ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

Θεόδωρος Τσατσαρώνης

Area & Coarea Formula

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννινα, 2024

UNIVERSITY OF IOANNINA
Department of Mathematics

Theodoros Tsatsaronis

Area & Coarea Formula

Master's Thesis

Ioannina, 2024

Αφιερώνεται στους γονείς μου, Γεώργιο και Παναγιώτα.

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Ανάλυση (Ειδίκευση Α΄: Μαθηματικά (Ανάλυση - Άλγεβρα - Γεωμετρία)), που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 02/07/2024 από την εξεταστική επιτροπή:

Ονοματεπώνυμο	**Βαθμίδα**
Ελευθέριος Νικολιδάκης	Επίκουρος Καθηγητής
Ιωάννης Πουρναράς	Καθηγητής
Ανδρέας Τόλιας	Αναπληρωτής Καθηγητής

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

"Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή."

Θεόδωρος Τσατσαρώνης

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της Μεταπτυχιακής μου Διατριβής, θα ήθελα να απευθύνω ένα θερμότατο ευχαριστώ, εκ μέσης καρδίας, προς όλους όσους στάθηκαν δίπλα μου σε αυτό το ταξίδι και με βοήθησαν να το φέρω εις πέρας. Εξαιρέτως, θα ήθελα να ευχαριστήσω τον επιβλέποντά μου, κ. Ελευθέριο Νικολιδάκη, Επίκουρο Καθηγητή του Πανεπιστημίου Ιωαννίνων, παροτρύνσει του οποίου ασχολήθηκα με αυτό το τόσο ενδιαφέρον θέμα. Με τις παρατηρήσεις του και τις συμβουλές του, έπαιξε καθοριστικό ρόλο στη διαμόρφωση αυτού του αποτελέσματος.

Επίσης, θα ήθελα να ευχαριστήσω τους κ.κ. Ιωάννη Πουρναρά και Ανδρέα Τόλια, των οποίων είχα την τύχη να υπάρξω μαθητής, τόσο σε Προπτυχιακό όσο και σε Μεταπτυχιακό επίπεδο, και οι οποίοι πλαισιώνουν την Εξεταστική Επιτροπή, για τον χρόνο που αφιέρωσαν και για την επιμέλεια με την οποία προσέγγισαν την εργασία αυτή.

Τέλος, το μεγαλύτερο ευχαριστώ το απευθύνω στους δικούς μου ανθρώπους, στην οικογένειά μου και στους φίλους μου. Η ανιδιοτελής αγάπη τους, η αμέριστη συμπαράστασή τους και η αδιάκοπη στήριξή τους κατά τη διάρκεια του Μεταπτυχιακού μου ήταν η κυριότερη πηγή δύναμης και έμπνευσης για εμένα, και δεν θα διστάσω να το πω, ότι θα ήταν αδύνατο να τα καταφέρω χωρίς αυτούς!

Περιληψη

Σκοπός της παρούσης μεταπτυχιακής διατριβής είναι η αυστηρή θεμελίωση στο πλαίσιο της Μαθηματικής Ανάλυσης δύο σημαντικών τύπων της Γεωμετρικής Θεωρίας Μέτρου, γνωστών ως Τύποι Area και Coarea.

Η δομή που θα ακολουθήσουμε είναι η εξής: Τα δύο πρώτα κεφάλαια της εργασίας μας είναι εισαγωγικά. Σε αυτά αναπτύσσουμε την απαραίτητη θεωρία που πρέπει να γνωρίζει ο αναγνώστης, προκειμένου να κατανοήσει το περιεχόμενο της προκειμένης εργασίας. Αναλυτικότερα, στο Πρώτο Κεφάλαιο παρουσιάζονται οι βασικές έννοιες της Θεωρίας Μέτρου και θεμελιώνονται τα εργαλεία πάνω στα οποία θα αναπτύξουμε την θεωρία μας. Στο Δεύτερο Κεφάλαιο, ορίζεται το Μέτρο Hausdorff, το οποίο πρωταγωνιστεί στους προαναφερθέντες τύπους και αποδεικνύονται αναλυτικά οι ιδιότητές του. Ακολούθως, παρουσιάζουμε την συμμετρικοποίηση Steiner, την οποία και αξιοποιούμε για να καταδείξουμε την λεγομένη Ισοδιαμετρική Ανισότητα, καταλήγοντας σε ένα εξαιρετικής σημασίας αποτέλεσμα· την ταύτιση του μέτρου Lebesgue με το n-διάστατο μέτρο Hausdorff.

Ακολούθως, στο Τρίτο Κεφάλαιο ορίζουμε την έννοια της απεικόνισης Lipschitz και πότε αυτή θα καλείται διαφορίσιμη και αποδεικνύουμε το Θεώρημα του Rademacher, το οποίο μας εξασφαλίζει την σχεδόν παντού διαφορισιμότητα μιας τέτοιας απεικόνισης. Το κεφάλαιο επισφραγίζεται με την παρουσίαση ορισμένων ιδιοτήτων των Γραμμικών Απεικονίσεων του $\mathbb{R}^n$, και με τη βοήθεια του Θεωρήματος της Πολικής Αναπαράστασης, καταλήγουμε σε μια κατάλληλη έννοια για την Ιακωβιανή μιας Lipschitz απεικόνισης.

Μετά από αυτή τη διαδρομή, μπορούμε να προχωρήσουμε στην απόδειξη του Τύπου Area, η οποία αποτελεί και τη θεματολογία του Τετάρτου Κεφαλαίου. Μελετούμε απεικονίσεις Lipschitz $f : \mathbb{R}^n \to \mathbb{R}^m$ για $n \leq m$ και εξάγουμε κάποιους χαρακτηριστικούς τύπους για το Ολοκλήρωμα της Ιακωβιανής τους. Αρχικώς αποδεικνύονται τα προπαρασκευαστικά Λήμματα και στη συνέχεια το κεντρικό θεώρημα. Το κεφάλαιο ολοκληρώνεται με την παράθεση ορισμένων χαρακτηριστικών εφαρμογών.

Το Πέμπτο Κεφάλαιο ασχολείται με την "δυϊκή" μορφή του προβλήματος,

δηλαδή την μελέτη απεικονίσεων Lipschitz $f : \mathbb{R}^n \to \mathbb{R}^m$ για $n \geq m$, αυτή τη φορά. Η δομή του προκειμένου Κεφαλαίου μιμείται το προηγηθέν Τέταρτο Κεφάλαιο: Παρουσιάζουμε αρχικά τα Λήμματα που μας οδηγούν στην απόδειξη του Τύπου Coarea και στη συνέχεια διατυπώνουμε και αποδεικνύουμε το θεώρημα. Τέλος,παρουσιάζουμε κάποιες χαρακτηριστικές εφαρμογές. Η διατριβή ολοκληρώνεται με την παράθεση κάποιων αποτελεσμάτων πέραν από την Γ.Θ.Μ., τα οποία στηρίζονται στους τύπους Area και Coarea και φανερώνουν την σημαντικότητα αυτών των εργαλείων σε κάθε πτυχή των Μαθηματικών.

Abstract

The aim of the present Master's Thesis is to establish rigorously, within the framework of Mathematical Analysis, two mathematical Formulas, known as Area and Coarea Formula. The structure of the Thesis is the following; The first two chapters are introductory. In them we offer a thorough overview of all the concepts the reader needs to be familiar with, in order to better understand the content of our work.

In particular, in the Chapter 1 we deal with elements of Measure Theory and we lay the groundwork for the tools on which our work will be based on. In Chapter 2, we define the Hausdorff Measure, which will play a leading part in the aforementioned formulas, and we prove its properties. We then introduce the Steiner Symmetrization, which we use in order to prove the so-called Isodiametric Inequality, reaching to a result of high importance; The identification of the Lebesgue Measure with the n-dimensional Hausdorff Measure on Rn.

Afterwards, in Chapter 3, we define the notion of a Lipschitz map and determine when that map is differentiable and in which sense and we prove Rademacher's Theorem, which ensures us that such a map is almost-everywhere differentiable. We end this Chapter by stating some properties of Linear maps of $\mathbb{R}^n$, and via the Polar Decomposition Theorem, we conclude with an appropriate notion for the Jacobian of a Lipschitz map.

After all of this journey, we are able to proceed in the proof of the Area Formula, which is the subject of Chapter 4. We study Lipschitz mappings of the form $f : \mathbb{R}^n \to \mathbb{R}^m$ for $n \leq m$ and we derive some special formulas regarding the Integral of their Jacobian. We begin by proving the preparatory Lemmas and then the main theorem. The Chapter is concluded with some characteristic applications.

Chapter Five deals with the "dual" form of the problem, i.e. the study of Lipschitz mappings $f : \mathbb{R}^n \to \mathbb{R}^m$ for$n \geq m$ this time. Its structure mimics the preceding Ch. 4; Firstly, we present in great detail the Lemmas which

guide us towards the proof of the Coarea Formula, and then we state and prove the Theorem. Finally, we present some typical applications. The thesis is culminated by presenting some extra results, beyond the G.T.M., which are based on the Area and Coarea formulas and highlight the importance of these tools, across every aspect of Mathematics.

Contents

CHAPTER 1

Protheoria I: Elements of Measure Theory

In this Chapter, we offer a basic overview of standard measure theory. We start by referencing some definitions of abstract measure and integration theory, reaching up to product measure and Fubini's theorem. We then quickly shift our focus on Radon measures. We establish the Differentiation Theorem for Radon measures and we state three important theorems: Lebesgue Differentiation theorem, Lebesgue Density theorem and an "Exhaustion" theorem of open sets with balls.

The content of the present Thesis is primarily influenced by the book of Lawrence C. Evans and Ronald F. Gariepy (see [8] and [7]). Our journey through Measure Theory follows the approach of H. Federer [10], in parallel with [24] and other bibliographic sources; [9, 15, 5].

1.1 Measures & measurable sets

Let X denote a non-empty set and 2^{X} the collection of all subsets of X.

Definition 1.1. *A mapping $\mu : 2^X \to [0, \infty]$ is called a* ***measure*** *on X provided that*

1. $\mu(\varnothing) = 0$ *and*
2. *if* $A \subseteq \bigcup_{k=1}^{\infty} A_k$ *then*

$$\mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k).$$

[1] Protheoria: The greek word "Protheoria" (Προθεωρία) referred to the introductory part of Medieval & Byzantine music codices, which used to include the key concepts & ideas, as well as some explanatory notes, for what was presented in the following sheets. Since we mimic the same pattern, we assumed its use in this Thesis.

REMARK. We are highly aware that the vast majority of mathematical texts would call such a mapping an **outer measure**, reserving the name `measure` for μ restricted to the collection of μ-measurable subsets of X (see the definition below).

However, we will adhere to this definition, due to the advantages we get by being able to "measure" even the *non-measurable* sets.

Definition 1.2. *A set $A \subseteq X$ is called **μ-measurable** if for each $B \subseteq X$ we have*

$$\mu(B) = \mu(B \cap A) + \mu\big(B \setminus A\big) .$$

Theorem 1.1 (**Elementary properties of measure**)**.** *Let μ be a measure on X.*

1. *If $A \subseteq B \subseteq X$, then $\mu(A) \leq \mu(B)$.*
2. *A set A is μ-measurable if and only if $X \setminus A$ is μ-measurable.*
3. *The sets $\varnothing$ and X are μ-measurable. More generally, if $\mu(A) = 0$, then A is μ-measurable.*
4. *For any $C \subseteq X$; Each μ-measurable set is also $\mu \llcorner C$-measurable, where by $\mu \llcorner C$ we denote the following*

$$(\mu \llcorner C)(A) = \mu(A \cap C) .$$

Theorem 1.2 (**Sequences of measurable sets**)**.** *Let $\{A_k\}_{k=1}^{\infty}$ be a sequence of μ-measurable sets.*

1. *The sets $\displaystyle\bigcup_{k=1}^{\infty} A_k$ and $\displaystyle\bigcap_{k=1}^{\infty} A_k$ are μ-measurable.*
2. *If the sets $\{A_k\}_{k=1}^{\infty}$ are **disjoint**, then*

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k) .$$

3. *If $A_1 \subseteq ... \subseteq A_k \subseteq A_{k+1} \subseteq ...$, then*

$$\lim_{k\to\infty} \mu(A_k) = \mu\left(\bigcup_{k=1}^{\infty} A_k\right) .$$

4. *If* $A_1 \supseteq ... \supseteq A_k \supseteq A_{k+1} \supseteq ...$ *with* $\mu(A_1) < \infty$, *then*

$$\lim_{k\to\infty} \mu(A_k) = \mu\left(\bigcap_{k=1}^{\infty} A_k\right).$$

Definition 1.3. *Let X be a non-empty set and $\mathcal{A}$ a collection of subsets of X. We say that $\mathcal{A}$ is a* ***σ-algebra*** *of X, provided that*

1. $\varnothing, X \in \mathcal{A}$,
2. $A \in \mathcal{A} \Rightarrow X \setminus A \in \mathcal{A}$,
3. $A_k \in \mathcal{A}$ $(k = 1, 2, ...) \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.

Theorem 1.3 (**Measurable sets constitute a σ-algebra**). *If μ is a measure on a non-empty set X, then the collection of all μ-measurable subsets of X is a σ-algebra.*

Definition 1.4.
If $\mathcal{C} \subseteq 2^X$ is any collection of subsets from X, the ***σ-algebra generated by*** *$\mathcal{C}$, denoted as $\sigma(\mathcal{C})$, is the smallest σ-algebra containing $\mathcal{C}$.*

Definition 1.5.

1. *The smallest σ-algebra containing the open sets of $\mathbb{R}^n$ is called* ***Borel σ-algebra****.*
2. *Its elements are called* ***Borel-measurable*** *sets.*
3. *We call μ a* ***Borel measure*** *if every Borel set is μ-measurable.*

Definition 1.6.

1. *A measure μ on X is* ***regular****, if for every set $A \subseteq X$ there exists a μ-measurable set B such that $A \subseteq B$ and $\mu(A) = \mu(B)$.*
2. *A measure μ on $\mathbb{R}^n$ is* ***Borel-regular****, if μ is Borel and for each set $A \subseteq \mathbb{R}^n$ there exists a Borel-measurable set B such that $A \subseteq B$ and $\mu(A) = \mu(B)$.*
3. *A measure μ on $\mathbb{R}^n$ is* ***Radon measure****, if μ is Borel regular and $\mu(K) < \infty$ for each compact set $K \subseteq \mathbb{R}^n$.*

Theorem 1.4.
Let μ be a regular measure on X. If $A_1 \subseteq \ldots \subseteq A_k \subseteq A_{k+1} \subseteq \ldots$, then

$$\lim_{k\to\infty} \mu(A_k) = \mu\left(\bigcup_{k=1}^{\infty} A_k\right).$$

REMARK. In contrast with the previous result, here, the sets $\{A_k\}_{k=1}^{\infty}$ need not be μ-measurable.

Theorem 1.5. *Let μ be a Borel measure on $\mathbb{R}^n$ and B a Borel set.*

1. *If $\mu(B) < \infty$, there exists, for each $\varepsilon > 0$, a closed set C such that*

$$C \subseteq B, \;\; \mu(B \setminus C) < \varepsilon.$$

2. *If μ is a Radon measure, there exists, for each $\varepsilon > 0$, an open set U such that*

$$B \subseteq U, \;\; \mu(U \setminus B) < \varepsilon.$$

Theorem 1.6 (Approximation by open and by compact sets).
Let μ be a Radon measure on $\mathbb{R}^n$. Then;

1. *For each set $A \subseteq \mathbb{R}^n$,*

$$\mu(A) = \inf\{\mu(U) \mid A \subseteq U, \; U \textit{open}\}.$$

2. *For each μ-measurable set $A \subseteq \mathbb{R}^n$,*

$$\mu(A) = \sup\{\mu(K) \mid K \subseteq A, \; K \textit{compact}\}.$$

The following criterion is a useful way to verify whether a measure μ is Borel.

Theorem 1.7 (Carathéodory's criterion). *Let μ be a measure on $\mathbb{R}^n$. If for all sets $A, B \subseteq \mathbb{R}^n$, we have*

$$\mu(A \cup B) = \mu(A) + \mu(B) \;\; \textit{whenever} \; \operatorname{dist}(A, B) > 0,$$

then μ is a Borel measure.

Proof. First, for clarification reasons, we shall state the specific notion of "set-theoretic" distance that we will use;
We denote

$$\operatorname{dist}(A,B) \coloneqq \inf\big\{d(\alpha,b) \mid \alpha \in A \text{ and } b \in B\big\}$$

for any metric d on $\mathbb{R}^n$.
Now, let $A, C \subseteq \mathbb{R}^n$ with C : closed. It suffices to show that

$$\mu(A) \geq \mu(A\cap C) + \mu(A \setminus C), \tag{$\star$}$$

since from sub-additivity we get that

$$\mu(A) = \mu\big((A\cap C) \cup (A \cup (\mathbb{R}^n \setminus C))\big\} \leq \mu(A\cap C) + \mu(A \cap (\mathbb{R}^n \setminus C)).$$

Observe that, if $\mu(A) = \infty$, then ($\star$) is obvious. Therefore, we continue assuming that $\mu(A) < \infty$. For $n = 1, 2, ...,$ we define sets

$$C_n \coloneqq \left\{ x \in \mathbb{R}^n \;\middle|\; \operatorname{dist}(x, C) \leq \frac{1}{n} \right\}.$$

Then, $\operatorname{dist}(A \setminus C_n, A \cap C) \geq \frac{1}{n}$, since for all $\alpha \in A \setminus C_n$ we have that $\operatorname{dist}(\alpha, C) > \frac{1}{n}$.
Therefore, our hypothesis implies that

$$\begin{aligned}\mu(A \setminus C_n) + \mu(A\cap C) &= \mu\big((A\setminus C_n) \cup (A\cap C)\big) \\ &\leq \mu\big(A \cup (A\cap C)\big) \leq \mu(A). \end{aligned}\tag{$\star\star$}$$

`Claim:`

$$\lim_{n\to\infty} \mu(A \setminus C_n) = \mu(A\setminus C).$$

Proof of claim: For $k = 1, 2, ...,$ take

$$R_k \coloneqq \left\{ x \in A \;\middle|\; \frac{1}{k+1} < \operatorname{dist}(x, C) \leq \frac{1}{k} \right\}.$$

Note that; if $z \in \displaystyle\bigcup_{k=n}^{\infty} R_k$, then $z \in R_{k_o}$ for some $k_o \geq n$.
Therefore, $0 < \dfrac{1}{k_o+1} < \operatorname{dist}(z, C) \leq \dfrac{1}{k_o}$ and thus $z \notin C$. Consequentially,

$$\big(A \setminus C_n\big) \cup \bigcup_{k=n}^{\infty} R_k = A \setminus C.$$

Hence

$$\mu(A \setminus C_n) \le \mu(A \setminus C) = \mu\left((A \setminus C_n) \cup \bigcup_{k=n}^{\infty} R_k \right)$$
$$\le \mu(A \setminus C_n) + \sum_{k=n}^{\infty} \mu(R_k).$$

It suffices now to show that the countable sum $\sum_{k=1}^{\infty} \mu(R_k) < \infty$, and thus, the "tail" will converge to zero as $n \to \infty$, establishing the claim.

For $j \ge i + 2$, we have that $R_i \cap R_j = \varnothing$ and

$$\operatorname{dist}(R_i, R_j) = \frac{1}{i+1} - \frac{1}{j} = \frac{j-i-1}{j(i+1)} \ge \frac{1}{j(i+1)} > 0.$$

Summing on the indices, via our hypothesis, we get that

$$\sum_{k=1}^{m} \mu(R_{2k}) = \mu\left(\bigcup_{k=1}^{m} R_{2k} \right) \le \mu(A),$$

and, for the odd indices,

$$\sum_{k=0}^{m} \mu(R_{2k+1}) = \mu\left(\bigcup_{k=0}^{m} R_{2k+1} \right) \le \mu(A).$$

Now, we bring these results together and allow $m \to \infty$. Consequentially,

$$\sum_{k=1}^{\infty} \mu(R_k) \le 2\mu(A) < \infty.$$

This concludes our claim.

Combining the Claim and $(\star\star)$ gives us

$$\mu(A \setminus C) + \mu(A \cap C) = \lim_{n \to \infty} \mu(A \setminus C_n) + \mu(A \cap C) \overset{(\star\star)}{\le} \mu(A).$$

This proves $(\star)$. Hence, the closed set C is μ-measurable, and consequentially, all Borel sets are μ-measurable. □

REMARK. Let it be noted that the converse also holds true;
If μ is a Borel measure, then μ splits additively on positively separated sets.
Indeed, let $A, B \subseteq \mathbb{R}^n$ with $\operatorname{dist}(A, B) > 0$. Observe that

$$A = \big(A \cup B\big) \cap \overline{A} \quad \text{and} \quad B = \big(A \cup B\big) \setminus \overline{A}.$$

Therefore, we get that

$$\mu(A) + \mu(B) = \mu\big(\big(A \cup B\big) \cap \overline{A}\,\big) + \mu\big(\big(A \cup B\big) \setminus \overline{A}\,\big). \tag{$\star$}$$

Now, since $\overline{A}$ is Borel measurable, applying the definition on ($\star$), we get that

$$\mu(A) + \mu(B) = \mu(A \cup B).$$

Hence, μ is additive on A, B.

Notation. Henceforward, we will denote with $|\bullet|$ the Euclidean norm (the 2-norm) of $\mathbb{R}^n$. Circumstantially, when there is need for clarification on the dimension, we will turn to the "customary" notation of $\|\bullet\|_d$, where d will denote the dimension of the argument of the norm.

1.2 Measurable functions

We now extend the notion of measurability from sets to functions.
Let μ be a measure on a non-empty set X, and, let Y be a topological space

Definition 1.7.

1. *A function $f : X \to Y$ is called **$\boldsymbol{\mu}$-measurable** if for each open set $U \subseteq Y$, the set*
$$f^{-1}(U)$$
is μ-measurable.

2. *A function $f : \mathbb{R}^n \to Y$ is called **Borel-measurable** if for each open set $U \subseteq Y$, the set*
$$f^{-1}(U)$$
is Borel-measurable.

Theorem 1.8.

1. *If $f : X \to Y$ is μ-measurable, then $f^{-1}(B)$ is μ-measurable for each Borel set $B \subseteq Y$.*

2. *If $f : \mathbb{R}^n \to Y$ is continuous, then f is Borel-measurable.*

Definition 1.8 (**Measurability of functions on the extended real number line**)**.** *A function $f : X \to [-\infty, \infty]$ is μ-measurable if and only if*

$$f^{-1}([-\infty, \alpha))$$

is μ-measurable for each $\alpha \in \mathbb{R}$.

Theorem 1.9 (**Algebra of $\boldsymbol{\mu}$-measurable functions**)**.**

1. *If $f, g : X \to [-\infty, \infty]$ are μ-measurable functions, then so are*

$$f \pm g,$$

provided that $\mu(\{f = \pm\infty\}) = 0 = \mu(\{g = \pm\infty\})$, or (alternatively) that $f \pm g$ is assigned with a specific real value, whenever the "pathological" cases of $\infty - \infty$ and $-\infty + \infty$ occur.

2. *If $f, g : X \to [-\infty, \infty]$ are μ-measurable functions, then the functions*

$$fg,\ |f|,\ \min(f, g),\ \max(f, g)$$

are also μ-measurable.
The function $\dfrac{f}{g}$ is also μ-measurable, provided that $g \neq 0$ on X.

3. *If the functions $f_k : X \to [-\infty, \infty]$ are μ-measurable (k=1,2,...) then*

$$\inf_{k\geq 1} f_k,\ \sup_{k\geq 1} f_k,\ \liminf_{k\to\infty} f_k \text{ and } \limsup_{k\to\infty} f_k$$

are also μ-measurable.

REMARK. It is customary in Measure Theory to take $0 \cdot (\pm\infty) = 0$. However, an appropriate definition for $\infty \pm \infty$ is problematic, hence we imposed those extra conditions in (*1.*).

Theorem 1.10. *Assume $f : X \to [0, +\infty]$ is μ-measurable. There exists an (at-most) countable family of μ-measurable sets $\{A_k\}_{k=1}^{\infty}$ in X such that*

$$f = \sum_{k=1}^{\infty} \frac{1}{k} \chi_{A_k}$$

REMARK. Note that; The sets $\{A_k\}_{k=1}^{\infty}$ in the preceding Theorem need not be disjoint. Also, note that the assertion is valid, even if the image of f is not a countable set.

Proof. We shall use the so-called "strong" induction.

First, we define

$$A_1 \coloneqq \{x \in X \mid f(x) \geq 1\}$$

and inductively, for $k = 2, 3, ...$

$$A_k \coloneqq \left\{ x \in X \;\middle|\; f(x) \geq \frac{1}{k} + \sum_{j=1}^{k-1} \frac{1}{j} \chi_{A_j}(x) \right\}.$$

We will show that; For all $m = 1, 2, ...$, we have the estimate

$$f \geq \sum_{j=1}^{m} \frac{1}{j} \chi_{A_j}.$$

Assume that the hypothesis holds for all $m \leq k$. Then $f \geq \sum_{j=1}^{k} \frac{1}{j} \chi_{A_j}$.

For the $k + 1$ index, we have that; If $x \notin A_{k+1}$, then

$$\sum_{j=1}^{k+1} \frac{1}{j} \chi_{A_j}(x) = \frac{1}{k+1} \chi_{A_{k+1}}(x) + \sum_{j=1}^{k} \frac{1}{j} \chi_{A_j}(x) = \sum_{j=1}^{k} \frac{1}{j} \chi_{A_j}(x) \leq f(x)$$

and for $x \in A_{k+1}$, we get $f(x) \geq \frac{1}{k+1} + \sum_{j=1}^{k} \frac{1}{j} \chi_{A_j}(x)$

$$= \frac{1}{k+1} . \chi_{A_{k+1}}(x) + \sum_{j=1}^{k} \frac{1}{j} \chi_{A_j}(x) = \sum_{j=1}^{k+1} \frac{1}{j} \chi_{A_j}(x).$$

Hence, for every case we have that $f(x) \geq \sum_{j=1}^{k+1} \frac{1}{k} \chi_{A_j}(x)$ for all $x \in X$.

