TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

institute of telecommunications

DIPLOMARBEIT

Information Theory on Rectifiable Sets

ausgeführt am
Institute of Telecommunications
der Technischen Universität Wien

unter Anleitung von
Dipl.-Ing. Günther Koliander
Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Franz Hlawatsch

von

Georg Pichler
Wischerstr. 20
A-4040 Linz
georg.pichler@ztpichler.at

Wien, November 2013 ______________________

Acknowledgements

I want to express my gratitude to my supervisor, Günther Koliander, for his sincere critique, many valuable discussions and above all for so generously donating his time.

Also I wish to thank Franz Hlawatsch for his advice and helpful comments.

Finally, I would especially like to thank my family for their continuous support and for enabling me to pursue my studies of telecommunications.

Abstract

Entropy and—maybe even more so—mutual information are invaluable tools for analyzing properties of probability distributions, especially in coding theory. While there are general definitions for both concepts available for arbitrary probability distributions, these tend to be hard to work with and the literature (e.g., [CT06]) focuses on either discrete, or continuous random variables. In this thesis we extend the theory to singular probability distributions on suitably "smooth" lower-dimensional subsets of Euclidean space, where no p.d.f. or p.m.f. is readily available. We choose those prerequisites carefully in order to retain most properties from the discrete and/or continuous case.

The mathematical framework, this work is built upon, is the field of geometric measure theory. In particular, we make extensive use of the material found in [Fed69]. As it is the study of geometric properties of measures, and thereby closely related to probability theory as well, geometric measure theory proves fruitful for analyzing information theoretic properties of probability measures, when geometric restrictions are imposed.

We consider a random variable X on Euclidean space. The distribution (i.e. the induced measure) of X is required to be absolutely continuous with respect to the m-dimensional Hausdorff measure and to be concentrated on an m-dimensional rectifiable set E, i.e., the complement of E is a set of measure zero. Under these conditions we obtain expressions for the entropy $h(X)$ and develop the mutual information $I(X;Y)$ for two random variables when the combined random variable (X,Y) satisfies similar constraints. We give integral expressions for these quantities and show how to manipulate them using results from geometric measure theory. Another central result is the proof of the relation $I(X;Y) = h(X) + h(Y) - h(X,Y)$ between mutual information and entropy for our newly introduced definitions.

This work is intended as a theoretical starting point for further investigations. Possible applications are, e.g., the information theoretic treatment of sparse sources in source coding or of the vector interference channel in channel coding. In both examples singular distributions on "smooth" lower-dimensional subsets play a pivotal role. While this work was conducted with applications in coding theory in mind, the presented framework is inherently measure theoretic and might, therefore, be applied in other areas as well.

Zusammenfassung

Entropie und Transinformation stellen wichtige Werkzeuge dar, um Eigenschaften von Wahrscheinlichkeitsverteilungen, speziell in der Kodierungstheorie, zu studieren. Obwohl allgemeine Definitionen beider Konzepte für beliebige Wahrscheinlichkeitsverteilungen zur Verfügung stehen, sind diese schwer anwendbar, was zur Folge hat, dass sich der Großteil der Literatur (z.B. [CT06]) auf die Analyse von diskreten oder kontinuierlichen Zufallsvariablen beschränkt. In dieser Arbeit erweitern wir die Theorie von Entropie und Transinformation und betrachten dabei singuläre Verteilungen auf geeignet "glatten", niedrig dimensionalen Untermengen des Euklidischen Raumes, wo keine Wahrscheinlichkeitsfunktion oder Wahrscheinlichkeitsdichtefunktion zur Verfügung steht. Die Voraussetzungen, die an die Wahrscheinlichkeitsverteilungen gestellt werden, gestalten sich derart, dass viele wohlbekannte Eigenschaften der (differentiellen) Entropie des diskreten (kontinuierlichen) Falls erhalten bleiben.

Das Gebiet der geometrischen Maßtheorie bietet den Rahmen für unsere Überlegungen. Im Speziellen greifen wir häufig auf [Fed69] zurück. Die geometrische Maßtheorie beschäftigt sich mit den geometrischen Eigenschaften von Maßen und steht damit in enger Beziehung zur Wahrscheinlichkeitstheorie. Sie ist geeignet, Wahrscheinlichkeitsmaße unter geometrischen Einschränkungen informationstheoretisch zu untersuchen.

Wir betrachten eine Zufallsvariable X im Euklidischen Raum. Die Verteilung (d.h. das induzierte Maß) von X ist absolut stetig bezüglich des m-dimensionalen Hausdorff-Maßes und konzentriert auf einer m-dimensional rektifizierbaren Menge E, d.h. das Maß des Komplements von E ist Null. Unter diesen Voraussetzungen leiten wir Ausdrücke für die Entropie $h(X)$ her und betrachten außerdem die Transinformation $I(X;Y)$, wenn ähnliche Voraussetzungen für die kombinierte Zufallsvariable (X,Y) gelten. Für die genannten Größen werden Integralausdrücke präsentiert und wir zeigen, wie diese mit Mitteln der geometrischen Maßtheorie umgeformt werden können. Ein weiteres zentrales Resultat ist der Beweis der Relation $I(X;Y) = h(X) + h(Y) - h(X,Y)$ zwischen der Transinformation und der Entropie für die hier neu eingeführten Definitionen dieser Größen.

Diese Arbeit soll als theoretischer Ausgangspunkt für weitere Untersuchung dienen. Mögliche Anwendungsbereiche sind z.B. die informationstheoretische Untersuchung dünnbesetzter Quellen in der Quellenkodierung oder der Vektor-Interferenz Kanal in der Kanalkodierung. In beiden Beispielen spielen singuläre Verteilungen auf "glatten", niedrig dimensionalen Untermengen eine wichtige Rolle. Obwohl diese Arbeit auf Anwendungen in der Kodierungstheorie ausgerichtet ist, ist sie doch im Grunde der Maßtheorie zuzurechnen. Anwendungen in anderen Bereichen sind daher durchaus denkbar.

Contents

1. Introduction **1**
1.1. Motivation 1
1.2. Thesis Outline 2

2. Fundamentals and Notation **3**
2.1. Measures 3
2.2. Lebesgue Integration 6
2.3. The Radon-Nikodým Theorem 8
2.4. Fubini's Theorem 11
2.5. Probability Measures 11
2.6. Geometric Measure Theory 12

3. Rectifiable Measures and Entropy **17**
3.1. Rectifiable Measures 17
3.2. Densities 20
3.3. Entropy 24
3.4. Image of a Lipschitz Function 25

4. Combined Rectifiable Measures and Mutual Information **31**
4.1. Combined Rectifiable Measures 31
4.2. Mixtures 34
4.3. Mutual Information 38
4.4. Connection between Mutual Information and Entropy 43

5. Conclusions **51**

A. Appendix **53**
A.1. Proof of Lemma 2.6.8 53
A.2. Proof of Lemma 2.6.13 54
A.3. On the Decomposition of Mixtures 57

Index **61**

Bibliography **63**

Chapter 1.

Introduction

1.1. Motivation

This text is dedicated to developing a generalization of entropy and mutual information of random variables in Euclidean spaces, which assume values on lower-dimensional subsets. In what follows we will illustrate where the material in this text may be applied.

Let X be a random variable, assuming values in an alphabet $\mathscr{X}$. We will first focus on discrete random variables and define their entropy. Thus, let $\mathscr{X}$ be a finite or countable set. Then X is fully described by the probabilities $p_x := P(\{x\}) = P(X = x)$, $x \in \mathscr{X}$. Following [CT06, Section 2.1], we define the entropy of X as

$$H(X) := E_X[-\log p_X] = -\sum_{x\in\mathscr{X}} p_x \log p_x, \tag{1.1}$$

where $E_X[f(X)]$ denotes the expectation of $f(X)$ with respect to (w.r.t.) the random variable X. For continuous random variables the situation looks similar. For $\mathscr{X} = \mathbb{R}$ let X be a continuous random variable with density $p(x)$.[1] Referring to [CT06, Section 8.1], the differential entropy $h(X)$ is defined as

$$h(X) := E_X[-\log p(X)] = -\int_{-\infty}^{\infty} p(x) \log p(x)\,\mathrm{d}x. \tag{1.2}$$

Comparing (1.1) and (1.2), one notices that these equations look remarkably similar. We shall reformulate them in measure theoretic terms, starting with discrete entropy.

Let μ be the probability measure[2] on $(\mathscr{X}, \mathfrak{P}(\mathscr{X}))$, induced by a discrete random variable X, i.e., $|\mathscr{X}| \leq \aleph_0$. Denoting by ζ the counting measure on $\mathscr{X}$, we can write

$$H(X) = -\int \log \frac{\mathrm{d}\mu}{\mathrm{d}\zeta}\,\mathrm{d}\mu = -\int \frac{\mathrm{d}\mu}{\mathrm{d}\zeta} \log \frac{\mathrm{d}\mu}{\mathrm{d}\zeta}\,\mathrm{d}\zeta, \tag{1.3}$$

where $\frac{\mathrm{d}\mu}{\mathrm{d}\zeta}$ is the Radon-Nikodým derivative. Similarly, for the differential entropy (1.2), let λ denote the Lebesgue measure, and μ the measure induced by a continuous random variable on

[1] We assume $\mathscr{X} = \mathbb{R}$ for simplicity. The same results would hold for $\mathscr{X} = \mathbb{R}^M$ with $M \in \mathbb{N}$.

[2] We will use some concepts and notation in this section that are defined later in the text. The reader is referred to Chapter 2 and in particular Table 2.1.

$\mathscr{X} = \mathbb{R}$. Then

$$h(X) = -\int \log \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \,\mathrm{d}\mu = -\int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \log \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \,\mathrm{d}\lambda. \tag{1.4}$$

As one can see, we can embed both (1.3) and (1.4) into a common formulation: Let μ be a probability measure on $\mathscr{X}$ and ν some (different) measure on the same measure space such that the Radon-Nikodým derivative $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ exists. We can define the entropy of μ w.r.t. ν as

$$h_\nu(\mu) := -\int \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \,\mathrm{d}\mu = -\int \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \,\mathrm{d}\nu.$$

In the case of a discrete random variable, ν is conveniently chosen to be the counting measure and in the case of a continuous random variable, the Lebesgue measure is a reasonable choice. But what happens if μ is neither continuous nor discrete? The remainder of this text is devoted to answering this question for a certain type of such probability measures, namely ones concentrated on "smooth" lower-dimensional subsets of $\mathbb{R}^M$.

1.2. Thesis Outline

The rest of this thesis is organized as follows.

- Chapter 2 gives an overview of the mathematical background, to the extent needed in this text, and introduces the necessary notation. In Sections 2.1 to 2.5 an overview of measure theory is given, while Section 2.6 covers more advanced topics in geometric measure theory. In particular, Section 2.6 presents the novel concept of $\mathscr{B}$-countably m-rectifiable sets, which will be used extensively in subsequent chapters.

- Based on $\mathscr{B}$-countably m-rectifiable sets, our definition of $\mathscr{B}$-countably m-rectifiable measures is established in Chapter 3 and their properties are analyzed. Furthermore, we treat the density of $\mathscr{B}$-countably m-rectifiable measures with respect to the Hausdorff measure in Section 3.2 and give a transformation law for the entropy of a random variable under a (locally) Lipschitz function in Section 3.4.

- Chapter 4 states the main results of this thesis. The concepts that were introduced prior to this chapter are generalized to product spaces and mixtures of random variables, comprised of different dimensions. We show how to obtain mutual information in this context and prove a familiar, though intricate, connection between mutual information and entropy in Section 4.4.

- Final conclusions are drawn in Chapter 5 and we note some areas for possible further investigation.

Chapter 2.

Fundamentals and Notation

The analysis in this text involves measure theory and, in particular, geometric measure theory. A very extensive treatment of geometric measure theory can be found in [Fed69]. Although it does not only cover geometric measure theory in great detail, but also includes a full treatment of the foundations of measure theory in general in Section 2, the book [Fed69] can only be recommended to experienced readers. Readers not so familiar with measure theoretic concepts can, e.g., resort to [AL06].

Most relevant results from the literature will be restated in this chapter with references to proofs and further treatment. Some notation, which is not introduced in the text, but largely follows known practice may be found in Table 2.1.

2.1. Measures

The material in this section is taken from [AL06], where more detailed information on the subject can be found. A word of caution is in order as we will extensively refer to [Fed69] in subsequent sections. Federer uses a different, but equivalent definition of measures. He defines a measure as what is usually called an *outer measure*. An account of the differences can be found in [KP08, Remark 1.2.6].

Definition 2.1.1. *[AL06, Definitions 2.1.1 and 1.1.2] A pair* $(\Omega, \mathfrak{S})$ *is called a* measurable space *if* Ω *is a nonempty set and* $\mathfrak{S} \subseteq \mathfrak{P}(\Omega)$ *is a* σ-algebra *on* Ω*, i.e.,*

(i) $\varnothing \in \mathfrak{S}$,

(ii) $A \in \mathfrak{S} \implies A^c \in \mathfrak{S}$*, and*

(iii) $A_i \in \mathfrak{S}$ *for all* $i \in \mathbb{N} \implies \bigcup_{i \in \mathbb{N}} A_i \in \mathfrak{S}$.

An important σ-algebra on a metric space, e.g., $\Omega = \mathbb{R}$, is the σ-algebra of Borel sets. It is the smallest (w.r.t. inclusion) σ-algebra containing all open sets.

Definition 2.1.2. *[AL06, Definition 1.1.4] Let* (Ω, d) *be a metric space with metric* $d : \Omega \times \Omega \to [0, \infty)$*. Let* $\mathscr{T}$ *denote the* topology *on* Ω *induced by* d*, i.e., the set of all open subsets of* Ω*. Then*

Symbol	Definition
$\mathfrak{P}(A)$	Power set of the set A
$\varnothing$	Empty set $\varnothing = \{\}$
A^c	Complement of the set A
$\mathbb{R}^+ := [0,\infty)$	Nonnegative real numbers
$\bar{\mathbb{R}}^+ := [0,\infty]$	Nonnegative extended real numbers
$\bar{\mathbb{R}} := [-\infty,\infty]$	Extended real numbers
$f^{-1}(A)$	Preimage of the set A under the function f
$\mathbb{N} := \{1,2,3,\dots\}$	Set of natural numbers
$\mathbb{1}_A$	Indicator function of the set A
id_A	Identity function on the set A
$f \circ g$	Composition of functions f and g
$\lvert A\rvert$	Cardinality of the set A
$\aleph_0 := \lvert\mathbb{N}\rvert$	Cardinality of the natural numbers
$\bigcup\mathfrak{C} / \bigcap\mathfrak{C}$	Union/Intersection: $\bigcup\mathfrak{C} := \bigcup_{C\in\mathfrak{C}} C$ / $\bigcap\mathfrak{C} := \bigcap_{C\in\mathfrak{C}} C$
$\mathrm{dist}(\cdot,\cdot)$	Euclidean distance
$\mathrm{diam}(E)$	Diameter of a set: $\mathrm{diam}(E) := \sup_{x,y\in E}\mathrm{dist}(x,y)$
$B_r(x)$	Open ball: $B_r(x) = \{y : \mathrm{dist}(x,y) < r\}$
$\overline{A}$	Closure of the set A

Table 2.1.: Summary of notation

define the Borel σ-algebra *on* Ω *as*[1]

$$\mathscr{B}(\Omega) := \bigcap\{\mathfrak{S} \subseteq \mathfrak{P}(\Omega) : \mathfrak{S} \textit{ is } \sigma\textit{-algebra}, \mathscr{T} \subseteq \mathfrak{S}\}.$$

If the underlying metric space is the real line with the Euclidean distance, i.e., $(\Omega,d) = (\mathbb{R},\mathrm{dist}(\cdot,\cdot))$, we will use $\mathscr{B} := \mathscr{B}(\mathbb{R})$ for short to denote the Borel sets. Definition 2.1.2 is in particular also applicable to the product of metric spaces, when equipped with the product metric, e.g., higher dimensional Euclidean space $\mathbb{R}^M$, where $M \in \mathbb{N}$ and we will use $\mathscr{B}^M := \mathscr{B}(\mathbb{R}^M)$ for short.

Now we introduce measures on measurable spaces. A measure is a function from the σ-algebra $\mathfrak{S}$ to $\bar{\mathbb{R}}^+$ with certain properties.

Definition 2.1.3. *[AL06, Definitions 2.1.1 and 1.3.2] A triple* $(\Omega,\mathfrak{S},\mu)$ *is called a* measure space *if* $(\Omega,\mathfrak{S})$ *is a measurable space and the function* $\mu : \mathfrak{S} \to \bar{\mathbb{R}}^+$ *satisfies the following properties:*

(i) $\mu(\varnothing) = 0$ *and*

(ii) (σ-additivity) *for a countable family* $\{A_i\}_{i\in\mathbb{N}}$ *of pairwise disjoint sets* $A_i \in \mathfrak{S}$*, we have* $\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i)$.

[1] For a proof that this is a well-defined σ-algebra see [AL06, Definition 1.1.3].

The function μ *is called a* measure *on the measurable space* $(\Omega, \mathfrak{S})$.

Some important properties arising from Definition 2.1.3 will be summarized in the following proposition.

Proposition 2.1.4. *Let* $(\Omega, \mathfrak{S}, \mu)$ *be a measurable space as in Definition 2.1.3. Then the following properties hold:*

(i) (σ-subadditivity)*[AL06, Proposition 1.2.3(ii)] For a countable family of (not necessarily disjoint) sets* $\{A_i\}_{i\in\mathbb{N}}$, $A_i \in \mathfrak{S}$,

$$\mu\left(\bigcup_i A_i\right) \leq \sum_i \mu(A_i).$$

(ii) (monotonicity)*[AL06, Proposition 1.2.2(i)] For two sets* $A, B \in \mathfrak{S}$ *with* $A \subseteq B$ *we have* $\mu(A) \leq \mu(B)$.

(iii) (monotone continuity from below)*[AL06, Proposition 1.2.1(iii)$'_b$] For a countable family of sets* $\{A_i\}_{i\in\mathbb{N}}$, *with* $A_i \in \mathfrak{S}$ *and* $A_i \subseteq A_{i+1}$ *for all* $i \in \mathbb{N}$,

$$\mu\left(\bigcup_i A_i\right) = \lim_{i\to\infty} \mu(A_i).$$

An important measure on $(\mathbb{R}, \mathscr{B})$ is the *Lebesgue measure*, denoted λ. For a detailed account of its construction refer to [AL06, Definition 1.3.8]. Note that we do not consider the complete Lebesgue measure, but its restriction to the Borel sets, also called *Borel-Lebesgue measure* by some authors. Furthermore, we will make use of the higher dimensional Lebesgue measure on $\mathbb{R}^M$ and denote it by λ^M, where $M \in \mathbb{N}$. Again, a detailed account of the construction can be found in the literature, e.g., in [AL06, Section 1.3.3].

Definition 2.1.5. *The* counting measure, *denoted by* ζ, *is defined as*

$$\zeta(A) := \begin{cases} |A|, & A \text{ is finite} \\ \infty, & \text{otherwise} \end{cases}$$

for any set A.

2.2. Lebesgue Integration

This section is meant to serve as an introduction to Lebesgue integration and may be skipped by readers familiar with the concept. The text follows the introduction of the Lebesgue integral in [KP08, Sections 1.3.1 and 1.3.2], where the proofs of the stated theorems may be found.

Let $(\Omega, \mathfrak{S}, \mu)$ be a measure space throughout this section.

Definition 2.2.1. *[KP08, Definition 1.3.1] The term* μ-almost *can serve as an adjective or adverb in the following way:*

Let $\mathscr{P}(x)$ be a statement or formula that contains a free variable $x \in \Omega$. We say that $\mathscr{P}(x)$ holds for μ-almost every (μ-a.e.) $x \in \Omega$ *if*

$$\mu\Big(\{x \in \Omega : \mathscr{P}(x) \text{ is false}\}\Big) = 0.$$

If Ω is understood from the context, we simply say that $\mathscr{P}(x)$ holds μ-almost everywhere *(* μ-a.e.*).*

Definition 2.2.2. *[AL06, Definition 2.1.3] Given two measurable spaces $(\Omega_1, \mathfrak{S}_1)$ and $(\Omega_2, \mathfrak{S}_2)$, a function $f : \Omega_1 \to \Omega_2$ is called* $(\mathfrak{S}_1, \mathfrak{S}_2)$-measurable *if $f^{-1}(A) \in \mathfrak{S}_1$, whenever $A \in \mathfrak{S}_2$.*

Given a measure space $(\Omega_1, \mathfrak{S}_1, \mu)$, a measurable space $(\Omega_2, \mathfrak{S}_2)$ and a $(\mathfrak{S}_1, \mathfrak{S}_2)$-measurable function $f : \Omega_1 \to \Omega_2$, f and μ can be used to define a measure on $(\Omega_2, \mathfrak{S}_2)$ in a very natural way.

Definition 2.2.3. *[AL06, Definition 2.2.1] For a measure space $(\Omega_1, \mathfrak{S}_1, \mu)$ and a measurable space $(\Omega_2, \mathfrak{S}_2)$, let $f : \Omega_1 \to \Omega_2$ be a $(\mathfrak{S}_1, \mathfrak{S}_2)$-measurable function. The* induced measure μ_f *on $(\Omega_2, \mathfrak{S}_2)$ is defined by*

$$\mu_f(A) := \mu(f^{-1}(A))$$

for every $A \in \mathfrak{S}_2$.

