JOHANNES KEPLER
UNIVERSITÄT LINZ | JKU

TNF
Technisch-Naturwissenschaftliche
Fakultät

Stochastic Processes and Applications in the Geometry of Banach Spaces

MASTERARBEIT

zur Erlangung des akademischen Grades

Diplomingenieur

im Masterstudium

Mathematik in den Naturwissenschaften

Eingereicht von:
Daniel Temesvari

Angefertigt am:
Institut für Analysis

Beurteilung:
Asst.-Prof. Dr. Joscha Prochno
Dipl.-Ing.Dr. Markus Passenbrunner

Linz, March 2016

Eidesstattliche Erklärung

Ich, Daniel Temesvari, erkläre an Eides statt, dass ich die vorliegende Masterarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt beziehungsweise die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Die vorliegenden Masterarbeit ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, March 2016

Daniel Temesvari

Acknowledgements

I want to thank my supervisors Dr. Joscha Prochno and Dr. Markus Passenbrunner for giving me the opportunity to write this master thesis and for their guidance throughout this time. Numerous discussions and pointers in the right direction were very valuable in understanding the involved topics.

Finally, I would like to thank my family for supporting me throughout my studies.

Contents

Chapter 1

Introduction

In this thesis, we will take a look at two famous and deep theorems about stochastic processes, namely the Kolmogorov existence theorem and Talagrand's majorizing measure theorem, and applications of these in different fields of mathematics, i.e., stochastic processes, Banach space theory and asymptotic convex geometry.

We start our investigations in Chapter 2 with a rather general version of Kolmogorov's existence theorem. In particular, we will consider measurable spaces, so-called Borel spaces. We will show for a (possibly uncountable) family of Borel spaces and a projective family of probability measures that there exists a unique probability measure on the product space of the family of Borel spaces such that its family of finite-dimensional distributions are the given projective family of probability measures. (Theorem 2.28)

We provide an application of this theorem to the theory of stochastic processes, to be more precise, we show the existence of the Brownian motion (see appendix). Throughout this chapter and the appendix we will heavily rely on the concept of stochastic kernels, which will allow us to consider this problem in a very general manner. As a result we will see that there exist various processes similar to the Brownian motion.

The Kolmogorov existence theorem goes back to Kolmogorov's groundbreaking work "Grundbegriffe der Wahrscheinlichkeitstheorie" from 1932 [Ko]. It is worthwhile noting that sometimes this theorem is referred to as Daniell-Kolmogorov existence theorem. This is due to the fact that Percy J. Daniell, in essence, already showed this theorem in a paper written in 1919, which is part of a 10 papers series in which he developed his theory of Daniell-integrals. Yet, Kolmogorov's result is more general, since it allows index sets of arbitrary cardinality, whereas Daniell's is only for denumerable ones. In 1921, Daniell wrote some work on Brownian motion calling it "dynamic probability". His early works were also a starting point to Wiener's investigation on Brownian motion.

In Chapter 3 we derive two results of local Banach space geometry. We show for a real Banach space $(X, \|\cdot\|)$ that $\|\cdot\|^p$, $1 \leq p \leq 2$, is of negative type, if X is linear isomorphic to a subspace of $L_p(\Omega, \mathbb{P})$. We also show that if $\|\cdot\|^p$, $1 \leq p \leq 2$, is of negative type, then there exists, for each $0 < \beta < p$, a probability space such that X is linear isomorphic to a subspace of $L_\beta(\Omega, \mathbb{P})$. In the case of X being infinite dimensional, the Kolmogorov existence theorem will aid us in proving the second part. This part is based on *Applications de l'étude de certains formes linéaires aléatoires au plongement d'espaces de Banach dans les espaces* L_p by J.Bretagnolle and D.Dacunha-Castelle [B-D]. We follow the presentation of J.Prochno in *Charakteriserungen von Teilräumen von* L_p [P].

We will turn our attention to Talagrand's and Fernique's majorizing measures theorem in Chapter 4. This theorem gives us sharp upper and lower bounds for the supremum of stochastic processes with subgaussian tails.

By this theorem, the boundedness of a subgaussian process is characterized, since, due to H.J. Landau and L.A. Shepp in 1970 [L-S], the boundedness of a stochastic process $(X_t)_{t\in T}$ is equivalent to the boundedness of $\mathbb{E}\sup_{t\in T}|X_t|$. Furthermore, the boundedness of $\mathbb{E}\sup_{t\in T}|X_t|$ is equivalent to the boundedness of $\mathbb{E}\sup_{t\in T} X_t$, which can be easily seen by the inequality

$$\mathbb{E}\sup_{t\in T} X_t \leq \mathbb{E}\sup_{t\in T}|X_t| \leq \mathbb{E}|X_{t_0}| + \mathbb{E}2\sup_{t\in T} X_t,$$

for any $t_0 \in T$, and has the advantage that $\mathbb{E}\sup_{t\in T}(X_t + Y) = \mathbb{E}\sup_{t\in T} X_t$ holds, for any subgaussian random variable Y.

In 1967, R.M. Dudley showed in [D2] that the boundedness of a Gaussian process with a pseudometric space (T, d) as index set, where d is the canonical distance induced by the associated process of covariances (we will see what this exactly means), is implied by the boundedness of the Dudley entropy bound, i.e., by

$$\int_0^\infty \sqrt{\log N_\varepsilon}\, d\varepsilon < \infty,$$

where N_ε is the ε-entropy of T in the metric d. Yet, this fails to characterize the boundedness of the stochastic process, since the ε-entropy does not account for possible lack of homogeneity in (T, d) [T2].

The possible solution of this was established in 1975 by X. Fernique [F]. He proved that the boundedness of the process follows from the boundedness of

$$\sup_{t\in T}\int_0^\infty \log\sqrt{\frac{1}{\mu(B(t,\varepsilon))}}d\varepsilon,$$

where μ is a probability measure on T and $B(t,\varepsilon)$ is the ε-ball centered at t in the metric d, and conjectured that this might characterize the boundedness.

In 1987, Michel Talagrand showed in [T2] that this conjecture is indeed true and elaborated the topic further in a series of papers, which can be tracked down in the references of [T1].

We will be solely concerned with his approach to the proof of the majorizing measures theorem from his work: *Majorizing Measure: The Generic Chaining* [T1]. Astonishingly, the proof of this theorem is based upon a geometric argument, namely, regrouping nearly identical random variables of the stochastic process, which can be expressed through certain increasing sequence of partitions of the index set T. This is elaborated by the so-called "generic chaining"-argument, which goes back to at least Kolmogorov [T1], and which we will see in detail. Furthermore, we will introduce a construction scheme for such sequences of partitions.

Mainly based upon the works *Almost Orthogonal Submatrices of an Orthogonal Matrix* [R1] and *Contact Points of Convex Bodies* [R2] by Mark Rudelson, we will derive in Chapter 5 an application of the majorizing measures theorem to asymptotic convex geometry. We will prove the approximation result that for each n-dimensional convex body B and each $\varepsilon > 0$ there exists another n-dimensional convex body K such that the Banach-Mazur distance of B and K is less or equal than $1+\varepsilon$ and that K and B are having less then $C(\varepsilon)n\log n$ contact points. Furthermore, it will be shown that, if K is embedded in $\mathbb{R}^n$ such that B_2^n is it's John's ellipsoid, then K induces a John's decomposition of the identity operator in $\mathbb{R}^n$ by its contact points with B_2^n.

The classical theorem on John's decomposition, established by Fritz John in 1948 [J], was the starting point to this theorem. It was later, namely in 1992, refined by Keith Ball [Ba]. Yet, it could only bound the number of contact points by $n(n+1)/2$, for symmetric convex bodies, and $n(n+3)/2$, for general ones, respectively.

It is due to a series of papers by Mark Rudelson that this could be lowered significantly. In 1995 to 1997, in a first step, he could give the bound $C(\varepsilon)n(\log n)^3$ in [R2] and [R3]. Two years later, he was able to improve the bound to $C(\varepsilon)n\log n(\log\log n)^2$ before realizing, that he could actually bound it by $C(\varepsilon)n\log n$, which was given in [R1] and will be our main concern.

In 2009, N. Srivastava, D. Spielman and J. Batson managed to do even better. In [S-S-B], they showed that, for symmetric convex bodies, the bound can be lowered to $C(\varepsilon)n$. Moreover, they proved that there exists a convex body K for each convex body B such that their Banach-Mazur distance is smaller than 2.24 and the number of their contact points is bounded by $C(\varepsilon)n$.

Chapter 2

Kolmogorov Existence Theorem

In this chapter, we will provide a proof for the Kolmogorov existence theorem for a certain class of measurable spaces, so-called Borel spaces. First, we need to introduce definitions and notational devices in order to be able to address this problem and work out the proof in a clean manner.

Definition 2.1. (Stochastic process)
Let $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ be a probability space, $(\Omega_2, \mathcal{F}_2)$ be a measure space and T be an index set. A **stochastic process** X is a function

$$X : \Omega_1 \times T \to \Omega_2,\ (\omega, t) \mapsto X_t(\omega),$$

such that $X_t : \Omega_1 \to \Omega_2$ is $\mathcal{F}_1$-$\mathcal{F}_2$-measurable for each $t \in T$.

Remark. To emphasize the view of the stochastic process X as family of random variables we will also write $X = (X_t)_{t \in T}$.

Definition 2.2. (Product space)
Let $(\Omega_t)_{t \in T}$ be an arbitrary family of sets. Then, by the **product space**, $\Omega := \bigtimes_{t \in T} \Omega_t$, we denote the set of all maps

$$\omega : T \to \bigcup_{t \in T} \Omega_t$$

with the property that $\omega(t) \in \Omega_t$, for all $t \in T$.

Notation. If $J \subset T$, then $\Omega_J := \bigtimes_{j \in J} \Omega_j$, and if $T = \mathbb{N}_0$ and $i, j \in T$, then $\Omega_{\{i,\dots,j\}} := \bigtimes_{k=i}^{j} \Omega_k$ will be our notation of choice. In the case that there exists an Ω_0 such that $\Omega_t = \Omega_0$ for all $t \in T$, we write $\Omega = \bigtimes_{t \in T} \Omega_0 = \Omega_0^T$, $\Omega_J = \bigtimes_{j \in J} \Omega_0 = \Omega_0^J$ and $\Omega_{\{i,\dots,j\}} = \bigtimes_{k=i}^{j} \Omega_0 = \Omega_0^{\{i,\dots,j\}}$, respectively.

Definition 2.3. (Coordinate functions)
Let $t \in T$, then $\pi_t : \Omega \to \Omega_t,\ \omega \mapsto \omega(t)$ is the t-th **coordinate function**.
For $J \subset J' \subset T$ the mapping

$$\pi_J^{J'} : \Omega_{J'} \to \Omega_J, \quad \omega \mapsto \omega|_J$$

is called a **canonical projection**.
In the case that $J' = T$ we write π_J. Additionally, π_j^J will be used for the projection from Ω_J to Ω_j.

Definition 2.4. (Product σ-algebra)
Let $(\Omega_t, \mathcal{F}_t),\ t \in T$, be measurable spaces. The **product σ-algebra**

$$\mathcal{F} := \bigotimes_{t \in T} \mathcal{F}_t$$

is the smallest σ-algebra on Ω such that for all $t \in T$ the coordinate function π_t is $\mathcal{F}$-$\mathcal{F}_t$-measurable, i.e.,

$$\mathcal{F} = \sigma(\pi_t, t \in T) := \sigma(\pi_t^{-1}(\mathcal{F}_t), t \in T).$$

Notation. As for product spaces the notation will be $\mathcal{F}_J := \bigotimes_{j \in J} \mathcal{F}_j$ if $J \subset T$, and $\mathcal{F}_{\{i,\dots,j\}} := \bigotimes_{k=i}^{j} \mathcal{F}_k$ if $T = \mathbb{N}_0$. If there exists a $\mathcal{F}_0$ such that $\mathcal{F}_t = \mathcal{F}_0$ for all $t \in T$, we will use $\mathcal{F} = \bigotimes_{t \in T} \mathcal{F}_0 = \mathcal{F}_0^{\otimes T}$, $\mathcal{F}_J = \bigotimes_{j \in J} \mathcal{F}_0 = \mathcal{F}_0^{\otimes J}$ and $\mathcal{F}_{\{i,\dots,j\}} = \bigotimes_{k=i}^{j} \mathcal{F}_0 = \mathcal{F}_0^{\otimes \{i,\dots,j\}}$, respectively.

Now we introduce the concept of projective families of probability measures. As we will see this is central to our investigation, since this criterion tells us how tightly a family of probability measures has to be interlocked in order to induce the existence of some stochastic process.

Definition 2.5. (Joint distribution)
Let $(\Omega_j, \mathcal{F}_j, \mathbb{P}_j)$, $j \in J = \{1, ..., n\}$, $n \in \mathbb{N}$, be probability spaces. The **joint distribution** $\mathbb{P}_J$ is defined by

$$\forall A_j \in \mathcal{F}_j,\ j \in J : \mathbb{P}_J \left(\bigtimes_{j \in J} A_j \right) = \mathbb{P}(A_1 \in \mathcal{F}_1, ..., A_n \in \mathcal{F}_n).$$

Definition 2.6.
Let T be an arbitrary index set. Then by $\mathcal{E}(T) := \{J \subset T : J \text{ finite}\}$, we denote the set of all finite subsets of T.

Definition 2.7. (Projective family)
The family $(\mathbb{P}_J)_{J\in\mathcal{E}(T)}$ of probability measures on $(\Omega_J, \mathcal{F}_J)$ is called a **projective family**, if

$$\forall L \subset J;\ L, J \in \mathcal{E}(T):\ \mathbb{P}_J \circ \left(\pi_L^J\right)^{-1} = \mathbb{P}_L.$$

$\mathbb{P}_J$ is called a **finite-dimensional distribution** for $J \in \mathcal{E}(T)$ and is defined by

From the next corollary it follows that a stochastic process as defined above always admits a projective family of probability measures which are just the finite-dimensional distributions associated with the process. We will see that the Kolmogorov existence theorem tackles the opposite direction of this corollary, namely, that a prescribed projective family induces the existence of a stochastic process on some probability space.

Corollary 2.8.
Let X be a stochastic process from $(\Omega_1, \mathcal{F}_1, \mathbb{P}) \times T$ to $(\Omega_2, \mathcal{F}_2)$. Then the family $(\mathbb{P}_J)_{J\in\mathcal{E}(T)}$ of finite-dimensional distributions is projective.

Proof. Let $J, L \in \mathcal{E}(T)$, $L \subset J$ and define $A := \bigtimes_{\ell\in L} A_\ell$, where $A_\ell \in \mathcal{F}_2$ for all $\ell \in L$. For the joint distribution of $(X_\ell)_{\ell\in L}$ it follows that

$$\begin{aligned}\mathbb{P}_L(A) = \mathbb{P}(\omega_1 \in \Omega_1 : (X_\ell(\omega_1))_{\ell\in L} \in A)\\ &= \mathbb{P}(\omega_1 \in \Omega_1 : (X_j(\omega_1))_{j\in J} \in A \times \Omega_{J\setminus L})\\ &= \mathbb{P}_J(A \times \Omega_{J\setminus L}) = \left(\mathbb{P}_J \circ (\pi_L^J)^{-1}\right)(A).\end{aligned}$$

□

In order to prove results in probability theory concerning σ-algebras, one usually resorts to generators of the aforementioned σ-algebras. Here, we will introduce the cylinder sets. A specially constructed intersection of these cylinder sets, which, as we will see is, an algebra, will then generate the product σ-algebra.
In the proof of the Kolmogorov existence theorem we will make use of this, as this will guarantee that the requirements for Carathéodory's extension theorem (Appendix, Theorem A.12) are satisfied.

Definition 2.9. (Cylinder sets)
Let T be an index set and $J \subset T$. For each $A \in \mathcal{F}_J$ the preimage $\pi_J^{-1}(A) \subset \Omega := \bigtimes_{t\in T} \Omega_t$ of the canonical projection π_J from Ω to Ω_J, is called a **cylinder set** with basis J. The set of all cylinder sets with Basis J is denoted by $\mathcal{Z}_J$.

If $A = \bigotimes_{j\in J} A_j$, for some $A_j \in \mathcal{F}_j$, then $\pi_J^{-1}(A)$ is called a **rectangular cylinder** with basis J. We write $\mathcal{Z}_J^R$ for the set of all rectangular cylinders with basis J. By $\mathcal{Z}$ we denote the set of all cylinder sets with finite basis and by $\mathcal{Z}^R$ the set of all rectangular cylinder sets with finite basis, i.e.,

$$\mathcal{Z} := \bigcup_{J\in\mathcal{E}(T)} \mathcal{Z}_J, \quad \mathcal{Z}^R := \bigcup_{J\in\mathcal{E}(T)} \mathcal{Z}_J^R.$$

Lemma 2.10.
Every $\mathcal{Z}_J$, $J \subset T$, is a σ-algebra and $\mathcal{Z}$ is an algebra.

Proof. By definition $\mathcal{Z}_J$, $J \subset T$, is a set consisting of preimages of the elements of a σ-algebra $\mathcal{F}_J$. Since the preimage operation preserves the properties of a σ-algebra, $\mathcal{Z}_J$ is a σ-algebra since $\mathcal{F}_J$ is one.

Now, we show that $\mathcal{Z}$ is an algebra:

(1) Since, for each $J \subset T$, $\mathcal{Z}_J$ is a σ-algebra, $\Omega \in \mathcal{Z}_J$ for all $J \subset T$. By the definition of $\mathcal{Z}$ it follows that $\Omega \in \mathcal{Z}$.

(2) Let now $B \in \mathcal{Z}$. Then there exists some finite $J \in \mathcal{E}(T)$ such that $B \in \mathcal{Z}_J$. Since $\mathcal{Z}_J$ is a σ-algebra we have $B^c \in \mathcal{Z}_J$ and hence that $B^c \in \mathcal{Z}$.

(3) To show that $\mathcal{Z}$ is closed under finite unions let $B, B' \in \mathcal{Z}$. Then there exist $J, K \in \mathcal{E}(T)$ such that $B \in \mathcal{Z}_J$ and $B' \in \mathcal{Z}_K$. Define $L := J \cup K$. L is clearly finite. Because $\pi_K = \pi_K^L \circ \pi_L$ holds, we get that

$$\mathcal{Z}_K = \pi_K^{-1}(\mathcal{F}_K) = \pi_L^{-1}\left(\left(\pi_K^L\right)^{-1}(\mathcal{F}_K)\right) \subseteq \pi_L^{-1}(\mathcal{F}_L) = \mathcal{Z}_L,$$

and analogously that $\mathcal{Z}_J \subseteq \mathcal{Z}_L$. Hence, $\mathcal{Z}_J \cup \mathcal{Z}_K \subseteq \mathcal{Z}_L$ holds and there exists a finite $L \in \mathcal{E}(T)$ such that $B \cup B' \in \mathcal{Z}_L \subset \mathcal{Z}$.

□

Lemma 2.11.
The cylinder sets $\mathcal{Z}$ generate the product σ-algebra $\mathcal{F}$, i.e., $\sigma(\mathcal{Z}) = \mathcal{F}$.

Proof. First, we prove the inclusion $\mathcal{F} \subseteq \sigma(\mathcal{Z})$.
For all $t \in T$ it holds that $\pi_t = \pi_t^{\{t\}} \circ \pi_{\{t\}}$ and therefore

$$\mathcal{Z}_t = \pi_t^{-1}(\mathcal{F}_t) = \pi_{\{t\}}^{-1}\left(\left(\pi_t^{\{t\}}\right)^{-1}(\mathcal{F}_t)\right) = \pi_{\{t\}}^{-1}(\mathcal{F}_{\{t\}}) = \mathcal{Z}_{\{t\}} \subseteq \mathcal{Z} \subseteq \sigma(\mathcal{Z})$$

holds. Since $\mathcal{F} = \sigma(\pi_t^{-1}(\mathcal{F}_t), t \in T)$ one sees that $\mathcal{F} \subseteq \sigma(\mathcal{Z})$ is true.

To show the second inclusion, $\sigma(\mathcal{Z}) \subseteq \mathcal{F}$, let $J \in \mathcal{E}(T)$. Since π_J is $\mathcal{F}$-$\mathcal{F}_J$-measurable (see Lemma A.1 from the appendix) it holds that $\mathcal{Z}_J = \pi_J^{-1}(\mathcal{F}_J) \subseteq \mathcal{F}$. It follows that

$$\mathcal{Z} = \bigcup_{J \in \mathcal{E}(T)} \mathcal{Z}_J \subseteq \mathcal{F},$$

and consequently that $\sigma(\mathcal{Z}) \subseteq \sigma(\mathcal{F}) = \mathcal{F}$. Hence, $\sigma(\mathcal{Z}) = \mathcal{F}$. □

Lemma 2.12.
The set $\mathcal{Z}^R$ of all rectangular cylinder sets with finite basis is a generator of the product σ-algebra $\mathcal{F}$, i.e., $\sigma(\mathcal{Z}^R) = \mathcal{F}$, and is closed under finite intersections.

Proof. Since $\mathcal{Z}^R \subseteq \mathcal{Z}$, it follows with the second part of Lemma 2.11 that $\sigma(\mathcal{Z}^R) \subseteq \mathcal{F}$.
For the other inclusion we observe that $\mathcal{Z}_{\{t\}} = \mathcal{Z}^R_{\{t\}}$ and $\mathcal{Z}^R_{\{t\}} \subset \mathcal{Z}^R$, for all $t \in T$, and therefore, by the first part of Lemma 2.11, that $\mathcal{F} \subseteq \sigma(\mathcal{Z}^R)$. Hence, $\mathcal{F} = \sigma(\mathcal{Z}^R)$.

Let now $A, B \in \mathcal{Z}^R$, then there exist $J, K \in \mathcal{E}(T)$ and $A'_j \in \mathcal{F}_j$, $j \in J$, and $B'_k \in \mathcal{F}_k$, $k \in K$, such that for $A'_J := \bigtimes_{j \in J} A'_j$ and $B'_K := \bigtimes_{k \in K} B'_k$, it holds that $\pi_J^{-1}(A'_J) = A$ and $\pi_K^{-1}(B'_K) = B$. It follows that

$$\begin{aligned} A \cap B &= \pi_J^{-1}(A'_J) \cap \pi_K^{-1}(B'_K) = \pi_{J\cup K}^{-1}(A'_J \times \Omega_{K\setminus J}) \cap \pi_{J\cup K}^{-1}(B'_K \times \Omega_{J\setminus K}) \\ &= \pi_{J\cup K}^{-1}((A'_J \times \Omega_{K\setminus J}) \cap (B'_K \times \Omega_{J\setminus K})) = \pi_{J\cup K}^{-1}(A'_{J\setminus K} \times B'_{K\setminus J} \times C'_{J\cap K}), \end{aligned}$$

where $C'_{J\cap K} := \bigtimes_{i \in J\cap K} (A'_i \cap B'_i)$ with $A'_i \cap B'_i \in \mathcal{F}_i$, for all $i \in J \cap K$.
Therefore, $A \cap B \in \mathcal{Z}^R_{J\cup K}$ and consequently $A \cap B \in \mathcal{Z}^R$, giving us that $\mathcal{Z}^R$ is closed under finite intersection. □

Lemma 2.13.
The set

$$\widetilde{\mathcal{Z}} := \bigcup_{\substack{J \subset T \\ J \text{ count.}}} \mathcal{Z}_J$$

is a σ-algebra.

Proof. Since $\mathcal{Z}_J$ is a σ-algebra for every subset of T, the first two steps of the proof are the same as (1) and (2) from Lemma 2.10. It remains to show

the closure under countable union.
Let $B_i \in \widetilde{\mathcal{Z}}$, $i \in \mathbb{N}$. Then there exist countable $J_i \in T$ such that $B_{J_i} \in \mathcal{Z}_{J_i}$, for all $i \in \mathbb{N}$. Define $L := \bigcup_{i\in\mathbb{N}} J_i$. L is countable since it is the countable unions of countable sets.
Following (3) from Lemma 2.10 one sees that $\mathcal{Z}_{J_i} \subseteq \mathcal{Z}_L$, for all $i \in \mathbb{N}$ and consequently that $\bigcup_{i\in\mathbb{N}} \mathcal{Z}_{J_i} \subseteq \mathcal{Z}_L$. This means that there exists a countable $L \subset T$ such that

$$\bigcup_{i\in\mathbb{N}} B_i \in \mathcal{Z}_L \subset \bigcup_{\substack{J\subset T\\ J \text{ count.}}} \mathcal{Z}_J = \widetilde{\mathcal{Z}}.$$

□

As a next step, we will establish the concept of transition kernels, which will describe the transition probability of some state of a process at time t_1 to another state of the process at time t_2, and therefore generalizing the concept of transition matrices for Markovian processes.
In other words, transition kernels will give an answer to the question of how to describe a diffusion process which transitions the unit mass in some point $\omega \in \Omega$ into a mass distribution on Ω', prescribed by some measure $A \mapsto \mathbb{P}(\omega, A)$, where $A \in \mathcal{F}$.

Definition 2.14. (Transition kernel/ Stochastic kernel)
Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measure spaces. The map $\kappa : \Omega_1 \times \mathcal{F}_2 \to [0, \infty]$ is called a (σ-)finite **transition kernel** from Ω_1 to Ω_2 if

(i) $\omega_1 \mapsto \kappa(\omega_1, A_2)$ is $\mathcal{F}_1$-measurable for all $A_2 \in \mathcal{F}_2$, and

(ii) $A_2 \mapsto \kappa(\omega_1, A_2)$ is a (σ-)finite measure on $(\Omega_2, \mathcal{F}_2)$ for all $\omega_1 \in \Omega_1$.

If for all $\omega_1 \in \Omega_1$ the measure in (ii) is a probability measure, then κ is called a **stochastic kernel**.

Notation. We write $\kappa(d\omega_2) := \kappa(\omega_1, d\omega_2)$ if a kernel κ is independent of ω_1.

Corollary 2.15.
It suffices to claim the $\mathcal{F}_1$-measurability of $\omega_1 \mapsto \kappa(\omega_1, A_2)$ only for $\mathcal{E}$ in (i) of the definition of transition kernels, where $\mathcal{E}$ is a generator of $\mathcal{F}_2$ that is closed under finite intersections and such that there exists a sequence $(E_n)_{n\in\mathbb{N}}$ with $E_n \in \mathcal{E}$ and $E_n \uparrow \Omega$.

Proof. Showing that $\mathcal{D} := \{A_2 \in \mathcal{F}_2 : \omega_1 \mapsto \kappa(\omega_1, A_2) \text{ is } \mathcal{F}_1\text{-measurable}\}$ is a Dynkin system is the only thing that has to be done here (for the definition

of a Dynkin system see appendix, Definition A.2). From Dynkin's theorem (Appendix, Theorem A.3) it follows that $\delta(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{F}_2$.
Additionally, by $\mathcal{E} \subset \mathcal{D} \subseteq \mathcal{F}_2$, and therefore $\delta(\mathcal{E}) \subseteq \mathcal{D} \subseteq \mathcal{F}_2$, it follows that $\mathcal{D} = \mathcal{F}_2 = \sigma(\mathcal{E})$. The way $\mathcal{E}$ was chosen guarantees with Lemma A.4 from the appendix, that the kernel κ is uniquely determined.

Therefore, we need to show that $\mathcal{D}$ is a Dynkin system.

(i) Let $A', A \in \mathcal{D}$ with $A \subset A'$. For fixed $\omega_1 \in \Omega_1$, by (ii) of the definition, $\kappa(\omega_1, A' \setminus A)$ is the measure of the set $A' \setminus A$ w.r.t. the σ-finite measure corresponding to $\omega_1 \in \Omega_1$. It follows that

$$\forall \omega_1 \in \Omega_1 : \kappa(\omega_1, A' \setminus A) = \kappa(\omega_1, A') - \kappa(\omega_1, A),$$

and therefore that $\kappa(\,\cdot\,, A' \setminus A) = \kappa(\,\cdot\,, A') - \kappa(\,\cdot\,, A)$ is $\mathcal{F}_1$-measurable, because $\kappa(\,\cdot\,, A')$, $\kappa(\,\cdot\,, A)$ are $\mathcal{F}_1$-measurable. Hence, $A' \setminus A \in \mathcal{D}$.

(ii) Let $A_n \in \mathcal{D}$, $n \in \mathbb{N}$, pairwise disjoint and let $A := \bigcup_{n\in\mathbb{N}} A_n$. For fixed $\omega_1 \in \Omega_1$, it follows by the same argument as above and σ-additivity that

$$\forall \omega_1 \in \Omega_1 : \kappa(\omega_1, A) = \kappa(\omega_1, \bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \kappa(\omega_1, A_n),$$

and that $\kappa(\,\cdot\,, A) = \sum_{n\in\mathbb{N}} \kappa(\,\cdot\,, A_n)$ is $\mathcal{F}_1$-measurable, because it is the countable sum of $\mathcal{F}_1$-measurable functions. Hence, $A = \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{D}$.

(iii) Using the fact that $\mathcal{E} \subset \mathcal{D}$ and that there exists a sequence $(E_n)_{n\in\mathbb{N}}$ in $\mathcal{E}$ with $E_n \uparrow \Omega_2$, we can define disjoint sets $A_1 := E_1$ and $A_n := E_n \setminus E_{n-1}$ for $n \geq 2$. Clearly, $\bigcup_{n\in\mathbb{N}} A_n = \Omega_2$.
By (i), $\kappa(\,\cdot\,, A_1) = \kappa(\,\cdot\,, E_1)$ and $\kappa(\,\cdot\,, A_n) = \kappa(\,\cdot\,, E_n) - \kappa(\,\cdot\,, E_{n-1})$ are $\mathcal{F}_1$-measurable for $n \in \mathbb{N}$. By (ii) it follows that

$$\kappa(\,\cdot\,, \Omega_2) = \kappa(\,\cdot\,, \bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \kappa(\,\cdot\,, A_n)$$

is $\mathcal{F}_1$-measurable and therefore $\Omega_2 \in \mathcal{D}$.

□

Lemma 2.16. (Product of kernels)
Let $(\Omega_i, \mathcal{F}_i)$, $i = 0, 1, 2$, be measurable spaces and κ_1, κ_2 be finite kernels from $(\Omega_0, \mathcal{F}_0)$ to $(\Omega_1, \mathcal{F}_1)$ and from $(\Omega_0 \times \Omega_1, \mathcal{F}_0 \otimes \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$, respectively. If we denote by

$$(\kappa_1 \otimes \kappa_2)(\omega_0, A) := \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_A(\omega_1, \omega_2) \kappa_2((\omega_0, \omega_1), d\omega_2) \kappa_1(\omega_0, d\omega_1),$$

with $\omega_0 \in \Omega_0$ and $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$, the **product of the kernels** of κ_1 and κ_2, then $\kappa_1 \otimes \kappa_2$ forms a σ-finite kernel (not necessarily finite) from $(\Omega_0, \mathcal{F}_0)$ to $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, which is stochastic if κ_1 and κ_2 are stochastic.

To prove this lemma, we require another result.

Lemma 2.17. Let κ be a finite transition kernel from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ and let $f : \Omega_1 \times \Omega_2 \to [0, \infty]$ be $\mathcal{F}_1 \otimes \mathcal{F}_2$-$\mathcal{B}([0,\infty])$-measurable. Then

$$I_f : \Omega_1 \to [0, \infty], \quad \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\kappa(\omega_1, d\omega_2)$$

is well-defined and $\mathcal{F}_1$-measurable.

Proof. First, we show that $I_f(\omega_1)$ is well-defined.
Define $i_{\omega_1} : \Omega_2 \to \Omega_1 \times \Omega_2$ by $i_{\omega_1}(\omega_2) = (\omega_1, \omega_2)$ with some fixed $\omega_1 \in \Omega_1$. Since $\pi_2 \circ i_{\omega_1} = id_{\Omega_2}$ and $\pi_1 \circ i_{\omega_1} = \omega_1 \mathbb{1}_{\Omega_1}$ holds, it follows that i_{ω_1} is $\mathcal{F}_2$-$\mathcal{F}_2$- and $\mathcal{F}_2$-$\mathcal{F}_1$-measurable. We get that i_{ω_1} is $\mathcal{F}_2$-$\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable (Lemma A.1 from the appendix).
Together with the $\mathcal{F}_1 \otimes \mathcal{F}_2$-$\mathcal{B}([0,\infty])$-measurability of f it holds that $f \circ i_{\omega_1}$ is $\mathcal{F}_2$-$\mathcal{B}([0,\infty])$-measurable for every $\omega_1 \in \Omega_1$ and consequently that I_f is well-defined.

To show that I_f is $\mathcal{F}_1$-measurable define $g := \mathbb{1}_{A_1 \times A_2}$. Then it holds that

$$I_g(\omega_1) = \int_{\Omega_2} \mathbb{1}_{A_1 \times A_2}(\omega_1, \omega_2)\kappa(\omega_1, d\omega_2) = \mathbb{1}_{A_1}(\omega_1)\kappa(\omega_1, A_2)$$

is $\mathcal{F}_1$-measurable, because $\mathbb{1}_{A_1}(\omega_1)$ and $\kappa(\omega_1, A_2)$, for every $A_2 \in \mathcal{F}_2$, are $\mathcal{F}_1$-measurable.
Define $\mathcal{D} := \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 : I_{\mathbb{1}_A} \text{ is } \mathcal{F}_1\text{-measurable}\}$. $\mathcal{D}$ is a Dynkin system:

(i) Obviously $\Omega_1 \times \Omega_2 \in \mathcal{D}$.

(ii) Let $A, B \in \mathcal{D}$ with $A \subset B$, then we have that $I_{\mathbb{1}_{B \setminus A}} = I_{\mathbb{1}_B} - I_{\mathbb{1}_A}$ holds because κ is finite and $\mathbb{1}_{B \setminus A} = \mathbb{1}_B - \mathbb{1}_A$ holds. Thereby $B \setminus A \in \mathcal{D}$.

(iii) Let $A_n \in \mathcal{D}$, $n \in \mathbb{N}$, pairwise disjoint and $A := \bigcup_{n \in \mathbb{N}} A_n$. By the fact that $\mathbb{1}_A = \mathbb{1}_{\bigcup_{n \in \mathbb{N}} A_n} = \sum_{n \in \mathbb{N}} \mathbb{1}_{A_n}$, it follows that $I_{\mathbb{1}_A} = \sum_{n \in \mathbb{N}} I_{\mathbb{1}_{A_n}}$ and therefore that $A \in \mathcal{D}$.

By the definition of $\mathcal{D}$ one sees that the set of rectangular cylinders $\mathcal{Z}^R$ is contained in $\mathcal{D}$. From Lemma 2.12 we know that $\mathcal{Z}^R$ is a generator of the product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ and closed under finite intersections. By Dynkin's theorem (see Appendix, Theorem A.3) it follows that $\mathcal{D} = \mathcal{F}_1 \otimes \mathcal{F}_2$.
Hence, $\mathbb{1}_A$ is $\mathcal{F}_1$-measurable for each $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$. It follows for every step function g that I_g is $\mathcal{F}_1$-measurable. Let now $(f_n)_{n\in\mathbb{N}}$ be a monotonously growing sequence of step functions. By the monotone convergence theorem (see Appendix, Theorem A.5), we get that $I_f(\omega_1) = \lim_{n\to\infty} I_{f_n}(\omega_1)$, for every $\omega_1 \in \Omega_1$, and that I_f is measurable, since it is the limit of measurable functions. □

Having shown this lemma, we can now go over to show the result about products of kernels.

Proof. (Product of kernels)
Let $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$. By Lemma 2.17, it follows that the map

$$g_A : (\omega_0, \omega_1) \mapsto \int\limits_{\Omega_2} \mathbb{1}_A(\omega_1, \omega_2)\kappa_2((\omega_0, \omega_1), d\omega_2)$$

is well-defined and measurable w.r.t $\mathcal{F}_1 \otimes \mathcal{F}_2$. Again, by Lemma 2.17, the map

$$\omega_0 \mapsto (\kappa_1 \otimes \kappa_2)(\omega_0, A) = \int\limits_{\Omega_1} g_A(\omega_0, \omega_1)\kappa_1(\omega_0, d\omega_1)$$

is well-defined and $\mathcal{F}_0$-measurable.

Let now $A_n \in \mathcal{F}_1 \otimes \mathcal{F}_2$, $n \in \mathbb{N}$, be pairwise disjoint and $A := \bigcup_{n\in\mathbb{N}} A_n$. For fixed ω_0, we have that

$$\begin{aligned}(\kappa_1 \otimes \kappa_2)(\omega_0, A) &= \int\limits_{\Omega_1}\int\limits_{\Omega_2} \mathbb{1}_A(\omega_1, \omega_2)\kappa_1(\omega_0, d\omega_1)\kappa_2((\omega_0, \omega_1), d\omega_2)\\ &= \int\limits_{\Omega_1}\int\limits_{\Omega_2} \mathbb{1}_{\bigcup\limits_{n\in\mathbb{N}} A_n}(\omega_1, \omega_2)\kappa_1(\omega_0, d\omega_1)\kappa_2((\omega_0, \omega_1), d\omega_2)\\ &= \int\limits_{\Omega_1}\int\limits_{\Omega_2} \sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(\omega_1, \omega_2)\kappa_1(\omega_0, d\omega_1)\kappa_2((\omega_0, \omega_1), d\omega_2),\end{aligned}$$

and by the monotone convergence theorem (Appendix, Theorem A.5) that

$$= \sum_{n\in\mathbb{N}} \int\limits_{\Omega_1}\int\limits_{\Omega_2} \mathbb{1}_{A_n}(\omega_1, \omega_2)\kappa_1(\omega_0, d\omega_1)\kappa_2((\omega_0, \omega_1), d\omega_2) = \sum_{n\in\mathbb{N}} (\kappa_1 \otimes \kappa_2)(\omega_0, A_n).$$

Therefore, $A \mapsto (\kappa_1 \otimes \kappa_2)(\omega_0, A)$ is σ-additive and, hence, a measure.

