n\|} \middle| \mathcal{F}_n \right] \right\|_L \\
 &= \left\| \mathbb{E} \left[\frac{(X-m)}{\|X-m\|} \otimes \frac{(X-m)}{\|X-m\|} - \frac{(X-Z_n)}{\|X-Z_n\|} \otimes \frac{(X-Z_n)}{\|X-Z_n\|} \middle| \mathcal{F}_n \right] \right\|_L.
 \end{aligned}$$

Dabei benutzen wir in der zweiten Zeile, dass X unabhängig von $\mathcal{F}_n$ ist. Setze im folgenden $u := \frac{(X-m)}{\|X-m\|}$ und $v := \frac{(X-Z_n)}{\|X-Z_n\|}$. Bemerke, dass dann $\|u\| = \|v\| = 1$ gilt.

$$\begin{aligned}
 \|\mathbb{E}[u \otimes u - v \otimes v | \mathcal{F}_n]\|_L &\leq \mathbb{E}[\|u \otimes u - v \otimes v\|_L | \mathcal{F}_n] \\
 &= \mathbb{E}[\|u \otimes u - v \otimes u + v \otimes u - v \otimes v\|_L | \mathcal{F}_n] \\
 &= \mathbb{E}[\|(u-v) \otimes u + v \otimes (u-v)\|_L | \mathcal{F}_n] \\
 &\leq \mathbb{E}[\|(u-v) \otimes u\|_L + \|v \otimes (u-v)\|_L | \mathcal{F}_n] \\
 &\leq \mathbb{E}[\|(u-v)\| \|u\| + \|v\| \|(u-v)\| | \mathcal{F}_n] \\
 &= 2\mathbb{E}[\|v-u\| | \mathcal{F}_n].
 \end{aligned}$$

Dabei sind die ersten beiden Ungleichungen die Dreiecksungleichung. Die dritte Zeile ist das Distributivgesetz aus Lemma 4.2.20. Die vorletzte Zeile ist Lemma 4.2.19.

$$\begin{aligned}
 \|\mathbb{E}[u \otimes u - v \otimes v | \mathcal{F}_n]\|_L &\leq 2\mathbb{E}[\|v-u\| | \mathcal{F}_n] \\
 &= 2\mathbb{E} \left[\left\| \frac{(X-Z_n)}{\|X-Z_n\|} - \frac{(X-m)}{\|X-m\|} \right\| \middle| \mathcal{F}_n \right] \\
 &= 2\mathbb{E} \left[\frac{1}{\|X-m\|} \left\| \frac{\|X-m\|}{\|X-Z_n\|} (X-Z_n) - (X-m) \right\| \middle| \mathcal{F}_n \right] \\
 &= 2\mathbb{E} \left[\frac{1}{\|X-m\|} \| (X-Z_n) \right. \\
 &\quad \left. + \frac{\|X-m\| - \|X-Z_n\|}{\|X-Z_n\|} (X-Z_n) - (X-m) \right\| \middle| \mathcal{F}_n \right] \\
 &= 2\mathbb{E} \left[\frac{1}{\|X-m\|} \| (m-Z_n) \right. \\
 &\quad \left. + \frac{\|X-m\| - \|X-Z_n\|}{\|X-Z_n\|} (X-Z_n) \right\| \middle| \mathcal{F}_n \right] \\
 &\leq 2\mathbb{E} \left[\frac{1}{\|X-m\|} (\|m-Z_n\| \right. \\
 &\quad \left. + \left\| \frac{\|X-m\| - \|X-Z_n\|}{\|X-Z_n\|} (X-Z_n) \right\|) \middle| \mathcal{F}_n \right]
 \end{aligned}$$

Dabei benutzen wir bei der ersten Ungleichungen die gewöhnliche Dreiecksungleichung.

$$\begin{aligned}
 \|\mathbb{E}[u \otimes u - v \otimes v | \mathcal{F}_n]\|_L &\leq 2\mathbb{E} \left[\frac{1}{\|X - m\|} (\|m - Z_n\| \right. \\
 &\quad \left. + \left\| \frac{\|X - m\| - \|X - Z_n\|}{\|X - Z_n\|} (X - Z_n) \right\| \right) \Big| \mathcal{F}_n \right] \\
 &= 2\mathbb{E} \left[\frac{1}{\|X - m\|} (\|m - Z_n\| \right. \\
 &\quad \left. + |\|X - m\| - \|X - Z_n\||) \Big| \mathcal{F}_n \right] \\
 &\leq 2\mathbb{E} \left[\frac{1}{\|X - m\|} (\|m - Z_n\| + \|X - m - X - Z_n\|) \Big| \mathcal{F}_n \right] \\
 &= 4\mathbb{E} \left[\frac{\|Z_n - m\|}{\|X - m\|} \Big| \mathcal{F}_n \right] \\
 &= 4\mathbb{E} \left[\frac{1}{\|X - m\|} \right] \|Z_n - m\|.
 \end{aligned}$$

Die letzte ist die umgedrehte Dreiecksungleichung. In der letzten Zeile benutzen wir, dass Z_n messbar bezüglich $\mathcal{F}_n$ ist und wir deshalb Lemma 4.2.16 anwenden können.