Definition 2.2.4. *[KP08, Definition 1.3.8] By a nonnegative* simple function $f : \Omega \to \mathbb{R}^+$ *we mean a function that can be represented as a linear combination of indicator functions, i.e., f can be written as*

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}, \tag{2.1}$$

where $a_i \in \mathbb{R}^+$ and $A_i \subseteq \Omega$ for $n \in \mathbb{N}$ and for all $i \in \{1, \ldots, n\}$.

Lemma 2.2.5. *[KP08, Lemma 1.3.9] If $f : \Omega \to \mathbb{R}^+$ is $(\mathfrak{S}, \mathscr{B})$-measurable*[2]*, then there exists a*

[2] Noting that $\mathscr{B}$ is technically not a σ-algebra on $\mathbb{R}^+$, one can either interpret f as a function $\tilde{f} : \Omega \to \mathbb{R}$ and require $\tilde{f}$ to be $(\mathfrak{S}, \mathscr{B})$-measurable or require f to be $(\mathfrak{S}, \mathscr{B} \cap \mathbb{R}^+)$-measurable, where $\mathscr{B} \cap \mathbb{R}^+ := \{B \cap \mathbb{R}^+ : B \in \mathscr{B}\}$. As both approaches are equivalent and there is no danger of confusion, we will not distinguish between f and $\tilde{f}$.

sequence of $(\mathfrak{S}, \mathscr{B})$*-measurable, simple functions* $h_n : \Omega \to \mathbb{R}^+$, $n \in \mathbb{N}$ *such that*

(i) $0 \le h_1 \le h_2 \le \cdots \le f$*, and*

(ii) $\lim_{n\to\infty} h_n(x) = f(x)$*, for all* $x \in \Omega$*.*

Definition 2.2.6. *[KP08, Lemma 1.3.10 and 1.3.11] If* $f : \Omega \to \mathbb{R}$ *is a* $(\mathfrak{S}, \mathscr{B})$*-measurable function, then the* Lebesgue integral *of* f *w.r.t.* μ *is denoted by*

$$\int f \,\mathrm{d}\mu = \int f(x)\,\mathrm{d}\mu(x) \tag{2.2}$$

and is defined as follows:

(i) In case f *is a nonnegative, simple function (see Definition 2.2.4) written as in (2.1) with* $A_i \in \mathfrak{S}$*, we set*

$$\int f \,\mathrm{d}\mu := \sum_{i=1}^{n} a_i \mu(A_i).$$

(ii) If f *is a nonnegative function, we set*

$$\int f \,\mathrm{d}\mu := \sup\left\{\int h \,\mathrm{d}\mu : 0 \le h \le f, h \textit{ simple and } (\mathfrak{S}, \mathscr{B})\textit{-measurable}\right\}. \tag{2.3}$$

(iii) Otherwise, we define $f^+(x) := \max(f(x), 0)$ *and* $f^-(x) := \max(-f(x), 0)$ *and set*

$$\int f \,\mathrm{d}\mu := \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu \tag{2.4}$$

if at least one of the two $\int f^+ \,\mathrm{d}\mu$ *and* $\int f^- \,\mathrm{d}\mu$ *is finite. If* $\int f^+ \,\mathrm{d}\mu = \int f^- \,\mathrm{d}\mu = \infty$*, the integral* $\int f \,\mathrm{d}\mu$ *is undefined.*

Note in particular that (2.3), and thereby also (2.4), is well-defined as the set in (2.3) is non-empty by Lemma 2.2.5.

Remark 2.2.7. [KP08, Remark 1.3.2]

(i) For the purposes of measure and integration, two functions that are equal μ-almost everywhere are equivalent. This defines an equivalence relation.

(ii) For integration, it thus suffices if a function is defined μ-a.e. as it then belongs to a well-defined equivalence class w.r.t. equality μ-a.e. Thus, considering a function $f : D \subseteq \Omega \to \mathbb{R}$ where $\mu(\Omega \setminus D) = 0$, we can define, e.g., the function

$$\tilde{f} : \Omega \to \mathbb{R}$$

$$x \mapsto \begin{cases} f(x), & x \in D \\ 0, & \text{otherwise}, \end{cases}$$

which is equivalent to f.

Remark 2.2.8. On the right-hand side in (2.2) the dummy variable x is introduced. This notation will serve to distinguish which function is integrated w.r.t. which measure when dealing with nested integrals.

Definition 2.2.9. *[KP08, Lemma 1.3.11(1)] To integrate f over a set $A \in \mathfrak{S}$, we multiply f by the indicator function of A, i.e.,*

$$\int_A f \, \mathrm{d}\mu := \int \mathbb{1}_A f \, \mathrm{d}\mu.$$

We will summarize some properties of the Lebesgue integral in the following theorem.

Theorem 2.2.10. *[Fed69, Theorem 2.4.4(3,4,5)] Given two $(\mathfrak{S}, \mathscr{B})$-measurable functions $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$, such that the integrals $\int f \, \mathrm{d}\mu \in \bar{\mathbb{R}}$ and $\int g \, \mathrm{d}\mu \in \bar{\mathbb{R}}$ are defined (see Part (iii) of Definition 2.2.6). Then the following properties hold:*

(i) (Homogeneity) *For $c \in \mathbb{R}$, $\int cf \, \mathrm{d}\mu = c \int f \, \mathrm{d}\mu$.*

(ii) (Additivity) *$\int f + g \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$ if the sum is defined.*

(iii) (Monotonicity) *If $f(x) \geq g(x)$, μ-a.e. $x \in \Omega$, then $\int f \, \mathrm{d}\mu \geq \int g \, \mathrm{d}\mu$.*

2.3. The Radon-Nikodým Theorem

We will make extensive use of densities and product measures. To make these concepts precise, we have to restrict ourselves to σ-finite measures.

Definition 2.3.1. *[AL06, Definition 1.2.2] A measure space $(\Omega, \mathfrak{S}, \mu)$ is called* σ-finite *if there exists a countable family of sets $\{A_i\}_{i \in \mathbb{N}}$ in $\mathfrak{S}$ such that $\bigcup_i A_i = \Omega$ and $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$. A measure μ on a measurable space $(\Omega, \mathfrak{S})$ is called σ-finite if $(\Omega, \mathfrak{S}, \mu)$ is σ-finite.*

Definition 2.3.2. *[AL06, Definitions 4.1.1 and 4.1.2] Let μ and ν be two measures on the measurable space $(\Omega, \mathfrak{S})$.*

(i) We say that μ is absolutely continuous *w.r.t. ν, written $\mu \ll \nu$, if for all $A \in \mathfrak{S}$,*

$$\nu(A) = 0 \implies \mu(A) = 0.$$

(ii) *We call* μ *and* ν mutually singular, *written* $\mu \perp \nu$, *if there exists an* $A \in \mathfrak{S}$ *with*

$$\mu(A) = 0 \quad \text{and} \quad \nu(A^c) = 0.$$

Equipped with Definition 2.3.1 and Definition 2.3.2 we can state the Radon-Nikodým theorem and the Lebesgue decomposition theorem. Especially the Radon-Nikodým theorem is a basis of the subsequent considerations.

Theorem 2.3.3. *[AL06, Theorem 4.1.1] Let* $(\Omega, \mathfrak{S})$ *be a measurable space and let* μ *and* ν *be two* σ*-finite measures on* $(\Omega, \mathfrak{S})$.

(i) *(Lebesgue Decomposition Theorem). The measure* μ *can be uniquely decomposed as*

$$\mu = \mu_a + \mu_s,$$

where μ_a *and* μ_s *are* σ*-finite measures on* $(\Omega, \mathfrak{S})$*, such that* $\mu_a \ll \nu$ *and* $\mu_s \perp \nu$.

(ii) *(Radon-Nikodým Theorem). A nonnegative* $(\mathfrak{S}, \mathscr{B})$*-measurable function* $\frac{\mathrm{d}\mu_a}{\mathrm{d}\nu} : \Omega \to \mathbb{R}^+$ *exists, such that*

$$\mu_a(A) = \int_A \frac{\mathrm{d}\mu_a}{\mathrm{d}\nu} \,\mathrm{d}\nu \quad \text{for all } A \in \mathfrak{S}.$$

Furthermore, if a function $f : \Omega \to \mathbb{R}^+$ *satisfies*

$$\mu_a(A) = \int_A f \,\mathrm{d}\nu \quad \text{for all } A \in \mathfrak{S},$$

then $f = \frac{\mathrm{d}\mu_a}{\mathrm{d}\nu}$, ν*-a.e.*

Part (ii) of Theorem 2.3.3 states that $\frac{\mathrm{d}\mu_a}{\mathrm{d}\nu}$ is ν-almost unique. This result is mentioned in [AL06, p.118] and proved in Proposition 2.3.6.

Two properties of the Radon-Nikodým derivative we will need later on are as follows.

Proposition 2.3.4. *[AL06, Proposition 4.1.2(i,ii)] Let* μ_1, μ_2, *and* ν *be* σ*-finite measures on the measurable space* $(\Omega, \mathfrak{S})$ *and* $a, b \in \mathbb{R}^+$.

(i) *If* $\mu_1 \ll \mu_2$ *and* $\mu_2 \ll \nu$*, then* $\mu_1 \ll \nu$ *and*

$$\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu}(x) = \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}(x)\frac{\mathrm{d}\mu_2}{\mathrm{d}\nu}(x) \quad \text{for } \nu\text{-a.e. } x \in \Omega.$$

(ii) If $\mu_1 \ll \nu$ and $\mu_2 \ll \nu$, then $a\mu_1 + b\mu_2 \ll \nu$ and

$$\frac{\mathrm{d}(a\mu_1 + b\mu_2)}{\mathrm{d}\nu}(x) = a\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu}(x) + b\frac{\mathrm{d}\mu_2}{\mathrm{d}\nu}(x) \quad \textit{for } \nu\textit{-a.e. } x \in \Omega.$$

Definition 2.3.5. *Let $(\Omega, \mathfrak{S}, \mu)$ be a measure space and let $f : \Omega \to \mathbb{R}^+$ be a nonnegative $(\mathfrak{S}, \mathscr{B})$-measurable function, defined for μ-a.e. $x \in \Omega$. The measure $f\mu$ on $(\Omega, \mathfrak{S})$ is defined for all $A \in \mathfrak{S}$ as*

$$(f\mu)(A) := \int_A f \,\mathrm{d}\mu.$$

Proposition 2.3.6. *For a σ-finite measure space $(\Omega, \mathfrak{S}, \mu)$ and two nonnegative $(\mathfrak{S}, \mathscr{B})$-measurable functions $f : \Omega \to \mathbb{R}^+$ and $g : \Omega \to \mathbb{R}^+$, satisfying $f\mu = g\mu$, we have that*

$$g = f \quad \mu\textit{-a.e.} \tag{2.5}$$

Proof. First, we will prove (2.5) for a finite measure μ. If $A := \{x \in \Omega : f(x) \neq g(x)\}$, then A can be written as the disjoint union $A = A_1 \cup A_2$, where $A_1 = \{x \in \Omega : f(x) > g(x)\}$ and $A_2 = \{x \in \Omega : f(x) < g(x)\}$. Because addition is continuous and thus measurable and we can write, e.g., $A = (f - g)^{-1}(\{0\}^c)$, we have that $A, A_1, A_2 \in \mathfrak{S}$. We now define the sets $B_n \in \mathfrak{S}$ for $n \in \mathbb{N}$ as

$$B_n := \left\{x \in \Omega : g(x) + \frac{1}{n} < f(x) \text{ and } g(x) < n\right\}. \tag{2.6}$$

Note that by (2.6), $B_n \subseteq B_{n+1}$ for $n \in \mathbb{N}$, and also $\bigcup_{n=1}^{\infty} B_n = A_1$. We have that

$$\int_{B_n} g \,\mathrm{d}\mu \stackrel{(a)}{=} \int_{B_n} f \,\mathrm{d}\mu \stackrel{(b)}{\geq} \int_{B_n} \left(g + \frac{1}{n}\right) \mathrm{d}\mu \stackrel{(c)}{=} \int_{B_n} g \,\mathrm{d}\mu + \frac{1}{n}\mu(B_n), \tag{2.7}$$

where (a) follows from $f\mu = g\mu$, (b) follows from (2.6) and Part (ii) of Proposition 2.1.4, and (c) is a consequence of Parts (i) and (ii) of Theorem 2.2.10. As μ is assumed finite and $g < n$ on B_n, (2.7) gives us $\mu(B_n) = 0$, which yields

$$\mu(A_1) = \mu\left(\bigcup_n B_n\right) \stackrel{(a)}{=} \lim_{n\to\infty} \mu(B_n) = 0, \tag{2.8}$$

where Part (iii) of Proposition 2.1.4 was used in (a). Similarly one can show $\mu(A_2) = 0$.

For a σ-finite measure μ, applying (2.8) to the restricted measures $\mu|_{E_i}$, where $\{E_i\}_{i\in\mathbb{N}}$ is the covering of Ω by countably many sets of finite measure (see Definition 2.3.1) proves $\mu_{E_i}(A_1) = \mu(A_1 \cap E_i) = 0$ for all $i \in \mathbb{N}$. Thus, $\mu(A_1) = 0$ by Part (i) of Proposition 2.1.4. Analogously one proves $\mu(A_2) = 0$, which concludes the proof that $\mu(A) = \mu(A_1) + \mu(A_2) = 0$. □

2.4. Fubini's Theorem

Definition 2.4.1. *[AL06, Definition 5.1.1(c,d)] Given two measurable spaces* $(\Omega_1, \mathfrak{S}_1)$ *and* $(\Omega_2, \mathfrak{S}_2)$, *one defines their* product measurable space *as the measurable space*

$$(\Omega_1 \times \Omega_2, \mathfrak{S}_1 \times \mathfrak{S}_2),$$

where $\Omega_1 \times \Omega_2$ *denotes the Cartesian product and* $\mathfrak{S}_1 \times \mathfrak{S}_2$ *is the* product σ-algebra *given as*[3]

$$\mathfrak{S}_1 \times \mathfrak{S}_2 := \bigcap \{\mathfrak{S} \subseteq \mathfrak{P}(\Omega_1 \times \Omega_2) : \mathfrak{S} \textit{ is } \sigma\textit{-algebra}, T \times V \in \mathfrak{S} \textit{ for } T \in \mathfrak{S}_1 \textit{ and } V \in \mathfrak{S}_2\}.$$

Definition 2.4.2. *For two* σ*-finite measure spaces* $(\Omega_1, \mathfrak{S}_1, \mu)$ *and* $(\Omega_2, \mathfrak{S}_2, \nu)$, *the measure* $\mu \times \nu$ *is defined as the unique*[4] σ*-finite measure on the product measurable space* $(\Omega_1 \times \Omega_2, \mathfrak{S}_1 \times \mathfrak{S}_2)$ *with the property*

$$\mu \times \nu(A \times B) = \mu(A)\nu(B) \quad \textit{for all } A \in \mathfrak{S}_1, B \in \mathfrak{S}_2.$$

Now we can state Fubini's Theorem. The following is a slightly adapted variant of the one provided in [AL06, Theorem 5.2.2].

Theorem 2.4.3. (Fubini's Theorem) *Let* $(\Omega_1, \mathfrak{S}_1, \mu)$ *and* $(\Omega_2, \mathfrak{S}_2, \nu)$ *be two* σ*-finite measure spaces and* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ *is a* $(\mathfrak{S}_1 \times \mathfrak{S}_2, \mathscr{B})$*-measurable function. If at least one of the integrals*

$$\int |f| \,\mathrm{d}(\mu \times \nu), \quad \iint |f(x,y)| \,\mathrm{d}\mu(x) \,\mathrm{d}\nu(y), \quad \textit{or} \quad \iint |f(x,y)| \,\mathrm{d}\nu(y) \,\mathrm{d}\mu(x)$$

is finite, then

$$\int f \,\mathrm{d}(\mu \times \nu) = \iint f(x,y) \,\mathrm{d}\mu(x) \,\mathrm{d}\nu(y) = \iint f(x,y) \,\mathrm{d}\nu(y) \,\mathrm{d}\mu(x) \tag{2.9}$$

holds. In particular, the integrals in (2.9) are all defined.

2.5. Probability Measures

Now we can move on to a special type of measures, called probability measures, which are used to represent probability distributions.

[3] As in Definition 2.1.2, the reader is referred to [AL06, Definition 1.1.3] for a proof that this is indeed a well-defined σ-algebra.

[4] The uniqueness for σ-finite measure spaces is proved in [AL06, Theorem 5.1.2].

Definition 2.5.1. *[AL06, Definitions 1.2.2 and 2.1.1] A measure space* $(\Omega, \mathfrak{S}, \mu)$ *is called a* probability space *if* $\mu(\Omega) = 1$. *In this context the elements of* $\mathfrak{S}$ *are also called* events *and* μ *is called a* probability measure.

Definition 2.5.2. *For* $M \in \mathbb{N}$, *a* random variable[5] X *on a probability space* $(\Omega, \mathfrak{S}, \mu)$ *is a* $(\mathfrak{S}, \mathscr{B}^M)$*-measurable function* $X : \Omega \to \mathbb{R}^M$.

In particular, given a probability space $(\mathbb{R}^M, \mathscr{B}^M, \mu)$ for an $M \in \mathbb{N}$, the identity $\mathrm{id}_{\mathbb{R}^M}$ is a random variable.

Remark 2.5.3. Recalling Definition 2.2.3, we note that a random variable $X : \Omega \to \mathbb{R}^M$ induces the probability measure μ_X on $(\mathbb{R}^M, \mathscr{B}^M)$.

Remark 2.5.4. When dealing with two random variables, e.g., $X : \Omega \to \mathbb{R}^M$ and $Y : \Omega \to \mathbb{R}^N$, $M, N \in \mathbb{N}$ on a common probability space $(\Omega, \mathfrak{S}, \mu)$, they can be combined to the random variable $(X, Y) : \Omega \to \mathbb{R}^M \times \mathbb{R}^N \sim \mathbb{R}^{M+N}$.

2.6. Geometric Measure Theory

We will utilize some concepts from geometric measure theory. The most important definitions and theorems, which we will need for our subsequent considerations, will be restated with references to proofs and further treatment in the literature. The notation will mainly follow [Fed69].

In this section we will introduce the types of sets and measures that will form the basis of all subsequent considerations. To analyze the information theoretic properties of "lower-dimensional" probability measures on Euclidean space, we first need to specify the meaning of "lower-dimensional". To this end, we will use rectifiable sets and the Hausdorff measure, as provided by geometric measure theory.

The Hausdorff measure naturally permits measuring m-dimensional hypervolume in M-dimensional ($0 \le m \le M$) Euclidean space and the rectifiable sets are those, where the Hausdorff measure behaves "nicely". The probability measures under study are assumed to be concentrated on[6] a rectifiable set and absolutely continuous w.r.t. the Hausdorff measure.

The construction of the Hausdorff measure makes use of a more general concept for constructing measures, commonly referred to as Caratheodory's construction. We shall sketch this process and thereby introduce the Hausdorff measure $\mathscr{H}^m$. A more rigorous and complete treatment of this material can be found in [Fed69, Section 2.10] or [KP08, Section 2.1].

[5] Although a random variable is a function, it is customary to denote it using capital letters.
[6] "Concentrated on" means that the complement of the set in question has measure zero.

For $M \in \mathbb{N}$, let $\mathfrak{C} \subseteq \mathfrak{P}(\mathbb{R}^M)$ be a collection of subsets of $\mathbb{R}^M$ and $f : \mathfrak{C} \to \bar{\mathbb{R}}^+$. Then, for some $\delta > 0$ and $A \subseteq \mathbb{R}^M$, we define the approximating outer measure of size δ and gauge f as

$$\phi_\delta(A) := \inf\left\{ \sum_{Q \in \mathfrak{D}} f(Q) : \mathfrak{D} \subseteq \mathfrak{C} \cap \{Q : \operatorname{diam}(Q) \leq \delta\}, |\mathfrak{D}| \leq \aleph_0, \text{ and } A \subseteq \bigcup \mathfrak{D}. \right\}$$

Thus, to obtain $\phi_\delta(A)$, we take all countable coverings $\mathfrak{D}$ of A by sets from $\mathfrak{C}$, whose diameter does not exceed δ, and then take the infimum of the sums of the gauges $f(Q)$ for $Q \in \mathfrak{D}$. As, by construction, $\phi_\delta(A)$ increases as $\delta \to 0$ (we take the infimum of ever smaller sets), we may define another measure ϕ as

$$\phi(A) := \lim_{\delta \downarrow 0} \phi_\delta(A) = \sup_{\delta > 0} \phi_\delta(A). \tag{2.10}$$

This method for obtaining ϕ is called *Caratheodory's construction*. Once the collection of test sets $\mathfrak{C}$ and the gauge f is determined, the measure (2.10) is unique. This leads to the following definition of the Hausdorff outer measure.

Definition 2.6.1. *For $M \in \mathbb{N}$ and $m \geq 0$, the* Hausdorff outer measure *of dimension m on $\mathbb{R}^M$ is given by Caratheodory's construction with $\mathfrak{C} = \mathfrak{P}(\mathbb{R}^M) \setminus \{\emptyset\}$ and gauge*

$$f(E) = \alpha(m) \left(\frac{\operatorname{diam}(E)}{2} \right)^m,$$

where the constant $\alpha(m)$ is defined as

$$\alpha(m) := \frac{\left(\Gamma(\frac{1}{2})\right)^m}{\Gamma(\frac{m}{2} + 1)} \tag{2.11}$$

and denotes the m-hypervolume of the unit ball in m-dimensional Euclidean space, as given in [Fed69, 2.10.2].