For $\omega_0 \in \Omega_0$ and $n \in \mathbb{N}$ define $A_{\omega_0,n} := \{\omega_1 \in \Omega_1 : \kappa_2((\omega_0, \omega_1), \Omega_2) < n\}$. Since κ_2 is finite, it holds that $\bigcup_{n\in\mathbb{N}} A_{\omega_0,n} = \Omega_1$, for every $\omega_0 \in \Omega_0$. Furthermore, it holds that

$$\begin{aligned}(\kappa_1 \otimes \kappa_2)(\omega_0, A_{\omega_0,n} \times \Omega_2) &= \int\limits_{\Omega_1}\int\limits_{\Omega_2} \mathbb{1}_{A_{\omega_0,n}}(\omega_1)\mathbb{1}_{\Omega_2}(\omega_2)\kappa_1(\omega_0, d\omega_1)\kappa_2((\omega_0, \omega_1), d\omega_2)\\ &= \kappa_1(\omega_0, A_{\omega_0,n})\kappa_2((\omega_0, \omega_1), \Omega_2)\\ &< \kappa_1(\omega_0, A_{\omega_0,n}) \cdot n < \infty,\end{aligned}$$

where the last inequality follows from the finiteness of κ_1.
Therefore, $(\kappa_1 \otimes \kappa_2)(\omega_0, \cdot)$ is σ-finite and $\kappa_1 \otimes \kappa_2$ a transition kernel.

Let $\omega_0 \in \Omega_0$ still be fixed. Then, $\kappa_2((\omega_0, \cdot), \Omega_2)$ is not necessarily bounded, even if κ_2 is a finite kernel. Together with

$$(\kappa_1 \otimes \kappa_2)(\omega_0, \Omega_1 \times \Omega_2) = \kappa_1(\omega_0, \Omega_1) \int\limits_{\Omega_2} \mathbb{1}_{\Omega_2}(\omega_2)\kappa_2((\omega_0, \omega_1), d\omega_2),$$

one sees that $(\kappa_1 \otimes \kappa_2)(\omega_0, \cdot)$ is not necessarily finite and consequently the kernel $\kappa_1 \otimes \kappa_2$ is not necessarily finite.

If κ_1 and κ_2 are stochastic kernels, $\kappa_1(\omega_0, \Omega_1) = 1$ and $\kappa_2((\omega_0, \omega_1), \Omega_2) = 1$ holds. For fixed ω_0 it follows that

$$(\kappa_1 \otimes \kappa_2)(\omega_0, \Omega_1 \times \Omega_2) = \kappa_1(\omega_0, \Omega_1) \cdot \kappa_2((\omega_0, \omega_1), \Omega_2) = 1$$

and hence that $\kappa_1 \otimes \kappa_2$ is a stochastic kernel. □

Remark. For sets of the form $A = A_1 \times A_2$, $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$, the product simplifies to

$$(\kappa_1 \otimes \kappa_2)(\omega_0, A_1 \times A_2) := \int\limits_{A_1} \kappa_2((\omega_0, \omega_1), A_2)\kappa_1(\omega_0, d\omega_1).$$

If the kernel κ_1 is independent of ω_0, it is a probability measure on $(\Omega_1, \mathcal{F}_1)$ and the product $\kappa_1 \otimes \kappa_2$ is a probability measure on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$:

$$(\kappa_1 \otimes \kappa_2)(A) := \int\limits_{\Omega_1}\int\limits_{\Omega_2} \mathbb{1}_A(\omega_1, \omega_2)\kappa_2((\omega_0, \omega_1), d\omega_2)\kappa_1(d\omega_1),$$

where $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$. If, additionally, κ_2 is independent of ω_1, $\kappa_1 \otimes \kappa_2$ simplifies to the ordinary product measure of κ_1 and κ_2.

Remark. The definition of the product of two kernels can be extended inductively to more than two factors in the following way:

$$\bigotimes_{j=1}^{n} \kappa_j = \left(\bigotimes_{j=1}^{n-1} \kappa_j \right) \otimes \kappa_n,$$

where κ_j, for $j \geq 2$, is a kernel from $(\Omega_{\{1,...,j-1\}}, \mathcal{F}_{\{1,...,j-1\}})$ to $(\Omega_j, \mathcal{F}_j)$. Since this product is associative, it follows, for all $k = 1, ..., n-1$, that

$$\bigotimes_{j=1}^{n} \kappa_j = \left(\bigotimes_{j=1}^{k} \kappa_j \right) \otimes \left(\bigotimes_{j=k+1}^{n} \kappa_j \right).$$

In the proof of the Kolmogorov existence theorem, we will need that a random variable Y, together with a sub-σ-algebra $\mathcal{F}$, induces the existence of corresponding stochastic kernels. These kernels will be called regular versions of the conditional distribution, or in short: regular conditional distributions.

First, consider a random variable X with values in a measure space $(\Omega_2, \mathcal{F}_2)$. For $A \in \mathcal{F}_1$ we can express the conditional probability under X as $\mathbb{E}(A|X)$ (see appendix, Definition A.9). Now let Z be a $\sigma(X)$-measurable random variable. By the factorization lemma (Appendix, Theorem A.8) it follows that there exists a $\mathcal{F}_2 - \mathcal{B}(\mathbb{R})$-measurable map $\varphi : \Omega_2 \to \mathbb{R}$ with $\varphi(X) = Z$. If X is onto, then φ is uniquely determined and we write $Z \circ X^{-1} = \varphi$ (even if X^{-1} does not exist).
Thereby, we can explain the conditional expectation $\mathbb{E}(A|X = x)$ of A under $X = x$, for each $A \in \mathcal{F}_1$, as in the next definition and furthermore such that $\mathbb{E}(A|X) = \mathbb{E}(A|X = x)$ holds on $\{X = x\}$.

Definition 2.18. (Conditional expectation)
Let $Y \in \mathcal{L}^1(\Omega_1, \mathcal{F}_1, \mathbb{P})$ and let $X : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$. The **conditional expectation** of Y given $X = x$, in short $\mathbb{E}(Y|X = x)$, is defined as the function φ from the factorization theorem (Appendix, Theorem A.8) with $Z = \mathbb{E}(Y|X)$ as explained above.
Additionally, we define the **conditional probability** of $A \in \mathcal{F}_1$ given $X = x$ as $\mathbb{P}(A|X = x) = \mathbb{E}(\mathbb{1}_A|X = x)$.

Definition 2.19. (Regular conditional distribution)
Let Y be a random variable from $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ to $(\Omega_2, \mathcal{F}_2)$ and $\mathcal{F} \subset \mathcal{F}_1$ a sub-σ-algebra. A stochastic kernel $\kappa_{Y,\mathcal{F}}$ from $(\Omega_1, \mathcal{F})$ to $(\Omega_2, \mathcal{F}_2)$ is called a **regular version of the conditional distribution** of Y given $\mathcal{F}$, if it holds that $\kappa_{Y,\mathcal{F}}(\omega_1, B) = \mathbb{P}(\{Y \in B\}|\mathcal{F})(\omega_1)$ for $\mathbb{P}$-almost all $\omega_1 \in \Omega_1$ and for every

$B \in \mathcal{F}_2$.
Let now $\mathcal{F} = \sigma(X)$, where X is a another random variable. Then the kernel $(\omega_1, B) \mapsto \kappa_{Y,X}(\omega_1, B) = \mathbb{P}(\{Y \in B\}|X = x) = \kappa_{Y,\sigma(X)}(X^{-1}(\omega_1), B)$ is called a **regular version of the conditional distribution** of Y given X.

Theorem 2.20. (Regular conditional distribution)
Let Y be a random variable from $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\mathcal{F} \subset \mathcal{F}_1$ a sub-σ-algebra. Then there exists a regular version $\kappa_{Y,\mathcal{F}}$ of the conditional distribution $\mathbb{P}(\{Y \in \ \cdot \ \}|\mathcal{F})$.

Proof. We are going to construct a measurable version of the distribution function of the conditional distribution of Y by defining its values on $\mathbb{Q}$ (up to a Null set) and then extending this to $\mathbb{R}$.

Let $F(r, \cdot)$ be a version of the conditional probability $\mathbb{P}(Y \in (-\infty, r]|\mathcal{F})$ for $r \in \mathbb{Q}$. For $r \leq s$ it holds that $\mathbb{1}_{\{Y \in (-\infty, r]\}} \leq \mathbb{1}_{\{Y \in (-\infty, s]\}}$, and therefore, by the monotonicity of the conditional expectation (Appendix, Theorem A.10), that there exists a Null set $A_{r,s} \in \mathcal{F}$ and that following holds

$$\forall \omega_1 \in \Omega \setminus A_{r,s} : \ F(r, \omega_1) \leq F(s, \omega_1).$$

From the dominated convergence theorem for conditional expectations (Appendix, Theorem A.10) it follows that there exist Null sets $B_r \in \mathcal{F}$, $r \in \mathbb{Q}$, and $C \in \mathcal{F}$ such that for all $\omega_1 \in \Omega \setminus B_r$ it holds that

$$\begin{aligned}
\lim_{n \to \infty} F\left(r + \frac{1}{n}, \omega_1\right) &= \lim_{n \to \infty} \mathbb{P}\left(Y \in \left(-\infty, r + \frac{1}{n}\right] |\mathcal{F}\right) \\
&= \lim_{n \to \infty} \mathbb{P}\left(Y - \frac{1}{n} \in (-\infty, r] \,|\mathcal{F}\right) \\
&= \mathbb{P}\left(Y \in (-\infty, r] \,|\mathcal{F}\right) \\
&= F(r, \omega_1),
\end{aligned}$$

since $Y \in \mathcal{L}(\Omega, \mathcal{F}, \mathbb{P})$ and $Y - \frac{1}{n} \to Y$, and consequently that

$$\forall \omega_1 \in \Omega \setminus B_r : \ \lim_{n \to \infty} F\left(r + \frac{1}{n}, \omega_1\right) = F(r, \omega_1).$$

Furthermore, we have that

$$\forall \omega_1 \in \Omega \setminus C : \ \lim_{n \to \infty} F(-n, \omega_1) = 0 \qquad \text{and} \qquad \lim_{n \to \infty} F(n, \omega_1) = 1.$$

Set now $N := \bigcup_{r,s \in \mathbb{Q}} A_{r,s} \cup \bigcup_{r \in \mathbb{Q}} B_r \cup C$ and for all $\omega_1 \in \Omega_1 \setminus C$ define

$$\tilde{F}(z, \omega_1) := \inf\{F(r, \omega_1) : \ r \in \mathbb{Q}, \ r \geq z\}.$$

Since $F(\cdot, \omega_1)$ is monotonously growing and right-continuous, $\tilde{F}(\cdot, \omega_1)$ is also monotonously growing on $\mathbb{R} \setminus \mathbb{Q}$ and right-continuous on $\mathbb{Q}$. Consequently, $\tilde{F}(\cdot, \omega_1)$ is a distribution function for every $\omega_1 \in \Omega_1 \setminus N$.
For the remaining $\omega_1 \in N$ set $\tilde{F}(\cdot, \omega_1) := F_0$, where F_0 is an arbitrary but fixed distribution function.

We proceed by defining $\kappa(\omega_1, \cdot)$, for every $\omega_1 \in \Omega_1$, as the measure corresponding to the distribution function $\tilde{F}(\cdot, \omega_1)$ on $(\Omega_1, \mathcal{F})$.
For $r \in \mathbb{Q}$ and $B := (-\infty, r]$

$$\omega_1 \mapsto \kappa(\omega_1, B) = \mathbb{P}(Y \in B|\mathcal{F})\mathbb{1}_{N^c}(\omega_1) + F_0(r)\mathbb{1}_N(\omega_1) \tag{2.1}$$

is clearly $\mathcal{F}$-measurable.
Since $\{(-\infty, r] : r \in \mathbb{Q}\}$ is a generator of $\mathcal{B}(\mathbb{R})$, which is closed under finite intersections and contains the sequence $(-\infty, r] \uparrow \mathbb{R}$, we get by Corollary 2.15 that the $\mathcal{F}$-measurability holds for all of $\mathcal{B}(\mathbb{R})$. Hence, κ is a stochastic kernel.

It remains to show that κ is a version of the conditional distribution. For $A \in \mathcal{F}$, $r \in \mathbb{Q}$ and $B = (-\infty, r]$ it follows from equation (2.1) that

$$\int_A \kappa(\omega_1, B)\mathbb{P}(d\omega_1) = \int_A \mathbb{P}(Y \in B|\mathcal{F})d\mathbb{P} = \mathbb{P}(A \cap \{Y \in B\}).$$

As a function of B, both sides are finite measures on $\mathcal{B}(\mathbb{R})$, which coincide on the generator $\{(-\infty, r] : \ r \in \mathbb{Q}\}$. By the uniqueness theorem for measures (Appendix, Theorem A.11) it follows that for every $B \in \mathcal{B}(\mathbb{R})$ the equality holds and therefore that $\mathbb{P}$-a.s. $\kappa(\ \cdot\ , B) = \mathbb{P}(Y \in B|\mathcal{F})$. Hence, $\kappa = \kappa_{Y,\mathcal{F}}$. □

Lemma 2.21.
Let $(\Omega_i, \mathcal{F}_i)$, $i \in \mathbb{N}_0$, be measurable spaces and let $\mathbb{P}_0$ be a probability measure on $(\Omega_0, \mathcal{F}_0)$. Furthermore, let κ_i, $i \in \mathbb{N}$, be stochastic kernels from $(\Omega_{\{0,\dots,i-1\}}, \mathcal{F}_{\{0,\dots,i-1\}})$ to $(\Omega_i, \mathcal{F}_i)$. Define probability measures $\mathbb{P}_i$ on $(\Omega_{\{0,\dots,i\}}, \mathcal{F}_{\{0,\dots,i\}})$ recursively by $\mathbb{P}_i = \mathbb{P}_0 \otimes \bigotimes_{\ell=1}^{i} \kappa_\ell$. Then, for all $i, j \geq k$ and $A \in \mathcal{F}_{\{0,\dots,k\}}$, the following holds:

$$\mathbb{P}_i(A \times \Omega_{\{k+1,\dots,i\}}) = \mathbb{P}_j(A \times \Omega_{\{k+1,\dots,j\}})).$$

Proof. Since $\mathbb{P}_0$ is a probability measure on $(\Omega_0, \mathcal{F}_0)$, we know from the remark about stochastic kernels that $\mathbb{P}_k$ is likewise a probability measure on $(\Omega_{\{0,\dots,k\}}, \mathcal{F}_{\{0,\dots,k\}})$, for all $k \in \mathbb{N}$. Let $k, n \in \mathbb{N}_0$, $k \leq n$, then $A \times \Omega_{\{i+1,\dots,n\}} \in$

$\mathcal{F}_{\{0,...,n\}}$, for all $A \in \mathcal{F}_{\{0,...,i\}}$. Additionally, the following holds for all $n \geq k$

$$\begin{aligned}\mathbb{P}_n(A \times \Omega_{\{k+1,...,n\}}) &= \int\limits_{A\times\Omega_{\{k+1,...,n\}}} \mathbb{P}_n(d(\omega_0, ..., \omega_n)) \\ &= \int\limits_{A\times\Omega_{\{k+1,...,n\}}} \kappa_{n+1}(\omega_n, \Omega_{n+1})\mathbb{P}_n(d(\omega_0, ..., \omega_n)) \\ &= (\mathbb{P}_n \otimes \kappa_{n+1})(A \times \Omega_{\{k+1,...,n+1\}}) \\ &= \mathbb{P}_{n+1}(A \times \Omega_{\{k+1,...,n+1\}}),\end{aligned}$$

where we have used that $\kappa_{n+1}(\omega_n, \Omega_{n+1}) = 1$ for every $n \in \mathbb{N}$.
By induction we get the above lemma. □

With the tools we have gathered so far, we are now able to show Ionescu-Tulcea's theorem, which will be used to prove the Kolmogorov existence theorem.

Theorem 2.22. (Ionescu-Tulcea)
Let the assumptions from Lemma 2.21 hold. Then, there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{P}(A \times \Omega_{\{k+1,...,\infty\}}) = \mathbb{P}_k(A),$$

for all $k \in \mathbb{N}_0$ and $A \in \mathcal{F}_{\{0,...,k\}}$.

Proof. From Lemma 2.11 we know that the set of cylinder sets $\mathcal{Z}$ generates the product σ-algebra. Additionally we know by Lemma 2.10 that $\mathcal{Z}$ is an algebra and therefore closed under finite intersections. By the uniqueness theorem for measures (see Appendix, Theorem A.11) it follows that it suffices to show the unique existence of a probability measure for $\mathcal{Z}$.
We will show the existence: Define a function $\mathbb{P}$ on $\mathcal{Z}$ by

$$\forall k \in \mathbb{N}_0\ \forall A \in \mathcal{F}_{\{0,...,k\}} :\ \mathbb{P}(A \times \Omega_{\{k+1,...,\infty\}}) := \mathbb{P}_k(A).$$

First, we show that $\mathbb{P}$ is additive on $\mathcal{Z}$.
Let $k, \ell \in \mathbb{N}_0$, $k \leq l$, and let $A \in \mathcal{F}_{\{0,...,k\}}$, $B \in \mathcal{F}_{\{0,...,\ell\}}$ such that $A \cap B = \emptyset$.
By Lemma 2.21, we see that $\mathbb{P}_k(A) = \mathbb{P}_\ell(A \times \Omega_{\{k+1,...,\ell\}})$ and, therefore,

$$\begin{aligned}&\mathbb{P}((A \times \Omega_{\{k+1,...,\infty\}}) \cup (B \times \Omega_{\{\ell+1,...,\infty\}})) \\ &= \mathbb{P}_\ell((A \times \Omega_{\{k+1,...,\ell\}}) \cup B) \\ &= \mathbb{P}_\ell((A \times \Omega_{\{k+1,...,\ell\}})) + \mathbb{P}_\ell(B) \\ &= \mathbb{P}_k(A) + \mathbb{P}_\ell(B) \\ &= \mathbb{P}(A \times \Omega_{\{k+1,...,\infty\}}) + \mathbb{P}(B \times \Omega_{\{l+1,...,\infty\}}),\end{aligned}$$

where we have used that $(A \times \Omega_{\{k+1,...,\ell\}}) \cap B = \emptyset \iff A \cap B = \emptyset$ and that $\mathbb{P}_\ell$ is σ-additive.

Since $\mathbb{P}$ is an additive function on the algebra $\mathcal{Z}$, it is a content. If, additionally, it holds that $\mathbb{P}$ is σ-additive on $\mathcal{Z}$, then $\mathbb{P}$ is a pre-measure. By applying the Carathéodory extension theorem (Appendix, Theorem A.12), $\mathbb{P}$ can be uniquely extended to a measure on $\mathcal{F}_{\{0,...,k\}}$.
Therefore it is necessary to show the σ-additivity of $\mathbb{P}$. Since σ-additivity and $\emptyset$-continuity are equivalent for algebras, we can show $\emptyset$-continuity of $\mathbb{P}$ in order to proof this theorem. (For the definition of $\emptyset$-continuity and the lemma with the mentioned equivalence, see Appendix, Definition A.13 and Lemma A.14).
Let $A_0 \supset A_1 \supset ...$ be a sequence in $\mathcal{Z}$ and define $\alpha := \inf_{n\in\mathbb{N}_0} \mathbb{P}(A_n) > 0$. We need to show that

$$\bigcap_{n=0}^{\infty} A_n \neq \emptyset.$$

W.l.o.g. we can assume $A_n = A'_n \times \Omega_{\{n+1,...,\infty\}}$ for some $A'_n \in \mathcal{F}_{\{0,...,n\}}$. For $n \geq m$ we set

$$h_{m,n}(\omega_0, ..., \omega_m) := (\bigotimes_{k=m+1}^{n} \kappa_k)((\omega_0, ..., \omega_m), A'_n)$$

and $h_m := \inf_{n\geq m} h_{m,n}$. We show inductively that there exists a $\rho_i \in \Omega_i$, $i \in \mathbb{N}_0$, with

$$h_m(\rho_0, ..., \rho_m) \geq \alpha. \tag{2.2}$$

Because $A'_{n+1} \subset A'_n \times \Omega_{n+1}$, it holds that

$$\begin{aligned} h_{m,n+1}(\omega_0, ..., \omega_m) &= (\bigotimes_{k=m+1}^{n+1} \kappa_k)((\omega_0, ..., \omega_m), A'_{n+1}) \\ &\leq (\bigotimes_{k=m+1}^{n+1} \kappa_k)((\omega_0, ..., \omega_m), A'_n \times \Omega_{n+1}) \\ &= (\bigotimes_{k=m+1}^{n} \kappa_k)((\omega_0, ..., \omega_m), A'_n) \\ &= h_{m,n}(\omega_0, ..., \omega_m) \end{aligned}$$

Therefore, we have $h_{m,n} \downarrow h_m$ for $n \to \infty$. The monotone convergence theorem (Appendix, Theorem A.5) then implies

$$\int h_m d\mathbb{P}_m = \inf_{n\geq m} \int h_{m,n} d\mathbb{P}_m = \inf_{n\geq m} \mathbb{P}_n(A'_n) = \inf_{n\geq m} \mathbb{P}(A_n) \overset{m=0}{=} \alpha.$$

By this, equation (2.2) is shown for $m = 0$. Assume now that (2.2) is true for $m \in \mathbb{N}$, then, we get that

$$\begin{aligned}
&\int h_{m+1}(\rho_0, ..., \rho_m, \omega_{m+1})\kappa_{m+1}((\rho_0, ..., \rho_m), d\omega_{m+1}) \\
&= \inf_{n \geq m+1} \int h_{m+1,n}(\rho_0, ..., \rho_m, \omega_{m+1})\kappa_{m+1}((\rho_0, ..., \rho_m), d\omega_{m+1}) \\
&= \inf_{n \geq m+1} \int (\bigotimes_{k=m+2}^{n} \kappa_k)((\rho_0, ..., \rho_m, \omega_{m+1}), A'_n)\kappa_{m+1}((\rho_0, ..., \rho_m), d\omega_{m+1}),
\end{aligned}$$

and, by the first remark about stochastic kernels, that

$$\begin{aligned}
&\inf_{n \geq m+1} (\kappa_{m+1} \otimes \bigotimes_{k=m+2}^{n} \kappa_k)((\rho_0, ..., \rho_m), A'_n) \\
&= \inf_{n \geq m+1} (\bigotimes_{k=m+1}^{n} \kappa_k)((\rho_0, ..., \rho_m), A'_n) \\
&= \inf_{n \geq m+1} h_{m,n}(\rho_0, ..., \rho_m) = h_m(\rho_0, ..., \rho_m) \\
&\geq \alpha.
\end{aligned}$$

By induction we get that equation (2.2) holds for all $m \in \mathbb{N}_0$.
Let $\rho := (\rho_0, \rho_1, ...) \in \Omega$. Our construction tells us

$$\alpha \leq h_{m,m}(\rho_0, ..., \rho_m) = \mathbb{1}_{A'_m}(\rho_0, ..., \rho_m),$$

hence, $\rho \in A_m$ for all $m \in \mathbb{N}$. Thereby, it follows that

$$\bigcap_{i=0}^{\infty} A_i \neq \emptyset.$$

□

As mentioned in the beginning of this chapter, we want to show the Kolmogorov existence theorem for Borel spaces. The relevant definitions and lemmas about these Borel spaces are given here.

Definition 2.23. (Isomorphism)
Two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are **isomorphic**, if there exists a bijective mapping $\varphi : \Omega_1 \to \Omega_2$, such that φ is $\mathcal{F}_1$-$\mathcal{F}_2$-measurable and φ^{-1} is $\mathcal{F}_2$-$\mathcal{F}_1$-measurable. φ is called an isomorphism between measurable spaces. If, additionally, μ_1 and μ_2 are measures on $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, such that $\mu_2 = \mu_1 \circ \varphi^{-1}$, then φ is a measure space isomorphism and $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ are called isomorphic.

Definition 2.24. (Borel space)
A measurable space $(\Omega, \mathcal{F})$ is a **Borel space**, if there is a Borel set $B \in \mathcal{B}(\mathbb{R})$, such that $(\Omega, \mathcal{F})$ and $(B, \mathcal{B}(B))$ are isomorphic.
Analogously, a measurable probability space $(\Omega, \mathcal{F})$ is a Borel space if $(\Omega, \mathcal{F})$ and $(B, \mathcal{B}(B))$ are isomorphic, for some $B \in \mathcal{B}_{[0,1]}$.

Corollary 2.25.
The product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ of two Borel spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ is a Borel space.

Proof. Let $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ be Borel spaces. Then, there exist $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ and bijective $\varphi_1 : \Omega_1 \to B_1$, $\varphi_2 : \Omega_2 \to B_2$ such that φ_1 is $\mathcal{F}_1$-$\mathcal{B}(B_1)$-, φ_1^{-1} is $\mathcal{B}(B_1)$-$\mathcal{F}_1$-, φ_2 is $\mathcal{F}_2$-$\mathcal{B}(B_2)$- and φ_2^{-1} is $\mathcal{B}(B_2)$-$\mathcal{F}_2$-measurable.
Define $\varphi := (\varphi_1, \varphi_2)$. One immediately sees that $\varphi(\omega_1, \cdot) = (\varphi_1(\omega_1), \varphi_2(\cdot))$ is $\mathcal{F}_1 \otimes \mathcal{F}_2$-$\mathcal{B}(B_2)$-measurable and that $\varphi(\cdot, \omega_2) = (\varphi_1(\cdot), \varphi_2(\omega_1))$ is $\mathcal{F}_1 \otimes \mathcal{F}_2$-$\mathcal{B}(B_1)$-measurable. Therefore φ is $\mathcal{F}_1 \otimes \mathcal{F}_2$-$\mathcal{B}(B_1 \times B_2)$-measurable (see Appendix, Lemma A.1).
Analogously, one gets that φ^{-1} is $\mathcal{B}(B_1 \times B_2)$-$\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable.
Hence, $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is isomorphic to $(B_1 \times B_2, \mathcal{B}(B_1 \times B_2))$ and, thereby, a Borel space. □

Theorem 2.26. (Regular conditional distribution)
Let Y be a random variable from $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ to the Borel space $(\Omega_2, \mathcal{F}_2)$ and $\mathcal{F} \subset \mathcal{F}_1$ a sub-σ-algebra. Then there exists a regular version $\kappa_{Y,\mathcal{F}}$ of the conditional distribution $\mathbb{P}(\{Y \in \cdot\}|\mathcal{F})$.

Proof. Let $B \in \mathcal{B}(\mathbb{R})$ be the Borel set and $\varphi : \Omega_2 \to B$ be the measure space isomorphism corresponding to the Borel space $(\Omega_2, \mathcal{F}_2)$. With Theorem 2.20 we get the regular conditional distribution $\kappa_{Y',\mathcal{F}}$ of the real-valued random variable $Y' = \varphi \circ Y$. Setting $\kappa_{Y,\mathcal{F}}(\omega_1, A) = \kappa_{Y',\mathcal{F}}(\omega_1, \varphi(A))$, for all $A \in \mathcal{F}_2$, closes the proof. □

Now we are able, with the help of Ionescu-Tulcea's theorem, to prove the Kolmogorov existence theorem in the case that the index set T is countable.

Theorem 2.27.
Let T be a countable index set and let $(\Omega_t, \mathcal{F}_t)$, $t \in T$, be Borel spaces. Let $(\mathbb{P}_J)_{J \in \mathcal{E}(T)}$ be a projective family of probability measures. Then, there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ with the property

$$\mathbb{P}_J = \mathbb{P} \circ (\pi_J)^{-1},$$

for each finite $J \subset T$.

Proof. W.l.o.g let $T = \mathbb{N}_0$ and $\mathbb{P}_n := \mathbb{P}_{\{0,...,n\}}$. From Corollary 2.25 we know that the finite product of Borel spaces is a Borel space, i.e. $(\Omega_{\{0,...,n\}}, \mathcal{F}_{\{0,...,n\}})$ is a Borel space for all $n \in \mathbb{N}_0$.
Let $\widetilde{\mathcal{F}}_n := \{A \times \Omega_{n+1} | A \in \mathcal{F}_{\{0,...,n\}}\}$ and let the probability measure $\widetilde{\mathbb{P}}_n$ on $(\Omega_{\{0,...,n+1\}}, \widetilde{\mathcal{F}}_n)$ be defined by

$$\forall A \in \mathcal{F}_{\{0,...,n\}} : \; \widetilde{\mathbb{P}}_n(A \times \Omega_{n+1}) := \mathbb{P}_n(A)$$

The projectivity gives us that $\mathbb{P}_{n+1}|_{\widetilde{\mathcal{F}}_n} = \widetilde{\mathbb{P}}_n$. Since $\widetilde{\mathcal{F}}_n \subset \mathcal{F}_n$ is a sub-σ-algebra it follows, by Theorem 2.26, the theorem on the existence of regular conditional distributions, that a stochastic kernel κ'_{n+1} from $(\Omega_{\{0,...,n+1\}}, \widetilde{\mathcal{F}}_n)$ to $(\Omega_{n+1}, \mathcal{F}_{n+1})$ exists such that

$$\begin{aligned} &\mathbb{P}_{n+1}(A) \\ &= \int\int \mathbb{1}_A(\omega_0, ..., \omega_n, \widetilde{\omega}_{n+1}) \kappa'_{n+1}((\omega_0, ..., \omega_{n+1}), d\widetilde{\omega}_{n+1}) \widetilde{\mathbb{P}}_n(d(\omega_0, ..., \omega_{n+1})), \end{aligned}$$

for every $A \in \mathcal{F}_{\{0,...,n+1\}}$. Since $\kappa'_{n+1}(\;\cdot\;, A)$ is measurable w.r.t $\widetilde{\mathcal{F}}_n$, κ'_{n+1} does not depend on ω_{n+1}. Hence, by

$$\kappa_{n+1}((\omega_0, ..., \omega_n), \;\cdot\;) := \kappa'_{n+1}((\omega_0, ..., \omega_{n+1}), \;\cdot\;)$$

a stochastic kernel from $(\Omega_{\{0,...,n\}}, \mathcal{F}_{\{0,...,n\}})$ to $(\Omega_{n+1}, \mathcal{F}_{n+1})$ is defined with

$$\mathbb{P}_{n+1}(A) = \int\int \mathbb{1}_A(\omega_0, ..., \omega_{n+1}) \kappa_{n+1}((\omega_0, ..., \omega_n), d\omega_{n+1}) \mathbb{P}_n(d(\omega_0, ..., \omega_n)).$$

Consequently, $\mathbb{P}_{n+1} = \mathbb{P}_n \otimes \kappa_{n+1}$ holds and Ionescu-Tulcea's theorem can be applied to finish the proof. □

Finally, we can show the Kolmogorov existence theorem for arbitrary index sets. The proof will resort to the result we have just shown.

Theorem 2.28. (Kolmogorov existence theorem)
Let T be an arbitrary index set and let $(\Omega_t, \mathcal{F}_t)$, $t \in T$, be Borel spaces. Let $(\mathbb{P}_J)_{J \in \mathcal{E}(T)}$ be a projective family of probability measures. Then, there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, called **projective limit**, with the property

$$\mathbb{P} \circ (\pi_J)^{-1} = \mathbb{P}_J.$$

Proof. By the previous theorem we got that for countable $J \subset T$ there exists a unique measure $\mathbb{P}_J$ on $(\Omega_J, \mathcal{F}_J)$ with the property $\mathbb{P}_K = \mathbb{P}_J \circ (\pi^J_K)^{-1}$ for each $K \in \mathcal{E}(T)$. A measure on $(\Omega, \sigma(\pi_J))$ can be defined by

$$\widetilde{\mathbb{P}}_J((\pi_J)^{-1}(A_J)) := \mathbb{P}_J(A_J),$$

where $A_J \in \mathcal{F}_J$.
Let $J, J' \subset T$ be at most countable subsets and let $A \in \sigma(\pi_J) \cap \sigma(\pi_{J'}) \cap \mathcal{Z}$ be a $\sigma(\pi_J) \cap \sigma(\pi_{J'})$-measurable cylinder with finite Basis. Then, there exists a finite $K \subset J \cap J'$ and $A_K \in \mathcal{F}_K$ with $A = (\pi_K)^{-1}(A_K)$.
Hence, $\widetilde{\mathbb{P}}_J(A) = \mathbb{P}_K(A_K) = \widetilde{\mathbb{P}}_{J'}(A)$ for all $A \in \sigma(\pi_J) \cap \sigma(\pi_{J'})$. From Lemma 2.13 we know that $\widetilde{\mathcal{Z}}$, the set of all cylinder sets with countable basis, is a σ-algebra and furthermore that it holds that

$$\mathcal{F} = \bigotimes_{t \in T} \mathcal{F}_t = \bigcup_{\substack{J \subset T \\ J \ count.}} \mathcal{Z}_J = \widetilde{\mathcal{Z}},$$

which follows from the fact that $\mathcal{Z} \subseteq \widetilde{\mathcal{Z}} \subseteq \mathcal{F}$ and $\sigma(\mathcal{Z}) = \mathcal{F}$. Hence, there exists a countable $J \subset T$ with $A \in \sigma(\pi_J)$ for every $A \in \mathcal{F}$, and we can, uniquely and independently of the choice of J, define a function $\mathbb{P}$ on $\mathcal{F}$ by

$$\mathbb{P}(A) = \widetilde{\mathbb{P}}_J(A).$$

All that remains to show is that $\mathbb{P}$ is a probability measure. Obviously, $\mathbb{P}(\Omega) = 1$ holds. Let $A_1, A_2, ... \in \mathcal{F}$ pairwise disjoint sets with the property that $A := \bigcup_{n=1}^{\infty} A_n$, then there exist countable $J_n \subset T$ with $A_n \in \sigma(\pi_{J_n})$ for $n \in \mathbb{N}$. Set $J = \bigcup_{n \in \mathbb{N}} J_n$. Then, every $A_n \in \sigma(\pi_J)$ and, thereby, also $A \in \sigma(\pi_J)$, i.e.,

$$\mathbb{P}(A) = \widetilde{\mathbb{P}}_J(A) = \widetilde{\mathbb{P}}_J(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \widetilde{\mathbb{P}}_J(A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Therefore, $\mathbb{P}$ is a probability measure satisfying the theorem. □

Definition 2.29.
A topological space (Ω, τ) is **polish**, if it is countably generated and completely metrizable by a metric, i.e. if there exists a τ inducing metric d such that (Ω, d) is a complete, separable, metric space.

Definition 2.30.
A measurable space $(\Omega, \mathcal{F})$ is **polish**, if $\mathcal{F}$ is the Borel-σ-algebra w.r.t. a topology τ that makes (Ω, τ) a Polish space.

Theorem 2.31.
Every Polish space $(\Omega, \mathcal{F})$ is a Borel space.

Proof. Since the proof of this theorem does not lie within the scope of this work, we refer to [D1], Page 487 ff. □

Definition 2.32. (Consistency)
Let $(\Gamma, \leq)$ be a partially ordered set and $(\Omega_\gamma, \mathcal{F}_\gamma, \mathbb{P}_\gamma)$, $\gamma \in \Gamma$, be a family of probability spaces. For each pair $\gamma_1, \gamma_2 \in \Gamma$, $\gamma_1 \leq \gamma_2$, let $\phi_{\gamma_1}^{\gamma_2} : \Omega_{\gamma_2} \to \Omega_{\gamma_1}$ be a $\mathcal{F}_{\gamma_2}$-$\mathcal{F}_{\gamma_1}$-measurable mapping, such that

$$\forall \gamma_1, \gamma_2, \gamma_3 \in \Gamma, \gamma_1 \leq \gamma_2 \leq \gamma_3 : \; \phi_{\gamma_1}^{\gamma_2} \circ \phi_{\gamma_2}^{\gamma_3} = \phi_{\gamma_1}^{\gamma_3}.$$

The family of probability measures $(\mathbb{P}_\gamma)_{\gamma \in \Gamma}$ is **consistent**, if

$$\forall \phi_{\gamma_1}^{\gamma_2} \in \Gamma, \gamma_1 \leq \gamma_2 : \; \mathbb{P}_{\gamma_2} \circ (\phi_{\gamma_1}^{\gamma_2})^{-1} = \mathbb{P}_{\gamma_1}.$$

Definition 2.33.
Let the assumptions from Definition 2.32 hold and define the set Ω_Γ as follows:

$$\Omega_\Gamma := \{\omega \in \Omega \mid \forall \gamma_1 \leq \gamma_2 : \omega(\gamma_1) = \phi_{\gamma_1}^{\gamma_2}(\omega(\gamma_2))\}.$$

For each $\gamma \in \Gamma$ let $\pi_\gamma : \Omega_\Gamma \to \Omega_\gamma$ be the γ-th coordinate function of Ω_Γ, i.e., $\pi_\gamma(\omega) := \omega(\gamma)$, and analogously to the product-σ-algebra of Ω, define $\mathcal{B}_\Gamma$ as the smallest σ-algebra on Ω_Γ such that π_γ is $\mathcal{B}_\Gamma$-$\mathcal{B}_\gamma$-measurable for all $\gamma \in \Gamma$:

$$\mathcal{B}_\Gamma := \sigma(\pi_\gamma^{-1}(\mathcal{B}_\gamma), \gamma \in \Gamma).$$

Then, $(\Omega_\Gamma, \mathcal{B}_\Gamma)$ is the **projective limit of the spaces** $(\Omega_\gamma, \mathcal{B}_\gamma)$, $\gamma \in \Gamma$.

Corollary 2.34. Let $(\Gamma, \leq)$ be a partially ordered set und let $(\Omega_\gamma, \mathcal{B}_\gamma, \mathbb{P}_\gamma)$, $\gamma \in \Gamma$ a family of probability spaces, where Ω_γ is a Polish space and $\mathcal{B}_\gamma$ the corresponding σ-algebra. Let the probability measures $\mathbb{P}_\gamma$ be consistent. Then, there exists a unique probability measure $\mathbb{P}_\Gamma$ on $(\Omega_\Gamma, \mathcal{B}_\Gamma)$, such that

$$\forall \gamma \in \Gamma : \; \mathbb{P}_\Gamma \circ \pi_\gamma^{-1} = \mathbb{P}_\gamma.$$

Chapter 3

Characterization of Subspaces of L_p

Now we turn to an application of the Kolmogorov existence theorem in Banach space theory. We will prove the astonishing result that every real Banach space $(X, \|\cdot\|)$ is linear isometric to a subspace of $L_p(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, as long as $\|\cdot\|^p$, $1 \leq p \leq 2$, is of negative type. Concretely, the theorem reads like this:

Theorem 3.1.
Let $(X, \|\cdot\|)$ be a real Banach space and $1 \leq p \leq 2$.