Therefore, we can take the limit as $k \to \infty$, and conclude that

$$f \geq \sum_{k=1}^{\infty} \frac{1}{k} \chi_{A_k}$$

Now, it is clear that, if $f(x) = \infty$, then $x \in A_k$ for all k. Otherwise, if $0 \leq f(x) < \infty$, then the finiteness of the countable summation above implies that $x \notin A_n$ for infinitely many n. Hence

$$0 \leq f(x) \leq \frac{1}{n} + \sum_{k=1}^{n-1} \frac{1}{k} \chi_{A_k}(x)$$

for all such n. Taking the limit as $n \to \infty$ completes the proof. □

1.3 Integrals & Limit theorems

We now present some basic concepts in Integration Theory with respect to a measure.

For this section, we abide by the following **Notation**;

$$f^+ = \max(f, 0), \quad f^- = \max(-f, 0), \quad f = f^+ - f^-.$$

Let μ be a measure on a non-empty set X.

Definition 1.9. *A function $g : X \to [-\infty, \infty]$ is called a **simple function** if the image of g is countable.*

REMARK. Doing this, we allow for more functions to be taken into account.

Definition 1.10.

1. *If g is a non-negative and simple μ-measurable function, we define its **integral** to be*
$$\int g \, d\mu \coloneqq \sum_{0 \leq y \leq \infty} y \mu\big(g^{-1}\{y\}\big)\,.$$
2. *If g is a simple μ-measurable function for which either $\int g^+ d\mu < \infty$ or $\int g^- d\mu < \infty$, we call g a **μ-integrable simple function** and define its integral to be*
$$\int g \, d\mu \coloneqq \int g^+ \, d\mu - \int g^- \, d\mu.$$
It is clear that we allow the integral $\int g \, d\mu$ to take values $\pm\infty$.

Therefore, combining the two definitions, if g is a **μ-integrable simple function**, we get that

$$\int g \, d\mu := \sum_{-\infty \leq y \leq \infty} y\mu\big(g^{-1}\{y\}\big) \, .$$

To verify that, simply observe that we can decompose the inverse image of g into the union of two disjoint sets; The set of all arguments which give a non-negative value and the set of those arguments which yield a striktly negative argument. Thus, we only need to calculate that

$$\begin{aligned}
\int g \, d\mu &= \int g^{+} \, d\mu - \int g^{-} \, d\mu \\
&= \sum_{0 \leq y \leq \infty} y\mu\big((g^{+})^{-1}\{y\}\big) - \sum_{0 \leq y \leq \infty} y\mu\big((g^{-})^{-1}\{y\}\big) \\
&= \sum_{0 \leq y \leq \infty} y\mu\big((g^{+})^{-1}\{y\}\big) - \sum_{0 \leq -\mathcal{Y} \leq \infty} (-\mathcal{Y})\mu\big((g^{-})^{-1}\{-\mathcal{Y}\}\big) \\
&= \sum_{0 \leq y \leq \infty} y\mu\big((g^{+})^{-1}\{y\}\big) + \sum_{-\infty \leq \mathcal{Y} \leq 0} \mathcal{Y}\mu\big((g^{-})^{-1}\{-\mathcal{Y}\}\big) \\
&= \sum_{-\infty \leq y \leq \infty} y\Bigg(\mu\big((g^{+})^{-1}\{y\}\big) + \mu\big((g^{-})^{-1}\{-y\}\big) \Bigg) \\
&= \sum_{-\infty \leq y \leq \infty} y\mu\big((g^{+} - g^{-})^{-1}\{y\}\big) \\
&= \sum_{-\infty \leq y \leq \infty} y\mu\big(g^{-1}\{y\}\big) \, .
\end{aligned}$$

Notation. The expression

$$\mu - a.e.$$

is an abbreviation of the phrase **almost everywhere with respect to measure μ**, meaning that the aforementioned assertion is valid for all elements of the space X **except possibly** from a set A with $\mu(A) = 0$.

Of course this set could be the $\varnothing$, but that just means that the assertion is valid for the whole space X.

Definition 1.11.

1. *Let $f : X \to [-\infty, \infty]$. We define the* ***upper integral***

$$\int^{\star} f \, d\mu := \inf\left\{ \int g \, d\mu \;\middle|\; g\text{: } \mu\text{-integrable simple } \; g \geq f \;\, \mu - a.e. \right\}$$

*and the **lower integral***

$$\int_{\star} f \, d\mu := \sup\left\{ \int g \, d\mu \;\middle|\; g\text{: } \mu\text{-integrable simple } \; g \le f \;\; \mu - a.e. \right\}.$$

2. *A μ-measurable function $f : X \to [-\infty, \infty]$ is called **μ-integrable** if $\displaystyle\int^{\star} f \, d\mu = \int_{\star} f \, d\mu$ and, therefore, we write*

$$\int f \, d\mu := \int^{\star} f \, d\mu = \int_{\star} f \, d\mu.$$

REMARK. We shall specify that the term **integrable** differs from most texts. For our purposes, a function is `integrable` whenever **"it has an integral"**, even if this integral equals $+\infty$ or $-\infty$.

REMARK. It is immediate that a non-negative μ-measurable function is always μ-integrable.

First, assume that $\mu\big(\{f = \infty\}\big) = \mu\big(\{x \in X \mid f(x) = \infty\}\big) > 0$. Then, for any $t > 0$, we employ the simple function $\phi = t\,\chi_{\{f=\infty\}}$ and the definition of $\int_{\star} f \, d\mu$, in order to obtain

$$\int_{\star} f \, d\mu \ge \int t\,\chi_{\{f=\infty\}} = t\,\mu\big(\{f = \infty\}\big), \quad \text{for any } t > 0.$$

Thus $\displaystyle\int_{\star} f \, d\mu = \infty$ and since $\int^{\star} f \, d\mu \ge \int_{\star} f \, d\mu$, we also get that $\displaystyle\int^{\star} f \, d\mu = \infty$. Hence f is μ-integrable, with $\displaystyle\int f \, d\mu = \infty$.

Now, suppose that $\mu\big(\{f = \infty\}\big) = 0$. Then $f(x) < \infty$ for μ-a.e. $x \in X$. Let $t > 1$. We define

$$E_k := \big\{x \in X \mid t^k \le f(x) < t^{k+1}\big\}, \; k \in \mathbb{Z}.$$

Notice that the sets $\{E_k\}_{k \in \mathbb{Z}}$ are disjoint and μ-measurable. Furthermore, we define the simple function

$$g := \sum_{k \in \mathbb{Z}} t^k \, \chi_{E_k}.$$

Then $X \setminus \{f = 0\} = \displaystyle\bigcup_{k \in \mathbb{Z}} E_k$ and assigning the value 0 to g on $\{f = 0\}$, we have that $g(x) \le f(x) \le t g(x)$, μ-a.e. $x \in X$. We get that

$$\int^{\star} f \, d\mu \le \int t g(x) \, d\mu = t \int g(x) \, d\mu \le t \int_{\star} f \, d\mu$$

for all $t > 1$. Taking the limit $t \to 1^+$, we get that $\int^{\star} f\ d\mu \leq \int_{\star} f\ d\mu$, which yield the equality of the integrals, and hence, the μ-integrability of the function f.
Finally, from the estimate above, we get that $\int f\ d\mu \geq 0$.

Definition 1.12.

1. *A function $f : X \to [-\infty, \infty]$ is **μ-summable** if f is μ-integrable and*

$$\int |f|\ d\mu < \infty.$$

2. *We say that a function f: $X \to [-\infty, \infty]$ is **locally μ-summable** if $f|_K$ is μ-summable for each compact set $K \subseteq \mathbb{R}^n$.*

Theorem 1.11 (Fatou's lemma). *Let $f_k : X \to [0, \infty]$ be μ-measurable for $k = 1, 2, \ldots$. Then*

$$\int \liminf_{k\to\infty} f_k\ d\mu \leq \liminf_{k\to\infty} \int f_k\ d\mu.$$

Lemma 1.1. *Let $f_k : X \to [0, \infty]$ be an increasing sequence of not necessarily integrable functions, for which $f_1 \leq \ldots \leq f_k \leq f_{k+1} \leq \ldots$ μ-a.e. Then*

$$\lim_{k\to\infty} \int^{\star} f_k\ d\mu = \int^{\star} \lim_{k\to\infty} f_k\ d\mu$$

Proof. It is clear that, from the monotonicity of the sequence, the limit on the left-hand side exists and that

$$\lim_{k\to\infty} \int^{\star} f_k\ d\mu \leq \int^{\star} \lim_{k\to\infty} f_k\ d\mu$$

From the Infimum Property, we choose φ_k simple μ-integrable functions, such that $0 \leq f_k \leq \varphi_k$ and

$$\int \varphi_k\ d\mu \leq \int^{\star} f_k\ d\mu + \frac{1}{2^k}$$

This implies that

$$\int^{\star} \lim_{k\to\infty} f_k \leq \int \liminf_{k\to\infty} \varphi_k\ d\mu \leq \liminf_{k\to\infty} \int \varphi_k\ d\mu \leq \lim_{k\to\infty} \int^{\star} f_k\ d\mu$$

and the proof is complete. □

As a consequence, we get the following;

Theorem 1.12 (**Monotone Convergence Theorem**). *Let* $f_k : X \to [0,\infty]$ *be* μ*-measurable* $(k = 1, 2, ...)$, *with* $f_1 \leq ... \leq f_k \leq f_{k+1} \leq ....$ *Then*

$$\int \lim_{k\to\infty} f_k \; d\mu = \lim_{k\to\infty} \int f_k \; d\mu.$$

Theorem 1.13 (**Dominated Convergence Theorem**). *Assume* $g \geq 0$ *be a* μ*-summable function and* f, f_k*:* μ*-integrable. Suppose that*

$$f_k \to f \quad \mu - a.e.$$

and

$$|f_k| \leq g \;\; (k = 1, 2, ...)$$

Then;

$$\lim_{k\to\infty} \int |f_k - f| \; d\mu = 0$$

and so

$$\int f_k \, d\mu \to \int f \, d\mu.$$

Finally, in preparation of our groundwork, we shall include here a Proposition [Lemma] from Measure Theory, concerning the upper integral, which will be crucial towards the end of our thesis.

Lemma 1.2. *Let* $f_k : X \to [0,\infty]$ *be a decreasing sequence of not necessarily integrable functions, for which*

$$\limsup_{k\to\infty} \int^{\star}_{\mathbb{R}^m} f_k(y) dy = 0.$$

Then

$$f_k(y) \to 0 \;\; (\mathcal{L}^m - a.e. \; y \in \mathbb{R}^m).$$

Proof. Let us suppose that the conclusion does not hold. This implies that; There exists a subset $B_1 \subseteq \mathbb{R}^m$ of positive measure $\mathcal{L}^m(B_1) > 0$, such that

$$0 < \liminf_{k\to\infty} f_k(y) < \limsup_{k\to\infty} f_k(y) \text{ for all } y \in B_1.$$

Thus

$$B_1 = \bigcup_{\delta>0} \left\{ y \in B_1 \;\middle|\; \limsup_{k\to\infty} f_k(y) \geq \delta \right\} = \bigcup_{n\in\mathbb{N}} \left\{ y \in B_1 \;\middle|\; \limsup_{k\to\infty} f_k(y) \geq \frac{1}{n} \right\}$$

Therefore, there exists a $\delta > 0$ and a $B_2 \subseteq B_1 \subseteq \mathbb{R}^m$, such that;

$$\limsup_{k\to\infty} f_k(y) \geq \delta, \quad \text{for all } y \in B_2.$$

Recall now the definition of the limes superior and that f_k is a decreasing point-wise sequence of functions, and so;

$$\limsup_{k\to\infty} f_k(y) = \lim_{k\to\infty} f_k(y) \geq \delta, \quad \text{for all } y \in B_2.$$

Consequently,

$$f_k(y) \geq \delta, \quad \text{for all } y \in B_2,$$

and, for all $k = 1, 2, \ldots$. Therefore, we obtain that;

$$\int_{\mathbb{R}^m}^{\star} f_k \geq \int_{B_2}^{\star} f_k \geq \int_{B_2}^{\star} \delta = \delta \mathcal{L}^m(B_2) > 0,$$

hence,

$$\limsup_{k\to\infty} \int_{\mathbb{R}^m}^{\star} f_k \geq \delta \mathcal{L}^m(B_2) > 0,$$

which is a contradiction. The proof is complete. □

1.4 Product measures & Fubini's theorem

Consider non-empty sets X and Y.

Definition 1.13. *Let μ be a measure on X and ν be a measure on Y. We define the measure $\mu \times \nu : 2^{X\times Y} \to [0,\infty]$ by*

$$(\mu \times \nu)(S) := \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i) \right\},$$

for each $S \subseteq X \times Y$, where the infimum is taken over all collections of μ-measurable sets $A_i \subseteq X$ and ν-measurable sets $B_i \subseteq Y$ $(i = 1, 2...)$ such that

$$S \subseteq \bigcup_{i=1}^{\infty} (A_i \times B_i).$$

The measure $\mu \times \nu$ is called the ***product measure of μ and ν****.*

Definition 1.14. *Let X be a non-empty set and μ a measure on X.*

1. *A subset $A \subseteq X$ is* ***σ-finite with respect to μ*** *if it can be expressed as*
$$A = \bigcup_{k=1}^{\infty} B_k,$$
where each set B_k is μ-measurable with $\mu(B_k) < \infty$ for $k = 1, 2, \ldots$.

2. *A function $f : X \to [-\infty, \infty]$ is* ***σ-finite with respect to μ****, when f is μ-measurable and $\{x \mid f(x) \neq 0\}$ is σ-finite with respect to μ.*

Theorem 1.14 (**Fubini's theorem**). *Let μ be a measure on X and ν a measure on Y.*

1. *Then $\mu \times \nu$ is a regular measure on $X \times Y$, even if μ and ν are not regular.*

2. *If $A \subseteq X$ is μ-measurable and $B \subseteq Y$ is ν-measurable, then $A \times B$ is $(\mu \times \nu)$-measurable, with*
$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B).$$

3. *If $S \subseteq X \times Y$ is σ-finite with respect to $\mu \times \nu$, then the cross section*
$$S_y \coloneqq \{x \mid (x, y) \in S\}$$
is μ-measurable for ν-a.e. $y \in Y$, and
$$S_x \coloneqq \{y \mid (x, y) \in S\}$$
is ν-measurable for μ-a.e. $x \in X$.
Moreover, $y \mapsto \mu(S_y)$ is ν-integrable & $x \mapsto \nu(S_x)$ is μ-integrable, with
$$(\mu \times \nu)(S) = \int_Y \mu(S_y)\, d\nu(y) = \int_X \nu(S_x)\, d\mu(x).$$

4. *If f is $(\mu \times \nu)$-integrable and f is also σ-finite with respect to $\mu \times \nu$ (in particular, if f is $(\mu \times \nu)$-summable) then the mapping*
$$y \mapsto \int_X f(x, y)\, d\mu(x)$$

is ν-integrable, and the mapping

$$x \mapsto \int_Y f(x,y)\, d\nu(y)$$

is μ-integrable.

Moreover, we have that

$$\begin{aligned}\int_{X\times Y} f\, d(\mu\times\nu) &= \int_Y \left[\int_X f(x,y)\, d\mu(x)\right] d\nu(y)\\ &= \int_X \left[\int_Y f(x,y)\, d\nu(y)\right] d\mu(x).\end{aligned}$$

1.5 Lebesgue measure

Definition 1.15. *We define the **one-dimensional Lebesgue measure** on $\mathbb{R}^1$ as*

$$\begin{aligned}\mathcal{L}^1(A) &\coloneqq \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam} C_i \;\middle|\; A \subseteq \bigcup_{i=1}^{\infty} C_i,\ C_i \subseteq \mathbb{R}\right\}\\ &= \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam} I_i \;\middle|\; A \subseteq \bigcup_{i=1}^{\infty} I_i,\ I_i \text{ interval in } \mathbb{R}\right\}.\end{aligned}$$

Definition 1.16. *We define, inductively, the **n-dimensional Lebesgue measure** $\mathcal{L}^n$ on $\mathbb{R}^1$ by*

$$\mathcal{L}^n \coloneqq \mathcal{L}^{n-1} \times \mathcal{L}^1 = \mathcal{L}^1 \times \cdots \times \mathcal{L}^1 \quad (n \text{ times})$$

Theorem 1.15 (**Equivalent characterisation of Lebesgue measure**). *We have*

$$\mathcal{L}^n = \mathcal{L}^{n-k} \times \mathcal{L}^k$$

for each $k \in \{1, ..., n-1\}$.

Notation. We will write "dx", "dy" etc. rather than "$d\mathcal{L}^n$" in integrals taken with respect to $\mathcal{L}^n$. However, when we need to emphasize on the dimension and/or the variable of integration, we shall do so, by writing the "explicit" notation, like so $d\mathcal{L}^n(x)$.

We will now, for the sake of completeness, state a well-known Theorem concerning the Lebesgue measure of the image of a set under a linear transformation, without proof.

Theorem 1.16 (**Behavior of Lebesgue Measure under Linear Maps**). *Let $L : \mathbb{R}^n \to \mathbb{R}^n$ a linear map and A be a $\mathcal{L}^n$-measurable set. Then the image set $L(A)$ is also $\mathcal{L}^n$-measurable and it holds;*

$$\mathcal{L}^n\big(L(A)\big) = |\det A|\, \mathcal{L}^n(A).$$

1.6 Differentiation of Radon Measures

Let μ and ν be Radon Measures on $\mathbb{R}^n$.

Definition 1.17. *For each point $x \in \mathbb{R}^n$, we define*

$$\overline{D}_\mu \nu(x) \coloneqq \begin{cases} \limsup_{r\to 0} \dfrac{\nu(B(x,r))}{\mu(B(x,r))} & \text{if } \mu(B(x,r)) > 0 \text{ for all } r > 0 \\ +\infty & \text{if } \mu(B(x,r)) = 0 \text{ for some } r > 0 \end{cases}$$

and

$$\underline{D}_\mu \nu(x) \coloneqq \begin{cases} \liminf_{r\to 0} \dfrac{\nu(B(x,r))}{\mu(B(x,r))} & \text{if } \mu(B(x,r)) > 0 \text{ for all } r > 0 \\ +\infty & \text{if } \mu(B(x,r)) = 0 \text{ for some } r > 0 \end{cases}$$

Definition 1.18. *If $\overline{D}_\mu \nu(x) = \underline{D}_\mu \nu(x) < +\infty$, we say ν is **differentiable** with respect to μ at x and write*

$$D_\mu \nu(x) \coloneqq \overline{D}_\mu \nu(x) = \underline{D}_\mu \nu(x).$$

Therefore,

$$D_\mu \nu = \lim_{r\to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))}.$$

*$D_\mu \nu$ is the **derivative** of ν with respect to μ. We also call $D_\mu \nu$ the **density** of ν with respect to μ.*

Theorem 1.17 (**Differentiating measures**). *Let μ and ν be Radon Measures on $\mathbb{R}^n$. Then*

1. *$D_\mu \nu(x)$ exists and is finite μ-a.e., and,*
2. *$D_\mu \nu(x)$ is μ-measurable.*

Definition 1.19. *Let μ and ν be measures on $\mathbb{R}^n$. The measure ν is **absolutely continuous** with respect to μ, and we denote this as*

$$\nu << \mu$$

provided that $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \subseteq \mathbb{R}^n$

Definition 1.20. *Let μ and ν be Borel measures on $\mathbb{R}^n$. We say that μ and ν are **mutually singular**, and we denote this as*

$$\nu \perp \mu$$

if there exists a Borel $B \subseteq \mathbb{R}^n$ such that

$$\mu\big(\mathbb{R}^n \setminus B\big) = \nu(B) = 0.$$

Theorem 1.18 (**Radon-Nikodym Theorem**). *Let μ, ν be Radon measures on $\mathbb{R}^n$ with $\nu << \mu$. Then*

$$\nu(A) = \int_A D_\mu \nu \, d\mu$$

for all μ-measurable sets $A \subseteq \mathbb{R}^n$.

1.7 Lebesgue Differentiation & Density Theorem

Notation. 1. We denote by

$$L^1(X, \mu)$$

the set of all μ-summable functions on X, and by

$$L^1_{\text{loc}}(X, \mu)$$

the set of all locally μ-summable functions.

2. Similarly, if $1 < p < \infty$, we denote by

$$L^p(X, \mu)$$

the set of all μ-measurable functions f on X, such that $|f|^p$ is μ-summable, and by

$$L^p_{\text{loc}}(X, \mu)$$

the set of all μ-measurable functions f on X,such that $|f|^p$ is locally μ-summable.

Notation. We denote the **average value** of f over the set E with respect to a measure μ by

$$\fint_E f \, d\mu \coloneqq \frac{1}{\mu(E)} \int_E f \, d\mu,$$

provided that $0 < \mu(E) < \infty$ and the integral is defined.

Theorem 1.19 (**Lebesgue Differentiation Theorem**)**.** *Let μ be a Radon measure on $\mathbb{R}^n$ and $f \in L^1_{\text{loc}}(\mathbb{R}^n, \mu)$. Then*

$$\lim_{r \to 0} \fint_{B(x,r)} f \, d\mu = f(x)$$

for μ-a.e. $x \in \mathbb{R}^n$.

Theorem 1.20 (**Lebesgue Density Theorem**)**.** *Let $A \subseteq \mathbb{R}^n$ be $\mathcal{L}^n$-measurable. Then*

$$\lim_{r \to 0} \frac{\mathcal{L}^n(B(x,r) \cap A)}{\mathcal{L}^n(B(x,r))} = 1 \quad \mathcal{L}^n - a.e. \;\; x \in A$$

and

$$\lim_{r \to 0} \frac{\mathcal{L}^n(B(x,r) \cap A)}{\mathcal{L}^n(B(x,r))} = 0 \quad \mathcal{L}^n - a.e. \;\; x \in \mathbb{R}^n \setminus A.$$

Theorem 1.21 (**Exhaustion theorem: Filling open sets with balls**)**.** *Let $U \subseteq \mathbb{R}^n$ be open set and $\delta > 0$. There exists a countable collection $\mathfrak{C}$ of disjoint closed balls in U such that* diam $B < \delta$ *for all $B \in \mathfrak{C}$ and*

$$\mathcal{L}^n\left(U \setminus \bigcup_{B \in \mathfrak{C}} B\right) = 0$$

CHAPTER 2

Protheoria II: Hausdorff Measures

In this Chapter, we introduce certain "lower dimensional" measures on $\mathbb{R}^n$, which enable us to "measure" some "very small" subsets of $\mathbb{R}^n$. These are called Hausdorff measures. We begin by proving some fundamental properties and we proceed to show the isoperimetric inequality, an important tool in order to show that $\mathcal{H}^n = \mathcal{L}^n$ on $\mathbb{R}^n$.

For a deeper understanding of Hausdorff measures, we refer to [12], [20], [25] and [4, 9]. For a better visualisation Steiner Symmetrization, we suggest [28] and [15].

2.1 Definitions & elementary properties

Definition 2.1. *Let $A \subseteq \mathbb{R}^n$, $0 \leq s < \infty$, $0 < \delta \leq \infty$. We define*

$$\mathcal{H}^s_\delta(A) \coloneqq \inf\left\{ \sum_{j=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} C_j}{2} \right)^s \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j,\, \operatorname{diam} C_j \;\leq \delta \right\}$$

where

$$\alpha(s) = \frac{\pi^{\frac{s}{2}}}{\Gamma(\frac{s}{2}+1)}$$

and $\Gamma(s) \coloneqq \displaystyle\int_0^\infty e^{-x} x^{s-1}\, dx$ $(0 < s < \infty)$ *is the Gamma function.*

We call

$$\mathcal{H}^s(A) \coloneqq \lim_{\delta \to 0} \mathcal{H}^s_\delta(A) = \sup_{\delta > 0} \mathcal{H}^s_\delta(A)$$

*the **s-dimensional Hausdorff measure** of A on $\mathbb{R}^n$.*

REMARKS.

1. Our demand that $\delta \to 0$ forces the coverings to "follow the local geometry" on A.
2. Observe that $\mathcal{H}^s_\delta(A)$ is a decreasing sequence with respect to δ. Therefore, the limit and the supremum are well-defined.
3. Recall that, for $s > 0$ we have that $\Gamma(s+1) = s\Gamma(s)$. Therefore, if $s = n \in \mathbb{N}$, by induction, we have that $\Gamma(n) = (n-1)!,\ n = 1, 2, \ldots$.
4. Finally, observe that $\mathcal{L}^n(B(x,r)) = \alpha(n)r^n$ for every ball $B(x,r) \subseteq \mathbb{R}^n$. Especially, we shall demonstrate later on that, whenever $s = k \in \mathbb{N}$, the $\mathcal{H}^k$ agrees with the ordinary "k-dimensional surface area" on some "nice" sets, and this is the reason for "adding" $\alpha(s)$ to the definition, so as it serves as a normalising constant.

Theorem 2.1 (**Hausdorff measures are Borel**). *For all* $0 \le s < \infty$, $\mathcal{H}^s$ *is a Borel regular measure in* $\mathbb{R}^n$.

Proof. We will proceed in steps.

`Claim #1:` $\mathcal{H}^s_\delta$ is a measure for every $0 < \delta \le \infty$.

Proof of claim: Let $0 < \delta \le \infty$. Obviously, $\mathcal{H}^s_\delta(\varnothing) = 0$.

Suppose $\{A_k\}_{k=1}^\infty \subseteq \mathbb{R}^n$ and $A \subseteq \bigcup_{k=1}^\infty A_k$. For $\varepsilon > 0$ and $k = 1, 2, \ldots$ we consider a covering $\{C_j^k\}_{j=1}^\infty$ of A_k of the form $A_k \subseteq \bigcup_{j=1}^\infty C_j^k$ with $\operatorname{diam} C_j^k \le \delta$, so that $A \subseteq \bigcup_{k=1}^\infty \bigcup_{j=1}^\infty C_j^k$ and $\mathcal{H}^s_\delta(A_k) + \frac{\varepsilon}{2^k} \ge \sum_{j=1}^\infty \alpha(s)\left(\frac{\operatorname{diam} C_j^k}{2}\right)^s$.
Then

$$\varepsilon + \sum_{k=1}^\infty \mathcal{H}^s_\delta(A_k) = \sum_{k=1}^\infty \left(\frac{\varepsilon}{2^k} + \mathcal{H}^s_\delta(A_k)\right) \ge \sum_{k=1}^\infty \sum_{j=1}^\infty \alpha(s)\left(\frac{\operatorname{diam} C_j^k}{2}\right)^s$$
$$\ge \inf\{\bullet\} = \mathcal{H}^s_\delta(A).$$

Now, by letting $\varepsilon \to 0$ we get that

$$\mathcal{H}^s_\delta(A) \le \sum_{k=1}^\infty \mathcal{H}^s_\delta(A_k).$$

`Claim #2:` $\mathcal{H}^s$ is a measure.