Definition 2.6.2. *For $M \in \mathbb{N}$ and $m \in \{0, 1, \ldots, M\}$ let $\mathcal{H}^m$ denote the Hausdorff outer measure of dimension m on $\mathbb{R}^M$, as defined in Definition 2.6.1, but restricted to $\mathcal{B}^M$. The measure[7] $\mathcal{H}^m$ is the* Hausdorff measure *of dimension m on $\mathbb{R}^M$.*

Note that the dimension of the underlying space of the m-dimensional Hausdorff measure is suppressed in the notation and will be stated explicitly, when there is danger of confusion. Although, in general the m-dimensional Hausdorff measure is defined for any $m \in \mathbb{R}^+ \cup \{0\}$,

[7] The argument for $\mathcal{H}^m$ being a measure on $(\mathbb{R}^M, \mathcal{B}^M)$ as defined in Definition 2.1.3 is given in [Fed69, Section 2.10] or [KP08, Section 2.1].

we will only consider the case $m \in \mathbb{N} \cup \{0\}$. We state some properties of the Hausdorff measure in the following theorem.

Theorem 2.6.3. *Let $\mathcal{H}^m$ be the m-dimensional Hausdorff measure on $\mathbb{R}^M$, then*

(i) $m = 0 \implies \mathcal{H}^m = \mathcal{H}^0 = \zeta$,

(ii) $m = M \implies \mathcal{H}^m = \mathcal{H}^M = \lambda^M$,

(iii) For all $i \in \mathbb{N}$ let $A_i \in \mathcal{B}^M$ with $\mathcal{H}^m(A_i) < \infty$, then for all $m' \in \{m+1, m+2, \dots, M\}$ we have

$$\mathcal{H}^{m'}\left(\bigcup_i A_i\right) = 0.$$

Proof. Part (i) follows immediately from the construction and is remarked in [Fed69, 2.10.2(1)]. For Part (ii) see [Fed69, Theorem 2.10.35]. The remark in [Fed69, 2.10.2(1)] and the σ-subadditivity of measures (see Part (i) of Proposition 2.1.4) give us Part (iii). □

The rest of this section is dedicated to analyzing rectifiable sets and their connection with the Hausdorff measure. The rectifiable sets can be loosely characterized as those sets, where the Hausdorff outer measure behaves "nicely".

Definition 2.6.4. *Let $(\mathcal{X}_1, d_1)$ and $(\mathcal{X}_2, d_2)$ be metric spaces. A function $f : \mathcal{X}_1 \to \mathcal{X}_2$ is called* Lipschitz *if there exists an $L \geq 0$, such that*

$$d_2(f(x), f(y)) \leq L d_1(x, y) \quad \text{for all } x, y \in \mathcal{X}_1. \tag{2.12}$$

The smallest $L \geq 0$ satisfying (2.12) is called the Lipschitz constant *of f, denoted by* $\mathrm{Lip}(f)$.

Furthermore, we will use the following notation for the *multiplicity* of a function.

Definition 2.6.5. *For a function $f : A \to B$, define*[8]

$$N(f, y) := \zeta\big(f^{-1}(\{y\})\big) \quad \text{for } y \in B.$$

Definition 2.6.6. *For $m, n, k \in \mathbb{N}$ and a function $f : \mathbb{R}^m \to \mathbb{R}^n$ that is differentiable at $x \in \mathbb{R}^m$, let $J_k f(x)$ denote the k-dimensional Jacobian of f at x. The details of this definition may be found in [Fed69, 3.2.1] or [KP08, Definition 5.1.3].*

The Jacobian can be thought of as a quantity representing the deformation caused by the function f locally at x. Depending on m and n, it can be expressed using the differential (see

[8] Here ζ denotes the counting measure from Definition 2.1.5.

[Fed69, 3.1.1]) $Df(x)$ and is given by (from [KP08, Lemma 5.1.4])

$$J_{\min(m,n)}f(x) = \begin{cases} |\det(Df(x))| & m = n \\ \sqrt{\det[Df(x)^T Df(x)]} & m \leq n \\ \sqrt{\det[Df(x) Df(x)^T]} & m \geq n. \end{cases}$$

We will exclusively deal with the situation $m \leq n$, thus

$$J_m f(x) = \sqrt{\det[Df(x)^T Df(x)]}. \tag{2.13}$$

We can now define rectifiable sets. These sets will be the "support" for the probability distributions to be discussed. We extend the definition of rectifiability given in [Fed69, 3.2.14] to include the case $m = 0$, as suggested in the last paragraph of [Fed69, 3.2.14]. Furthermore, we define a restriction of the concept of rectifiability, which will prove useful for our setting.

Definition 2.6.7. *For $E \subseteq \mathbb{R}^M$ and $m \in \{0, 1, \ldots, M\}$ we define:*

*(i) For $m \neq 0$, E is m-*rectifiable *if there exists a Lipschitz function $f : A \to \mathbb{R}^M$ such that $f(A) = E$ for a bounded set $A \subseteq \mathbb{R}^m$. E is 0-rectifiable if it is finite.*

*(ii) E is $\mathscr{B}$-*countably m-rectifiable *if E equals the union of a countable family of m-rectifiable Borel sets.*

Lemma 2.6.8. *Let $E \subseteq \mathbb{R}^M$ be an m-rectifiable Borel set, then $\mathscr{H}^m(E) < \infty$. Furthermore, for a $\mathscr{B}$-countably m-rectifiable set F the measure $\mathscr{H}^m|_F$ is σ-finite.*

The proof of this lemma can be found in Appendix A.1.

Definition 2.6.7 has the nice property that any kind of rectifiability immediately extends to subsets.

Corollary 2.6.9. *Let $E, F \subseteq \mathbb{R}^M$, $F \subseteq E$ and $m \in \{0, 1, \ldots, M\}$. Then the following properties hold.*

(i) E is m-rectifiable $\Longrightarrow$ F is m-rectifiable,

(ii) $F \in \mathscr{B}^M$ and E is $\mathscr{B}$-countably m-rectifiable $\Longrightarrow$ F is $\mathscr{B}$-countably m-rectifiable.

Proof. To see Part (i) for $m = 0$, note that $|F| \leq |E| < \infty$ implies that F is 0-rectifiable. For $m \in \{1, \ldots, M\}$, let f be a Lipschitz function on a bounded $A \subseteq \mathbb{R}^m$ and $f(A) = E$. The set $B := f^{-1}(F)$ is bounded as $B \subseteq A$, the function $f|_B$ is Lipschitz, and $f(B) = F$.

For showing Part (ii) we write $E = \bigcup_{i\in\mathbb{N}} E_i$ where each E_i is m-rectifiable. Then $F = \bigcup_{i\in\mathbb{N}}(F \cap E_i)$, where for all $i \in \mathbb{N}$, $(F \cap E_i) \subseteq E_i$ is Borel as intersection of two Borel sets and m-rectifiable because of Part (i). □

Remark 2.6.10. For a $\mathscr{B}$-countably m-rectifiable set E, the m-rectifiable Borel sets covering E can be chosen disjoint by subtracting all preceding sets and using Part (i) of Corollary 2.6.9, as the Borel sets are closed under countable subtraction.

Remark 2.6.11. The countable union of $\mathscr{B}$-countably m-rectifiable sets is again $\mathscr{B}$-countably m-rectifiable.

Corollary 2.6.12. *For $m \in \{0, 1, \dots, M\}$, let the set E be $\mathscr{B}$-countably m-rectifiable. Then, for $m' > m$, $\mathscr{H}^{m'}(E) = 0$.*

Proof. According to Part (ii) of Definition 2.6.7, $E = \bigcup_i E_i$, where $E_i \in \mathscr{B}^M$ is m-rectifiable for $i \in \mathbb{N}$. Hence, by Lemma 2.6.8, $\mathscr{H}^m(E_i) < \infty$ for all $i \in \mathbb{N}$. Part (iii) of Theorem 2.6.3 now yields $\mathscr{H}^{m'}(E) = 0$. □

Lemma 2.6.13. *For two numbers $m_1, m_2 \in \{0, 1, \dots, M\}$, let $C \subseteq \mathbb{R}^M$ be $\mathscr{B}$-countably m_1-rectifiable and $D \subseteq \mathbb{R}^M$ be $\mathscr{B}$-countably m_2-rectifiable, then the set $E := C \times D \subseteq \mathbb{R}^{2M}$ is $\mathscr{B}$-countably $(m_1 + m_2)$-rectifiable and*[9]

$$\mathscr{H}^{m_1}|_C \times \mathscr{H}^{m_2}|_D = \mathscr{H}^{m_1+m_2}|_E. \tag{2.14}$$

The proof of this lemma is rather lengthy. It may be found in Appendix A.2.

[9] It should be noted that (2.14) denotes the equality of two measures.

Chapter 3.

Rectifiable Measures and Entropy

In this chapter we present the basic definitions and theorems, the rest of this thesis is built upon. In particular, we introduce $\mathscr{B}$-countably m-rectifiable measures in Section 3.1, a new notion that encompasses both discrete (0-rectifiable) and continuous probability distributions.

We discuss the density (w.r.t. the Hausdorff measure) and entropy of rectifiable measures in Sections 3.2 and 3.3, and conclude with a transformation formula for entropy under a (locally) Lipschitz function in Section 3.4.

3.1. Rectifiable Measures

We will now move to discussing properties of finite measures on $\mathbb{R}^M$, which are concentrated on a rectifiable set E, i.e., the set E^c is a null set.

Definition 3.1.1. *Let $m \in \{0, 1, \ldots, M\}$ and let $(\mathbb{R}^M, \mathscr{B}^M, \mu)$ be a measure space. The measure μ is an m-*rectifiable measure ($\mathscr{B}$-countably m-rectifiable measure) *if there exists a set $E \subseteq \mathbb{R}^M$ such that the following properties hold:*

(i) $\mu(\mathbb{R}^M) < \infty$,

(ii) E is m-rectifiable and Borel ($\mathscr{B}$-countably m-rectifiable), and

(iii) $\mu \ll \mathscr{H}^m|_E$.

We call a set E, satisfying these requirements, a support *of μ.*

Remark 3.1.2. For an $m \in \{0, 1, \ldots, M\}$ and all $i \in I$, where I is finite, let μ_i be a $\mathscr{B}$-countably m-rectifiable measure with support E_i and p_i a nonnegative real number. Then the measure

$$\mu := \sum_i p_i \mu_i$$

is again $\mathscr{B}$-countably m-rectifiable with support $E := \bigcup_i E_i$ (see Remark 2.6.11).

Lemma 3.1.3. *For $m \in \{0, 1, \ldots, M\}$, let the measure μ on $(\mathbb{R}^M, \mathscr{B}^M)$ be $\mathscr{B}$-countably m-rectifiable with support E. Then there exists a $(\mathscr{B}^M, \mathscr{B})$-measurable function $\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} : \mathbb{R}^M \to \mathbb{R}^+$, which satisfies*

$$x \notin E \implies \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E}(x) = 0 \tag{3.1}$$

and, using Definition 2.3.5,

$$\mu = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} \mathscr{H}^m. \tag{3.2}$$

Proof. The measure μ is finite and $\mathscr{H}^m|_E$ is σ-finite by Lemma 2.6.8. This enables us to apply Part (ii) of Theorem 2.3.3 which guarantees the existence of a function $\frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{H}^m|_E}$ such that

$$\mu = \frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{H}^m|_E} \mathscr{H}^m|_E.$$

The function

$$\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} := \mathbb{1}_E \frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{H}^m|_E}$$

satisfies (3.1) and (3.2). □

Remark 3.1.4. Using $\mu = \mu|_E$ (by Part (iii) of Definition 3.1.1) in (3.2), one can see that

$$\mu = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} \mathscr{H}^m = \left(\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} \mathscr{H}^m \right)\Bigg|_E = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} \mathscr{H}^m|_E,$$

which proves that $\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E}$ is in fact a Radon-Nikodým derivative, conforming with Part (ii) of Theorem 2.3.3 and justifying the notation.

Note that the support of a $\mathscr{B}$-countably m-rectifiable measure is not unique. Nevertheless, the function $\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E}$ is unique except for an $\mathscr{H}^m$-null set, irrespective of E, as we will show now.

Lemma 3.1.5. *For $m \in \{0, 1, \dots, M\}$, let the measure μ on $(\mathbb{R}^M, \mathscr{B}^M)$ be $\mathscr{B}$-countably m-rectifiable. Let E and E' denote two supports of μ, then*

$$\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_{E'}} \quad \mathscr{H}^m\text{-a.e.} \tag{3.3}$$

Thus, we can drop the explicit dependence on the support E whenever changes on an $\mathscr{H}^m$-null set are of no concern and simply write

$$\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m}.$$

Proof. We will first show

$$\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_{E'}} \quad \mathscr{H}^m|_{E'}\text{-a.e.} \tag{3.4}$$

To show (3.4), it suffices to prove

$$\mu(A) = \int_A \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} \,\mathrm{d}\mathscr{H}^m|_{E'} \quad \text{for all } A \in \mathscr{B}^M,$$

since (3.4) then follows from Part (ii) of Theorem 2.3.3. Using $\mu(A) = \mu(A \cap E')$, as $\mu(E'^c) = 0$ by Part (iii) of Definition 3.1.1, we obtain for any $A \in \mathscr{B}^M$,

$$\int_A \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} \,\mathrm{d}\mathscr{H}^m|_{E'} = \int_{A\cap E'} \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} \,\mathrm{d}\mathscr{H}^m \overset{(a)}{=} \mu(A \cap E') = \mu(A),$$

where (a) follows from (3.2). This shows (3.4).

We can now argue that in analogy to (3.4) also

$$\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_{E'}} \quad \mathscr{H}^m|_E\text{-a.e.}$$

holds, which, together with (3.4), provides

$$\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E} = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_{E'}} \quad \mathscr{H}^m|_{E\cup E'}\text{-a.e.}$$

This already proves (3.3) since $\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E}(x) = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_{E'}}(x) = 0$ for $x \in (E \cup E')^c$ by (3.1). □

Note, however, that $\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m}$ is not necessarily represented by a particular support of μ, i.e., there may be a function f that is $\mathscr{H}^m$-a.e. equal to $\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E}$ for any support E of μ, but no support E' exists with $f = \frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_{E'}}$.

Lemma 3.1.6. *Let $\mathfrak{M} \subseteq \{0, 1, \ldots, M\}$ and for every $m \in \mathfrak{M}$, let I_m be a finite set. Let p_m^i be a nonnegative real number and μ_m^i be a $\mathscr{B}$-countably m-rectifiable measure on $(\mathbb{R}^M, \mathscr{B}^M)$ with support E_m^i for every $m \in \mathfrak{M}$ and $i \in I_m$. Assume that the sets $E_m := \bigcup_{i\in I_m} E_m^i$ are disjoint and set*

$$\mu := \sum_{m\in\mathfrak{M}} \sum_{i\in I_m} p_m^i \mu_m^i.$$

Then, using Definition 2.3.5,

$$\mu = \left(\sum_{m\in\mathfrak{M}} \sum_{i\in I_m} p_m^i \frac{\dot{\mathrm{d}}\mu_m^i}{\mathrm{d}\mathscr{H}^m|_{E_m^i}} \right) \left(\sum_{m\in\mathfrak{M}} \mathscr{H}^m|_{E_m} \right). \tag{3.5}$$

Proof. We define $\nu := \sum_{m\in\mathfrak{M}} \mathscr{H}^m|_{E_m}$. The measure μ is finite and ν—as the finite sum of

σ-finite measures—is also σ-finite. Applying Part (ii) of Theorem 2.3.3 and using Part (ii) of Proposition 2.3.4, we can write

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu} = \sum_{m\in\mathfrak{M}} \sum_{i\in I_m} p_m^i \frac{\mathrm{d}\mu_m^i}{\mathrm{d}\nu}.$$

Note that $\mu_m^i = \mu_m^i|_{E_m^i}$ and that the sets E_m are disjoint, which implies $\nu|_{E_m^i} = \mathscr{H}^m|_{E_m^i}$, as $E_m^i \subseteq E_m$. We thus have $\frac{\mathrm{d}\mu_m^i}{\mathrm{d}\nu} = \frac{\mathrm{d}\mu_m^i}{\mathrm{d}\mathscr{H}^m|_{E_m^i}}$ ν-a.e., which yields (3.5). □

3.2. Densities

In this section we will obtain an explicit expression for the density $\frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{H}^m|_E}$ of a $\mathscr{B}$-countably m-rectifiable measure μ. To this end, we will employ the Besicovitch differentiation theorem.

We recall from Table 2.1 that the closed ball with radius r, centered at x is denoted $\overline{B_r(x)} \subseteq \mathbb{R}^M$, i.e., for $r > 0$, $x \in \mathbb{R}^M$,

$$\overline{B_r(x)} := \{y \in \mathbb{R}^M : \mathrm{dist}(x,y) \le r\},$$

where $\mathrm{dist}(\cdot,\cdot)$ denotes the Euclidean distance.

Definition 3.2.1. *[Fed69, 2.10.19]For $m \in \{0,1,\dots,M\}$ let μ be a measure on $(\mathbb{R}^M, \mathscr{B}^M)$. Then the m-density $\Theta^m(\mu,x)$ of μ at $x \in \mathbb{R}^M$ is defined as*

$$\Theta^m(\mu,x) := \lim_{r\downarrow 0} \frac{\mu(\overline{B_r(x)})}{\alpha(m)r^m}, \tag{3.6}$$

if the limit on the right in (3.6) exits. It may also assume the value $+\infty$. The constant $\alpha(m)$ is defined in (2.11).

It should be noted that the density from Definition 3.2.1 is linear in the measure. This will be precisely formulated in the following corollary.

Corollary 3.2.2. *For $m \in \{0,1,\dots,M\}$ let $(\mu_i)_{i\in\mathbb{N}}$ be a sequence of measures on $(\mathbb{R}^M, \mathscr{B}^M)$ and let $(c_i)_{i\in\mathbb{N}}$ be a sequence of nonnegative real numbers. Assume that for an $x \in \mathbb{R}^M$, and for all $i \in \mathbb{N}$, the density $\Theta^m(\mu_i,x)$ exits. Then the density of the measure $\mu := \sum_i c_i\mu_i$ can be expressed as*

$$\Theta^m(\mu,x) = \sum_i c_i \Theta^m(\mu_i,x).$$

Proof. The result follows immediately from the definition of the m-density in (3.6). □

Definition 3.2.3. *For $m \in \{0,1,\dots,M\}$, we define the m-*density of a Borel set *$E \in \mathscr{B}^M$ as the m-density of the measure $\mathscr{H}^m|_E$, i.e.,*

$$\Theta^m(E,x) := \Theta^m(\mathscr{H}^m|_E, x) = \lim_{r\downarrow 0} \frac{\mathscr{H}^m|_E(\overline{B_r(x)})}{\alpha(m)r^m} = \lim_{r\downarrow 0} \frac{\mathscr{H}^m(E \cap \overline{B_r(x)})}{\alpha(m)r^m}. \tag{3.7}$$

Lemma 3.2.4. *Let μ be a measure on $(\mathbb{R}^M, \mathscr{B}^M)$, $m \in \{0,1,\dots,M\}$ and $E \in \mathscr{B}^M$ with $\mu(E) < \infty$, then*

$$\Theta^m(\mu|_E, x) = \mathbb{1}_E \Theta^m(\mu|_E, x) \quad \textit{for } \mathscr{H}^m\textit{-a.e. } x \in \mathbb{R}^M. \tag{3.8}$$

Proof. In [Fed69, 2.10.19(4)] it is shown that[1] $\Theta^m(\mu|_E, x) = 0$ for $\mathscr{H}^m$-a.e. $x \in E^c$, which proves (3.8). □

Theorem 3.2.5. *For $m \in \{0,1,\dots,M\}$, let the measure μ on $(\mathbb{R}^M, \mathscr{B}^M)$ be m-rectifiable. Then the function*

$$f(x) := \Theta^m(\mu, x)$$

is defined $\mathscr{H}^m$-a.e. and f is a density of μ w.r.t. $\mathscr{H}^m$ in the sense that

$$\mu = f\mathscr{H}^m.$$

Proof. Let E be a support of μ. The measures μ and $\mathscr{H}^m|_E$ (see Lemma 2.6.8) are finite, therefore also σ-finite, and inner regular by [AFP00, Proposition 1.43]. Additionally, as both measures are finite, every compact set trivially has finite measure. As a result μ and $\mathscr{H}^m|_E$ fulfill the requirements for Radon measures as defined in [AFP00, Definition 1.40]. We can therefore apply the Besicovitch differentiation theorem [AFP00, Theorem 2.22]. It guarantees $\mathscr{H}^m|_E$-a.e. the existence of

$$\tilde{f}(x) := \lim_{r\downarrow 0} \frac{\mu(\overline{B_r(x)})}{\mathscr{H}^m|_E(\overline{B_r(x)})} = \lim_{r\downarrow 0} \frac{\mu(\overline{B_r(x)})}{\mathscr{H}^m(E \cap \overline{B_r(x)})}, \tag{3.9}$$

and it ensures that $\tilde{f}$ is the Radon-Nikodým derivative of μ w.r.t. $\mathscr{H}^m|_E$ because $\mu \ll \mathscr{H}^m|_E$ by Part (iii) of Definition 3.1.1, i.e., we have

$$\mu = \tilde{f}\mathscr{H}^m|_E. \tag{3.10}$$

[1] Federer proves the stated result for the spherical measure $\mathscr{S}^m$, but $\mathscr{S}^m \geq \mathscr{H}^m$ by [Fed69, 2.10.6].