(i) If $\|\cdot\|^p$ is a function of negative type on X, then it holds that for every β, with $0 < \beta < p$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that X is linear isometric to a subspace of $L_p(\Omega, \mathcal{F}, \mathbb{P})$.

(ii) If $(X, \|\cdot\|)$ is linear isometric to a subspace of $L_p(\Omega, \mathcal{F}, \mathbb{P})$, then $\|\cdot\|^p$ is a function of negative type on X.

We start our investigation with clearing up the basic properties of functions of negative type and with lemmas that we will need in order to prove Theorem 3.1.

Definition 3.2. (Function of Negative Type)
Let $T \neq \emptyset$ be a set. The map $\phi : T \times T \to \mathbb{R}$ is of **negative type**, if

(i) $\forall t \in T : \phi(t,t) = 0$,

(ii) $\forall s,t \in T : \phi(s,t) = \phi(t,s)$,

(iii) $\forall n \in \mathbb{N}\ \forall t_1, ..., t_n \in T\ \forall a_1, ..., a_n \in \mathbb{R}$, $\sum_{i=1}^{n} a_i = 0$: $\sum_{i,j=1}^{n} \phi(t_i, t_j) a_i a_j \leq 0$.

Lemma 3.3.
Let $T \neq \emptyset$ be a set and $\phi : T \times T \to \mathbb{R}$ be a non-negative function. Then, the following two statements are equivalent:

(i) ϕ is of negative type,

(ii) For each $\lambda > 0$ the function $e^{-\lambda\phi}$ is positive semidefinite and for all $t \in T$ it holds that $\phi(t,t) = 0$.

Lemma 3.4.
Let $\phi : T \times T \to \mathbb{R}$ be a non-negative function of negative type and μ be a measure on $\mathbb{R}_+$ such that

$$\int_0^\infty \min\Big\{1, \frac{1}{x}\Big\} d\mu(x) < \infty.$$

Then,

$$(t,s) \mapsto \int_0^\infty \frac{1}{x}(1 - e^{-x\phi(t,s)}) d\mu(x)$$

is of negative type. In particular ϕ^α is of negative type for all $0 < \alpha \leq 1$.

Lemma 3.5.
Let $0 \leq p \leq 2$. Then the function $t \mapsto |t|^p$ is of negative type.

For proofs to these three lemmas see [P].

Lemma 3.6.
The norms $\|\cdot\|_p$ of the spaces L_p, $1 \leq p \leq 2$, are of negative type.

Proof. Let $1 \leq p \leq 2$. It holds that

$$\begin{aligned}\sum_{i,j=1}^n \|x_i - x_j\|_p^p a_i a_j &= \sum_{i,j=1}^n \int_\Omega |x_i(\omega) - x_j(\omega)|^p d\mu(\omega) a_i a_j \\ &= \int_\Omega \sum_{i,j=1}^n |x_i(\omega) - x_j(\omega)|^p a_i a_j d\mu(\omega).\end{aligned}$$

Since Lemma 3.5 gives us that $t \mapsto |t|^p$ is of negative type and, therefore, that

$$\sum_{i,j=1}^n |x_i(\omega) - x_j(\omega)|^p a_i a_j \leq 0,$$

if $\sum_{i=1}^{n} a_i = 0$, we got that

$$\sum_{i,j=1}^{n} \|x_i - x_j\|_p^p a_i a_j \leq 0$$

and that $\|\cdot\|_p^p$ is of negative type. From Lemma 3.4, together with the choice of $\alpha := \frac{1}{p}$, it follows that $(\|\cdot\|_p^p)^\alpha = (\|\cdot\|_p^p)^{\frac{1}{p}} = \|\cdot\|_p^{p\frac{1}{p}} = \|\cdot\|_p$ is of negative type. $\square$

Theorem 3.7.
Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, $X : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a random variable with distribution $\mathbb{P}$, distribution function $F_\mathbb{P}$ and a λ-integrable Fourier transform $\phi_\mathbb{P}$, where λ denotes the Lebesgue measure. Then, it holds that

(i)

$$\forall a, b \in \mathbb{R},\ a < b: \ F_\mathbb{P}(b) - F_\mathbb{P}(a) \ = \int_{-\infty}^{\infty} \frac{e^{-iat} - e^{-ibt}}{2\pi i t} \phi_\mathbb{P}(t) dt.$$

(ii) X has a bound and continuous λ-density $f_\mathbb{P}$ such that for all $x \in \mathbb{R}$

$$f_\mathbb{P}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \phi_\mathbb{P}(t) dt.$$

(iii)

$$\forall t \in \mathbb{R}: \ \phi_\mathbb{P}(t) \ = \int_{-\infty}^{\infty} e^{ixt} f_\mathbb{P}(x) dx.$$

Proof. (i): Since $e^{-i\cdot} : \mathbb{R} \to \mathbb{C}$, $x \mapsto e^{-ix}$ is Lipschitz continuous, there exists a constant C such that for all $a, b, t \in \mathbb{R}$, $t \neq 0$, it holds that

$$\left| \frac{e^{-iat} - e^{-ibt}}{\pi i t} \right| \leq C|b - a|.$$

For $b \to a$ we get from Lévy's inversion formula (Appendix, Lemma A.36) that

$$\lim_{b \to a} \left(\mathbb{P}(a < X < b) + \frac{\mathbb{P}(X = a) + \mathbb{P}(X = b)}{2} \right)$$
$$= \lim_{b \to a} \mathbb{P}(a < X < b) + \lim_{b \to a} \frac{\mathbb{P}(X = a) + \mathbb{P}(X = b)}{2}$$

$$
\begin{aligned}
&= \lim_{b \to a} \mathbb{P}(a < X < b) + \mathbb{P}(X = a) \\
&\overset{Levy}{=} \lim_{b \to a} \lim_{c \to \infty} \int_{-c}^{c} \frac{e^{-iat} - e^{-ibt}}{2\pi i t} \phi_{\mathbb{P}}(t) dt.
\end{aligned}
$$

Since $\lim_{b \to a} \mathbb{P}(a < X < b) \geq 0$ we get for all $a \in \mathbb{R}$ that

$$
\begin{aligned}
\mathbb{P}(X = a) &\leq \lim_{b \to a} \lim_{c \to \infty} \int_{-c}^{c} \left| \frac{e^{-iat} - e^{-ibt}}{2\pi i t} \phi_{\mathbb{P}}(t) \right| dt \\
&< \lim_{b \to a} \lim_{c \to \infty} \int_{-c}^{c} \left| \frac{e^{-iat} - e^{-ibt}}{\pi i t} \right| |\phi_{\mathbb{P}}(t)| dt \\
&= \lim_{b \to a} \int_{-\infty}^{\infty} \left| \frac{e^{-iat} - e^{-ibt}}{\pi i t} \right| |\phi_{\mathbb{P}}(t)| dt \\
&\overset{Lipschitz}{\leq} \lim_{b \to a} C|b - a| \int_{-\infty}^{\infty} |\phi_{\mathbb{P}}(t)| dt = 0
\end{aligned}
$$

and, therefore, that $\mathbb{P}(X = a) = 0$.
Analogously, one gets $\mathbb{P}(X = b) = 0$ for all $b \in \mathbb{R}$ from using Lévy inversion formula for $a \to b$. Once again, with Lévy's inversion formula it follows for every $a < b$ that

$$
\begin{aligned}
\lim_{c \to \infty} \int_{-c}^{c} \frac{e^{-iat} - e^{-ibt}}{2\pi i t} \phi_{\mathbb{P}}(t) dt &\overset{Levy}{=} \mathbb{P}(a < X < b) + \frac{\mathbb{P}(X = a) + \mathbb{P}(X = b)}{2} \\
&= \mathbb{P}(a < X < b) = F_{\mathbb{P}}(b) - F_{\mathbb{P}}(a).
\end{aligned}
$$

(ii): Using (i) together with the dominated convergence theorem (see Appendix, Lemma A.6), we get that

$$
\begin{aligned}
F'_{\mathbb{P}}(a) &= \lim_{b \to a} \frac{F_{\mathbb{P}}(b) - F_{\mathbb{P}}(a)}{b - a} \\
&= \lim_{b \to a} \lim_{c \to \infty} \int_{-c}^{c} \frac{e^{-iat} - e^{-ibt}}{2\pi i t (b - a)} \phi_{\mathbb{P}}(t) dt \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi i t} \left(\lim_{b \to a} \frac{e^{-iat} - e^{-ibt}}{b - a} \right) \phi_{\mathbb{P}}(t) dt
\end{aligned}
$$

$$= \int_{-\infty}^{\infty} -\frac{1}{2\pi it}\left(\lim_{b\to a}\frac{e^{-ibt}-e^{-iat}}{b-a}\right)\phi_{\mathbb{P}}(t)dt$$

$$= \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} -\frac{1}{it}(-it)e^{-iat}\phi_{\mathbb{P}}(t)dt$$

$$= \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-iat}\phi_{\mathbb{P}}(t)dt =: \; f_{\mathbb{P}}(a), \;\forall a\in\mathbb{R}.$$

Thereby, we have shown the differentiability of $F_{\mathbb{P}}$ with $F'_{\mathbb{P}} = f_{\mathbb{P}}$ and that $f_{\mathbb{P}}$ is a λ-density of X.
Using a continuity lemma (see Appendix, Theorem A.7), one sees that the continuity of $f_{\mathbb{P}}$ follows. Namely, for $f(t,x) := f_{\mathbb{P}}(t)\mathbb{1}_{\mathbb{R}}(x) = \frac{1}{2\pi}e^{-iat}\phi_{\mathbb{P}}(t)\mathbb{1}_{\mathbb{R}}(x)$ it holds for every $t\in\mathbb{R}$ that, $t\mapsto f(t,x)$ is in $\mathcal{L}^1(\mu)$ for all $x\in\mathbb{R}$ and that $x\mapsto f(t,x)$ is continuous in t, for all $t\in\mathbb{R}$. Additionally $t\mapsto \sup_{x\in\mathbb{R}}|f(t,x)| = |f(t,x)|$ is in $\mathcal{L}^1(\mu)$ and the assumptions from Theorem A.7 are fulfilled, which then gives the continuity of $f_{\mathbb{P}}$, for each $t\in\mathbb{R}$.

The boundedness of $f_{\mathbb{P}}$, for each $x\in\mathbb{R}$, follows from

$$|f_{\mathbb{P}}(x)| = \left|\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-ixt}\phi_{\mathbb{P}}(t)dt\right| \le \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} |e^{-ixt}||\phi_{\mathbb{P}}(t)|dt \le \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} |\phi_{\mathbb{P}}(t)|dt < \infty,$$

since $\phi_{\mathbb{P}}\in L_1(\mathbb{R},\mathcal{B},\lambda)$.
(iii) easily follows from

$$\phi_{\mathbb{P}}(t) = \mathbb{E}e^{iXt} = \int\limits_{\mathbb{R}} e^{ixt}d\mathbb{P}(x) = \int\limits_{\mathbb{R}} e^{ixt}f_{\mathbb{P}}(x)dx.$$

□

Corollary 3.8.
Let $0<\alpha\le 2$. Then, there exists an integrable function $h_\alpha:\mathbb{R}\to\mathbb{R}$ with

$$\forall u\in\mathbb{R}: e^{-|u|^\alpha} = \int\limits_{-\infty}^{\infty} e^{ius}h_\alpha(s)ds.$$

Furthermore, it holds for every β with $0<\beta<\alpha$ that

$$\int\limits_{-\infty}^{\infty} h_\alpha(s)|s|^\beta ds < \infty.$$

Proof. The function $u \mapsto e^{-|u|^\alpha}$ is positive semidefinite. From Lemma 3.5 we get that $u \mapsto |u|^p$ is of negative type for $0 < p \leq 2$. The positive semidefiniteness of this function follows from Lemma 3.3.
With the Lemma of Bôchner (Appendix, Lemma A.35) we see that $u \mapsto e^{-|u|^\alpha}$ is the characteristic function of a Borel probability measure $\mathbb{P}_\alpha$ on $\mathbb{R}$.
The existence of a density h_α with $\mathbb{P}_\alpha = h_\alpha ds$, coming from Theorem 3.7, closes the proof. □

Corollary 3.9.
From the Lemma of Bôchner (Appendix Lemma A.35) it follows that there exists a Borel probability measure $\mathbb{P}_\alpha$ on $\mathbb{R}$ such that

$$e^{-|u|^\alpha} = \int_{-\infty}^{\infty} e^{ius} d\mathbb{P}_\alpha(s)$$

is the Fourier transform or characteristic function of $\mathbb{P}_\alpha$.

Lemma 3.10.
Let $0 < \beta < 2$. Then, for all $u \in \mathbb{R}$ it holds that

$$\int_0^\infty \frac{1-\cos(\lambda u)}{\lambda^{\beta+1}} d\lambda = |u|^\beta \int_0^\infty \frac{1-\cos(\lambda)}{\lambda^{\beta+1}} d\lambda.$$

Proof. For every $\lambda \in \mathbb{R}$ it holds that $1 - \cos(\lambda) \leq 2$ and $\cos(\lambda) \geq 1 - \frac{\lambda^2}{2}$. Hence, it holds that $1 - \cos(\lambda) \leq \frac{\lambda^2}{2}$. It follows that

$$\int_0^\infty \frac{1-\cos(\lambda u)}{\lambda^{\beta+1}} d\lambda = \int_0^1 \frac{1-\cos(\lambda u)}{\lambda^{\beta+1}} d\lambda + \int_1^\infty \frac{1-\cos(\lambda u)}{\lambda^{\beta+1}} d\lambda$$

$$\leq \int_0^1 \frac{\frac{\lambda^2}{2}}{\lambda^{\beta+1}} d\lambda + \int_1^\infty \frac{2}{\lambda^{\beta+1}} d\lambda = \frac{1}{2}\int_0^1 \lambda^{1-\beta} d\lambda + \int_1^\infty \frac{2}{\lambda^{\beta+1}} d\lambda \overset{0<\beta<2}{<} \infty$$

and, therefore, that the integral exists.
To show the equality we first note that the cosine is a symmetric function and it therefore suffices to show the equality for $|u|$. Using the substitution $y(\lambda) := |u|\lambda$ we get

$$\int_0^\infty \frac{1-\cos(\lambda|u|)}{\lambda^{\beta+1}} d\lambda = \int_0^\infty \frac{1-\cos(y)}{(\frac{y}{|u|})^{\beta+1}} \frac{1}{|u|} dy = |u|^\beta \int_0^\infty \frac{1-\cos(y)}{y^{\beta+1}} dy$$

□

The next lemma is the central lemma of this chapter since it proves (i) of Theorem 3.1. Two proves will be given of this, namely, a proof for the finite dimensional case that makes no use of the Kolmogorov existence theorem but is quite insightful, and another proof for the infinite dimensional case making use of the Kolmogorov existence theorem. Although the proof for the infinite dimensional case also holds in the finite dimensional case, these distinctions will allow us to see where the first proof breaks down and the second holds.

Lemma 3.11.
Let $\|\cdot\|$ be a norm on a real Banach space X such that for some α, with $0 < \alpha \leq 2$, $\|\cdot\|^\alpha$ is of negative type. Then, for each β, $0 < \beta < \alpha$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that X is linear isometric to a subspace of $L_\beta(\Omega, \mathcal{F}, \mathbb{P})$. (If $\alpha = 2$, X is isometric to a subspace of $L_\beta(\Omega, \mathcal{F}, \mathbb{P})$ for every $\beta > 0$).

Proof. (finite-dimensional case)
Let X be a finite dimensional Banach space such that $\|\cdot\|^\alpha$ is of negative type. From Lemma 3.3 we have that $\exp(-\|x\|^\alpha)$ is positive semidefinite. Let $x_1, ..., x_n$ be a basis of X. Then the function $g_{x_1,...,x_n} : \mathbb{R}^n \to \mathbb{R}$ with

$$g_{x_1,...,x_n}(t) = \exp\left(-\left\|\sum_{i=1}^n t_i x_i\right\|^\alpha\right)$$

is positive semidefinite. From the Lemma of Bôchner (Appendix, Lemma A.35) it follows that $g_{x_1,...,x_n}$ is the Fourier transform, or characteristic function respectively, of a measure $\mathbb{P}$ on $\mathbb{R}^n$, i.e.,

$$g_{x_1,...,x_n}(t) = \int_{\mathbb{R}^n} e^{i\langle t,y\rangle} d\mathbb{P}(y).$$

Hence, we got

$$\int_{\mathbb{R}^n} e^{i\lambda\langle t,y\rangle} d\mathbb{P}(y) = \exp\left(-\lambda^\alpha \left\|\sum_{i=1}^n t_i x_i\right\|^\alpha\right).$$

Since this is a pure real expression, we get from Euler's formula that

$$\int_{\mathbb{R}^n} \cos(\lambda\langle t,y\rangle) d\mathbb{P}(y) = \exp\left(-\lambda^\alpha \left\|\sum_{i=1}^n t_i x_i\right\|^\alpha\right).$$

From Corollary 3.8 we know that there exists a function $h_\alpha : \mathbb{R} \to \mathbb{R}$ such that for all $u \in \mathbb{R}$

$$e^{-|u|^\alpha} = \int_{-\infty}^{\infty} e^{ius} h_\alpha(s) ds \overset{Euler}{=} \int_{-\infty}^{\infty} \cos(us) h_\alpha(s) ds$$

and, hence,

$$\int_{\mathbb{R}^n} \cos(\lambda\langle t, y\rangle) d\mathbb{P}(y) = \int_{-\infty}^{\infty} \cos\left(\lambda s \left\|\sum_{i=1}^{n} t_i x_i\right\|\right) h_\alpha(s) ds.$$

We have $\int_{\mathbb{R}^n} d\mathbb{P} = 1$ and $1 = e^{-|0|^\alpha} = \int_{-\infty}^{\infty} e^{i\cdot 0\cdot s} h_\alpha(s) ds = \int_{-\infty}^{\infty} h_\alpha(s) ds$ and, thereby, that

$$\int_{\mathbb{R}^n} 1 - \cos(\lambda\langle t, y\rangle) d\mathbb{P}(y) = \int_{-\infty}^{\infty} 1 - \cos\left(\lambda s \left\|\sum_{i=1}^{n} t_i x_i\right\|\right) h_\alpha(s) ds.$$

As next step we divide both sides by $\lambda^{\beta+1}$ and integrate w.r.t. λ to get

$$\int_0^{\infty} \int_{\mathbb{R}^n} \frac{1 - \cos(\lambda\langle t, y\rangle)}{\lambda^{\beta+1}} d\mathbb{P}(y) d\lambda = \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1 - \cos\left(\lambda s \left\|\sum_{i=1}^{n} t_i x_i\right\|\right)}{\lambda^{\beta+1}} h_\alpha(s) ds d\lambda.$$

Since $f : [0, \infty] \times \mathbb{R}^n \to \bar{\mathbb{R}}$, $(\lambda, y) \mapsto \frac{1-\cos(\lambda\langle t, y\rangle)}{\lambda^{\beta+1}}$ is continuous up to the Null set $\{0\} \times \mathbb{R}$ and non-negative, Tonelli's Theorem (Appendix, Theorem A.37) yields

$$\int_{\mathbb{R}^n} \int_0^{\infty} \frac{1 - \cos(\lambda\langle t, y\rangle)}{\lambda^{\beta+1}} d\lambda d\mathbb{P}(y) = \int_{-\infty}^{\infty} \int_0^{\infty} \frac{1 - \cos\left(\lambda s \left\|\sum_{i=1}^{n} t_i x_i\right\|\right)}{\lambda^{\beta+1}} h_\alpha(s) d\lambda ds.$$

Now we can use the identity from Lemma 3.10 to get

$$\begin{aligned}
\int_{\mathbb{R}^n} |\langle t, y\rangle|^\beta \int_0^{\infty} \frac{1 - \cos(\lambda)}{\lambda^{\beta+1}} d\lambda d\mathbb{P}(y) &= \int_{\mathbb{R}^n} \int_0^{\infty} \frac{1 - \cos(\lambda\langle t, y\rangle)}{\lambda^{\beta+1}} d\lambda d\mathbb{P}(y) \\
&= \int_{-\infty}^{\infty} \int_0^{\infty} \frac{1 - \cos\left(\lambda s \left\|\sum_{i=1}^{n} t_i x_i\right\|\right)}{\lambda^{\beta+1}} h_\alpha(s) d\lambda ds \\
&= \int_{-\infty}^{\infty} |s|^\beta \left\|\sum_{i=1}^{n} t_i x_i\right\|^\beta \int_0^{\infty} \frac{1 - \cos(\lambda)}{\lambda^{\beta+1}} h_\alpha(s) d\lambda ds.
\end{aligned}$$

Hence, we got

$$\int_{\mathbb{R}^n} |\langle t, y\rangle|^\beta d\mathbb{P}(y) = \left\| \sum_{i=1}^n t_i x_i \right\|^\beta \int_{-\infty}^{\infty} |s|^\beta h_\alpha(s) ds.$$

The finiteness of these expressions follows from Corollary 3.8.

For the map $\Phi : X \to L_\beta(\mathbb{R}^n, \mathbb{P})$ with $\Phi(\sum_{i=1}^n t_i x_i) := \frac{1}{c}\sum_{i=1}^n t_i \phi_i$, where $c := (\int_{-\infty}^{\infty} |s|^\beta h_\alpha(s) ds)^{\frac{1}{\beta}}$ and $\phi_i(w) := w_i$, it holds that

$$\begin{aligned}
\left\| \sum_{i=1}^n t_i x_i \right\|^\beta &= \frac{1}{c^\beta} \int_{\mathbb{R}^n} |\langle t, y\rangle|^\beta d\mathbb{P}(y) \\
&= \frac{1}{c^\beta} \int_{\mathbb{R}^n} \left| \sum_{i=1}^n t_i y_i \right|^\beta d\mathbb{P}(y) \\
&= \int_{\mathbb{R}^n} \left| \frac{1}{c} \sum_{i=1}^n t_i \phi_i(y) \right|^\beta d\mathbb{P}(y) \\
&= \int_{\mathbb{R}^n} \left| \Phi(\sum_{i=1}^n t_i x_i)(y) \right|^\beta d\mathbb{P}(y) \\
&= \left\| \Phi(\sum_{i=1}^n t_i x_i) \right\|_\beta^\beta
\end{aligned}$$

Hence, Φ is an isometry.
We show the linearity. Let $v, w \in X$. Then, there exist $t_i, s_i \in \mathbb{R}$, $i = 1, .., n$ such that $v = \sum_{i=1}^n t_i x_i$ and $w = \sum_{i=1}^n s_i x_i$. It follows that

$$\begin{aligned}
\Phi(v + w) &= \Phi(\sum_{i=1}^n t_i x_i + \sum_{i=1}^n s_i x_i) = \Phi(\sum_{i=1}^n (t_i + s_i) x_i) = \sum_{i=1}^n (t_i + s_i) \phi_i \\
&= \sum_{i=1}^n t_i \phi_i + \sum_{i=1}^n s_i \phi_i = \Phi(\sum_{i=1}^n t_i x_i) + \Phi(\sum_{i=1}^n s_i x_i) = \Phi(v) + \Phi(w).
\end{aligned}$$

Furthermore, it holds for $\lambda \in \mathbb{R}$ that

$$\Phi(\lambda v) = \Phi(\lambda \sum_{i=1}^n t_i x_i) = \Phi(\sum_{i=1}^n \lambda t_i x_i) = \sum_{i=1}^n \lambda t_i \phi_i = \lambda \sum_{i=1}^n t_i \phi_i = \lambda \Phi(v).$$

Therefore Φ is a linear isometry. □

Proof. (infinite dimensional case)
Let X be a real Banach space such that $\|\cdot\|^\alpha$ is of negative type. From Lemma 3.3 it follows that $\exp(-\|\cdot\|^\alpha)$ is positive semidefinite. Let $x_1, ..., x_n \in X$. Then the function $g_{x_1,...,x_n} : \mathbb{R}^n \to \mathbb{R}$ with

$$g_{x_1,...,x_n}(t) = \exp\left(-\left\|\sum_{i=1}^n t_i x_i\right\|^\alpha\right)$$

is positive semidefinite and it holds that $g_{x_1,...,x_n}(0) = 1$. From the Lemma of Bôchner (Appendix, Lemma A.35) it follows that $g_{x_1,...,x_n}$ is the Fourier transform, or characteristic function respectively, of a measure $\mathbb{P}_{\{x_1,...,x_n\}}$ on $\mathbb{R}^n$, i.e.,

$$g_{x_1,...,x_n}(t) = \int\limits_{\mathbb{R}^n} e^{i\langle t,y\rangle} d\mathbb{P}_{\{x_1,...,x_n\}}(y).$$

Hence, we obtain

$$\int\limits_{\mathbb{R}^n} e^{i\lambda\langle t,y\rangle} d\mathbb{P}_{\{x_1,...,x_n\}}(y) = \exp\left(-\lambda^\alpha\left\|\sum_{i=1}^n t_i x_i\right\|^\alpha\right).$$

Thus, we have a family of probability spaces $(\Omega_{\{x_1,...,x_n\}}, \mathcal{B}_{\{x_1,...,x_n\}}, \mathbb{P}_{\{x_1,...,x_n\}})$, where actually $\Omega_{\{x_1,...,x_n\}} = \mathbb{R}^n$ holds.

Consider a semi ordering on the index set $\Gamma := \{(x_1, ..., x_n) | \; n \in \mathbb{N}, \; x_i \in X\}$ defined by $\forall n, k \in \mathbb{N} : (x_1, ..., x_n) \leq (y_1, ..., y_k) \iff \{x_1, ..., x_n\} \subseteq \{y_1, ..., y_k\}$.

For $k \geq n$, the maps $\phi^{\{x_1,...,x_k\}}_{\{x_1,...,x_n\}} : \Omega_{\{x_1,...,x_k\}} \to \Omega_{\{x_1,...,x_n\}}, \; (t_1, ..., t_k) \mapsto (t_1, ..., t_n)$, are the projections onto the first n coordinates. If the coordinates appear in a different order we get a corresponding projection.
We show now that the measures $\mathbb{P}_{\{x_1,...,x_n\}}$ are consistent (Definition 2.32), i.e.,

$$\mathbb{P}_{\{x_1,...,x_k\}} \circ \left(\phi^{\{x_1,...,x_k\}}_{\{x_1,...,x_n\}}\right)^{-1} = \mathbb{P}_{\{x_1,...,x_n\}}.$$

The theorem on the uniqueness of the Fourier transform (Appendix, Theorem A.34) tells us that it suffices to show the following.
We know that for all $t = (t_1, ..., t_n) \in \mathbb{R}^n$ it holds that

$$\int\limits_{\mathbb{R}^n} e^{i\lambda\langle t,y\rangle} d\mathbb{P}_{\{x_1,...,x_n\}}(y) = \exp\left(-\lambda^\alpha\left\|\sum_{i=1}^n t_i x_i\right\|^\alpha\right). \tag{3.1}$$

On the other hand we get from the transformation formula (Appendix, Theorem A.18) that

$$\begin{aligned}
&\int\limits_{\mathbb{R}^n} e^{i\lambda\langle t,y\rangle} d\mathbb{P}_{\{x_1,\ldots,x_k\}} \circ \left(\phi^{\{x_1,\ldots,x_k\}}_{\{x_1,\ldots,x_n\}}\right)^{-1}(y) \\
&= \int\limits_{\mathbb{R}^k} e^{i\lambda\langle t,\cdot\rangle} \circ \phi^{\{x_1,\ldots,x_k\}}_{\{x_1,\ldots,x_n\}}(z) d\mathbb{P}_{\{x_1,\ldots,x_k\}}(z) \\
&= \int\limits_{\mathbb{R}^k} \exp(i\lambda \sum_{\ell=1}^{n} t_\ell \phi^{\{x_1,\ldots,x_k\}}_{\{x_1,\ldots,x_n\}}(z)(\ell)) d\mathbb{P}_{\{x_1,\ldots,x_k\}}(z) \\
&= \int\limits_{\mathbb{R}^k} e^{i\lambda \sum_{\ell=1}^{n} t_\ell z_\ell} d\mathbb{P}_{\{x_1,\ldots,x_k\}}(z) \\
&= \int\limits_{\mathbb{R}^k} e^{i\lambda\langle t,z\rangle} d\mathbb{P}_{\{x_1,\ldots,x_k\}}(z) \\
&\overset{(4.1)}{=} \exp\left(-\lambda^\alpha \left\|\sum_{i=1}^{k} t_i x_i\right\|^\alpha\right) \\
&= \exp\left(-\lambda^\alpha \left\|\sum_{i=1}^{n} t_i x_i\right\|^\alpha\right),
\end{aligned}$$

where the last equation holds because of $t_{n+1} = \ldots = t_k = 0$.
Therefore, we have that $\mathbb{P}_{\{x_1,\ldots,x_k\}} \circ \left(\phi^{\{x_1,\ldots,x_k\}}_{\{x_1,\ldots,x_n\}}\right)^{-1}$ and $\mathbb{P}_{\{x_1,\ldots,x_n\}}$ have the same Fourier transform and, hence, that the consistency of the measures $\mathbb{P}_{\{x_1,\ldots,x_n\}}$ follows from the uniqueness theorem (Appendix, Theorem A.34).
It follows now, with Corollary 2.34, as a direct consequence of the Kolmogorov existence theorem, that there exists a measure space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\Omega = \bigotimes_{i=1}^{\infty} \Omega_{\{x_1,\ldots,x_i\}}$ and projections $\pi_{\{x_1,\ldots,x_n\}} : \Omega \to \mathbb{R}^n$ for which the holds that

$$\mathbb{P} \circ (\pi_{\{x_1,\ldots,x_n\}})^{-1} = \mathbb{P}_{\{x_1,\ldots,x_n\}}.$$

In particular, there exists for each $x \in X$ a measure $\mathbb{P}_x$ on $\mathbb{R}$ with

$$\begin{aligned}
e^{-\|tx\|^\alpha} &= \int\limits_{\mathbb{R}^n} e^{its} d\mathbb{P}_x(s) = \int\limits_{\mathbb{R}} e^{its} d\mathbb{P} \circ (\pi_x)^{-1}(s) = \int\limits_{\Omega} e^{it\pi_x(\omega)} d\mathbb{P}(\omega) \\
&\overset{Euler}{=} \int\limits_{\Omega} \cos(it\pi_x(\omega)) d\mathbb{P}(\omega).
\end{aligned}$$

Now we define the embedding $I : X \to L_\beta(\Omega, \mathbb{P})$ by $I(x) := \frac{1}{c}\pi_x$, where $c := (\int_{-\infty}^{\infty} |s|^\beta h_\alpha(s)ds)^{\frac{1}{\beta}}$. We show that I is a well-defined linear isometry.

We have $e^{-\|tx\|^\alpha} = \int_\Omega \cos(it\pi_x(\omega))d\mathbb{P}(\omega)$ and $e^{-|\lambda u|^\alpha} = \int_{-\infty}^{\infty} \cos(\lambda us)h_\alpha(s)ds$, and, thereby, for $\|x\| = |u|$ that

$$\int_\Omega \cos(it\pi_x(\omega))d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} \cos(\|\lambda x\|s)h_\alpha(s)ds$$

holds. From $\int_{\mathbb{R}^n} d\mathbb{P} = 1$ and $1 = \int_{-\infty}^{\infty} h_\alpha(s)ds$ one sees that

$$\int_\Omega 1 - \cos(it\pi_x(\omega))d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} 1 - \cos(\lambda\|x\|s)h_\alpha(s)ds.$$

Dividing by $\lambda^{\beta+1}$ and integrating w.r.t λ gives that

$$\int_0^\infty \int_\Omega \frac{1 - \cos(it\pi_x(\omega))}{\lambda^{\beta+1}}d\mathbb{P}(\omega)d\lambda = \int_0^\infty \int_{-\infty}^{\infty} \frac{1 - \cos(\lambda\|x\|s)}{\lambda^{\beta+1}}h_\alpha(s)dsd\lambda.$$

Analogously, as in the proof for the finite dimensional case we get by Tonelli's theorem (Appendix, Theorem A.37) that

$$\int_\Omega \int_0^\infty \frac{1 - \cos(it\pi_x(\omega))}{\lambda^{\beta+1}}d\lambda d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} \int_0^\infty \frac{1 - \cos(\lambda\|x\|s)}{\lambda^{\beta+1}}h_\alpha(s)d\lambda ds$$

and furthermore by Lemma 3.10 that

$$\int_\Omega |\pi_x(\omega)|^\beta \int_0^\infty \frac{1 - \cos(\lambda)}{\lambda^{\beta+1}}d\lambda d\mathbb{P}(\omega) = \int_\Omega \int_0^\infty \frac{1 - \cos(it\pi_x(\omega))}{\lambda^{\beta+1}}d\lambda d\mathbb{P}(\omega)$$

$$\int_{-\infty}^{\infty} \int_0^\infty \frac{1 - \cos(\lambda\|x\|s)}{\lambda^{\beta+1}}h_\alpha(s)d\lambda ds = \int_{-\infty}^{\infty} |s|^\beta \|x\|^\beta \int_0^\infty \frac{1 - \cos(\lambda)}{\lambda^{\beta+1}}h_\alpha(s)d\lambda ds.$$

Hence,

$$\int_\Omega |\pi_x(\omega)|^\beta d\mathbb{P}(\omega) = \|x\|^\beta \int_{-\infty}^{\infty} |s|^\beta h_\alpha(s)ds$$

holds and, in particular, we get that

$$\left(\int_{\Omega} |\pi_x(\omega)|^\beta d\mathbb{P}(\omega)\right)^{\frac{1}{\beta}} = \|x\| \left(\int_{-\infty}^{\infty} |s|^\beta h_\alpha(s)ds\right)^{\frac{1}{\beta}},$$

which is, following from Corollary 3.8, finite and therefore well-defined.

From this result and the fact that $\|\pi_x\|_\beta = \int_\Omega |\pi_x(\omega)|^\beta d\mathbb{P}(\omega)$ it follows immediately that I is an isometry, since

$$\|I(x)\|_\beta^\beta = \left\|\frac{1}{c}\pi_x\right\|_\beta^\beta = \frac{1}{c^\beta}\|\pi_x\|_\beta^\beta = \frac{1}{c^\beta}\|x\|^\beta \int_{-\infty}^{\infty} |s|^\beta h_\alpha(s)ds = \|x\|^\beta.$$

To show the linearity we consider that

$$\int_{\mathbb{R}^k} \exp(i\lambda \sum_{\ell=1}^n t_\ell \pi_{x_\ell}(\omega))d\mathbb{P}(\omega) = exp(-\lambda^\alpha \| \sum_{i=1}^n t_i x_i\|^\alpha) \tag{3.2}$$

and

$$\int_{\mathbb{R}^n} e^{i\lambda\langle t,y\rangle} d\mathbb{P}_{\{x_1,\dots,x_n\}}(y) = exp(-\lambda^\alpha \| \sum_{i=1}^n t_i x_i\|^\alpha)$$

holds. Thereby, it follows, together with the consistency of $\mathbb{P}_{\{x_1,\dots,x_n\}}$, that

$$\begin{aligned} exp(-\lambda^\alpha \| \sum_{i=1}^n t_i x_i\|^\alpha) &= \int_{\mathbb{R}^n} e^{i\lambda\langle t,y\rangle} d\mathbb{P}\circ(\pi_{\{x_1,\dots,x_n\}})^{-1}(y) \\ &= \int_\Omega e^{i\lambda\langle t,\pi_{\{x_1,\dots,x_n\}}(\omega)\rangle} d\mathbb{P}(\omega) \\ &= \int_\Omega \exp(i\lambda \sum_{\ell=1}^n t_\ell(\pi_{\{x_1,\dots,x_n\}}(\omega))(\ell))d\mathbb{P}(\omega). \end{aligned}$$

It remains to observe that $(\pi_{\{x_1,\dots,x_n\}}(\omega))(\ell) = \pi_{x_l}(\omega)$ holds, which follows form

$$\forall \gamma_1 \le \gamma_2 : \omega(\gamma_1) = \phi_{\gamma_1}^{\gamma_2}(\omega(\gamma_2)), \tag{3.3}$$

and the definition of π by $\pi_\gamma(\omega) = \omega(\gamma)$, since we have that

$$\pi_{x_l}(\omega) \overset{(4.3)}{=} \omega(x_l) = \phi_{x_l}^{\{x_1,\dots,x_n\}}(\omega(x_1,\dots,x_n))$$

$$= \phi_{x_l}^{\{x_1,\dots,x_n\}}(\pi_{\{x_1,\dots,x_n\}}(\omega)) = (\pi_{\{x_1,\dots,x_n\}}(\omega))(\ell).$$

First, we show $I(tx) = tI(x)$. For all $\lambda \in \mathbb{R}$ it holds that

$$1 = e^{-\|\lambda(tx-tx)\|^\alpha} = \int\limits_\Omega e^{i\lambda(t\pi_x(\omega)-\pi_{tx}(\omega))} d\mathbb{P}(\omega) = \int\limits_\Omega \cos(\lambda(t\pi_x(\omega)-\pi_{tx}(\omega)))d\mathbb{P}(\omega)$$

and, in particular, for $\lambda = 1$ that

$$1 = \int_\Omega \cos(t\pi_x(\omega) - \pi_{tx}(\omega))d\mathbb{P}(\omega).$$

Hence, for $\mathbb{P}$-almost all $\omega \in \Omega$, it has to hold that

$$t\pi_x(\omega) - \pi_{tx}(\omega) = 2\pi k(\omega), \ k(\omega) \in \mathbb{Z} \setminus \{0\},$$

and, thereby, for $\lambda = \frac{1}{2\pi}$ that

$$\begin{aligned} 1 &= \int\limits_\Omega \cos(\frac{1}{2\pi}(t\pi_x(\omega) - \pi_{tx}(\omega)))d\mathbb{P}(\omega) \\ &= \int\limits_\Omega \cos(\frac{1}{2\pi}2\pi k(\omega))d\mathbb{P}(\omega) = \int\limits_\Omega \cos(k(\omega))d\mathbb{P}(\omega), \end{aligned}$$

and, consequently, that $k(\omega)$ has to be $\mathbb{P}$-almost everywhere equal to 0, since else it would follow from $k(\omega) \in \mathbb{Z}$ that $\pi \in \mathbb{Q}$, which would be a contradiction. Hence, we have that $t\pi_x(\omega) = \pi_{tx}(\omega)$

Finally, we show that $I(x + y) = I(x) + I(y)$. Following the proof of $I(tx) = tI(x)$, we have for all $\lambda \in \mathbb{R}$ that

$$\begin{aligned} 1 &= e^{-\|\lambda(x+y-x-y)\|^\alpha} \\ &= \int\limits_\Omega e^{i\lambda(\pi_{x+y}(\omega)-\pi_x(\omega)-\pi_y(\omega))} d\mathbb{P}(\omega) \\ &= \int\limits_\Omega \cos(\lambda(\pi_{x+y}(\omega) - \pi_x(\omega) - \pi_y(\omega)))d\mathbb{P}(\omega) \end{aligned}$$

holds and particularly for $\lambda = 1$ that

$$1 = \int\limits_\Omega \cos(\pi_{x+y}(\omega) - \pi_x(\omega) - \pi_y(\omega))d\mathbb{P}(\omega).$$

Just as before it has to hold, for $\mathbb{P}$-almost all $\omega \in \Omega$, that

$$\pi_{x+y}(\omega) - \pi_x(\omega) - \pi_y(\omega) = 2\pi k(\omega),\ \ k(\omega) \in \mathbb{Z} \setminus \{0\},$$

and, analogously, for $\lambda = \frac{1}{2\pi}$ that

$$\begin{aligned} 1 &= \int_\Omega \cos(\frac{1}{2\pi}(\pi_{x+y}(\omega) - \pi_x(\omega) - \pi_y(\omega)))d\mathbb{P}(\omega) \\ &= \int_\Omega \cos(\frac{1}{2\pi}2\pi k(\omega))d\mathbb{P}(\omega) = \int_\Omega \cos(k(\omega))d\mathbb{P}(\omega), \end{aligned}$$

With the same argument as before we see that $k(\omega) = 0$ $\mathbb{P}$-almost everywhere and that $\pi_{x+y}(\omega) = \pi_x(\omega) + \pi_y(\omega)$ holds. □

We close this chapter with the short proof of Theorem 3.1 by putting together what we have so far.