Insgesamt folgt $\|\Sigma - \mathbb{E}[U_{n+1} \otimes U_{n+1} | \mathcal{F}_n]\|_L \leq 4\mathbb{E} \left[\frac{1}{\|X - m\|} \right] \|Z_n - m\|$. Nach Bemerkung 3.1.1 ist $\mathbb{E} \left[\frac{1}{\|X - m\|} \right]$ endlich, da wir in Annahme 2 $m \notin D$ voraussetzen. Wie auch oben schon wissen wir nach Satz 5.2.2, dass Z_n fast sicher gegen m konvergiert und deshalb konvergiert $\|\Sigma - \mathbb{E}[U_{n+1} \otimes U_{n+1} | \mathcal{F}_n]\|_L$ fast sicher gegen 0. Insgesamt folgt also

$$\|\Sigma - \Sigma_n\|_L \leq \|\Sigma - \mathbb{E}[U_{n+1} \otimes U_{n+1} | \mathcal{F}_n]\|_L + \|\Phi(Z_n) \otimes \Phi(Z_n)\|_L \xrightarrow{\mathbb{P}\text{-f.s.}} 0 + 0.$$

□

Lemma 5.4.3. Für $i, j \in \mathbb{N}$ gilt

$$\frac{1}{n} \sum_{k=1}^n \langle \xi_{k+1}, e_i \rangle \langle \xi_{k+1}, e_j \rangle \xrightarrow{\mathbb{P}\text{-f.s.}} \langle \Sigma e_i, e_j \rangle.$$

Beweis. Fixiere $i, j \in \mathbb{N}$. Es gilt

$$\begin{aligned}
 \langle \Sigma_n e_i, e_j \rangle &= \langle \mathbb{E}[\xi_{n+1} \otimes \xi_{n+1} | \mathcal{F}_n] e_i, e_j \rangle \\
 &= \mathbb{E}[\langle (\xi_{n+1} \otimes \xi_{n+1}) e_i, e_j \rangle | \mathcal{F}_n] = \mathbb{E}[\langle \xi_{n+1}, e_i \rangle \langle \xi_{n+1}, e_j \rangle | \mathcal{F}_n].
 \end{aligned} \tag{5.29}$$

Setze $\psi_k := \langle \xi_k, e_i \rangle \langle \xi_k, e_j \rangle$. Es gilt

$$\begin{aligned}
 &\frac{1}{n} \sum_{k=1}^n \langle \xi_{k+1}, e_i \rangle \langle \xi_{k+1}, e_j \rangle - \langle \Sigma e_i, e_j \rangle \\
 &= \frac{1}{n} \sum_{k=1}^n \psi_{k+1} - \langle \Sigma e_i, e_j \rangle \\
 &= \frac{1}{n} \sum_{k=1}^n \psi_{k+1} - \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\psi_{k+1} | \mathcal{F}_k] + \frac{1}{n} \sum_{k=1}^n \langle \Sigma_n e_i, e_j \rangle - \langle \Sigma e_i, e_j \rangle \\
 &= \frac{1}{n} \sum_{k=1}^n (\psi_{k+1} - \mathbb{E}[\psi_{k+1} | \mathcal{F}_k]) + \frac{1}{n} \sum_{k=1}^n \langle (\Sigma_n - \Sigma) e_i, e_j \rangle.
 \end{aligned}$$

Dabei ist die dritte Gleichheit Gleichung 5.29. Wir möchten nun für beide Summen zeigen, dass sie fast sicher gegen 0 konvergieren. Bei der linken Summe wollen wir Satz 4.4.4 (Chows Lemma) anwenden. Setze $\eta_{k+1} := \psi_{k+1} - \mathbb{E}[\psi_{k+1}|\mathcal{F}_k]$. Wir prüfen die Bedingungen:

1. Wegen Linearität ist klar, dass für alle $k \in \mathbb{N}$ $\mathbb{E}[\eta_{k+1}|\mathcal{F}_k] = \mathbb{E}[\psi_{k+1} - \mathbb{E}[\psi_{k+1}|\mathcal{F}_k]|\mathcal{F}_k] = 0$ gilt.
2. Zu zeigen: $\sup_{n \in \mathbb{N}} \mathbb{E}[\|\eta_{n+1}\|_L^a|\mathcal{F}_n] < \infty$ für ein $1 < a \leq 2$.

Wegen Lemma 5.1.7 und Cauchy-Schwarz gilt $\|\psi_n\| = \|\langle \xi_k, e_i \rangle \langle \xi_k, e_j \rangle\| \leq \|\xi_k\|^2 \leq 4$. Entsprechend folgt

$$\begin{aligned} \mathbb{E}[\|\eta_{n+1}\|^2|\mathcal{F}_n] &= \mathbb{E}[\|\psi_{n+1} - \mathbb{E}[\psi_{n+1}|\mathcal{F}_n]\|^2|\mathcal{F}_n] \\ &\leq \mathbb{E}[(\|\psi_{n+1}\| + \mathbb{E}[\|\psi_{n+1}\||\mathcal{F}_n])^2|\mathcal{F}_n] \\ &\leq \mathbb{E}[(4 + 4)^2|\mathcal{F}_n] \leq 64 < \infty. \end{aligned}$$