By [Fed69, Theorem 3.2.19], we have that

$$\Theta^m(E,x)=1 \quad \text{for } \mathcal{H}^m\text{-a.e. } x\in E. \tag{3.11}$$

Using (3.11), we obtain for $\mathcal{H}^m$-a.e. $x\in E$:

$$\begin{aligned}
\tilde{f}(x) &\overset{(3.9)}{=} \lim_{r\downarrow 0} \frac{\mu(\overline{B_r(x)})}{\mathcal{H}^m(E\cap\overline{B_r(x)})}\Theta^m(E,x)= \\
&\overset{(3.6)}{=} \lim_{r\downarrow 0} \frac{\mu(\overline{B_r(x)})}{\mathcal{H}^m(E\cap\overline{B_r(x)})} \lim_{r\downarrow 0} \frac{\mathcal{H}^m(E\cap\overline{B_r(x)})}{\alpha(m)r^m} = \\
&= \lim_{r\downarrow 0} \frac{\mu(\overline{B_r(x)})\mathcal{H}^m(E\cap\overline{B_r(x)})}{\mathcal{H}^m(E\cap\overline{B_r(x)})\alpha(m)r^m} = \\
&\overset{(3.7)}{=} \Theta^m(\mu,x)
\end{aligned} \tag{3.12}$$

Note that by Part (iii) of Definition 3.1.1, $\mu=\mu|_E$. Using Lemma 3.2.4, we get

$$\mu \overset{(3.10)}{=} \tilde{f}\mathcal{H}^m|_E \overset{(3.12)}{=} \mathbb{1}_E\Theta^m(\mu,\cdot)\mathcal{H}^m \overset{3.2.4}{=} \Theta^m(\mu,\cdot)\mathcal{H}^m = f\mathcal{H}^m.$$

□

Corollary 3.2.6. *For $m\in\{0,1,\dots,M\}$, let the measure μ on $(\mathbb{R}^M,\mathcal{B}^M)$ be $\mathcal{B}$-countably m-rectifiable, then the function*

$$f(x):=\Theta^m(\mu,x)$$

is defined $\mathcal{H}^m$-a.e. and f is a density of μ w.r.t. $\mathcal{H}^m$ in the sense that

$$\mu = f\mathcal{H}^m. \tag{3.13}$$

Proof. Let E be a support of μ and $\{E_i\}_{i\in\mathbb{N}}$ be a partition of E into m-rectifiable Borel sets which exists according to Part (ii) of Definition 2.6.7 and Remark 2.6.10.

First, we want to show for all $i\in\mathbb{N}$ that $\Theta^m\left(\mu|_{E_i},x\right)$ is well-defined for $\mathcal{H}^m$-a.e. $x\in E_i$, that

$$\Theta^m\left(\mu|_{E_i},x\right)=\Theta^m(\mu,x) \quad \text{for } \mathcal{H}^m\text{-a.e. } x\in E_i \tag{3.14}$$

and, that

$$\mu|_{E_i}=\Theta^m\left(\mu|_{E_i},\cdot\right)\mathcal{H}^m. \tag{3.15}$$

Checking Definition 3.1.1 reveals that $\mu|_{E_i}$ is an m-rectifiable measure with support E_i. This permits application of Theorem 3.2.5, which guarantees that $\Theta^m\left(\mu|_{E_i},\cdot\right)$ is well-defined

$\mathscr{H}^m$-a.e., and yields (3.15).

As $\{E_i\}_{i\in\mathbb{N}}$ is a partition of E, using Definition 3.2.1, we can rewrite

$$\begin{aligned}\Theta^m(\mu,x) &= \lim_{r\downarrow 0}\frac{\mu(\overline{B_r(x)})}{\alpha(m)r^m}\\ &= \lim_{r\downarrow 0}\frac{\mu(E_i\cap\overline{B_r(x)})+\mu\left(\left(\bigcup_{j\neq i}E_j\right)\cap\overline{B_r(x)}\right)}{\alpha(m)r^m}\\ &= \Theta^m\left(\mu|_{E_i},x\right)+\Theta^m\left(\mu|_{\bigcup_{j\neq i}E_j},x\right).\end{aligned}$$

Using Lemma 3.2.4, we see that $\Theta^m\left(\mu|_{\bigcup_{j\neq i}E_j},x\right)=0$ for $\mathscr{H}^m$-a.e. $x\in E_i$, thus proving (3.14).

Now, using these properties, we can prove (3.13) by

$$\begin{aligned}\mu &= \sum_i \mu|_{E_i}\\ &\overset{(3.15)}{=} \sum_i \Theta^m\left(\mu|_{E_i},\cdot\right)\mathscr{H}^m\\ &\overset{(a)}{=} \sum_i \mathbb{1}_{E_i}\Theta^m\left(\mu|_{E_i},\cdot\right)\mathscr{H}^m\\ &\overset{(3.14)}{=} \sum_i \mathbb{1}_{E_i}\Theta^m(\mu,\cdot)\mathscr{H}^m\\ &\overset{(b)}{=} \mathbb{1}_E\Theta^m(\mu,\cdot)\mathscr{H}^m\\ &\overset{(c)}{=} \Theta^m(\mu,\cdot)\mathscr{H}^m,\end{aligned}$$

where (a) holds because of Lemma 3.2.4, (b) follows from [Fed69, 2.4.8], and (c) holds by Lemma 3.2.4 and because $\mu=\mu|_E$ (see Part (iii) of Definition 3.1.1). □

Comparing now (3.2) and (3.13), we note that

$$\Theta^m(\mu,x)=\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E}(x)\quad\text{for }\mathscr{H}^m\text{-a.e. }x\in\mathbb{R}^M,$$

i.e.,

$$\frac{\dot{\mathrm{d}}\mu}{\mathrm{d}\mathscr{H}^m|_E}=\mathbb{1}_E\Theta^m(\mu,\cdot).$$

satisfies both (3.1) and (3.2).

3.3. Entropy

In this section we will define entropy in a general setting and then apply it to the case of a $\mathscr{B}$-countably m-rectifiable probability measure.

Definition 3.3.1. *Let $(\Omega, \mathfrak{S}, \mu)$ be a probability space and let ν be a σ-finite measure on the same measurable space $(\Omega, \mathfrak{S})$ with the property $\mu \ll \nu$. Then the Radon-Nikodým derivative $\frac{d\mu}{d\nu}$ exists (see Part (ii) of Theorem 2.3.3) and the* entropy of μ w.r.t. ν *is given by*

$$h_\nu(\mu) := \int \log\left(\frac{d\mu}{d\nu}\right)^{-1} d\mu, \tag{3.16}$$

if the integral in (3.16) is defined.

As a special case of Definition 3.3.1 we define the entropy of a random variable on $(\mathbb{R}^M, \mathscr{B}^M)$.

Definition 3.3.2. *Let $(\mathbb{R}^M, \mathscr{B}^M, \mu)$ be a probability space and let the random variable $X = \mathrm{id}_{\mathbb{R}^M}$ be the identity on $\mathbb{R}^M$. Furthermore, let ν denote a σ-finite measure on $(\mathbb{R}^M, \mathscr{B}^M)$ with the property $\mu \ll \nu$. Then the* entropy of X w.r.t. ν *is defined as the entropy of μ w.r.t. ν, i.e.,*

$$h_\nu(X) := h_\nu(\mu) = \int \log\left(\frac{d\mu}{d\nu}\right)^{-1} d\mu, \tag{3.17}$$

if the integral in (3.17) is defined.

We can already give a useful expression for the entropy of a $\mathscr{B}$-countably m-rectifiable probability measure.

Lemma 3.3.3. *Let μ be a $\mathscr{B}$-countably m-rectifiable probability measure on $(\mathbb{R}^M, \mathscr{B}^M)$, where $m \in \{0, 1, \ldots, M\}$. If E is a support of μ, the entropy of μ w.r.t. $\mathscr{H}^m|_E$ satisfies*

$$h_{\mathscr{H}^m|_E}(\mu) = \int \frac{\dot{d}\mu}{d\mathscr{H}^m} \log\left(\frac{\dot{d}\mu}{d\mathscr{H}^m}\right)^{-1} d\mathscr{H}^m, \tag{3.18}$$

if the integral in (3.18) is defined. It is in particular independent of the choice of the support E.

Proof. As the dotted Radon-Nikodým derivative is also a Radon-Nikodým derivative by Remark 3.1.4, inserting $\nu = \mathscr{H}^m|_E$ in (3.17), yields

$$h_{\mathscr{H}^m|_E}(\mu) = \int \frac{\dot{d}\mu}{d\mathscr{H}^m|_E} \log\left(\frac{\dot{d}\mu}{d\mathscr{H}^m|_E}\right)^{-1} d\mathscr{H}^m|_E.$$

We can drop the restriction of the integration because of (3.1) and dropping the supports in the dotted Radon-Nikodým derivatives is justified by Lemma 3.1.5. □

The fact that $h_{\mathscr{H}^m|_E}(\mu)$ does not depend on the particular support E justifies the following definition.

Definition 3.3.4. *Let μ be a $\mathscr{B}$-countably m-rectifiable measure on $(\mathbb{R}^M, \mathscr{B}^M)$, where $m \in \{0, 1, \ldots, M\}$. Let E be an arbitrary support of μ. Then the entropy of μ is defined as*

$$h(\mu) := h_{\mathscr{H}^m|_E}(\mu), \tag{3.19}$$

if $h_{\mathscr{H}^m|_E}(\mu)$ is defined.

Remark 3.3.5. It is worthwhile pointing out that Definition 3.3.4 generalizes entropy of discrete random variables (see [CT06, Section 2.1]) and differential entropy (see [CT06, Section 8.1]) of continuous random variables. If μ is a discrete probability measure on $(\mathbb{R}^M, \mathscr{B}^M)$, it takes values in a countable set $E \subseteq \mathbb{R}^M$, i.e., μ is $\mathscr{B}$-countably 0-rectifiable by Part (i) of Theorem 2.6.3 with support E and

$$h(\mu) = h_{\zeta|_E}(\mu) = \sum_{x \in E} \mu(\{x\}) \log \mu(\{x\})^{-1}.$$

Given a probability measure μ on $(\mathbb{R}^M, \mathscr{B}^M)$ with $\mu \ll \lambda^M$, Part (ii) of Theorem 2.6.3 shows that μ is $\mathscr{B}$-countably M-rectifiable with support $E = \mathbb{R}^M$ and

$$h(\mu) = h_{\lambda^M}(\mu) = \int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^M} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^M}\right)^{-1} \mathrm{d}\lambda^M.$$

3.4. Image of a Lipschitz Function

In this section we will discuss the entropy of the image of a random variable under a (locally) Lipschitz function. The discussion will give rise to an elegant conversion law for entropy and the simplest case, a linear transformation (without full rank), will also be treated.

Much of the subsequent considerations will rely on the following two fundamental theorems.

Theorem 3.4.1. *[Fed69, Theorem 2.10.43]* (Kirszbraun's Theorem) *For $M \in \mathbb{N}$ and $m \in \{1, \ldots, M\}$, if $S \subseteq \mathbb{R}^m$ and $f : S \to \mathbb{R}^M$ is Lipschitz, then there exists a Lipschitz function $g : \mathbb{R}^m \to \mathbb{R}^M$ satisfying $f(x) = g(x)$ for all $x \in S$ and* $\mathrm{Lip}(g) = \mathrm{Lip}(f)$.

Theorem 3.4.2. *[Fed69, Theorem 3.2.3]* (Area Formula) *For $M \in \mathbb{N}$ and $m \in \{1, \ldots, M\}$, let $f : \mathbb{R}^m \to \mathbb{R}^M$ be a Lipschitz function.*

(i) If $A \in \mathscr{B}^m$, then

$$\int_A J_m f(x)\,\mathrm{d}\lambda^m(x) = \int N(f|_A, y)\,\mathrm{d}\mathscr{H}^m(y). \tag{3.20}$$

(ii) If $u : \mathbb{R}^m \to \mathbb{R}^+$ is a $(\mathscr{B}^m, \mathscr{B})$-measurable function, then

$$\int u(x) J_m f(x)\,\mathrm{d}\lambda^m(x) = \int \sum_{x \in f^{-1}(\{y\})} u(x)\,\mathrm{d}\mathscr{H}^m(y). \tag{3.21}$$

The domain of the integrals on the left-hand side in (3.20) and (3.21) is $\mathbb{R}^m$ and the domain of the integrals on the right-hand side is $\mathbb{R}^M$.

Define the random variable $X = \mathrm{id}_{\mathbb{R}^m}$ on the probability space $(\mathbb{R}^m, \mathscr{B}^m, \mu)$, satisfying $\mu \ll \lambda^m$ for $m \in \mathbb{N}$, and let f be a function $f : \mathbb{R}^m \to \mathbb{R}^M$ for $M \geq m$. If the function f is measurable, we obtain the random variable $Y = f(X)$, which induces $\mu_Y = \mu_{f \circ X} = \mu_f$ on $(\mathbb{R}^M, \mathscr{B}^M)$, as given in Remark 2.5.3. In general μ_f will not be $\mathscr{B}$-countably m-rectifiable, however, we will now state sufficient conditions and develop an expression for $\frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{H}^m|_{f(\mathbb{R}^m)}}$, following this analysis.

Theorem 3.4.3. *For $M, m \in \mathbb{N}$ and $m < M$ let $(\mathbb{R}^m, \mathscr{B}^m, \mu)$ be a probability space satisfying $\mu \ll \lambda^m$. A function $f : \mathbb{R}^m \to \mathbb{R}^M$ is given, for which the following properties hold.*

(i) For λ^m-a.e. $x \in \mathbb{R}^m$, there exists an open set $x \in U_x$, such that $f|_{U_x}$ is Lipschitz, i.e., f is locally Lipschitz λ^m-a.e. Let B be the λ^m-null set where no such open set exists.

(ii) For λ^m-a.e. $x \in \mathbb{R}^m$, the Jacobian $J_m f(x) > 0$.

Then the induced measure μ_f is $\mathscr{B}$-countably m-rectifiable and there exists a support E of μ_f satisfying $f(B^c) \subseteq E$.

Proof. We check the properties in Definition 3.1.1. Part (i) follows from $\mu_f(\mathbb{R}^M) = \mu(\mathbb{R}^m) = 1$.

To show Part (ii) we proceed as follows. For every $x \in B^c$ we choose $r_x > 0$ such that $V_x := B_{r_x}(x) \subseteq \overline{B_{r_x}(x)} \subseteq U_x$ (U_x is the open set from Part (i) of Theorem 3.4.3.). $\{V_x\}_{x \in B^c}$ constitutes an open[2] cover of B^c. By [Mun00, Theorem 30.2], B^c is second-countable, as the topological subspace of the second-countable space $\mathbb{R}^M$. Thus, B^c also has the Lindelöf property by [Mun00, Theorem 30.3] and a countable subcover $\{V_{x_i}\}_{i \in \mathbb{N}}$ suffices to cover B^c. For $i \in \mathbb{N}$, $\overline{V_{x_i}}$ is bounded and closed, i.e., compact. The set $E_i := f(\overline{V_{x_i}})$ is the image of a compact set under a continuous function and thus compact by [Mun00, Theorem 26.5]. Every compact set is in particular a Borel set and therefore E_i is Borel and the image of the bounded set $\overline{V_{x_i}}$ under

[2] B^c is equipped with the subspace topology.

the Lipschitz function f, i.e., m-rectifiable. The set $E := \bigcup_i E_i$ is consequently $\mathscr{B}$-countably m-rectifiable and satisfies $f(B^c) \subseteq E$.

For showing Part (iii), let $C \in \mathscr{B}^M$ with $\mathscr{H}^m|_E(C) = 0$. We need to show $\mu_f(C) = \mu(f^{-1}(C)) = 0$. By Part (ii) of Proposition 2.1.4, $\mathscr{H}^m(C \cap E_i) = 0$ for all $i \in \mathbb{N}$. Using Theorem 3.4.1, we can extend $f|_{E_i}$ to a Lipschitz function with domain $\mathbb{R}^m$ and apply Part (i) of Theorem 3.4.2 (setting $A = f^{-1}(C \cap E_i)$) yielding

$$\int_{f^{-1}(C\cap E_i)} J_m f \,\mathrm{d}\lambda^m = \int N(f|_{f^{-1}(C\cap E_i)}, y)\,\mathrm{d}\mathscr{H}^m(y) = 0,$$

where we used $N(f|_{f^{-1}(C\cap E_i)}, y) = 0$ for $\mathscr{H}^m$-a.e. $y \in \mathbb{R}^M$ because $N(f|_{f^{-1}(C\cap E_i)}, y) \neq 0$ only for $y \in (C \cap E_i)$, which is an $\mathscr{H}^m$-null set. As $J_m f(x) > 0$ for λ^m-a.e. $x \in \mathbb{R}^m$, $f^{-1}(C \cap E_i)$ has to be a λ^m-null set.

We note that

$$\begin{aligned}
\lambda^m\left((f^{-1}(E))^c\right) &\overset{(a)}{=} \lambda^m\left(\left(f^{-1}\left(\bigcup_i f(\overline{V_{x_i}})\right)\right)^c\right) \\
&= \lambda^m\left(\left(\bigcup_i f^{-1}(f(\overline{V_{x_i}}))\right)^c\right) \\
&\overset{(b)}{\leq} \lambda^m\left(\left(\bigcup_i \overline{V_{x_i}}\right)^c\right) \\
&\overset{(c)}{\leq} \lambda^m(B) = 0,
\end{aligned} \tag{3.22}$$

where (a) follows from the definition of E, and (b) and (c) follow from the monotonicity of measures (Part (ii) of Proposition 2.1.4), using additionally $B^c \subseteq \bigcup_i V_{x_i}$ in (c). We conclude

$$\begin{aligned}
\lambda^m\left(f^{-1}(C)\right) &\overset{(a)}{=} \lambda^m\left(f^{-1}(C) \cap f^{-1}(E)\right) \\
&= \lambda^m\left(f^{-1}(C \cap E)\right) \\
&\overset{(b)}{=} \lambda^m\left(\bigcup_i f^{-1}(C \cap E_i)\right) \\
&\overset{(c)}{\leq} \sum_i \lambda^m(f^{-1}(C \cap E_i) = 0,
\end{aligned}$$

where (a) follows from (3.22), (b) follows from the definition of E, and (c) is a consequence of the σ-subadditivity (Part (i) of Proposition 2.1.4). Now $\mu \ll \lambda^m$ yields $\mu_f(C) = \mu(f^{-1}(C)) = 0$. □

The Area Formula also provides us with an expression for the Radon-Nikodým derivative of μ_f w.r.t. the Hausdorff measure $\mathscr{H}^m|_E$.

Theorem 3.4.4. *In the situation of Theorem 3.4.3, for any $C \in \mathscr{B}^M$,*

$$\mu_f(C) = \int_C \sum_{x \in f^{-1}(\{y\})} \frac{\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x)}{J_m f(x)} \,\mathrm{d}\mathscr{H}^m|_E(y),$$

i.e.,

$$\frac{\mathrm{d}\mu_f}{\mathrm{d}\mathscr{H}^m|_E}(y) = \sum_{x \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)}. \tag{3.23}$$

Proof. The proof will use the Area Formula in a similar fashion as the proof of Theorem 3.4.3.

Choose E, E_i, and V_{x_i} as in the proof of Theorem 3.4.3. By removing all preceding sets in the countable covering, we obtain the partition

$$\tilde{V}_i := V_{x_i} \setminus \bigcup_{i' < i} V_{x_i}.$$

For every $i \in \mathbb{N}$, the function $f|_{\tilde{V}_i}$ is Lipschitz. We can extend it using Theorem 3.4.1 and apply Part (ii) of Theorem 3.4.2, yielding

$$\int_{\tilde{V}_i} u(x) J_m f(x) \,\mathrm{d}\lambda^m(x) = \int_{f(\tilde{V}_i)} \sum_{x \in f^{-1}(\{y\})} u(x) \,\mathrm{d}\mathscr{H}^m(y) \tag{3.24}$$

for a $(\mathscr{B}^m, \mathscr{B})$-measurable function $u : \mathbb{R}^m \to \mathbb{R}^+$. For $C \in \mathscr{B}^M$, we can choose u as

$$u(x) = \mathbb{1}_{f^{-1}(C)}(x) \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)} \tag{3.25}$$

observing that $J_m f(x) \neq 0$ for λ^m-a.e. $x \in \mathbb{R}^m$. Substituting (3.25) in (3.24) yields

$$\begin{aligned} \mu(f^{-1}(C) \cap \tilde{V}_i) &= \int_{\tilde{V}_i} \mathbb{1}_{f^{-1}(C)}(x) \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \,\mathrm{d}\lambda^m(x) \\ &\overset{(3.24)}{=} \int_{f(\tilde{V}_i)} \sum_{x \in f^{-1}(\{y\})} \mathbb{1}_{f^{-1}(C)}(x) \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)} \,\mathrm{d}\mathscr{H}^m(y) \\ &= \int_{f(\tilde{V}_i) \cap C} \sum_{x \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)} \,\mathrm{d}\mathscr{H}^m(y). \end{aligned} \tag{3.26}$$

Using $\mu\left(\left(\bigcup_i \tilde{V}_i\right)^c\right) = \mu\left(\left(\bigcup_i V_{x_i}\right)^c\right) \leq \mu(B) = 0$, we can write

$$\begin{aligned}
\mu_f(C) &= \mu(f^{-1}(C)) \\
&\overset{(a)}{=} \sum_i \mu(f^{-1}(C) \cap \tilde{V}_i) \\
&\overset{(3.26)}{=} \sum_i \int_{f(\tilde{V}_i)\cap C} \sum_{x \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)} \, \mathrm{d}\mathcal{H}^m(y) \\
&\overset{(b)}{=} \int_{E\cap C} \sum_{x \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)} \, \mathrm{d}\mathcal{H}^m(y),
\end{aligned}$$

where (a) follows from $B^c \subseteq \bigcup_i \tilde{V}_i$ by the σ-additivity (Part (ii) of Definition 2.1.3) and (b) follows from [Fed69, 2.4.8]. □

If f is injective, Theorem 3.4.4 directly provides an expression for the entropy of μ_f. When inserting (3.23) into Definition 3.3.2, Definition 3.3.4 gives us

$$\begin{aligned}
h(\mu_f) &\overset{(3.17)}{=} \int \log\left(\frac{\mathrm{d}\mu_f}{\mathrm{d}\mathcal{H}^m|_E}(y)\right)^{-1} \mathrm{d}\mu_f(y) \\
&\overset{(3.23)}{=} \int \log\left(\sum_{x \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)}\right)^{-1} \mathrm{d}\mu_f(y) \\
&= \int \sum_{x \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)} \log\left(\sum_{\tilde{x} \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(\tilde{x}) \frac{1}{J_m f(\tilde{x})}\right)^{-1} \mathrm{d}\mathcal{H}^m|_E(y) \\
&\overset{(a)}{=} \int \sum_{x \in f^{-1}(\{y\})} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)}\right)^{-1} \mathrm{d}\mathcal{H}^m|_E(y) \\
&\overset{(b)}{=} \int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \frac{1}{J_m f(x)}\right)^{-1} \mathrm{d}\lambda^m(x) \\
&\overset{(c)}{=} \int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x)\right)^{-1} \mathrm{d}\lambda^m(x) + \int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \log\left(\frac{1}{J_m f(x)}\right)^{-1} \mathrm{d}\lambda^m(x) \\
&= h(\mu) + \int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \log(J_m f(x)) \, \mathrm{d}\lambda^m(x),
\end{aligned} \tag{3.27}$$

where (a) follows from f being injective, Part (ii) of Theorem 3.4.2 is applied in (b), and (c) follows from Part (ii) of Theorem 2.2.10 if the integrals in (3.27) and their sum are defined. Thus, we proved the following theorem.