Proof of claim: Again, it is obvious that $\mathcal{H}^s(\varnothing) = 0$.
Let $\{A_k\}_{k=1}^{\infty} \subseteq \mathbb{R}^n$ and $A \subseteq \bigcup_{k=1}^{\infty} A_k$. Then, for every $0 < \delta \leq \infty$, we have that

$$\mathcal{H}^s_\delta(A) \leq \sum_{k=1}^{\infty} \mathcal{H}^s_\delta(A_k) \leq \sum_{k=1}^{\infty} \sup_{\delta'>0} \mathcal{H}^s_{\delta'}(A_k) = \sum_{k=1}^{\infty} \mathcal{H}^s(A_k).$$

Now the right-hand side does not depend on δ and is an upper bound for $\mathcal{H}^s_\delta(A)$. Consequently,

$$\mathcal{H}^s(A) \leq \sum_{k=1}^{\infty} \mathcal{H}^s(A_k).$$

`Claim #3:` $\mathcal{H}^s$ is a Borel measure.

Proof of claim: We are going to use Carathéodory's criterion. For this, let us choose sets $A, B \subseteq \mathbb{R}^n$ with dist$(A,B)> 0$. Select $0 < \delta < \frac{1}{4}\text{dist}(A,B)$ and suppose that $A \cup B \subseteq \bigcup_{k=1}^{\infty} C_k$ with $\operatorname{diam} C_k \leq \delta$.

Notice that, for $z \in A$, we get that $z \in \bigcup_{k=1}^{\infty} C_k$, hence $z \in C_\bullet$ for possibly more than one indices. The same holds for any $w \in B$. We collect those members of our initial cover and form families

$$\mathcal{A} \coloneqq \{\, C_j \mid C_j \cap A \neq \varnothing\} \quad \text{and} \quad \mathcal{B} \coloneqq \{\, C_j \mid C_j \cap B \neq \varnothing\}.$$

Hence, we have that

$$A \subseteq \bigcup_{C_j \in \mathcal{A}} C_j \quad \text{and} \quad B \subseteq \bigcup_{C_j \in \mathcal{B}} C_j, \quad \text{with} \quad C_i \cap C_j = \varnothing.$$

for $C_i \in \mathcal{A}$ and $C_j \in \mathcal{B}$. Therefore

$$\begin{aligned}\sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s &\geq \sum_{C_j \in \mathcal{A}} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s + \sum_{C_j \in \mathcal{B}} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s \\ &\geq \mathcal{H}^s_\delta(A) + \mathcal{H}^s_\delta(B).\end{aligned}$$

Taking the infimum over all such sets $\{C_k\}_{k=1}^{\infty}$, we find that $\mathcal{H}^s_\delta(A\cup B)\geq \mathcal{H}^s_\delta(A)+\mathcal{H}^s_\delta(B)$ provided that $0<4\delta<\operatorname{dist}(A,B)$.
Letting $\delta\to 0$, we obtain

$$\mathcal{H}^s(A\cup B)\geq \mathcal{H}^s(A)+\mathcal{H}^s(B)$$

for all $A,B\subseteq\mathbb{R}^n$ with $\operatorname{dist}(A,B)>0$.

The reverse inequality follows from Claim 2, since $\mathcal{H}^s$ is a measure. Carathéodory's criterion implies that $\mathcal{H}^s$ is a Borel measure.

`Claim #4:` $\mathcal{H}^s$ is a Borel-regular measure.

Proof of claim: We are familiar with the property that $\operatorname{diam}\overline{C}=\operatorname{diam}C$ for all C; hence

$$\mathcal{H}^s_\delta(A)\coloneqq\inf\left\{\sum_{j=1}^{\infty}\alpha(s)\left(\frac{\operatorname{diam}C_j}{2}\right)^s\ \middle|\ A\subseteq\bigcup_{j=1}^{\infty}C_j,\ \operatorname{diam}C_j\leq\delta,\ C_j\text{ closed}\right\}.$$

Choose $A\subseteq\mathbb{R}^n$ such that $\mathcal{H}^s(A)<\infty$. Then, it is obvious that $\mathcal{H}^s_\delta(A)<\infty$ for all $\delta>0$.
For each $k\geq 1$,choose closed sets $\{C_j^k\}_{j=1}^{\infty}$ so that $A\subseteq\bigcup_{j=1}^{\infty}C_j^k$ with $\operatorname{diam}C_j^k\leq\frac{1}{k}$ for which

$$\sum_{j=1}^{\infty}\alpha(s)\left(\frac{\operatorname{diam}C_j^k}{2}\right)^s\leq\mathcal{H}^s_{\frac{1}{k}}(A)+\frac{1}{k}.$$

Now, letting $A_k\coloneqq\bigcup_{j=1}^{\infty}C_j^k$ and $B\coloneqq\bigcap_{k=1}^{\infty}A_k$, B becomes a Borel set. Moreover, $A\subseteq A_k$ for each k and so $A\subseteq B$.
Furthermore,

$$\mathcal{H}^s_{\frac{1}{k}}(B)=\mathcal{H}^s_{\frac{1}{k}}\left(\bigcap_{k=1}^{\infty}A_k\right)\leq\mathcal{H}^s_{\frac{1}{k}}(A_k)=\inf\{\bullet\}\leq\sum_{j=1}^{\infty}\alpha(s)\left(\frac{\operatorname{diam}C_j^k}{2}\right)^s$$
$$\leq\mathcal{H}^s_{\frac{1}{k}}(A)+\frac{1}{k}.$$

Therefore, we obtain that

$$\mathcal{H}^s(B)=\lim_{\delta\to 0}\mathcal{H}^s_\delta(B)\ \underset{\delta\to 0^+\Rightarrow k\to\infty}{\overset{\delta=\frac{1}{k}}{=\!=\!=}}\ \lim_{k\to\infty}\mathcal{H}^s_{\frac{1}{k}}(B)\leq\lim_{k\to\infty}\left(\mathcal{H}^s_{\frac{1}{k}}(A)+\frac{1}{k}\right)=\mathcal{H}^s(A)$$

Finally, recall that $A \subseteq B$, and thus $\mathcal{H}^s(A) \le \mathcal{H}^s(B)$, since $\mathcal{H}^s$ is a measure (Claim 2). Hence $\mathcal{H}^s(B) = \mathcal{H}^s(A)$.

$\square$

REMARK. In the proof of Assertion (*4.*) we used a slightly "different" definition for $\mathcal{H}^s_\delta$, namely that

$$\mathcal{H}^s_\delta(A) \coloneqq \inf\left\{ \sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j,\ \operatorname{diam} C_j \le \delta,\ C_j \text{ closed} \right\}.$$

Truly, the equality holds.

Let $A \subseteq \mathbb{R}^n$. Define

$$\Psi^s_\delta(A) \coloneqq \inf\left\{ \sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} F_j}{2}\right)^s \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} F_j,\ \operatorname{diam} F_j \le \delta,\ F_j \text{ closed} \right\}.$$

It is immediate that since we restrict ourselves in the sub-set of closed coverings, that

$$\begin{aligned}\Psi^s_\delta(A) &\coloneqq \inf\left\{ \sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} F_j}{2}\right)^s \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} F_j, \operatorname{diam} F_j \le \delta,\ F_j \text{ closed} \right\} \\ &\ge \mathcal{H}^s_\delta(A).\end{aligned}$$

Moreover, since $\operatorname{diam} \overline{C} = \operatorname{diam} C$ for any set C, we can treat some of the F_j sets as being the closures of other sets, not necessarily closed or open or none of the above. Hence, for any cover $A \subseteq \bigcup_{j=1}^{\infty} C_j \subseteq \bigcup_{j=1}^{\infty} \overline{C_j}$ with $\operatorname{diam} C_j \le \delta$ consisting now of closed sets, we have that

$$\Psi^s_\delta(A) \le \sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} \overline{C_j}}{2}\right)^s = \sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s.$$

In this case, $\Psi^s_\delta(A)$ simply becomes a lower bound for $\mathcal{H}^s_\delta(A)$, and thus $\Psi^s_\delta(A) \le \mathcal{H}^s_\delta(A)$, proving the equality.

REMARK. $\mathcal{H}^s$ is NOT a Radon measure if $0 \le s < n$, since $\mathbb{R}^n$ is not σ-finite with respect to $\mathcal{H}^s$. (see the **REMARK** following **Theorem 2.3** for the justification).

Theorem 2.2 (**Properties of the Hausdorff measure**).

1. *$\mathcal{H}^0$ is the counting measure.*
2. *$\mathcal{H}^1 = \mathcal{L}^1$ on $\mathbb{R}^1$.*
3. *$\mathcal{H}^s \equiv 0$ on $\mathbb{R}^n$ for all $s > n$.*
4. *$\mathcal{H}^s(\lambda A) = \lambda^s \mathcal{H}^s(A)$ for all $\lambda > 0, A \subseteq \mathbb{R}^n$.*
5. *$\mathcal{H}^s(L(A)) = \mathcal{H}^s(A)$ for all affine isometries $L : \mathbb{R}^n \to \mathbb{R}^n, A \subseteq \mathbb{R}^n$.*

Proof.
1. It is easy to calculate that $\alpha(0) = 1$ and so $\mathcal{H}^0(\{\alpha\}) = 1$, for each $\alpha \in \mathbb{R}^n$. Now, (*1.*) follows.

2. Choose $A \subseteq \mathbb{R}$ and $\delta > 0$. Observe that

$$\begin{aligned}\mathcal{L}^1(A) &= \inf\left\{ \sum_{j=1}^{\infty} \operatorname{diam} C_j \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j \right\} \\ &\leq \inf\left\{ \sum_{j=1}^{\infty} \operatorname{diam} C_j \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j,\ \operatorname{diam} C_j \leq \delta \right\} \\ &= \mathcal{H}^1_\delta(A).\end{aligned}$$

since $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$ and $\alpha(1) = 2$. Hence $\mathcal{L}^1(A) \leq \mathcal{H}^1(A)$.

For the reverse inequality, we choose sets $\{C_j\}_{j=1}^{\infty}$ such that $A \subseteq \bigcup_{j=1}^{\infty} C_j$. Let $I_k \coloneqq [k\delta, (k+1)\delta]$, for $k \in \mathbb{Z}$. Then, for all j, k we have that

$$\operatorname{diam}\big(C_j \cap I_k\big) \leq \delta \quad \text{and} \quad \sum_{k=-\infty}^{\infty} \operatorname{diam}\big(C_j \cap I_k\big) \leq \operatorname{diam} C_j \;\; (j : \text{fixed}).$$

Now, simply observe that

$$\begin{aligned}A \subseteq \bigcup_{j=1}^{\infty} C_j = \bigcup_{j=1}^{\infty} (C_j \cap \mathbb{R}) = \bigcup_{j=1}^{\infty} \left(C_j \cap \bigcup_{k=-\infty}^{\infty} I_k \right) &= \bigcup_{j=1}^{\infty} \left(\bigcup_{k=-\infty}^{\infty} C_j \cap I_k \right) \\ &= \bigcup_{\substack{j=1 \\ k=-\infty}}^{\infty} \big(C_j \cap I_k\big).\end{aligned}$$

Hence,

$$\mathcal{H}^1_\delta(A) \leq \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} \operatorname{diam}\big(C_j \cap I_k\big) \leq \sum_{j=1}^{\infty} \operatorname{diam} C_j \,.$$

and so $\mathcal{H}^1_\delta(A)$ becomes a lower bound for $\mathcal{L}^1(A)$, since

$$\mathcal{L}^1(A) \coloneqq \inf\left\{ \sum_{j=1}^{\infty} \operatorname{diam} C_j \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j \right\}.$$

Therefore, we end up with $\mathcal{H}^1_\delta(A) \leq \mathcal{L}^1(A)$ for all $\delta > 0$, hence

$$\mathcal{H}^1(A) \leq \mathcal{L}^1(A),$$

which, by taking into account the reverse inclusion from above and that this holds for all $A \subseteq \mathbb{R}$, provides us with the equality we were aiming for, namely that $\mathcal{L}^1 = \mathcal{H}^1$ on $\mathbb{R}^1$.

3. Fix an integer $m \geq 1$. The unit cube Q in $\mathbb{R}^n$ can be decomposed into m^n cubes with side $\frac{1}{m}$ and diameter (length of body diagonal) $\frac{\sqrt{n}}{m}$. Therefore

$$\mathcal{H}^s_{\frac{\sqrt{n}}{m}}(Q) \leq \sum_{i=1}^{m^n} \alpha(s) \left(\frac{\sqrt{n}}{m} \right)^s = \alpha(s)\, n^{\frac{s}{2}}\, m^{n-s},$$

where the last term tends to zero, as $m \to \infty$, for $s > n$.
Hence $\mathcal{H}^s(Q) = 0$, and by "exhausting" $\mathbb{R}^n$ with homocentric scaled versions of Q, say of the form $\{ kQ \}_{k=1}^{\infty}$, we get that $\mathcal{H}^s(\mathbb{R}^n) = 0$.

4. Fix $\lambda > 0$. Then, for an arbitrary but fixed $\delta > 0$, we get that

$$\begin{aligned}
\mathcal{H}^s_\delta(\lambda A) &= \inf\left\{ \sum_{k=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} C_k}{2} \right)^s \;\middle|\; \lambda A \subseteq \bigcup_{k=1}^{\infty} C_k,\ \operatorname{diam} C_k \leq \delta \right\} \\
&= \inf\left\{ \sum_{k=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} C_k}{2} \right)^s \;\middle|\; A \subseteq \bigcup_{k=1}^{\infty} \left(\frac{C_k}{\lambda} \right),\ \operatorname{diam} C_k \leq \delta \right\}.
\end{aligned}$$

Set $\widetilde{C_k} = \dfrac{C_k}{\lambda}$. Then $\operatorname{diam} \widetilde{C_k} = \dfrac{1}{\lambda} \operatorname{diam} C_k \leq \dfrac{\delta}{\lambda}$ and so

$$\mathcal{H}^s_\delta(\lambda A) = \inf\left\{ \sum_{k=1}^{\infty} \alpha(s) \left(\frac{\lambda \operatorname{diam} \widetilde{C_k}}{2} \right)^s \;\middle|\; A \subseteq \bigcup_{k=1}^{\infty} \widetilde{C_k},\ \operatorname{diam} \widetilde{C_k} \leq \frac{\delta}{\lambda} \right\}$$

$$= \inf\left\{ \sum_{k=1}^{\infty} \alpha(s) \frac{\lambda^s \,(\operatorname{diam} \widetilde{C_k}\,)^s}{2^s} \;\middle|\; A \subseteq \bigcup_{k=1}^{\infty} \widetilde{C_k},\ \operatorname{diam} \widetilde{C_k} \ \leq \frac{\delta}{\lambda} \right\}$$
$$= \lambda^s \inf\left\{ \sum_{k=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} \widetilde{C_k}}{2} \right)^s \;\middle|\; A \subseteq \bigcup_{k=1}^{\infty} \widetilde{C_k},\ \operatorname{diam} \widetilde{C_k} \ \leq \frac{\delta}{\lambda} \right\}$$
$$= \lambda^s \, \mathcal{H}^s_{\frac{\delta}{\lambda}}(A).$$

Sending $\delta \to 0$, and since the above hold true for any $A \subseteq \mathbb{R}^n$ and any $\lambda > 0$, we get the equality we were aiming for, namely $\mathcal{H}^s(\lambda A) = \lambda^s \mathcal{H}^s(A)$.

5. Let an affine isometry $L : \mathbb{R}^n \to \mathbb{R}^n$ and $A \subseteq \mathbb{R}^n$. A well-known result from Analytic Geometry gives us a beautiful and complete description for all these maps;

Any affine isometry of $\mathbb{R}^n$ is given as

$$L(x) = \mathbb{O}x + \mathfrak{b},$$

where $\mathbb{O}$ is an orthogonal matrix and $\mathfrak{b}$ a fixed vector of $\mathbb{R}^n$.

Let $\delta > 0$. Take sets $\{C_k\}_{k=1}^{\infty}$ such that $A \subseteq \bigcup_{k=1}^{\infty} C_k$ with $\operatorname{diam} C_k \ \leq \delta$. Consequently,

$$L(A) \subseteq L\left(\bigcup_{k=1}^{\infty} C_k \right) = \bigcup_{k=1}^{\infty} L(C_k) = \bigcup_{k=1}^{\infty} \big(\mathbb{O}\, C_k + \mathfrak{b} \big).$$

Let $\widetilde{C_k} = \mathbb{O}\, C_k + \mathfrak{b}$.

Then

$$\operatorname{diam} \widetilde{C_k} \ = \sup_{x,y \in C_k} \big| (\mathbb{O}x + \mathfrak{b}) - (\mathbb{O}y + \mathfrak{b}) \big|$$
$$= \sup_{x,y \in C_k} \big| \mathbb{O}(x-y) \big| = \sup_{x,y \in C_k} \big| x - y \big| = \operatorname{diam} C_k \,.$$

Hence, we get that

$$\mathcal{H}^s_\delta(L(A)) = \inf\left\{ \sum_{k=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} B_k}{2} \right)^s \;\middle|\; L(A) \subseteq \bigcup_{k=1}^{\infty} B_k,\ \operatorname{diam} B_k \ \leq \delta \right\}$$

$$\leq \sum_{k=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} \widetilde{C_k}}{2} \right)^s = \sum_{k=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} C_k}{2} \right)^s.$$

Therefore, $\mathcal{H}^s_\delta(L(A))$ becomes a lower bound for $\mathcal{H}^s_\delta(A)$, hence

$$\mathcal{H}^s_\delta(L(A)) \leq \mathcal{H}^s_\delta(A).$$

Letting $\delta \to 0$, gives us

$$\mathcal{H}^s(L(A)) \leq \mathcal{H}^s(A).$$

Now, the proof is essentially complete, since, L is also an epimorphism, and the inverse map L^{-1} is also an affine isometry ($L^{-1}(y) = \mathbb{O}^{-1}y - \mathfrak{b}', \mathfrak{b}' = \mathbb{O}^{-1}\mathfrak{b}$) and hence

$$\mathcal{H}^s(L(A)) \leq \mathcal{H}^s(A) = \mathcal{H}^s(L^{-1}(L(A))) \leq \mathcal{H}^s(L(A))$$

and so, $\mathcal{H}^s(L(A)) = \mathcal{H}^s(A)$. □

Lemma 2.1. *Suppose $A \subseteq \mathbb{R}^n$ and $\mathcal{H}^s_\delta(A) = 0$ for some $0 < \delta < \infty$. Then $\mathcal{H}^s(A) = 0$.*

Proof. First of all, for $s = 0$, the conclusion is obvious;
As we proved in the previous Lemma, $\mathcal{H}^0$ is the counting measure.

Now, assume that $A \neq \varnothing$. Then there exists $\alpha \in A$, and so

$$\mathcal{H}^0_\delta(A) \geq \mathcal{H}^0_\delta(\{\alpha\}) = 1.$$

Hence

$$\mathcal{H}^0_\delta(A) \geq 1 \text{ for all } \delta > 0.$$

We reached a contradiction. Hence, $A = \varnothing$. The conclusion is immediate.

Now, we study the case of $s > 0$. Fix $\varepsilon > 0$. Then there exist sets $\{C_j\}_{j=1}^\infty$ with $\operatorname{diam} C_j \leq \delta$, such that $A \subseteq \bigcup_{j=1}^\infty C_j$ and

$$\sum_{j=1}^\infty \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s \leq \varepsilon.$$

Now, for each i we get that

$$\operatorname{diam} C_i \leq 2\left(\frac{\varepsilon}{\alpha(s)}\right)^{\frac{1}{s}} = \delta(\varepsilon).$$

Hence

$$\mathcal{H}^s_{\delta(\varepsilon)}(A) \leq \varepsilon.$$

Since $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$, we get that $\mathcal{H}^s(A) = 0$. □

Lemma 2.2. *Let $A \subseteq \mathbb{R}^n$ and $0 \leq s < t < \infty$.*

1. *If $\mathcal{H}^s(A) < \infty$, then $\mathcal{H}^t(A) = 0$.*
2. *$\mathcal{H}^t(A) > 0$, then $\mathcal{H}^s(A) = +\infty$*

Proof. 1. Let $\mathcal{H}^s(A) < \infty$ and $\delta > 0$. From the infimum characterisation and the definition of Hausdorff measure, there exist sets $\{C_j\}_{j=1}^{\infty}$ such that $A \subseteq \bigcup_{j=1}^{\infty} C_j$ with $\operatorname{diam} C_j \leq \delta$ and

$$\sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s \leq \mathcal{H}^s_\delta(A) + 1 \leq \mathcal{H}^s(A) + 1.$$

Hence, we have that

$$\begin{aligned}
\mathcal{H}^t_\delta(A) &\leq \sum_{j=1}^{\infty} \alpha(t)\left(\frac{\operatorname{diam} C_j}{2}\right)^t \\
&= \frac{\alpha(t)}{\alpha(s)} 2^{s-t} \sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s (\operatorname{diam} C_j)^{t-s} \\
&\leq \frac{\alpha(t)}{\alpha(s)}\, 2^{s-t}\, \delta^{t-s}\, \big(\mathcal{H}^s(A) + 1\big).
\end{aligned}$$

By sending $\delta \to 0$, we conclude that $\mathcal{H}^t(A) = 0$. This proves *(1.)*

2. Now, let $\mathcal{H}^t(A) > 0$. For $s < t$ we get that

$$\begin{aligned}
\mathcal{H}^s(A) &= \lim_{\delta \to 0} \mathcal{H}^s_\delta(A) \\
&= \lim_{\delta \to 0} \inf\left\{ \sum_{j=1}^{\infty} \alpha(s)\left(\frac{\operatorname{diam} C_j}{2}\right)^s \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j,\ \operatorname{diam} C_j \leq \delta \right\} \\
&= \lim_{\delta \to 0} \inf\left\{ \sum_{j=1}^{\infty} \alpha(t)\, \frac{\alpha(s)}{\alpha(t)}\, \frac{2^t}{2^s}\left(\frac{\operatorname{diam} C_j}{2}\right)^t (\operatorname{diam} C_j)^{s-t} \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j, \right. \\
&\qquad\qquad \left. \operatorname{diam} C_j \leq \delta \right\} \\
&\geq \lim_{\delta \to 0} \frac{\alpha(s)}{\alpha(t)}\, 2^{t-s} \delta^{s-t}\, \inf\left\{ \sum_{j=1}^{\infty} \alpha(t)\left(\frac{\operatorname{diam} C_j}{2}\right)^t \;\middle|\; A \subseteq \bigcup_{j=1}^{\infty} C_j, \right.
\end{aligned}$$

$$\operatorname{diam} C_j \le \delta \Bigg\}$$

$$= \lim_{\delta\to 0} \frac{1}{\delta^{t-s}} \frac{\alpha(s)2^{t-s}}{\alpha(t)} \inf\Bigg\{ \sum_{j=1}^{\infty} \alpha(t) \left(\frac{\operatorname{diam} C_j}{2}\right)^t \,\Bigg|\, A \subseteq \bigcup_{j=1}^{\infty} C_j,$$

$$\operatorname{diam} C_j \le \delta \Bigg\}$$

$$= +\infty$$

□

Definition 2.2. *The **Hausdorff dimension** of a set $A \subseteq \mathbb{R}^n$ is*

$$\begin{aligned} H_{dim}(A) &\coloneqq \inf\big\{ 0 \le s < \infty \,|\, \mathcal{H}^s(A) = 0 \,\big\} \\ &= \sup\big\{ 0 \le s < \infty \,|\, \mathcal{H}^s(A) = +\infty \,\big\}. \end{aligned}$$

REMARKS. 1. We saw in the previous Lemma that if there exists $s \ge 0$ so that $\mathcal{H}^s(A) < \infty$, then $\mathcal{H}^t(A) = 0$ for all $t > s$. Hence the set of indices for which $\mathcal{H}^\bullet(A) = 0$ is bounded from below, and thus the first definition is well posed.

2. For the second definition, the justification is similar; From the previous Lemma, we saw that if there exist a $t \ge 0$ such that $0 < \mathcal{H}^t(A) < \infty$, then

$$\begin{aligned} &\mathcal{H}^{s'}(A) = +\infty, \quad s' < t \quad (\star) \\ &\mathcal{H}^{s''}(A) = 0, \qquad t < s''. \end{aligned}$$

From $(\star)$ we get that the set of indices where $\{\mathcal{H}^s(A) = +\infty\}$ is bounded from above, with $H_{\mathrm{dim}}(A)$ being an upper bound, hence

$$\sup\big\{ 0 \le s < \infty \,|\, \mathcal{H}^s(A) = +\infty \,\big\} \le H_{\mathrm{dim}}(A) \quad \text{and} \quad H_{\mathrm{dim}}(A) = t.$$

Now, it shall be perfectly clear that the inequality above "collapses" into equality, since, had the inequality been strict, there would have been $\hat{s}$ with $\sup\{\bullet\} < \hat{s} < t$ where (*2.*) from Lemma would imply $\mathcal{H}^{\hat{s}}(A) = +\infty$, which is a contradiction to the definition of the supremum.

Note that if there is no $s \geq 0$ so that $0 < \mathcal{H}^s(A) < +\infty$, the above quantities "collapse" into a minimum/maximum (respectively) and, again, the point where the "discontinuity" of the map $d \mapsto \mathcal{H}^d(A)$ occurs, is the Hausdorff dimension.

Finally, in the case that $H_{\text{dim}}(A) = \inf\big\{0 \leq s < \infty \,|\, \mathcal{H}^s(A) = 0\big\} = 0$, for example, when A is a finite set, the set $\{\mathcal{H}^\bullet = \infty\}$ is empty and we ignore it (equivalently we "adopt" the convention that $\sup\{\varnothing\} = 0$).

This concludes the proof of the equality (and consequentially, the equivalence) between the two definitions.