Somit ist gezeigt, dass die linke Summe gegen 0 konvergiert. Bei der rechten Summe stellen wir fest, dass wir bereits in Lemma 5.4.2 $\|\Sigma_n - \Sigma\|_L \rightarrow 0$ $\mathbb{P}$ -f.s. gezeigt haben. Deshalb gibt es eine Nullmenge M_0 , sodass $\Sigma_n \mathbb{1}_{M_0^C} \rightarrow \Sigma \mathbb{1}_{M_0^C}$ überall punktweise gilt. Aus Lemma 6.10 auf Seite 208 in [Pet95] folgt hieraus, dass $\frac{1}{n} \sum_{k=1}^n \Sigma_k \mathbb{1}_{M_0} \xrightarrow{n \rightarrow \infty} \Sigma \mathbb{1}_{M_0}$. Daraus folgt wiederum, dass $\frac{1}{n} \sum_{k=1}^n \Sigma_k \xrightarrow{n \rightarrow \infty} \Sigma$ $\mathbb{P}$ -f.s.. $\square$

Satz 5.4.4. *Es gilt*

$$\frac{\tilde{M}_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

Beweis. Wir möchten Satz 4.4.3 verwenden. Setze $X_{n,k} := \frac{\xi_{k+1}}{\sqrt{n}}$ und $\mathcal{F}_{n,k} := \mathcal{F}_{k+1}$. Nach Bemerkung 5.1.5 wissen wir, dass $\mathbb{E}[\xi_{k+1}|\mathcal{F}_k] = 0$. Deshalb ist $\{X_{n,k}, \mathcal{F}_{n,k}\}_{n,k \in \mathbb{N}}$ ein Martingaldifferenzarray (vgl. 4.3.6). Es genügt nun, die drei Bedingungen aus Satz 4.4.3 zu zeigen.

1.

$$\begin{aligned} \mathbb{E} \left[\sup_{1 \leq k \leq n} \|X_{n,k}\| \right] &= \frac{1}{\sqrt{n}} \mathbb{E} \left[\sup_{1 \leq k \leq n} \|\xi_{k+1}\| \right] \\ &\leq \frac{1}{\sqrt{n}} \mathbb{E}[2] \leq \frac{2}{\sqrt{n}} \rightarrow 0. \end{aligned}$$

Dabei benutzen wir, dass nach Lemma 5.1.7 $\|\xi_{n+1}\| \leq 2$ $\mathbb{P}$ -f.s. gilt.

2. Das ist genau die Aussage von Lemma 5.4.3.

3. Für $\varepsilon > 0$ gilt

$$\begin{aligned} \lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{k=1}^n r_N^2(X_{n,k}) > \varepsilon \right) &= \lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{k=1}^n \sum_{i=N}^{\infty} \langle X_{n,k}, e_i \rangle^2 > \varepsilon \right) \\ &= \lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{k=1}^n \frac{1}{n} \sum_{i=N}^{\infty} \langle \xi_{k+1}, e_i \rangle^2 > \varepsilon \right) \\ &\leq \lim_{N \rightarrow \infty} \mathbb{P} \left(\sup_{k \in \mathbb{N}} \sum_{i=N}^{\infty} \langle \xi_{k+1}, e_i \rangle^2 > \varepsilon \right). \end{aligned}$$

Wir wissen nach Korollar 2.1.15 und Lemma 5.1.7, dass für alle $k \in \mathbb{N}$:

$$\sum_{i=1}^{\infty} \langle \xi_{k+1}, e_i \rangle^2 = \|\xi_{k+1}\|^2 \leq 4 \text{ } \mathbb{P}\text{-f.s.}. \text{ Deshalb gilt}$$

$$\begin{aligned} \forall k \in \mathbb{N} : \lim_{N \rightarrow \infty} \sum_{i=N}^{\infty} \langle \xi_{k+1}, e_i \rangle^2 &= 0 \text{ } \mathbb{P}\text{-f.s.} \quad \Rightarrow \lim_{N \rightarrow \infty} \sup_{k \in \mathbb{N}} \sum_{i=N}^{\infty} \langle \xi_{k+1}, e_i \rangle^2 = 0 \text{ } \mathbb{P}\text{-f.s.} \\ \Rightarrow \lim_{N \rightarrow \infty} \mathbb{P} \left(\sup_{k \in \mathbb{N}} \sum_{i=N}^{\infty} \langle \xi_{k+1}, e_i \rangle^2 > \varepsilon \right) &= 0. \end{aligned}$$

Nach Satz 4.4.3 folgt, dass $\frac{\tilde{M}_n}{\sqrt{n}} = \sum_{k=1}^n \frac{\xi_{k+1}}{\sqrt{n}} = \sum_{k=1}^n X_{n,k} \xrightarrow{d} \mathcal{N}(0, \Sigma)$. $\square$

Das wichtigste Ergebnis dieser Arbeit, auf das alles bisher abgezielt hat:

Korollar 5.4.5.

$$\sqrt{n} (\bar{Z}_n - m) \xrightarrow{d} \mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1})$$

mit Σ wie in 5.4.