Proof. (For Theorem 3.1)
(i) This is exactly the statement of Lemma 3.11.

(ii) Since X is isometric to a subspace of $L_p(\Omega, \mathcal{F}, \mathbb{P})$, $1 \leq p \leq 2$, it holds that $\| \cdot \|^p = \| \cdot \|_{L_p}$ and, therefore, by Lemma 3.6 that $\| \cdot \|^p$ is of negative type on X. □

Chapter 4

Majorizing Measures Theorem

The following chapter is dedicated to Michel Talagrand's and Xavier Fernique's majorizing measures theorem, a deep and powerful result in the theory of stochastic processes that provides bounds for the expectation of the supremum of a stochastic process with subgaussian tails.

We start by introducing out setting: we consider a collection of mean-zero random variables $X := (X_t)_{t\in T}$, i.e., $\mathbb{E}X_t = 0$, for all $t \in T$, where the index set T is equipped with a distance d such that (T, d) becomes a metric space. Moreover, X is assumed to have subgaussian tails, i.e.,

$$\forall s, t \in T \ \forall u > 0 : \ \mathbb{P}\left(|X_t - X_s| > u\right) \leq 2 \exp\left(-\frac{u^2}{d^2(t,s)}\right).$$

Consider for a moment the special case that X is a Gaussian process, i.e., a process where each finite-dimensional distribution is a multivariate normal distribution. Then, this process is fully characterized by the family of covariances, i.e., $\left(\mathbb{E}\left(X_s X_t\right)\right)_{s,t\in T}$.
By considering the L_2-metric, the natural distance that comes to mind is the following:

$$d(s,t) = \left(\mathbb{E}\,|X_s - X_t|^2\right)^{1/2}.$$

Note that this is only a pseudometric [L], i.e., $d(s,t) = 0$ does in general not imply that $X_s - X_t = 0$. Thus, we are able to completely characterize the process X by its distances $\left(d(s,t)\right)_{s,t\in T}$, up to translation by a Gaussian random variable X_{t_0}, $t_0 \in T$, since the process $\left(X_t - X_{t_0}\right)_{t\in T}$ will yield the same distances as X.
If we want to have a clear picture in mind of what is going on, we will turn to this case.

By $B(t, \varepsilon)$ we denote the closed ball centered at $t \in T$ with radius $\varepsilon > 0$ in the metric d and by $\operatorname{diam}(A)$ the diameter of the set $A \subseteq T$ w.r.t d, i.e., $\operatorname{diam}(A) = \sup\{d(x, y) : \ x, y \in A\}$. By C we refer to a positive absolute constant, which may change from line to line.

Our first goal will be to find upper bounds for $\mathbb{E} \sup_{t \in T} X_t$. Note that we use the definition $\mathbb{E} \sup_{t \in T} X_t := \sup\{\mathbb{E} \sup_{t \in F} X_t : F \subset T, F \text{ finite}\}$ to avoid problems with the supremum over possibly uncountably many random variables. The picture we should have in mind (the figure shown below is taken from [L]), is the following: for a finite $F \subset \mathbb{R}^n$, $n \in \mathbb{N}$, and a standard Gaussian g on $\mathbb{R}^n$, we can write $X_t = \langle g, t \rangle$ and think of the process X as the projections of F onto a uniformly random direction. This is possible, because the orthogonal projection onto a direction $g \in \mathbb{R}^n$ can be written as $\langle g, t \rangle \frac{g}{\|g\|_2}$ and, thus, $\langle g, t \rangle$ determines the position of process on the line $\mathbb{R} \cdot g \subseteq \mathbb{R}^n$. The picture we should have in mind is thus the following:

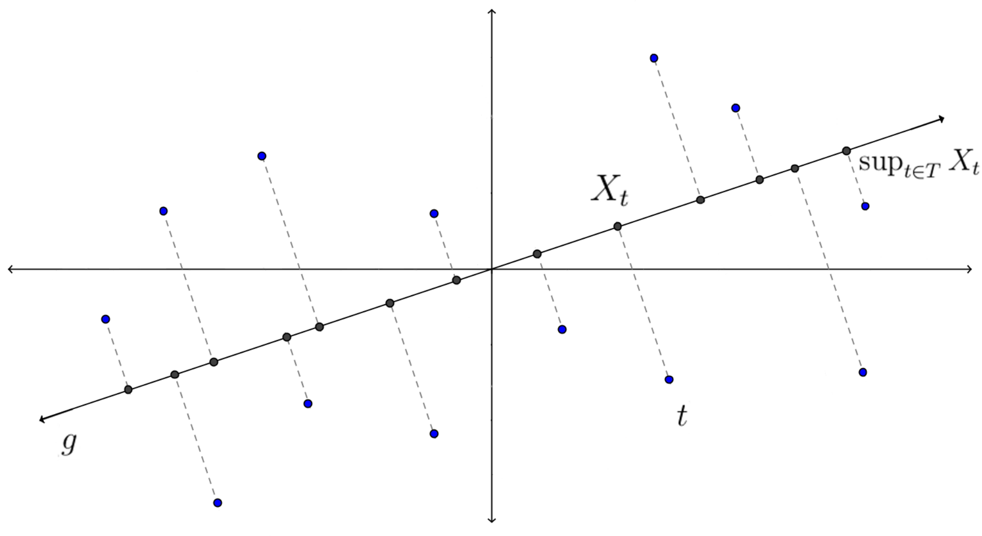

Figure 4.1: Points in $F \subset \mathbb{R}^2$ projected onto a random direction.

Talagrand's majorizing measures theorem provides such upper bounds for $\mathbb{E} \sup_{t \in T} X_t$. Its statement is the following and can be found in [T1], in particular, Proposition 2.3.:

Theorem 4.1. (Majorizing measures theorem)
Let $X = (X_t)_{t \in T}$ be a stochastic process as above and let μ be a probability measure on T. Let $r \geq 2$ and $i \in \mathbb{N}$ the largest natural number such that $\operatorname{diam}(T) \leq 2r^{-i}$. Let $(\mathcal{A}_j)_{j \geq i}$ be an increasing sequence of partitions of T so

that $\mathcal{A}_i = T$ and $\operatorname{diam}(A_j) \leq 2r^{-j}$ for each $j \geq i$, $A \in \mathcal{A}_j$. Then, there exists a constant $C > 0$ only depending on r such that

$$\mathbb{E} \sup_{t \in T} X_t \leq C(r) \sup_{t \in T} \sum_{j > i} r^{-j} \sqrt{\log \frac{1}{\mu(A_j(t))}},$$

where $A_j(t)$ denotes the unique element of $\mathcal{A}_j$ containing t.

Remark. Note that the quality of this estimate depends on the probability measure μ on T. A probability measure μ that provides a "decent" bound is referred to as a *majorizing measure*.

Remark. In [T1], Talagrand shows the following estimate, which is the most elaborate form of the majorizing measures theorem: there exists a universal constant $K > 0$ such that

$$\frac{1}{K} \gamma(T, d) \leq \mathbb{E} \sup_{t \in T} X_t \leq K \gamma(T, d),$$

where

$$\gamma(T, d) := \inf\{\sup_{t \in T} \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t, \varepsilon))}} \, d\varepsilon \ : \ \mu \text{ prob. measure on } T\}.$$

The upper bound can be shown by a combination of Lemma 4.2 and Lemma 4.5 below. The lower bound is derived by the Gaussian isoperimetric inequality and the famous Sudakov minoration. For details we refer to [T1].

In the second part of this chapter, we will turn to the question of how to construct those sequences of partitions. For this purpose we will use a partition scheme going back to Talagrand.

We start our investigation with Talagrands proof of Theorem 4.1, which uses the so-called "generic chaining", an argument that already appears in the work of Kolmogorov.

Proof. (Majorizing measures theorem)
As a first step, we introduce the reformulation $\mathbb{E} \sup_{t \in T} X_t = \mathbb{E} \sup_{t \in T} (X_t - X_{t_0})$, which is possible since we are dealing with mean-zero random variables. The random variable $Y := \sup_{t \in F} (X_t - X_{t_0})$, $F \subset T$ finite, is now non-negative, so we can write

$$\mathbb{E} Y = \int_0^\infty \mathbb{P}(Y \geq u) du.$$

Hence, we are interested in bounding $\mathbb{P}(\sup_{t\in T}(X_t - X_{t_0}) \geq u)$.
The natural estimate that comes to mind is the union bound,

$$\mathbb{P}(\sup_{t\in F}(X_t - X_{t_0}) \geq u) \leq \sum_{t\in F} \mathbb{P}(X_t - X_{t_0} \geq u). \tag{4.1}$$

If there only a few uncorrelated random variables this bound is quite good, it is very bad for correlated ones (and especially for a big number of them), since then, for each random variable a term of roughly the same size will appear and the bound is vastly overestimating the supremum.
This is where the idea of the generic chaining comes into play by regrouping the random variables X_t which are nearly identical. We illustrate the idea with T consisting of two finite subsets. At the first level, we consider the finite subsets $T_1, T_2 \subset T$. The setting is shown in Figure 4.2, which was also taken from [L]. In each of the two subsets T_1 and T_2 some Gaussians are clustered together and therefore highly correlated. Thus, projecting them onto a random direction will project them closely together. We choose now representatives $p_1(t) \in T_1$ and $p_2(t) \in T_2$ for $t \in T$.

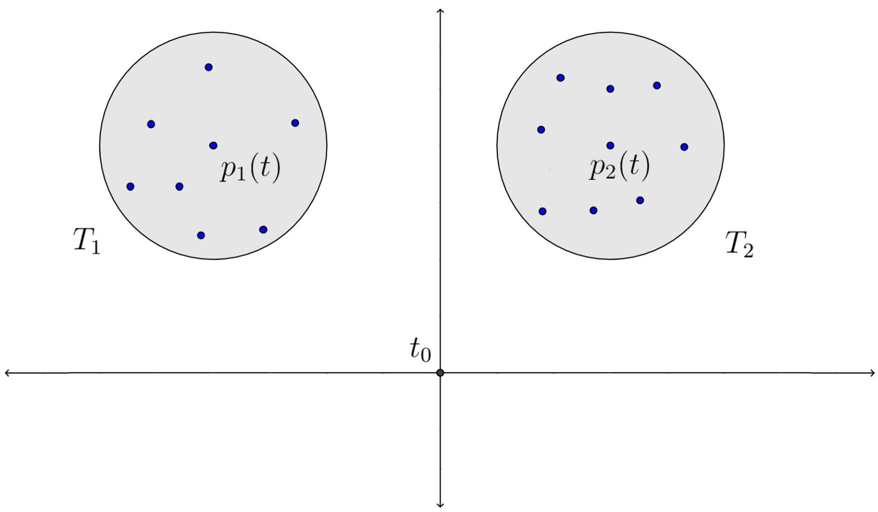

Figure 4.2: Gaussian random variables clustered together.

Therefore, we can write

$$X_t - X_{t_0} = X_t - X_{p_1(t)} + X_{p_1(t)} - X_{t_0}, \tag{4.2}$$

$$X_t - X_{t_0} = X_t - X_{p_2(t)} + X_{p_2(t)} - X_{t_0}, \tag{4.3}$$

where the terms $X_{p_1(t)} - X_{t_0}$ and $X_{p_2(t)} - X_{t_0}$ can be handled with (4.1), i.e.,

$$\mathbb{P}\left(\sup_{t\in F}(X_t - X_{t_0}) \geq u\right) \leq \mathbb{P}\left(X_{p_1(t)} - X_{t_0} \geq \frac{u}{2}\right) + \mathbb{P}\left(X_{p_2(t)} - X_{t_0} \geq \frac{u}{2}\right)$$
$$+ \sum_{t\in T_1} \mathbb{P}\left(X_t - X_{p_1(t)} \geq \frac{u}{2}\right) + \sum_{t\in T_2} \mathbb{P}\left(X_t - X_{p_2(t)} \geq \frac{u}{2}\right).$$

There is no reason to stop at one step and we can repeat this procedure on subsets of T_1, T_2, and the terms $X_t - X_{p_1(t)}$, $X_t - X_{p_2(t)}$. Clearly, this prescription works for any finite number of subsets and also for a different splitting than $u = u/2 + u/2$.
So we can approximate t_0 by a "chain" of representatives lying in a finite, ascending sequence of subsets of T, i.e., $\{t_0\} = T_0 \subset ... \subset T_n \subset T$, $n \in \mathbb{N}$. Hence, the name "generic chaining".
A way to choose these representative would be for example the "best approximations" of t in each of the subsets.
The quality of this approximation will be measured through comparison with numbers r^{-i}, $i \in \mathbb{Z}$, $r \geq 2$. For the set T we choose the largest $i \in \mathbb{Z}$ such that $\operatorname{diam}(T) \leq 2r^{-i}$.

We work out this prescription more elaborately now. For $j \geq i$, we consider finite sets $\Pi_j \subset T$ and points $\pi_j(t) \in \Pi_j$ for $t \in T$ such that the points $\pi_j(t)$ are successive approximations of t. Obviously, $\pi_i(t) = t_0$ has to hold. Hence, $X_t - X_{t_0}$ can be decomposed moving from t_0 to t along the "chain" $\pi_j(t)$, as

$$X_t - X_{t_0} = \sum_{j>i} X_{\pi_j(t)} - X_{\pi_{j-1}(t)}. \tag{4.4}$$

To ensure the convergence of (4.4) the mild condition

$$\lim_{j\to\infty} d(t, \pi_j(t)) = 0 \tag{4.5}$$

suffices, while in order to guarantee that $\pi_j(t)$ approximates t, it is enough to assume

$$\forall t \in T\ \forall j > i:\ d(\pi_j(t), \pi_{j-1}(t)) \leq 2r^{-j+1}. \tag{4.6}$$

Additionally, we assume the following two minor restrictions:

$$\forall s, t \in T:\ \pi_j(t) = \pi_j(s) \implies \pi_{j-1}(t) = \pi_{j-1}(s) \tag{4.7}$$

and

$$\forall v \in \Pi_j:\ \pi_j(v) = v. \tag{4.8}$$

Introducing (4.7) lowers the number of possible increments to control, while both of them combined give $\pi_{j-1}(t) = \pi_{j-1}(\pi_j(t))$. Hence, controlling $X_{\pi_j(t)} - X_{\pi_{j-1}(t)}$ means controlling $X_v - X_{\pi_{j-1}(v)}$, for all $v \in \Pi_j$.

Assume now that for certain numbers $a_j(v)$, depending on j and v, we have, for some $u > 0$,

$$\forall v \in \Pi_j : \ X_v - X_{\pi_{j-1}(v)} \leq u a_j(v). \tag{4.9}$$

Setting $S = \sup_{t \in T} \sum_{j>i} a_j(\pi_j(t))$, we get, by (4.5) and (4.9), that

$$\forall t \in F : \ X_t - X_{t_0} \leq uS.$$

Using now (4.4), (4.6) and the subgaussian tail estimate, one can show

$$\begin{aligned} \mathbb{P}(\sup_{t \in F}(X_t - X_{t_0}) \geq uS) &\leq \sum_{j>i} \sum_{v \in \Pi_j} \mathbb{P}(X_v - X_{\pi_{j-1}(v)} \geq u a_j(v)) \\ &\leq \sum_{j>i} \sum_{v \in \Pi_j} 2 \exp\left(-\frac{u^2 a_j^2(v)}{(2r^{-j+1})^2}\right). \end{aligned} \tag{4.10}$$

We require now that the right-hand side of (4.10) is less than or equal to 1 to make sense. The most natural choice is to require that this is true for $u = 1$. Hence,

$$\sum_{j>i} \sum_{v \in \Pi_j} 2 \exp\left(-\frac{a_j^2(v)}{(2r^{-j+1})^2}\right) \leq 1.$$

We set

$$w_j(v) = 2 \exp\left(-\frac{a_j^2(v)}{(2r^{-j+1})^2}\right) \tag{4.11}$$

and want to have that $\sum w_j(v) \leq 1$. This allows us to incorporate probability in this investigation.
Consider a probability measure μ on T and set

$$\forall j > 1 \ \forall v \in \Pi_j : \ w_j(v) = \mu(\{v\}).$$

Then, by (4.11), $a_j(v) = 2r^{-j+1}\sqrt{\log \frac{2}{\mu(\{v\})}}$ and

$$S = 2 \sup_{t \in T} \sum_{j>i} r^{-j+1} \sqrt{\log \frac{2}{\mu(\{\pi_j(t)\})}}.$$

For $u \geq 1$, the right-hand side of (4.10) gives $\sum_{j>i} \sum_{v \in \Pi_j} 2(\frac{w_j(v)}{2})^{u^2} \leq 2^{1-u^2}$, since $2(\frac{w_j(v)}{2})^{u^2} \leq w_j(v) 2^{1-u^2}$, ensuring $\mathbb{E} \sup_{t \in F}(X_t - X_{t_0}) \leq CS$. Thus,

$$\mathbb{E} \sup_{t \in T} X_t \leq C \sup_{t \in T} \sum_{j>i} r^{-j+1} \sqrt{\log \frac{2}{\mu(\{\pi_j(t)\})}}. \tag{4.12}$$

Consider now the partition $(\mathcal{A}_j)_{j\geq i}$ and choose, for each $j \geq i$ and $A \in \mathcal{A}_j$, an arbitrary point $x_A \in A$. For each $t \in T$ define

$$\pi_j(t) = x_{A_j(t)}. \tag{4.13}$$

Then, (4.5),(4.6),(4.7) and (4.8) hold, since $(\mathcal{A}_j)_{j\geq i}$ is an increasing sequence. From $\sum_{j>i}\sum_{A\in\mathcal{A}_j} 2^{-j+i}\mu(A) \leq 1$, we get that there exists a probability measure μ' on T such that $\mu'(\{x_A\}) > 2^{-j+i}\mu(A)$ for all $j > i$ and $A \in \mathcal{A}_j$. Applying (4.12) to μ' then gives

$$\mathbb{E}\sup_{t\in T} \leq \sup_{t\in T}\sum_{j>i} r^{-j+1}\sqrt{\log\frac{2^{j-i+1}}{\mu(A_j(t))}}. \tag{4.14}$$

Using the inequality $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$, we see that

$$\sqrt{\log\frac{2^{j-i+1}}{\mu(A_j(t))}} \leq \sqrt{j-i+1}\sqrt{\log 2} + \sqrt{\log\frac{1}{\mu(A_j(t))}},$$

and by using $\sum_{j>i} r^{-j}\sqrt{j-i+1} \leq C(r)r^{-i}$, we conclude that the right-hand side of (4.14) is at most

$$C(r)\left(r^{-i} + \sup_{t\in T}\sum_{j>i}\sqrt{\log\frac{1}{\mu(A_j(t))}}\right).$$

By the definition of i, we see that $\operatorname{diam}(T) > 2r^{-j-1}$. Assume that $\operatorname{card}(\mathcal{A}_{i+1}) = 1$. Then the only $A \in \mathcal{A}_{i+1}$ fulfills $\operatorname{diam}(A) \leq 2r^{-j-1}$ and implies that $\operatorname{diam}(T) \leq 2r^{-j-1}$. This is a contradiction and therefore $\operatorname{card}(\mathcal{A}_{i+1}) > 1$. Consequently, there exists an $A \in \mathcal{A}_{i+1}$ with $\mu(A) \leq 1/2$. Then, if $t \in A$, we have

$$r^{-i} \leq C(r)\left(r^{-i-1}\sqrt{\log\frac{1}{\mu(A_{i+1}(t))}}\right),$$

and subsequently

$$\mathbb{E}\sup_{t\in T} X_t \leq C(r)\sup_{t\in T}\sum_{j>i} r^{-j}\sqrt{\log\frac{1}{\mu(A_j(t))}}.$$

□

We now turn to the partitioning scheme, which will allow us to construct a partition as above. We will use this partition scheme in Lemma 4.3. This scheme will produce an increasing sequence of partitions of T, for which we can guarantee that there exists a uniform bound to the maximal diameter of the elements of each partition, and that the number of elements of the partitions are bounded in some sense. In particular, it will guarantee the finiteness of each partition $\mathcal{A}_j$.

We are going to work in the following framework: assume that we have maps $\varphi_j : T \to \mathbb{R}^+$, $j \in \mathbb{Z}$, for which

$$S = \sup\{\varphi_j(t) \ : \ j \in \mathbb{Z},\ t \in T\} < \infty \tag{4.15}$$

holds. Furthermore, assume a function $\theta : \mathbb{N} \to \mathbb{R}^+$ exists such that

$$\lim_{n\to\infty} \theta(n) = \infty. \tag{4.16}$$

Additionally, we assume for certain numbers $r \geq 4$, $\beta > 0$, that for any point $s \in T$, any $j \in \mathbb{Z}$ and any $n \in \mathbb{N}$ we have: for any point $t_1, ..., t_n \in B(s, r^{-j})$ for which

$$\forall p, q \leq n \ : \ p \neq q \implies d(t_p, t_q) \geq r^{-j-1} \tag{4.17}$$

holds, we have

$$\varphi_j(s) \geq r^{-\beta j}\theta(n) + \min_{\ell \leq n} \varphi_{j+2}(t_\ell). \tag{4.18}$$

In Figure 4.3, a generic instance of this framework is displayed. Given fixed r and β, the point a shown in the picture has to be below $\varphi_j(s)$ for any s,j,n and $t_1, ..., t_n$ as specified above.

Remark. The maps φ_j serve the purpose of being a measure of the size of $B(s, 3r^{-j})$. Since the numbers $\varphi_{j+2}(t_\ell)$ depend only on the respective $B(t_\ell, 3r^{-j-2})$ balls, we know on the one hand, by $d(t_\ell, t_{\ell'}) \geq r^{-j-1}$, for $\ell \neq \ell'$ and $r \geq 8$, that they are well separated from each other and on the other hand, by $t_\ell \in B(s, r^{-j})$ that they lie well inside of $B(s, 3r^{-j})$.
This is a rather weak condition, since we choose "min" instead of "max" and φ_{j+2} instead of φ_{j+1}.
To determine such functions one essentially has to guess them using their dependence on the geometry of T [T1].

Remark. As we will see later on, this setting is slightly more general than the one we will need. Namely, we will only use the case $\beta = 1$. Nevertheless, this does not change the proofs we will present, so we will not omit it.

As first step we give an example of such a sequence for $\beta = 1$.

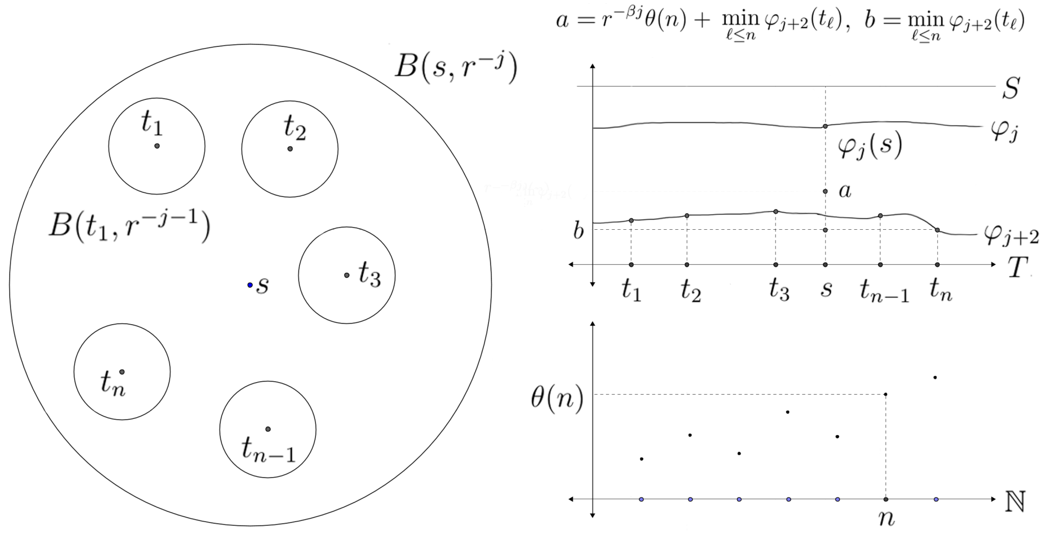

Figure 4.3: Illustration of the framework.

Lemma 4.2.
Let μ be a probability measure on T and let

$$S = \sup_{t \in T} \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t, \varepsilon))}}\, d\varepsilon < \infty.$$

By taking $r = 8$, $\beta = 1$ and defining

$$\varphi_j(t) = \sup \left\{ \int_0^{r^{-j}} \sqrt{\log \frac{1}{\mu(B(u, \varepsilon))}}\, d\varepsilon : \ d(t, u) \le 2r^{-j} \right\}$$

the conditions (4.15)-(4.18) hold with $\theta(n) = (1/r^2)\sqrt{\log n}$.

Proof. One easily sees that $\varphi_j(t) \le S < \infty$, for all $j \in \mathbb{Z}$ and all $t \in T$, holds, from which condition (4.15) follows. Moreover, for the choice of $\theta(n) = (1/r^2)\sqrt{\log n}$ also condition (4.16) holds, i.e., $\lim_{n\to\infty} \theta(n) = \infty$.
To prove that condition (4.18) is fulfilled, let $s, t_1, ..., t_n$ be as in (4.17).
For $\ell \le n$ let $s_\ell \in B(t_\ell, 2r^{-j-2})$. By the triangle inequality we get

$$d(s, s_\ell) \le d(s, t_\ell) + d(t_\ell, s_\ell) \overset{(4.17)}{\le} r^{-j} + 2r^{-j-2} \le 2r^{-j}.$$

We also see that for $\ell \ne k$ we have again by the triangle inequality

$$d(t_\ell, t_k) \le d(t_\ell, s_\ell) + d(s_\ell, s_k) + d(s_k, t_k) \le d(s_\ell, s_k) + 4r^{-j-2},$$

which gives us with the choice of $r = 8$, that

$$d(s_\ell, s_k) \geq d(t_\ell, t_k) - 4r^{-j-2} \overset{(4.17)}{\geq} r^{-j-1} - 4r^{-j-2} \overset{r=8}{\geq} 4r^{-j-2}.$$

This means that the (open) balls $B(s_\ell, 2r^{-j-2})$ are disjoint for $\ell \leq n$ and we can find ℓ such that $\mu(B(s_\ell, 2r^{-j-2})) \leq 1/n$, which simply follows from the inequality $1 \geq \mu(B(s, r^{-j})) \geq \sum_{\ell=1}^n \mu(B(s_\ell, 2r^{-j-2})) \geq n \min_{\ell=1,\ldots,n} \mu(B(s_\ell, 2r^{-j-2}))$. Since $s_\ell \in B(s, 2r^{-j-2})$, we have

$$\begin{aligned}
\varphi_j(t) &= \sup\left\{ \int_0^{r^{-j}} \sqrt{\log \frac{1}{\mu(B(u,\varepsilon))}}\, d\varepsilon : \ d(t,u) \leq 2r^{-j} \right\} \\
&\geq \int_0^{r^{-j}} \sqrt{\log \frac{1}{\mu(B(s_\ell,\varepsilon))}}\, d\varepsilon \\
&\geq \int_0^{2r^{-j-2}} \sqrt{\log \frac{1}{\mu(B(s_\ell,\varepsilon))}}\, d\varepsilon \\
&\geq \int_0^{r^{-j-2}} \sqrt{\log \frac{1}{\mu(B(s_\ell,\varepsilon))}}\, d\varepsilon + r^{-j-2}\sqrt{\log n},
\end{aligned}$$

where the last inequality follows from the integrand being at least $\sqrt{\log n}$ for $r^{-j-2} \leq \varepsilon \leq 2r^{-j-2}$. Thus, we have

$$\varphi_j(s) \geq \frac{1}{r^2} r^{-j}\sqrt{\log n} + \min_{\ell \leq n} \int_0^{r^{-j-2}} \sqrt{\log \frac{1}{\mu(B(s_\ell,\varepsilon))}}\, d\varepsilon,$$

from which we get

$$\varphi_j(s) \geq r^{-j}\theta(n) + \min_{\ell \leq n} \varphi_{j+2}(t_\ell)$$

by taking the supremum over all possible s_ℓ. □

The following two lemmas are the ones incorporating the scheme for constructing partitions as required in Theorem 4.1. In fact, the second lemma is just an alteration of the first, exchanging the condition (4.18).

Lemma 4.3.
Assume (4.15)-(4.18) holds. Let T have finite diameter, i.e., let $i \in \mathbb{Z}$ be the

largest integer such that $\mathrm{diam}(T) \leq 2r^{-i}$. Then there exists an increasing sequence of partitions $(\mathcal{A}_j)_{j\geq i}$ of T, and one can find a natural number $\ell_j(A)$ for each $A \in \mathcal{A}_j$, such that the following holds:
Each set of $\mathcal{A}_j$ has diameter at most $2r^{-j}$, and, given $j \geq i$, any two sets A and B of $\mathcal{A}_{j+1}$, which are contained in an element of $\mathcal{A}_j$, fulfill

$$\ell_{j+1}(A) \neq \ell_{j+1}(B) \tag{4.19}$$

and

$$\forall t \in T: \sum_{j\geq i} r^{-\beta j}\theta(\ell_{j+1}(A_{j+1}(t))) \leq 4S. \tag{4.20}$$

Remark. The conditions (4.19) and (4.20) combined with the assumptions (4.17) and (4.18) yield the finiteness of the partitions $\mathcal{A}_j$.

Proof. We will show this by inductively constructing the partition: for $j = i$, we set $\mathcal{A}_i = \{T\}$, $\ell_i(T) = 1$ and we choose a distinguished point $u_i(T)$ such that

$$\varphi_{i+2}(u_i(T)) \geq \sup\{\varphi_{i+2} : t \in T\} - \frac{S}{2}.$$

Assume now that the j-th partition $\mathcal{A}_j$ has already been constructed and distinguished points $u_j(A)$, for each $A \in \mathcal{A}_j$, have been chosen for which

$$\forall t \in A: \ d(t, u_j(A)) \leq r^{-j} \tag{4.21}$$

holds. In particular this means that $\mathrm{diam}(A) \leq 2r^{-j}$.
We will now go on by constructing the elements of $\mathcal{A}_{j+1}$ from the elements of $\mathcal{A}_j$ by an exhaustion argument which goes as follows: at the first step, see Figure 4.4, we pick $t_1 \in A$ such that

$$\varphi_{j+2}(t_1) \geq \sup\{\varphi_{j+2}(t) : t \in A\} - 2^{i-j-1}S,$$

and define the set $D_1 := A \cap B(t_1, r^{-j-1})$.

We repeat this procedure on $A \setminus D_1$ instead of A until the set A is exhausted, i.e., we construct points $t_1, ..., t_p \in A$ such that

$$t_p \in A \setminus \bigcup_{\ell<p} B(t_\ell, r^{-j-1}),$$

$$\varphi_{j+2}(t_p) \geq \sup\left\{\varphi_{j+2}(t) : t \in A \setminus \bigcup_{\ell<p} B(t_\ell, r^{-j-1})\right\} - 2^{i-j-1}S. \tag{4.22}$$

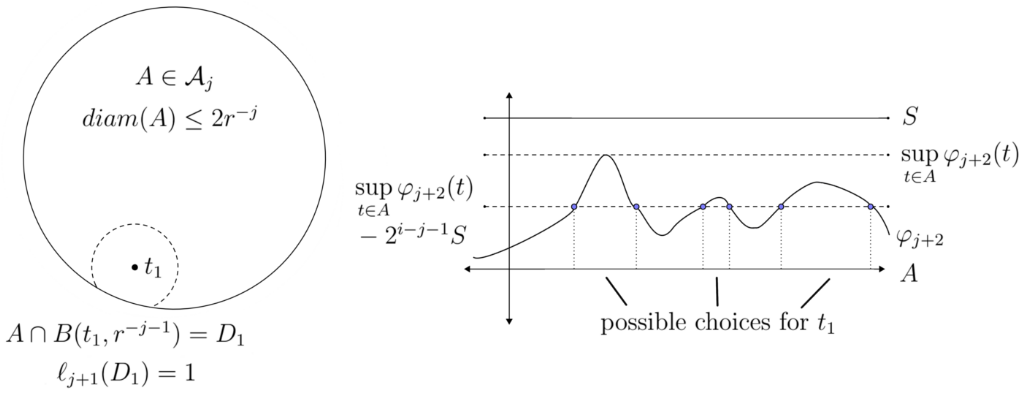

Figure 4.4: Illustration of the first selection.

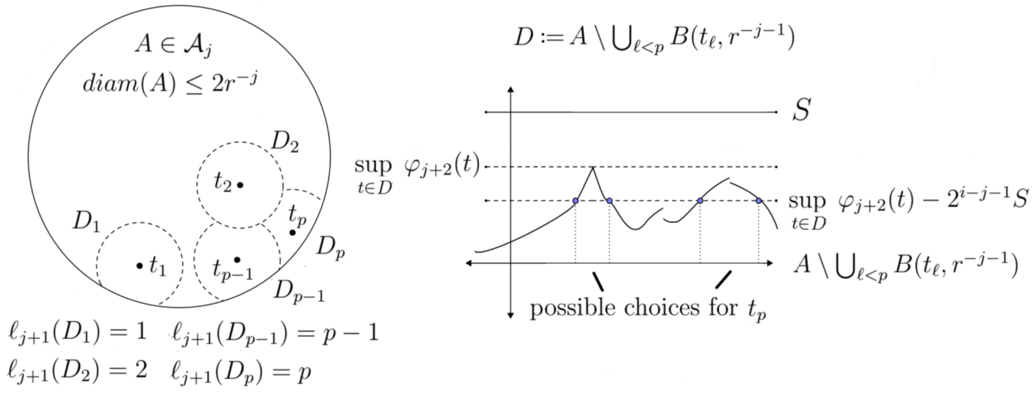

Figure 4.5: Illustration of the p-th selection.

and define the set $D_p := B(t_p, r^{-j-1}) \cap \Big(A \setminus \bigcup_{\ell<p} B(t_\ell, r^{-j-1})\Big)$. See Figure 4.5 at this point

The set containing the sets D_p is thereby a finite partition $(D_p)_{p\geq 1}$ of the set A, where each element D_p has diameter at most $2r^{-j-1}$. Partitioning each of the sets $A \in \mathcal{A}_j$ in this way and collecting them into a single set gives the new partition $\mathcal{A}_{j+1}$.

Choosing now distinguished points and numbers

$$u_{j+1}(D_p) = t_p \qquad \text{and} \qquad \ell_{j+1}(D_p) = p, \tag{4.23}$$

we see that (4.19) is indeed fulfilled for the partition $\mathcal{A}_{j+1}$.

It remains to prove (4.20).