3. An immediate observation is that $H_{\text{dim}}(A) \leq n$;

Suppose that $H_{\text{dim}}(A) > n$, strictly. Then we immediately stumble upon a contradiction, since in Assertion (*3.*) of Theorem 2.2 we saw that $\mathcal{H}^s \equiv 0$ on $\mathbb{R}^n$ for all $s > n$, and in that case $H_{\text{dim}}(A)$ would not be an infimum.

4. Let $s = H_{\text{dim}}(A)$. Then, from the definition of Hausdorff dimension, we get that $\mathcal{H}^t(A) = 0$ for all $t > s$ and from the second assertion of the previous Lemma, we also get that $\mathcal{H}^t(A) = +\infty$ for all $t < s$.

An intuition behind this is that the volume of a painting on a sheet of paper is zero, and that the "length" of a surface, let's say of a prism for example, is infinite.

5. At the borderline case of $s = H_{\text{dim}}(A)$ we cannot have any general non-trivial information about the value of $\mathcal{H}^s(A)$; all three cases are possible.

6. Based on the above, we can say that, for a fixed set E, the function $d \mapsto \mathcal{H}^d(E)$ is decreasing and attains a finite non-zero value at most once.

Theorem 2.3 (**Properties of the Hausdorff dimension**).

1. *Let $A, B \subseteq \mathbb{R}^n$. If $A \subseteq B$, then $H_{\text{dim}}(A) \leq H_{\text{dim}}(B)$.*
2. *Let $\{A_i\}_{i=1}^{\infty} \subseteq \mathbb{R}^n$. Then* $H_{\text{dim}}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sup\big\{\, H_{\text{dim}}(A_i) \mid i \in \mathbb{N}\,\big\}$.

Proof.

1. Let $s > H_{\text{dim}}(B)$. From the sub-additivity of the $\mathcal{H}^s$-measure and the definition of the Hausdorff dimension, we get that $\mathcal{H}^s(A) \leq \mathcal{H}^s(B) = 0$. Therefore, $H_{\text{dim}}(A) \leq s$. Since this is true for all $s > H_{\text{dim}}(B)$, we immediately get that $H_{\text{dim}}(A) \leq H_{\text{dim}}(B)$.

2. First, we notice that for every $j = 1, 2, \ldots$, we have that $A_j \subseteq \bigcup_{i=1}^{\infty} A_i$. Hence, by passing onto the supremum, we get

$$\sup_i \{ H_{\dim}(A_i) \} \leq H_{\dim}\left(\bigcup_{i=1}^{\infty} A_i\right).$$

For the reverse inequality, let $s > \sup_i \{ H_{\dim}(A_i) \}$. Then, for all $i = 1, 2, ...$ we have that $\mathcal{H}^s(A_i) = 0$ and by the sub-additivity of the $\mathcal{H}^s$-measure, we get that $\mathcal{H}^s\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathcal{H}^s(A_i) = 0$. Therefore, $H_{\dim}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq s$.
Taking infimum over all such s, implies that

$$H_{\dim}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sup_i \{ H_{\dim}(A_i) \}.$$

Hence, we get the desired equality. □

REMARKS. 1. Essentially, what the above theorem tells us is that Hausdorff dimension behaves nicely, namely, that it preserves monotonicity in the $\subseteq$-order and is stable with respect to countable unions.

2. Needles to say that in the finite case, the supremum "collapses" into maximum, namely; $H_{\dim}\left(\bigcup_{i=1}^{k} A_i\right) = \max_{i=1,\ldots,k} \{ H_{\dim}(A_i) \}$.

REMARK. ($\mathbb{R}^n$ is not a σ-finite with respect to $\mathcal{H}^s$ for $s < n$)
Having established the groundwork, we are now ready to present the proof of this Claim we stated earlier in this Chapter.

Let us suppose, momentarily, that $\mathbb{R}^n$ is σ-finite with respect to $\mathcal{H}^s$, for $s < n$. Then $\mathbb{R}^n$ can be decomposed as

$$\mathbb{R}^n = \bigcup_{k=1}^{\infty} A_k, \quad \text{where we have that } \mathcal{H}^s(A_k) < \infty \ (k = 1, 2, ...).$$

However, this would imply that $H_{\dim}(A_k) \leq s$, for all $k = 1, 2, ...$, thus from the above Theorem we get that

$$H_{\dim}(\mathbb{R}^n) = H_{\dim}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sup_k \{ H_{\dim}(A_k)\} \leq s < n.$$

Hence, we have reached a contradiction, thus proving our claim.

2.2 Isodiametric inequality

Our goal in this section is to prove that $\mathcal{H}^n = \mathcal{L}^n$ on $\mathbb{R}^n$. This is not obvious at all, since $\mathcal{L}^n$ is defined as the n-fold product of the one dimensional Lebesgue measure $\mathcal{L}^1$ and therefore

$$\mathcal{L}^n(A) = \inf\left\{ \sum_{i=1}^{\infty} \mathcal{L}^n(Q_i) \;\middle|\; Q_i \,\text{cubes},\ A \subseteq \bigcup_{i=1}^{\infty} Q_i \right\}.$$

Let it be noted that, the above justification would imply the use of rectangular coverings, induced by the Cartesian product of intervals. However, since cubes are a sub-class of rectangles & rectangles can be decomposed into cubes, we can transition into the above definition of $\mathcal{L}^n$.

On the other hand, $\mathcal{H}^n$ is computed with use of arbitrary coverings of small diameter.

REMARK. In the definition of $\mathcal{L}^n$, we could even take balls as coverings.

Lemma 2.3. *Let $f : \mathbb{R}^n \to [0, \infty]$ be $\mathcal{L}^n$-measurable. Then the region "under the graph of f"*

$$A \coloneqq \big\{(x, y) \;\big|\; x \in \mathbb{R}^n,\, y \in \mathbb{R},\, 0 \le y \le f(x)\big\}$$

is $\mathcal{L}^{n+1}$-measurable.

Proof. Consider a function $g : \mathbb{R}^n \times \mathbb{R} \to [0, \infty]$ defined as

$$g(x, y) = f(x) - y$$

with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}$. Then g is $\mathcal{L}^{n+1}$-measurable and thus

$$A = \big\{(x, y) \,\big|\, y \ge 0\big\} \cap \big\{(x, y) \,\big|\, g(x, y) \ge 0\big\}$$

is $\mathcal{L}^{n+1}$-measurable. □

Notation. Fix $\alpha, b \in \mathbb{R}^n$, with $|\alpha| = 1$. We define

$$L_b^a \coloneqq \{b + t\alpha \mid t \in \mathbb{R}\}$$

the line passing through b in the direction α, and

$$P_\alpha \coloneqq \{x \in \mathbb{R}^n \mid x \cdot \alpha = 0\}$$

the plane through the origin perpendicular to α.

Definition 2.3. *Fix an* $\alpha \in \mathbb{R}^n$*, with* $|\alpha| = 1$*, and let* $A \subseteq \mathbb{R}^n$*. We define the* ***Steiner symmetrization*** *of* A *with respect to the plane* P_α *to be the set*

$$S_\alpha(A) \coloneqq \bigcup_{\substack{b \in P_\alpha \\ A \cap L_b^a \neq \varnothing}} \left\{ b + t\alpha \;\middle|\; |t| \leq \frac{1}{2}\mathcal{H}^1(A \cap L_b^a) \right\}.$$

Theorem 2.4 (Properties of Steiner Symmetrization).

1. $\operatorname{diam} S_\alpha(A) \leq \operatorname{diam} A$ *.*

2. If A *is* $\mathcal{L}^n$*-measurable, then so is* $S_a(A)$*, and*

$$\mathcal{L}^n(S_\alpha(A)) = \mathcal{L}^n(A).$$

Proof. 1. Clearly, if $\operatorname{diam} A = \infty$, the inequality holds trivially.
Therefore, we will assume that $\operatorname{diam} A < \infty$ and, without loss of generality, we may suppose that A is closed.

Fix $\varepsilon > 0$ and select $x, y \in S_\alpha(A)$ such that

$$\operatorname{diam} S_\alpha(A) \leq |x - y| + \varepsilon.$$

Set

$$b \coloneqq x - (x \cdot \alpha)\alpha \quad \text{and} \quad c \coloneqq y - (y \cdot \alpha)\alpha.$$

Then $b, c \in P_\alpha$, since

$$b \cdot \alpha = (x - (x \cdot \alpha)\alpha) \cdot \alpha = x \cdot \alpha - (x \cdot \alpha)|\alpha|^2 \overset{|\alpha|=1}{=\!=\!=} x \cdot \alpha - x \cdot \alpha = 0.$$

In the exact same way, we prove that $c \cdot \alpha = 0$, thus $c \in P_\alpha$, as well.

Let

$$\begin{aligned} r &\coloneqq \inf\{t \,|\, b + t\alpha \in A\}, \\ s &\coloneqq \sup\{t \,|\, b + t\alpha \in A\}, \\ u &\coloneqq \inf\{t \,|\, c + t\alpha \in A\}, \\ v &\coloneqq \sup\{t \,|\, c + t\alpha \in A\}. \end{aligned}$$

Then, by construction, we get that $x = b + (x \cdot \alpha)\alpha \in S_\alpha(A)$ and also that $y = c + (y \cdot \alpha)\alpha \in S_\alpha(A)$, hence

$$|x \cdot \alpha| \leq \frac{1}{2}\mathcal{H}^1(A \cap L_b^a) \quad \text{and} \quad |y \cdot \alpha| \leq \frac{1}{2}\mathcal{H}^1(A \cap L_c^a),$$

and, also that

$$s - r = \sup\{t \,|\, b + t\alpha \in A\} - \inf\{t \,|\, b + t\alpha \in A\} \geq \mathcal{H}^1(A \cap L_b^a)$$

and

$$v - u = \sup\{t \,|\, c + t\alpha \in A\} - \inf\{t \,|\, c + t\alpha \in A\} \geq \mathcal{H}^1(A \cap L_c^a).$$

Here, without any loss in generality, we may assume that we have already chosen our points in such a way, that $v - r \geq s - u$. We have

$$\begin{aligned} v - r &\geq \frac{1}{2}(v - r) + \frac{1}{2}(s - u) \\ &= \frac{1}{2}(s - r) + \frac{1}{2}(v - u) \\ &\geq \frac{1}{2}\mathcal{H}^1(A \cap L_b^a) + \frac{1}{2}\mathcal{H}^1(A \cap L_c^a) \\ &\geq |x \cdot \alpha| + |y \cdot \alpha| \\ &\geq |x \cdot \alpha - y \cdot \alpha|. \end{aligned}$$

Moreover,

$$\begin{aligned} (\operatorname{diam} S_\alpha(A) \, - \varepsilon)^2 &\leq |x - y|^2 \\ &= |(b + (x \cdot \alpha)\alpha) - (c + (y \cdot \alpha)\alpha)|^2 \\ &= |(b - c) + (x \cdot \alpha - y \cdot \alpha)\alpha|^2 \\ &= |b - c|^2 + |x \cdot \alpha - y \cdot \alpha|^2 \, |\alpha|^2 + 2(x \cdot \alpha - y \cdot \alpha)\,(b - c) \cdot \alpha \\ &\qquad \text{by the Pythagorean Theorem} \\ &= |b - c|^2 + |x \cdot \alpha - y \cdot \alpha|^2 \quad \text{because } b,c \in P_\alpha \text{ plane} \\ &\leq |b - c|^2 + (v - r)^2 \\ &= |(b + r\alpha) - (c + v\alpha)|^2 \quad \text{via the Pythagorean Theorem and because } b,c \in P_\alpha \text{ plane} \\ &\leq (\operatorname{diam} A\,)^2, \end{aligned}$$

since A is closed and $b + r\alpha, c + v\alpha \in A$.

It follows that $\operatorname{diam} S_\alpha(A) \, - \varepsilon < \operatorname{diam} A$. Since ε is arbitrary, we end up with the desired inequality

$$\operatorname{diam} S_\alpha(A) \, \leq \operatorname{diam} A\,.$$

2. Since $\mathcal{L}^n$ is rotation invariant, we are going to assume that $\alpha = e_n = (0, ..., 0, 1)$. Then $P_\alpha = P_{e_n} = \mathbb{R}^{n-1}$. Since $\mathcal{L}^1 = \mathcal{H}^1$ on $\mathbb{R}$ and $\mathcal{L}^n = \mathcal{L}^1 \times \mathcal{L}^{n-1}$,

we employ Fubini's Theorem and get

$$\begin{aligned}\mathcal{L}^n(A) &= \int \chi_A \, d\mathcal{L}^n \\ &= \int_{\mathbb{R}^{n-1}\times\mathbb{R}} \chi_A(x,y)\, d\mathcal{L}^n(x,y) \\ &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \chi_A(x,y)\, d\mathcal{L}^1(y)\, d\mathcal{L}^{n-1}(x).\end{aligned}$$

Now let $A_x \coloneqq \{y \in \mathbb{R} \,|\, (x,y) \in A\}$. Then

$$\chi_{A_x}(y) = \begin{cases} 1, & y \in A_x \\ 0, & y \notin A_x \end{cases} = \begin{cases} 1, & (x,y) \in A \\ 0, & (x,y) \notin A \end{cases} = \chi_A(x,y).$$

Since the nested integral in the equality above is independent of x, we can continue our calculations as follows

$$\begin{aligned}\mathcal{L}^n(A) &= \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} \chi_{A_x}(y)\, d\mathcal{L}^1(y) \right) d\mathcal{L}^{n-1}(x) \\ &= \int_{\mathbb{R}^{n-1}} \mathcal{L}^1(A_x)\, d\mathcal{L}^{n-1}(x).\end{aligned}$$

Let the map $f : \mathbb{R}^{n-1} \to \mathbb{R}$ be defined as

$$f(b) = \mathcal{H}^1(A \cap L_b^a),$$

where $a, b \in \mathbb{R}^n$, with $|\alpha| = 1$. It is clear that f is $\mathcal{L}^{n-1}$-measurable. Now, recall from Measure Theory[2] that $\mathcal{L}^1$ is translation invariant, thus we get

$$\begin{aligned}\mathcal{L}^n(A) &= \int_{\mathbb{R}^{n-1}} \mathcal{L}^1(A_x)\, d\mathcal{L}^{n-1}(x) \\ &= \int_{\mathbb{R}^{n-1}} \mathcal{L}^1(A \cap L_b^\alpha)\, d\mathcal{L}^{n-1}(b) \\ &= \int_{\mathbb{R}^{n-1}} \mathcal{H}^1(A \cap L_b^\alpha)\, d\mathcal{L}^{n-1}(b) \\ &= \int_{\mathbb{R}^{n-1}} f(b)\, d\mathcal{L}^{n-1}(b).\end{aligned}$$

[2]See [24], Proposition 4.6.i.

Notice, also, that

$$\begin{aligned}
S_\alpha(A) &\coloneqq \bigcup_{\substack{b\in P_\alpha\\ A\cap L_b^a\neq\varnothing}} \left\{ b+t\alpha \,\middle|\, |t| \leq \frac{1}{2}\mathcal{H}^1(A\cap L_b^a)\right\}\\
&= \left\{(b,y) \,\middle|\, -\frac{1}{2}\mathcal{H}^1(A\cap L_b^\alpha) \leq y \leq \frac{1}{2}\mathcal{H}^1(A\cap L_b^\alpha)\right\}\\
&\qquad\qquad\qquad\setminus \left\{ (b,0) \,\middle|\, L_b^a \cap A = \varnothing\right\}\\
&= \left\{(b,y) \,\middle|\, \frac{-f(b)}{2} \leq y \leq \frac{f(b)}{2}\right\} \setminus \left\{(b,0) \,\middle|\, A\cap L_b^a = \varnothing\right\}.
\end{aligned}$$

From Lemma 2.3, it follows that the first part of the union is $\mathcal{L}^n$-measurable, as the union of two $\mathcal{L}^n$-measurable sets, namely "The region under the graph" of our function f and its reflection with respect to $\mathbb{R}^{n-1}$. Let

$$B \coloneqq \left\{(b,0) \,\middle|\, A\cap L_b^a = \varnothing\right\}.$$

Then $B^c = \left\{ (b,0) \,\middle|\, A \cap L_b^a \neq \varnothing\right\} = proj_{\mathbb{R}^{n-1}}(A)$, where $(\bullet)^c$ denotes the complement of a set into its ambient space and $proj_{\mathbb{R}^{n-1}}(A)$ is the projection onto the "floor" of $\mathbb{R}^n$, i.e. $\mathbb{R}^{n-1}$, of the set A. This is an $\mathcal{L}^n$-measurable set, hence B is also $\mathcal{L}^n$-measurable.

This concludes the $\mathcal{L}^n$-measurability of the set $S_\alpha(A)$.

Let

$$\widetilde{B} \coloneqq \left\{b\in\mathbb{R}^{n-1} \,\middle|\, A\cap L_b^a \neq \varnothing\right\}$$

Observe that; For $b\in\mathbb{R}^{n-1}\setminus\widetilde{B}$, we have $f(b) = \mathcal{H}^1(A\cap L_b^\alpha) = \mathcal{H}^1(\varnothing) = 0$ and $\mathcal{L}^n(B) = 0$, since it belongs in a hyperplane of $\mathbb{R}^n$, namely $B\subseteq\mathbb{R}^{n-1}\times\{0\}$.

Consequentially, we have

$$\begin{aligned}
\mathcal{L}^n\big(S_\alpha(A)\big) &= \mathcal{L}^n\Bigg(\left\{(b,y)\in\mathbb{R}^{n-1}\times\mathbb{R} \,\middle|\, \frac{-f(b)}{2} \leq y \leq \frac{f(b)}{2}\right\}\\
&\qquad\qquad\qquad \setminus \left\{ (b,0) \,\middle|\, L_b^a\cap A = \varnothing\right\}\Bigg)\\
&= \mathcal{L}^n\Bigg(\left\{(b,y)\in\mathbb{R}^{n-1}\times\mathbb{R} \,\middle|\, \frac{-f(b)}{2} \leq y \leq \frac{f(b)}{2}\right\}\setminus B\Bigg)\\
&= \mathcal{L}^n\Bigg(\left\{(b,y)\in\widetilde{B}\times\mathbb{R} \,\middle|\, \frac{-f(b)}{2} \leq y \leq \frac{f(b)}{2}\right\}\Bigg)
\end{aligned}$$

$$= \int_{\mathbb{R}} \chi_{\left\{(b,y)\in \widetilde{B}\times\mathbb{R} \,\middle|\, \frac{-f(b)}{2} \le y \le \frac{f(b)}{2}\right\}} \, d\mathcal{L}^n$$

$$= \int_{\mathbb{R}^{n-1}\times\mathbb{R}} \chi_{\widetilde{B}}(b) \cdot \chi_{\left[\frac{-f(b)}{2}, \frac{f(b)}{2}\right]}(y) \, d\mathcal{L}^{n-1}(b) \, d\mathcal{L}^1(y)$$

where by employing Fubini′s Theorem, we get

$$= \int_{\mathbb{R}^{n-1}} \chi_{\widetilde{B}}(b) \left(\int_{\mathbb{R}} \chi_{\left[\frac{-f(b)}{2}, \frac{f(b)}{2}\right]}(y) \, d\mathcal{L}^1(y) \right) d\mathcal{L}^{n-1}(b)$$

$$= \int_{\mathbb{R}^{n-1}} \chi_{\widetilde{B}}(b) \, \mathcal{L}^1\left(\left[\frac{-f(b)}{2}, \frac{f(b)}{2} \right] \right) d\mathcal{L}^{n-1}(b)$$

$$= \int_{\mathbb{R}^{n-1}} \chi_{\widetilde{B}}(b) \, f(b) \, d\mathcal{L}^{n-1}(b)$$

$$= \int_{\mathbb{R}^{n-1}} \left(\chi_{\widetilde{B}}(b) f(b) + \chi_{\mathbb{R}^{n-1}\setminus\widetilde{B}}(b) f(b) \right) d\mathcal{L}^{n-1}(b)$$

$$= \int_{\mathbb{R}^{n-1}} \left(\chi_{\widetilde{B}}(b) + \chi_{\mathbb{R}^{n-1}\setminus\widetilde{B}}(b) \right) f(b) \, d\mathcal{L}^{n-1}(b)$$

$$= \int_{\mathbb{R}^{n-1}} \chi_{\mathbb{R}^{n-1}} \, f(b) \, d\mathcal{L}^{n-1}(b)$$

$$= \int_{\mathbb{R}^{n-1}} f(b) \, d\mathcal{L}^{n-1}(b).$$

Hence, we ended up with our desired equality, namely

$$\mathcal{L}^n(S_\alpha(A)) = \int_{\mathbb{R}^{n-1}} f(b) \, db = \mathcal{L}^n(A).$$

□

Theorem 2.5 (Isodiametric inequality). *For all sets $A \subseteq \mathbb{R}^n$,*

$$\mathcal{L}^n(A) \le \alpha(n) \left(\frac{\operatorname{diam} A}{2} \right)^n.$$

Proof. If $\operatorname{diam} A = \infty$ the inequality is trivial. Hence, we will safely assume that $\operatorname{diam} A < \infty$. Let $\{e_1, ..., e_n\}$ be the standard basis for $\mathbb{R}^n$. Define

$$A_1 \coloneqq S_{e_1}(A), \quad A_2 \coloneqq S_{e_2}(A_1), \quad ..., \quad A_n \coloneqq S_{e_n}(A_{n-1})$$

Write $A^\star = A_n$.

`Claim #1:` $A^\star$ is a symmetric with respect to the origin.

Proof of claim: Clearly, A_1 is symmetric with respect to P_{e_1}. Let $1 \leq k < n$ and suppose that A_k is symmetric with respect to $P_{e_1}, ..., P_{e_k}$. We will prove that A_{k+1} is symmetric with respect to $P_{e_1}, ..., P_{e_{k+1}}$.
First, by definition, we have that $A_{k+1} = S_{e_{k+1}}(A_k)$ is symmetric with respect to $P_{e_{k+1}}$. We fix $1 \leq j \leq k$ and let $S_j : \mathbb{R}^n \to \mathbb{R}^n$ be the reflection through P_{e_j}. Let $b \in P_{e_{k+1}}$. Since we assumed symmetry of A_k with respect to P_{e_j}, we have that $S_j(A_k) = A_k$. Moreover

$$\begin{aligned}\mathcal{H}^1(A_k \cap L_b^{e_{k+1}}) = \mathcal{H}^1\big(S_j(A_k \cap L_b^{e_{k+1}})\big) = \mathcal{H}^1\big(S_j(A_k) \cap S_j(L_b^{e_{k+1}})\big)& \\ = \mathcal{H}^1(A_k \cap L_{S_jb}^{e_{k+1}}).&\end{aligned}$$

Notice that, by definition, we have

$$A_{k+1} = S_{e_{k+1}}(A_k) = \bigcup_{\substack{b \in P_{e_{k+1}} \\ A_k \cap L_b^{e_{k+1}} \neq \varnothing}} \left\{ b + te_{k+1} \;\middle|\; |t| \leq \frac{1}{2}\mathcal{H}^1(A_k \cap L_b^{e_{k+1}}) \right\}.$$

Also, from $S_j(A_k) = A_k$, we get the following expression

$$\begin{aligned} A_{k+1} = S_{e_{k+1}}(A_k) = S_{e_{k+1}}(S_j(A_k))& \\ = \bigcup_{\substack{A_k = S_j(A_k) \ni \hat{b} = S_j b, \\ \hat{b} \in P_{e_{k+1}} \\ A_k \cap L_{S_j b}^{e_{k+1}} \neq \varnothing}} &\left\{ S_j b + te_{k+1} \;\middle|\; |t| \leq \frac{1}{2}\mathcal{H}^1(A_k \cap L_{S_j b}^{e_{k+1}}) \right\} \\ = \bigcup_{\substack{S_j b \in P_{e_{k+1}} \\ A_k \cap L_{S_j b}^{e_{k+1}} \neq \varnothing}} &\left\{ S_j b + te_{k+1} \;\middle|\; |t| \leq \frac{1}{2}\mathcal{H}^1(A_k \cap L_b^{e_{k+1}}) \right\}. \end{aligned}$$

Consequently,

$$\{t \mid b + te_{k+1} \in A_{k+1}\} = \{t \mid S_j b + te_{k+1} \in A_{k+1}\}.$$

Thus $S_j(A_{k+1}) = A_{k+1}$, which implies that A_{k+1} is symmetric to P_{e_j}. Therefore, from the "strong" induction, we get that $A^\star = A_n$ is symmetric with respect to $P_{e_1}, ..., P_{e_n}$. Hence, it is symmetric with respect to the origin, since each point in A_n ends in its catercorner position after being reflected iteratively through all $P_{e_1}, ..., P_{e_n}$.

`Claim #2:` $\mathcal{L}^n(A^\star) \le \alpha(n)\left(\dfrac{\operatorname{diam} A^\star}{2}\right)^n.$

Proof of claim: Choose $x \in A^\star$. Then $-x \in A^\star$ by Claim #1, and so $\operatorname{diam} A \ge 2|x|$. Thus $A \subseteq B\left(0, \dfrac{\operatorname{diam} A^\star}{2}\right)$ and consequentially

$$\mathcal{L}^n(A^\star) \le \mathcal{L}^n\left(B\left(0, \frac{\operatorname{diam} A^\star}{2}\right)\right) = \alpha(n)\left(\frac{\operatorname{diam} A^\star}{2}\right)^n.$$

`Claim #3:` $\mathcal{L}^n(A) \le \alpha(n)\left(\dfrac{\operatorname{diam} A}{2}\right)^n.$

Proof of claim:
Since $\overline{A}$ is $\mathcal{L}^n$-measurable, by an iterative application of Theorem 2.4 we get that

$$\mathcal{L}^n(\overline{A}) = \mathcal{L}^n\big(S_{e_1}(\overline{A})\big) = \mathcal{L}^n\big(\overline{A_1}\big) = \mathcal{L}^n\big(S_{e_2}(\overline{A_1})\big) = \cdots = \mathcal{L}^n\big(\overline{A_n}\big) = \mathcal{L}^n\big((\overline{A})^\star\big),$$

and, doing the same for the diameter of $(\overline{A})^\star$, we end up with

$$\mathcal{L}^n\big((\overline{A})^\star\big) = \mathcal{L}^n(A) \quad \text{and} \quad \operatorname{diam}(\overline{A})^\star \le \operatorname{diam}\overline{A} = \operatorname{diam} A\,.$$

Therefore, we have

$$\begin{aligned}\mathcal{L}^n(A) \le \mathcal{L}^n(\overline{A}) = \mathcal{L}^n\big((\overline{A})^\star\big) &\le \alpha(n)\left(\frac{\operatorname{diam} A^\star}{2}\right)^n \\ &\le \alpha(n)\left(\frac{\operatorname{diam} \overline{A}}{2}\right)^n = \alpha(n)\left(\frac{\operatorname{diam} A}{2}\right)^n,\end{aligned}$$

which proves our assertion. □

REMARK. We do not require A to be enclosed in a ball of diameter $\operatorname{diam} A$. In fact, there exist sets for which this is not possible.
Take, for example, the equilateral triangle of side length ℓ. Its diameter, i.e. the largest distance between two of its points, is

$$\operatorname{diam}(\mathit{triangle}) = \text{side length} = \ell.$$

Yet, the smallest ball containing the set has the circumcircle of the triangle as great circle, thus having a diameter of

$$\operatorname{diam}(\mathit{ball}) = \operatorname{diam}(\mathit{circumcircle}) = \frac{\text{side length}}{\mathit{sin}(\text{facing angle})} = \frac{\ell}{\frac{\sqrt{3}}{2}} = \frac{2\ell}{\sqrt{3}}.$$

Therefore diam $(ball) >$ diam $(triangle)$, which means that we cannot cover the equilateral triangle with a ball of the same diameter.