Beweis. Rufen wir die Gleichung 5.27 in Erinnerung

$$\sqrt{n} \Gamma_m \bar{T}_n = \frac{1}{\sqrt{n}} \tilde{M}_{n+1} + \frac{1}{\sqrt{n}} \left(\frac{T_1}{\gamma_1} - \frac{T_{n+1}}{\gamma_n} + \sum_{k=2}^n T_k \left[\frac{1}{\gamma_k} - \frac{1}{\gamma_{k-1}} \right] - \sum_{k=1}^n \delta_k \right).$$

Nach Lemma 5.4.1 konvergiert der rechte Summand gegen 0 und nach Satz 5.4 der linke Summand gegen $\mathcal{N}(0, \Sigma)$ in Verteilung. Somit haben wir $\sqrt{n} \Gamma_m \bar{T}_n \xrightarrow{d} \mathcal{N}(0, \Sigma)$. Γ_m ist nach Bemerkung 3.3.7 invertierbar. Dann folgt die Aussage direkt aus den Rechenregeln der multivariaten Normalverteilung, da Γ_m symmetrisch ist. $\square$

5.5 Nicht-asymptotische Konfidenzbereiche

Wir geben Konfidenzbälle von den beiden Algorithmen für endliches n . Die Resultate stammen aus Abschnitt 4 in [CCG17].

In der Praxis ist es relevanter, für ein endliches n eine Aussage über die Verteilung des Schätzers zu geben. Ein wesentliches Güte Merkmal eines Schätzers ist der Konfidenzball.

Satz 5.5.1. *Es gibt eine positive Konstante C , sodass es für alle $\delta \in (0, 1)$ einen Rang n_δ gibt, sodass für alle $n \geq n_\delta$*

$$\mathbb{P} \left[\|Z_n - m\| \leq \frac{C}{n^{\alpha/2}} \log \left(\frac{4}{\delta} \right) \right] \geq 1 - \delta \text{ und}$$

$$\mathbb{P} \left[\|\bar{Z}_n - m\| \leq \frac{4}{\lambda_{\min}} \left(\frac{2}{3n} + \frac{1}{\sqrt{n}} \right) \log \left(\frac{4}{\delta} \right) \right] \geq 1 - \delta$$

gilt, wobei $\lambda_{\min}$ der $\liminf$ der Eigenwerte der Hesse-Matrix Γ_m ist.

Der Beweis findet sich in Abschnitt 4 in [CCG17]. Er wird hier nicht ausgeführt, da er den Rahmen dieser Arbeit überschreiten würde.

Kapitel 6

Simulation

In diesem Kapitel möchten wir ein Resultat aus dem letzten Kapitel mit einer Simulation in der Programmiersprache R illustrieren.

Wir wählen im Folgenden $\alpha = \frac{3}{4}$ und $c_\gamma = 2,7$ (vgl. Unterabschnitt 5.1.1). Sei $X \sim U_{(-2,2)}^{\otimes 3}$, wobei $U_{(-2,2)}$ für die stetige Gleichverteilung auf dem Intervall $[-2, 2]$ steht. Sei $n \in \mathbb{N}$ eine natürliche Zahl und seien $X_0, X_1, \dots, X_n$ unabhängige identische Kopien von X . Wir können mit R leicht Werte für $X_0, X_1, \dots, X_n$ simulieren. Nach Beispiel 3.1.3 erfüllt X dann die Annahmen 2 und 3. Es ist auch leicht zu zeigen, dass Annahme 1 erfüllt ist, aber wir verzichten an dieser Stelle darauf. Dann gilt nach Korollar 5.4.5

$$W := \sqrt{n} (\bar{Z}_n - m) \xrightarrow{d} \mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1}) \quad (6.1)$$

mit

$$\Gamma_m = \mathbb{E} \left[\frac{1}{\|X - m\|} \left(\mathbf{I}_{\mathcal{H}} - \frac{(X - m) \otimes (X - m)}{\|X - m\|^2} \right) \right], \quad \Sigma = \mathbb{E} \left[\frac{(X - m)}{\|X - m\|} \otimes \frac{(X - m)}{\|X - m\|} \right].$$

Wir möchten diesen Effekt nun mit R visualisieren. Wir berechnen für eine bestimmte Stichprobengröße n in 1000 Durchläufen mit jeweils neu generierten $X_0, X_1, \dots, X_n$ den Wert für W in Gleichung 6.1. Wir projizieren alle diese Werte auf ihre ersten beiden Koordinaten $\pi_1(W)$ und $\pi_2(W)$. Die Verteilung von diesen 1000 Werten von W in den ersten beiden Dimensionen können wir mit dem Kerndichteschätzer aus dem Paket „stats“ in R visualisieren. Die Graphen finden sich unten in der Tabelle 6.1.

Zum Abgleich möchten wir $\mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1})$ bestimmen und daneben zeichnen. Dazu müssen wir zunächst für $X \sim U_{[-2,2]}^{\otimes 3}$ Γ_m und Σ bestimmen. Eine ähnliche Rechnung wie in Lemma 3.3.4 gibt uns,

$$\begin{aligned} \langle e_i, \Gamma_m e_j \rangle &= \mathbb{E} \left[\frac{1}{\|X - m\|} \left(\langle e_i, e_j \rangle - \frac{\langle X - m, e_j \rangle \langle e_i, X - m \rangle}{\|X - m\|^2} \right) \right] \\ &= \delta_{i,j} \mathbb{E} \left[\frac{1}{\|X - m\|} \right] - \mathbb{E} \left[\frac{\langle X - m, e_j \rangle \langle e_i, X - m \rangle}{\|X - m\|^3} \right]. \end{aligned} \quad (6.2)$$

Für $X \sim U_{[-2,2]}^{\otimes 3}$ ist $m = 0$. Wir führen eine Fallunterscheidung durch.