From (4.21) we see that $d(u_j(A), t_p) \leq r^{-j}$ holds for each p. Moreover, we have constructed the partition in such a way that $d(t_\ell, t_k) \geq r^{-j-1}$ for all $\ell < k$. Since, by assumption, condition (4.18) holds, we get for each p

$$\varphi_j(u_j(A)) \geq r^{-\beta j}\theta(p) + \min_{\ell \leq p} \varphi_{j+2}(t_\ell). \tag{4.24}$$

We also know that $A \setminus \bigcup_{k<p} B(t_k, r^{-j-1}) \subset A \setminus \bigcup_{k<\ell} B(t_k, r^{-j-1})$ and therefore

$$\begin{aligned} &\sup\left\{\varphi_{j+2}(t) : t \in A \setminus \bigcup_{k<\ell} B(t_k, r^{-j-1})\right\} \\ &\geq \sup\left\{\varphi_{j+2}(t) : t \in A \setminus \bigcup_{k<p} B(t_k, r^{-j-1})\right\}. \end{aligned}$$

Since $t_p \in A \setminus \bigcup_{\ell<p} B(t_\ell, r^{-j-1})$, we get by applying (4.22) to t_ℓ, $\ell \leq p$, that

$$\begin{aligned} \varphi_{j+2}(t_\ell) &\geq \sup\left\{\varphi_{j+2}(t) : t \in A \setminus \bigcup_{k<\ell} B(t_k, r^{-j-1})\right\} - 2^{i-j-1}S \\ &\geq \varphi_{j+2}(t_p) - 2^{i-j-1}S. \end{aligned}$$

Combining this with (4.24) we get

$$\varphi_j(u_j(A)) \geq \varphi_{j+2}(t_p) + r^{-\beta j}\theta(p) - 2^{i-j-1}S. \tag{4.25}$$

Let $t \in D_p$. Setting $A := A_j(t)$ and $A_{j+1}(t) := D_p$ we have $\ell_{j+1}(A_{j+1}(t)) = \ell_{j+1}(D_p) = p$, and we can rewrite (4.25) as

$$\varphi_j(u_j(A_j(t))) \geq \varphi_{j+2}(t_p) + r^{-\beta j}\theta(\ell_{j+1}(A_{j+1}(t))) - 2^{i-j-1}S. \tag{4.26}$$

Setting $u := u_{j+2}(A_{j+2}(t))$ and observing that $u \in A_{j+2}(t) \subset A_{j+1}(t) = D_p$, we can conclude that

$$\begin{aligned} \varphi_{j+2}(t_p) &\geq \sup\left\{\varphi_{j+2}(t) : t \in A \setminus \bigcup_{\ell<p} B(t_\ell, r^{-j-1})\right\} - 2^{i-j-1}S \\ &\geq \sup\{\varphi_{j+2}(t) : t \in D_p\} - 2^{i-j-1}S \geq \varphi_{j+2}(u) - 2^{i-j-1}S. \end{aligned}$$

The combination of this with (4.26) yields

$$\varphi_j(u_j(A_j(t))) \geq \varphi_{j+2}(u_{j+2}(A_{j+2}(t))) + r^{-\beta j}\theta(\ell_{j+1}(A_{j+1}(t))) - 2^{i-j}S.$$

Now we can rearrange this inequality and sum up over $j \geq i$ to get

$$\begin{aligned} &\sum_{j\geq i} r^{-\beta j}\theta(\ell_{j+1}(A_{j+1}(t))) \\ &\leq \sum_{j\geq i} \varphi_j(u_j(A_j(t))) - \sum_{j\geq i} \varphi_{j+2}(u_{j+2}(A_{j+2}(t))) + \sum_{j\geq i} 2^{i-j}S \\ &\leq \varphi_i(u_i(A_i(t))) + \varphi_{i+1}(u_{i+1}(A_{i+1}(t))) + 2S \leq 4S, \end{aligned}$$

for any $t \in T$, which closes the proof. □

Lemma 4.4.
Let (T,d) be a metric space. For each $j \in \mathbb{Z}$ let $\psi_j : T \to \mathbb{R}^+$ be such that $\psi_j(t) \leq S$ for each $j \in \mathbb{Z}$ and $t \in T$. Under (4.17) let, instead of (4.18), the following hold:

$$\max_{\ell \leq n} \psi_{j+2}(t_\ell) \geq \psi_j(s) + r^{-\beta j}\theta(n). \tag{4.27}$$

Then the conclusions of Lemma 4.3 hold.

Proof. By defining $\varphi_j(t) := S - \psi_j(t)$ one sees that (4.27) yields (4.18), i.e., $\varphi_j(s) \geq r^{-\beta j}\theta(n) + \min_{\ell \leq n} \varphi_{j+2}(t_\ell)$. Applying now Lemma 4.3 gives the lemma. □

The next theorem allows to further estimate the right-hand side of Theorem 4.1, given the sequence of partitions $(\mathcal{A}_j)_{j \geq i}$ satisfies the property (4.19).

Lemma 4.5.
Let (T,d) be a metric space and $(\mathcal{A}_j)_{j \geq i}$ be an increasing sequence of partitions of T such that each $A \in \mathcal{A}_j$, $j \geq i$, has an associated number $\ell_j(A) \in \mathbb{N}$ satisfying (4.19). Then there exists a probability measure μ on T such that, given $\alpha, \beta > 0$,

$$\sup_{t \in T} \sum_{j > i} r^{-\beta j} \sqrt[\alpha]{\log \frac{1}{\mu(A_j(t))}} \leq C \left(r^{-\beta i} + \sup_{t \in T} \sum_{j > i} r^{-\beta j} \sqrt[\alpha]{\log \ell_j(A_j(t))} \right), \tag{4.28}$$

with a constant $C > 0$ depending only on α, β and r.

Remark. Again, this is a slightly more general result than we need. We will only use this lemma with $\beta = 1$ and $\alpha = 2$. Here we will only show the proof for the case that $\alpha \geq 1$. For the general proof see [T1].

Proof. Let $j \geq i$ and $A \in \mathcal{A}_j$. We will construct numbers $w_j(A)$ inductively as follows:
Set $w_i(T) = 1$. Assume that numbers $w_{j-1}(A)$ have been constructed for $A \in \mathcal{A}_{j-1}$. Let now $B \in \mathcal{A}_j$ such that $B \subset A$ and define the number

$$w_j(B) = \frac{1}{4\ell_j(B)^2} w_{j-1}(A). \tag{4.29}$$

From the property (4.19) of the partition we see that we can use the simple inequality $\sum_{\ell \geq 1} \ell^{-2} \leq 2$ to give, together with (4.29), the following estimate

$$\sum_{\substack{B \in \mathcal{A}_j \\ B \subset A}} w_j(B) = \frac{1}{4} w_{j-1}(A) \sum_{\substack{B \in \mathcal{A}_j \\ B \subset A}} \frac{1}{\ell_j(B)^2} \leq \frac{1}{2} w_{j-1}(A).$$

From this we get, by induction over j, that

$$\sum_{j>i}\sum_{A\in\mathcal{A}_j} w_j(A) \le \sum_{j>i} 2^{i-j} w_i(T) = \sum_{j>i} 2^{i-j} \overset{i\ge 1}{\le} 1.$$

Hence, there exists a probability measure μ on T such that

$$\forall j \ge i\ \forall A \in \mathcal{A}_j :\ \mu(A) \ge w_j(A),$$

and in particular

$$\log\frac{1}{\mu(A_j(t))} \le \log\frac{1}{w_j(A_j(t))}. \tag{4.30}$$

Now let $t \in T$ and $j > i$. Again by (4.29), we see that

$$w_j(A_j(t)) = \frac{1}{4\ell_j(A_j(t))^2} w_{j-1}(A_{j-1}(t)) = ... = 4^{i-j}\prod_{k>i}^{j} \ell_k(A_k(t))^{-2},$$

from which we get, together with (4.30), that

$$\log\frac{1}{\mu(A_j(t))} \le \log\left(4^{i-j}\prod_{k>i}^{j}\ell_k(A_k(t))^{-2}\right) = (j-i)\log 4 + 2\sum_{k>i}^{j}\log \ell_k(A_k(t))$$

Let now $\alpha \ge 1$. Then the basic inequality $\sqrt[\alpha]{x+y} \le \sqrt[\alpha]{x} + \sqrt[\alpha]{y}$ implies

$$\sqrt[\alpha]{\log\frac{1}{\mu(A_j(t))}} \le C\sqrt[\alpha]{j-i} + \sqrt[\alpha]{2}\sum_{k>i}^{j}\sqrt[\alpha]{\log \ell_k(A_k(t))},$$

which we can rearrange by changing the order of summation as

$$\sum_{j>i} r^{-\beta j}\sqrt[\alpha]{\log\frac{1}{\mu(A_j(t))}}$$

$$\le C\sum_{j>i} r^{-\beta j}\sqrt[\alpha]{j-i} + \sqrt[\alpha]{2}\sum_{k>i}\left(\sum_{j\ge k} r^{-\beta j}\right)\sqrt[\alpha]{\log \ell_k(A_k(t))}.$$

Since we can find constants such that $\sum_{j>i} r^{-\beta j}\sqrt[\alpha]{j-i} \le C(\alpha,\beta,r)\cdot r^{-\beta i}$ and $\sqrt[\alpha]{2}\sum_{j\ge k} r^{-\beta j} \le C(\alpha,\beta,r)\cdot r^{-\beta k}$, we can further estimate the above expression by

$$\sum_{j>i} r^{-\beta j}\sqrt[\alpha]{\log\frac{1}{\mu(A_j(t))}} \le C(\alpha,\beta,r)\left(r^{-\beta i} + \sum_{k>i} r^{-\beta k}\sqrt[\alpha]{\log \ell_k(A_k(t))}\right).$$

Taking the supremum over all $t \in T$ completes the proof. □

As an application of the majorizing measures theorem we introduce a theorem which we will use in the next chapter to provide an approximation result for convex bodies.
It states that for a process X and a sequence as in (4.15)-(4.18), $\mathbb{E}\sup_{t\in T} X_t$ is bounded by a constant only depending on r.

Remark. Note that so far we have shown everything for metric spaces, but in fact all the theorems, except the example given in form of Lemma 4.2, also hold in semimetric spaces.

Theorem 4.6.
Let (T,d) be a semimetric space and let $(X_t)_{t\in T}$ be a collection of centered random variables with subgaussian tail estimate

$$\mathbb{P}\left(|X_t - X_s| > a\right) \le \exp\left(-c\frac{a^2}{d^2(t,s)}\right), \qquad a > 0.$$

Let $r > 1$ and let k_0 be a natural number such that the $\operatorname{diam}(T) < r^{-k_0}$. Let $\{\varphi_k\}_{k=k_0}^{\infty}$ be a sequence of functions from T to $\mathbb{R}^+$, uniformly bounded by a constant depending only on r. Assume the existence of $\sigma > 0$ such that for any k the functions φ_k satisfy for any $s \in T$ and for all $t_1, ..., t_N \in B_{t^{-k}}(s)$ with mutual distances at least r^{-k-1} one has

$$\max_{j=1,\dots,N} \varphi_{k+2}(t_j) \ge \varphi_k(s) + \sigma r^{-k}\sqrt{\log N}. \tag{4.31}$$

Then, we have

$$\mathbb{E}\sup_{t\in T} X_t \le C(r)\sigma^{-1}.$$

Proof. Since we know that $\operatorname{diam}(T) < r^{-k_0}$, Lemma 4.4 gives us the existence of an increasing sequence of partitions $(\mathcal{A}_k)_{k\ge k_0}$ of T and natural numbers $\ell_k(A)$, for each $A \in \mathcal{A}_j$.
This partition is such that each $A \in \mathcal{A}_j$ has $\operatorname{diam}(A) \le 2r^{-k}$ and that, given $k \ge k_0$, for any two sets $A, B \in \mathcal{A}_{j+1}$ which are contained in the same element of $\mathcal{A}_j$, the numbers satisfy $\ell_{j+1}(A) \ne \ell_{j+1}(B)$. Moreover, the lemma gives us for each $t \in T$ the estimate (note: $\beta = 1$)

$$\sum_{k\ge k_0} r^{-k}\sqrt{\log \ell_{k+1}(A_{k+1}(t))} \le 4S\sigma^{-1}. \tag{4.32}$$

Furthermore, we obtain from this

$$\begin{aligned} r^{-k_0} &\le r^{-k_0} + \sqrt{\log \ell_{k_0+1}(A_{k_0+1}(t))}\sum_{k>k_0} r^{-k}\sqrt{\log \ell_{k+1}(A_{k+1}(t))} \\ &\le 4S\sigma^{-1}\sqrt{\log \ell_{k_0+1}(A_{k_0+1}(t))}. \end{aligned}$$

Since there exists a $p \in \mathbb{N}$ such that $\ell_{k_0+1}(A_{k_0+1}(t) \leq p$ for all $t \in T$, we have

$$r^{-k_0} \leq 4S\sigma^{-1}\sqrt{\log p}. \tag{4.33}$$

Using now Theorem 4.1 and Lemma 4.5 we obtain that

$$\begin{aligned}
\mathbb{E}\sup_{t\in T} X_t &\overset{Theorem\ 4.1}{\leq} C(r)\sup_{t\in T}\sum_{k>k_0} r^{-k}\sqrt{\log\frac{1}{\mu(A_k(t))}} \\
&\overset{Lemma\ 4.5}{\leq} C(r)\left(r^{-k_0} + \sup_{t\in T}\sum_{k>k_0} r^{-k}\sqrt{\log \ell_k(A_k(t))}\right) \\
&\overset{(4.32),(4.33)}{\leq} C(r)\sigma^{-1}.
\end{aligned}$$

□

Chapter 5

Contact points of convex bodies

In this chapter we will see an application of the majorizing measures theorem in convex geometry. Namely, we will improve the bound for the number of contact points of the famous theorem of John and derive that for each convex body B in $\mathbb{R}^M$ there exists another convex body K in $\mathbb{R}^M$ arbitrarily "close" to it, which has less than $CM \log M$ contact points with its maximal volume ellipsoid for an absolute constant $C > 0$ (Theorem 5.3).

This goal will be achieved by deriving two crucial results along the way:
The first result, Lemma 5.7, which incorporates the majorizing measures theorem, will be concerned with estimating the expectation of the supremum of the absolute value of a weighted "Euclidean scalar product" over the Euclidean unit ball of an n-dimensional Subspace W of $\mathbb{R}^M$. More precisely, each of the squares of the components of a vector in the sum defining the Euclidean scalar product will be weighted with independent Rademacher random variables. As we will see, this can be bounded from above by means of the logarithm of the dimension of the space, i.e., $\sqrt{\log M}$, and by the operator norm of the orthogonal projection onto the subspace W, considered with the ℓ_1^M-norm for the domain and the ℓ_2^M-norm for the image, i.e., $\|P_W : \ell_1^M \to \ell_2^M\|$.

The second result, Theorem 5.8, which derives from the aforementioned one, will show that an $n \times M$-Matrix, $n \leq M$, with orthonormal rows and specifically bounded columns, namely, their Euclidean norm should be bounded by some $t \geq \sqrt{n/M}$, possesses submatrices which are "almost" orthogonal. The size of theses almost orthogonal submatrices depends upon the quality of approximation, t and n. The dependence on n is of the type $n \log n$. This result from [R1] is equivalent to our goal, as we will use this to explicitly construct the convex body K from Theorem 5.3.

Let us first introduce some preliminary notions. By a convex body $K \subseteq \mathbb{R}^M$ we mean a convex and compact set with nonempty interior. With $\mathcal{K}^M$ we denote the set of all M-dimensional convex bodies. Let $K \subseteq L \subseteq \mathbb{R}^M$ be two convex bodies, then we call a point $x \in \partial K \cap \partial L$ a contact point of the two bodies.

Throughout this chapter we write $e_1, ..., e_M$ for the canonical basis in $\mathbb{R}^M$. By $\|\cdot\|_p$, $1 \leq p \leq \infty$ we refer to the ℓ_p^M-norms, where we use $\|\cdot\| := \|\cdot\|_2$ for the ℓ_2^M-norm as shorthand notation. Moreover, we denote by B_p^M the closed unit ball in ℓ_p^M, for all $1 \leq p \leq \infty$. We will also use the notation $B_\rho(w)$ for the closed ρ-ball, $\rho > 0$, centered at $w \in W$ in the (semi-)metric d on W, where W is a n-dimensional subspace of $\mathbb{R}^M$.
Furthermore, we will use an extended definition of the Banach-Mazur distance of convex bodies to determine how “similar” the shapes of two convex bodies are. Note that the classical Banach-Mazur distance measures how close two normed spaces are to being isometrically isomorphic, which can be translated to symmetric convex bodies, since those are the unit balls of normed spaces. From this translation one can go on to extend the definition to non-symmetric convex bodies. We have the following definitions:

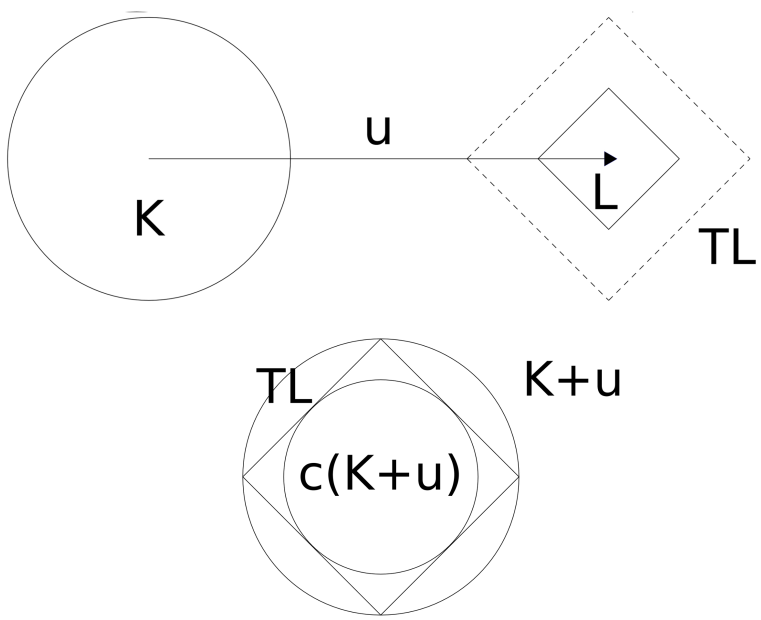

Figure 5.1: Illustration of the Banach-Mazu distance for two convex bodies K and L and vector u and operator T for which the constant c is attained.

Definition 5.1. (Banach-Mazur distance)
For two convex bodies $K, L \in \mathcal{K}^M$ we define the **Banach-Mazur distance** as

$$d(K, L) := \inf\{c > 0 | K + u \subset TL \subset c(K + u)\},$$

where the infimum is taken over all vectors $u \in \mathbb{R}^M$ and all invertible operators T.

Definition 5.2. (John ellipsoid) Let $K \in \mathcal{K}^M$. The **John ellipsoid** of K is the maximum volume ellipsoid contained in K.

The goal of this chapter is to show the following theorem, which is due to M. Rudelson. We will follow his approach from [R1] and [R2]:

Theorem 5.3.
Let $B \in \mathcal{K}^M$ and let $\varepsilon > 0$. Then there exists $K \in \mathcal{K}^M$ such that $d(K, B) \leq 1 + \varepsilon$ and the number of contact points of K with its John ellipsoid is less then

$$m(M, \varepsilon) = C(\varepsilon) M \log M.$$

Furthermore, if the John ellipsoid of K is B_2^M, then the identity operator id_M in $\mathbb{R}^M$ has the decomposition

$$\mathrm{id}_M = \sum_{i=1}^{m} a_i u_i \otimes u_i, \tag{5.1}$$

where $m \leq m(M, \varepsilon)$, $u_1, ..., u_m$ are the only contact points of K with B_2^M,

$$\sum_{i=1}^{m} a_i u_i = 0 \tag{5.2}$$

and, for every $i = 1, ..., m$, $\frac{m}{M} a_i \in [1 - \varepsilon, 1 + \varepsilon]$.

This is an astonishing result, since it means that for any convex body B there can be found another convex body K with geometric distorsion less than $1 + \varepsilon$ such that the number of the contact points of K with its John ellipsoid is bounded by a constant depending only on ε and growing at most like $M \log M$ with the dimension M.

Remark. This result has already been improved. Namely, Srivastava, Spielman and Batson showed in [S-S-B] that this theorem holds with $m(M, \varepsilon) \leq C(\varepsilon) M$ for symmetric convex bodies. Moreover, they showed that for any convex body K there exists a convex body L such that $d(K, L) \leq 2.24$ and the number of contact points is bounded by $m(M, \varepsilon) \leq C(\varepsilon) M$.

We will start our investigation by introducing a few lemmas, which will help us to obtain an intermediate result on almost orthogonal submatrices of orthogonal matrices on $\mathbb{R}^M$ (see Theorem 5.8). In [R2], Mark Rudelson states the equivalence of Theorem 5.8 and Theorem 5.3. We will only show that Theorem 5.8 implies Theorem 5.3.

Definition 5.4. (ε-Entropy)
Let (X, d) be a (semi-)metric space and let A, B be subsets of X. By $N(B, d, \varepsilon)$, the so-called ε**-entropy**, we denote the number of closed ε-balls needed to cover B in the (semi-)metric d.
Additionally, we denote by $N(A, B)$ the translates of the set B by elements of the set A to cover A.

In the following parts of this chapter we will make use of orthogonal projections. For the definition, see Definition A.39 in the appendix.

Lemma 5.5.
Let W be an n-dimensional subspace of $\mathbb{R}^M$, let P_W be the orthogonal projection onto W. Let $a_1, ..., a_M \in \mathbb{R} \setminus \{0\}$ and let, for all $x \in \mathbb{R}^M$, $\|\cdot\|_{\mathcal{E}}$ be the norm defined by

$$\|x\|_{\mathcal{E}} := \left(\sum_{i=1}^{M} x^2(i) a_i^2 \right)^{1/2}.$$

Then, for all $\varepsilon > 0$,

(i) $\varepsilon \sqrt{\log N(B_2^M \cap W, \|\cdot\|_\infty, \varepsilon)} \leq C \left\| P_W : \ell_1^M \to \ell_2^M \right\| \sqrt{\log M}$, and

(ii) $\varepsilon \sqrt{\log N(B_2^M \cap W, \|\cdot\|_{\mathcal{E}}, \varepsilon)} \leq C \left\| P_W : \ell_1^M \to \ell_2^M \right\| \left(\sum_{i=1}^M a_i^2 \right)^{1/2}$.

Remark. We write the norm $\|\cdot\|_{\mathcal{E}}$ with an $\mathcal{E}$ to denote the fact that its unit ball, also denoted $\mathcal{E}$, is an ellipsoid, i.e., $\mathcal{E} := \{x \in \mathbb{R}^M \ : \ \sum_{i=1}^M x^2(i) a_i^2 \leq 1\}$ with semi-axis $1/|a_i|$, $i = 1, ..., M$. This norm is particularly useful when considering a John ellipsoid of a convex body.

Proof. Let g be a standard Gaussian random vector in $\mathbb{R}^M$. Then $P_W g$ is a standard Gaussian vector in W (Appendix, Lemma A.40).
For a standard Gaussian vector we can apply the dual Sudakov minoration (see Appendix, Theorem A.41 and subsequent remarks) which gives for (i)

that

$$\begin{aligned}
\sup_{\varepsilon>0} \varepsilon \sqrt{\log N(B_2^M \cap W, \|\cdot\|_\infty, \varepsilon)} &= \sup_{\varepsilon>0} \varepsilon \sqrt{\log N(B_2^M \cap W, \varepsilon(B_\infty^M \cap W))} \\
&\leq C\mathbb{E} \sup_{t\in(B_\infty^M \cap W)^\circ} \left| \sum_{i=1}^n (P_W g)_i t_i \right| \\
&= C\mathbb{E} \sup_{t\in(B_\infty^M \cap W)^\circ} |\langle P_W g, t\rangle| \\
&= C\mathbb{E} \sup_{t\in P_W\left(B_1^M\right)} |\langle P_W g, t\rangle| .
\end{aligned}$$

The last equality follow from the fact that $(B_\infty^M \cap W)^\circ = P_W(B_1^M)$. (Appendix, Theorem A.41).
Since for every $t \in P_W(B_1^M)$ there exists an $s \in B_1^M$ such that $P_W(s) = t$ and the orthogonal projection P_W is self-adjoint and idempotent, i.e., $P_W^2 = P_W$, we get

$$\begin{aligned}
C\mathbb{E} \sup_{t\in P_W\left(B_1^M\right)} |\langle P_W g, t\rangle| &= C\mathbb{E} \sup_{s\in B_1^M} |\langle P_W g, P_W s\rangle| \\
&= C\mathbb{E} \sup_{s\in B_1^M} |\langle P_W P_W g, s\rangle| \\
&= C\mathbb{E} \sup_{s\in B_1^M} |\langle P_W g, s\rangle|
\end{aligned}$$

Now we use the extreme point estimate

$$\begin{aligned}
\sup_{s\in B_1^M} |\langle P_W g, s\rangle| &= \sup_{s\in B_1^M} \left| \sum_{i=1}^M (P_W g)(i) s_i \right| \\
&\leq \sup_{s\in B_1^M} \sum_{i=1}^M |(P_W g)(i)||s_i| \\
&\leq \max_{j=1,\ldots,M} |(P_W g)(j)| \sup_{s\in B_1^M} \sum_{i=1}^M |s_i| \\
&= \max_{j=1,\ldots,M} |\langle P_W g, e_j\rangle| \sup_{s\in B_1^M} \|s\|_1 \\
&= \max_{j=1,\ldots,M} |\langle P_W g, e_j\rangle|
\end{aligned}$$

to show that

$$C\mathbb{E} \sup_{s\in B_1^M} |\langle P_W g, s\rangle| \leq C\mathbb{E} \max_{j=1,\ldots,M} |\langle P_W g, e_j\rangle| = C\mathbb{E} \max_{j=1,\ldots,M} |\langle g, P_W e_j\rangle| .$$

From Lemma A.44, in the appendix, we have, for all $k = 1, ..., M$, that

$$\langle g, P_W e_j \rangle = \sum_{i=1}^{M} (P_W e_j)(i) g_i \overset{\mathcal{D}}{=} \left(\sum_{i=1}^{M} (P_W e_j)(i)^2 \right)^{1/2} g_k = \|P_W e_j\| \cdot g_k,$$

which gives us that

$$\begin{aligned} C \cdot \mathbb{E} \max_{j=1,\dots,M} |\langle g, P_W e_j \rangle| &= C \cdot \mathbb{E} \max_{j=1,\dots,M} |\|P_W e_j\| \cdot g_k| \\ &\leq C \max_{j=1,\dots,M} \|P_W e_j\| \cdot \mathbb{E} \max_{k=1,\dots,M} |g_k|. \end{aligned}$$

To estimate this further, note that for independent folded standard normal distributions $|Z_i|$, we have $\mathbb{E} \max_{i=1,\dots,M} |Z_i| \leq C\sqrt{\log M}$ (Appendix, Lemma A.43). Using this, we obtain

$$\begin{aligned} C \max_{j=1,\dots,M} \|P_W e_j\| \cdot \mathbb{E} \max_{k=1,\dots M} |g_k| &\leq C \max_{j=1,\dots,M} \|P_W e_j\| \sqrt{\log M} \\ &\leq C \left\| P_W : \ell_1^M \to \ell_2^M \right\| \sqrt{\log M}. \end{aligned}$$

Since this chain of inequality holds for the supremum over all $\varepsilon > 0$, this holds for any $\varepsilon > 0$ and we have shown part (i) of this lemma.

Analogously, we get for (ii) that

$$\begin{aligned} &\sup_{\varepsilon>0} \varepsilon \sqrt{\log N(B_2^M \cap W, \|\cdot\|_{\mathcal{E}}, \varepsilon)} = \sup_{\varepsilon>0} \varepsilon \sqrt{\log N(B_2^M \cap W, \varepsilon(B_{\mathcal{E}}^M \cap W))} \\ &\leq C\mathbb{E} \sup_{t \in \left(B_{\mathcal{E}}^M \cap W\right)^\circ} |\langle P_W g, t \rangle| = C\mathbb{E} \sup_{t \in P_W\left(\left(B_{\mathcal{E}}^M\right)^\circ\right)} |\langle P_W g, t \rangle| \\ &= C\mathbb{E} \sup_{s \in \left(B_{\mathcal{E}}^M\right)^\circ} |\langle P_W g, P_W s \rangle| = C\mathbb{E} \sup_{s \in \left(B_{\mathcal{E}}^M\right)^\circ} |\langle P_W g, s \rangle| \\ &= C\mathbb{E} \|P_W g\|_{\mathcal{E}} \leq C \left(\mathbb{E} \|P_W g\|_{\mathcal{E}}^2 \right)^{1/2} \\ &= C \left(\mathbb{E} \sum_{i=1}^{M} \langle P_W g, e_i \rangle^2 a_i^2 \right)^{1/2} = C \left(\mathbb{E} \sum_{i=1}^{M} \langle g, P_W e_i \rangle^2 a_i^2 \right)^{1/2}. \end{aligned}$$

As above we use the consequence from Lemma A.44, for all $k = 1, ..., M$, that $\langle g, P_W e_i \rangle \overset{D}{=} \|P_W e_i\| \cdot g_k$ to show that

$$\begin{aligned}
C\left(\mathbb{E}\sum_{i=1}^{M}\langle g, P_W e_i\rangle^2 a_i^2\right)^{1/2} &\leq C\left(\mathbb{E}\sum_{i=1}^{M}\|P_W e_i\|^2 a_i^2 g_k^2\right)^{1/2} \\
&= C\left(\sum_{i=1}^{M}\|P_W e_i\|^2 a_i^2 \mathbb{E} g_k^2\right)^{1/2} \\
&\overset{\mathbb{E}g_k^2=1}{\leq} C\left(\sum_{i=1}^{M}\|P_W e_i\|^2 a_i^2\right)^{1/2} \\
&\leq C \max_{j=1,\ldots,M}\|P_W e_j\|\left(\sum_{i=1}^{M} a_i^2\right)^{1/2} \\
&\leq C\left\|P_W : \ell_1^M \to \ell_2^M\right\|\left(\sum_{i=1}^{M} a_i^2\right)^{1/2}.
\end{aligned}$$

□

Lemma 5.6.
Let W be an n-dimensional subspace of $\mathbb{R}^M$. $w \in W$ and $\rho > 0$. Consider the closed ρ-ball $B_\rho(w)$ in the semimetric

$$d(v, w) := \left(\sum_{i=1}^{M}(v(i) - w(i))^2(v(i)^2 + w(i)^2)\right)^{1/2}.$$

Then, we have

$$\operatorname{conv} B_\rho(w) \subset B_{4\rho}(w).$$

Proof. First, we check that d is a semimetric.
Let $v, w \in W$. From the definition of d we see that $d(v, w) \geq 0$ and that $d(v, w) = 0$ if and only if $v = w$. Also the symmetry of d follows immediately from the symmetries of $(v(i) - w(i))^2$ and $v(i)^2 + w(i)^2$.
On the other hand we see for $a, b \in \mathbb{R}$ that the inequality

$$(a - b)^2(a^2 + b^2) = a^4 + b^4 + 2a^2b^2 - 2a^3b - 2ab^3 \leq a^4 + b^4$$

does not hold true in general. Hence, the triangle inequality does not hold and d is indeed a semimetric.

Let now $v \in B_\rho(w)$. Obviously, we have

$$\left(\sum_{i=1}^{M}(v(i) - w(i))^2 w^2(i)\right)^{1/2} \leq d(v, w) \leq \rho.$$

Furthermore, from the inequality $a^2 + b^2 - 2ab \leq 2(a^2 + b^2)$, for all $a, b \in \mathbb{R}$, and the definition of the semimetric d we easily see

$$\begin{aligned}\left(\sum_{i=1}^{M}(v(i)-w(i))^4\right)^{1/2} &= \left(\sum_{i=1}^{M}(v(i)-w(i))^2\left(v^2(i)+w^2(i)-2v(i)w(i)\right)\right)^{1/2}\\ &\leq \left(2\sum_{i=1}^{M}(v(i)-w(i))^2\left(v^2(i)+w^2(i)\right)\right)^{1/2}\\ &= \sqrt{2}d(v,w) \leq \sqrt{2}\rho.\end{aligned}$$

These inequalities also hold for any $v \in \operatorname{conv} B_\rho(w)$. To see this, let $n \in \mathbb{N}$, $\alpha_1, ..., \alpha_n \in [0,1]$ with $\sum_{i=1}^{n} \alpha_i = 1$ and $v_1, ..., v_n \in B_\rho(w)$, such that

$$v = \sum_{i=1}^{n} \alpha_i v_i.$$

From the convexity of the function $f(x) := x^{2p}$ on $\mathbb{R}$ with $p \in \mathbb{N}$, i.e., $f(\sum \alpha_i x_i) \leq \sum \alpha_i f(x_i)$ for any finite convex combination of real numbers x_i, we can derive

$$\begin{aligned}\left(\sum_{i=1}^{n}\alpha_i v_i(j) - w(j)\right)^{2p} &= \left(\sum_{i=1}^{n}\alpha_i v_i(j) - \sum_{i=1}^{n}\alpha_i w(j)\right)^{2p}\\ &= \left(\sum_{i=1}^{n}\alpha_i (v_i(j) - w(j))\right)^{2p}\\ &\overset{convex.}{\leq} \sum_{i=1}^{n}\alpha_i (v_i(j) - w(j))^{2p}.\end{aligned}$$

We consider the case $p = 2$ to get

$$\begin{aligned}\left(\sum_{i=1}^{M}(v(i)-w(i))^4\right)^{1/2} &= \left(\sum_{i=1}^{M}\left(\sum_{j=1}^{n}\alpha_j v_j(i) - w(i)\right)^4\right)^{1/2}\\ &\leq \left(\sum_{j=1}^{n}\alpha_j\left(\sum_{i=1}^{M}(v_j(i)-w(i))^4\right)\right)^{1/2}\\ &\leq \left(2\rho^2\sum_{j=1}^{n}\alpha_j\right) = \sqrt{2}\rho\end{aligned}$$

and analogously, from the case $p = 1$,

$$\left(\sum_{i=1}^{M} (v(i) - w(i))^2 w^2(i) \right)^{1/2} \leq \rho.$$

Finally, we get from the elementary inequality $a^2 + b^2 \leq 4a^2 + 2(a-b)^2$, for all $a, b \in \mathbb{R}$, that

$$\begin{aligned} d(v, w) &= \left(\sum_{i=1}^{M} (v(i) - w(i))^2 (v(i)^2 + w(i)^2) \right)^{1/2} \\ &\leq \left(\sum_{i=1}^{M} (v(i) - w(i))^2 (4w^2(i) + 2(v(i) - w(i))^2) \right)^{1/2} \\ &\leq 2 \left(\sum_{i=1}^{M} (v(i) - w(i))^2 w_i^2(i) \right)^{1/2} + \sqrt{2} \left(\sum_{i=1}^{M} (v(i) - w(i))^4 \right)^{1/2} \\ &\leq 4\rho, \end{aligned}$$

for every $v \in \operatorname{conv} B_\rho(w)$. □

Lemma 5.7.
Let W be an n-dimensional subspace of $\mathbb{R}^M$. Let $\varepsilon_1, ..., \varepsilon_M$ be independent random variables having values in $\{-1, 1\}$ with probability $1/2$ each. Let $P_W : \mathbb{R}^M \to \mathbb{R}^M$ be the orthogonal projection onto W. Then

$$\mathbb{E} \sup_{w \in W \cap B_2^M} \left| \sum_{i=1}^{M} \varepsilon_i w^2(i) \right| \leq C \sqrt{\log M} \cdot \left\| P_W : \ell_1^M \to \ell_2^M \right\|.$$

Proof. Denote $W_1 = B_2^M \cap W$. We have to estimate the expectation of the supremum over all $w \in W_1$ of a random process

$$V_w = \sum_{i=1}^{M} \varepsilon_i w^2(i).$$

Let $w, \bar{w} \in W_1$. Then we see that $V_w - V_{\bar{w}}$ is a symmetric random variable, since V_w and $V_{\bar{w}}$ are symmetric. Consequently, we get

$$\mathbb{P}(V_w - V_{\bar{w}} > a) = \mathbb{P}(V_{\bar{w}} - V_w > a) = \frac{1}{2} \cdot \mathbb{P}(|V_w - V_{\bar{w}}| > a).$$

From the classical subgaussian tail estimate for Rademacher random variables (see Appendix C, Theorem A.42) we see that the process V_w has a subgaussian tail estimate

$$\mathbb{P}(V_w - V_{\bar{w}} > a) = \frac{1}{2}\mathbb{P}(|V_w - V_{\bar{w}}| > a) \leq \exp\left(-c\frac{a^2}{\tilde{d}^2(w,\bar{w})}\right),$$

with the metric

$$\tilde{d}^2(w,\bar{w}) = \left(\sum_{i=1}^{M}\left(w^2(i) - \bar{w}^2(i)\right)^2\right)^{1/2}.$$

We estimate the metric $\tilde{d}$ by the simpler to control semimetric defined in Lemma 5.6:

$$\frac{1}{\sqrt{2}}\tilde{d}(w,\bar{w}) \leq d(w,\bar{w}) = \left(\sum_{i=1}^{M}(w(i) - \bar{w}(i))^2(w^2(i) + \bar{w}^2(i))\right)^{1/2},$$

which follows from the simple inequality $(w(i) + \bar{w}(i))^2 \leq 2(w^2(i) + \bar{w}^2(i))$, for all $i = 1, ..., M$.
We already saw in the proof of Lemma 5.6 that the triangle inequality does not hold for d. However, a generalized triangle inequality for d can be derived from

$$\begin{aligned} d(w,\bar{w}) &= \left(\sum_{i=1}^{M}\frac{1}{2}(w(i) - \bar{w}(i))^2\left((w(i) + \bar{w}(i))^2 + (w(i) - \bar{w}(i))^2\right)\right)^{1/2} \\ &\leq \frac{1}{\sqrt{2}}\left(\tilde{d}(w,\bar{w}) + \|w - \bar{w}\|_{\ell_4^M}^2\right) \\ &\leq \sqrt{2}\tilde{d}(w,\bar{w}). \end{aligned}$$

Namely, for all $u, w, \bar{w} \in W$ we have

$$d(w,\bar{w}) \leq \sqrt{2}\tilde{d}(w,\bar{w}) \leq \sqrt{2}(\tilde{d}(w,u) + \tilde{d}(u,\bar{w})) \leq 2(d(w,u) + d(u,\bar{w})). \quad (5.3)$$

Note that balls in the semimetric d are not convex. This problem will later be overcome by Lemma 5.6.

Let now $r \in \mathbb{N}$ to be chosen later. Let k_0 and k_1 be the largest natural numbers such that

$$r^{-k_0} \geq \operatorname{diam}(W_1, \|\cdot\|_\infty) =: Q \quad \text{and} \quad r^{-k_1} \geq \frac{Q}{\sqrt{n}}.$$

This is equivalent to $-k_0 \geq (\log r)^{-1}Q$ and $k_1 \leq -(\log Q - \frac{1}{2}\log n)(\log r)^{-1}$, from which

$$k_1 - k_0 \leq (2\log r)^{-1}\log n \tag{5.4}$$

follows. Additionally, define functions $\varphi_k : W_1 \to \mathbb{R}$ by

$$\varphi_k(w) = \begin{cases} \min\{\|u\|^2 \mid u \in \operatorname{conv} B_{2r^{-k}}(w)\} + \frac{k-k_0}{\log M}, & \text{if } k_0 \leq k \leq k_1 \\ 1 + \frac{1}{2\log r} + \sum_{\ell=k_1}^{k} r^{-l}\frac{\sqrt{n\log(1+2\sqrt{2}r^l)}}{Q\sqrt{\log M}}, & \text{if } k > k_1. \end{cases}$$

We see, for any $w \in W_1$, that the sequence $\{\varphi_k(w)\}_{k=k_0}^{\infty}$ is nonnegative and nondecreasing. Moreover it is bounded by an absolute constant depending only on r, since for $k \leq k_1$ we have by (5.4)

$$\varphi_k(w) \leq 1 + \frac{1}{2\log r}\cdot\frac{\log n}{\log M}, \tag{5.5}$$

and for $k > k_1$ we have

$$\begin{aligned} \varphi_k(w) &\leq 1 + \frac{1}{2\log r} + \sum_{\ell=k_1}^{\infty} r^{-l}\frac{\sqrt{n\log(1+2\sqrt{2}r^l)}}{Q\sqrt{\log M}} \\ &\leq 1 + \frac{1}{2\log r} + c(r)r^{-k_1}\cdot\frac{\sqrt{n}}{Q}\cdot\frac{\sqrt{\log(1+2\sqrt{2}r^{k_1})}}{\sqrt{\log M}} \\ &\leq C(r), \end{aligned}$$

where the second to last inequality follows from the fact that this infinite series is a geometric series.