Theorem 2.6 (The n-dimensional Hausdorff & Lebesgue measure). *We have*

$$\mathcal{H}^n = \mathcal{L}^n \ \text{ on } \ \mathbb{R}^n.$$

Proof. We will proceed in steps.

`Claim #1:` $\mathcal{L}^n(A) \leq \mathcal{H}^n(A)$ for all $A \subseteq \mathbb{R}^n$.

Proof of claim: Fix $\delta > 0$. We choose sets $\{C_j\}_{j=1}^{\infty}$ so that $A \subseteq \bigcup_{j=1}^{\infty} C_j$, with $\operatorname{diam} C_j \leq \delta$. Now, from the Isodiametric Inequality (Thm. 2.5), we get

$$\mathcal{L}^n(A) \leq \sum_{j=1}^{\infty} \mathcal{L}^n(C_j) \leq \sum_{j=1}^{\infty} \alpha(n) \left(\frac{\operatorname{diam} C_j}{2} \right)^n.$$

Taking infimum, we find that $\mathcal{L}^n(A) \leq \mathcal{H}^n_\delta(A)$, and thus $\mathcal{L}^n(A) \leq \mathcal{H}^n(A)$.

Furthermore, from the definition[3] of $\mathcal{L}^n$ as $\mathcal{L}^1 \times \cdots \times \mathcal{L}^1$, we can deduce that, for all $A \subseteq \mathbb{R}^n$ and $\delta > 0$,

$$\mathcal{L}^n(A) = \inf \left\{ \sum_{i=1}^{\infty} \mathcal{L}^n(Q_i) \;\middle|\; Q_i \text{ cubes }, A \subseteq \bigcup_{i=1}^{\infty} Q_i, \operatorname{diam} Q_i \leq \delta \right\}.$$

Hence, from now on we will consider only cubes with vertices parallel to the coordinate axes of $\mathbb{R}^n$.

`Claim #2:` $\mathcal{H}^n$ is absolutely continuous with respect to $\mathcal{L}^n$.

Proof of claim: Observe that, for any cube $Q \subseteq \mathbb{R}^n$ of side length ℓ, we have

$$\mathcal{L}^n(Q) = \ell^n = \left(\frac{\ell\sqrt{n}}{\sqrt{n}} \right)^n = \left(\frac{\operatorname{diam} Q}{\sqrt{n}} \right)^n.$$

Take $C_n \coloneqq \alpha(n) \left(\frac{\sqrt{n}}{2} \right)^n$. Then for each cube $Q \subseteq \mathbb{R}^n$, we have that

$$\alpha(n) \left(\frac{\operatorname{diam} Q}{2} \right)^n = C_n \, \mathcal{L}^n(Q).$$

[3]See more on [24], Chapter 3.

Thus, since we are restricting ourselves to countable coverings consisting of cubes, we have

$$\begin{aligned}\mathcal{H}^n_\delta(A) &\le \inf\left\{ \sum_{i=1}^\infty \alpha(n)\left(\frac{\operatorname{diam} Q_i}{2}\right)^n \;\middle|\; A \subseteq \bigcup_{i=1}^\infty Q_i,\ \operatorname{diam} Q_i \le \delta \right\} \\ &= \inf\left\{ \sum_{i=1}^\infty C_n\,\mathcal{L}^n(Q_i) \;\middle|\; A \subseteq \bigcup_{i=1}^\infty Q_i,\ \operatorname{diam} Q_i \le \delta \right\} \\ &= C_n \inf\left\{ \sum_{i=1}^\infty \mathcal{L}^n(Q_i) \;\middle|\; A \subseteq \bigcup_{i=1}^\infty Q_i,\ \operatorname{diam} Q_i \le \delta \right\} = C_n\,\mathcal{L}^n(A).\end{aligned}$$

Now, by implementing the definition of $\mathcal{H}^n$, we see that the right-hand side is an upper bound for $\mathcal{H}^n_\delta(A)$, and so, we end up with

$$\mathcal{H}^n(A) \le C_n\,\mathcal{L}^n(A).$$

`Claim #3:` $\mathcal{H}^n(A) \le \mathcal{L}^n(A)$ for all $A \subseteq \mathbb{R}^n$.

Proof of claim: Fix $\delta > 0$ and $\varepsilon > 0$. We can select cubes $\{Q_i\}_{i=1}^\infty$ so that $A \subseteq \bigcup_{i=1}^\infty Q_i$ with $\operatorname{diam} Q_i < \delta$ and $\sum_{i=1}^\infty \mathcal{L}^n(Q_i) \le \mathcal{L}^n(A) + \varepsilon$.

Now, according to Theorem 1.21, for each i there exist disjoint closed balls $\{B^i_k\}_{k=1}^\infty$ contained in Q_i° (= interior of Q_i) such that

$$\operatorname{diam} B^i_k \le \delta \quad \text{and} \quad \mathcal{L}^n\left(Q_i \setminus \bigcup_{k=1}^\infty B^i_k\right) = \mathcal{L}^n\left(Q_i^\circ \setminus \bigcup_{k=1}^\infty B^i_k\right) = 0$$

From Claim 2, we get that $\mathcal{H}^n\left(Q_i \setminus \bigcup_{k=1}^\infty B^i_k\right) = 0$. Thus

$$\begin{aligned}\mathcal{H}^n_\delta(A) &\le \sum_{i=1}^\infty \mathcal{H}^n_\delta(Q_i) = \sum_{i=1}^\infty \mathcal{H}^n_\delta\left(\left(Q_i \setminus \bigcup_{k=1}^\infty B^i_k\right) \cup \bigcup_{k=1}^\infty B^i_k\right)\right) \\ &\le \sum_{i=1}^\infty \mathcal{H}^n_\delta\left(\bigcup_{k=1}^\infty B^i_k\right) \le \sum_{i=1}^\infty\sum_{k=1}^\infty \mathcal{H}^n_\delta(B^i_k) \le \sum_{i=1}^\infty\sum_{k=1}^\infty \alpha(n)\left(\frac{\operatorname{diam} B^i_k}{2}\right)^n \\ &= \sum_{i=1}^\infty\sum_{k=1}^\infty \mathcal{L}^n(B^i_k) = \sum_{i=1}^\infty \mathcal{L}^n\left(\bigcup_{k=1}^\infty B^i_k\right) = \sum_{i=1}^\infty \mathcal{L}^n(Q_i) \le \mathcal{L}^n(A) + \varepsilon\end{aligned}$$

Letting $\delta, \varepsilon \to 0$ completes the proof. □

CHAPTER 3

Lipschitz functions & Linear mappings

In the first part of this chapter, we define Lipschitz functions and prove an important Theorem that connects them with Hausdorff measures and then proceed with the proof of Rademacher's Theorem. In the later part, we state some definitions and properties of linear functions and give our definition of the Jacobian.

A comprehensive exposition on Lipschitz functions can be found in [9, 20]. We also suggest [2],[27] and [3] for a detailed substantiation on topics from Linear Algebra.

3.1 An Extension Theorem

Definitions 3.1.1.

1. *Let* $A \subseteq \mathbb{R}^n$*. A function* $f : A \to \mathbb{R}^n$ *is called* ***Lipschitz continuous*** *(or sometimes simply "Lipschitz") provided that*

$$|f(x) - f(y)| \leq C|x - y| \qquad (\star)$$

for some constant C *and all* $x, y \in A$*.*

2. *The smallest constant* C *such that* $(\star)$ *holds for all* x, y *is denoted as*

$$\operatorname{Lip}(f) \coloneqq \sup\left\{ \frac{|f(x) - f(y)|}{|x - y|} \;\middle|\; x, y \in A,\, x \neq y \right\}$$

Thus

$$|f(x) - f(y)| \leq \operatorname{Lip}(f)|x - y| \quad (x, y \in A)$$

3. *A function* $f : A \to \mathbb{R}^n$ *is called* ***locally Lipschitz continuous*** *if for each compact* $K \subseteq A$*, there exists a constant* C_K*, such that*

$$|f(x) - f(y)| \leq C_K|x - y|$$

for all $x, y \in K$.

Theorem 3.1 (**Extension of Lipschitz mappings**). *Assume* $A \subseteq \mathbb{R}^n$ *and let* $f : A \to \mathbb{R}^m$ *be a Lipschitz function. There exists a Lipschitz continuous function* $\overline{f} : \mathbb{R}^n \to \mathbb{R}^m$ *such that*

1. $\overline{f} = f$ *on* A*, and*
2. $\operatorname{Lip}(\overline{f}) \leq \sqrt{m}\operatorname{Lip}(f)$.

Proof. First, we are going to assume that $f : A \to \mathbb{R}$. Define

$$\overline{f}(x) := \inf_{\alpha \in A}\big\{f(\alpha) + \operatorname{Lip}(f)|x - \alpha|\big\} \quad (x \in \mathbb{R}^n).$$

Let $b \in A$.
Since f is Lipschitz on A, we deduce with ease that; For every $\alpha \in A$,

$$f(b) - f(\alpha) \leq |f(b) - f(\alpha)| \leq \operatorname{Lip}(f)|b - \alpha|.$$

Thus,

$$f(\alpha) + \operatorname{Lip}(f)|b - \alpha| \geq f(b).$$

Taking the infimum over all $\alpha \in A$, we get $\overline{f}(b) \geq f(b)$. For the reverse inequality, we observe that (since $b \in A$)

$$\overline{f}(b) = \inf_{b \in A}\big\{f(b) + \operatorname{Lip}(f)|x - b|\big\} \leq f(b) + \operatorname{Lip}(f)|b - b| = f(b).$$

Hence, we get the desired equality on elements of A.
Moreover, if $x, y \in \mathbb{R}^n$, then

$$\begin{aligned}
\overline{f}(x) &= \inf_{\alpha \in A}\big\{f(\alpha) + \operatorname{Lip}(f)|x - \alpha|\big\} \\
&\leq \inf_{\alpha \in A}\big\{f(\alpha) + \operatorname{Lip}(f)(\,|y - \alpha| + |x - y|\,)\big\} \\
&= \inf_{\alpha \in A}\big\{f(\alpha) + \operatorname{Lip}(f)|y - \alpha| + \operatorname{Lip}(f)|x - y|\big\} \\
&= \inf_{\alpha \in A}\big\{f(\alpha) + \operatorname{Lip}(f)|y - \alpha|\big\} + \operatorname{Lip}(f)|x - y| \quad \text{since the last term does not involve } \alpha \\
&= \overline{f}(y) + \operatorname{Lip}(f)|x - y|.
\end{aligned}$$

In a symmetrical way, we can also see that

$$\overline{f}(y) \leq \overline{f}(x) + \mathrm{Lip}(f)|x-y|.$$

Hence, the extension $\overline{f}:\mathbb{R}^n \to \mathbb{R}$ is also a Lipschitz function with constant

$$\mathrm{Lip}(\overline{f}) \leq \mathrm{Lip}(f).$$

In fact, we have something stronger; From the definition of the Lipschitz constant, we get that

$$\begin{aligned}\mathrm{Lip}(\overline{f}) &\coloneqq \sup\left\{ \frac{|\overline{f}(x)-\overline{f}(y)|}{|x-y|} \;\middle|\; x,y\in\mathbb{R}^n,\, x\neq y \right\} \\ &\underset{\overline{f}=f \text{ on } A}{\geq} \sup\left\{ \frac{|f(x)-f(y)|}{|x-y|} \;\middle|\; x,y\in A,\, x\neq y \right\} = \mathrm{Lip}(f).\end{aligned}$$

Therefore, we get that for the Lipschitz constant extension $\overline{f}:\mathbb{R}^n \to \mathbb{R}$ satisfies

$$\mathrm{Lip}(\overline{f}) = \mathrm{Lip}(f).$$

For the general case, let $f : A \to \mathbb{R}^m$ be a Lipschitz function. We can decompose f as $f = (f_1, f_2, ..., f_m)$, where each map $f_i : A \to \mathbb{R}$.
Notice that

$$|f_i(x) - f_i(y)| \leq |f(x) - f(y)| \leq \mathrm{Lip}(f)|x-y|.$$

Therefore, the components f_i are Lipschitz functions with constants $\mathrm{Lip}(f_i)$, for which we get the estimate

$$\mathrm{Lip}(f_i) \leq \mathrm{Lip}(f).$$

We employ the "baby-case" from above m-times, for each function f_i;

There exists Lipschitz continuous extensions $\overline{f_i} : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, ..., m)$, with $\mathrm{Lip}(\overline{f_i}) = \mathrm{Lip}(f_i)$. Therefore

$$\begin{aligned}|\overline{f}(x)-\overline{f}(y)|^2 = \sum_{i=1}^m |\overline{f_i}(x)-\overline{f_i}(y)|^2 &\leq \sum_{i=1}^m \mathrm{Lip}(f_i)^2|x-y|^2 \\ &\leq \sum_{i=1}^m \mathrm{Lip}(f)^2|x-y|^2 \\ &= m\,\mathrm{Lip}(f)^2|x-y|^2.\end{aligned}$$

We have demonstrated that

$$|\overline{f}(x) - \overline{f}(y)| \leq \sqrt{m}\,\mathrm{Lip}(f)|x - y|.$$

Hence,

$$\mathrm{Lip}(\overline{f}) \leq \sqrt{m}\,\mathrm{Lip}(f).$$

□

REMARKS.

1. Of course, the extension is NOT unique. We could also define $\overline{f}$ as

$$\overline{f}(x) = \sup_{\alpha \in A} \big\{ f(\alpha) - \mathrm{Lip}(f)|x - \alpha| \big\}.$$

   and attain the exactly same result.
   One can also verify with ease, that these two extensions are not at all similar, that is of course outside of the set A.

2. At last, Kirszbraun's Thorem asserts that, in fact, there exists an extension $\overline{f}$ with the `same` Lipschitz constant. Its proof differs substantially from what we have presented above, therefore, it is omitted.

Theorem 3.2 (**Hausdorff measure under Lipschitz maps**).

1. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous, $A \subseteq \mathbb{R}^n$ and $0 \leq s < \infty$. Then*

$$\mathcal{H}^s\big(f(A)\big) \leq \big(\mathrm{Lip}(f)\big)^s\, \mathcal{H}^s(A).$$

2. *Suppose $n > k$ and let $P : \mathbb{R}^n \to \mathbb{R}^k$ denote the projection. Assume $A \subseteq \mathbb{R}^n$ and $0 \leq s < \infty$. Then*

$$\mathcal{H}^s\big(P(A)\big) \leq \mathcal{H}^s(A).$$

Proof. 1. Fix $\delta > 0$ and choose sets $\{C_i\}_{i=1}^{\infty}$ so that $A \subseteq \bigcup_{i=1}^{\infty} C_i$, with $\operatorname{diam} C_i \leq \delta$. Now, we have that

$$\operatorname{diam} f(C_i) \leq \mathrm{Lip}(f) \operatorname{diam} C_i \leq \mathrm{Lip}(f)\delta \quad \text{and} \quad f(A) \subseteq \bigcup_{i=1}^{\infty} f(C_i).$$

Thus

$$\mathcal{H}^s_{\mathrm{Lip}(f)\delta}(f(A)) \leq \sum_{i=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} f(C_i)}{2} \right)^s \leq (\mathrm{Lip}(f))^s \sum_{i=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} C_i}{2} \right)^s.$$

Taking infima over all sets $\{C_i\}_{i=1}^{\infty}$, we find

$$\mathcal{H}^s_{\mathrm{Lip}(f)\delta}(f(A)) \leq (\mathrm{Lip}(f))^s\, \mathcal{H}^s_\delta(A).$$

Send $\delta \to 0$ to finish the proof.

2. Assertion (*2.*) follows immediately from (*1.*), since $\mathrm{Lip}(P) = 1$.
To verify that, simply take two distinct points of $\mathbb{R}^n$, namely $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$.

Then, since P is the projection onto the first k-coordinates, we get that

$$\begin{aligned}\|P(x) - P(y)\|_k &= \|(x_1 - y_1, ..., x_k - y_k)\|_k \\ &= \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \leq \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \|x - y\|_n\,.\end{aligned}$$

Simplifying our notation, we write

$$|P(x) - P(y)| \leq |x - y|.$$

Thus $\mathrm{Lip}(P) \leq 1$.

Moreover, from the definition, by taking $x' = (x_1, ..., x_k, 0, ..., 0) \in \mathbb{R}^n$ and $y' = (y_1, ..., y_k, 0, ..., 0) \in \mathbb{R}^n$ we get that

$$\begin{aligned}\mathrm{Lip}(P) &\coloneqq \sup\left\{ \frac{|P(x) - P(y)|}{|x - y|} \;\middle|\; x, y \in \mathbb{R}^n,\, x \neq y \right\} \\ &= \sup\left\{ \frac{\|P(x) - P(y)\|_k}{\|x - y\|_n} \;\middle|\; x, y \in \mathbb{R}^n,\, x \neq y \right\} \\ &\geq \frac{\|P(x') - P(y')\|_k}{\|x' - y'\|_n} = \frac{\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}}{\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2 + 0^2 + \cdots + 0^2}} = 1.\end{aligned}$$

Hence, $\mathrm{Lip}(P) \geq 1$, thus proving the equality. □

Definition 3.1. *For $f : \mathbb{R}^n \to \mathbb{R}^m$ and $A \subseteq \mathbb{R}^n$, we denote the* ***graph*** *of f over A by*

$$G(f; A) \coloneqq \left\{ (x, f(x)) \mid x \in A \right\} \subseteq \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}.$$

Theorem 3.3 (Hausdorff dimension of graphs). *Assume that $f : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathcal{L}^n(A) > 0$.*

1. *$H_{\dim}(G(f; A)) \geq n$.*

2. If f is also Lipschitz continuous, then $H_{\dim}(G(f;A)) = n$.

Proof. 1. Let $P : \mathbb{R}^{n+m} \to \mathbb{R}^n$ denote the standard projection. Then

$$\mathcal{H}^n(G(f;A)) \geq \mathcal{H}^n\big(P(G(f;A))\big) = \mathcal{H}^n(A) > 0$$

and thus $H_{\dim}(G(f;A)) \geq n$.

2. Let Q denote any cube $\mathbb{R}^n$ of side length 1. Subdivide Q into k^n subcubes of side length $\frac{1}{k}$. We name these subcubes $Q_1, ..., Q_{k^n}$ and observe that $\operatorname{diam} Q_i = \frac{\sqrt{n}}{k}$. Define

$$a_j^i \coloneqq \min_{x \in Q_j} f_i(x) \text{ and } b_j^i \coloneqq \max_{x \in Q_j} f_i(x) \quad (i = 1, ..., m).$$

Since, f is Lipschitz continuous, we get that

$$|b_j^i - a_j^i| \leq \operatorname{Lip}(f) \operatorname{diam} Q_j = \operatorname{Lip}(f) \frac{\sqrt{n}}{k}.$$

We now define $C_j \coloneqq Q_j \times \prod_{i=1}^{m} (a_j^i, b_j^i)$. Then for any $x \in Q_j$ we get that $a_j^i \leq f_i(x) \leq b_j^i$ for $i = 1, ..., m$. Thus

$$G(f; A \cap Q_j) \coloneqq \big\{(x, f(x)) \,|\, x \in Q_j \cap A\big\} \subseteq C_j.$$

Moreover, letting $\Omega \coloneqq \prod_{i=1}^m (a_j^i, b_j^i)$ we have that

$$\operatorname{diam} \Omega^2 = \sum_{i=1}^m |b_j^i - a_j^i|^2 \leq \sum_{i=1}^m \operatorname{Lip}(f)^2 \frac{n}{k^2} = m\operatorname{Lip}(f)^2 \frac{n}{k^2}.$$

Therefore,

$$\operatorname{diam} C_j{}^2 \leq \operatorname{diam} Q_j{}^2 + \operatorname{diam} \Omega^2 = \frac{n}{k^2} + m\operatorname{Lip}(f)^2 \frac{n}{k^2} = n\big(1 + m\operatorname{Lip}(f)^2\big)\frac{1}{k^2}$$

Since $G(f; A \cap Q) = \bigcup_{j=1}^{k^n} G(f; A \cap Q_j) \subseteq \bigcup_{j=1}^{k^n} C_j$, for which $\operatorname{diam} C_j < \dfrac{C}{k}$, where $C = \sqrt{n(1 + m\operatorname{Lip}(f)^2)}$, we have

$$\begin{aligned}\mathcal{H}^n_{\frac{C}{k}}(G(f; A \cap Q)) &\leq \sum_{j=1}^{k^n} \alpha(n) \left(\frac{\operatorname{diam} C_j}{2}\right)^n \\ &\leq k^n \alpha(n) \left(\frac{C}{2k}\right)^n = \alpha(n) \left(\frac{C}{2}\right)^n.\end{aligned}$$

Now, if we let $k \to \infty$, because the right-hand side of the inequality is a bounded quantity independent of k, by application of the definition of Hausdorff measure, we find that

$$\mathcal{H}^n\big(G(f; A \cap Q)\big) < \infty.$$

Consequentially, $H_{\dim}(G(f; A \cap Q)) \leq n$. Since we work this estimate for any cube of $\mathbb{R}^n$ of side length 1, we can "exhaust" A with an (at-most)countable collection of such cubes, and by use of Theorem 2.3, we get eventually that $H_{\dim}(G(f; A)) \leq n$. □

3.2 Rademacher's Theorem

Definition 3.2. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called **differentiable** at $x \in \mathbb{R}^n$, if there exists a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{y \to x} \frac{|f(y) - f(x) - L(y - x)|}{|y - x|} = 0,$$

or, using the little-o notation,

$$f(y) = f(x) + L(y - x) + o(|y - x|) \text{ as } y \to x.$$

NOTATION - REMARK. If such a mapping L exists, it is unique, and we will denote it as

$$Df(x)$$

We call $Df(x)$ the **derivative** of f at x.

Proof: Suppose that there exist two linear maps $L_1, L_2 : \mathbb{R}^n \to \mathbb{R}^m$ such that the equation above is satisfied.

Fix x and take any $u \in \mathbb{R}^n$ with $|u| = 1$. Let $y = x + tu$. Then $|y - x| = |tu| = |t|$ and so $y \to x$ becomes $t \to 0$. We now have that

$$\lim_{t \to 0} \frac{|f(x + tu) - f(x) - L_1(tu)|}{|t|} = 0$$

and

$$\lim_{t \to 0} \frac{|f(x + tu) - f(x) - L_2(tu)|}{|t|} = 0.$$

Observe that;

$$\lim_{t\to 0}\frac{|L_1(tu)-L_2(tu)|}{|t|}=$$
$$\lim_{t\to 0}\frac{\big|\big(f(x+tu)-f(x)-L_1(tu)\big)-\big(f(x+tu)-f(x)-L_2(tu)\big)\big|}{|t|}$$
$$\leq \lim_{t\to 0}\frac{|f(x+tu)-f(x)-L_1(tu)|}{|t|}+\lim_{t\to 0}\frac{|f(x+tu)-f(x)-L_2(tu)|}{|t|}=0.$$

Therefore

$$\lim_{t\to 0}\frac{|L_1(tu)-L_2(tu)|}{|t|}=\lim_{t\to 0}\frac{|t(L_1u-L_2u)|}{|t|}$$
$$=\lim_{t\to 0}\frac{|t||L_1u-L_2u|}{|t|}=\lim_{t\to 0}|L_1u-L_2u|=0.$$

Hence, we end up with

$$L_1(u)=L_2(u)\ \text{ for all } u\in\mathbb{R}^n\ \text{ with } |u|=1.$$

For the general case; Let $x\in\mathbb{R}^n$ $(x\neq\vec{0})$. Then, by the linearity of the maps and the preceding relation, we get;

$$L_1(x)=L_1\left(|x|\frac{x}{|x|}\right)=|x|L_1\left(\frac{x}{|x|}\right)$$
$$=|x|L_2\left(\frac{x}{|x|}\right)=L_2\left(|x|\frac{x}{|x|}\right)=L_2(x).$$

We have demonstrated that the two maps we contended earlier are identical. This concludes our proof. □

Theorem 3.4 (**Rademacher's Theorem**). *Assume that $f:\mathbb{R}^n\to\mathbb{R}^m$ is a locally Lipschitz function. Then f is differentiable $\mathcal{L}^n$-a.e.*

Proof. Without loss of generality, we may assume at first, that $m=1$ and that f is Lipschitz continuous, since differentiability is a local property.

`Step 1:` Fix any $u\in\mathbb{R}^n$ with $|u|=1$, and define

$$D_uf(x)\coloneqq\lim_{t\to 0}\frac{f(x+tu)-f(x)}{t}\quad (x\in\mathbb{R}^n),$$

provided that the limit exists.

`Claim #1:` $D_u f(x)$ exists for $\mathcal{L}^n$-a.e. x.