Theorem 3.4.5. *For $M, m \in \mathbb{N}$ and $m < M$ let $(\mathbb{R}^m, \mathcal{B}^m, \mu)$ be a probability space satisfying $\mu \ll \lambda^m$. An injective function $f : \mathbb{R}^m \to \mathbb{R}^M$ is given and the following properties hold.*

(i) *For λ^m-a.e. $x \in \mathbb{R}^m$, there exists an open set $x \in U_x$, such that $f|_{U_x}$ is Lipschitz, i.e., f is locally Lipschitz λ^m-a.e.*

(ii) *For λ^m-a.e. $x \in \mathbb{R}^m$, the Jacobian $J_m f(x) > 0$.*

Then the entropy of the induced measure μ_f is given by

$$h(\mu_f) = h(\mu) + \int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^m}(x) \log(J_m f(x))\, \mathrm{d}\lambda^m(x), \tag{3.28}$$

if the entropy $h(\mu)$ of μ w.r.t. the Lebesgue measure λ^m, the integral and their sum in (3.28) *are defined.*

In particular, if f is a linear function, represented by the $(M \times m)$-matrix D,

$$h(\mu_f) = h(\mu) + \log \sqrt{\det[D^T D]},$$

if $h(\mu)$ is defined.

Proof. The first part has been proved already. The second part follows from (3.28) by μ being a probability measure and using (2.13). □

Chapter 4.

Combined Rectifiable Measures and Mutual Information

So far we focused on one probability measure and its properties. This chapter is dedicated to generalizing the concepts developed in Chapter 3 to the product space of two Euclidean spaces. That will permit the analysis of two random variables by means of their combined distribution and will result in an expression for mutual information.

Furthermore, mixtures of random variables, comprised of different dimensions, are treated in Section 4.2 and all subsequent results are given for those. In Section 4.4 we will show a connection between mutual information and entropy, as defined earlier in Chapter 3.

4.1. Combined Rectifiable Measures

We will now consider finite measures on $\mathbb{R}^N \times \mathbb{R}^N \cong \mathbb{R}^{2N}$ for $N \in \mathbb{N}$.

Choosing different dimensions, i.e., $\mathbb{R}^{N'} \times \mathbb{R}^{M'}$ and allowing for $N' \neq M'$ would not provide any benefit, as simply choosing $N = \max(N', M')$ and defining measures on[1] $\{0\} \times \mathbb{R}^{N'+M'}$ yields the same result.

Definition 4.1.1. *For $n_1, n_2 \in \{0, 1, \ldots, N\}$, a measure μ on $(\mathbb{R}^{2N}, \mathscr{B}^{2N})$, is called a* combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure *if there exist two sets $C, D \in \mathscr{B}^N$, such that the following properties hold:*

(i) $\mu(\mathbb{R}^{2N}) < \infty$

(ii) *C is $\mathscr{B}$-countably n_1-rectifiable*

(iii) *D is $\mathscr{B}$-countably n_2-rectifiable*

(iv) $\mu \ll \mathscr{H}^{n_1}|_C \times \mathscr{H}^{n_2}|_D$

We call two sets C and D, fulfilling these requirements, a first *and* second support *of μ, respectively.*

Lemma 4.1.2. *For $n_1, n_2 \in \{0, 1, \ldots, N\}$, let μ be a combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure on $\mathbb{R}^{2N}$ with C and D denoting a first and second support, respectively. Then μ is also a $\mathscr{B}$-countably $(n_1 + n_2)$-rectifiable measure on $\mathbb{R}^{2N}$ and $C \times D$ is a support of μ.*

[1] Here 0 denotes the all-zero vector of appropriate dimension $|N' - M'|$.

Proof. As μ is a measure on $(\mathbb{R}^{2N}, \mathscr{B}^{2N})$ by definition, we have to check the three conditions in Definition 3.1.1:

Part (i) holds because of Part (i) of Definition 4.1.1, Part (ii) holds because of Lemma 2.6.13, and Part (iii) is equivalent to Part (iv) of Definition 4.1.1 by Lemma 2.6.13. □

Definition 4.1.3. *Let $(\mathbb{R}^{2N}, \mathscr{B}^{2N}, \mu)$ be a measure space.*

(i) The measure μ_1 on $(\mathbb{R}^N, \mathscr{B}^N)$, defined as $\mu_1(A) := \mu(A \times \mathbb{R}^N)$ for a set $A \in \mathscr{B}^N$ is called the first marginal *of μ.*

(ii) The measure μ_2 on $(\mathbb{R}^N, \mathscr{B}^N)$, defined as $\mu_2(A) := \mu(\mathbb{R}^N \times A)$ for a set $A \in \mathscr{B}^N$ is called the second marginal *of μ.*

Lemma 4.1.4. *For $n_1, n_2 \in \{0, 1, \ldots, N\}$ let μ be a combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure on $\mathbb{R}^{2N}$ with C and D denoting a first and second support, respectively. For $\xi \in \{1, 2\}$, let μ_ξ be the first/second marginal of μ as in Definition 4.1.3. Then μ_ξ is a $\mathscr{B}$-countably n_ξ-rectifiable measure. Furthermore, C is a support of μ_1 and D is a support of μ_2.*

Proof. We will prove Lemma 4.1.4 for $\xi = 1$. The case $\xi = 2$ follows analogously. Apparently, μ_1 is a measure on $(\mathbb{R}^N, \mathscr{B}^N)$ and we have to check the three conditions in Definition 3.1.1:

Part (i) holds as $\mu_1(\mathbb{R}^N) = \mu(\mathbb{R}^N \times \mathbb{R}^N) < \infty$ because of Part (i) of Definition 4.1.1. Part (ii) follows from Part (ii) of Definition 4.1.1 and Part (iii) can be seen as follows. For any $A \in \mathscr{B}^N$

$$\begin{aligned} &\mathscr{H}^{n_1}|_C(A) = 0 \implies \mathscr{H}^{n_1}|_C \times \mathscr{H}^{n_2}|_D(A \times \mathbb{R}^N) = 0 \\ &\overset{(a)}{\implies} \mu(A \times \mathbb{R}^N) = 0 \Leftrightarrow \mu_1(A) = 0, \end{aligned}$$

where (a) follows from Part (iv) of Definition 4.1.1. □

Lemma 4.1.5. *For $n_1, n_2 \in \{0, 1, \ldots, N\}$ let μ_1 be a $\mathscr{B}$-countably n_1-rectifiable measure on $(\mathbb{R}^N, \mathscr{B}^N)$ with support C and let μ_2 be a $\mathscr{B}$-countably n_2-rectifiable measure on $(\mathbb{R}^N, \mathscr{B}^N)$ with support D. Then the product measure $\mu_1 \times \mu_2$ is a combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure on $(\mathbb{R}^{2N}, \mathscr{B}^{2N})$ with C and D as first and second support. Furthermore,*

$$\frac{\dot{\mathrm{d}}\mu_1 \times \mu_2}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{C\times D}}(x, y) = \frac{\dot{\mathrm{d}}\mu_1}{\mathrm{d}\mathscr{H}^{n_1}|_C}(x) \frac{\dot{\mathrm{d}}\mu_2}{\mathrm{d}\mathscr{H}^{n_2}|_D}(y) \tag{4.1}$$

can be chosen in compliance with Lemma 3.1.3.

Proof. We check Definition 4.1.1:

For Part (i), we have $\mu_1 \times \mu_2(\mathbb{R}^{2N}) = \mu_1(\mathbb{R}^N)\mu_2(\mathbb{R}^N) < \infty$, as both μ_1 and μ_2 are finite by Part (i) of Definition 3.1.1. Parts (ii) and (iii) are equivalent to Part (ii) of Definition 3.1.1. To show Part (iv), we apply Theorem 2.4.3 (Fubini's Theorem) for an arbitrary $A \in \mathscr{B}^{2N}$,

$$\begin{aligned}
\mu_1 \times \mu_2(A) &= \int \mathbb{1}_A \,\mathrm{d}(\mu_1 \times \mu_2) \\
&= \iint \mathbb{1}_A(x,y) d\mu_1(x)\,\mathrm{d}\mu_2(y) \\
&= \iint \mathbb{1}_A(x,y) \frac{\mathrm{d}\mu_1}{\mathrm{d}\mathscr{H}^{n_1}|_C}(x) \frac{\mathrm{d}\mu_2}{\mathrm{d}\mathscr{H}^{n_2}|_D}(y)\,\mathrm{d}\mathscr{H}^{n_1}|_C(x)\,\mathrm{d}\mathscr{H}^{n_2}|_D(y) \\
&= \int \mathbb{1}_A(x,y) \frac{\mathrm{d}\mu_1}{\mathrm{d}\mathscr{H}^{n_1}|_C}(x) \frac{\mathrm{d}\mu_2}{\mathrm{d}\mathscr{H}^{n_2}|_D}(y)\,\mathrm{d}(\mathscr{H}^{n_1}|_C \times \mathscr{H}^{n_2}|_D)(x,y).
\end{aligned}$$

As A was arbitrary this proves that

$$\begin{aligned}
\mu_1 \times \mu_2 &= \frac{\mathrm{d}\mu_1}{\mathrm{d}\mathscr{H}^{n_1}|_C}(x) \frac{\mathrm{d}\mu_2}{\mathrm{d}\mathscr{H}^{n_2}|_D}(y)(\mathscr{H}^{n_1}|_C \times \mathscr{H}^{n_2}|_D)(x,y) \\
&\overset{(a)}{=} \frac{\mathrm{d}\mu_1}{\mathrm{d}\mathscr{H}^{n_1}|_C}(x) \frac{\mathrm{d}\mu_2}{\mathrm{d}\mathscr{H}^{n_2}|_D}(y)(\mathscr{H}^{n_1+n_2}|_{C\times D}(x,y),
\end{aligned} \tag{4.2}$$

where (a) follows from Lemma 2.6.13. As a result, an $\mathscr{H}^{n_1}|_C \times \mathscr{H}^{n_2}|_D$-null set is also a $\mu_1 \times \mu_2$-null set.

Now define $\frac{\dot{\mathrm{d}}\mu_1\times\mu_2}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{C\times D}}$, as in (4.1). Considering Remark 3.1.4, the dotted Radon-Nikodým derivatives can also be used in (4.2), which shows (3.2), as we already have $(x,y) \notin C \times D \implies \frac{\dot{\mathrm{d}}\mu_1}{\mathrm{d}\mathscr{H}^{n_1}|_C}(x) = 0$ or $\frac{\dot{\mathrm{d}}\mu_2}{\mathrm{d}\mathscr{H}^{n_2}|_D}(y) = 0$. This reasoning also shows that $\frac{\dot{\mathrm{d}}\mu_1\times\mu_2}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{C\times D}}$ satisfies (3.1). □

The following corollary now follows immediately.

Corollary 4.1.6. *For $n_1, n_2 \in \{0, 1, \ldots, N\}$ let μ be a combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure on $\mathbb{R}^{2N}$ with C and D denoting a first and second support, respectively. Let μ_1 and μ_2 denote the first and second marginal of μ, respectively, as defined in Definition 4.1.3. Then the product measure $\mu_1 \times \mu_2$ is a combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure on $(\mathbb{R}^{2N}, \mathscr{B}^{2N})$ with C and D as first and second support, respectively. Furthermore,*

$$\frac{\dot{\mathrm{d}}\mu_1 \times \mu_2}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{C\times D}}(x,y) = \frac{\dot{\mathrm{d}}\mu_1}{\mathrm{d}\mathscr{H}^{n_1}|_C}(x) \frac{\dot{\mathrm{d}}\mu_2}{\mathrm{d}\mathscr{H}^{n_2}|_D}(y)$$

can be chosen in compliance with Lemma 3.1.3.

Proof. The result follows immediately from Lemmas 4.1.4 and 4.1.5. □

In Corollary 4.1.6 we are constructing a measure from the two marginals of a combined

$\mathscr{B}$-countably (n_1, n_2)-rectifiable measure. We combine them to a new combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure as if they were independent. This construction will be useful for defining mutual information later on, as mutual information may be defined as the "deviation" (Kullback-Leibler divergence) from independence.

4.2. Mixtures

With the aim of deriving more general results, we will now consider linear combinations of $\mathscr{B}$-countably m-rectifiable measures for different $m \in \{0, 1, \ldots, M\}$. We will restrict ourselves to probability measures as our primary object of investigation, although the results in this section hold for finite measures in general and the restriction is not used in any of the proofs in this section.

First, we will introduce some notation. Let $M \in \mathbb{N}$ and for every $m \in \{0, 1, \ldots, M\}$,

(a1) μ_m is a $\mathscr{B}$-countably m-rectifiable measure on $(\mathbb{R}^M, \mathscr{B}^M)$,

(a2) $\mu_m(\mathbb{R}^M) = 1$,

(a3) E_m is a support of the measure μ_m, with

$$E_m \cap E_{m'} = \varnothing \quad \text{for } m \neq m', \text{ and} \tag{4.3}$$

(a4) $p_m \in \mathbb{R}^+$ with $\sum_m p_m = 1$.

(a5) We define the measure μ as

$$\mu := \sum_m p_m \mu_m$$

and introduce the random variable $X = \mathrm{id}_{\mathbb{R}^M}$. Note that by Assumptions (a2) and (a4), μ is a probability measure.

Remark 4.2.1. Note that (4.3) can be satisfied without loss of generality considering Lemma A.3.1.

Theorem 4.2.2. *Under Assumptions (a1) to (a5), μ can be written as*

$$\mu = \left(\sum_m p_m \frac{\mathrm{d}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}} \right) \left(\sum_m \mathscr{H}^m|_{E_m} \right).$$

Proof. The result is a special case of Lemma 3.1.6 with $\mathfrak{M} = \{0, 1, \ldots, M\}$, $I_m = \{0\}$, and $p_m^0 = p_m$. □

Adapting the setup given in Assumptions (a1) to (a5) to account for combined measures, we introduce the following notation, which we will use extensively throughout the rest of this section.

Let $N \in \mathbb{N}$ and for all $(n_1, n_2) \in \{0, 1, \ldots, N\}^2$,

(b1) the measure μ_{n_1,n_2} is a combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable measure on $(\mathbb{R}^{2N}, \mathscr{B}^{2N})$,

(b2) C_{n_1,n_2} and D_{n_1,n_2} are a first and second support of the measure μ_{n_1,n_2}, respectively,

(b3) $\mu_{n_1,n_2}(\mathbb{R}^{2N}) = 1$ for all $(n_1, n_2) \in \{0, 1, \ldots, N\}^2$,

(b4) $\mu^1_{n_1,n_2}$ and $\mu^2_{n_1,n_2}$ denote the first and second marginal of μ_{n_1,n_2}, respectively,

(b5) $p_{n_1,n_2} \in \mathbb{R}^+$, with $\sum_{n_1,n_2} p_{n_1,n_2} = 1$, and

(b6) we introduce the following notation[2]

$$
\begin{aligned}
E_{n_1,n_2} &:= C_{n_1,n_2} \times D_{n_1,n_2}, \\
E_n &:= \bigcup_{n=n_1+n_2} E_{n_1,n_2}, \qquad (4.4)\\
\mu_n &:= \sum_{n=n_1+n_2} p_{n_1,n_2} \mu_{n_1,n_2}, \\
C_{n_1} &:= \bigcup_{n_2} C_{n_1,n_2}, \qquad (4.5)\\
D_{n_2} &:= \bigcup_{n_1} D_{n_1,n_2}, \\
F_{n_1,n_2} &:= C_{n_1} \times D_{n_2} = \bigcup_{n_2',n_1'} C_{n_1,n_2'} \times D_{n_1',n_2}, \text{ and} \qquad (4.6)\\
F_n &:= \bigcup_{n=n_1+n_2} F_{n_1,n_2}, \qquad (4.7)
\end{aligned}
$$

where we require

$$
F_{n_1,n_2} \cap F_{m_1,m_2} = \varnothing \quad \text{for } (n_1, n_2) \neq (m_1, m_2). \tag{4.8}
$$

(b7) Furthermore, we define the measure μ as

$$
\mu := \sum_{n_1,n_2} p_{n_1,n_2} \mu_{n_1,n_2} \tag{4.9}
$$

[2] We use the symbol $\bigcup_{n=n_1+n_2}$ $\left(\sum_{n=n_1+n_2}\right)$ to denote the union (sum) of sets indexed by $\{(n_1, n_2) \in \{0, 1, \ldots, N\}^2 : n_1 + n_2 = n\}$.

Measure	Support
μ_{n_1,n_2}	E_{n_1,n_2}
$\mu^1_{n_1,n_2}$	C_{n_1,n_2}
$\mu^2_{n_1,n_2}$	D_{n_1,n_2}
μ_n	E_n
$\sum_{n_2} p_{n_1,n_2}\mu^1_{n_1,n_2}$	C_{n_1}
$\sum_{n_1} p_{n_1,n_2}\mu^2_{n_1,n_2}$	D_{n_2}
$\sum_{n_1',n_2'} p_{n_1,n_2'} p_{n_1',n_2} \mu^1_{n_1,n_2'} \times \mu^2_{n_1',n_2}$	F_{n_1,n_2}
$\sum_{n=n_1+n_2} \sum_{n_1',n_2'} p_{n_1,n_2'} p_{n_1',n_2} \mu^1_{n_1,n_2'} \times \mu^2_{n_1',n_2}$	F_n

Table 4.1.: Measures and their respective supports in Assumptions (b1) to (b8)

and additionally μ_1 and μ_2 shall denote the first and second marginal of μ (see Definition 4.1.3).

(b8) Two random variables $X : \mathbb{R}^{2N} \to \mathbb{R}^N$ and $Y : \mathbb{R}^{2N} \to \mathbb{R}^N$ are defined as the projections on the first N and the second N components, respectively. These induce the measures $\mu_X = \mu_1$ and $\mu_Y = \mu_2$ on $(\mathbb{R}^N, \mathscr{B}^N)$.

Note that by Assumptions (b3) and (b5) μ (as defined in (4.9)) is a probability measure

Remark 4.2.3. The sets introduced in Assumption (b6) will serve as supports for certain measures. Table 4.1 lists the measures and their corresponding supports. This can be seen using Remark 3.1.2.

Remark 4.2.4. Note that (4.8) can be satisfied without loss of generality considering Lemma A.3.2.

Remark 4.2.5. Furthermore, (4.8) also implies

$$F_n \cap F_m = \varnothing \quad \text{for } n \neq m, \tag{4.10}$$

$$E_n \cap E_m = \varnothing \quad \text{for } n \neq m, \tag{4.11}$$

$$E_{n_1,n_2} \cap E_{m_1,m_2} = \varnothing \quad \text{for } (n_1,n_2) \neq (m_1,m_2), \tag{4.12}$$

$$C_{n_1} \cap C_{m_1} = \varnothing \quad \text{for } n_1 \neq m_1, \text{ and}$$

$$D_{n_2} \cap D_{m_2} = \varnothing \quad \text{for } n_2 \neq m_2.$$

We can now prove the following theorem.