„ $i = j$ “ :

$$\begin{aligned}\langle e_i, \Gamma_m e_i \rangle &= \mathbb{E} \left[\frac{1}{\|X\|} \right] - \mathbb{E} \left[\frac{\langle e_i, X \rangle^2}{\|X\|^3} \right] \\ &= \frac{1}{64} \int_{-2}^2 \int_{-2}^2 \int_{-2}^2 \frac{1}{\sqrt{x^2 + y^2 + z^2}} - \frac{x^2}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} dx dy dz \\ &\approx \frac{1}{64} (38,0812 - 12,6937) = 0,3966.\end{aligned}$$

Dabei wurden die Integrale numerisch mit Hilfe von „*wolframalpha.com*“ bestimmt.

„ $i \neq j$ “ :

$$\begin{aligned}\langle e_i, \Gamma_m e_j \rangle &= -\mathbb{E} \left[\frac{\langle e_i, X \rangle \langle e_j, X \rangle}{\|X\|^3} \right] \\ &= -\frac{1}{64} \int_{-2}^2 \int_{-2}^2 \int_{-2}^2 \frac{xy}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} dx dy dz = 0.\end{aligned}$$

Wenden wir uns Σ zu.

$$\begin{aligned}\langle e_i, \Sigma e_j \rangle &= \left\langle e_i, \mathbb{E} \left[\frac{(X - m)}{\|X - m\|} \otimes \frac{(X - m)}{\|X - m\|} \right] e_j \right\rangle \\ &= \mathbb{E} \left[\frac{\langle X - m, e_j \rangle \langle e_i, X - m \rangle}{\|X - m\|^2} \right].\end{aligned}\tag{6.3}$$

Der wahre Median m ist 0. Auch hier führen wir eine Fallunterscheidung durch.

„ $i = j$ “ :

$$\begin{aligned}\langle e_i, \Sigma e_i \rangle &= \mathbb{E} \left[\frac{\langle e_i, X \rangle^2}{\|X\|^2} \right] \\ &= \frac{1}{64} \int_{-2}^2 \int_{-2}^2 \int_{-2}^2 \frac{x^2}{x^2 + y^2 + z^2} dx dy dz = \frac{1}{3}.\end{aligned}$$

„ $i \neq j$ “ :

$$\begin{aligned}\langle e_i, \Sigma e_j \rangle &= \mathbb{E} \left[\frac{\langle e_j, X \rangle \langle e_i, X \rangle}{\|X\|^2} \right] \\ &= \frac{1}{64} \int_{-2}^2 \int_{-2}^2 \int_{-2}^2 \frac{xy}{x^2 + y^2 + z^2} dx dy dz = 0.\end{aligned}$$

Somit gilt für $X \sim U_{[-2,2]}^{\otimes 3}$

$$\Gamma_m \approx \begin{pmatrix} 0,3966 & 0 & 0 \\ 0 & 0,3966 & 0 \\ 0 & 0 & 0,3966 \end{pmatrix}, \quad \Sigma \approx \begin{pmatrix} 0,3333 & 0 & 0 \\ 0 & 0,3333 & 0 \\ 0 & 0 & 0,3333 \end{pmatrix},$$

und entsprechend

$$\Gamma_m^{-1} \Sigma \Gamma_m^{-1} \approx \begin{pmatrix} \frac{0,3333}{0,3966^2} & 0 & 0 \\ 0 & \frac{0,3333}{0,3966^2} & 0 \\ 0 & 0 & \frac{0,3333}{0,3966^2} \end{pmatrix} \approx \begin{pmatrix} 2,119 & 0 & 0 \\ 0 & 2,119 & 0 \\ 0 & 0 & 2,119 \end{pmatrix}.$$

Die Dichte von $\mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1})$ findet sich in der letzten Zeile von Tabelle 6.1. Wir bemerken, dass sich die Verteilung von W sich in der Tat immer näher der von $\mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1})$ annähert.

Tabelle 6.1: Verteilung des Schätzers für $X \sim U_{[-2,2]}^{\otimes 3}$ bei 1000 Durchläufen für verschiedene Stichprobengrößen n