Proof of claim: Since f is continuous,

$$\overline{D}_u f(x) \coloneqq \limsup_{t\to 0} \frac{f(x+tu)-f(x)}{t} = \lim_{k\to\infty} \sup_{\substack{0<|t|<\frac{1}{k} \\ t\in\mathbb{Q}}} \frac{f(x+tu)-f(x)}{t}$$

is Borel measurable. The same holds for

$$\underline{D}_u f(x) \coloneqq \liminf_{t\to 0} \frac{f(x+tu)-f(x)}{t}.$$

Thus

$$\begin{aligned} A_u &= \big\{x\in\mathbb{R}^n \mid D_u f(x) \text{ does not exist}\big\} \\ &= \big\{x\in\mathbb{R}^n \mid \underline{D}_u f(x) < \overline{D}_u f(x)\big\} \end{aligned}$$

is Borel measurable.

For each $x, u \in \mathbb{R}^n$ with $|u| = 1$, we define $\phi : \mathbb{R} \to \mathbb{R}$ by

$$\phi(t) \coloneqq f(x+tu) \quad (t\in\mathbb{R}).$$

It is easy to see that ϕ is Lipschitz continuous, thus absolutely continuous, and thus differentiable $\mathcal{L}^1$-a.e. Hence

$$\begin{aligned} \phi'(t) &\coloneqq \lim_{h\to 0} \frac{\phi(t+h)-\phi(t)}{h} \\ &= \lim_{h\to 0} \frac{f(x+(t+h)u)-f(x+tu)}{h} \\ &= \lim_{h\to 0} \frac{f((x+tu)+hu)-f(x+tu)}{h} \\ &= D_u f(x+tu). \end{aligned}$$

exists $\mathcal{H}^1$-a.e.
In other words, $D_u f(\mathcal{X})$ exists $\mathcal{H}^1$-a.e for $\mathcal{X} \in L_x = \{x+tu \,|\, t\in\mathbb{R}\}$ line, and since x is arbitrary, we can deduce that

$$\mathcal{H}^1\big(A_u \cap L\big) = 0$$

for each line L parallel to u. Fubini's Theorem then implies that

$$\mathcal{L}^n(A_u) = 0.$$

Indeed, we have that

$$\begin{aligned}
\mathcal{L}^n(A_u) &= \int \chi_{A_u}\, d\mathcal{L}^n \\
&= \int_{\mathbb{R}^{n-1}\times\mathbb{R}} \chi_{A_u}(y,z)\, d\mathcal{L}^n(y,z) \\
&= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \chi_{A_u}(y,z)\, d\mathcal{L}^1(z)\, d\mathcal{L}^{n-1}(y).
\end{aligned}$$

Now let $A_u^y \coloneqq \{z \in \mathbb{R} \,|\, (y,z) \in A_u\}$. Then

$$\chi_{A_u^y}(z) = \begin{cases} 1, & z \in A_u^y \\ 0, & z \notin A_u^y \end{cases} = \begin{cases} 1, & (y,z) \in A_u \\ 0, & (y,z) \notin A_u \end{cases} = \chi_{A_u}(y,z).$$

Since the nested integral in the equality above is independent of x, we can continue our calculations as follows

$$\begin{aligned}
\mathcal{L}^n(A_u) &= \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} \chi_{A_u^y}(z)\, d\mathcal{L}^1(z) \right) d\mathcal{L}^{n-1}(y) \\
&= \int_{\mathbb{R}^{n-1}} \mathcal{L}^1(A_u^y)\, d\mathcal{L}^{n-1}(y).
\end{aligned}$$

For each fixed $y \in \mathbb{R}^{n-1}$, we define the map

$$\begin{aligned}
\phi_y : A_u^y &\to A_u \cap L \\
z &\mapsto (y,z)
\end{aligned}$$

where L is the line passing from $(y,\bullet) \in \mathbb{R}^n$ and parallel to u. It is clear that ϕ_y is an isometry. Therefore

$$\mathcal{L}^n(A_u) = \int_{\mathbb{R}^{n-1}} \mathcal{H}^1(A_u \cap L)\, d\mathcal{L}^{n-1} = 0.$$

An immediate consequence of this is that

$$\operatorname{grad} f(x) \coloneqq \big(f_{x_1}(x), ..., f_{x_n}(x)\big)$$

exists for $\mathcal{L}^n$-a.e. point x.

`Step 2:` We will show that

$$D_u f(x) = u \cdot \operatorname{grad} f(x) \quad \text{for } \mathcal{L}^n - a.e. \,\text{point } x$$

Write $u = (u_1, ..., u_n)$ and let $\zeta \in C_c^\infty(\mathbb{R}^n;\mathbb{R})$. It is easy to confirm that the following equality holds true

$$\int_{\mathbb{R}^n} \left[\frac{f(x+tu)-f(x)}{t}\right]\zeta(x)dx = -\int_{\mathbb{R}^n} f(x)\left[\frac{\zeta(x)-\zeta(x-tu)}{t}\right]dx. \qquad (\star)$$

Indeed, simply by performing a linear change of variables, namely the translation $x \mapsto x - tu$, we see that;

$$\int_{\mathbb{R}^n} f(x+tu)\zeta(x)\;dx = \int_{\mathbb{R}^n} f(x)\zeta(x-tu)\;dx.$$

Multiplying both sides with $\dfrac{1}{t}$ and then substracting the term $\displaystyle\int_{\mathbb{R}^n} \frac{f(x)\zeta(x)}{t}\;dx$ implies $(\star)$.

Consider now the following sequence of functions; We define $\phi_k : \mathbb{R}^n \to \mathbb{R}$ as

$$\phi_k(x) \coloneqq \frac{f(x+\frac{1}{k}u)-f(x)}{\frac{1}{k}}\zeta(x).$$

Observe that;

$$\begin{aligned} |\phi_k(x)| &= \left|\frac{f(x+\frac{1}{k}u)-f(x)}{\frac{1}{k}}\zeta(x)\right| \\ &\leq k\,\mathrm{Lip}(f)\left|x+\frac{1}{k}u-x\right||\zeta(x)| = \mathrm{Lip}(f)|u||\zeta(x)| = \mathrm{Lip}(f)|\zeta(x)|, \end{aligned}$$

where

$$\int_{\mathbb{R}^n} \mathrm{Lip}(f)|\zeta(x)|\,dx = \int_{supp(\zeta)} \mathrm{Lip}(f)|\zeta(x)|\,dx < +\infty,$$

since $\zeta \in C_c^\infty(\mathbb{R}^n;\mathbb{R})$, and finaly that

$$\lim_{k\to\infty} \phi_k(x) \underset{t\to 0}{\overset{t=\frac{1}{k}}{=\!=\!=}} \lim_{t\to 0} \frac{f(x+tu)-f(x)}{t}\zeta(x) = D_u f(x)\zeta(x).$$

Thus, all of the requirements of the Dominated Convergence Theorem are fulfilled, and so we can invoke the Theorem alongside with $(\star)$, in order to deduce that;

$$\begin{aligned}\int_{\mathbb{R}^n} D_u f(x)\zeta(x)\, dx &= \lim_{k\to\infty}\int_{\mathbb{R}^n} \phi_k(x)\, dx \\ &= \lim_{k\to\infty}\int_{\mathbb{R}^n} \frac{f(x+\frac{1}{k}u)-f(x)}{\frac{1}{k}}\zeta(x)\, dx \\ &\overset{(\star)}{=} -\lim_{k\to\infty}\int_{\mathbb{R}^n} f(x)\left[\frac{\zeta(x)-\zeta(x-\frac{1}{k}u)}{\frac{1}{k}}\right] dx \\ &= -\int_{\mathbb{R}^n} f(x)D_u\zeta(x)\, dx.\end{aligned}$$

where the last equality stems by employing the Dominated Convergence Theorem on the right-hand side of $(\star)$, in an analogous setting.

Therefore, we can continue our calculations, and get that;

$$\begin{aligned}\int_{\mathbb{R}^n} D_u f(x)\zeta(x)\, dx &= -\int_{\mathbb{R}^n} f(x)D_u\zeta(x)\, dx \\ &= -\int_{\mathbb{R}^n} f(x)\left(\sum_{i=1}^{n} u_i\zeta_{x_i}(x)\right) dx \\ &= -\sum_{i=1}^{n} u_i\int_{\mathbb{R}^n} f(x)\zeta_{x_i}(x)\, dx \\ &= \sum_{i=1}^{n} u_i\int_{\mathbb{R}^n} f_{x_i}(x)\zeta(x)\, dx \\ &= \int_{\mathbb{R}^n}\left(\sum_{i=1}^{n} u_i f_{x_i}(x)\right)\zeta(x)\, dx \\ &= \int_{\mathbb{R}^n}\Big(u\cdot \operatorname{grad} f(x)\Big)\zeta(x)\, dx.\end{aligned}$$

where we also made use of Fubini's Theorem and the absolute continuity of f on lines. Since the above equality holds for all $\zeta \in C_c^\infty(\mathbb{R}^n;\mathbb{R})$, we get that

$$D_u f(x) = u\cdot \operatorname{grad} f(x) \ \text{ for } \mathcal{L}^n - a.e\ x.$$

Indeed, by setting;

$$\mathcal{T}(x) = D_u f(x) - u\cdot \operatorname{grad} f(x) \ \ (x\in\mathbb{R}^n).$$

we have shown that; $\displaystyle\int_{\mathbb{R}^n} \mathcal{T}(x)\zeta(x)\, dx = 0$, for all $\zeta \in C_c^\infty(\mathbb{R}^n;\mathbb{R})$.

Since $D_u f(x) = \lim_{t\to 0} \dfrac{f(x+tu)-f(x)}{t}$ exists $\mathcal{L}^n - a.e.\ x \in \mathbb{R}^n$, we deduce with ease that;

$$\|D_u f(x)\|_{L^\infty} \leq \text{Lip}(f) \text{ for all } u \in \mathbb{R}^n \text{ such that } |u| = 1.$$

Moreover,

$$\|\text{grad}\, f(x)\|_{L^\infty} = \sup_{1\leq i\leq n} \{|f_{x_i}|\} = \sup_{1\leq i\leq n} \{|D_{e_i} f|\} \leq \text{Lip}(f).$$

Hence, we get that;

$$\begin{aligned}\|\mathcal{T}(x)\|_{L^\infty} &= \|D_u f(x) - u \cdot \text{grad}\, f(x)\|_{L^\infty} \\ &\leq \|D_u f(x)\|_{L^\infty} + \|\text{grad}\, f(x)\|_{L^\infty} = 2\text{Lip}(f) < +\infty.\end{aligned}$$

Therefore, $\mathcal{T} \in L^\infty(\mathbb{R}^n)$ and so $\mathcal{T} \in L^1_{\ell oc}(\mathbb{R}^n)$ with $\int_{\mathbb{R}^n} \mathcal{T}(x)\zeta(x)\, dx = 0$, for all $\zeta \in C_c^\infty(\mathbb{R}^n;\mathbb{R})$. This evokes the Lebesgue Differentiation Theorem (Theorem 1.19), which, once employed here, gives us

$$\lim_{r\to 0} ⨍_{B(x,r)} \mathcal{T}\, d\mathcal{L}^n = \mathcal{T}(x)$$

for $\mathcal{L}^n - a.e.\ x \in \mathbb{R}^n$, namely;

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} \mathcal{T}(y)\, dy = \mathcal{T}(x).$$

for $\mathcal{L}^n - a.e.\ x \in \mathbb{R}^n$, where $|\bullet|$ was used to denote the Lebesgue measure, in order to simplify the notation.

Notice now that; For all $n \in \mathbb{N}$ we can find a suitable $\zeta_n \in C_c^\infty(\mathbb{R}^n;\mathbb{R})$ with $supp(\zeta_n) \subseteq B\left(x, \dfrac{1}{n}\right)$ such that;

$$\left| \frac{1}{|B(x,1/n)|} \int_{B(x,1/n)} \mathcal{T}(y)\, dy - \frac{1}{|B(x,1/n)|} \int_{B(x,1/n)} \mathcal{T}(y)\zeta_n(y)\, dy \right| < \frac{1}{n^2}.$$

However, $\int_{B(x,1/n)} \mathcal{T}(y)\zeta_n(y)\, dy = \int_{\mathbb{R}^n} \mathcal{T}\zeta_n\, d\mathcal{L}^n = 0$, leaving us with;

$$\left| \frac{1}{|B(x,1/n)|} \int_{B(x,1/n)} \mathcal{T}(y)\, dy \right| < \frac{1}{n^2}.$$

for all $n \in \mathbb{N}$. Therefore, by the Lebesgue Differentiation Theorem, we obtain that;

$$|\mathcal{T}(x)| \leq \frac{1}{n^2} \ \mathcal{L}^n - a.e.\ x \in \mathbb{R}^n.$$

Hence,

$$\mathcal{T}(x) = 0 \ \mathcal{L}^n - a.e.\ x \in \mathbb{R}^n,$$

which concludes the proof of this step.

`Step 3:` We will show that f is differentiable $\mathcal{L}^n$-a.e.

We begin by choosing $\Omega \coloneqq \{u_k\}_{k=1}^{\infty}$ to be a countable, dense subset of $\partial B(1)$ ($\coloneqq$ the topological border of the closed ball of $\mathbb{R}^n$ of center $\vec{\mathbf{0}}$ and radius 1).

`Claim #2:` Let $\eta > 0$. There exists a finite subset $\Omega_\eta \subseteq \Omega$, which is η-dense in $\partial B(1)$.

Proof of claim: Fix $\eta > 0$. Since $\partial B(1)$ is compact in $\mathbb{R}^n$, it is totally bounded, hence there exist $M \in \mathbb{N}$ and $v_1, ..., v_M \in \partial B(1)$ such that

$$\partial B(1) = B\Big(v_1, \frac{\eta}{2}\Big) \cup \cdots \cup B\Big(v_M, \frac{\eta}{2}\Big).$$

Since Ω dense, there exists $z_i \in \Omega$ $(i = 1, ..., M)$ such that $|v_i - z_i| < \dfrac{\eta}{2}$. Define $\Omega_\eta \coloneqq \{z_1, ..., z_M\}$. Then Ω_η is a finite subset of Ω and for all $v \in \partial B(1)$, from the total-boundedness, there exists $1 \leq i \leq M$, such that $v \in B\Big(v_i, \dfrac{\eta}{2}\Big)$. Therefore

$$|v - z_i| \leq |v - v_i| + |v_i - z_i| < \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$

This concludes the proof of the Claim.

Let, for $k = 1, 2, ...,$

$$A_k \coloneqq \big\{x \in \mathbb{R}^n \ \big| \ D_{u_k} f(x) \ \&\ \mathrm{grad}\, f(x) : \text{exist},\ D_{u_k} f(x) = u_k \cdot \mathrm{grad}\, f(x)\big\}$$

and define

$$A \coloneqq \bigcap_{k=1}^{\infty} A_k.$$

Notice that, Step 2 implies $\mathcal{L}^n(\mathbb{R}^n \setminus A_k) = 0$ $(k = 1, 2, ...)$. Immediately, we can deduce that

$$\mathcal{L}^n(\mathbb{R}^n \setminus A) = 0.$$

It suffices to show that;

`Claim #3:` f is differentiable at each point $x \in A$.

Proof of claim: Fix any $x \in A$. Choose $u \in \partial B(1),\ t \in \mathbb{R} \setminus \{0\}$ and define the quantity

$$Q(x,u,t) \coloneqq \frac{f(x+tu)-f(x)}{t} - u \cdot \operatorname{grad} f(x).$$

Then, for any $w \in \partial B(1)$, we have that

$$\begin{aligned}
&|Q(x,u,t)-Q(x,w,t)| \\
&= \left| \left(\frac{f(x+tu)-f(x)}{t} - u \cdot \operatorname{grad} f(x) \right) \right. \\
&\qquad\qquad \left. - \left(\frac{f(x+tw)-f(x)}{t} - w \cdot \operatorname{grad} f(x) \right) \right| \\
&\leq \left| \frac{f(x+tu)-f(x+tw)}{t} \right| + \left| (u-w) \cdot \operatorname{grad} f(x) \right| \\
&\leq \operatorname{Lip}(f)|u-w| + |\operatorname{grad} f(x)||u-w| \\
&\leq (\sqrt{n}+1)\operatorname{Lip}(f)\,|u-w|. \qquad (\star)
\end{aligned}$$

Let it be noted that for the last step, we used the estimate

$$|\operatorname{grad} f(x)| \leq \sqrt{n}\operatorname{Lip}(f).$$

Indeed, we have that $\operatorname{grad} f(x) \coloneqq \big(f_{x_1}(x), ..., f_{x_n}(x)\big)$ and for each component we get that

$$\left|\frac{\partial f}{\partial x_i}\right| = \lim_{t \to 0} \left| \frac{f(x_1, ..., x_i + t, x_{i+1}, ..., x_n) - f(x_1, ..., x_i, x_{i+1}, ..., x_n)}{t} \right| \leq \operatorname{Lip}(f).$$

Hence

$$|\operatorname{grad} f(x)|^2 = \sum_{i=1}^{n} \left| \frac{\partial f}{\partial x_i} \right|^2 \leq n \operatorname{Lip}(f)^2.$$

Now, fix $\varepsilon > 0$. From Claim 2, by letting $\eta = \dfrac{\varepsilon}{2(\sqrt{n}+1)\operatorname{Lip}(f)}$ we obtain a finite η-dense subset of $\partial B(1)$, meaning that;

For each $u \in \partial B(1)$, there exists u_k, $k \in \{1, ..., N(\eta)\}$, such that

$$|u-u_k| \leq \frac{\varepsilon}{2(\sqrt{n}+1)\operatorname{Lip}(f)}. \qquad (\star\star)$$

Substituting $(\star\star)$ in $(\star)$ for $w = u_k$ gives us

$$|Q(x,u,t)-Q(x,u_k,t)| < \frac{\varepsilon}{2}. \qquad (\star\star\star)$$

Moreover, since $x \in A_k$, by construction, we get

$$\lim_{t\to 0} Q(x, u_k, t) = 0 \quad (k = 1, ..., N).$$

and thus, there exists $\delta > 0$ so that

$$|Q(x, u_k, t)| < \frac{\varepsilon}{2} \quad \text{for all } 0 < |t| < \delta,\ k = 1, ..., N \qquad (\star\star\star\star)$$

Simply, choose $\delta = \min\{\delta_k \,|\, k = 1, ..., N\}$.
Consequently, taking into account $(\star\star\star)$-$(\star\star\star\star)$, we get that for each $u \in \partial B(1)$, there exists $k \in \{1, .., N\}$ such that

$$|Q(x, u, t)| \leq |Q(x, u_k, t)| + |Q(x, u, t) - Q(x, u_k, t)| < \varepsilon$$

for $0 < |t| < \delta$. Note also that the same $\delta > 0$ holds for all $u \in \partial B(1)$.

Finally, choose any $y \in \mathbb{R}^n$, $y \neq x$. Write $u = \dfrac{y-x}{|y-x|}$, so that y can be expressed as $y = x + tu$ for $t = |y - x|$. Then

$$\begin{aligned} f(y) - f(x) - \operatorname{grad} f(x) \cdot (y - x) &= f(x + tu) - f(x) - tu \cdot \operatorname{grad} f(x) \\ &= o(t) \\ &= o(|y - x|). \end{aligned}$$

Hence, f is differentiable at x, with

$$Df(x) = \operatorname{grad} f(x).$$

For the general case; Let us decompose our map $f : \mathbb{R}^n \to \mathbb{R}^m$ into its components $f_i : \mathbb{R}^n \to \mathbb{R}$ $(1 \leq i \leq m)$. As we have seen, each map f_i is also a Lipschitz map. Therefore, we can apply Rademacher's Theorem on each one of them, and so we get that each f_i is differentiable $\mathcal{L}^n$-a.e., with

$$Df_i(x) = \operatorname{grad} f_i(x) = \left(\frac{\partial f_i}{\partial x_1}, ..., \frac{\partial f_i}{\partial x_n}\right), \quad i = 1, ..., m.$$

Hence, we may define

$$L = \begin{pmatrix} \operatorname{grad} f_1 \\ \operatorname{grad} f_2 \\ \vdots \\ \operatorname{grad} f_m \end{pmatrix}$$

Observe now that; For $y \in \mathbb{R}^n$, $y \neq x$ we have;

$$\frac{\|f(y) - f(x) - L(y-x)\|_{\mathbb{R}^m}}{\|y-x\|_{\mathbb{R}^n}} =$$

$$= \frac{\big\|\big(f_1(y), ..., f_m(y)\big) - \big(f_1(x), ..., f_m(x)\big) - \big(\nabla f_1(x)(y-x), ..., \nabla f_m(x)(y-x)\big)\big\|}{\|y-x\|}$$

$$= \frac{\big\|\big(..., f_i(y) - f_i(x) - \nabla f_i(x)(y-x), ...\big)\big\|}{\|y-x\|}$$

$$= \frac{1}{\|y-x\|}\left[\sum_{i=1}^{m} |f_i(y) - f_i(x) - \nabla f_i(x)(y-x)|^2\right]^{1/2}$$

$$= \sum_{i=1}^{m}\left[\left(\frac{|f_i(y) - f_i(x) - \nabla f_i(x)(y-x)|}{\|y-x\|}\right)^2\right]^{1/2}$$

However, we also have that; For $i = 1, ..., m$, and for any $x \in \mathbb{R}^n$ where f_i is differentiable, we get;

$$\lim_{y \to x} \frac{|f(y) - f(x) - \nabla f_i(x)(y-x)|}{\|y-x\|} = 0.$$

Hence, for $\mathcal{L}^n - a.e.\ x \in \mathbb{R}^n$ we end up with

$$\lim_{y \to x} \frac{\|f(y) - f(x) - L(y-x)\|_{\mathbb{R}^m}}{\|y-x\|_{\mathbb{R}^n}} = 0,$$

which concludes our proof. □

Theorem 3.5 (**Differentiability on level sets**).

1. *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$ *be a locally Lipschitz continuous function and*

$$Z \coloneqq \big\{x \in \mathbb{R}^n \mid f(x) = 0\big\}.$$

Then $Df(x) = 0$ *for* $\mathcal{L}^n$*-a.e. point* $x \in Z$.

2. *Let* $f, g : \mathbb{R}^n \to \mathbb{R}^n$ *be a locally Lipschitz continuous and*

$$Y \coloneqq \big\{x \in \mathbb{R}^n \mid g(f(x)) = x\big\}.$$

Then

$$Dg(f(x))Df(x) = I \quad \text{for} \quad \mathcal{L}^n - a.e.\ x \in Y.$$

Proof. 1. We may assume, without loss of generality, that $m = 1$ and $\mathcal{L}^n(Z) > 0$. Choose $x \in Z$ so that $Df(x)$ exists and

$$\lim_{r \to 0} \frac{\mathcal{L}^n(Z \cap B(x,r))}{\mathcal{L}^n(B(x,r))} = 1. \tag{$\star$}$$

Here, the Lebesgue Density Theorem reassures us that $\mathcal{L}^n$-a.e. point $x \in Z$ will do. Then

$$\begin{aligned} f(y) &= f(x) + Df(x) \cdot (y - x) + o(|y - x|) \\ &= Df(x) \cdot (y - x) + o(|y - x|), \quad \text{as } y \to x. \end{aligned} \tag{$\star\star$}$$

We will denote $\alpha \coloneqq Df(x)$ and assume that $\alpha \neq 0$, and define the set

$$S \coloneqq \left\{ u \in \partial B(1) \;\middle|\; \alpha \cdot u \geq \frac{1}{2}|\alpha| \right\}.$$

Moreover, for for each $r > 0$, we define the set

$$S_r \coloneqq \big\{ \lambda u \mid 0 < \lambda \leq r,\ u \in S \big\}.$$

It is immediate that $S_r \subseteq B(r)$ and that $S_r = rS_1$.
For each $u \in S$ and $t > 0$, substituting $y = x + tu$ in ($\star\star$), we get

$$f(x + tu) = \alpha \cdot tu + o(|tu|) \geq \frac{t|\alpha|}{2} + o(t) > 0, \quad \text{as } t \to 0.$$

Hence, there exists $R > 0$ such that

$$f(x + tu) > 0, \quad 0 < t < R,\ u \in S.$$

In particular, for all $0 < r < R$, we get that $f > 0$ on $x + S_r$, thus

$$Z \cap B(x,r) \subseteq B(x,r) \setminus (x + S_r).$$

Consequently, for all $0 < r < R$, we get

$$\begin{aligned} \frac{\mathcal{L}^n(Z \cap B(x,r))}{\mathcal{L}^n(B(x,r))} \leq \frac{\mathcal{L}^n\big(B(x,r) \setminus (x + S_r)\big)}{\mathcal{L}^n(B(x,r))} &= 1 - \frac{\mathcal{L}^n\big(x + S_r\big)}{\mathcal{L}^n(B(x,r))} \\ &= 1 - \frac{\mathcal{L}^n\big(S_r\big)}{\mathcal{L}^n(B(x,r))} \\ &= 1 - \frac{\mathcal{L}^n\big(rS_1\big)}{\mathcal{L}^n(B(x,r))} \end{aligned}$$

$$= 1 - \frac{r^n \mathcal{L}^n\big(S_1\big)}{r^n \alpha(n)} = 1 - \frac{\mathcal{L}^n(S_1)}{\alpha(n)}.$$

Hence, $\limsup_{r\to 0} \frac{\mathcal{L}^n(Z \cap B(x,r))}{\mathcal{L}^n(B(x,r))} \leq 1 - \frac{\mathcal{L}^n(S_1)}{\alpha(n)}$, which, in view of ($\star$), implies that $1 - \frac{\mathcal{L}^n(S_1)}{\alpha(n)} \geq 1$, thus, $\mathcal{L}^n(S_1) = 0$. However, S_1 has non-empty interior, therefore, we have reached a contradiction. The assertion is proved.