Theorem 4.2.6. *Under Assumptions (b1) to (b8), the measure μ can be expressed as*

$$\mu = \left(\sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}} \right) \left(\sum_n \mathcal{H}^n|_{E_n} \right). \tag{4.13}$$

Proof. From Lemma 4.1.2 we know that μ_{n_1,n_2} is a $\mathcal{B}$-countably (n_1+n_2)-rectifiable measure. Together with (4.11) this shows that all the requirements for Lemma 3.1.6 are met and we can apply it with $\mathfrak{M} = \{0,1,\ldots,2N\}$, $I_m = I_n = \{(n_1,n_2) \in \{0,1,\ldots,N\}^2 : n = n_1+n_2\}$, and $p_m^i = p_n^{(n_1,n_2)} = p_{n_1,n_2}$. This gives us

$$\mu = \left(\sum_n \sum_{n=n_1+n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}} \right) \left(\sum_n \mathcal{H}^n|_{E_n} \right),$$

using (4.4). Rearranging the finite sum yields (4.13). □

Explicit use of the supports E_{n_1,n_2} in (4.13) is necessary, i.e., substituting $\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}$ for $\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}}$ is not possible. This is due to the fact that the measure is comprised of several Hausdorff measures with different dimensions. Thus, changing the function on an $\mathcal{H}^m$-null set, A, might influence integration as A may not be a $\mathcal{H}^{m'}$-null set for $m' < m$.

Theorem 4.2.7. *Under Assumptions (b1) to (b8), the measure $\mu_1 \times \mu_2$ can be written as*

$$\mu_1 \times \mu_2 = \left(\sum_{n_1,n_2',n_1',n_2} p_{n_1,n_2'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathcal{H}^{n_1}|_{C_{n_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathcal{H}^{n_2}|_{D_{n_1',n_2}}}(y) \right) \left(\sum_n \mathcal{H}^n|_{F_n}(x,y) \right). \tag{4.14}$$

Proof. From the definition of the marginal (see Definition 4.1.3) we know that for all $A \in \mathcal{B}^N$,

$$\mu_1(A) = \mu(A \times \mathbb{R}^N) = \sum_{n_1,n_2} p_{n_1,n_2} \mu_{n_1,n_2}(A \times \mathbb{R}^N) = \sum_{n_1,n_2} p_{n_1,n_2} \mu^1_{n_1,n_2}(A). \tag{4.15}$$

Therefore, $\mu_1 = \sum_{n_1,n_2} p_{n_1,n_2} \mu^1_{n_1,n_2}$ and accordingly $\mu_2 = \sum_{n_1,n_2} p_{n_1,n_2} \mu^2_{n_1,n_2}$. As the product measure $\mu_1 \times \mu_2$ is unique (see Definition 2.4.2) and for any $A, B \in \mathcal{B}^N$,

$$\begin{aligned} \mu_1 \times \mu_2(A \times B) &= \mu_1(A)\mu_2(B) \\ &\overset{(4.15)}{=} \left(\sum_{n_1,n_2} p_{n_1,n_2} \mu^1_{n_1,n_2}(A) \right) \left(\sum_{n_1,n_2} p_{n_1,n_2} \mu^2_{n_1,n_2}(B) \right) \\ &= \sum_{n_1,n_2',n_1',n_2} p_{n_1,n_2'} \mu^1_{n_1,n_2'}(A) p_{n_1',n_2} \mu^2_{n_1',n_2}(B) \end{aligned}$$

$$= \sum_{n_1,n_2',n_1',n_2} p_{n_1,n_2'} p_{n_1',n_2} (\mu^1_{n_1,n_2'} \times \mu^2_{n_1',n_2})(A \times B),$$

we obtain that

$$\mu_1 \times \mu_2 = \sum_{n_1,n_2',n_1',n_2} p_{n_1,n_2'} p_{n_1',n_2} (\mu^1_{n_1,n_2'} \times \mu^2_{n_1',n_2}).$$

Corollary 4.1.6, applied to the products $\mu^1_{n_1,n_2'} \times \mu^2_{n_1',n_2}$, guarantees their combined $\mathscr{B}$-countably (n_1, n_2)-rectifiability and gives us

$$\frac{\dot{\mathrm{d}}(\mu^1_{n_1,n_2'} \times \mu^2_{n_1',n_2})}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{C_{n_1,n_2'} \times D_{n_1',n_2}}}(x,y) = \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}|_{D_{n_1',n_2}}}(y).$$

Together with (4.10) this guarantees that the prerequisites of Lemma 3.1.6 are satisfied. Applying Lemma 3.1.6 with $\mathfrak{M} = \{0, 1, \dots, 2N\}$, $I_m = \{(n_1, n_2) \in \{0, 1, \dots, N\}^2 : n_1 + n_2 = n\}$, and $p^i_m = p_n^{(n_1,n_2)} = p_{n_1,n_2}$, and substituting (4.7) yields (4.14). □

4.3. Mutual Information

We will present a definition of mutual information from [Gra13] in a very general setting and then apply it in the context of combined rectifiable probability measures. To this end we will first introduce some definitions, we will then use to define mutual information.

Definition 4.3.1. *[Gra13, Section 2.3] Let μ and ν denote probability measures on the measurable space $(\Omega, \mathfrak{S})$ and let $\mathscr{Q} \subseteq \mathfrak{S}$, be a finite measurable partition of Ω, i.e., the elements of $\mathscr{Q}$ are disjoint, and their union equals Ω. The* relative entropy *of $\mathscr{Q}$ with measure μ w.r.t. ν is defined as*

$$H_{\mu\|\nu}(\mathscr{Q}) = \sum_{q \in \mathscr{Q}} \mu(q) \log \frac{\mu(q)}{\nu(q)}. \tag{4.16}$$

It should be noted that the sum in (4.16) is finite and thus, $H_{\mu\|\nu}(\mathscr{Q})$ is well-defined although it may assume the value $+\infty$ in case $\mu(q) > 0$ and $\nu(q) = 0$ for some $q \in \mathscr{Q}$.

Definition 4.3.2. *[Gra13, Section 5.2, (5.1)] Let μ and ν denote probability measures on the measurable space $(\Omega, \mathfrak{S})$. The* divergence *of μ* w.r.t. *ν is defined as*

$$D(\mu\|\nu) = \sup_{\mathscr{Q}} H_{\mu\|\nu}(\mathscr{Q}),$$

where the supremum is taken over all finite measurable partitions of Ω.

Again, $D(\mu||\nu) = +\infty$ may occur.

Definition 4.3.3. *[Gra13, Section 5.5, (5.29)] Let* $(\Omega, \mathfrak{S}, \mu)$ *be a probability space and with* $N \in \mathbb{N}$ *let* $X : \Omega \to \mathbb{R}^N$ *and* $Y : \Omega \to \mathbb{R}^N$ *be two random variables (see Definition 2.5.2) on that space. The* mutual information *of* X *and* Y *is defined as*[3]

$$I(X;Y) = D(\mu_{(X,Y)}||\mu_X \times \mu_Y). \tag{4.17}$$

To obtain an explicit expression for the mutual information, given in Definition 4.3.3, one resorts to the Perez-Yaglom-Gelfand theorem which is stated in the following lemma.

Lemma 4.3.4. *[Gra13, Lemma 5.2.3]* (Perez-Yaglom-Gelfand Theorem) *Given two probability measures* μ *and* ν *on a common measurable space* $(\Omega, \mathfrak{S})$, *if* $\mu \not\ll \nu$, *then*

$$D(\mu||\nu) = \infty.$$

If $\mu \ll \nu$, *then the Radon-Nikodým derivative* $f = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ *exists and*

$$D(\mu||\nu) = \int \log f \,\mathrm{d}\mu = \int f \log f \,\mathrm{d}\nu.$$

This directly amounts to a formula for the mutual information of X and Y as given in Assumption (b8). We will use the symbol $H_2(p_{n_1,n_2})$ to denote

$$H_2(p_{n_1,n_2}) := \sum_{n_1,n_2} p_{n_1,n_2} \log \frac{1}{p_{n_1,n_2}}. \tag{4.18}$$

Theorem 4.3.5. *Under Assumptions (b1) to (b8),* $I(X;Y)$ *can be expressed as*

$$I(X;Y) = -H_2(p_{n_1,n_2}) + \sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}}(x,y) \log \left(\frac{\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}}(x,y)}{\sum_{n_1',n_2'} p_{n_1,n_2'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}}(y)} \right) \mathrm{d}\mathscr{H}^{n_1+n_2}(x,y) \tag{4.19}$$

[3]The measures with a random variable in the subscript are induced, as given in Definition 2.2.3 and (X,Y) denotes the combined random variable as in Remark 2.5.4.

if the integrand in (4.19) is defined, otherwise

$$I(X;Y)=\infty.$$

Proof. Note that the combined random variable (X,Y) equals the identity, thus $\mu_{(X,Y)}=\mu$.

Case $\mu \ll \mu_1 \times \mu_2$. We define

$$\nu := \sum_n \mathscr{H}^n|_{F_n}. \tag{4.20}$$

The sets $F_{n_1,n_2}=C_{n_1}\times D_{n_2}$ are $\mathscr{B}$-countably (n_1+n_2)-rectifiable by Lemma 2.6.13. Thus, by the definition of F_n in (4.7) and Remark 2.6.11, F_n is $\mathscr{B}$-countably n-rectifiable. By Lemma 2.6.8, $\mathscr{H}^n|_{F_n}$ is σ-finite and, thus, also is ν as it is the finite sum of σ-finite measures.

Note that $\sum_n \mathscr{H}^n|_{E_n}$ can be replaced by $\sum_n \mathscr{H}^n|_{F_n}$ in Theorem 4.2.6. This is justified by

$$\sum_{n_1,n_2} p_{n_1,n_2}\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x)=0 \quad \text{for all } x\notin \bigcup_n E_n.$$

Together with Theorem 4.2.7, this gives us expressions for the Radon-Nikodým derivatives of μ and $\mu_1\times\mu_2$ w.r.t. ν. We can write the Radon-Nikodým derivative of μ w.r.t. $\mu_1\times\mu_2$ using Part (i) of Proposition 2.3.4 as

$$\frac{d\mu}{d(\mu_1\times\mu_2)}=\frac{d\mu}{d\nu}\left(\frac{d(\mu_1\times\mu_2)}{d\nu}\right)^{-1}.$$

because $\frac{\mathrm{d}(\mu_1\times\mu_2)}{\mathrm{d}\nu}\neq 0$, $\mu_1\times\mu_2$-a.e.

By (4.17) and Lemma 4.3.4 we then have

$$\begin{aligned}
I(X;Y)&=D(\mu||\mu_1\times\mu_2)\\
&=\int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\left(\frac{\mathrm{d}(\mu_1\times\mu_2)}{\mathrm{d}\nu}\right)^{-1}\right)\mathrm{d}\mu\\
&=\int \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\left(\frac{\mathrm{d}(\mu_1\times\mu_2)}{\mathrm{d}\nu}\right)^{-1}\right)\mathrm{d}\nu.
\end{aligned}\tag{4.21}$$

Inserting (4.13) and (4.14) into (4.21) yields

$$
\begin{aligned}
I(X;Y) = &\int \left(\sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y) \right) \\
&\cdot \log \left(\frac{\sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y)}{\sum_{n_1,n_2,n_1',n_2'} p_{n_1,n_2'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathcal{H}^{n_1}|_{C_{n_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathcal{H}^{n_2}|_{D_{n_1',n_2}}}(y)} \right) \mathrm{d}\nu(x,y) \\
\overset{(a)}{=} &\sum_{n_1,n_2} \int \left(\sum_{\tilde{n}_1,\tilde{n}_2} p_{\tilde{n}_1,\tilde{n}_2} \frac{\dot{\mathrm{d}}\mu_{\tilde{n}_1,\tilde{n}_2}}{\mathrm{d}\mathcal{H}^{\tilde{n}_1+\tilde{n}_2}|_{E_{\tilde{n}_1,\tilde{n}_2}}}(x,y) \right) \\
&\cdot \log \left(\frac{\sum_{\tilde{n}_1,\tilde{n}_2} p_{\tilde{n}_1,\tilde{n}_2} \frac{\dot{\mathrm{d}}\mu_{\tilde{n}_1,\tilde{n}_2}}{\mathrm{d}\mathcal{H}^{\tilde{n}_1+\tilde{n}_2}|_{E_{\tilde{n}_1,\tilde{n}_2}}}(x,y)}{\sum_{\tilde{n}_1,\tilde{n}_2,n_1',n_2'} p_{\tilde{n}_1,n_2'} p_{n_1',\tilde{n}_2} \frac{\dot{\mathrm{d}}\mu^1_{\tilde{n}_1,n_2'}}{\mathrm{d}\mathcal{H}^{\tilde{n}_1}|_{C_{\tilde{n}_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',\tilde{n}_2}}{\mathrm{d}\mathcal{H}^{\tilde{n}_2}|_{D_{n_1',\tilde{n}_2}}}(y)} \right) \mathrm{d}\mathcal{H}^{n_1+n_2}|_{F_{n_1,n_2}}(x,y) \\
\overset{(b)}{=} &\sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y) \\
&\cdot \log \left(\frac{p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y)}{\sum_{\tilde{n}_1,\tilde{n}_2,n_1',n_2'} p_{\tilde{n}_1,n_2'} p_{n_1',\tilde{n}_2} \frac{\dot{\mathrm{d}}\mu^1_{\tilde{n}_1,n_2'}}{\mathrm{d}\mathcal{H}^{\tilde{n}_1}|_{C_{\tilde{n}_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',\tilde{n}_2}}{\mathrm{d}\mathcal{H}^{\tilde{n}_2}|_{D_{n_1',\tilde{n}_2}}}(y)} \right) \mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}(x,y) \\
\overset{(c)}{=} &\sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y) \\
&\cdot \log \left(\frac{p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y)}{\sum_{n_1',n_2'} p_{n_1,n_2'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathcal{H}^{n_1}|_{C_{n_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathcal{H}^{n_2}|_{D_{n_1',n_2}}}(y)} \right) \mathrm{d}\mathcal{H}^{n_1+n_2}|_{E_{n_1,n_2}}(x,y)
\end{aligned}
$$

$$\overset{(d)}{=} \sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y)$$
$$\cdot \log\left(\frac{\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}}(x,y)}{\sum_{n_1',n_2'} p_{n_1,n_2'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}|_{D_{n_1',n_2}}}(y)} \right) \mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}(x,y)$$
$$-H_2(p_{n_1,n_2}) \tag{4.22}$$

where (a) follows from (4.20). For justifying (b), assume $(\tilde{n}_1,\tilde{n}_2) \neq (n_1,n_2)$. Then for $(x,y) \in E_{n_1,n_2}$, $\frac{\dot{\mathrm{d}}\mu_{\tilde{n}_1,\tilde{n}_2}}{\mathrm{d}\mathscr{H}^{\tilde{n}_1+\tilde{n}_2}|_{E_{\tilde{n}_1,\tilde{n}_2}}}(x,y) = 0$, by (3.1) and (4.12). Part (i) of Theorem 2.2.10 is used to pull p_{n_1,n_2} out of the integral. To justify (c), assume $\tilde{n}_1 \neq n_1$ and $x \in E_{n_1,n_2}$. Then $x \in C_{n_1}$ and $\frac{\dot{\mathrm{d}}\mu^1_{\tilde{n}_1,n_2'}}{\mathrm{d}\mathscr{H}^{\tilde{n}_1}|_{C_{\tilde{n}_1,n_2'}}}(x) = 0$ for any n_2', which follows from the fact that C_{n_1} and $C_{\tilde{n}_1}$ are disjoint (see Remark 4.2.5). The case $\tilde{n}_2 \neq n_2$ can be treated the same way. In (d) the additive term $\log p_{n_1,n_2}$ is split from the logarithm and we use that μ_{n_1,n_2} is a probability measure by Assumption (b3).

To prove (4.19), it is only left to argue that we can drop the supports in all the dotted Radon-Nikodým derivatives in (4.22). We can argue with Lemma 3.1.5 that $\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}} = \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}}$ for the purpose of this integration, i.e., $\mathscr{H}^{n_1+n_2}$-a.e. We have to show that the set

$$G := \left\{ (x,y) \in E_{n_1,n_2} : \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2'}}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}|_{D_{n_1',n_2}}}(y) \neq \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}}(y) \right\}$$

satisfies $\mathscr{H}^{n_1+n_2}(G) = 0$. We write $G \subseteq (G_1 \times D_{n_1,n_2}) \cup (C_{n_1,n_2} \times G_2)$ with

$$G_1 := \left\{ x \in C_{n_1,n_2} : \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2'}}}(x) \neq \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}}(x) \right\} \quad \text{and}$$
$$G_2 := \left\{ y \in D_{n_1,n_2} : \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}|_{D_{n_1',n_2}}}(y) \neq \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}}(y) \right\}.$$

The set $G_1 \subseteq C_{n_1,n_2}$ is Borel, which can be seen by writing it like[4]

$$G_1 = C_{n_1,n_2} \cap \left(\frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2'}}} - \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}} \right)^{-1} (\{0\}).$$

Thus, G_1 is $\mathscr{B}$-countably n_1-rectifiable by Part (ii) of Corollary 2.6.9. Analogously one proves the $\mathscr{B}$-countably n_2-rectifiability of G_2. Lemma 2.6.13 then guarantees $0 = \mathscr{H}^{n_1+n_2}(G_1 \times D_{n_1,n_2}) = \mathscr{H}^{n_1+n_2}(C_{n_1,n_2} \times G_2)$, which shows $\mathscr{H}^{n_1+n_2}(G) = 0$ by the monotonicity of measures, Part (ii) of Proposition 2.1.4.

And, finally, Lemma 3.1.3 allows us to integrate w.r.t. $\mathscr{H}^{n_1+n_2}$ instead of $\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}$.

Case $\mu \not\ll \mu_1 \times \mu_2$. This is equivalent to the existence of a set $A \in \mathscr{B}^{2N}$ with $\mu(A) > 0$ and $\mu_1 \times \mu_2(A) = 0$. We can conclude that for the set $A' := \{x \in A : \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) > 0\}$, $\nu(A') > 0$ holds. But as $\frac{\mathrm{d}\mu_1 \times \mu_2}{\mathrm{d}\nu}(x) = 0$ for ν-a.e. $x \in A' \subseteq A$, the integrand in (4.21) is not defined on a set of positive measure and, by Lemma 4.3.4, $I(X;Y) = \infty$. □

Remark 4.3.6. It is worthwhile mentioning that the distinction between the two cases, when the integrand in (4.19) is defined, and when it is not, is just a consequence of the definition of measurable functions that we introduced. If we allowed for measurable functions taking values in the extended reals $\bar{\mathbb{R}}$, the integral in (4.19) would be defined in any case and yield $+\infty$, in the case $\mu \not\ll \mu_1 \times \mu_2$.

4.4. Connection between Mutual Information and Entropy

In this section we will develop expressions for the entropy of linear combinations of combined rectifiable probability measures and use them to show a connection between the mutual information, as presented in Section 4.3 in the same setting, and entropy, as treated in Section 3.3. This intricate connection also serves as an additional motivation and justification for our definition of entropy (Definition 3.3.2).

We will use the notation

$$H_2(p_m) := \sum_m p_m \log \frac{1}{p_m}$$

as in (4.18).

[4] Here $f^{-1}(\cdot)$ denotes the preimage.

Lemma 4.4.1. *Under Assumptions (a1) to (a5), if the entropy of μ (entropy of X) w.r.t. $\nu := \sum_m \mathscr{H}^m|_{E_m}$ (as defined in Definition 3.3.2) is defined, it can be expressed as*

$$h_\nu(X) = h_\nu(\mu) = H_2(p_m) + \sum_m p_m h(\mu_m). \tag{4.23}$$

Proof. By Theorem 4.2.2,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu} = \sum_m p_m \frac{\dot{\mathrm{d}}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}}. \tag{4.24}$$

This gives us

$$\begin{aligned}
h_\nu(\mu) &\overset{(3.17)}{=} \int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)^{-1} \mathrm{d}\mu \\
&\overset{(4.24)}{=} \int \left(\sum_m p_m \frac{\dot{\mathrm{d}}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}}\right) \log\left(\sum_m p_m \frac{\dot{\mathrm{d}}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}}\right)^{-1} d\nu \\
&\overset{(a)}{=} \sum_m \int \left(\sum_{\tilde{m}} p_{\tilde{m}} \frac{\dot{\mathrm{d}}\mu_{\tilde{m}}}{\mathrm{d}\mathscr{H}^{\tilde{m}}|_{E_{\tilde{m}}}}\right) \log\left(\sum_{\tilde{m}} p_{\tilde{m}} \frac{\dot{\mathrm{d}}\mu_{\tilde{m}}}{\mathrm{d}\mathscr{H}^{\tilde{m}}|_{E_{\tilde{m}}}}\right)^{-1} \mathrm{d}\mathscr{H}^m|_{E_m} \\
&\overset{(b)}{=} \sum_m p_m \int \frac{\dot{\mathrm{d}}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}} \log\left(p_m \frac{\dot{\mathrm{d}}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}}\right)^{-1} \mathrm{d}\mathscr{H}^m|_{E_m} \\
&\overset{(c)}{=} H_2(p_m) + \sum_m p_m \int \frac{\dot{\mathrm{d}}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}} \log\left(\frac{\dot{\mathrm{d}}\mu_m}{\mathrm{d}\mathscr{H}^m|_{E_m}}\right)^{-1} \mathrm{d}\mathscr{H}^m|_{E_m},
\end{aligned} \tag{4.25}$$

where (a) follows from the definition of ν, (b) follows from the fact that $\frac{\dot{\mathrm{d}}\mu_{\tilde{m}}}{\mathrm{d}\mathscr{H}^{\tilde{m}}|_{E_{\tilde{m}}}}(x) = 0$ for any $x \in E_m$ if $\tilde{m} \neq m$ and from Part (i) of Theorem 2.2.10. In (c), the additive term $\log p_m$ is split from the logarithm and the fact that μ_m is a probability measure (Assumption (a2)) is used. Substituting (3.18) and (3.19) in (4.25) completes the proof. □

Lemma 4.4.1 can in particular be applied to a discrete continuous mixture on $\mathbb{R}$, i.e., $\mu = p_1\mu_1 + (1-p_1)\mu_0$. The entropy of μ is given by Lemma 4.4.1 as

$$h(\mu) = -p_1 \log(p_1) - (1-p_1)\log(1-p_1) + p_1 h(\mu_1) + (1-p_1)h(\mu_0), \tag{4.26}$$

where $h(\mu_0)$ and $h(\mu_1)$ are the entropy of a discrete random variable and differential entropy, respectively, by Remark 3.3.5. The entropy given in (4.26) coincides with Rényi-Entropy, introduced by Alfréd Rényi in [Ré59], by virtue of [Ré59, Theorem 3].