In order to prove this lemma we have to show that the condition (4.6) holds for $\{\varphi_k(w)\}_{k=k_0}^{\infty}$ with $\sigma = (c \cdot Q \cdot \sqrt{\log M})^{-1}$.
So, let $x \in W_1$ and suppose that the points $x_1, ..., x_N \in B_{r^{-k}}(x)$ satisfy $d(x_j, x_\ell) \geq 6r^{-k-1}$ for all $j \neq \ell$.

For $k \geq k_1 - 1$ condition (4.6) follows from the simple volume estimate

$$\begin{aligned} N &\leq N\left(W_1, d, 6r^{-k-1}\right) \leq N\left(W_1, d, r^{-k-1}\right) \\ &\leq N\left(W_1, \|\cdot\|_\infty, \frac{r^{-k-1}}{\sqrt{2}}\right) \overset{B_\infty^n \supseteq B_2^n}{\leq} N\left(W_1, \|\cdot\|, \frac{r^{-k-1}}{\sqrt{2}}\right) \\ &\leq \left(1 + \frac{2\sqrt{2}}{r^{-k-1}}\right)^n, \end{aligned}$$

where the last inequality follows for sufficiently large r, which, as we will see in the case $k_0 \leq k < k_1 - 1$, will surely be fulfilled because we have to make the choice $r \geq 16$.
To see that this volume estimate implies (4.6), we note first that by the independence of $\varphi_{k+2}(x_j)$ from x_j, for all $j = 1, ..., N$, we have

$$\max_{j=1,\dots,N} \varphi_{k+2}(x_j) = 1 + \frac{1}{2\log r} + \sum_{\ell=k_1}^{k+2} r^{-l} \frac{\sqrt{n\log(1+2\sqrt{2}r^l)}}{Q\sqrt{\log M}}.$$

In the case that $k_1 + 2 \geq k \geq k_1 - 1$ we get by (5.5) that

$$\begin{aligned}
&\max_{j=1,\dots,N} \varphi_{k+2}(x_j) \\
&\geq 1 + \frac{1}{2\log r}\frac{\log n}{\log M} + (Q\sqrt{\log M})^{-1} r^{-k} \sum_{\ell=k_1}^{k+2} r^{k-l}\sqrt{n\log(1+2\sqrt{2}r^l)} \\
&\geq \varphi_k(x) + (Q\sqrt{\log M})^{-1} r^{-k}\sqrt{\log N}\sum_{\ell=k_1}^{k+2} r^{k-l}.
\end{aligned}$$

For the case $k > k_1 + 2$ we have

$$\begin{aligned}
&\max_{j=1,\dots,N} \varphi_{k+2}(x_j) \\
&= 1 + \frac{1}{2\log r} + \sum_{\ell=k_1}^{k} r^{-l}\frac{\sqrt{n\log(1+2\sqrt{2}r^l)}}{Q\sqrt{\log M}} + \sum_{\ell=k+1}^{k+2} r^{-l}\frac{\sqrt{n\log(1+2\sqrt{2}r^l)}}{Q\sqrt{\log M}} \\
&= \varphi_k(x) + \sum_{\ell=k+1}^{k+2} r^{-l}\frac{\sqrt{n\log(1+2\sqrt{2}r^l)}}{Q\sqrt{\log M}} \\
&\geq \varphi_k(x) + (Q\sqrt{\log M})^{-1} r^{-k}\sqrt{\log N}\sum_{\ell=k+1}^{k+2} r^{k-l}.
\end{aligned}$$

Choosing now $c^{-1} := \sum_{\ell=k+1}^{k+2} r^{k-l}$ we see that

$$\max_{j=1,\dots,N} \varphi_{k+2}(x_j) \geq \varphi_k(x) + (cQ\sqrt{\log M})^{-1} r^{-k}\sqrt{\log N}$$

and condition (4.6) holds for $k \geq k_1 - 1$.

Suppose now that $k_0 \leq k < k_1 - 1$. For $j = 1, ..., N$, denote by z_j the point of $\operatorname{conv} B_{3r^{-k-2}}(x_j)$ for which the minimum of $\|z\|$ is attained, and denote by u the similar point of $\operatorname{conv} B_{3r^{-k}}(x)$. By (5.3) and Lemma 5.6 we have for all $j \neq \ell$

$$\begin{aligned} 6r^{-k-1} &\leq d(x_j, x_\ell) \leq 2(d(x_j, z_j) + d(z_j, x_\ell)) \\ &\leq 2(d(x_j, z_j) + 2(d(z_j, z_\ell) + d(z_\ell, x_\ell))) \\ &\leq 4 \cdot (d(x_j, z_j) + d(z_j, z_\ell) + d(z_\ell, x_\ell)) \\ &\leq 4 \cdot \left(16 \cdot r^{-k-2} + d(z_j, z_\ell)\right), \end{aligned}$$

so, $d(z_j, z_\ell) \geq \frac{1}{2} r^{-k-1}$ if $r \geq 16$.
Under the same assumptions on r we also have

$$\begin{aligned} d(z_j, x) &\leq 2\left(d(z_j, x_j) + d(x_j, x)\right) \\ &\leq 2(8r^{-k-2} + r^{-k}) \\ &\leq \left(\frac{1}{16 \cdot 256} + 2\right) r^{-k} \leq 3r^{-k}. \end{aligned}$$

Denote $\theta := \max_{j=1,...,N} \|z_j\|^2 - \|u\|$. We have to prove that

$$r^{-k} \cdot (c \cdot Q \cdot \sqrt{\log M})^{-1} \cdot \sqrt{\log N} \leq \max_{j=1,...,N} \varphi_{k+2}(x_j) - \varphi_k(x) = \theta + \frac{2}{\log M}. \quad (5.6)$$

From $\frac{z_j+u}{2} \in \operatorname{conv} B_{3r^{-k}}(x)$ we know, $\|\frac{z_j+u}{2}\| \geq \|u\|$. Additionally, we have by definition of u and z_j that $\|u\| \leq \|z_j\|$ for all $j = 1, ..., N$. Using these, together with the parallelogram law $\|z_j + u\|^2 + \|z_j - u\|^2 = 2(\|z_j\|^2 + \|u\|^2)$, we get

$$\begin{aligned} \left\|\frac{z_j - u}{2}\right\|^2 &= \frac{1}{2}\|z_j\|^2 + \frac{1}{2}\|u\|^2 - \left\|\frac{z_j + u}{2}\right\|^2 \\ &\leq \|z_j\|^2 - \left\|\frac{z_j + u}{2}\right\|^2 \\ &\leq \|z_j\|^2 - \|u\|^2 \\ &\leq \max_{j=1,...,N} \|z_j\|^2 - \|u\|^2 \\ &= \theta, \end{aligned}$$

and, therefore,

$$\|z_j - u\| \leq 2\sqrt{\theta}. \quad (5.7)$$

Thus, $z_1, ..., z_N$ is contained in $u + 2\sqrt{\theta} B_2^M \cap W$. From $d(x_j, x_l) \geq 6r^{-k-1}$, for all $j \neq l$, we see that the $\frac{1}{2} r^{-k-1}$-entropy of the set $K = u + 2\sqrt{\theta} B_2^M \cap W$

in the quasimetric d, namely, the number of $\frac{1}{2}r^{-k-1}$-balls in the quasimetric d needed to cover K, gives us an upper bound for N.
To estimate this entropy we partition the set K into S disjoint subsets having diameter less than $\frac{1}{16}r^{-k-1}\theta^{-1/2}$ in the ℓ_∞ metric. By part (i) of Lemma 5.5 we may assume that

$$\frac{1}{16}r^{-k-1}\theta^{-1/2}\sqrt{\log S} \leq c \cdot Q \cdot \sqrt{\theta} \cdot \sqrt{\log M}. \tag{5.8}$$

In the first case, $S \geq \sqrt{N}$, we can rewrite this as $\sqrt{2}\sqrt{\log S} \geq \sqrt{\log N}$ and rearrange (5.8) in the form

$$\frac{1}{16}r^{-k-1}(c \cdot Q \cdot \sqrt{\log M})^{-1}\sqrt{2}\sqrt{\log S} \leq \theta.$$

So we get (5.6), since

$$\frac{1}{16\sqrt{2}}r^{-k-1}\cdot(c\cdot Q\cdot\sqrt{\log M})^{-1}\cdot\sqrt{\log N} \leq \frac{1}{16}r^{-k-1}\cdot(c\cdot Q\cdot\sqrt{\log M})^{-1}\cdot\sqrt{\log S} \leq \theta.$$

Pulling $\frac{1}{16\sqrt{2}}r^{-1}$ into the constant c we have

$$r^{-k} \cdot (c \cdot Q \cdot \sqrt{\log M})^{-1} \cdot \sqrt{\log N} \leq \theta \leq \theta + \frac{2}{\log M},$$

and, therefore, shown (5.6) for this case.

Suppose now, as second case, that $S \leq \sqrt{N}$. We see that there exists an element of the partition containing at least $\sqrt{N}$ points z_j. Let $J \subset \{1, ..., N\}$ be the set of the indices of these points. We have

$$\|z_j - z_\ell\|_\infty \leq \frac{1}{16}r^{-k-1}\theta^{-1/2} \tag{5.9}$$

for all $j, \ell \in J$, $j \neq \ell$.
Using the estimate $d(z_j, z_\ell) \geq \frac{1}{2}r^{-k-1}$ from before, and temporarily using the short hand notations $Z_j := \{i \ : \ |z_j(i)| \geq 2|u(i)|\}$ and $Z_\ell := \{i \ : \ |z_\ell(i)| \geq 2|u(i)|\}$, we have

$$\begin{aligned}\left(\frac{1}{2}r^{-k-1}\right)^2 &\leq \sum_{i=1}^{M} (z_j(i) - z_\ell(i))^2 \cdot \left(z_j^2(i) + z_\ell^2(i)\right) \\ &\leq \sum_{i=1}^{M} (z_j(i) - z_\ell(i))^2 \cdot \left(8u^2(i) + z_j^2(i)\mathbb{1}_{Z_j}(i) + z_\ell^2(i)\mathbb{1}_{Z_\ell\}}(i)\right).\end{aligned} \tag{5.10}$$

Then, (5.7) implies

$$\sum_{i=1}^{M} z_j^2(i)\mathbb{1}_{Z_j}(i) \leq 4\sum_{i=1}^{M}(z_j(i) - u(i))^2 \leq 16 \cdot \theta. \tag{5.11}$$

Combining (5.9) and (5.11) we get that (5.10) is bounded by

$$\left(\frac{1}{2}r^{-k-1}\right)^2 \leq \sum_{i=1}^{M} (z_j(i) - z_\ell(i))^2 \cdot \left(8u^2(i) + z_j^2(i)\mathbb{1}_{Z_j}(i) + z_\ell^2(i)\mathbb{1}_{Z_\ell}(i)\right)$$
$$\leq 8\sum_{i=1}^{M} (z_j(i) - z_\ell(i))^2 u^2(i) + \|z_j - z_l\|_\infty^2 \cdot \left(\sum_{i=1}^{M} z_j^2(i)\mathbb{1}_{Z_j}(i) + \sum_{i=1}^{M} z_\ell^2(i)\mathbb{1}_{Z_\ell}(i)\right)$$
$$\leq 8\sum_{i=1}^{M} (z_j(i) - z_\ell(i))^2 u^2(i) + \left(\frac{\theta^{-1/2}}{16}r^{-k-1}\right)^2 \cdot 2 \cdot 16\theta$$

Thus, by rearranging this inequality, we have for all $j, \ell \in J$, $j \neq \ell$

$$\left(\sum_{i=1}^{M} (z_j(i) - z_\ell(i))^2 \cdot u^2(i)\right)^{1/2} \geq \frac{1}{8}r^{-k-1}.$$

Then, part (ii) of Lemma 5.5 implies

$$\frac{1}{8}r^{-k-1}\sqrt{\log |J|} \leq c \cdot \sqrt{\theta} \cdot Q \cdot \left(\sum_{i=1}^{M} u^2(i)\right)^{1/2} \leq c \cdot \sqrt{\theta} \cdot Q.$$

Since, for all $\theta > 0$,

$$2\sqrt{\theta} \leq \sqrt{\log M} \cdot \theta + \frac{1}{\sqrt{\log M}}$$

holds, and we know from $|J| \geq \sqrt{N}$ that $\sqrt{\log |J|} \geq \sqrt{\log N}/\sqrt{2} \geq \sqrt{\log N}/2$, we get

$$\frac{1}{16}r^{-k-1}\sqrt{\log N} \leq \frac{1}{8}r^{-k-1}\sqrt{\log |J|} \leq c \cdot Q \cdot \sqrt{\log M} \cdot \left(\theta + \frac{1}{\log M}\right).$$

Pulling again $\frac{1}{16}r^{-1}$ into the constant c and rearranging this inequality, we obtain

$$r^{-k}(c \cdot Q \cdot \sqrt{\log M})^{-1}\sqrt{\log N} \leq \theta + \frac{1}{\log M} \leq \theta + \frac{2}{\log M}.$$

Therefore, we have also shown (5.6) for the second case.
Altogether, we have proven that condition (4.6) holds for $\{\varphi_k(w)\}_{k=k_0}^{\infty}$ with $\sigma = (c \cdot Q \cdot \sqrt{\log M})^{-1}$ and that we can finally apply Theorem 4.6 to get

$$\mathbb{E} \sup_{w \in W \cap B_2^M} \left| \sum_{i=1}^{M} \varepsilon_i w^2(i) \right| \leq C \sqrt{\log M} \left\| P_W : \ell_1^M \to \ell_2^M \right\|.$$

□

The next theorem was proposed by B. Kashin and L. Tsafriri [Kas-Tz]. As already mentioned, this theorem will imply Theorem 5.3.

Theorem 5.8.
Let $A = (a_{i,j})$ be an $n \times M$ matrix with orthonormal rows and let $t \geq 1$. Let the inequality

$$\sqrt{\frac{M}{n}} \left(\sum_{i=1}^{n} a_{i,j}^2 \right)^{1/2} \leq t$$

hold for all $j = 1, ..., M$.
Then, for every $\varepsilon > 0$, there exists a set $I \subset \{1, ..., M\}$ such that

$$|I| \leq C \frac{t^2}{\varepsilon^2} n \log \frac{nt^2}{\varepsilon^2} \tag{5.12}$$

and for all $x \in \mathbb{R}^n$

$$(1 - \varepsilon) \cdot \|x\| \leq \sqrt{\frac{M}{|I|}} \cdot \left\| R_I A^T x \right\| \leq (1 + \varepsilon) \cdot \|x\|, \tag{5.13}$$

where $R_I : \mathbb{R}^M \to \mathbb{R}^M$ denotes the orthogonal projection onto the space span $\{e_i : i \in I\}$.

Proof. Let $M \geq C \frac{t^2}{\varepsilon^2} n \log n$ for some absolute constant C. Let $A = (a_{i,j})$ be an $n \times M$ matrix for which

$$\sqrt{\frac{M}{n}} \left(\sum_{i=1}^{n} a_{i,j}^2 \right)^{1/2} \leq t$$

holds for all $j = 1, ..., M$. Define a sequence $\{\varepsilon_i\}_{i=1}^M$ of Rademacher random variables and set $I_1 = \{i : \varepsilon_i = 1\}$. From $0 \leq |I_1| \leq M$ we can deduce that

$$\mathbb{P}\left(|I_1| \leq \frac{M}{2} \right) \geq \frac{1}{2} \text{ and } \mathbb{P}\left(\frac{M}{2}\left(1 - \frac{1}{\sqrt{M}}\right) \leq |I_1| \right) \geq \mathbb{P}\left(\frac{M}{2} \leq |I_1| \right) \geq \frac{1}{2},$$

which gives us with probability at least 1/4 that

$$\frac{M}{2}\left(1-\frac{1}{\sqrt{M}}\right) \leq |I_1| \leq \frac{M}{2}. \tag{5.14}$$

Let $w(1), ..., w(M)$ be the coordinates of a vector $w \in W := A^T\mathbb{R}^n$. Note that from the fact that $W \cap B_2^M$ is the euclidean unit ball in W, we know that $\sup\limits_{w\in W\cap B_2^M} \|w\| = \sup\limits_{x\in B_2^n} \|x\|$, from which follows that

$$\sup_{x\in B_2^n} \left|2\|R_{I_1}A^Tx\|^2 - \|x\|^2\right| = \sup_{w\in W\cap B_2^M} \left|2\|R_{I_1}w\|^2 - \|w\|^2\right|$$

$$= \sup_{w\in W\cap B_2^M} \left|2\sum_{i\in I_1} w^2(i) - \sum_{i=1}^{M} w^2(i)\right| = \sup_{w\in W\cap B_2^M} \left|\sum_{i=1}^{M} \varepsilon_i w^2(i)\right|.$$

Note that $w_i := A^Te_i$, $i = 1, ..., n$, is a basis of W, since the columns of A^T are orthonormal. From this, together with the inequality $\sqrt{\frac{M}{n}}\left(\sum_{i=1}^n a_{i,j}^2\right)^{1/2} \leq t$, we get

$$\|P_W : \ell_1^M \to \ell_2^M\| = \sup_{x\in B_1^M} \|P_W x\| = \sup_{x\in B_1^M} \|\sum_{i=1}^{n}\langle x, w_i\rangle w_i\|.$$

Since the supremum is attained in an extreme point of B_1^M, there exists a $j \in \{1, ..., n\}$ such that

$$\begin{aligned}
&\sup_{x\in B_1^M} \|\sum_{i=1}^{n}\langle x, w_i\rangle w_i\| = \|\sum_{i=1}^{n} a_{j,i}w_i\| \\
&= \|A^T(a_{i,j})_{i=1}^n\| \leq \|A^T\|_2 \cdot \|(a_{i,j})_{i=1}^n\|_{\ell_2^n} \\
&= \|(a_{i,j})_{i=1}^n\|_{\ell_2^n} = \left(\sum_{i=1}^{n} a_{i,j}^2\right)^{1/2} \\
&\leq t\sqrt{\frac{n}{M}}.
\end{aligned}$$

In the last, line we used $\|A^T\|_2 = \sqrt{\lambda_{\max}(AA^T)} = \sqrt{\lambda_{\max}(I_n)} = 1$.

By Markov's inequality we have

$$\mathbb{P}\left(\sup_{w\in W\cap B_2^M} \left|\sum_{i=1}^{M} \varepsilon_i w^2(i)\right| \geq 4\mathbb{E}\sup_{w\in W\cap B_2^M} \left|\sum_{i=1}^{M} \varepsilon_i w^2(i)\right|\right) \leq \frac{1}{4}$$

and consequently, with probability at least 3/4, that

$$\sup_{w\in W\cap B_2^M}\left|\sum_{i=1}^M \varepsilon_i w^2(i)\right| \le 4\mathbb{E}\sup_{w\in W\cap B_2^M}\left|\sum_{i=1}^M \varepsilon_i w^2(i)\right|.$$

From this, together with Lemma 5.7, we have with probability greater than 3/4.

$$\sup_{x\in B_2^n}\left|2\|R_{I_1}A^T x\|^2-\|x\|^2\right| = \sup_{w\in W\cap B_2^M}\left|\sum_{i=1}^M \varepsilon_i w^2(i)\right| \le Ct\cdot\sqrt{\frac{n}{M}}\cdot\sqrt{\log M} \tag{5.15}$$

Hence, there exists a set $I_1\subset\{1,...,M\}$ satisfying (5.14) and (5.15).

Repeating these steps inductively, we obtain a decreasing sequence of sets $\{1,...,M\} =: I_0\supset I_1\supset ...\supset I_s$ such that with probability greater 1/4, analogously as in (5.14),

$$\frac{|I_k|}{2}\left(1-\frac{1}{\sqrt{|I_k|}}\right)\le |I_{k+1}|\le\frac{|I_k|}{2}. \tag{5.16}$$

Analogously, as in (5.15), we get by Markov's inequality,

$$\mathbb{P}\left(\sup_{w\in W_k\cap B_2^m}\left|\sum_{i=1}^m \varepsilon_i w^2(i)\right|\ge 4\mathbb{E}\sup_{w\in W_k\cap B_2^m}\left|\sum_{i=1}^m \varepsilon_i w^2(i)\right|\right)\le\frac{1}{4},$$

that with probability at least 3/4

$$\sup_{w\in W_k\cap B_2^m}\left|\sum_{i=1}^m \varepsilon_i w^2(i)\right| \le 4\mathbb{E}\sup_{w\in W_k\cap B_2^m}\left|\sum_{i=1}^m \varepsilon_i w^2(i)\right|.$$

Furthermore, we have at each step of induction

$$\frac{1}{2}\cdot\|x\|_{\ell_2^n}\le 2^{\frac{k-1}{2}}\cdot\|R_{I_{k-1}}A^T x\|\le\frac{3}{2}\cdot\|x\|_{\ell_2^n}. \tag{5.17}$$

W.l.o.g. we can assume that $I_{k-1}=\{1,...,m\}$ for some $m<M$. Let $W_k=R_{I_{k-1}}A^T\mathbb{R}^n\subset\mathbb{R}^m$ and let $P_{W_k}:\mathbb{R}^m\to\mathbb{R}^m$ be the orthogonal projection onto W_k. Then, by (5.17),

$$2^{\frac{k-1}{2}}R_{I_{k-1}}A^T B_2^n\subset\frac{3}{2}B_2^m\cap W_k$$

holds and we have for a random set $I_k \subset I_{k-1}$,

$$\begin{aligned}
&\mathbb{E} \sup_{x \in B_2^n} \left(2^k \|R_{I_k} A^T x\|^2 - 2^{k-1} \|R_{I_{k-1}} A^T x\|^2\right) \\
&\leq \mathbb{E} \sup_{w \in W_k \cap \frac{3}{2} B_2^m} \left| 2 \sum_{i \in I_k} w_i^2 - \sum_{i \in I_{k-1}} w_i^2 \right| \\
&= \mathbb{E} \sup_{w \in W_k \cap \frac{3}{2} B_2^m} \left| \sum_{i=1}^m \varepsilon_i w_i^2 \right| \\
&\leq \frac{9}{4} \cdot \mathbb{E} \sup_{w \in W_k \cap B_2^m} \left| \sum_{i=1}^m \varepsilon_i w_i^2 \right| .
\end{aligned}$$

We also need to estimate $\|P_{W_k} : \ell_1^m \to \ell_2^m\|$, which is possible by (5.17) and gives

$$\|P_{W_k} : \ell_1^m \to \ell_2^m\| \leq 2 \cdot \left\| \left(2^{\frac{k-1}{2}} R_{I_{k-1}} A^T\right)^T : \ell_1^M \to \ell_2^n \right\| \leq 2^{\frac{k+1}{2}} \cdot t \cdot \sqrt{\frac{n}{M}}.$$

Again, by Lemma 5.7 we get with probability greater than 3/4

$$\sup_{x \in B_2^n} \left(2^k \|R_{I_k} A^T x\|^2 - 2^{k-1} \|R_{I_{k-1}} A^T x\|^2\right) \leq Ct \sqrt{\frac{n}{M/2^k}} \cdot \sqrt{\log |I_{k-1}|}.$$

Note that we have absorbed a factor $\sqrt{2}$ from $\|P_{W_k} : \ell_1^m \to \ell_2^m\| \leq 2^{\frac{k+1}{2}} \cdot t \cdot \sqrt{\frac{n}{M}}$ into the constant C.

From these inequalities and (5.15), we obtain by the triangle inequality, together with $|I_k| \leq M/2^k$, that

$$\begin{aligned}
&\sup_{x \in B_2^n} \left|2^s \|R_{I_s} A^T x\|^2 - \|x\|^2\right| \\
&\leq \sup_{x \in B_2^n} \left|2 \|R_{I_1} A^T x\|^2 - \|x\|^2\right| + \sum_{k=2}^s \sup_{x \in B_2^n} \left|2^k \|R_{I_k} A^T x\|^2 - 2^{k-1} \|R_{I_{k-1}} A^T x\|^2\right| \\
&\leq Ct \left(\sqrt{\frac{n}{M}} \cdot \sqrt{\log M} + \sum_{k=2}^s \sqrt{\frac{n}{M/2^k}} \cdot \sqrt{\log |I_{k-1}|} \right) \\
&\leq sCt \cdot \sqrt{\frac{n}{M/2^s}} \cdot \sqrt{\log |I_s|} \leq Ct \cdot \sqrt{\frac{n}{M/2^s}} \cdot \sqrt{\log \frac{M}{2^s}} .
\end{aligned} \tag{5.18}$$

We break this procedure if the last expression of the above chain of inequalities is smaller than $\varepsilon/2$ and get in this case that

$$c \cdot \frac{t^2}{\varepsilon^2} \cdot n \cdot \log \frac{nt^2}{\varepsilon^2} \leq \frac{M}{2^s} \leq C \cdot \frac{t^2}{\varepsilon^2} \cdot n \cdot \log \frac{nt^2}{\varepsilon^2}. \tag{5.19}$$

Indeed, under the assumption that $\log \frac{M}{2^s} \geq 1$ we have

$$\log \frac{nt^2}{\varepsilon^2} \leq \log \frac{Mt^2}{\varepsilon^2} \leq \log \frac{M}{2^s} + \log \frac{2^s t^2}{\varepsilon^2} \leq C \log \frac{M}{2^s},$$

which gives us, together with

$$Ct \cdot \sqrt{\frac{n}{M/2^s}} \cdot \sqrt{\log \frac{M}{2^s}} \leq \frac{\varepsilon}{2} \iff C \log \frac{nt^2}{\varepsilon^2} \log \frac{M}{2^s} \leq \frac{M}{2^s},$$

the lower bound of (5.19).
Assume now that there exists no such upper bound as in (5.19). Note that we still can find a bound with a constant $C(n, M)$ depending on n and M. Assume that $\log \frac{nt^2}{\varepsilon^2} \geq 1$, giving us

$$\log \left(c \frac{nt^2}{\varepsilon^2} \log \frac{nt^2}{\varepsilon^2} \right) \geq \log c + \log \frac{nt^2}{\varepsilon^2} \geq c \log \frac{nt^2}{\varepsilon^2}.$$

Then we have

$$\begin{aligned} \frac{\varepsilon}{2} &\geq Ct \cdot \sqrt{\frac{n}{M/2^s}} \cdot \sqrt{\log \frac{M}{2^s}} \\ &\geq Ct \cdot \sqrt{\frac{n}{C(n,M) \cdot \frac{t^2}{\varepsilon^2} \cdot n \cdot \log \frac{nt^2}{\varepsilon^2}}} \cdot \sqrt{c \log \frac{nt^2}{\varepsilon^2}} \\ &\geq C(n,M)\varepsilon. \end{aligned}$$

This tells us that $\frac{1}{2} \geq C(n, M)$, contradicting our assumption that we can't find an upper bound of (5.19) independent of n and M.

With (5.14) and (5.16) we can deduce that

$$\frac{M}{2^s} \prod_{k=0}^{s-1} \left(1 - \frac{1}{\sqrt{|I_s|}} \right) \leq |I_s| \leq \frac{M}{2^s}.$$

This term can be estimated from below by the Weierstrass product inequality, i.e., $\prod_{i=1}^{n}(1-u_i) \geq 1 - \sum_{i=1}^{n} u_i$ for $u_i \in [0,1]$, $i = 1, ..., n$, yielding

$$\begin{aligned}\prod_{k=0}^{s-1}\left(1-\frac{1}{\sqrt{|I_s|}}\right) &\geq 1-\sum_{k=0}^{s-1}\frac{1}{\sqrt{|I_k|}} \overset{|I_{k-1}|\leq|I_k|/2}{\geq} 1-\sum_{k=0}^{s-1}\frac{1}{\sqrt{2^{s-k}|I_s|}} \\ &= 1-\frac{1}{|I_s|}\sum_{k=0}^{s-1}\frac{1}{\sqrt{2^{s-k}}} = 1-\frac{1}{\sqrt{|I_s|}}\frac{(1+\sqrt{2})(\sqrt{2^s}-1)}{\sqrt{2^s}} \\ &\geq 1-\frac{4}{\sqrt{|I_s|}}\end{aligned}$$

and, subsequently,

$$\frac{M}{2^s}\left(1-\frac{4}{\sqrt{|I_s|}}\right) \leq |I_s| \leq \frac{M}{2^s}, \tag{5.20}$$

which tells us, together with (5.19), that the set $|I_s|$ fulfills (5.12).
From the lower bound, together with (5.19) and the fact that $nt^2/\varepsilon^2 \geq c \cdot n$, one sees that

$$|I_s| \geq c\cdot\frac{t^2}{\varepsilon^2}\cdot n\cdot\log\frac{nt^2}{\varepsilon^2}\cdot\left(1-\frac{4}{\sqrt{|I_s|}}\right) \geq c\cdot\frac{t^2}{\varepsilon^2}\cdot n\cdot\log n.$$

Hence, we can reformulate (5.20) as

$$\frac{M}{|I_s|}\left(1-\left(c\cdot\frac{t^2}{\varepsilon^2}\cdot n\cdot\log n\right)^{-1/2}\right) \leq 2^s \leq \frac{M}{|I_s|}.$$

Thus, from (5.18) we have

$$\sup_{x\in B_2^n}\left|\frac{M}{|I_s|}\|R_{I_s}A^Tx\|^2 - \|x\|^2\right| \leq \varepsilon.$$

□

We go on now by using this theorem to show that for each convex body, with the euclidean unit ball as its John ellipsoid, there exists a decomposition of the identity operator in terms of the contact points of these two bodies. This is in general not a John's decomposition, but the number of contact points needed is bounded by $C(\varepsilon)M\log M$, where M denotes the dimension of the space.
Having shown this Lemma, we will use it to prove Theorem 5.11, which is an approximate version of the John's decomposition. By this approximate

John's decomposition it is possible to show Theorem 5.3. Namely, one can construct the convex body K required in Theorem 5.3.
Let's turn to the aforementioned lemma. We will make use of the notation of rank-1 orthogonal projections (appendix, Definition A.39 and the following remark). Its statement is the following:

Lemma 5.9.
Let $K \in \mathcal{K}^M$ with John ellipsoid B_2^M and let $\varepsilon > 0$. Then, there exist

$$m \leq C(\varepsilon) M \log M$$

contact points $x_1, ..., x_m$ and positive numbers $c_1, ..., c_m$ so that the identity operator in $\mathbb{R}^M$ has the decomposition

$$\mathrm{id}_M = \frac{N}{m} \sum_{i=1}^{m} c_i x_i \otimes x_i + S,$$

where $\|S : \ell_2^M \to \ell_2^M\| \leq \varepsilon$ and $N \in \mathbb{N}$ with $N \leq M(M+3)/2$.

Proof. Let $\varepsilon > 0$. Let $K \in \mathcal{K}^M$ with John ellipsoid B_2^M. Then, by John's theorem, there exist $N \leq M(M+3)/2$ contact points $x_1, ..., x_N$ and positive numbers $c_1, ..., c_N$ such that the identity operator in $\mathbb{R}^M$ can be decomposed as

$$\mathrm{id}_M = \sum_{i=1}^{N} c_i x_i \otimes x_i \tag{5.21}$$

and

$$\sum_{i=1}^{N} c_i x_i = 0 \tag{5.22}$$

holds.
Define the $M \times N$-matrix $A := (a_{i,j}) := (\sqrt{c_j} x_j(i))$. Due to (5.21), we have

$$\begin{aligned}
\sum_{k=1}^{M} x(k) e_k = x = \left(\sum_{i=1}^{N} c_i x_i \otimes x_i \right)(x) &= \sum_{i=1}^{N} c_i \langle x_i, x \rangle x_i \\
&= \sum_{i=1}^{N} c_i \left(\sum_{j=1}^{M} x_i(j) x(j) \right) \left(\sum_{k=1}^{M} x_i(k) e_k \right) \\
&= \sum_{j,k=1}^{M} \left(\sum_{i=1}^{N} \sqrt{c_i} x_i(j) \sqrt{c_i} x_i(k) \right) x(j) e_k.
\end{aligned}$$

Hence, $\sum_{i=1}^N \sqrt{c_i}x_i(j)\sqrt{c_i}x_i(k) = \delta_{j,k}$ and, thereby, that the rows of A are orthonormal. With $t := \sqrt{\frac{N}{M}} \max_{j=1,\dots,N} \sqrt{c_j}$ we have $\sqrt{\frac{N}{M}} \left(\sum_{i=1}^M a_{i,j}^2\right)^{1/2} \leq t$. Thus, by Theorem 5.7,

$$\sup_{x \in B_2^M} \left| \frac{N}{|I|} \|R_I A^T x\|^2 - \|x\|^2 \right| \leq \varepsilon$$

for a set $I \subset \{1, ..., N\}$ with $|I| \leq C(\varepsilon) M \log M$. Since

$$\sup_{x \in B_2^M} \left| \frac{N}{|I|} \|R_I A^T x\|^2 - \|x\|^2 \right| = \left\| \frac{N}{|I|} \sum_{i \in I} c_i x_i \otimes x_i - id_M \right\|,$$

we get by setting $m = |I|$ and renaming $(x_i)_{i=1,\dots N}$ so that $(x_i)_{i \in I}$ are the first m elements, that

$$\left\| \mathrm{id}_M - \frac{N}{m} \sum_{i=1}^m c_i x_i \otimes x_i \right\| \leq \varepsilon.$$

Taking the trace of this expression yields the statement of the theorem. □

Before we go on, we introduce another lemma which will be needed to show Theorem 5.11. We will only give the idea of proof and relegate the reader to Lemma 3.2 of [R2] for further details.
It allows us to estimate the expectation of the norm of a randomized sum of rank-1 orthogonal projections in terms of the norm of the sum off all rank-1 orthogonal projections which are partaking in the randomized sum. More precisely, we have

Lemma 5.10.
Let $y_1, ..., y_n \in \mathbb{R}^M$ and $\varepsilon_1, ..., \varepsilon_n$ be independent Bernoulli random variables. Then, for some constant C,

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i y_i \otimes y_i \right\| \leq C \log M \sqrt{\log n} \max_{i=1,\dots,n} \|y_i\| \left\| \sum_{i=1}^n y_i \otimes y_i \right\|^{1/2}.$$

Proof. (Idea)
By an estimate of Dudley [L-T], one can show

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i y_i \otimes y_i \right\| \leq C \int_0^\infty \left(\log N(B_2^M, \delta, u)\right)^{1/2} du,$$

with the metric $\delta(x,y) = \left(\sum_{i=1}^n \left(\langle x, y_i\rangle^2 - \langle y, y_i\rangle^2\right)^2\right)^{1/2}$.
One can estimate further with elementary arguments and show

$$N(B_2^M, \delta, u) \leq N\left(B_2^M, \|\cdot\|_Y, \frac{1}{\rho}u\right),$$

where $\|\cdot\|_Y := \sup_{i=1,\dots,n} |\langle \cdot, y_i\rangle|$ and $\rho = 2\,||\sum_{i=1}^n y_i \otimes y_i||^{1/2}$, and, furthermore,

$$\mathbb{E}\left|\left|\sum_{i=1}^n \varepsilon_i y_i \otimes y_i\right|\right| \leq C\rho \int_0^\infty \left(\log N\left(B_2^M, \|\cdot\|_Y, v\right)\right)^{1/2} dv.$$

For $v > 1$ one has $N\left(B_2^M, \|\cdot\|_Y, v\right) = 1$. By a standard volume estimate

$$N\left(B_2^M, \|\cdot\|_Y, v\right) \leq N\left(B_2^M, \|\cdot\|, \frac{1}{\rho}u\right) \leq \left(1 + \frac{2}{v}\right)^M.$$

Additionally, by an inequality of Pajor and Tomczak-Jaegermann [P-T], one has

$$\left(\log N\left(B_2^M, \|\cdot\|_Y, v\right)\right)^{1/2} \leq \frac{c}{v}\mathbb{E}\|g\|_Y \leq Cn^{1/\log n}\sqrt{\log n},$$

where g denotes a standard gaussian vector in $\mathbb{R}^M$.
Combining these, one gets

$$\begin{aligned}
&\int_0^\infty \left(\log N\left(B_2^M, \|\cdot\|_Y, v\right)\right)^{1/2} dv \\
&\leq \int_0^{M^{-1/2}} \left(M \log\left(1 + \frac{2}{v}\right)\right)^{1/2} dv + \int_{M^{-1/2}}^1 C\sqrt{\log n}\frac{dv}{v} \\
&\leq \log(1 + 2\sqrt{M}) + C\sqrt{\log n}\log\sqrt{M}.
\end{aligned}$$

□

We now show the already mentioned theorem about the approximate John's decomposition. Note that the derivation of this theorem is necessary, since the vectors x_i and positive numbers c_i of the approximate decomposition of the identity operator id_M in $\mathbb{R}^M$ in Lemma 5.9 do not necessarily sum up to zero, i.e., $\sum_{i=1}^m c_i x_i \neq 0$.
However, in this theorem we construct a vector $u \in \mathbb{R}^M$ such that we get yet another approximate decomposition of the identity operator in $\mathbb{R}^M$ and, additionally, that the vectors constituting this decomposition add up to zero. This is a crucial condition for the construction of the body K in the proof of Theorem 5.3.