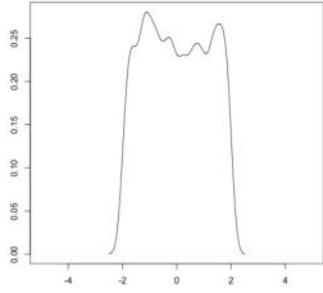
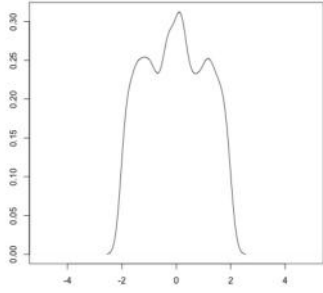
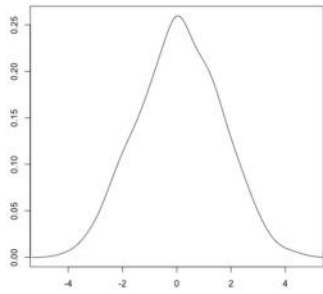
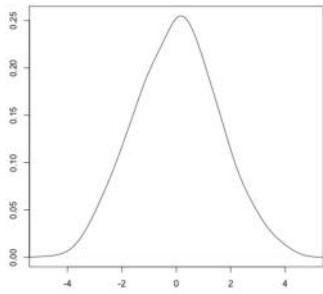
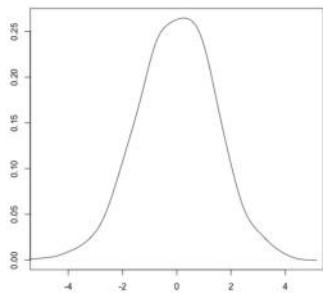
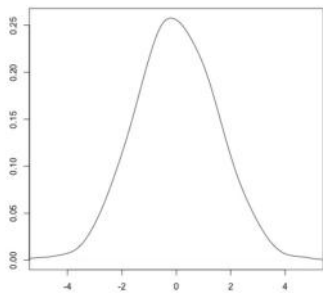
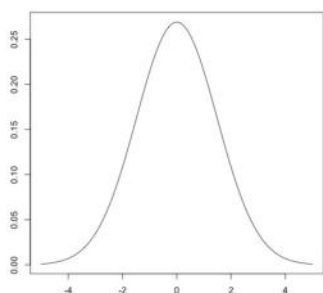
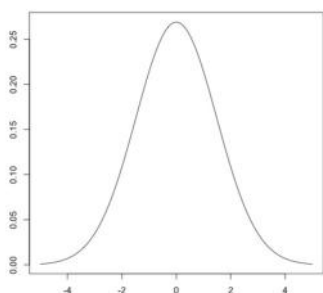
Stichprobengröße	Koordinate 1, $\pi_1(W)$	Koordinate 2, $\pi_2(W)$
$n = 1$		
$n = 10$		
$n = 10^5$		
$\mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1})$		

Tabelle 6.2: Verteilung des Schätzers für $X \sim \text{Exp}_1^{\otimes 3}$ bei 1000 Durchläufen für verschiedene Stichprobengrößen n

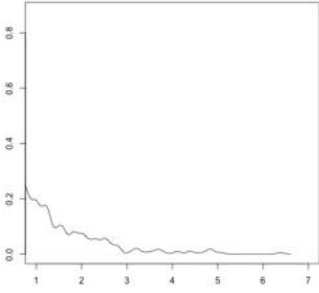
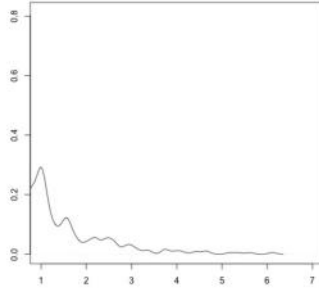
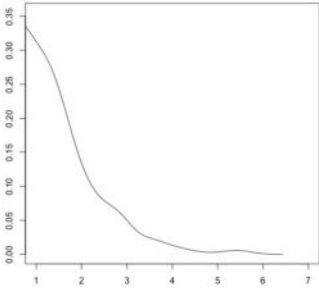
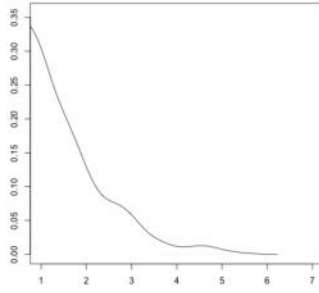
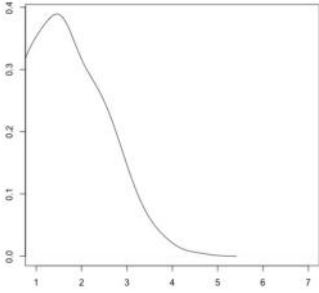
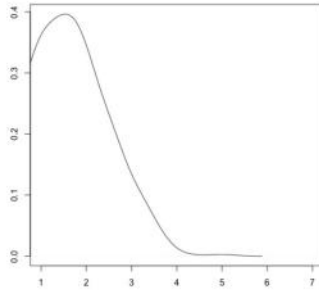
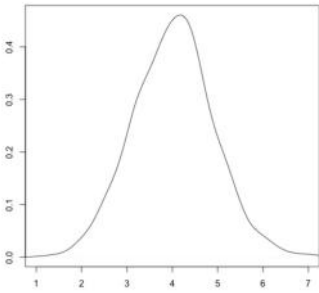
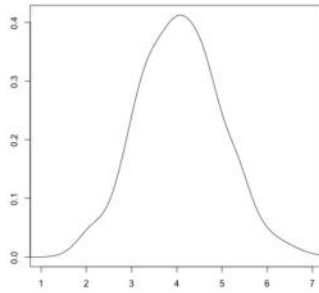
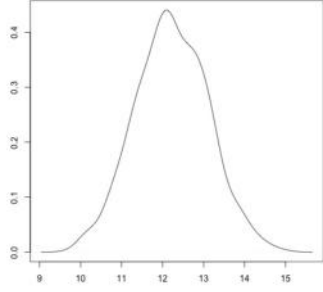
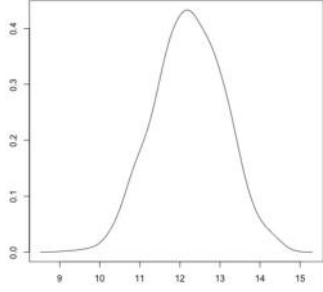
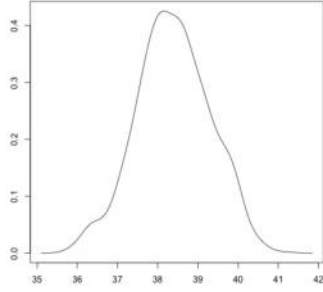
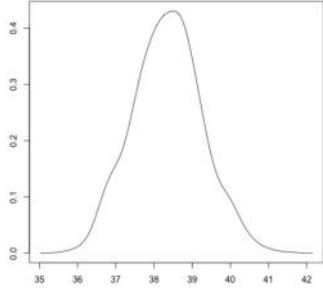
Stichprobengröße	Koordinate 1, $\pi_1(W)$	Koordinate 2, $\pi_2(W)$
$n = 1$		
$n = 10$		
$n = 100$		
$n = 1000$		