2. To prove assertion 2. we first define sets

$$A \coloneqq \{x \mid Df(x) \text{ exists}\} \text{ and } B \coloneqq \{x \mid Dg(x) \text{ exists}\}.$$

Moreover, define

$$X \coloneqq Y \cap A \cap f^{-1}(B).$$

Now, if $z \in Y \setminus X$, then $z \in Y$ and $z \notin X$, thus $z \notin A$ or $z \notin f^{-1}(B)$. Therefore, if $z \notin A$, we get

$$z \in Y \setminus f^{-1}(B),$$

hence

$$f(z) \in \mathbb{R}^n \setminus B,$$

and so

$$z = g(f(z)) \in g(\mathbb{R}^n \setminus B).$$

Combining all of the above

$$z \in (\mathbb{R}^n \setminus A) \cup g(\mathbb{R}^n \setminus B).$$

and thus we end up with

$$Y \setminus X \subseteq (\mathbb{R}^n \setminus A) \cup g(\mathbb{R}^n \setminus B). \tag{$\star\star\star$}$$

Now, since f and g are locally Lipschitz functions, according to Rademacher's Theorem, they are differentiable almost-everywhere on any compact subset of $\mathbb{R}^n$, and by "exhaustion", differentiable almost-everywhere on $\mathbb{R}^n$. Therefore,

$$\mathcal{L}^n\big(\mathbb{R}^n \setminus A\big) = 0 \quad \text{and} \quad \mathcal{L}^n\big(\mathbb{R}^n \setminus B\big) = 0.$$

Moreover, since g is locally Lipschitz, we can apply Theorem 3.2 locally on compact sets, and again by "exhaustion", so as to obtain that

$$\mathcal{H}^n\big(g(\mathbb{R}^n \setminus B)\big) = 0.$$

which, in view of $(\star\star\star)$, implies

$$\mathcal{L}^n(Y \setminus X) = 0.$$

Finally, if $x \in X$, then $Dg(f(x))$ and $Df(x)$ exist; We then apply the Chain rule, and so

$$Dg(f(x))Df(x) = D(g \circ f)(x)$$

exists. Also, on Y we have that $(g \circ f)(x) - x = 0$, and assertion *1.* implies

$$D(g \circ f) = I \quad \mathcal{L}^n - a.e. \text{ on } Y.$$

□

3.3 Linear mappings & Jacobians

Definitions 3.3.1.

1. *A linear map $O : \mathbb{R}^n \to \mathbb{R}^m$ is* ***orthogonal*** *if*

$$(Ox) \cdot (Oy) = x \cdot y$$

for all $x, y \in \mathbb{R}^n$.

2. *A linear map $S : \mathbb{R}^n \to \mathbb{R}^n$ is* ***symmetric*** *if*

$$x \cdot (Sy) = (Sx) \cdot y$$

for all $x, y \in \mathbb{R}^n$.

3. *A linear map $D : \mathbb{R}^n \to \mathbb{R}^n$ is* ***diagonal*** *if there exist $d_1, ..., d_n \in \mathbb{R}$ such that*

$$Dx = (d_1x_1, ..., d_nx_n)$$

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

4. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be linear. The* ***adjoint*** *of A is the linear map $A^* : \mathbb{R}^m \to \mathbb{R}^n$ defined by the relation*

$$(Ax) \cdot y = x \cdot (A^*y)$$

for all $x \in \mathbb{R}^n, y \in \mathbb{R}^m$

We continue by stating some standard facts from Linear Algebra, even though we presume them to be familiar to all readers.

Theorem 3.6.

1. *$A^{**} = A$ for any $A : \mathbb{R}^n \to \mathbb{R}^m$ linear map.*
2. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ and $B : \mathbb{R}^n \to \mathbb{R}^n$ be linear maps. Then*
$$(A \circ B)^* = B^* \circ A^*.$$
3. *$O^* = O^{-1}$ if $O : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal.*
4. *$S^* = S$ if $S : \mathbb{R}^n \to \mathbb{R}^n$ is symmetric.*
5. *If $S : \mathbb{R}^n \to \mathbb{R}^n$ is symmetric, there exists an orthogonal map O: $\mathbb{R}^n \to \mathbb{R}^n$ and a diagonal map $D : \mathbb{R}^n \to \mathbb{R}^n$ such that*
$$S = O \circ D \circ O^{-1}.$$
6. *If $O : \mathbb{R}^n \to \mathbb{R}^m$ is orthogonal, then for $n \leq m$, we have*
$$\begin{aligned} O^* \circ O &= I \text{ on } \mathbb{R}^n, \\ O \circ O^* &= I \text{ on } O(\mathbb{R}^n) \subseteq \mathbb{R}^m. \end{aligned}$$

REMARK. Essentially, what assertion (5.) says, is that all symmetric real matrices are orthogonally diagonalizable.

Proof. Since the proof of the first four Assertions is a direct consequence of the Definition of the Adjoint, we shall omit them, and focus only on Assertion 6.

Let $O : \mathbb{R}^n \to \mathbb{R}^m$ be an orthogonal map. Since O is an isometry, therefore a 1-1 map, we get that $Ker\, O = \{\vec{0}_{\mathbb{R}^n}\}$. Hence, from the First Isomorphism Theorem[4], we obtain that;
$$\dim \left(\mathbb{R}^n \big/ Ker\, O\right) = n = \dim Im(O) \leq m.$$
Moreover, from the Defining Property of the Adjoint, we get that; For all $x, y \in \mathbb{R}^n$
$$x \cdot y = Ox \cdot Oy = x \cdot (O^* \circ Oy).$$
Hence,
$$O^* \circ Oy = y \text{ for all } y \in \mathbb{R}^n.$$

[4]Refer to [27] for a detailed exposition on Isomorphism Theorems and other topics on Linear Algebra.

which concludes the first part of this proof, namely that;

$$O^* \circ O = I_n, \text{ on } \mathbb{R}^n.$$

For the second part, we will need to show that

$$O \circ O^* w = w, \text{ for all } w \in O(\mathbb{R}^n) \subseteq \mathbb{R}^m.$$

Therefore, we will need to show that;

$$v \cdot (O \circ O^* w) = v \cdot w$$

for all $v \in \mathbb{R}^m$ and all $w \in O(\mathbb{R}^n) \subseteq \mathbb{R}^m$. We proceed in steps.

Take any $v \in O(\mathbb{R}^n) \subseteq \mathbb{R}^m$. Then, there exists $x \in \mathbb{R}^n$ such that $v = Ox$. Set $y = O^* w \in \mathbb{R}^n$. Therefore;

$$v \cdot (O \circ O^* w) = Ox \cdot Oy = x \cdot y = x \cdot O^* w = Ox \cdot w = v \cdot w.$$

Now, take any $v \notin O(\mathbb{R}^n)$. Then v can be written as $v = v_1 + v_2$, where $v_1 \in Im(O)$ and $v_2 \perp Im(O)$.[5]

Since $v_1 \in Im(O)$, we already have that $v_1 \cdot (O \circ O^* w) = v_1 \cdot w$, for all $w \in O(\mathbb{R}^n)$. Moreover $v_2 \cdot (O \circ O^* w) = 0$, in view of $O(O^* w) \in Im(O)$. Hence;

$$\begin{aligned} v \cdot (O \circ O^* w) &= (v_1 + v_2) \cdot (O \circ O^* w) \\ &= v_1 \cdot (O \circ O^* w) + v_2 \cdot (O \circ O^* w) \\ &= v_1 \cdot w. \end{aligned}$$

However, we also have that;

$$v \cdot w = (v_1 + v_2) \cdot w = v_1 \cdot w + v_2 \cdot w = v_1 \cdot w,$$

since $v_2 \perp Im(O)$. Therefore

$$v \cdot (O \circ O^* w) = v_1 \cdot w = v \cdot w,$$

which concludes our proof. □

[5]See more on Orthogonal Complement in the Remark following Lemma 5.1.

Theorem 3.7 (**Polar decomposition**). *Let* $L : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear map.*

1. *If* $n \leq m$*, there exists a symmetric map* $S : \mathbb{R}^n \to \mathbb{R}^n$ *and an orthogonal map* $O : \mathbb{R}^n \to \mathbb{R}^m$ *such that*
$$L = O \circ S.$$
2. *If* $n \geq m$*, there exists a symmetric map* $S : \mathbb{R}^m \to \mathbb{R}^m$ *and an orthogonal map* $O : \mathbb{R}^m \to \mathbb{R}^n$ *such that*
$$L = S \circ O^*.$$

Proof. (*1.*) Define $C = L^* \circ L$; then $C : \mathbb{R}^n \to \mathbb{R}^n$. We immediately observe that
$$(Cx) \cdot y = (L^* \circ Lx) \cdot y = Lx \cdot Ly = x \cdot (L^* \circ Ly) = x \cdot (Cy)$$
and also
$$(Cx) \cdot x = (L^* \circ Lx) \cdot x = Lx \cdot Lx = \|Lx\|^2 \geq 0.$$
Thus C is symmetric and non-negative definite.Hence,there exist $\mu_1, ..., \mu_n \geq 0$ and an orthogonal basis $\{x_k\}_{k=1}^n$ of $\mathbb{R}^n$ such that
$$Cx_k = \mu_k x_k \quad (k = 1, ..., n).$$
Since all $\{\mu_k\}_{k=1}^n$ are non-negative,we can represent them as $\mu_k = \lambda_k^2, \lambda_k \geq 0$.
`Claim:` There exists an orthonormal set $\{z_k\}_{k=1}^n$ in $\mathbb{R}^m$ such that
$$Lx_k = \lambda_k z_k \quad (k = 1, ..., n).$$

Proof of claim: If $\lambda_k \neq 0$, define $z_k \coloneqq \frac{1}{\lambda_k} Lx_k$. Then, if $\lambda_k, \lambda_\ell \neq 0$,
$$\begin{aligned} z_k z_\ell = \frac{1}{\lambda_k \lambda_\ell} Lx_k Lx_\ell = \frac{1}{\lambda_k \lambda_\ell} (L^* \circ Lx_k) \cdot x_\ell &= \frac{1}{\lambda_k \lambda_\ell} (Cx_k) \cdot x_\ell \\ &= \frac{\lambda_k^2}{\lambda_k \lambda_\ell} x_k \cdot x_\ell = \frac{\lambda_k}{\lambda_\ell} x_k \cdot x_\ell = \frac{\lambda_k}{\lambda_\ell} \delta_{k\ell} \end{aligned}$$
where δ_{kl} is Kronecker's delta.
Thus the set $\{z_k \,|\, \lambda_k \neq 0\}$ is orthonormal. Finally, in the case that there is a $\lambda_k = 0$ we get that $\mu_\kappa = 0$, and so, $Cx_k = 0$. Consequently,
$$\|Lx_k\|^2 = Cx_k \cdot x_k = 0.$$
Thus, $Lx_k = 0$ and this is consistent with our claim, in a trivial way. Therefore, in this case, we can assign to that index any unit vector z_k, so that the set $\{z_k\}_{k=1}^n$ is orthonormal.

Now, define $S : \mathbb{R}^n \to \mathbb{R}^n$ by

$$Sx_k = \lambda_k x_k \quad (k = 1, ..., n)$$

and $O : \mathbb{R}^n \to \mathbb{R}^m$ by

$$Ox_k = z_k \quad (k = 1, ..., n).$$

Then $O \circ Sx_k = O(\lambda_k x_k) = \lambda_k(Ox_k) = \lambda_k z_k = Lx_k$, and so

$$L = O \circ S.$$

Observe that the mapping S is symmetric; Let $x = \sum_{k=1}^{n} \alpha_k x_k$ and $y = \sum_{\ell=1}^{n} \beta_\ell x_\ell$. Then

$$\begin{aligned} x \cdot S(y) = \left(\sum_{k=1}^{n} \alpha_k x_k\right) \cdot S\left(\sum_{\ell=1}^{n} \beta_\ell x_\ell\right) &= \left(\sum_{k=1}^{n} \alpha_k x_k\right) \cdot \left(\sum_{\ell=1}^{n} \beta_\ell \, S(x_\ell)\right) \\ &= \sum_{k,\ell=1}^{n} \alpha_k \beta_\ell \, x_k \cdot S(x_\ell) \\ &= \sum_{k,\ell=1}^{n} \alpha_k \beta_\ell \, x_k \cdot \lambda_\ell x_\ell = \sum_{k=1}^{n} \alpha_k \beta_k \lambda_k \|x_k\|^2 \end{aligned}$$

since $\{x_k\}_{k=1}^n$ is an orthogonal basis. Also, we have that

$$S(x) \cdot y = \left(\sum_{k=1}^{n} \alpha_k S(x_k)\right) \cdot \left(\sum_{\ell=1}^{n} \beta_\ell x_\ell\right) = \sum_{k,\ell=1}^{n} \alpha_k \beta_\ell \, S(x_k) \cdot x_\ell = \sum_{k=1}^{n} \alpha_k \beta_k \lambda_k \|x_k\|^2$$

Hence, we end up with; $S(x) \cdot y = x \cdot S(y)$, thus proving that S is symmetrical. In a similar way, we can demonstrate that O is orthogonal, which concludes the proof for this assertion.

2. The proof is analogous to the preceding case, when applied to $L^* : \mathbb{R}^m \to \mathbb{R}^n$. □

Definition 3.3. *Let* $L : \mathbb{R}^n \to \mathbb{R}^m$ *be linear.*

1. *If* $n \leq m$*, we write* $L = O \circ S$ *as above, and we define the* ***Jacobian*** *of* L *to be*

$$[\![L]\!] = |\det S|.$$

2. *If* $n \geq m$*, we write* $L = S \circ O^*$ *as above, and we define the* ***Jacobian*** *of* L *to be*

$$[\![L]\!] = |\det S|.$$

REMARK. An immediate observation is that

$$[\![L]\!] = [\![L^*]\!] .$$

Theorem 3.8 (Jacobians and adjoints).

1. *If $n \leq m$,*

$$[\![L]\!]^2 = \det (L^* \circ L).$$

2. *If $n \geq m$,*

$$[\![L]\!]^2 = \det (L \circ L^*).$$

REMARK. A consequence of this Theorem is that the definition of $[\![L]\!]$ is **independent** of the particular choices of O and S.

Proof. 1. Assume $n \leq m$ and write $L = O \circ S$ & $L^* = S \circ O^*$. We then have

$$L^* \circ L = S \circ O^* \circ O \circ S = S \circ I \circ S = S^2,$$

since O is orthogonal. Therefore,

$$\det (L^* \circ L) = (\det S)^2 = [\![L]\!]^2 .$$

2. Assertion (2.) follows easily. □

Theorem 3.9 (Norm of the Adjoint). *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then*

$$\|L\|_{op} = \|L^*\|_{op}$$

Proof. Indeed, take any $x \in \mathbb{R}^n$ with $|x| = 1$. Then;

$$\begin{aligned} |Lx|^2 = Lx \cdot Lx = (L^* \circ L)x \cdot x &\leq |(L^* \circ L)x| \, |x| \\ &\leq \|L^* \circ L\| \, |x|^2 \\ &\leq \|L^*\| \, \|L\| . \end{aligned}$$

Thus, we end up with

$$\|L\|_{op}^2 \leq \|L^*\|_{op} \|L\|_{op} .$$

Hence

$$\|L\|_{op} \leq \|L^*\|_{op} .$$

Consequently, by substituting L^* in place of L and by the property $L^{**} = L$ (Theorem 3.6) we get that;

$$\|L^*\|_{op} \leq \|L^{**}\|_{op} = \|L\|_{op} .$$

This concludes our proof. □

Jacobians of Lipschitz maps

Now, let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous. By Rademacher's Theorem, applied component-wise, f is differentiable $\mathcal{L}^n$-a.e.. Therefore, $Df(x)$ exists $\mathcal{L}^n$-a.e. $x \in \mathbb{R}^n$ and can be regarded as a linear mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$.

Notation. For $f : \mathbb{R}^n \to \mathbb{R}^m$, $f = (f^1, ..., f^m)$, we write the gradient matrix

$$Df(x) = \begin{pmatrix} f^1_{x_1} & \cdots & f^1_{x_n} \\ \vdots & \ddots & \vdots \\ f^m_{x_1} & \cdots & f^m_{x_n} \end{pmatrix}_{m \times n}$$

at each point where $Df(x)$ exists.

Definition 3.4. *For $\mathcal{L}^n$-a.e. point x, we define the* ***Jacobian*** *of f to be*

$$Jf(x) \coloneqq \llbracket Df(x) \rrbracket \,.$$

3.4 Binet-Cauchy formula

Notation.
Let $n \leq m$. We denote by $\Phi(m,n)$ the set of all maps $\{1, ..., n\} \to \{1, ..., m\}$. Moreover, we define

$$\Sigma(m,n) \coloneqq \big\{\lambda \in \Phi(m,n) \mid \lambda : \text{injective}\big\}.$$

Especially, when $m = n$, we will use the abreviation $\Sigma_n \coloneqq \Sigma(n,n)$, i.e., Σ_n is the set of premutations of $\{1, ..., n\}$.
Finally, we define the set of indicatrices as

$$\Lambda(m,n) \coloneqq \big\{\lambda : \{1, ..., n\} \to \{1, ..., m\} \mid \lambda : \text{strictly increasing}\big\},$$

and for each $\lambda \in \Lambda(m,n)$, the indexed projection $P_\lambda : \mathbb{R}^m \to \mathbb{R}^n$ as

$$P_\lambda(x_1, ..., x_m) = (x_{\lambda(1)}, ..., x_{\lambda(n)}).$$

Theorem 3.10 (**Binet-Cauchy formula**). *Let $n \leq m$ and $L : \mathbb{R}^n \to \mathbb{R}^m$ a linear map. Then*

$$[\![L]\!]^2 = \sum_{\lambda \in \Lambda(m,n)} \big(\det (P_\lambda \circ L)\big)^2.$$

REMARK. What this Theorem essentially tells us is that; We can calculate the $[\![L]\!]^2$ by adding the squares of the determinants of all $(n \times n)$-submatrices of the "larger" $(m \times n)$-matrix identifying the linear map L.

Proof. Let $(L_{ij})_{m \times n}$ be the corresponding matrix induced by the linear map L, with respect to the standard coordinate basis.

We define the $(n \times n)$-matrix $A \coloneqq L^* \circ L$, having elements (A_{ij}), given as

$$A_{ij} = \sum_{k=1}^{m} (L^*)_{ik} L_{kj} = \sum_{k=1}^{m} L_{ki} L_{kj}.$$

Recall that, the determinant of any $(n \times n)$-matrix M with entries (m_{ij}) is given - via the Leibniz formula - as

$$\det M = \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} m_{i\sigma(i)}.$$

Hence, we proceed with the calculations. We have that

$$\begin{aligned}
[\![L]\!]^2 = \det A &= \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i\sigma(i)} = \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \sum_{k=1}^{m} L_{ki} L_{k\sigma(i)} \\
&= \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) \sum_{\phi \in \Phi(m,n)} \prod_{i=1}^{n} L_{\phi(i)i} L_{\phi(i)\sigma(i)} \\
&\overset{(\dagger)}{=} \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) \sum_{\phi \in \Sigma(m,n)} \prod_{i=1}^{n} L_{\phi(i)i} L_{\phi(i)\sigma(i)}.
\end{aligned}$$

Where we passed with equality in $(\dagger)$, because for a non-injective map $\phi \in \Phi(m,n)$, we would get

$$\sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} L_{\phi(i)i} L_{\phi(i)\sigma(i)} = 0,$$

which would does not infect the "general" sum.

Notice also that, each $\phi \in \Sigma(m,n)$ can be written uniquely as $\phi = \lambda \circ \theta$, where $\lambda \in \Lambda(m,n)$ and $\theta \in \Sigma_n$. Hence, we can continue our calculations, as follows;

$$\begin{aligned}
[\![L]\!]^2 &= \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \prod_{i=1}^{n} L_{\lambda\circ\theta(i),i} L_{\lambda\circ\theta(i),\sigma(i)} \\
&= \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \prod_{\{i=\theta^{-1}(j)\,|\,1\le j\le n\}} L_{\lambda\circ\theta(i),i} L_{\lambda\circ\theta(i),\sigma(i)} \\
&= \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \prod_{j=1}^{n} L_{\lambda(j),\theta^{-1}(j)} L_{\lambda(j),\sigma\circ\theta^{-1}(j)} \\
&= \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} L_{\lambda(i),\theta(i)} L_{\lambda(i),\sigma\circ\theta(i)} \\
&= \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \sum_{\{\sigma=\rho\circ\theta^{-1}\,|\,\rho\in\Sigma_n\}} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} L_{\lambda(i),\theta(i)} L_{\lambda(i),\sigma\circ\theta(i)} \\
&\qquad\qquad \text{Note that } \operatorname{sgn}(\sigma) = \operatorname{sgn}(\rho)\cdot\operatorname{sgn}(\theta). \text{ Hence} \\
&= \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \sum_{\rho\in\Sigma_n} \operatorname{sgn}(\rho)\cdot\operatorname{sgn}(\theta) \prod_{i=1}^{n} L_{\lambda(i),\theta(i)} L_{\lambda(i),\rho(i)} \\
&= \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \operatorname{sgn}(\theta) \sum_{\rho\in\Sigma_n} \operatorname{sgn}(\rho) \prod_{i=1}^{n} L_{\lambda(i),\theta(i)} L_{\lambda(i),\rho(i)} \\
&= \sum_{\lambda\in\Lambda(m,n)} \sum_{\theta\in\Sigma_n} \operatorname{sgn}(\theta) \left[\sum_{\rho\in\Sigma_n} \operatorname{sgn}(\rho) \prod_{i=1}^{n} L_{\lambda(i),\rho(i)} \right] \prod_{i=1}^{n} L_{\lambda(i),\theta(i)} \\
&= \sum_{\lambda\in\Lambda(m,n)} \left(\sum_{\theta\in\Sigma_n} \operatorname{sgn}(\theta) \prod_{i=1}^{n} L_{\lambda(i),\theta(i)} \right)^2 \qquad (\dagger\dagger) \\
&= \sum_{\lambda\in\Lambda(m,n)} \left(\det P_\lambda \circ L\right)^2.
\end{aligned}$$

Note that the equality in (††) stems from the Leibniz formula; For a fixed $\lambda \in \Lambda(m,n)$, we get that

$$\left(P_\lambda \circ L\right)_{ij} = \sum_{k=1}^{m} (P_\lambda)_{ik} (L)_{kj} = L_{\lambda(i)j} \text{ since } (P_\lambda)_{ij} = \begin{cases} 1, & \text{when } j = \lambda(i) \\ 0, & \text{elsewhere.} \end{cases}$$

□

REMARK. The Binet-Cauchy equality admits an elegant geometric interpretation;

Indeed, let us consider a set Ω with unitary Lebesgue measure. We identify the linear maps with the matrices they induce. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map and take $C = L(\Omega)$. Then, using the notation we established earlier, $P_\lambda \circ L$ is but the projection from $\mathbb{R}^m$ to the n-dimensional subspace spanned by the canonical basis vectors $\{e_{\lambda(1)}, ..., e_{\lambda(n)}\}$. Therefore, up to sign, the $\det P_\lambda \circ L$ is the measure of the projection $P_\lambda(C)$, and the Binet-Cauchy formula can be restated as

$$\mathcal{L}^n(C)^2 = \sum_{\lambda \in \Lambda(m,n)} \mathcal{L}^n\big(P_\lambda(C)\big)^2 \tag{$\star$}$$

Now, the above equation reads as follows; The squared volume of an n-dimensional parallepiped contained in $\mathbb{R}^m$ is the sum of the squared volumes of its projections to all possible subspaces.

This brings us to the beauty of the special case where $n = 1$. Here, the parallepiped collapses to an interval and $P_\lambda(C)$ declare the projections to the coordinate axes. Hence, equation ($\star$) can be interpreted as a multidimensional analogue, of "algebraic" nature, of the Pythagorean Theorem.

3.5 Hadamard's inequality

We now turn our attention to an important tool of Linear Algebra, the so-called **Hadamard's inequality**, which will prove itself useful later on. Algebraically, it is a bound on the determinant of a matrix in terms of the lenghths of its column vectors. Geometrically, we can say that Hadamard's inequality gives us an upper bound for the volume of a parallepiped indicated by vectors $u_1, ..., u_n$ of $\mathbb{R}^n$, which is the product of the lengths of those vectors.

Theorem 3.11 (Hadamard's inequality). *Let A be a $(n \times n)$-matrix and denote by a_i $(1 \leq i\ eqn)$ its i-th column. Then*

$$|\det A| \leq \|a_1\| \cdots \|a_n\|\,.$$

Proof. First, if the matrix A is singular, i.e. not invertible, then the result holds trivially. Hence, we can safely assume that $\det A \neq 0$ and so we can write A as

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}.$$

By dividing each column by its length, we get the induced matrix

$$M = \begin{bmatrix} \frac{a_1}{\|a_1\|} & \frac{a_2}{\|a_2\|} & \cdots & \frac{a_n}{\|a_n\|} \end{bmatrix}.$$

where each column has lenght 1. Here, the Hadamard's inequality, once proven, gives us that

$$|\det M\,| \leq 1. \tag{$\star$}$$

Now, the geneality is achieved once we consider that

$$|\det A\,| = \left(\prod_{i=1}^{n} \|a_i\|\right) |\det M\,| \leq \prod_{i=1}^{n} \|a_i\|\,.$$

Therefore, it suffices to show that ($\star$) holds.

Indeed, let us consider the matrix $P = M^*M$. We immediately see that P is a symmetric real matrix, therefore P is diagonalisable (from the Spectral Theorem) with eigenvalues $\lambda_1, ..., \lambda_n$. Moreover, for $1 \leq i \leq n$ we have that;

$$(P)_{ii} = \sum_{k=1}^{n} (M^*)_{ik}(M)_{ki} = \sum_{k=1}^{n} (M_{ki})^2 = \sum_{k=1}^{n} \frac{a_{ki}^2}{\|a_i\|^2} = 1.$$

Since every element of the diagonal of P is equal to 1, we have that the trace of P is equal to n. Hence, by the famous Arithmetic-Geometric Means inequality, we get that

$$\det P = \prod_{i=1}^{n} \lambda_i \leq \left(\frac{1}{n}\sum_{i=1}^{n} \lambda_i\right)^n = \left(\frac{1}{n}\operatorname{tr} P\right)^n = 1^n = 1.$$

which essentially concludes our proof, since $\det M = \sqrt{\det P} = 1$. □

CHAPTER 4

The Area Formula

In the proceeding two Chapters, we will study Lipschitz continuous mappings of the form

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

and derive some special formulas regarding the integral of the Jacobian.

We begin by breaking the problem into two parts, according to the relative size of n and m. For $n \leq m$, we get the Area formula. This is what we will study in this Chapter. We start by proving some introductory lemmas, and then the aforementioned formula. We conclude by presenting some important applications.

This Chapter is still primarily influenced by Evans & Gariepy [8, 7], who, in their own words, follow the work of Hardt in [13] whose work is in turn built upon Federer [10]. We have also consulted the exposition of [18] and [12].