Theorem 5.11. (Approximate John's decomposition)
Let $K \in \mathcal{K}^M$ with John ellipsoid B_2^M and let $\varepsilon > 0$. Then there exist

$$m \leq C(\varepsilon) M \log M$$

contact points $x_1, ..., x_m$ and a vector u, $\|u\| \leq C(\varepsilon)(M \log M)^{-1/2}$, so that the identity operator id_M in $\mathbb{R}^M$ has the decomposition

$$\mathrm{id}_M = \frac{M}{m} \sum_{i=1}^{m} (x_i + u) \otimes (x_i + u) + S, \tag{5.23}$$

where

$$\sum_{i=1}^{m} (x_i + u) = 0 \tag{5.24}$$

and

$$\|S : \ell_2^M \to \ell_2^M\| < \varepsilon. \tag{5.25}$$

Proof. Let $\varepsilon > 0$. Let $K \in \mathcal{K}^M$ with John ellipsoid B_2^M. Then, by Lemma 5.9, there exist a decomposition of the identity operator id_M in $\mathbb{R}^M$ in terms of the contact points $\bar{x}_1, ..., \bar{x}_n$ of K and B_2^M and positive numbers $c_1, ..., c_n$, where $n \leq C(\varepsilon) M \log M$, such that

$$\mathrm{id}_M = \frac{N}{n} \sum_{i=1}^{n} c_i \bar{x}_i \otimes \bar{x}_i + S,$$

where $\|S\| \leq \frac{\varepsilon}{8}$ and $N \in \mathbb{N}$ with $N \leq \frac{M(M+3)}{2}$.
By defining $\bar{c}_i := \frac{N}{n} c_i$, we can write

$$\mathrm{id}_M = \sum_{i=1}^{n} \bar{c}_i \bar{x}_i \otimes \bar{x}_i + S.$$

Set now $L = \lfloor \frac{8Mn}{\varepsilon} \rfloor$ and, for $i = 1, ..., n$, $N_i = \lfloor \frac{\bar{c}_i L}{M} \rfloor$ or $N_i = \lfloor \frac{\bar{c}_i L}{M} \rfloor + 1$, such that we have $\sum_{i=1}^{n} N_i = L$. Furthermore, define the sequence $x_1, ..., x_L$ by N_i times repeating $\bar{x}_i$. Define

$$\bar{T}_0 := \frac{M}{L} \sum_{i=1}^{L} x_i \otimes x_i + S.$$

Then we have

$$\begin{aligned}
\|\operatorname{id}_M - \bar{T}_0\| &= \left\| \sum_{i=1}^{n} \left(\bar{c}_i - \frac{N_i M}{L} \right) \bar{x}_i \otimes \bar{x}_i + S \right\| \\
&\leq \left\| \sum_{i=1}^{n} \left(\bar{c}_i \frac{L}{M} - N_i \right) \frac{M}{L} \bar{x}_i \otimes \bar{x}_i + S \right\| \\
&\leq \frac{M}{L} \sum_{i=1}^{n} \left| \bar{c}_i \frac{L}{M} - N_i \right| \|\bar{x}_i \otimes \bar{x}_i\| + \|S\| \\
&\leq \frac{M}{L} \sum_{i=1}^{n} \|\bar{x}_i \otimes \bar{x}_i\| + \|S\| \\
&\leq \frac{Mn}{L} + \frac{\varepsilon}{8} \leq \frac{\varepsilon}{4}
\end{aligned}$$

and

$$\begin{aligned}
\left\| \frac{M}{L} \sum_{i=1}^{L} x_i \right\| &= \left\| \sum_{i=1}^{n} \left(\bar{c}_i - \frac{N_i M}{L} \right) \bar{x}_i \right\| \\
&\leq \frac{M}{L} \sum_{i=1}^{n} \left| \bar{c}_i \frac{L}{M} - N_i \right| \|\bar{x}_i\| \\
&\leq \frac{M}{L} \sum_{i=1}^{n} \|\bar{x}_i\| \leq \frac{Mn}{L} \leq \frac{\varepsilon}{8}.
\end{aligned}$$

Set now $u_0 := -\frac{1}{L} \sum_{i=1}^{n} \bar{x}_i$ and $T_0 := \frac{M}{L} \sum_{i=1}^{L} (x_i + u_0) \otimes (x_i + u_0)$. We get

$$\begin{aligned}
\|T_0 - \bar{T}_0\| &= \left\| \frac{M}{L} \sum_{i=1}^{L} (x_i \otimes u_0 + u_0 \otimes x_i) + M u_0 \otimes u_0 - S \right\| \\
&\leq 2 \left\| \left(\frac{M}{L} \sum_{i=1}^{L} x_i \right) \otimes u_0 \right\| + M \|u_0 \otimes u_0\| + \|S\| \\
&\leq 2 \frac{\varepsilon}{8} \cdot \frac{\varepsilon}{8M} + M \frac{n^2}{N^2} + \frac{\varepsilon}{8} \leq \frac{\varepsilon^2}{32M} + \frac{\varepsilon^2}{16M} + \frac{\varepsilon}{8} \leq \frac{\varepsilon}{4},
\end{aligned}$$

where the last expression follows for sufficiently small ε.
Let $\{\varepsilon_i\}_{i=1}^{N}$ be a sequence of independent Bernoulli random variables and $I_1 = \{i : \varepsilon_i = 1\}$ be the set of indices of random variables with value 1.

Clearly, with probability greater than 3/4, we have $L/4 \leq |I_1| \leq 3L/4$. Define the operator

$$\bar{T}_1 := 2\frac{M}{L}\sum_{i\in I_1}(x_i+u_0)\otimes(x_i+u_0).$$

Set $I_0 := \{1,...,L\}$ and momentarily $y_i := x_i + u_0$. We have, by Markov's inequality, that

$$\mathbb{P}\left(\|\bar{T}_1 - T_0\| \geq \mathbb{E}\|\bar{T}_1 - T_0\|\right) \leq \frac{1}{2}$$

and therefore, by Lemma 5.10, with probability at least 1/2

$$\begin{aligned}
\|\bar{T}_1 - T_0\| &\leq \mathbb{E}\|\bar{T}_1 - T_0\| \\
&= \frac{M}{L}\mathbb{E}\left\|\sum_{i\in I_1} y_i\otimes y_i - \sum_{i\in I_0\setminus I_1} y_i\otimes y_i\right\| \\
&\leq \frac{M}{L}\left(\mathbb{E}\left\|\sum_{i\in I_1} y_i\otimes y_i\right\| + \mathbb{E}\left\|\sum_{i\in I_0\setminus I_1} y_i\otimes y_i\right\|\right) \\
&= \frac{M}{L}\left(\mathbb{E}\left\|\sum_{i=1}^{L} \varepsilon_i y_i\otimes y_i\right\| + \mathbb{E}\left\|\sum_{i=1}^{L} (1-\varepsilon_i) y_i\otimes y_i\right\|\right) \\
&\overset{Lemma\ 5.10}{\leq} 2\frac{M}{L}C\log M\sqrt{\log L}\max_{i=1,\ldots,L}\|y_i\|\left\|\sum_{i=1}^{L} y_i\otimes y_i\right\|^{1/2}.
\end{aligned}$$

Since $\max_{i=1,\ldots,L}\|y_i\| \leq C$, for some constant C, and

$$\left\|\sum_{i=1}^{L} y_i\otimes y_i\right\| = \frac{L}{M}\,||T_0|| \leq \frac{L}{M}\left(\|T_0 - \bar{T}_0\| + \|\bar{T}_0 - id_M\| + \|id_M\|\right) \leq 2\frac{L}{M},$$

we can further estimate this to get

$$\|\bar{T}_1 - T_0\| \leq 4C\sqrt{\frac{M}{L}}\log M\sqrt{\log L} \tag{5.26}$$

Set

$$u_1 := -\frac{1}{|I_1|}\sum_{i\in I_1}(x_i+u_0).$$

Since $\sum_{i=1}^{L}(x_i+u_0) = 0$, we have $\sum_{i\in I_1}(x_i+u_0) = -\sum_{i\in I_0\setminus I_1}(x_i+u_0)$. This implies, by Jensen's inequality, linearity of the expectation, independence of the ε_i, and $\mathbb{E}(\varepsilon_i - 1/2) = 0$, that

$$\begin{aligned}
\mathbb{E}\left\|\sum_{i=1}^{L} \varepsilon_i(x_i+u_0)\right\| &= \mathbb{E}\left\|\sum_{i=1}^{L}\left(\varepsilon_i - \frac{1}{2}\right)(x_i+u_0)\right\| \\
&\leq \sqrt{\sum_{i=1}^{L} \mathbb{E}\left(\varepsilon_i - \frac{1}{2}\right)^2 \|x_i+u_0\|} \\
&\leq \sqrt{\sum_{i=1}^{L} \frac{1}{2}} = \sqrt{\frac{L}{2}}.
\end{aligned}$$

Therefore, we can choose I_1 such that $L/4 \leq |I_1| \leq 3L/4$, (5.26) holds and

$$\|u_1\| = \frac{1}{|I_1|}\left\|\sum_{i=1}^{L} \varepsilon_i(x_i+u_0)\right\| \leq \frac{\sqrt{M}}{\sqrt{2}|I_1|} \leq \frac{\sqrt{2}}{\sqrt{|I_1|}}.$$

Set

$$T_1 := 2\frac{M}{L}\sum_{i\in I_1}(x_i+u_0+u_1)\otimes(x_i+u_0+u_1).$$

We have

$$\begin{aligned}
\left\|\frac{M}{L}\sum_{i\in I_1}(x_i+u_0)\right\| &= \left\|\frac{M}{L}\sum_{i\in I_1} x_i + \frac{M|I_1|}{L}u_0\right\| \\
&\leq \frac{M}{L}\sum_{i\in I_1}\|x_i\| + \frac{M|I_1|}{L}\|u_0\| \\
&\leq \frac{M|I_1|}{L}\left(1+\frac{\varepsilon}{8M}\right),
\end{aligned}$$

and, thereby,

$$\begin{aligned}
\|T_1 - \bar{T}_1\| &\leq 2\frac{M}{L}\left(\left\|\sum_{i\in I_1}\left((x_i+u_0)\otimes u_1 + u_1\otimes(x_i+u_0)\right)\right\| + |I_1|\|u_1\otimes u_1\|\right) \\
&\leq 2\frac{M|I_1|}{L}\left(1+\frac{\varepsilon}{8M}\right)\cdot\frac{\sqrt{2}}{\sqrt{|I_1|}} + 2\frac{M|I_1|}{L}\cdot\frac{2}{|I_1|} \\
&\leq \frac{4\sqrt{2|I_1|}M}{L} + \frac{4M}{L} \\
&\leq \left(1+\sqrt{2|I_1|}\right)\cdot\frac{\varepsilon}{2n}.
\end{aligned}$$

Repeating these steps inductively, we get a sequence of sets $I_s \subset ... \subset I_1 \subset I_0 = \{1, ..., L\}$ and a sequence of vectors $u_0, u_1, ..., u_s$ for which

$$\frac{1}{4}|I_k| \leq |I_{k+1}| \leq \frac{3}{4}|I_k|,$$

$$\|u_k\| \leq \frac{\sqrt{2}}{\sqrt{I_k}}$$

and

$$\sum_{i\in I_k}(x_i + u_0 + ... + u_k) = 0$$

holds. Furthermore, we have for the operator

$$T_k := 2^k \frac{M}{L} \sum_{i\in I_k}(x_i + u_0 + ... + u_k) \otimes (x_i + u_0 + ... + u_k)$$

that

$$\|T_{k+1} - T_k\| \leq C\sqrt{\frac{M}{|I_k|}} \log M \sqrt{\log |I_k|}. \tag{5.27}$$

Now we sum up the inequalities (5.27) and get

$$\| \operatorname{id}_M - T_s\| \leq \| \operatorname{id}_M - T_0\| + \sum_{k=0}^{s-1} \|T_k - T_{k+1}\| \leq \frac{\varepsilon}{4} + CM \log M \sum_{k=0}^{s-1} \frac{\sqrt{\log |I_k|}}{\sqrt{|I_k|}}.$$

Choosing s such that the last expression of these inequalities is smaller than $\varepsilon/2$ one would be able to conclude that $|I_s| \leq C(\varepsilon)M \log^3 M$. But, since $n \leq C(\varepsilon)M \log M$ and $|I_s| \leq n$, we actually even have $|I_s| \leq C(\varepsilon)M \log M$. It follows that

$$\|u_0 + u_1 + ... + u_s\| \leq \frac{C(\varepsilon)}{\sqrt{M \log M}}.$$

By setting $m = |I_s|$, $u = u_0 + ... + u_s$ and renaming $\{1, ..., L\}$ such that I_s becomes the initial sequence $\{1, ..., m\}$, we obtain (5.24) and

$$\left\| \operatorname{id}_M - \frac{2^s M}{L} \sum_{i=1}^{m} (x_i + u) \otimes (x_i + u) \right\| \leq \frac{\varepsilon}{2}.$$

By this, $\left|M - \frac{2^s M}{L} m\right| < \frac{\varepsilon}{2} M$ holds and we can further estimate this to get

$$\left\| id_M - \frac{M}{m} \sum_{i=1}^{m} (x_i + u) \otimes (x_i + u) \right\| \leq \varepsilon.$$

Taking the trace of this returns (5.23) and (5.25). □

With the help of this Theorem we can finally show Theorem 5.3.

Proof. (Theorem 5.3)
Let B be embedded in $\mathbb{R}^M$ so that B_2^M is its John ellipsoid and let $\varepsilon > 0$. We use Theorem 5.11, namely that B has an approximate John's decomposition, to construct the convex body K. Define $\bar{B} := B + u$, $y_i := x_i + u$ and

$$T = \mathrm{id}_M - S = \frac{M}{m} \sum_{i=1}^{m} y_i \otimes y_i,$$

where $\|S\| < \varepsilon/8$.
We see that (5.24) can be written as

$$\sum_{i=1}^{m} y_i = 0. \tag{5.28}$$

Let $v \in \mathbb{R}^M$ be a vector with $\|v\| \leq \varepsilon/\sqrt{M}$, which will be defined later. Furthermore, set

$$T_v = \frac{M}{m} \sum_{i=1}^{m} (y_i + v) \otimes (y_i + v),$$

$R_v = T_v^{1/2}$ and $\mathcal{E} := \mathcal{E}_v := R_v B_2^M$.
By (5.28), we have

$$\begin{aligned} \|T_v - \mathrm{id}_M\| &\leq \|T_v - T\| + \|S\| \\ &\leq \left\| \frac{M}{m} \sum_{i=1}^{m} (y_i \otimes v + v \otimes y_i) \right\| + M\|v \otimes v\| + \|S\| \\ &= M\|v \otimes v\| + \|S\| \leq \varepsilon^2 + \frac{\varepsilon}{8} < \frac{\varepsilon}{4} \end{aligned} \tag{5.29}$$

for sufficiently small ε. Thus, $(1 - \frac{\varepsilon}{4})\mathcal{E} \subset B_2^M \subset (1 + \frac{\varepsilon}{4})\mathcal{E}$.
Denote

$$z_i := \frac{1}{\|y_i + v\|_{\mathcal{E}}} (y_i + v)$$

and set

$$\tilde{K} := \mathrm{conv}\left(\frac{1}{1+\varepsilon} (\bar{B} + v), z_1, ..., z_m \right).$$

Since $B \subset B_2^M$ and $\|v\| \leq \varepsilon/\sqrt{M}$, we get that the only contact points of $\tilde{K}$ with $\mathcal{E}$ are $z_1, ..., z_m$. We go on to show

$$\frac{1}{1+\varepsilon} (\bar{B} + v) \subset \tilde{K} \subset (1 + 2\varepsilon)(\bar{B} + v).$$

The first inclusion follows by the definition of $\tilde{K}$. To show the other inclusion, let $x \in \tilde{K}$ and consider the decomposition

$$x = \frac{\alpha_0}{1+\varepsilon} b + \sum_{i=1}^{m} \alpha_i z_i,$$

where $b \in \bar{B} + v$, $\alpha_i \geq 1$ and $\sum_{i=0}^{m} \alpha_i \leq 1$.
From $x_i \in \partial B$ and $y_i + v \in \partial(\bar{B} + v)$ we get $\|y_i + v\|_{\bar{B}+v} = 1$. Since $\|v\| \leq \varepsilon/\sqrt{M}$ and $\|u\| \leq C(\varepsilon)(M \log M)^{-1/2}$, we have, for sufficiently large M, that $\|u\| + \|v\| \leq \varepsilon/2$ and can conclude, that

$$\begin{aligned}
\|y_i + v\|_{\mathcal{E}} &\geq \left(\frac{1}{1+\frac{\varepsilon}{4}}\right) \|y_i + v\| \\
&\geq \left(\frac{1}{1+\frac{\varepsilon}{4}}\right) (\|x_i\| - \|u\| - \|v\|) \\
&\geq \left(\frac{1}{1+\frac{\varepsilon}{4}}\right) \left(1 - \frac{\varepsilon}{2}\right) \\
&\geq 1 - \varepsilon
\end{aligned}$$

Using now the triangle inequality, we can show that

$$\|x\| \leq \frac{\alpha_0}{1+\varepsilon}\|b\| + \sum_{i=1}^{m} \alpha_i \frac{\|y_i + v\|_{\bar{B}+v}}{\|y_i + v\|_{\mathcal{E}}} \leq \frac{\alpha_0}{1+\varepsilon} + \frac{1}{1-\varepsilon} \sum_{i=1}^{m} \alpha_i \leq 1 + 2\varepsilon.$$

Thus, we can define the following decomposition of the identity operator in $\mathbb{R}^M$

$$\mathrm{id}_M = R_v^{-1} \circ T \circ R_v^{-1} = \sum_{i=1}^{m} \frac{M}{m} \|y_i + v\|_{\mathcal{E}}^2 (R_v^{-1} z_i) \otimes (R_v^{-1} z_i).$$

Setting $a_i := \frac{M}{m}\|y_i + v\|_{\mathcal{E}}^2$ and $u_i := R_v^{-1} z_i$, for $i = 1, ..., m$, this reads as

$$\mathrm{id}_M = \sum_{i=1}^{m} a_i u_i \otimes u_i. \tag{5.30}$$

By defining $K := R_v^{-1}\tilde{K}$, we get a body $K \subset B_2^M$ that has $u_1, ..., u_m$ as the only contact points with B_2^M. It remains to choose the vector v, so that

$$\sum_{i=1}^{m} a_i u_i = 0, \tag{5.31}$$

since then, (5.30) and (5.31) become a John's decomposition of K and B_2^M is its John ellipsoid.
To do so, we rewrite (5.31) in the following way

$$\begin{aligned} 0 &= \sum_{i=1}^m a_i u_i \\ &= \frac{M}{m} R_v^{-1} \left(\sum_{i=1}^m \left|\left|R_v^{-1}(y_i+v)\right|\right|^2 \frac{y_i+v}{||R_v^{-1}(y_i+v)||} \right) \\ &= \frac{M}{m} R_v^{-1} \left(\sum_{i=1}^m \langle y_i+v, T_v^{-1}(y_i+v)\rangle^{1/2}(y_i+v) \right). \end{aligned}$$

Define, with the above reformulation, the function $F : \frac{\varepsilon}{\sqrt{M}} B_2^M \to \mathbb{R}^M$ by

$$F(v) = -\left(\sum_{i=1}^m \langle y_i+v, T_v^{-1}(y_i+v)\rangle^{1/2} \right)^{-1} \left(\sum_{i=1}^m \left(\langle y_i+v, T_v^{-1}(y_i+v)\rangle^{1/2} - 1 \right) y_i \right).$$

Recall that the Brouwer fixed point theorem states, that a continuous mapping G from the Euclidean unit ball into itself has a fixed point, i.e., a point w for which $G(w) = w$. Obviously, the same holds true for a stretched or shrinked version of the unit ball.
Hence, if we can show that for all w, with $\|w\| \le \varepsilon/\sqrt{M}$, it holds that $\|F(w)\| \le \varepsilon/\sqrt{M}$, then there exists a vector v, with $\|v\| \le \varepsilon/\sqrt{M}$, that fulfills (5.31).

Let $u \in B_2^M$ be a vector and $\alpha_1, ..., \alpha_m$, then by Cauchy-Schwarz

$$\begin{aligned} \left| \left\langle \sum_{i=1}^m \alpha_i y_i, u \right\rangle \right| &= \left| \sum_{i=1}^m \alpha_i \langle y_i, u\rangle \right| \\ &\le \left(\sum_{i=1}^m \alpha_i^2 \right)^{1/2} \left(\sum_{i=1}^m \langle y_i, u\rangle^2 \right)^{1/2} \\ &\le \sqrt{m} \max_{i=1,...,m} |\alpha_i| \left(\sum_{i=1}^m \langle y_i, u\rangle^2 \right)^{1/2} \\ &\le \frac{m}{\sqrt{M}} \max_{i=1,...,m} |\alpha_i| \|T\| \\ &\le \frac{m}{\sqrt{M}} \max_{i=1,...,m} |\alpha_i|(1+\varepsilon), \end{aligned} \tag{5.32}$$

since $\|T\| = \|id_M - S\| \leq \|id_M\| + \|S\| \leq 1 + \varepsilon/8 \leq 1 + \varepsilon$.
Furthermore, we have that

$$\begin{aligned} &\left|\left\langle y_i + w, T_w^{-1}(y_i + w)\right\rangle^{1/2} - 1\right| \\ &\leq \left|\langle y_i + w, y_i + w\rangle^{1/2} - 1\right| + \left|\langle y_i + w, y_i + w\rangle^{1/2} - \left\langle y_i + w, T_w^{-1}(y_i + w)\right\rangle^{1/2}\right|. \end{aligned}$$

With (5.29), i.e., $\|T_w - id_M\| \leq \varepsilon/4$, and $\|y_i + w\| \leq 2$ we can estimate further and conclude, for sufficiently large M, that this is smaller than

$$\left|\langle y_i + w, y_i + w\rangle^{1/2} - 1\right| + 2\|id_M - T_w\| \leq \frac{2\varepsilon}{\sqrt{M}} + 2\frac{\varepsilon}{4} \leq \frac{2\varepsilon}{3}. \tag{5.33}$$

We just showed in the estimate (5.33)

$$\left|\langle y_i + w, T_w^{-1}(y_i + w)\rangle^{1/2} - 1\right| \leq 2\|id_M - T_w\| \leq \varepsilon/2,$$

which, by the triangle and inverse triangle inequality, implies

$$1 - \varepsilon \leq \langle y_i + w, T_w^{-1}(y_i + w)\rangle^{1/2} \leq 1 + \varepsilon.$$

On the one hand, we get from this that

$$\sum_{i=1}^{m} \langle y_i + w, T_w^{-1}(y_i + w)\rangle^{1/2} \geq m(1 - \varepsilon), \tag{5.34}$$

and on the other hand, since $\frac{m}{M} a_i = \langle y_i + w, T_w^{-1}(y_i + w)\rangle^{1/2}$, that

$$\frac{m}{M} a_i \in [1 - \varepsilon, 1 + \varepsilon].$$

Thus, by (5.34), (5.32) and (5.33), we have for $w \in \frac{\varepsilon}{\sqrt{M}} B_2^M$

$$\begin{aligned} \|F(w)\| &= \left|\sum_{i=1}^{m} \langle y_i + w, T_w^{-1}(y_i + w)\rangle^{1/2}\right|^{-1} \left\|\sum_{i=1}^{m} \left(\langle y_i + w, T_w^{-1}(y_i + w)\rangle^{1/2} - 1\right) y_i\right\| \\ &\leq \frac{1}{\sqrt{M}} \frac{1 + \varepsilon}{1 - \varepsilon} \cdot \frac{2}{3} \varepsilon \leq \frac{\varepsilon}{\sqrt{M}}. \end{aligned}$$

Hence, there exist a vector v with $\|v\| \leq \frac{\varepsilon}{\sqrt{M}}$ satisfying (5.31). $\square$

Appendix A

In this appendix we provide results which are used throughout this work. References to proofs of these theorems and lemmas are given.

Probability theory

Lemma A.1.
Let $T \neq \emptyset$ be an index set and $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$ and $(\Omega_t, \mathcal{F}_t)$ be measure spaces, $t \in T$. Furthermore, let $(X_t)_{t \in T}$ be a family of measurable maps $X_t : \Omega' \to \Omega_t$ such that $\mathcal{F}' = \sigma(X_t : t \in T)$ holds. Then, the map $Y : \Omega \to \Omega'$ is $\mathcal{F}$-$\mathcal{F}'$-measurable, if and only if $X_t \circ Y$ is $\mathcal{F}$-$\mathcal{F}_i$-measurable for every $t \in T$.

Proof. see [K], Page 35. □

Definition A.2. (Dynkin system)
Let Ω be a set. Then, a set $\mathcal{D} \subset \mathcal{P}(\Omega)$ is called a **Dynkin system**, if

(i) $\Omega \in \mathcal{D}$

(ii) $A \in \mathcal{D} \implies A^C \in \mathcal{D}$

(iii) $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{D}$ disjoint $\implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}$.

Theorem A.3. (Dynkin's theorem)
Let $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ be closed under finite intersection. Then, it holds that

$$\delta(\mathcal{E}) = \sigma(\mathcal{E}),$$

where $\delta(\mathcal{E})$ denotes the generated Dynkin system.

Proof. see [K], Page 7. □

Lemma A.4. (Uniqueness by generators with finite intersection property)
Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space and let $\mathcal{E} \subset \mathcal{F}$ be a generator of $\mathcal{F}$ which is closed under finite intersection and such, that there exists a sequence $(E_n)_{n\in\mathbb{N}}$ with $E_n \in \mathcal{E}$, $\mu(E_n) < \infty$, for all $n \in \mathbb{N}$ and $E_n \uparrow \Omega$.
Then, μ is uniquely determined by the values $\mu(E)$ for $E \in \mathcal{E}$.
If μ is a probability measure, the lemma holds without the existence of the sequence $(E_n)_{n\in\mathbb{N}}$.

Proof. see [K], Page 19. □

Theorem A.5. (Monotone convergence theorem)
Let $(f_n)_{n\in\mathbb{N}}$ be a monotonously growing sequence of positive, measurable functions. Then, following holds

$$\int_\Omega \sup_{n\in\mathbb{N}} f_n d\mu = \sup_{n\in\mathbb{N}} \int_\Omega f_n d\mu.$$

Proof. see [K], Page 85. □

Lemma A.6. (Dominated convergence theorem)
Let f be measurable and $(f_n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{L}^1(\mu)$ with $f_n \overset{n\to\infty}{\to} f$ converging in probability. Let $0 \leq g \in \mathcal{L}^1(\mu)$ be such that $|f_n| \leq g$ almost everywhere for all $n \in \mathbb{N}$. Then, it holds that $f \in \mathcal{L}^1(\mu)$ and $f_n \overset{n\to\infty}{\to} f$ in L^1, i.e.,

$$\int f_n d\mu \overset{n\to\infty}{\to} \int f d\mu.$$

Proof. See [K], Page 135. □

Theorem A.7.
Let (E, d) be a metric space, $x_0 \in E$ and $f : \Omega \times E \to \mathbb{R}$ a map with the properties that

(i) for every $x \in E$ the map $\omega \mapsto f(\omega, x)$ is in $\mathcal{L}^1(\mu)$,

(ii) for almost all $\omega \in \Omega$ the map $x \mapsto f(\omega, x)$ is continuous in x_0,

(iii) the map $h : \omega \mapsto \sup_{x\in E} |f(\omega, x)|$ is in $\mathcal{L}^1(\mu)$.

Then, the map $F : E \to \mathbb{R}$, $x \mapsto \int f(\omega, x)\mu(d\omega)$ is continuous in x_0.

Proof. see [K], Page 136. □

Theorem A.8. (Factorization theorem)
Let $(\Omega', \mathcal{F}')$ be a measure space and let Ω be a non-empty set. Let $f : \Omega \to \Omega'$ be a map. A map $g : \Omega \to \bar{\mathbb{R}}$ is $\sigma(f)$-$\mathcal{B}(\bar{\mathbb{R}})$-measurable, if and only if there exists a measurable map $\varphi : (\Omega', \mathcal{F}') \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ with $g = \varphi \circ f$.

Proof. see [K], Page 40. □

Definition A.9. (Conditional expectation, conditional probability)
A random variable Y is called **conditional expectation** of a random variable X given the σ-algebra $\mathcal{F}$, denoted by $Y = \mathbb{E}(X|\mathcal{F})$, if

(i) Y is $\mathcal{F}$-measurable

(ii) $\forall A \in \mathcal{F} : \mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)$.

For each $B \in \mathcal{F}$, the **conditional probability** of B given $\mathcal{F}$ is defined as $\mathbb{P}(B|\mathcal{F}) := \mathbb{E}(\mathbb{1}_B|\mathcal{F})$.

Theorem A.10. (Properties of the conditional expectation)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ be random variables. Furthermore, let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then, we have that

(i) **(Monotonicity)** $\mathbb{E}(X|\mathcal{G}) \geq \mathbb{E}(Y|\mathcal{G})$, if $X \geq Y$ almost surely.

(ii) **(Dominated Convergence)** $\lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$ almost surely and in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, if $Y \geq 0$ and $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables with $|X_n| \leq Y$, $n \in \mathbb{N}$, and $X_n \to X$ almost surely, as $n \to \infty$.

Proof. see [K], Page 170. □

Theorem A.11. (Uniqueness theorem for measures)
Let $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ be a σ-algebra and let $\mathcal{E}$ be such, that it is closed under finite intersection and that $\sigma(\mathcal{E}) = \mathcal{F}$. If $\mu, \nu : \mathcal{F} \to [0, \infty]$ are measures with $\mu|_{\mathcal{E}} = \nu|_{\mathcal{E}}$ and $\mu|_{\mathcal{E}}$ σ-finite, then $\mu = \nu$ holds and μ is σ-finite.

Proof. see [S], Page 63, 64. □

Theorem A.12. (Carathéodory extension theorem)
Let $\mathcal{R} \subseteq \mathcal{P}(\Omega)$ be a ring and μ be a σ-finite pre-measure on $\mathcal{R}$. Then, there exists a unique measure $\widetilde{\mu}$ on $\sigma(\mathcal{R})$ such that $\widetilde{\mu}|_{\mathcal{R}} = \mu$ and $\widetilde{\mu}$ is σ-finite.

Proof. see [K], Page 19. □

Definition A.13. ($\emptyset$-continuity)
Let μ be a content on the ring $\mathcal{R}$. μ is called **$\emptyset$-continuous**, if for every sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{R}$ with $\mu(A_n) < \infty$ and $A_n \supset A_{n+1}$, for all $n \in \mathbb{N}$, and $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$, it holds that $\mu(A_n) \to 0$ as $n \to \infty$.

Lemma A.14.
Let μ be a finite content on the ring $\mathcal{R}$. Then, the following two statements are equivalent

(i) μ is σ-additive (and therefore a pre-measure),

(ii) μ is $\emptyset$-continuous.

Proof. see [S], Page 54. □

Theorem A.15.
Let $(\mathcal{E}_t)_{t\in T}$ be an independent family of sets $\mathcal{E}_t \subset \mathcal{F}$ which are closed under finite intersection and let $(T_j)_{j\in J}$ be a partition of T. If $\mathcal{F}_j := \sigma\left(\bigcup_{i\in I_j} \mathcal{E}_i\right)$ are the σ-algebras generated by all $\mathcal{E}_i$ with $i \in I_j$, then the family $(\mathcal{F}_j)_{j\in J}$ is independent.

Proof. see [B], Page 45 and 46. □

Definition A.16. (Hölder-continuity)
Let (E, d), (E', d') be metric spaces and $\gamma \in (0, 1]$. A mapping $\varphi : E \to E'$ is **Hölder-continuous of order γ in a point** $r \in E$, if there exists an $\epsilon > 0$ and a $C < \infty$, such that for all $s \in E$ with $d(s, r) < \epsilon$ it holds that

$$d'(\varphi(r), \varphi(s)) \leq C \cdot d(r, s)^{\gamma}. \tag{A.1}$$

φ is called **locally Hölder-continuous of order** γ, if for every $t \in E$ an $\epsilon > 0$ and a $C = C(t, \epsilon) > 0$ exists, such that for all $s, r \in E$ with $d(s, t) < \epsilon$ and $d(r, t) < \epsilon$, inequality (A.1) holds.
φ is called **Hölder-continuous of order** γ, if there exists a C, such that (A.1) holds for all $s, r \in E$.

Theorem A.17.
Let $\{(\Omega_i, \mathcal{F}_i, \mu_i)\}_{1\leq i\leq n}$, $n \in \mathbb{N}$, be a family of σ-finite measure spaces. There exist a unique measure $\mu : \mathcal{F}_{\{1,\dots,n\}} \mapsto [0, \infty]$ such that

$$\forall \bigtimes_{\{1,\dots,n\}} A_i \in \mathcal{F}_{\{1,\dots,n\}} : \mu\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mu_i(A_i)$$

holds. The measure μ is σ-finite. μ is a probability measure, if all μ_i are probability measures.

Proof. see [S], Page 173. □

Theorem A.18. (Integration with respect to push-forward measure)
Let $(\Omega_1, \mathcal{F}_1)$ be a measurable space, $(\Omega_2, \mathcal{F}_2, \mu)$ be a measure space and $T : \Omega_2 \to \Omega_1$ be a measurable map. Furthermore, let $f : \Omega_1 \to [-\infty, \infty]$ be a measurable function. Then

$$\int f d(\mu \circ T^{-1}) = \int f \circ T d\mu$$

holds if either integral is defined.

Proof. See [D1], Page 121. □

Existence of the Brownian Motion

The main concern of this chapter will be to prove the existence of the Brownian motion on $\mathbb{R}$. The way this is done here is more extensive than necessary, but will eventually allow us to apply a similar result to a wider range of stochastic processes. Namely, we will see that the construction we use only relies on a certain convolution behavior a family of probability measures can have.

Definition A.19.
A stochastic process $X = (X_t)_{t\in T}$ with values in Ω_2 is

(i) **real-valued**, if $\Omega_2 = \mathbb{R}$,

(ii) a process with **independent increments**, if for every $n \in \mathbb{N}$ and all $t_0, ..., t_n \in T$, with $0 = t_0 < ... < t_n$, it holds that

$$X_{t_0},\ X_{t_1} - X_{t_0}, ...,\ X_{t_n} - X_{t_{n-1}} \text{ is independent,}$$

(iii) a process with **stationary increments**, if for all $r, s, t \in T$ it holds that

$$\mathbb{P}_{X_{r+s+t}-X_{s+t}} = \mathbb{P}_{X_{r+s}-X_s}.$$

If $0 \in T$, this simplifies to $\mathbb{P}_{X_{r+s}-X_s} = \mathbb{P}_{X_r-X_0}$ for all $r, s \in T$.

Remark. The independence as demanded in (ii) implies that the independence holds for any choice of times $0 \leq t_0 < ... < t_n$.
In the case that $t_0 > 0$ it suffices to add $t_{-1} := 0$ to the sequence of points of time. From $X_{t_0} = X_{t_{-1}} + (X_{t_0} - X_{t_{-1}})$ one sees that X_{t_0} is measurable w.r.t to the σ-algebra $\mathcal{C}$ generated by $X_{t_{-1}}$ and $X_{t_0} - X_{t_{-1}}$, i.e. $\sigma(X_{t_0}) \subset \mathcal{C}$.
If $X_{t_{-1}}$, $X_{t_0} - X_{t_{-1}}$,..., $X_{t_n} - X_{t_{n-1}}$ are independent, it follows from Theorem A.15, in the appendix, that $\mathcal{C}$, $\sigma(X_{t_1} - X_{t_0})$,...,$\sigma(X_{t_n} - X_{t_{n-1}})$ are independent and consequently that X_{t_0}, $X_{t_1} - X_{t_0}$,..., $X_{t_n} - X_{t_{n-1}}$ are independent

Usually one uses the motion of a pollen in resting water to give an idea how to abstract a Brownian motion from a real life application. A convenient description is given in [D1] or [M-P].

Definition A.20. (Standard Brownian motion)
A real-valued stochastic process $B = (B_t)_{t\in[0,\infty)}$ is a **Brownian motion**, if

(i) $B_0 = 0$ almost surely,

(ii) B has stationary and independent increments,

(iii) B_t is normally distributed with mean 0 and variance t, i.e. $B_t \sim \mathcal{N}_{0,t}$

(iv) the **paths** $t \mapsto B_t$ of B are $\mathbb{P}$-almost surely continuous, for all $t \in T$.

Remark. When we refer to Brownian motion throughout this work, we always refer to the Standard Brownian motion.

From here on, as already mentioned, we will study the existence of the Brownian motion in an exhaustive way. This will further illustrate the power of transition kernels, when working with stochastic processes.

Definition A.21. Concatenation of kernels
Let $(\Omega_i, \mathcal{F}_i)$, $i = 0, 1, 2$, be measurable spaces and let κ_i be stochastic kernels from $(\Omega_{i-1}, \mathcal{F}_{i-1})$ to $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$. The **Concatenation** of κ_1 and κ_2 is defined by

$$\begin{aligned}\kappa_1 \cdot \kappa_2 : \Omega_0 \times \mathcal{F}_2 &\to [0, \infty)\\ (\omega_0, A_2) &\mapsto \int_{\Omega_1} \kappa_1(\omega_0, d\omega_1)\kappa_2(\omega_1, A_2).\end{aligned}$$

Remark. The concatenation of kernels is just a special case of the product of kernels. With the first remark on the product of stochastic kernels this can be written as $(\kappa_1 \cdot \kappa_2)(\omega_0, A_2) = (\kappa_1 \otimes \kappa_2)(\omega_0, \pi_2^{-1}(A_2))$, for all $A_2 \in \mathcal{F}_2$, where π_2 denotes the projection to Ω_2.

Definition A.22. (Consistent family of kernels)
Let E be a Polish space, $T \subset \mathbb{R}$ a non-empty index set and let $(\kappa_{s,t} : s, t \in T, s < t)$ be a family of stochastic kernels from E to E. The family is called **consistent**, if for all $r, s, t \in T$, with $r < s < t$, it holds that

$$\kappa_{r,s} \cdot \kappa_{s,t} = \kappa_{r,t}.$$

Definition A.23.
Let E be a Polish space and let $T \subset [0, \infty)$ be a semigroup. A family $(\kappa_t)_{t \in T}$ of stochastic kernels from E to E is a **semigroup of stochastic kernels** or **Markovian semigroup**, if the **Chapman-Kolmogorov equation** holds, i.e.,

$$\forall s, t \in T: \ \kappa_s \cdot \kappa_t = \kappa_{s+t}.$$

Lemma A.24.
If $(\kappa_t)_{t \in T}$ is a Markovian semigroup, then the family of kernels defined by $\tilde{\kappa}_{s,t} := \kappa_{t-s}$, $t > s$, is consistent.