Tabelle 6.3: Verteilung des Schätzers für $X \sim \text{Exp}_1^{\otimes 3}$ bei 1000 Durchläufen für verschiedene Stichprobengrößen n

Stichprobengröße	Koordinate 1, $\pi_1(W)$	Koordinate 2, $\pi_2(W)$
$n = 10000$		
$n = 10^5$		

Wir bemerken, dass sich im Falle der Exponentialverteilung $X \sim \text{Exp}_1^{\otimes 3}$ die Verteilung von W sich durchaus der Glockenform einer Gaußkurve nähert. Jedoch wandert diese immer weiter nach rechts. Das bedeutet, dass der geschätzte Wert für den Median mit steigender Stichprobengröße sich immer weiter vom wahren Wert entfernt. Die Annäherung an eine Normalverteilung um den Nullpunkt ist also bei dieser Stichprobengröße nicht erkenntlich.

Wir können das gleiche für $X \sim \mathcal{N}(0, A)$ mit

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & -0.5 \\ 1 & -0.5 & 2 \end{pmatrix}$$

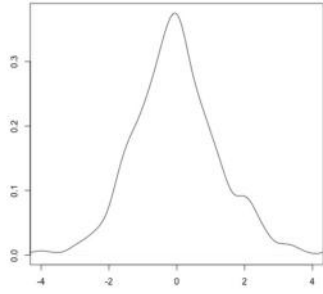
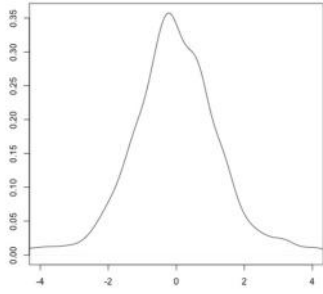
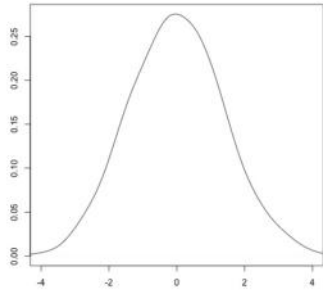
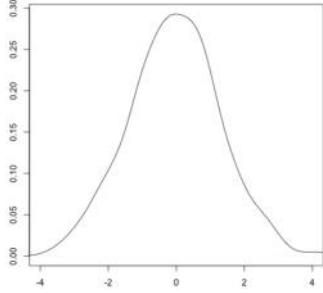
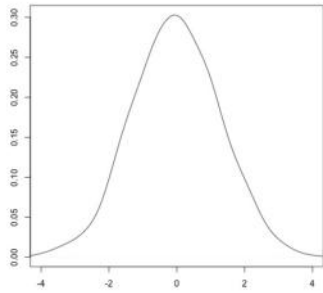
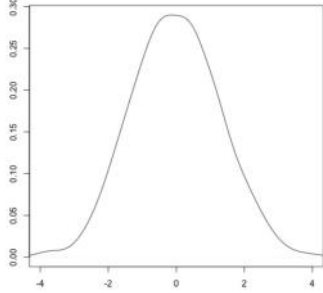
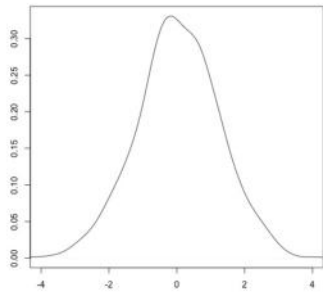
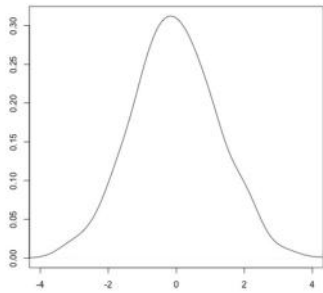
tun und $X \sim t_5^{\otimes 3}$. Die entsprechenden Graphen finden sich in Tabelle 6.4 und Tabelle 6.5.

Tabelle 6.4: Verteilung des Schätzers für $X \sim \mathcal{N}(0, A)$ bei 1000 Durchläufen für verschiedene Stichprobengrößen n .

Stichprobengröße	Koordinate 1, $\pi_1(W)$	Koordinate 3, $\pi_3(W)$
$n = 1$		
$n = 100$		
$n = 10000$		
$n = 10^5$		

Bei der Normalverteilung ist die Gauß-Kurve schon von Anfang an erkenntlich.