It is worthwhile mentioning that (4.23) shows that the entropy of μ w.r.t. ν does not depend on the individual supports E_m. It does not, however, fully justify the definition of an entropy of μ without referring to ν, as it was done for a single measure in Definition 3.3.4. This is due to the fact that entropy might still depend on the particular representation of μ as linear combination of rectifiable measures. We will show in the following that this is not the case.

Lemma 4.4.2. *Under Assumptions (a1) to (a5), given $p'_m \in [0,1]$ and $\mathscr{B}$-countably m-rectifiable measures μ'_m for $m \in \{0,1,\ldots,M\}$, such that*

$$\mu = \sum_m p_m \mu_m = \sum_m p'_m \mu'_m, \tag{4.27}$$

then necessarily $p'_m = p_m$ and, whenever $p_m \neq 0$, $\mu'_m = \mu_m$ for every $m \in \{0,1,\ldots,M\}$.

Proof. We do a proof by induction. Let E'_m denote a support of μ'_m. For some $k \in \{0,1,\ldots,M\}$, suppose $p_m = p'_m$ and if $p_m \neq 0$, then $\mu_m = \mu'_m$ for all $m \in \{0,1,\ldots,k-1\}$. For the base case, $k=0$, this is satisfied since $\{0,1,\ldots,k-1\} = \varnothing$. The induction step is done as follows. We have for any set $S \in \mathscr{B}^M$,

$$\begin{aligned}
\sum_m p'_m \mu'_m(S \cap (E_k \cup E'_k)) &\overset{(4.27)}{=} \sum_m p_m \mu_m(S \cap (E_k \cup E'_k)) \\
\sum_{m \geq k} p'_m \mu'_m(S \cap (E_k \cup E'_k)) &\overset{(a)}{=} \sum_{m \geq k} p_m \mu_m(S \cap (E_k \cup E'_k)) \\
p'_k \mu'_k(S) = p'_k \mu'_k(S \cap (E_k \cup E'_k)) &\overset{(b)}{=} p_k \mu_k(S \cap (E_k \cup E'_k)) = p_k \mu_k(S),
\end{aligned} \tag{4.28}$$

where (a) follows from the induction hypothesis and the fact that the measures μ_m are finite and in (b), Corollary 2.6.12 is used to provide $\mu'_m(S \cap (E_k \cup E'_k)) = \mu_m(S \cap (E_k \cup E'_k)) = 0$ for $m > k$. Equation (4.28) holds for all $S \in \mathscr{B}^M$, thus in particular for $S = \mathbb{R}^M$. This yields $p_k = p'_k$, as both μ_k and μ'_k, are probability measures. That in turn results in $\mu_k = \mu'_k$ if $p_k \neq 0$.

Thus, proceeding inductively from $k=0$ to $k=M$ proves $p_m = p'_m$ and, if $p_m \neq 0$, also $\mu_m = \mu'_m$ for all $m \in \{0,1,\ldots,M\}$. □

This justifies the following definition.

Definition 4.4.3. *Under Assumptions (a1) to (a5), the entropy of μ (entropy of X) is defined as*

$$h(X) := h(\mu) := h_{\sum_m \mathscr{H}^m|_{E_m}}(\mu), \tag{4.29}$$

if $h_{\sum_m \mathscr{H}^m|_{E_m}}(\mu)$ is defined.

The entropy in (4.29) is well-defined as the representation $\mu = \sum_m p_m \mu_m$ is unique by Lemma 4.4.2 and thus (4.23) is also unique.

It should be noted that also under Assumptions (b1) to (b8), Definition 4.4.3 already supplies us with an entropy of μ, μ_1, and μ_2. This can be seen as follows. In Assumptions (b1) to (b8),

μ can be represented as $\mu = \sum_m p_m \mu_m$, like in Assumptions (a1) to (a5), choosing

$$p_m = \sum_{m=n_1+n_2} p_{n_1,n_2}, \text{ and}$$
$$\mu_m = \frac{1}{p_m} \sum_{m=n_1+n_2} p_{n_1,n_2} \mu_{n_1,n_2}$$

for $m \in \{0, 1, \ldots, 2N\}$. As one can see in Table 4.1, E_m (notation of Assumption (b6)) is a support of μ_m. Similarly, we can represent μ_1 in Assumption (b7) like μ in Assumption (a5). Using (4.15), μ_1 is given by

$$p_m = \sum_{n_2'} p_{m,n_2'}, \text{ and}$$
$$\mu_m = \frac{1}{p_m} \sum_{n_2'} p_{m,n_2'} \mu^1_{m,n_2'}$$

for $m \in \{0, 1, \ldots, N\}$. In Table 4.1, C_m is given as a support of μ_m. Also μ_2 can be expressed the same way.

Thus, Definition 4.4.3 is directly applicable to μ, μ_1, and μ_2, which justifies the following definition.

Definition 4.4.4. *Under Assumptions (b1) to (b8), we define the entropy of μ (entropy of (X, Y)) as*

$$h(X, Y) := h(\mu) := h_{\sum_n \mathscr{H}^n|_{E_n}}(\mu),$$

if $h_{\sum_n \mathscr{H}^n|_{E_n}}(\mu)$ is defined. Furthermore, the entropy of μ_1 (entropy of X) and the entropy of μ_2 (entropy of Y) are given by

$$h(X) := h(\mu_1) := h_{\sum_{n_1} \mathscr{H}^{n_1}|_{C_{n_1}}}(\mu_1), \text{ and}$$
$$h(Y) := h(\mu_2) := h_{\sum_{n_2} \mathscr{H}^{n_2}|_{D_{n_2}}}(\mu_2),$$

respectively, if the respective entropies $h_{\sum_{n_1} \mathscr{H}^{n_1}|_{C_{n_1}}}(\mu_1)$ and $h_{\sum_{n_2} \mathscr{H}^{n_2}|_{D_{n_2}}}(\mu_2)$ are defined.

The next results extend Lemma 4.4.1 to combined distributions and develop expressions for the entropies in Definition 4.4.4.

Lemma 4.4.5. *Under Assumptions (b1) to (b8), if the entropy of μ (entropy of (X, Y)) is defined, it satisfies*

$$h(X, Y) = h(\mu) = H_2(p_{n_1,n_2}) + \sum_{n_1,n_2} p_{n_1,n_2} h(\mu_{n_1,n_2}). \tag{4.30}$$

Proof. The proof follows the same steps as the proof of Lemma 4.4.1. We set $\nu := \sum_n \mathscr{H}^n|_{E_n}$ and thus have $h(\mu) = h_\nu(\mu)$ from Definition 4.4.4. By Theorem 4.2.6, we can insert

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu} = \sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}} \tag{4.31}$$

into (3.17), which results in

$$\begin{aligned}
h_\nu(\mu) &\overset{(3.17)}{=} \int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)^{-1} \mathrm{d}\mu \\
&\overset{(4.31)}{=} \int \left(\sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}}\right) \log\left(\sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}}\right)^{-1} \mathrm{d}\nu \\
&\overset{(a)}{=} \sum_{n_1,n_2} \int \left(\sum_{\tilde{n}_1,\tilde{n}_2} p_{\tilde{n}_1,\tilde{n}_2} \frac{\dot{\mathrm{d}}\mu_{\tilde{n}_1,\tilde{n}_2}}{\mathrm{d}\mathscr{H}^{\tilde{n}_1+\tilde{n}_2}|_{E_{\tilde{n}_1,\tilde{n}_2}}}\right) \log\left(\sum_{\tilde{n}_1,\tilde{n}_2} p_{\tilde{n}_1,\tilde{n}_2} \frac{\dot{\mathrm{d}}\mu_{\tilde{n}_1,\tilde{n}_2}}{\mathrm{d}\mathscr{H}^{\tilde{n}_1+\tilde{n}_2}|_{E_{\tilde{n}_1,\tilde{n}_2}}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}} \\
&\overset{(b)}{=} \sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}} \log\left(p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}} \\
&\overset{(c)}{=} H_2(p_{n_1,n_2}) + \sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}} \log\left(\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_1+n_2}|_{E_{n_1,n_2}},
\end{aligned}$$

where one can argue (a), (b), and (c) as for (4.25) in the proof of Lemma 4.4.1. Using Lemmas 3.1.5 and 3.3.3 gives us (4.30). □

Lemma 4.4.6. *Under Assumptions (b1) to (b8), the entropies of μ_1 (X) and μ_2 (Y) satisfy*

$$h(X) = h(\mu_1) = \sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1}} \log\left(\sum_{n_2'} p_{n_1,n_2'} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_1} \text{ and} \tag{4.32}$$

$$h(Y) = h(\mu_2) = \sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu^2_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_2}} \log\left(\sum_{n_1'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_2}, \tag{4.33}$$

respectively, if the integrals and sums in (4.32) and (4.33) are defined.

Proof. We will only prove (4.32), as (4.33) follows analogously.

Using $\nu_1 := \sum_{n_1} \mathscr{H}^{n_1}|_{C_{n_1}}$ gives us $h(\mu_1) = h_{\nu_1}(\mu_1)$ by Definition 4.4.4. From Lemma 4.1.4 we know that $\mu^1_{n_1,n_2}$ is a $\mathscr{B}$-countably n_1-rectifiable probability measure. Furthermore, we know $\mu_1 = \sum_{n_1,n_2} p_{n_1,n_2} \mu^1_{n_1,n_2}$ from (4.15). We can therefore apply Lemma 3.1.6, yielding

$$\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu_1} = \sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2}}}. \tag{4.34}$$

And when substituting in (3.17),

$$\begin{aligned}
h_{\nu_1}(\mu_1) &\overset{(3.17)}{=} \int \log\left(\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu_1}\right)^{-1} \mathrm{d}\mu_1 \\
&\overset{(4.34)}{=} \int \left(\sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2}}}\right) \log\left(\sum_{n_1,n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2}}}\right)^{-1} \mathrm{d}\nu_1 \\
&\overset{(a)}{=} \sum_{n_1} \int \left(\sum_{\tilde{n}_1,n_2} p_{\tilde{n}_1,n_2} \frac{\dot{\mathrm{d}}\mu^1_{\tilde{n}_1,n_2}}{\mathrm{d}\mathscr{H}^{\tilde{n}_1}|_{C_{\tilde{n}_1,n_2}}}\right) \log\left(\sum_{\tilde{n}_1,n_2} p_{\tilde{n}_1,n_2} \frac{\dot{\mathrm{d}}\mu^1_{\tilde{n}_1,n_2}}{\mathrm{d}\mathscr{H}^{\tilde{n}_1}|_{C_{\tilde{n}_1,n_2}}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1}} \\
&\overset{(b)}{=} \sum_{n_1} \int \left(\sum_{n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2}}}\right) \log\left(\sum_{n_2} p_{n_1,n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2}}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1}} \\
&\overset{(c)}{=} \sum_{n_1,n_2} p_{n_1,n_2} \int \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2}}} \log\left(\sum_{n_2'} p_{n_1,n_2'} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2'}}}\right)^{-1} \mathrm{d}\mathscr{H}^{n_1}|_{C_{n_1,n_2}},
\end{aligned} \tag{4.35}$$

where (a) follows from the definition of ν_1, (b) from the fact that $\frac{\dot{\mathrm{d}}\mu^1_{\tilde{n}_1,n_2}}{\mathrm{d}\mathscr{H}^{\tilde{n}_1}|_{C_{\tilde{n}_1,n_2}}}(x) = 0$ for $x \in C_{n_1}$ and $\tilde{n}_1 \neq n_1$. To justify (c), we use the linearity of the integral (Parts (i) and (ii) of Theorem 2.2.10) and $C_{n_1,n_2} \subseteq C_{n_1}$ with (3.1). We can drop the restriction on the domain of integration in (4.35) because of (3.1) and we can drop the supports in the dotted Radon-Nikodým derivatives, as the integration is w.r.t. $\mathscr{H}^{n_1}$ and the functions are thus considered equivalent (see Part (i) of Remark 2.2.7). This yields (4.32). □

Note that in the expression for $I(X;Y)$ in (4.19), we can factor the finite sum

$$\sum_{n_1',n_2'} p_{n_1,n_2'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}}(x) \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}}(y) = \left(\sum_{n_2'} p_{n_1,n_2'} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathscr{H}^{n_1}}(x)\right)\left(\sum_{n_1'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathscr{H}^{n_2}}(y)\right),$$

and thereby split up the logarithm, which yields

$$
\begin{aligned}
I(X;Y) = -H_2(p_{n_1,n_2}) + \sum_{n_1,n_2} p_{n_1,n_2} \int \Bigg[&\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}(x,y) \log\left(\frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}(x,y) \right) \\
&- \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}(x,y) \log\left(\sum_{n_2'} p_{n_1,n_2'} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathcal{H}^{n_1}}(x) \right) \\
&- \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}(x,y) \log\left(\sum_{n_1'} p_{n_1',n_2} \frac{\dot{\mathrm{d}}\mu^2_{n_1',n_2}}{\mathrm{d}\mathcal{H}^{n_2}}(y) \right) \Bigg] \mathrm{d}\mathcal{H}^{n_1+n_2}(x,y). \qquad (4.36)
\end{aligned}
$$

If all the entropies $h(X)$, $h(Y)$, and $h(X,Y)$ are defined and also their sum is defined, we can use the linearity of the Lebesgue integral (see Parts (i) and (ii) of Theorem 2.2.10) to split the integral and use Theorem 2.4.3 (Fubini's theorem) to show, e.g., for the third term in (4.36),

$$
\begin{aligned}
&\int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}(x,y) \log\left(\sum_{n_2'} p_{n_1,n_2'} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathcal{H}^{n_1}}(x) \right) \mathrm{d}\mathcal{H}^{n_1+n_2}(x,y) \\
&\overset{(a)}{=} \int \log\left(\sum_{n_2'} p_{n_1,n_2'} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathcal{H}^{n_1}}(x) \right) \left(\int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}(x,y) \, \mathrm{d}\mathcal{H}^{n_2}(y) \right) \mathrm{d}\mathcal{H}^{n_1}(x) \\
&\overset{(b)}{=} \int \log\left(\sum_{n_2'} p_{n_1,n_2'} \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2'}}{\mathrm{d}\mathcal{H}^{n_1}}(x) \right) \frac{\dot{\mathrm{d}}\mu^1_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1}}(x) \, \mathrm{d}\mathcal{H}^{n_1}(x) = h(X),
\end{aligned}
$$

where we used Lemma 2.6.13 and Theorem 2.4.3 in (a), and (b) can be justified by applying Fubini's Theorem (Theorem 2.4.3) to

$$
\mu^1_{n_1,n_2}(A) = \int_{A\times\mathbb{R}^N} \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}} \, \mathrm{d}\mathcal{H}^{n_1+n_2} = \int_A \int \frac{\dot{\mathrm{d}}\mu_{n_1,n_2}}{\mathrm{d}\mathcal{H}^{n_1+n_2}}(x,y) \, \mathrm{d}\mathcal{H}^{n_2}(y) \, \mathrm{d}\mathcal{H}^{n_1}(x).
$$

This gives us the following theorem.

Theorem 4.4.7. *Under Assumptions (b1) to (b8), if $h(X)$, $h(Y)$, and $h(X,Y)$ are defined then*

$$
I(X;Y) = h(X) + h(Y) - h(X,Y), \qquad (4.37)
$$

if the sum in (4.37) is defined.

Chapter 5.

Conclusions

In this thesis we introduced $\mathscr{B}$-countably m-rectifiable measures and $\mathscr{B}$-countably m-rectifiable sets as their support. The properties of these sets and measures were discussed and we subsequently gave a definition of entropy for $\mathscr{B}$-countably m-rectifiable probability measures. This definition generalizes entropy of discrete random variables, differential entropy, and also Rényi entropy, for discrete continuous mixtures.

We generalized $\mathscr{B}$-countably m-rectifiability to product spaces, which led to combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable sets and measures. Based on the well-known general definition of mutual information, those enabled us to study the mutual information of mixtures of $\mathscr{B}$-countably (n_1, n_2)-rectifiable probability measures, comprised of different dimensions. This discussion concluded with the proof of the relation $I(X;Y) = h(X) + h(Y) - h(X,Y)$, for our newly defined entropy, giving additional justification for the definitions we introduced.

We hope that the presented results will lead to future research in this topic. Possible applications, where singular distributions on "smooth" lower-dimensional subsets play an important role, are, e.g., the following areas:

- *Almost lossless and lossy (analog) signal compression:* Almost lossless compression is concerned with compression by a mapping from a higher-dimensional Euclidean space $\mathbb{R}^M$ to a lower-dimensional space $\mathbb{R}^m$, satisfying certain regularity conditions. Here singular, lower-dimensional probability distributions play an important role, and their information theoretic properties give rise to bounds on the efficiency of such compression schemes. Based on Rényi information dimension, recovery thresholds and optimal phase transition were derived in [WV10] and [WV12], respectively.

- *Vector interference channel:* The optimum input distribution for the vector interference channel is singular and concentrated in a subspace of the ambient signal space, as shown in [SB12].

- *Block-fading channels:* The number of degrees of freedom of the block-fading channel was treated in [KRDH13]. The received vector distribution was shown to be singular and again concentrated on a lower-dimensional subset of the ambient space.

The material presented in this thesis can provide for a finer information theoretic analysis than purely dimension or degrees-of-freedom based approaches, by taking entropy itself, and not merely its dimension into account.

A. Appendix

A.1. Proof of Lemma 2.6.8

Before proving Lemma 2.6.8, we present a result on the Jacobian of a Lipschitz function, which will be used in the proof. To this end we will need the following remarkable theorem concerning Lipschitz functions from [Fed69].

Theorem A.1.1. *[Fed69, Theorem 3.1.6]* (Rademacher's Theorem) *For $m, M \in \mathbb{N}$, if $f : \mathbb{R}^m \to \mathbb{R}^M$ is Lipschitz, then f is differentiable at λ^m-a.e. point of $\mathbb{R}^m$.*

Lemma A.1.2. *Let $f : \mathbb{R}^m \to \mathbb{R}^M$ be Lipschitz. Then there exists a constant $L < \infty$, such that for λ^m-a.e. $x \in \mathbb{R}^m$,*

$$J_m f(x) \leq L. \tag{A.1}$$

Proof. Theorem A.1.1 (Rademacher's Theorem) guarantees that f is differentiable at λ^m-a.e. point in $\mathbb{R}^m$. We conclude that $Df(x)$, and therefore also $J_m f(x) := ||\Lambda_m Df(x)||$ are well-defined for λ^m-a.e. $x \in \mathbb{R}^m$.

As given in [Fed69, 3.1.1], the differential of f at $x \in A$, applied to $v \in \mathbb{R}^m$, can be written as

$$Df(x)(v) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}. \tag{A.2}$$

Using the definition of convergence, we can write (A.2) as

$$\lim_{t \to 0} \left\| Df(x)(v) - \frac{f(x + tv) - f(x)}{t} \right\| = 0, \tag{A.3}$$

and applying the reverse triangle inequality $\big| ||a|| - ||b|| \big| \leq ||a - b||$, gives us

$$0 \overset{(A.3)}{=} \lim_{t \to 0} \left\| Df(x)(v) - \frac{f(x + tv) - f(x)}{t} \right\| \geq \lim_{t \to 0} \left| ||Df(x)(v)|| - \left\| \frac{f(x + tv) - f(x)}{t} \right\| \right|.$$

Since $||Df(x)(v)||$ does not depend on t, it follows that

$$\lim_{t \to 0} \left\| \frac{f(x + tv) - f(x)}{t} \right\| = ||Df(x)(v)||. \tag{A.4}$$

Using the Lipschitz property (Definition 2.6.4) of f with Lipschitz constant $\mathrm{Lip}(f) = \widetilde{L}$ gives us

$$||Df(x)(v)|| \overset{(A.4)}{=} \lim_{t \to 0} \left\| \frac{f(x + tv) - f(x)}{t} \right\| \overset{(2.12)}{\leq} \lim_{t \to 0} \frac{\widetilde{L}||tv||}{|t|} = \widetilde{L}||v||.$$

This results in a bound on the operator norm,

$$||Df(x)|| \leq \widetilde{L}. \tag{A.5}$$

Finally, [Fed69, 1.7.6] and (A.5) yield

$$J_m f(x) = ||\Lambda_m Df(x)|| \leq ||Df(x)(v)||^m \leq \widetilde{L}^m,$$

which ensures that $L := \widetilde{L}^m$ satisfies (A.1). □

Proof of Lemma 2.6.8. For $m = 0$, the statement follows from Part (i) of Theorem 2.6.3 as, by Part (i) of Definition 2.6.7, $|E| < \infty$. For $m > 0$, let $A \subseteq \mathbb{R}^m$ be bounded, $f: A \to \mathbb{R}^M$ be Lipschitz and $f(A) = E$. As E is a Borel set, so is $f^{-1}(E) = A$, because f is continuous and thus $(\mathscr{B}^m, \mathscr{B}^M)$-measurable. We extend f to $\widetilde{f}: \mathbb{R}^m \to \mathbb{R}^M$, using Theorem 3.4.1. Applying Part (i) of Theorem 3.4.2, we obtain

$$\int_A J_m \widetilde{f}(x)\, \mathrm{d}\lambda^m(x) = \int N(f, y)\, \mathrm{d}\mathscr{H}^m(y). \tag{A.6}$$

Because of $N(f, y) \geq 1$ for $y \in E$ and $N(f, y) = 0$ for $y \notin E$, we have by (A.6) that

$$\int_A J_m \widetilde{f}(x)\, \mathrm{d}\lambda^m(x) \overset{(a)}{\geq} \int \mathbb{1}_E(y)\, \mathrm{d}\mathscr{H}^m(y) = \mathscr{H}^m(E), \tag{A.7}$$

where (a) holds by Part (iii) of Theorem 2.2.10. Since A is bounded, $\lambda^m(A) < \infty$, which yields

$$\int_A J_m \widetilde{f}(x)\, \mathrm{d}\lambda^m(x) \overset{(a)}{\leq} \lambda^m(A) L < \infty, \tag{A.8}$$

where (a) holds for some $L \geq 0$ because of Lemma A.1.2, and Part (iii) of Theorem 2.2.10. Combining (A.7) and (A.8) yields $\mathscr{H}^m(E) < \infty$.