Proof. Let $r, s, t \in T$ with $r < s < t$. Since $(\kappa_t)_{t \in T}$ is a Markovian semigroup the Chapman-Kolmogorov equation holds and we have that

$$\tilde{\kappa}_{r,s} \cdot \tilde{\kappa}_{s,t} = \kappa_{r-s} \cdot \kappa_{s-t} = \kappa_{r-s+s-t} = \kappa_{r-t} = \tilde{\kappa}_{r,t}.$$

□

The following result gives the existence of a stochastic kernel on the product space of, up to uncountably many, polish spaces under the assumption that the kernels defined on each of these polish spaces form a consistent family of stochastic kernels, as defined above. This theorem utilizes the Kolmogorov existence theorem and is the core result of this chapter.

Theorem A.25. (Kernel given by a consistent family of kernels)
Let $T \subset [0, \infty)$ and let $(\kappa_{s,t} : s, t \in T, s < t)$ be a consistent family of stochastic kernels on a Polish space E. Then, there exists a kernel κ from $(E, \mathcal{B}(E))$ to $(E^T, \mathcal{B}(E)^{\otimes T})$, such that for each $x \in E$ and $J := \{j_0, ..., j_n\} \in \mathcal{E}(T)$, with $0 = j_0 < ... < j_n$, it holds that

$$\kappa(x, \ \cdot \ ) \circ \pi_J^{-1} = \left(\bigotimes_{k=0}^{n-1} \kappa_{j_k, j_{k+1}} \right) (x, \ \cdot \ ).$$

Proof. In order to prove this we have to show that κ exists and that it is a stochastic kernel.
The Kolmogorov existence theorem guarantees the existence of the kernel κ, if the family $\mathbb{P}_J := (\bigotimes_{k=0}^{n-1} \kappa_{j_k, j_{k+1}})(x, \ \cdot \ )$, $J \in \mathcal{E}(T)$ with $0 \in J$, is projective. Let $0 \in L \subset J \subset T$, $L, J \in \mathcal{E}(T)$. We have to show that $\mathbb{P}_J \circ (\pi_L^J)^{-1} = \mathbb{P}_L$. W.l.o.g. we can assume that $L = J \setminus \{j_\ell\}$ with $\ell = 1, ..., n$, since the general case follows inductively.
We have to distinguish two case. First, let $\ell = n$, let $A_{j_0}, ..., A_{j_{n-1}} \in \mathcal{B}(E)$

and define $A := \bigtimes_{j\in L} A_j$. Then, we have that

$$(\mathbb{P}_J \circ (\pi_L^J)^{-1})(A) = \mathbb{P}_J(A \times E) = (\mathbb{P}_L \otimes \kappa_{j_{n-1},j_n})(A \times E)$$
$$= \int_A \mathbb{P}_L(d(\omega_0, ..., \omega_{n-1}))\kappa_{j_{n-1},j_n}(\omega_{n-1}, E) = \int_A \mathbb{P}_L(d(\omega_0, ..., \omega_{n-1})) = \mathbb{P}_L(A).$$

For the second case, let $\ell \in \{1, ..., n-1\}$, let $A_j \in \mathcal{B}(E)$ for all $j \in L$ and define the sets $A := \bigtimes_{j\in L} A_j$, $A_1 := \bigtimes_{k=1}^{\ell-1} A_{j_k}$ and $A_2 := \bigtimes_{k=\ell+1}^{n} A_{j_k}$. To further simplify the notation we introduce $f_i(\omega_i) := (\bigotimes_{k=i}^{n} \kappa_{j_k,j_{k+1}})(\omega_i, A_{j_{i+1}} \times ... \times A_{j_n})$ for all $i = 0, ..., n-1$.
Since we are dealing with a consistent family of kernels and Fubini's theorem holds true for finite transition kernels (see [K], Page 270) we get that

$$f_{\ell-1}(\omega_{\ell-1}) = (\bigotimes_{k=\ell-1}^{n} \kappa_{j_k,j_{k+1}})(\omega_{\ell-1}, E \times A_2)$$
$$= \int_E \kappa_{j_{\ell-1},j_\ell}(\omega_{\ell-1}, d\omega_\ell) \int_{A_{j_{\ell+1}}} \kappa_{j_\ell,j_{\ell+1}}(\omega_\ell, d\omega_{\ell+1})(\bigotimes_{k=\ell+1}^{n} \kappa_{j_k,j_{k+1}})(\omega_{\ell+1}, A_{j_{\ell+2}} \times ... \times A_{j_n})$$
$$\overset{Fubini}{=} \int_{A_{j_{\ell+1}}} \int_E \kappa_{j_{\ell-1},j_\ell}(\omega_{\ell-1}, d\omega_\ell)\kappa_{j_\ell,j_{\ell+1}}(\omega_\ell, d\omega_{\ell+1}) f_{\ell+1}(\omega_{\ell+1})$$
$$\overset{Consistency}{=} \int_{A_{j_{\ell+1}}} \kappa_{j_{\ell-1},j_{\ell+1}}(\omega_{\ell-1}, d\omega_{\ell+1}) f_{\ell+1}(\omega_{\ell+1}).$$

From this, it follows that

$$(\mathbb{P}_J \circ (\pi_L^J)^{-1})(A) = \mathbb{P}_J(A_1 \times E \times A_2) = (\bigotimes_{k=0}^{n} \kappa_{j_k,j_{k+1}})(x, A_1 \times E \times A_2)$$
$$= \int_{A_1} (\bigotimes_{k=0}^{\ell-2} \kappa_{j_k,j_{k+1}})(x, d(\omega_1, ..., \omega_{\ell-1}))(\bigotimes_{k=\ell-1}^{n} \kappa_{j_k,j_{k+1}})(d\omega_{l-1}, E \times A_2)$$
$$= \int_{A_1} (\bigotimes_{k=0}^{\ell-2} \kappa_{j_k,j_{k+1}})(x, d(\omega_1, ..., \omega_{\ell-1})) \int_{A_{j_{\ell+1}}} \kappa_{j_{\ell-1},j_{\ell+1}}(\omega_{\ell-1}, d\omega_{\ell+1}) f_{\ell+1}(\omega_{\ell+1})$$
$$= (\bigotimes_{k=0}^{\ell-2} \kappa_{j_k,j_{k+1}} \otimes \kappa_{j_{\ell-1},j_{\ell+1}} \otimes \bigotimes_{k=\ell+1}^{n} \kappa_{j_k,j_{k+1}})(x, A) = P_L(A),$$

and, therefore, that $\mathbb{P}_J := (\bigotimes_{k=0}^{n} \kappa_{j_k,j_{k+1}})(x, \ \cdot \ )$, $J \in \mathcal{E}(T)$ with $0 \in J$, is projective.

To finish the proof we need to show that κ is a stochastic kernel, i.e., that $x \mapsto \kappa(x, A)$ is $\mathcal{B}(E)$-$\mathcal{B}(E)^{\otimes T}$-measurable. From Corollary 2.15, we know that it suffices to show this for rectangular cylinders with finite Basis $\mathcal{Z}^{\mathcal{R}}$, since Lemma 2.12 gives that this is a generator of $\mathcal{B}(E)^{\otimes T}$ with the finite intersection property.
Let $0 = t_0 < ... < t_n$, $B_0, ..., B_n \in \mathcal{B}(E)$ and $A := \bigcap_{i=0}^{n} \pi_{t_i}^{-1}(B_i)$.
From Lemma 2.16, the lemma about products of kernels, we know that the finite product of stochastic kernels is stochastic and, therefore, that

$$x \mapsto \mathbb{P}_x(A) = (\bigotimes_{i=0}^{n-1} \kappa_{t_{i+1}-t_i})(x, \bigtimes_{i=0}^{n} B_i)$$

is measurable for every rectangular cylinder with finite basis. □

To be able to use this central theorem, we have to reformulate it for Markovian semigroups. The reformulation will tell us how a probability measure on a product space of uncountably many polish spaces, depending on a "starting distribution", has to look.

Corollary A.26. (Measures by Markovian semigroups)
Let $(\kappa_t)_{t\in T}$ be a Markovian semigroup on a Polish space E. Then, there exists a unique stochastic kernel κ from $(E, \mathcal{B}(E))$ to $(E^T, \mathcal{B}(E)^{\otimes T})$ such that for all $J := \{t_0, ..., t_n\}$, $0 = t_0 < ... < t_n \in T$,

$$\forall x \in E \ \forall J: \ \kappa(x, \cdot) \circ \pi_J^{-1} = \left(\bigotimes_{k=0}^{n-1} \kappa_{t_{k+1}-t_k} \right)(x, \cdot).$$

For every probability measure μ on E there exists a unique probability measure $\mathbb{P}^\mu$ on $(E^T, \mathcal{B}(E)^{\otimes T})$ such that

$$\forall J: \ \mathbb{P}^\mu \circ \pi_J^{-1} = \mu \otimes \bigotimes_{k=0}^{n-1} \kappa_{t_{k+1}-t_k}$$

Proof. Since $(\kappa_t)_{t\in T}$ is a Markovian semigroup, we have from Lemma A.24 that $\tilde{\kappa}_{s,r} := \kappa_{r-s}$, $r > s$, is a consistent family of kernels. The unique existence of the kernel κ follows from Theorem A.25. From the first remark on products of stochastic kernels it follows that for any probability measure μ on E the product $\mu \otimes \bigotimes_{k=0}^{n-1} \kappa_{t_{k+1}-t_k}$ is a probability measure. □

Definition A.27.
A Markovian semigroup $(\kappa_t)_{t\in T}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d)$ is called **translation invariant** if, for every $t \in T$, it holds that

$$\forall x, z \in \mathbb{R}^d,\ A \in \mathcal{B}(\mathbb{R})^d :\ \kappa_t(x, A) = \kappa_t(x + z, A + z).$$

Lemma A.28.
Let $(\mathbb{P}_t)_{t\in\mathbb{R}_+}$ be a family of probability measures on $\mathbb{R}^d$ such that for all $s, t \in \mathbb{R}_+$ it holds that $\mathbb{P}_s * \mathbb{P}_t = \mathbb{P}_{s+t}$. Then, $\kappa_t(x, A) := \mathbb{P}_t(A - x)$ defines a translation invariant Markovian semigroup $(\kappa_t)_{t\in\mathbb{R}_+}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d)$.

Proof. By definition $\kappa_t(x+z, A+z) = \mathbb{P}_t(A+z-x-z) = \mathbb{P}_t(A-x) = \kappa_t(x, A)$ holds for all $x, z \in \mathbb{R}^d$, $A \in \mathcal{B}(\mathbb{R})^d$, $t \in \mathbb{R}_+$ and, therefore, that $(\kappa_t)_{t\in\mathbb{R}_+}$ is translation invariant.
Obviously, $\kappa_t(x, A) = \mathbb{P}_t(A - x)$ is a probability measure for each $x \in \mathbb{R}^d$. Since, for each $A \in \mathcal{B}(\mathbb{R})^d$, the map $(x, y) \mapsto \mathbb{1}_A(x + y)$ is the composition of the continuous map $(x, y) \mapsto x + y$ with the $\mathcal{B}(\mathbb{R})^d$-measurable indicator function $\mathbb{1}_A$, it is $\mathcal{B}(\mathbb{R})^d \otimes \mathcal{B}(\mathbb{R})^d$-measurable. The $\mathcal{B}(\mathbb{R})^d$-measurability of

$$x \ \mapsto \int \mathbb{1}_A(x + y)\mathbb{P}_t(dy) \ = \int \mathbb{1}_{A-x}(y)\mathbb{P}_t(dy) = \mathbb{P}_t(A - x)$$

then follows by Fubini's theorem. Therefore, $(\kappa_t)_{t\in\mathbb{R}_+}$ are stochastic kernels. Furthermore, $(\kappa_t)_{t\in\mathbb{R}_+}$ is a Markovian semigroup, which follows from the fact that the Chapman-Kolmogorov equation holds:

$$\kappa_{s+t}(x, A) = \mathbb{P}_{s+t}(A - x) = (\mathbb{P}_s * \mathbb{P}_t)(A - x) = \int \mathbb{P}_s(dy)\mathbb{P}_t(A - y - x)$$

$$= \int \kappa_s(x, d(x + y))\kappa_t(y + x, A) = (\kappa_s \cdot \kappa_t)(x, A).$$

□

As one sees now, it suffices to have a family of probability measures $(\mathbb{P}_t)_{t\in\mathbb{R}_+}$ with $\mathbb{P}_s * \mathbb{P}_t = \mathbb{P}_{s+t}$ for all $s, t \in \mathbb{R}_+$, for all the result we have so far to be true. Furthermore, with the help of this property we can derive the second crucial theorem of this chapter.

Theorem A.29.
Let $(\Omega, \mathcal{F}, \mathbb{P}^\mu, (X_t)_{t\geq 0})$ be the process coming from a translation invariant Markovian semigroup $(\mathbb{P}_t)_{t\geq 0}$ on $\mathbb{R}^d$ and a probability measure μ, where X_t is the t-th coordinate function.
Then, the process $(X_t)_{t\geq 0}$ has stationary and independent increments. Furthermore, it holds for $s, t \in \mathbb{R}_+$, with $s \leq t$, that

$$\mathbb{P}^\mu_{X_t - X_s} = \mathbb{P}_{t-s}.$$

Proof. First, define $Y := (X_s, X_t)$ and the map $q(x_1, x_2) := x_2 - x_1$. Clearly, q is continuous and thereby Borel-measurable. With these definitions $\mathbb{P}^{\mu}_{X_t - X_s}$ is the distribution of $q \circ Y$ w.r.t. $\mathbb{P}^{\mu}$, i.e.,

$$\mathbb{P}^{\mu}_{X_t - X_s}(B) = \mathbb{P}^{\mu}\{X_t - X_s \in B\} = \mathbb{P}^{\mu}\{q \circ Y \in B\} = \mathbb{P}^{\mu}\{Y \in q^{-1}(B)\}$$

and, therefore, we have

$$\begin{aligned}
&\mathbb{P}^{\mu}_{X_t - X_s}(B) = \int\int\int \mathbb{1}_{q^{-1}(B)}(x_1, x_2)\kappa_{t-s}(x_1, dx_2)\kappa_s(x, dx_1)\mu(dx) \\
&= \int\int\int \mathbb{1}_{x_1 + B}(x_2)\kappa_{t-s}(x_1, dx_2)\kappa_s(x, dx_1)\mu(dx) \\
&= \int\int \kappa_{t-s}(x_1, x_1 + B)\kappa_s(x, dx_1)\mu(dx) \\
&\overset{trans.inv.}{=} \mathbb{P}_{t-s}(B) \int\int \kappa_s(x, dx_1)\mu(dx) \\
&= \mathbb{P}_{t-s}(B).
\end{aligned}$$

By this, we have that $\mathbb{P}^{\mu}_{X_{t+s} - X_s} = \mathbb{P}_{(t+s)-s} = \mathbb{P}_{t-0} = \mathbb{P}^{\mu}_{X_t - X_0}$ and consequently that $(X_t)_{t \geq 0}$ is stationary.

In the following let $J = t_0, ..., t_n$ with $0 = t_0 < ... < t_n$.
To see the independence of the increments we make the observation that $\kappa_t(x, A) = \mathbb{P}_t(A - x)$ can be written as $\int \mathbb{1}_A(y)\kappa_t(x, dy) = \int \mathbb{1}_A(x + y)\mathbb{P}_t(dy)$. For a Borel-measurable function $f : (\mathbb{R}^d)^n \mapsto [0, \infty]$ successive application of this yields

$$\begin{aligned}
&\int f(x_1, ..., x_n)\mathbb{P}_J(d(x_1, ..., x_n)) \\
&= \int ... \int f(x_0 + x_1, ..., x_0 + ... + x_n)\mathbb{P}_{t_n - t_{n-1}}(dx_n)...\mathbb{P}_{t_1}(dx_1)\mu(dx_0).
\end{aligned}$$

From this and the transformation formula (Appendix, Theorem A.18), together with defining $Y_0 := X_{t_0}$, $Y_1 := X_{t_1} - X_{t_0}$,...,$Y_n := X_{t_n} - X_{t_{n-1}}$, and $X_{t_{-1}} := 0$, it follows that

$$\begin{aligned}
&\mathbb{P}^{\mu}_{(Y_0, ..., Y_n)}(A_0 \times ... \times A_n) \\
&= \int \mathbb{1}_{A_0 \times ... \times A_n} d\mathbb{P}^{\mu}_{(Y_0, ..., Y_n)} \\
&= \int \mathbb{1}_{A_0 \times ... \times A_n} \circ (Y_0, ..., Y_n) d\mathbb{P}^{\mu} \\
&= \int \prod_{j=0}^{n} (\mathbb{1}_{A_j} \circ Y_j) d\mathbb{P}^{\mu}
\end{aligned}$$

$$
\begin{aligned}
&= \int \prod_{j=0}^{n} \mathbb{1}_{A_j} \circ (X_{t_j} - X_{t_{j-1}}) d\mathbb{P}^{\mu} \\
&= \int \mathbb{1}_{A_0}(x_0) \prod_{j=1}^{n} \mathbb{1}_{A_j}(x_j - x_{j-1}) \mathbb{P}_J(d(x_0, ..., x_n)) \\
&= \int ... \int \mathbb{1}_{A_0}(x_{-1} + x_0) \prod_{j=1}^{n} \mathbb{1}_{A_j}(x_j) \mathbb{P}_{t_n - t_{n-1}}(dx_n) ... \mathbb{P}_{t_0}(dx_0) \mu(dx_{-1}) \\
&= \left(\prod_{j=1}^{n} \int \mathbb{1}_{A_j}(x_j) \mathbb{P}_{t_j - t_{j-1}} \right) \int \mathbb{1}_{A_0}(x_{-1} + x_0) \mu(dx_{-1}) \\
&= \mu(A_0) \prod_{j=1}^{n} \mathbb{P}_{t_j - t_{j-1}}(A_j).
\end{aligned}
$$

Since $\mathbb{P}_0 = \delta_0$ and $Y_0 = X_{t_0} = X_0$, it holds that

$$
\begin{aligned}
\mathbb{P}^{\mu}_{Y_0}(A_0) &= \int \kappa_0(x, A_0) \mu(dx) \\
&= \int \mathbb{P}_0(A_0 - x) \mu(dx) \\
&= \int \delta_0(A_0 - x) \mu(dx) \\
&= \mu(A_0)
\end{aligned}
$$

and, together with the stationarity, that

$$
\begin{aligned}
\mu(A_0) \prod_{j=1}^{n} \mathbb{P}_{t_j - t_{j-1}}(A_j) &= \mathbb{P}^{\mu}_{Y_0}(A_0) \prod_{j=1}^{n} \mathbb{P}_{t_j - t_{j-1}}(A_j) \\
&\overset{stat.}{=} \mathbb{P}^{\mu}_{Y_0}(A_0) \prod_{j=1}^{n} \mathbb{P}^{\mu}_{X_{t_j} - X_{t_{j-1}}}(A_j) \\
&= \prod_{j=0}^{n} \mathbb{P}^{\mu}_{Y_j}(A_j)
\end{aligned}
$$

□

Definition A.30. (Modification/Version)
Let X and Y be stochastic processes from $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ to $(\Omega_2, \mathcal{F}_2)$ with the index set T. Then X and Y are called **modifications** or **versions** of each other, if

$$\forall t \in T: \;\; X_t = Y_t \quad \mathbb{P}\text{-almost surely}$$

Theorem A.31. (Kolmogorov-Chentsov)
Let $X = (X_t)_{t\in[0,\infty)}$ be a real-valued process. For each $t > 0$ let there be numbers $\alpha, \beta, C > 0$ with

$$\forall s,t \in [0,T]: \ \mathbb{E}\left(|X_t - X_s|^\alpha\right) \leq C|t-s|^{1+\beta}.$$

Then, there exists a modification $\widetilde{X} = (\widetilde{X}_t)_{t\in[0,\infty)}$ of X, which has locally Hölder-continuous paths of every order $\gamma \in (0, \frac{\beta}{\alpha})$.

Proof. see [K], Page 432-434. □

Theorem A.32.
There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Brownian motion B on $(\Omega, \mathcal{F}, \mathbb{P})$. The paths of B are almost surely locally Hölder-γ-continous with $\gamma < \frac{1}{2}$.

Proof. Let $T = [0,\infty)$, $\Omega_t = \mathbb{R}$ and $\mathcal{B}_t = \mathcal{B}(\mathbb{R})$ for all $t \in T$. Then, $\Omega = \mathbb{R}^{[0,\infty)}$ and $\mathcal{B} = \mathcal{B}(\mathbb{R})^{\otimes[0,\infty)}$ holds. Define $B_t := \pi_t$ as the t-th coordinate function and $B := (B_t)_{t\in T}$.

For every $t \in T$ define $\mathbb{P}_t := \mathcal{N}_{0,t}$ with $\mathcal{N}_{0,0} = \delta_0$.
From $\mathcal{N}_{0,t} * \mathcal{N}_{0,s} = \mathcal{N}_{0,t+s}$ it follows that $\mathbb{P}_t * \mathbb{P}_s = \mathbb{P}_{t+s}$ and, therefore, by Lemma A.28, that $\kappa_t(x, A) := \mathbb{P}_t(A-x)$, $t \in T$ defines a translation invariant Markovian semigroup.
Hence, Corollary A.26 together with setting $\mu = \delta_0$ gives the unique existence of a probability measure $\mathbb{P}^{\delta_0}$ on $(\Omega, \mathcal{B})$ for which

$$\mathbb{P}^{\delta_0} \circ \pi_J^{-1} = \delta_0 \otimes \bigotimes_{k=0}^{n-1} \kappa_{t_{k+1}-t_k}$$

holds. For every $A \in \mathcal{B}(\mathbb{R})$ and each $t \in T$ it follows that

$$\mathbb{P}^{\delta_0}_{B_t}(A) = \mathbb{P}^{\delta_0}(B_t \in A) = \int \kappa_t(x,A)\delta_0(dx) = \int \mathcal{N}_{0,t}(A-x)\delta_0(dx) = \mathcal{N}_{0,t}(A),$$

hence, that $B_0 \sim \delta_0$, which implies $B_0 = 0$ $\mathbb{P}^{\delta_0}$-almost surely, and $B_t \sim \mathcal{N}_{0,t}$, for $t > 0$. Furthermore, from Theorem A.29, we see that the process $(B_t)_{t\in T}$ on $(\Omega, \mathcal{B}, \mathbb{P}^{\delta_0})$ has stationary and independent increments and additionally that $\mathbb{P}^{\delta_0}_{B_t - B_s} = \mathbb{P}_{t-s} = \mathcal{N}_{0,t-s}$ holds, for $t > s$. This can be written as

$$\mathbb{P}^{\delta_0}_{B_t-B_s} = \mathcal{N}_{0,t-s} = \sqrt{t-s}\ \mathcal{N}_{0,1} = \sqrt{t-s}\ \mathbb{P}_1 = \sqrt{t-s}\ \mathbb{P}^{\delta_0}_{B_1}.$$

For $n \in \mathbb{N}$ define $C_n := \mathbb{E}(B_1^{2n}) = \frac{(2n)!}{2^n n!} < \infty$, which are the $2n$-th moments of the standard normal distribution $\mathcal{N}_{0,1}$. Then, for every $n \in \mathbb{N}$, it holds that

$$\mathbb{E}\left((B_t - B_s)^{2n}\right) = \mathbb{E}\left((\sqrt{t-s}\ B_1)^{2n}\right) = C_n|t-s|^n.$$

Let $n \geq 2$ and $\gamma \in \left(0, \frac{n-1}{2n}\right)$. From Theorem A.31 we get the existence of a version of B with paths that are locally Hölder-continuous of order γ. Since all continuous versions of a process are equivalent, B is Hölder-continuous of order γ for every $\gamma \in \left(0, \frac{n-1}{2n}\right)$ and each $n \geq 2$, hence for every $\gamma \in \left(0, \frac{1}{2}\right)$. □

Remark. The d-dimensional Brownian motion is constructed as d-tuple of one dimensional, independent Brownian motions on the probability space $(\Omega^{\{1,\ldots,d\}}, \mathcal{F}^{\otimes\{1,\ldots,d\}}, \bigotimes_{i=1}^{n} \mathbb{P})$, where $\bigotimes_{i=1}^{n} \mathbb{P}$ denotes the unique product measure for a finite family of σ-finite measures (see Theorem A.17).

Remark. Parts of this construction, namely, everything except the continuity of the paths, holds for any family of probability measures $(\mathbb{P}_t)_{t\in T}$ with the property $\mathbb{P}_t * \mathbb{P}_s = \mathbb{P}_{t+s}$, for all $s, t \in T$. The Normal, Gamma, Cauchy, Binomial, negative Binomial and Poisson distribution have this property (see [K], Page 291), and therefore that these distributions omit a similar process. In particular, the existence of the Poisson process is clear from this construction.

Fourier Transform and Characteristic Function

Definition A.33.
Let μ be a probability measure on $\mathbb{R}^n$. The **characteristic function** or **Fourier transform** $f : \mathbb{R}^n \to \mathbb{C}$ of μ is defined by

$$f(\xi) = \int_{\mathbb{R}^n} e^{i(\xi,x)} d\mu(x).$$

Theorem A.34.
If μ and ν are measures on $\mathbb{R}^n$ with the same Fourier transform or characteristic function $f : \mathbb{R}^n \to \mathbb{C}$, then $\mu = \nu$.

Proof. see [D1], Page 303. □

Lemma A.35. (Bôchner)
A continuous function $f : \mathbb{R}^n \to \mathbb{C}$ is the characteristic function of a Borel probability measure $\mathbb{P}$ on $\mathbb{R}^n$, if and only if f is positive semidefinite and $f(0) = 1$.

Proof. see [Ka], Page 377. □

Lemma A.36. (Lévy inversion formula)
Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, $X : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a random variable with distribution $\mathbb{P}$ and Fourier transform $\phi_{\mathbb{P}}$. For all $a, b \in \mathbb{R}$, with $-\infty < a < b < \infty$, it holds that

$$\lim_{c\to\infty} \int_{-c}^{c} \frac{e^{-iat} - e^{-ibt}}{2\pi i t} \phi_{\mathbb{P}}(t) dt = \mathbb{P}(a < X < b) + \frac{\mathbb{P}(X = a) + \mathbb{P}(X = b)}{2}.$$

Proof. see [A] □

Theorem A.37. (Tonelli's theorem)
Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ-finite and let $f : \Omega_1 \times \Omega_2 \to [0, \infty]$ be a $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function or $f \in \mathcal{L}^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$. Then,

$$\int f d(\mu_1 \otimes \mu_2) = \int \int f(x,y) d\mu_1(x) d\mu_2(y) = \int \int f(x,y) d\mu_2(y) d\mu_1(x)$$

holds and $\int f(x,y) d\mu_1(x)$ is defined for μ_2-almost all y and $\int f(x,y) d\mu_2(y)$ for μ_1-almost all x.

Proof. see [D1], Page 137. □

Convex Geometry and Related Theorems

Theorem A.38.
Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and $W \subset H$ be a closed subspace. Then,

$$\forall z \in H \; \exists! \; x \in W \; \exists! \; y \in W^{\perp} : z = x + y,$$

where $W^{\perp} := \{w \in H : \langle w, v \rangle = 0 \; \forall v \in W\}$.

Definition A.39. (Orthogonal projection)
Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and $W \subset H$ be a closed subspace. The **orthogonal projection** P_W onto W is defined as the map

$$\begin{aligned} P_W : H &\to H \\ z &\mapsto x, \end{aligned}$$

where x is as in Theorem A.38.

Remark. Let $u \in \mathbb{R}^M$ and $W := \operatorname{span}\{u\}$. We define rank-1 orthogonal projections as $u \otimes u := P_W$. We have that

$$\forall x \in \mathbb{R}^M : (u \otimes u)(x) = \langle u, x \rangle u.$$

Lemma A.40.
Let W be an n-dimensional subspace of $\mathbb{R}^M$ and let g be the standard Gaussian random vector in $\mathbb{R}^M$. Then, $P_W g$ is the standard Gaussian random vector in W.

Proof. Let $\{f_1, ..., f_n\}$ be an orthonormal basis in W. Then, we can extend this basis with Gram-Schmidt to an orthonormal basis $\{f_1, ..., f_M\}$ in $\mathbb{R}^M$. Let now g_i, $i = 1, ..., M$, be standard Gaussian random variables. Then, the

standard Gaussian vector in $\mathbb{R}^M$ can be written as $g = \sum_{i=1}^M g_i f_i$.
For the orthogonal projection P_W we have

$$P_W g = P_W \left(\sum_{i=1}^M g_i f_i \right) = \sum_{i=1}^n g_i f_i,$$

which is the standard Gaussian vector in W. □

Theorem A.41. (Dual Sudakov minoration)
Let $(g_i)_{i=1,\dots,n}$ be the standard Gaussian vector on $\mathbb{R}^M$. Let T be a symmetric convex body in $\mathbb{R}^M$ with elements $t = (t_i)_{i=1,\dots,M} \in T$ and $X = (X_t)_{t\in T}$ be a Gaussian process on $\mathbb{R}^M$ defined by $X_t = \sum_{i=1}^M g_i t_i$.
Let $T^\circ := \{x \in \mathbb{R}^M : \forall y \in T : \langle x, y\rangle \leq 1\}$ be the polar of T and denote by $N(A, B)$ the number of translates of the set B by elements of the set A necessary to cover A. Then

$$\sup_{\varepsilon>0} \varepsilon \sqrt{\log N(B_2^M, \varepsilon T^\circ)} \leq C \mathbb{E} \sup_{t\in T} |X_t|.$$

Proof. see [L-T], page 82. □

Remark.

(i) To apply this theorem to the ε-Entropy of B_2^M in $\mathbb{R}^M$ w.r.t. some norm $\|\cdot\|$ and corresponding unit ball B^M note that

$$N(B_2^M, \varepsilon B^M) = N(B_2^M, \|\cdot\|, \varepsilon).$$

(ii) For a convex body $K \subseteq \mathbb{R}^M$, i.e., K is a convex and compact set with nonempty interior, and a subspace $E \subseteq \mathbb{R}^M$ the following holds

$$(K \cap E)^\circ = P_E(K^\circ)$$

(iii) In Banach space $X = (\mathbb{R}^M, \|\cdot\|)$ with unit ball B_X we have

$$B_{X^*} = \{y \in \mathbb{R}^M : \forall x \in B_X : \langle x, y\rangle \leq 1\} = (B_X)^\circ.$$

So the unit ball of the dual space X^* is the polar body of B_X, $(B_X)^\circ$. This can be seen by the Riesz representation theorem.

Theorem A.42. (Subgaussian tail estimate)
Let $(\alpha_i)_{i\in\mathbb{N}}$, $\alpha_i \in \mathbb{R}$ for all $i \in \mathbb{N}$, and let $(\varepsilon_i)_{i\in\mathbb{N}}$ be a sequence of independent Rademacher random variables, i.e., taking values in $\{-1, 1\}$ with probability 1/2 each. Then for every $t > 0$ we have

$$\mathbb{P}\left(\left| \sum_{i\in\mathbb{N}} \varepsilon_i \alpha_i \right| > t \right) \leq 2 \exp\left(-\frac{1}{2} \frac{t^2}{\sum_{i\in\mathbb{N}} \alpha_i^2} \right).$$

Proof. see [L-T], page 92. □

Lemma A.43.
Let Z_i, $i = 1, ..., M$, be independent folded standard normal distributed random variables on $\mathbb{R}^+$, i.e., $Z_i = |X_i|$ with $X_i \sim \mathcal{N}(0,1)$ for $i = 1, ..., M$ and independent. Then, there exists a constant $C > 0$ such that

$$\mathbb{E} \max_{i=1,...,M} Z_i \leq C\sqrt{\log M}.$$

Proof. Let $Z := \max_{i=1,...,M} Z_i$. Then we have

$$\exp(t\mathbb{E}Z) \overset{Jensen}{\leq} \mathbb{E}\exp(tZ) = \mathbb{E} \max_{i=1,...,M} \exp(tZ_i) \leq \sum_{i=1}^{M} \mathbb{E}\exp(tZ_i).$$

The moment generating function of a folded standard normal distributed random variable Z_i is

$$\mathbb{E}\exp(tZ_i) = 2e^{t^2/2}\left(\frac{1}{2} - \frac{1}{2}erf\left(\frac{-t}{\sqrt{2}}\right)\right) \leq 2e^{t^2/2},$$

where $erf(t) := \frac{2}{\sqrt{\pi}}\int_0^t e^{-x^2}dx$ is the error-function.
Using this, we get

$$\exp(t\mathbb{E}Z) \leq 2Me^{-t^2/2},$$

which we can rearrange as

$$\mathbb{E}Z \leq \frac{\log 2}{t} + \frac{\log M}{t} + \frac{t}{2}.$$

With the choice of $t = \sqrt{2\log M}$ we see that

$$\mathbb{E}Z \leq \sqrt{2}\sqrt{\log M} + \frac{\log 2}{\sqrt{2\log M}} = \sqrt{2}\sqrt{\log M} + \log 2\frac{\sqrt{\log M}}{\log M},$$

which we further estimate by the inequality $\log M \geq \log 2$, for all $M \geq 2$, as

$$\leq \sqrt{2}\sqrt{\log M} + \sqrt{\log M} = (\sqrt{2}+1)\sqrt{\log M}.$$

□

Lemma A.44.
Let $Z_i \sim \mathcal{N}(0,1)$, $i = 1, ..., M$, independent. Let $a_1, ..., a_M \in \mathbb{R} \setminus \{0\}$. Then, for all $j = 1, ..., M$,

$$\sum_{i=1}^{M} a_i Z_i \overset{\mathcal{D}}{=} \left(\sum_{i=1}^{M} a_i^2\right)^{1/2} Z_j.$$

Proof.
We have $a_i Z_i \sim \mathcal{N}(0, a_i^2)$ and therefore $\sum_{i=1}^{M} a_i Z_i \sim \mathcal{N}(0, \sum_{i=1}^{M} a_i^2)$.
On the other hand we know

$$\left(\sum_{i=1}^{M} a_i^2 \right)^{1/2} Z_j \sim \mathcal{N} \left(0, \sum_{i=1}^{M} a_i^2 \right)$$

□

Bibliography

[A] G. Alsmeyer. *Wahrscheinlichkeitstheorie, Skripten zur Mathematischen Statistik Nr.30.*
Münster, 2005.

[B] H. Bauer. *Wahrscheinlichkeitstheorie.* 4. Auflage.
Walter de Gruyter, Berlin, New York, 1991.

[B-D] J. Bretagnolle, D. Dacunha-Castelle. *Applications de l'étude de certains formes linéaires aléatoires au plongement d'espaces de Banach dans les espaces L_p.*
Ann. Sci. École Norm. Sup. (4) 2, 1969.

[Ba] K.M. Ball. *Ellipsoids of maximal volume in convex bodies.*
Geometriae Dedicata, February 1992, Volume 41, Issue 2, pp. 241-250.

[D1] R.M. Dudley. *Real Analysis and Probability.*
Cambridge University Press, Cambridge, 2004.

[D2] R.M. Dudley. *The size of compact subsets of Hilbert spaces and continuity of Gaussian processes.*
J. Functional Analysis, Volume 1, pp. 290-330, 1967.

[F] X. Fernique. *Des résultats nouveaux sur les processus gaussiens.*
Séminar de probabilités de Strasbourg, Volume 9, pp. 318-335, 1975.

[J] F. John. *Extremum problems with inequalities as subsidiary conditions.*
Courant Anniversary Volume, Interscience, New York, pp. 187-204, 1948.

[K] A. Klenke. *Wahrscheinlichkeitstheorie.*
Springer-Verlag, Berlin Heidelberg, 2006.

[Ka] T. Kawata. *Fourier Analysis in Probability Theory.*
Academic Press, New York and London, 1972.

[Kas-Tz] B. Kashin, L. Tzafriri. *Some remarks on the restrictions of operators to coordinate subspaces.*
Preprint no. 12, Hebrew University of Jerusalem, Institute of Mathematics, Jerusalem (Israel), 1994 , 14 pp.

[Ko] A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitstheorie.*
Springer, Berlin, 1933.

[L] J.R. Lee. *Notes on Gaussian processes and majorizing measures.*
https://homes.cs.washington.edu/~jrl/mm.pdf.

[L-S] H.J. Landau, L.A. Shepp. *On the supremum of Gaussian processes.*
Sankhya A., Volume 32, pp. 369-378, 1970.

[L-T] M. Ledoux, M. Talagrand. *Probability in Banach spaces.*
Ergeb. Math. Grenzgeb., 3. Folge, vol. 23, Springer, Berlin, 1991.

[M-P] P. Mörters, Y. Peres. *Brownian Motion.*
Cambridge University Press, New York, 2010.

[P] J. Prochno. *Charakterisierung von Teilräumen von* L_p, 2008.

[P-T] A. Pajor, N. Tomczak-Jaegermann. *Subspaces of small codimension of finite-dimensional Banach spaces.*
Proceedings of the American Mathematical Society, 1986, Volume 97, Number 4, pp. 637-642.

[R1] M. Rudelson. *Almost Orthogonal Submatrices of an Orthogonal Matrix.*
Israel Journal of Mathematics, December 1999, Volume 111, Issue 1, pp. 143-155.

[R2] M. Rudelson. *Contact Points of Convex Bodies.*
Israel Journal of Mathematics, December 1997, Volume 101, Issue 1, pp. 93-124.

[R3] M. Rudelson. *Approximate John's Decomposition.*
Geometric aspect of Functional Analysis, Isreal Seminar (GAFA) 1992-1994, Volume 77 of the series Operator Theory Advances and Applications, pp. 245-249.

[S] K.D. Schmidt. *Maß und Wahrscheinlichkeit*, 2. Auflage.
Springer-Verlag, Berlin Heidelberg, 2011.

[S-S-B] N. Srivastava, D. Spielman, J. Batson. *Twice-Ramanujan Sparsifiers.* Society for Industrial and Applied Mathematics, Vol. 56, No. 2, pp. 315-334, 2009.

[T1] M. Talagrand. *Majorizing Measures: The Generic Chaining.* The Annals of Probability 1996, Vol. 24, Nr.3, Institute of Mathematical Statistics.

[T2] M. Talagrand. *Regularity of Gaussian processes.* Acta Mathematica, December 1987, Vol. 159, Issue 1, pp. 99-149.