Tabelle 6.5: Verteilung des Schätzers für $X \sim t_5^{\otimes 3}$ (der studentschen t-Verteilung mit 5 Freiheitsgraden) bei 1000 Durchläufen für verschiedene Stichprobengrößen n .

Stichprobengröße	Koordinate 1, $\pi_1(W)$	Koordinate 2, $\pi_2(W)$
$n = 1$		
$n = 100$		
$n = 1000$		
$n = 10^5$		

Bei der studentschen t-Verteilung ist es mit steigender Stichprobengröße deutlich zu erkennen, dass sich der Graph immer mehr einer Gauß-Kurve annähert.

Literatur

- [Bal+11] Juan Lucas Bali u. a. „Robust functional principal components: A projection-pursuit approach“. In: *The Annals of Statistics* 39.6 (2011), S. 2852–2882.
- [Bre10] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.
- [CC+14] Anirvan Chakraborty, Probal Chaudhuri u. a. „The spatial distribution in infinite dimensional spaces and related quantiles and depths“. In: *The Annals of Statistics* 42.3 (2014), S. 1203–1231.
- [CCG17] Hervé Cardot, Peggy Cénac und Antoine Godichon-Baggioni. „Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls“. In: *The Annals of Statistics* 45.2 (2017), S. 591–614.
- [CCZ13] Hervé Cardot, Peggy Cénac und Pierre-André Zitt. „Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm“. In: *Bernoulli* 19.1 (2013), S. 18–43.
- [CG12] Mohamed Chaouch und Camelia Goga. „Using Complex Surveys to Estimate the L1-Median of a Functional Variable: Application to Electricity Load Curves“. In: *International Statistical Review* 80.1 (2012), S. 40–59.
- [Duf13] Marie Duffo. *Random iterative models*. Bd. 34. Springer Science & Business Media, 2013.
- [FVJ09] P Thomas Fletcher, Suresh Venkatasubramanian und Sarang Joshi. „The geometric median on Riemannian manifolds with application to robust atlas estimation“. In: *NeuroImage* 45.1 (2009), S143–S152.
- [Ger08] Daniel Gervini. „Robust functional estimation using the median and spherical principal components“. In: *Biometrika* 95.3 (2008), S. 587–600.
- [Gos11] Maurice A de Gosson. „Hilbert–Schmidt and Trace Class Operators“. In: *Symplectic Methods in Harmonic Analysis and in Mathematical Physics*. Springer, 2011, S. 185–203.
- [Jak88] Adam Jakubowski. „TIGHTNESS CRITERIA FOR RANDOM MEASURES WITH APPLICATION TO THE PRINCIPLE OF CONDITIONING IN I-ULBEWT SPACES“. In: (1988).

- [Kem87] JHB Kemperman. „The median of a finite measure on a Banach space“. In: *Statistical data analysis based on the L1-norm and related methods (Neuchâtel, 1987)* (1987), S. 217–230.
- [Kle06] Achim Klenke. *Wahrscheinlichkeitstheorie*. Bd. 1. Springer, 2006.
- [KP12] David Kraus und Victor M Panaretos. „Dispersion operators and resistant second-order functional data analysis“. In: *Biometrika* 99.4 (2012), S. 813–832.
- [Liu90] Regina Y Liu. „On a notion of data depth based on random simplices“. In: *The Annals of Statistics* (1990), S. 405–414.
- [LM92] Carlos Antonio León und Jean-Claude Massé. „A counterexample on the existence of the L1-median“. In: *Statistics & probability letters* 13.2 (1992), S. 117–120.
- [Loc+99] N Locantore u. a. „Robust principal component analysis for functional data“. In: *Test* 8.1 (1999), S. 1–73.
- [Min+14] Stanislav Minsker u. a. „Scalable and robust Bayesian inference via the median posterior“. In: *International conference on machine learning*. 2014, S. 1656–1664.
- [Min+15] Stanislav Minsker u. a. „Geometric median and robust estimation in Banach spaces“. In: *Bernoulli* 21.4 (2015), S. 2308–2335.
- [Naz07] Alexander I Nazarov. „Exact L₂-small ball asymptotics of Gaussian processes and the spectrum of boundary value problems with non-separated boundary conditions“. In: *arXiv preprint arXiv:0710.1408* (2007).
- [Oja83] Hannu Oja. „Descriptive statistics for multivariate distributions“. In: *Statistics & Probability Letters* 1.6 (1983), S. 327–332.
- [Pet95] Valentin V Petrov. „Limit theorems of probability theory: sequences of independent random variables“. In: (1995).
- [RH13] Peter Rousseeuw und Mia Hubert. „High-breakdown estimators of multivariate location and scatter“. In: *Robustness and complex data structures*. Springer, 2013, S. 49–66.
- [Sim15] Barry Simon. *A comprehensive course in analysis*. American Mathematical Society Providence, Rhode Island, 2015.
- [Sma90] Christopher G Small. „A survey of multidimensional medians“. In: *International Statistical Review/Revue Internationale de Statistique* (1990), S. 263–277.
- [Web09] Alfred Weber. *Ueber den Standort der Industrien*. Bd. 1. Ripol Klassik, 1909.