According to Part (ii) of Definition 2.6.7, a $\mathscr{B}$-countably m-rectifiable set F can be written as a countable union $F = \bigcup_i F_i$, where F_i is m-rectifiable and Borel for all $i \in \mathbb{N}$. For all $i \in \mathbb{N}$ we already showed $\mathscr{H}^m(F_i) < \infty$, which concludes the proof. □

A.2. Proof of Lemma 2.6.13

To prove Lemma 2.6.13, we will need the following weaker version of a result from [Fed69].

Theorem A.2.1. *[Fed69, Theorem 3.2.23] Let $m, n, M, N \in \mathbb{N}$, $W \subseteq \mathbb{R}^M$ be an m-rectifiable*

Borel set, and $Z \subseteq \mathbb{R}^N$ *an* n*-rectifiable Borel set. Then*[1]

$$\mathscr{H}^{m+n}|_{W\times Z} = \mathscr{H}^m|_W \times \mathscr{H}^n|_Z. \tag{A.9}$$

As Federer does not use 0-rectifiable sets in [Fed69], we have to prove an extension of Theorem A.2.1, which is given in the following lemma.

Lemma A.2.2. *Let* $N \in \mathbb{N}$, $n \in \{1,\dots,N\}$, $a \in \mathbb{R}^N$ *and* $E \subseteq \mathbb{R}^N$, *then*

$$\mathscr{H}^n(E) = \mathscr{H}^n(\{a\} \times E),$$

where $\mathscr{H}^n$ *is used to denote both, the* n*-dimensional Hausdorff measure on* $(\mathbb{R}^N, \mathscr{B}^N)$ *and on* $(\mathbb{R}^{2N}, \mathscr{B}^{2N})$.

Proof. Following Carathéodory's construction ([Fed69, 2.10.1]) of the Hausdorff measure ([Fed69, 2.10.2(1)]), one notes that a countable covering of E by nonempty sets $\{E_i\}_{i\in I}$ in $\mathbb{R}^N$ corresponds to the covering $\{\{a\} \times E_i\}_{i\in I}$ of $\{a\} \times E$ in $\mathbb{R}^{2N}$. As $\operatorname{diam}(E_i) = \operatorname{diam}(\{a\} \times E_i)$, it follows that

$$\mathscr{H}^n(\{a\} \times E) \le \mathscr{H}^n(E),$$

as this inequality holds for all size $\delta > 0$ approximating measures.

On the other hand assume $\{E_i\}_{i\in I}$ to be a countable covering of $\{a\} \times E$ in $\mathbb{R}^{2N}$ by nonempty sets. Let $\pi_2 \colon \mathbb{R}^{2N} \to \mathbb{R}^N$ denote the canonical projection onto the second N coordinates and for a set $A \in \mathbb{R}^{2N}$ we define $[A]_a := A \cap (\{a\} \times \mathbb{R}^N)$. Then the sets $\{\widetilde{E}_i := \pi_2([E_i]_a)\}_{i\in I:\widetilde{E}_i\neq\varnothing}$ give us a nonempty countable covering of E in $\mathbb{R}^N$. As clearly $\operatorname{diam}(\widetilde{E}_i) \le \operatorname{diam}(E_i)$,

$$\mathscr{H}^n(E) \le \mathscr{H}^n(\{a\} \times E)$$

also holds, as it is valid for all size $\delta > 0$ approximating measures. □

Proof of Lemma 2.6.13. Let $\{C_i\}_{i\in\mathbb{N}}$ and $\{D_i\}_{i\in\mathbb{N}}$ be partitions of C and D into countably many m_1-rectifiable/m_2-rectifiable Borel sets. By Remark 2.6.10 we can assume both, $\{C_i\}_{i\in\mathbb{N}}$ and $\{D_i\}_{i\in\mathbb{N}}$, to be mutually disjoint, i.e., $C_i \cap C_{i'} = \varnothing$ and $D_i \cap D_{i'} = \varnothing$ for $i \neq i'$.

Case $m_1 = m_2 = 0$. This is equivalent to $|C|, |D| \le \aleph_0$. The Cartesian product of two countable sets is again countable, therefore E is $\mathscr{B}$-countably 0-rectifiable and (2.14) holds true for the counting measure.

Case $m_1 = 0$, $m_2 > 0$. For every $j \in \mathbb{N}$, there exists a Lipschitz function $g_j \colon B_j \subseteq \mathbb{R}^{m_2} \to \mathbb{R}^M$ with B_j bounded, and $g_j(B_j) = D_j$. By Part (ii) of Definition 2.6.7, $|C| \le \aleph_0$. We can

[1]Note that (A.9) establishes the equality of two measures.

write $C = \bigcup_i \{c_i\}$. Then E is given as

$$E = \left(\bigcup_i \{c_i\}\right) \times D = \bigcup_{i,j} \big(\{c_i\} \times D_j\big).$$

The sets $E_{i,j} := \{c_i\} \times D_j \in \mathscr{B}^{2M}$ are m_2-rectifiable, because the function

$$\begin{aligned} g_{i,j} : D_j &\to \mathbb{R}^{2M} \\ x &\mapsto (c_i, g_j(x)) \end{aligned}$$

is Lipschitz. Furthermore, $g_{i,j}(D_j) = E_{i,j}$ and, thus, E is $\mathscr{B}$-countably m_2-rectifiable. As both, $\mathscr{H}^0|_C$ and $\mathscr{H}^{m_2}|_D$ are σ-finite by Lemma 2.6.8, their product, as defined in Definition 2.4.2, is unique. Therefore, to prove (2.14) we have to show that for any two sets $A, B \in \mathscr{B}^M$,

$$\mathscr{H}^{m_2}|_E(A \times B) = \mathscr{H}^0|_C(A)\mathscr{H}^{m_2}|_D(B).$$

Indeed,

$$\begin{aligned} \mathscr{H}^{m_2}|_E(A \times B) &\overset{(a)}{=} \sum_{i \in \mathbb{N}} \mathscr{H}^{m_2}([A \cap \{c_i\}] \times [B \cap D]) \\ &\overset{(b)}{=} \sum_{i \in \mathbb{N}} \mathbb{1}_A(c_i)\mathscr{H}^{m_2}|_D(B) \\ &= \mathscr{H}^0|_C(A)\mathscr{H}^{m_2}|_D(B), \end{aligned}$$

where (a) holds because of Part (ii) of Definition 2.1.3 and (b) follows from Lemma A.2.2.

Case $m_1 > 0$, $m_2 = 0$. Follows in analogy to the case $m_1 = 0$, $m_2 > 0$.

Case $m_1 > 0$, $m_2 > 0$. For all $i, j \in \mathbb{N}$, there exist Lipschitz functions $f_i : A_i \subseteq \mathbb{R}^{m_1} \to \mathbb{R}^M$ and $g_j : B_j \subseteq \mathbb{R}^{m_2} \to \mathbb{R}^M$ with $f_i(A_i) = C_i$ and $g_j(B_j) = D_j$, where A_i, B_j are bounded. As both $\mathscr{H}^{m_1}|_C$ and $\mathscr{H}^{m_2}|_D$ are σ-finite by Lemma 2.6.8, their product as defined in Definition 2.4.2 is unique, i.e., to prove (2.14), we have to show that $\mathscr{H}^{m_1+m_2}|_E(A \times B) = \mathscr{H}^{m_1}|_C(A)\mathscr{H}^{m_2}|_D(B)$ for any two Borel sets A and B.

We have

$$\begin{aligned} E &= C \times D \\ &= \bigcup_i C_i \times \bigcup_j D_j \\ &= \bigcup_{i,j} C_i \times D_j. \end{aligned} \tag{A.10}$$

As the union in (A.10) is disjoint, applying Theorem A.2.1 and the σ-additivity of measures (Part (ii) of Definition 2.1.3) yields

$$\mathcal{H}^{m_1+m_2}|_E = \sum_{i,j} \mathcal{H}^{m_1+m_2}|_{C_i \times D_j} = \sum_{i,j} \mathcal{H}^{m_1}|_{C_i} \times \mathcal{H}^{m_2}|_{D_j}. \tag{A.11}$$

Thus, for any two Borel sets A and B,

$$\begin{aligned}
\mathcal{H}^{m_1+m_2}|_E(A \times B) &\overset{(A.11)}{=} \sum_{i,j} \mathcal{H}^{m_1}|_{C_i}(A) \mathcal{H}^{m_2}|_{D_j}(B) = \\
&= \left(\sum_i \mathcal{H}^{m_1}|_{C_i}(A)\right)\left(\sum_j \mathcal{H}^{m_2}|_{D_j}(B)\right) = \\
&= \left(\sum_i \mathcal{H}^{m_1}(C_i \cap A)\right)\left(\sum_j \mathcal{H}^{m_2}(D_j \cap B)\right) = \\
&\overset{(a)}{=} \mathcal{H}^{m_1}|_C(A) \mathcal{H}^{m_2}|_D(B),
\end{aligned}$$

where (a) follows from the σ-additivity of measures, Part (ii) of Definition 2.1.3.

We define the function

$$\begin{aligned}
h_{i,j} : A_i \times B_j \subseteq \mathbb{R}^{m_1+m_2} &\to \mathbb{R}^{2M} \\
(x, y) &\mapsto (f_i(x), g_j(y))
\end{aligned}$$

for all $(i, j) \in \mathbb{N}^2$. As the sets $A_i \times B_j$ are bounded and $h_{i,j}$ is Lipschitz, with Lipschitz constant $\mathrm{Lip}(h_{i,j}) = \max(\mathrm{Lip}(f_i), \mathrm{Lip}(g_j))$, the Borel sets $h_{i,j}(A_i \times B_j) = C_i \times D_j$ are $(m_1 + m_2)$-rectifiable and their union E is therefore $\mathscr{B}$-countably $(m_1 + m_2)$-rectifiable.

□

A.3. On the Decomposition of Mixtures

In Section 4.2 we analyzed mixtures of rectifiable measures of different dimensions. For several results, it was crucial that the supports of these different measures be disjoint. The following two lemmas show that such supports can always be found.

Lemma A.3.1. *In the setup given in Assumptions (a1) to (a5), drop (4.3) ($E_m \cap E_{m'} = \varnothing$ for $m \neq m'$). Then it is still possible to find a new support $\widetilde{E}_m$ for each measure μ_m, such that,*

$$\widetilde{E}_m \cap \widetilde{E}_{m'} = \varnothing \quad \text{for } m \neq m' \tag{A.12}$$

holds.

Proof. Define

$$\widetilde{E}_m := E_m \setminus \bigcup_{m'<m} E_{m'}.$$

Then the sets $\widetilde{E}_m$ satisfy (A.12).

We need to check that $\widetilde{E}_m$ fulfills the requirements for a support in Definition 3.1.1. It is noted that $\widetilde{E}_m$ is Borel as a finite intersection of Borel sets and therefore $\mathscr{B}$-countably m-rectifiable according to Part (ii) of Corollary 2.6.9, which ensures Part (ii). Because for $m' < m$, $\mathscr{H}^m(E_{m'}) = 0$, as proved in Corollary 2.6.12, we have $\mathscr{H}^m(\bigcup_{m'<m} E_{m'}) = 0$ by Part (i) of Proposition 2.1.4. This yields $\mathscr{H}^m|_{E_m} = \mathscr{H}^m|_{\widetilde{E}_m}$, which establishes Part (iii). □

Lemma A.3.2. *In the setup given in Assumptions (b1) to (b8), drop (4.8) ($F_{n_1,n_2} \cap F_{m_1,m_2} = \varnothing$ for $(n_1,n_2) \neq (m_1,m_2)$). It is then still possible to choose first and second supports $\widetilde{C}_{n_1,n_2}$ and $\widetilde{D}_{n_1,n_2}$ of μ_{n_1,n_2} for each $(n_1,n_2) \in \{0,1,\ldots,N\}^2$, such that*

(i)

$$\widetilde{C}_{n_1} \cap \widetilde{C}_{m_1} = \varnothing \quad \text{for } n_1 \neq m_1,$$

(ii)

$$\widetilde{D}_{n_2} \cap \widetilde{D}_{m_2} = \varnothing \quad \text{for } n_2 \neq m_2, \text{ and}$$

(iii)

$$\widetilde{F}_{n_1,n_2} \cap \widetilde{F}_{m_1,m_2} = \varnothing \quad \text{for } (n_1,n_2) \neq (m_1,m_2).$$

Proof. Let

$$\widetilde{C}_{n_1,n_2} := C_{n_1,n_2} \setminus \bigcup_{n_1'<n_1} C_{n_1'} \tag{A.13}$$

and

$$\widetilde{D}_{n_1,n_2} := D_{n_1,n_2} \setminus \bigcup_{n_2'<n_2} D_{n_2'}.$$

To show Parts (i) and (ii) we proceed as follows. Let, without loss of generality, $n_1 > m_1$. We assume $x \in (\widetilde{C}_{m_1} \cap \widetilde{C}_{n_1})$, which will lead to a contradiction. According to (4.5), $x \in \widetilde{C}_{m_1,m_2}$ for some m_2 and hence, by (A.13), $x \in C_{m_1,m_2}$ and therefore, using again (4.5), $x \in C_{m_1}$. Because $m_1 < n_1$, according to (A.13), $x \notin \widetilde{C}_{n_1,n_2}$ for all $n_2 \in \{0,1,\ldots,N\}$, which establishes $x \notin \widetilde{C}_{n_1}$, contradicting our initial assumption. Part (ii) follows analogously. For showing Part (iii) let m_1, m_2, n_1, n_2 be such that $(n_1,n_2) \neq (m_1,m_2)$. Then either $n_1 \neq m_1$ or $n_2 \neq m_2$. Without loss of generality we assume $n_1 \neq m_1$. By Part (i), $\widetilde{C}_{n_1} \cap \widetilde{C}_{m_1} = \varnothing$, which establishes $\widetilde{F}_{n_1,n_2} \cap \widetilde{F}_{m_1,m_2} = \varnothing$ using (4.6).

It remains to show that the newly defined sets are in fact valid supports of their respective measures. The sets $\widetilde{C}_{n_1,n_2}$ and $\widetilde{D}_{n_1,n_2}$ are Borel as intersections of Borel sets, hence they are $\mathscr{B}$-countably n_1-rectifiable/$\mathscr{B}$-countably n_2-rectifiable, as proved in Part (ii) of Corollary 2.6.9. It remains to show that $\mu_{n_1,n_2} \ll \mathscr{H}^{n_1}|_{\widetilde{C}_{n_1,n_2}} \times \mathscr{H}^{n_2}|_{\widetilde{D}_{n_1,n_2}}$. We will achieve this by showing $\mathscr{H}^{n_1+n_2}|_{C_{n_1,n_2} \times D_{n_1,n_2}} = \mathscr{H}^{n_1}|_{\widetilde{C}_{n_1,n_2}} \times \mathscr{H}^{n_2}|_{\widetilde{D}_{n_1,n_2}}$. For the two sets $R := \bigcup_{n_1' < n_1} C_{n_1'}$ and $S := \bigcup_{n_2' < n_2} D_{n_2'}$, by Part (iii) of Theorem 2.6.3, $\mathscr{H}^{n_1}(R) = 0$ and $\mathscr{H}^{n_2}(S) = 0$. Together with Lemma 2.6.13 this results in

$$\begin{aligned}
\mathscr{H}^{n_1+n_2}|_{C_{n_1,n_2} \times D_{n_1,n_2}} &\overset{(2.14)}{=} \mathscr{H}^{n_1}|_{C_{n_1,n_2}} \times \mathscr{H}^{n_2}|_{D_{n_1,n_2}} \\
&= (\mathscr{H}^{n_1}|_{\widetilde{C}_{n_1,n_2}} + \mathscr{H}^{n_1}|_{R \cap C_{n_1,n_2}}) \times (\mathscr{H}^{n_2}|_{\widetilde{D}_{n_1,n_2}} + \mathscr{H}^{n_2}|_{S \cap D_{n_1,n_2}}) \\
&= \mathscr{H}^{n_1}|_{\widetilde{C}_{n_1,n_2}} \times \mathscr{H}^{n_2}|_{\widetilde{D}_{n_1,n_2}}.
\end{aligned}$$

□

Index

absolutely continuous, 8
A^c, 4
$\alpha(m)$, 13
Area Formula, 25

$\mathscr{B}(\Omega)$, *see* Borel σ-algebra
Borel σ-algebra, 4
Borel-Lebesgue measure, 5
$B_r(x)$, 4

Caratheodory's construction, 13
counting measure, 5

divergence, 38
$\frac{d\mu}{d\mathscr{H}^m|_E}$, 17

entropy, 24
  of a $\mathscr{B}$-countably m-rectifiable measure, 25
event, 12

$h_\nu(\mu)$, *see* entropy
$H_{\mu||\nu}(\mathscr{Q})$, *see* relative entropy
$H_2(p_{n_1,n_2})$, 39
$H_2(p_m)$, 43
Hausdorff measure, 13
Hausdorff outer measure, 13

$I(X;Y)$, *see* mutual information
induced measure, 6
integral, *see* Lebesgue integral

λ, *see* Lebesgue measure
Lebesgue integral, 7
Lebesgue measure, 5
Lipschitz constant, 14
Lipschitz function, 14

m-density, 20
  of a Borel set, 21
marginal, 32
$(\mathfrak{S}_1, \mathfrak{S}_2)$-measurable function, 6
measurable space, 3
measure, 5
  m-rectifiable, 17
  $\mathscr{B}$-countably m-rectifiable, 17
  combined $\mathscr{B}$-countably (n_1, n_2)-rectifiable, 31
measure space, 4
monotone continuity from below, 5
monotonicity, 5
μ-a.e., *see* μ-almost every or μ-almost everywhere
μ-almost every, 6
μ-almost everywhere, 6
multiplicity, 14
mutual information, 39
mutually singular, 9

$N(f,y)$, *see* multiplicity

outer measure, 3

$\mathfrak{P}(A)$, 4
probability measure, 12
probability space, 12
product σ-algebra, 11
product measurable space, 11

$\mathbb{R}^+$, 4
$\bar{\mathbb{R}}$, 4
$\bar{\mathbb{R}}^+$, 4
random variable, 12
relative entropy, 38

set
  m-rectifiable, 15
  $\mathscr{B}$-countably m-rectifiable, 15
σ-additivity, 4
σ-algebra, 3
σ-finite, 8
σ-subadditivity, 5
simple function, 6
support
  of an m-rectifiable measure, 17

$\mathscr{T}$, *see* topology
theorem
  Besicovitch differentiation, 21
  Fubini's, 11
  Kirszbraun's, 25
  Lebesgue Decomposition, 9
  Perez-Yaglom-Gelfand, 39
  Rademacher's, 53
  Radon-Nikodým, 9
$\Theta^m(E,x)$, *see* m-density, of a Borel set
$\Theta^m(\mu,x)$, *see* m-density
topology, 3

w.r.t., 1

ζ, *see* counting measure

Bibliography

[AFP00] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press on Demand, 2000.

[AL06] Krishna B. Athreya and Soumendra N. Lahiri. *Measure Theory and Probability Theory*. Springer Texts in Statistics. Springer, 2006.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, 2006.

[Fed69] Herbert Federer. *Geometric measure theory*. Grundlehren der mathematischen Wissenschaften. Springer, 1969.

[Gra13] Robert M. Gray. *Entropy and information theory*. Springer-Verlag, first and corrected edition, 2013.

[KP08] Steven G. Krantz and Harold R. Parks. *Geometric Integration Theory*. Cornerstones. Birkhäuser Boston, 2008.

[KRDH13] Günther Koliander, Erwin Riegler, Giuseppe Durisi, and Franz Hlawatsch. Generic correlation increases noncoherent mimo capacity. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 2084–2088, July 2013.

[Mun00] James R. Munkres. *Topology*. Prentice Hall, 2000.

[Ré59] Alfréd Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Hungarica*, 10:193–215, March 1959.

[SB12] David Stotz and Helmut Bölcskei. Degrees of freedom in vector interference channels. In *Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA*, pages 1755–1760, October 2012.

[WV10] Yihong Wu and Sergio Verdú. Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Transactions on Information Theory*, 56(8):3721–3748, August 2010.

[WV12] Yihong Wu and Sergio Verdú. Optimal phase transitions in compressed sensing. *IEEE Transactions on Information Theory*, 58(10):6241–6263, October 2012.