Throughout this Chapter, we assume

$$n \leq m.$$

4.1 Preliminaries

Lemma 4.1. *Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, with $[\![L]\!] > 0$. We consider*

$$\nu(A) = \mathcal{H}^n\big(L(A)\big) \text{ for } A \subseteq \mathbb{R}^n.$$

Then ν is a Radon measure.

Proof. We will proceed in steps.

`Step 1:` ν is a measure of $\mathbb{R}^n$.

We immediately observe that

$$\nu(\varnothing) = \mathcal{H}^n\big(L(\varnothing)\big) = \mathcal{H}^n(\varnothing) = 0,$$

and, if $A \subseteq \mathbb{R}^n$ with $A \subseteq \bigcup_{i=1}^{\infty} A_i$, we have that

$$\begin{aligned}\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathcal{H}^n\left(L\left(\bigcup_{i=1}^{\infty} A_i\right)\right) = \mathcal{H}^n\left(\bigcup_{i=1}^{\infty} L(A_i)\right) &\le \sum_{i=1}^{\infty} \mathcal{H}^n\big(L(A_i)\big)\\ &= \sum_{i=1}^{\infty} \nu(A_i).\end{aligned}$$

`Step 2:` ν is a Borel measure.

From Theorem 3.7, we have the following decomposition

$$L = O \circ S,$$

for a symmetric map $S : \mathbb{R}^n \to \mathbb{R}^n$ and an orthogonal map $O : \mathbb{R}^n \to \mathbb{R}^m$. Moreover, according to Definition 3.3, $[\![L]\!] = |\det S\,| > 0$.

Let $B \subseteq \mathbb{R}^n$, a Borel set. Now, for every $X \subseteq \mathbb{R}^n$, we get that

$$\begin{aligned}
&\nu(X \cap B) + \nu(X \setminus B)\\
&\quad= \mathcal{H}^n(L(X \cap B)) + \mathcal{H}^n(L(X \setminus B))\\
&\quad= \mathcal{H}^n(O \circ S(X \cap B)) + \mathcal{H}^n(O \circ S(X \setminus B)) \quad \text{since } \mathcal{H}^n \text{ is invariant under isometries}\\
&\quad= \mathcal{H}^n(S(X \cap B)) + \mathcal{H}^n(S(X \setminus B))\\
&\quad= \mathcal{L}^n(S(X \cap B)) + \mathcal{L}^n(S(X \setminus B)) \quad \text{from Theorem 2.6}\\
&\quad= \mathcal{L}^n(S(X) \cap S(B)) + \mathcal{L}^n(S(X) \setminus S(B)) \quad \text{since } S{:}1{-}1\\
&\quad= \mathcal{L}^n(S(X)) \quad \text{since } S\text{: continuous \& invertible and } S{=}(S^{-1})^{-1},\ S(B) \text{ is Borel, thus } \mathcal{L}^n{-}\text{measurable}\\
&\quad= \mathcal{H}^n(S(X)) \quad \text{since } S(X) \subseteq \mathbb{R}^n\\
&\quad= \mathcal{H}^n(O \circ S(X))\\
&\quad= \mathcal{H}^n(L(X)) = \nu(X).
\end{aligned}$$

Hence, the Borel set B we started with is ν-measurable, and since B is chosen arbitrarily, this holds for all Borel sets. Thus, ν is a Borel measure.

`Step 3:` ν is a Borel-regular measure.

Let $A \subseteq \mathbb{R}^n$. Then, since $\mathcal{L}^n$ is Borel-regular, there exists a Borel-measurable set $\tilde{B}$ such that $\tilde{B} \supseteq S(A)$ and $\mathcal{L}^n(\tilde{B}) = \mathcal{L}^n(S(A))$.

Set $B \coloneqq S^{-1}(\tilde{B})$. Now, B is Borel and $A \subseteq B$, with

$$\nu(A) = \mathcal{H}^n(L(A)) = \mathcal{H}^n(O \circ S(A)) = \mathcal{L}^n(S(A)) = \mathcal{L}^n(\tilde{B}) = \mathcal{L}^n(S(B)) = \nu(B)$$

`Step 4:` ν is a Radon measure.

Let $K \subseteq \mathbb{R}^n$, K: compact. It is easy to see that

$$\nu(K) = \mathcal{H}^n(L(K)) = \mathcal{H}^n(O \circ S(K)) = \mathcal{L}^n(S(K)) < \infty,$$

since S is continuous,ergo $S(K)$ is compact, and $\mathcal{L}^n$ is a Radon measure. □

Lemma 4.2. *Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping. Then*

$$\mathcal{H}^n\big(L(A)\big) = [\![L]\!]\, \mathcal{L}^n(A)$$

for all $A \subseteq \mathbb{R}^n$.

Proof. Using the Polar Decomposition Theorem (Thm. 3.7) we get that L can be expressed as $L = O \circ S$ for a symmetric map $S : \mathbb{R}^n \to \mathbb{R}^n$ and an orthogonal map $O : \mathbb{R}^n \to \mathbb{R}^m$, with $[\![L]\!] = |\det S|$.
We explore the following two cases;

`Case 1:` $[\![L]\!] = 0$.

In this case, we get $|\det S| = 0$. Now, recall the dimension formula for linear maps; $n = \dim Ker(S) + \dim Im(S)$. Since $\det S = 0$, S is not invertible, hence S is not one-to-one, and so $Ker(S) \neq \{\vec{\mathbf{0}}\}$. Consequently, $\dim Ker(S) \geq 1$, which in turn implies that $\dim Im(S) = \dim S(\mathbb{R}^n) \leq n-1$. Hence $\dim L(\mathbb{R}^n) \leq n - 1 < n$. Therefore $\mathcal{H}^n\big(L(\mathbb{R}^n)\big) = 0$.

`Case 2:` $[\![L]\!] > 0$.

Now, we have that ($x \in \mathbb{R}^n$, $r > 0$)

$$\begin{aligned}
\frac{\mathcal{H}^n\big(L(B(x,r))\big)}{\mathcal{L}^n\big(B(x,r)\big)} &= \frac{\mathcal{H}^n\big(O \circ S(B(x,r))\big)}{\mathcal{L}^n\big(B(x,r)\big)} = \frac{\mathcal{H}^n\big(S(B(x,r))\big)}{\mathcal{L}^n\big(B(x,r)\big)} \\
&= \frac{\mathcal{L}^n\big(S(B(x,r))\big)}{\mathcal{L}^n\big(B(x,r)\big)} = \frac{\mathcal{L}^n\big(S(B(1))\big)}{\alpha(n)} \\
&= |\det S| = [\![L]\!]\,,
\end{aligned}$$

where we have used the rotation invariance of $\mathcal{H}^n$ and Theorems 1.16 & 2.6.

Defining $\nu(A) = \mathcal{H}^n\big(L(A)\big)$ for $A \subseteq \mathbb{R}^n$ as in the above lemma, we get that ν is a Radon measure, with $\nu << \mathcal{L}^n$. Indeed;

Since $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map between finite dimensional spaces, we can employ the Operator norm and get that;

$$|L(u)| \leq \|L\|_{op}\,|u|, \text{ for all } u \in \mathbb{R}^n$$

And so, simply by taking $u = x - y$ for $x, y \in \mathbb{R}^n$, from the linearity of L stems that

$$|L(x) - L(y)| \leq \|L\|_{op}\,|x - y|.$$

Hence, L is a Lipschitz map with $\mathrm{Lip}(L) = \|L\|_{op} < +\infty$, the latter following immediately from the definition of the Lipschitz constant and the Operator norm.

Now, let $E \subseteq \mathbb{R}^n$ such that $\mathcal{L}^n(E) = 0$. Then $\mathcal{H}^n(E) = 0$ (Theorem 2.6) and Theorem 3.2 tells us that

$$\nu(E) = \mathcal{H}^n\big(L(E)\big) \leq \big(\mathrm{Lip}(L)\big)^n\,\mathcal{H}^n(E) = 0.$$

which concludes our assertion.

Notice that;

$$D_{\mathcal{L}^n}\nu(x) = \lim_{r\to 0} \frac{\mathcal{H}^n\big(L(B(x,r))\big)}{\mathcal{L}^n\big(B(x,r)\big)} = [\![\, L \,]\!]\,.$$

Thus, for all Borel sets $B \subseteq \mathbb{R}^n$, Theorem 1.18 implies that

$$\nu(B) = \int_B D_{\mathcal{L}^n}\nu(x)\, d\mathcal{L}^n(x) = \int_B [\![\, L \,]\!]\; d\mathcal{L}^n = [\![\, L \,]\!]\,\mathcal{L}^n(B)$$

Since both ν and $[\![\, L \,]\!]\mathcal{L}^n$ are Radon measures, which coincide on Borel sets, we get the desired equality

$$\mathcal{H}^n\big(L(A)\big) = [\![\, L \,]\!]\,\mathcal{L}^n(A)$$

for all $A \subseteq \mathbb{R}^n$. □

REMARK. For the last argument in the proof earlier, we used a small **Proposition** from Measure Theory, which states that;

Proposition 4.1. *Two Borel-regular measures coincide on $\mathbb{R}^n$, provided that they do so on all Borel subsets of $\mathbb{R}^n$.*

Proof. Let μ_1 and μ_2 be Borel-regular measures on $\mathbb{R}^n$ and $A \subseteq \mathbb{R}^n$. There exists a Borel set $B \subseteq \mathbb{R}^n$, $B \supseteq A$ for which $\mu_1(B) = \mu_1(A)$. Then $\mu_1(A) = \mu_1(B) = \mu_2(B) \geq \mu_2(A)$. In a similar way, there exists a Borel set $\tilde{B} \subseteq \mathbb{R}^n$, $\tilde{B} \supseteq A$ for which $\mu_2(\tilde{B}) = \mu_2(A)$. Thus, $\mu_2(A) = \mu_2(\tilde{B}) = \mu_1(\tilde{B}) \geq \mu_1(A)$. Hence $\mu_1(A) = \mu_2(A)$, for all $A \subseteq \mathbb{R}^n$. □

Lemma 4.3.
Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz function and $A \subseteq \mathbb{R}^n$ a $\mathcal{L}^n$-measurable set. Then

1. *$f(A)$ is $\mathcal{H}^n$-measurable,*
2. *the mapping $y \mapsto \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)$ is $\mathcal{H}^n$-measurable on $\mathbb{R}^m$, and*
3. $$\int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n \leq (\mathrm{Lip}(f))^n \mathcal{L}^n(A).$$

REMARK. The mapping $y \mapsto \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)$ is often referred as the **multiplicity function**.

Proof. Without loss of generality, we may assume that A is bounded. Generality can be achieved, eventually, by "gluing" together "copies" of the basic case. From Theorem 1.6, there exist compact sets $K_i \subseteq A$, $(i = 1, 2, ...)$ such that

$$\mathcal{L}^n(K_i) \geq \mathcal{L}^n(A) - \frac{1}{i}.$$

Since $\mathcal{L}^n(A) < \infty$ and A is $\mathcal{L}^n$-measurable, we get that $\mathcal{L}^n(A \setminus K_i) \leq \dfrac{1}{i}$, thus

$$\mathcal{L}^n\left(A \setminus \bigcup_{i=1}^{\infty} K_i\right) = 0.$$

Moreover, since f is continuous, $f(K_i)$ is compact and thus $\mathcal{H}^n$-measurable. Hence $f\left(\bigcup_{i=1}^{\infty} K_i\right) = \bigcup_{i=1}^{\infty} f(K_i)$ is $\mathcal{H}^n$-measurable, and so

$$\mathcal{H}^n\left(f(A)\setminus f\left(\bigcup_{i=1}^{\infty}K_i\right)\right)\leq \mathcal{H}^n\left(f\left(A\setminus\bigcup_{i=1}^{\infty}K_i\right)\right)$$
$$\leq (\mathrm{Lip}(f))^n\mathcal{L}^n\left(A\setminus\bigcup_{i=1}^{\infty}K_i\right)=0.$$

Thus $f(A)$ is $\mathcal{H}^n$-measurable, and this proves (*1.*)

(*2.*) For $k = 1, 2, \dots$ we define sets

$$B_k \coloneqq \left\{Q \;\middle|\; Q = \bigtimes_{i=1}^{n}\left(\frac{c_i}{2^k}, \frac{c_i+1}{2^k}\right],\ c_i \in \mathbb{Z}\right\},$$

i.e. the collection of half-open/closed dyadic cubes of $\mathbb{R}^n$. We immediately notice that each B_k contains countably many cubes, and so, we can adopt an "enumeration" $B_k = \{(Q_i)_{i\in\mathbb{N}} \mid Q_i \text{ as described above}\}$ and follow it whenever necessary. Note also that for a fixed k, any cube $Q^{(k)} \in B_k$ can be "decomposed" as;

$$Q^{(k)} = \bigcup_{i=1}^{2^n} Q_i^{(k+1)},\ \text{ with } Q_i^{(k+1)} \in B_{k+1},$$

and, that

$$\mathbb{R}^n = \bigcup B_k = \bigcup_{Q_i\in B_k} Q_i,$$

where the unions above are disjoint.

Now, we define functions $g_k : \mathbb{R}^m \to \mathbb{R}$

$$g_k(y) \coloneqq \sum_{i\in\mathbb{N}} \chi_{f(A\cap Q_i^k)}(y).$$

At first, we notice that (*1.*) ensures the $\mathcal{H}^n$-measurability of all g_k functions. Therefore, we shall dive deeper and explore their properties.

An keen observer notices immediately that g_k acts like an "enumerator", meaning that, for $y \in \mathbb{R}^m$,

$$g_k(y) = \text{number of cubes } Q \in B_k \text{ such that } f^{-1}\{y\} \cap (A \cap Q) \neq \varnothing$$

`Claim 1:` $(g_k)_{k\in\mathbb{N}}$ is a point-wise increasing sequence.

Indeed, fix an index k; For every $y \in \mathbb{R}^m$, we have that

$$\begin{aligned} g_k(y) &= \# \left\{ Q \in B_k \ : \ Q \cap (f^{-1}\{y\} \cap A) \neq \varnothing \right\} \\ &\leq \# \left\{ Q_i \in B_{k+1} \ : \ \left(\bigcup_{i=1}^{2^n} Q_i \right) \cap (f^{-1}\{y\} \cap A) \neq \varnothing \right\} \\ &\leq \# \left\{ Q' \in B_{k+1} \ : \ Q' \cap (f^{-1}\{y\} \cap A) \neq \varnothing \right\} = g_{k+1}(y). \end{aligned}$$

`Claim 2:` $g_k(y) \leq \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)$ for all $y \in \mathbb{R}^m$.

Let $y \in f(A \cap Q_i^k)$. Then there exists $x \in A \cap Q_i^k$ such that $f(x) = y$. This implies that $x \in A \cap Q_i^k \cap f^{-1}\{y\}$.

On the other hand, for $y \notin f(A \cap Q_i^k)$, we get $\chi_{f(A \cap Q_i^k)}(y) = 0$ and $A \cap Q_i^k \cap f^{-1}\{y\} = \varnothing$. Hence;

$$\mathcal{H}^0\big(A \cap f^{-1}\{y\} \cap Q_i^k\big) \geq \chi_{f(A \cap Q_i^k)}(y).$$

Therefore, we get that

$$\begin{aligned} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big) = \mathcal{H}^0\left(\bigcup_{i \in \mathbb{N}} A \cap f^{-1}\{y\} \cap Q_i^k \right) &= \sum_{i \in \mathbb{N}} \mathcal{H}^0\big(A \cap f^{-1}\{y\} \cap Q_i^k\big) \\ &\geq \sum_{i \in \mathbb{N}} \chi_{f(A \cap Q_i^k)}(y) = g_k(y). \end{aligned}$$

`Claim 3:` $\mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)$ is the point-wise supremum of $g_k(y)$ for all $y \in \mathbb{R}^m$. We will demonstrate that; For all $y \in \mathbb{R}^m$ and for all $M \in \mathbb{N}$ such that $M \leq \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)$, there exists $k \in \mathbb{N}$ such that $g_k(y) \geq M$.

Indeed; Since $\mathcal{H}^0\big(A \cap f^{-1}\{y\}\big) \geq M$, we can find M distinct points $x_1, ..., x_M \in A \cap f^{-1}\{y\}$. Take k large enough, such that $\left\| x_p - x_{p'} \right\| > \dfrac{\sqrt{n}}{2^k}$, for all indices $1 \leq p < p' \leq M$. Since the cubes which are contained in B_k are disjoint and have a diameter of $\dfrac{\sqrt{n}}{2^k}$, each point x_p is contained in exactly one cube, for all $1 \leq p \leq M$. Let us denote that cube as $Q_{i(p)}^k$, where the indicatrix $p \overset{i}{\mapsto} i(p)$ is an $1-1$ map. Consequently;

$$g_k(y) = \sum_{i \in \mathbb{N}} \chi_{f(A \cap Q_i^k)}(y) \geq \sum_{1 \leq p \leq M} \chi_{f(A \cap Q_{i(p)}^k)} = M.$$

Consequentially, we have obtained that; As $k \to \infty$,

$$g_k(y) \to \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)$$

for each $y \in \mathbb{R}^m$; and so $y \mapsto \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)$ is $\mathcal{H}^n$-measurable, as the limit of $\mathcal{H}^n$-measurable maps.

(*3.*) From the Monotone Convergence Theorem, we get that

$$\begin{aligned}
\int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n &= \lim_{k\to\infty} \int_{\mathbb{R}^m} g_k \, d\mathcal{H}^n \\
&= \lim_{k\to\infty} \int_{\mathbb{R}^m} \sum_{i\in\mathbb{N}} \chi_{f(A\cap Q_i^k)}(y)\, d\mathcal{H}^n \\
&= \lim_{k\to\infty} \sum_{i\in\mathbb{N}} \int_{\mathbb{R}^m} \chi_{f(A\cap Q_i^k)}(y)\, d\mathcal{H}^n \\
&= \lim_{k\to\infty} \sum_{i\in\mathbb{N}} \mathcal{H}^n\big(f(A\cap Q_i^k)\big) \\
&\leq \limsup_{k\to\infty} \sum_{i\in\mathbb{N}} (\mathrm{Lip}(f))^n \mathcal{L}^n(A\cap Q_i^k) \\
&= (\mathrm{Lip}(f))^n \mathcal{L}^n(A).
\end{aligned}$$

□

REMARK.

From (*3.*) we deduce that $f^{-1}\{y\}$ is at-most countable for $\mathcal{H}^n$-a.e. $y \in \mathbb{R}^m$.

Proof. Fix any compact set $K \subseteq \mathbb{R}^n$. Then, K is closed and bounded, and so is its image under f (since f is continuous, it preserves compactness). Hence by Assertion (*3.*) we get that

$$\int_{f(K)} \mathcal{H}^0\big(K \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n \leq \int_{\mathbb{R}^m} \mathcal{H}^0\big(K \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n \leq \mathrm{Lip}(f)^n \mathcal{L}^n(K) < \infty$$

Consequentially, for $\mathcal{H}^n$-a.e. $y \in f(K)$ we get that

$$\mathcal{H}^0\big(K \cap f^{-1}\{y\}\big) < \infty,$$

since,otherwise,the multiplicity function would take infinite values for a set of positive measure,and thus the aforementioned integral would not be finite. Hence, the set $K \cap f^{-1}\{y\}$ contains finitely many elements for $\mathcal{H}^n$-a.e.$y \in \mathbb{R}^m$.

The final step consists of exhausting $\mathbb{R}^n$ with an increasing union of compact sets. Then, each intersection with $f^{-1}\{y\}$ will be the empty set or a set containing a finite number of elements. Elementary results from Set Theory and Measure Theory imply that the union of such sets is at-most countable and has Lebesgue measure zero. This concludes the proof of the remark. □

The next Lemma we are about to present plays an important role in the proof of both Area and Coarea formula.

The brilliant idea presented here, introduced initially by Federer in [10], is that we can utilise linear automorphisms in order to "approximate" - in a sense - a Lipschtz map, the same way we do in fundamental Calculus, with linear functions and C^1 maps, where the continuity of the gradient is employed, so as to deduce that the latter are locally constant.

Finaly, we shall state that, for reasons still to be clarified, the following Lemma, along with its many congener results, are generally known as **Linearisation Lemmas** for Lipschitz maps.

REMARK. A last Remark before proceding to the Lemma, of Algebraic & Computational nature. Given a $(n \times n)$-matrix L and considering the Operator norm on the induced linear map, i.e. taking $\|L\|_{op} \coloneqq \sup\{\|Lx\| : \|x\| = 1\}$ we observe that;

$$\|Lx\| = \|L(x_1e_1 + \cdots + x_ne_n)\| = \|x_1Le_1 + \cdots + x_nLe_n\| \\ \leq |x_1| \, \|Le_1\| + \cdots + |x_n| \, \|Le_n\| \leq \sum_{j=1}^{n} \|Le_j\| \, .$$

Therefore, given a matrix L we have an estimate of the "size" of its Operator norm via its columns, given as;

$$\|L\|_{op} \leq \sum_{j=1}^{n} \|Le_j\| \, .$$

Without further a do, we proceed to the Lemma.

Lemma 4.4. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz function and $t > 1$. Define*

$$B \coloneqq \{x \mid Df(x) : \text{exists}, \ \textit{and}, \ Jf(x) > 0\}$$

Then there exists a countable collection $\{E_k\}_{k=1}^{\infty}$ of Borel subsets of $\mathbb{R}^n$ such that

1. *$B = \bigcup_{k=1}^{\infty} E_k$,*
2. *$f|_{E_k}$ is one-to-one $(k = 1, 2, \ldots)$, and*
3. *for each $k = 1, 2, \ldots$ there exists a symmetric automorphism $T_k : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$\operatorname{Lip}((f|_{E_k}) \circ T_k^{-1}) \leq t, \ \operatorname{Lip}(T_k \circ (f|_{E_k})^{-1}) \leq t, \\ t^{-n} |\det T_k| \leq Jf|_{E_k} \leq t^n |\det T_k|.$$

Proof. (*1.*) Fix $\varepsilon > 0$ such that

$$\frac{1}{t} + \varepsilon < 1 < t - \varepsilon.$$

Since $\mathbb{R}^n$ is separable, take $\mathcal{C}$ to be a countable dense subset of B. Now consider the space of symmetric automorphisms of $\mathbb{R}^n$. We endow the space with the operator norm. We will construct a countable dense subset, as follows;

Let $S = (s_{ij})$ be a symmetric automorphism of $\mathbb{R}^n$. We define the symmetric matrix $S^{(1)}$, such that $q_1 \coloneqq (S^{(1)})_{1\,1} \in \mathbb{Q}$, keeping all other entries the same as in S. Due to the Rationals being dense in $\mathbb{R}$, we can choose a suitable q_1 such that $\det S^{(1)} \neq 0$ and

$$\left\|S - S^{(1)}\right\|_{op} \leq \sum_{j=1}^{n} \left\|(S - S^{(1)})e_j\right\| = |s_{1,1} - q_1| < \varepsilon.$$

We repeat the process, with $S^{(1)}$ in place of S, meaning that; We induce a symmetric matrix $S^{(2)}$ such that $q_2 \coloneqq (S^{(2)})_{1\,2} = (S^{(2)})_{2\,1} \in \mathbb{Q}$, keeping all other entries the same as in $S^{(1)}$. Again, we shall choose a suitable q_2, in order to ensure that $\det S^{(2)} \neq 0$ and $\left\|S^{(1)} - S^{(2)}\right\|_{op} < \varepsilon$. Finally, after a finite number of steps, we will have ended up with a symmetric matrix S' consisting of rational entries, for which $\det S' \neq 0$, and such that

$$\left\|S' - S\right\|_{op} < \varepsilon.$$

Gathering all such matrices, we end up with a countable subset, let's call it $\mathcal{S}$, of symmetric automorphism of $\mathbb{R}^n$, which is dense in the Operator norm.

For each $c \in \mathcal{C}, T \in \mathcal{S}$ and $i = 1, 2, ...$, we define set $\boldsymbol{E(c,T,i)}$ to be the set of all $b \in B \cap B\left(c, \frac{1}{i}\right)$ satisfying

$$\left(\frac{1}{t} + \varepsilon\right)|Tu| \leq |Df(b)u| \leq (t - \varepsilon)|Tu| \tag{$\star$}$$

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

$$|f(\alpha) - f(b) - Df(b) \cdot (\alpha - b)| \leq \varepsilon |T(\alpha - b)| \tag{$\star\star$}$$

for all $\alpha \in B(b, \frac{2}{i})$.
Note that $E(c, T, i)$ is a Borel set, since Df is Borel measurable.

From $(\star)$ and $(\star\star)$ follows that

$$\begin{aligned}|f(\alpha)-f(b)| &\leq |f(\alpha)-f(b)-Df(b)\cdot(\alpha-b)|+|Df(b)\cdot(\alpha-b)| \\ &\underset{(\star\star)}{\overset{(\star)}{\leq}} \varepsilon|T(\alpha-b)|+(t-\varepsilon)|T(\alpha-b)| \\ &= t|T(\alpha-b)|.\end{aligned}$$

In a similar way, using the so-called "reverse" triangular inequality, we get that

$$|f(\alpha)-f(b)| \geq t^{-1}|T(\alpha-b)|.$$

Hence, we have the estimate

$$t^{-1}|T(\alpha-b)| \leq |f(\alpha)-f(b)| \leq t|T(\alpha-b)| \qquad (\star\star\star)$$

for $b \in E(c,T,i), \alpha \in B(b,\frac{2}{i})$.

`Claim:` If $b \in E(c,T,i)$, then

$$(t^{-1}+\varepsilon)^n|\det T\,| \leq Jf(b) \leq (t-\varepsilon)^n|\det T\,|.$$

Proof of claim:
By the Decomposition Theorem, we have that $Df(b)=L=O\circ S$, and so

$$Jf(b)=[\![\,Df(b)\,]\!]=|\det S\,|.$$

According to $(\star)$, we have that

$$\left(\frac{1}{t}+\varepsilon\right)|Tu| \leq |(O\circ S)u|=|Su| \leq (t-\varepsilon)|Tu|$$

for $u \in \mathbb{R}^n$, and so, by setting $Tu=v$ and again renaming the result back to u-notation, we have that

$$\left(\frac{1}{t}+\varepsilon\right)|u| \leq |(S\circ T^{-1})u| \leq (t-\varepsilon)|u| \ \ (u \in \mathbb{R}^n).$$

Thus

$$(S\circ T^{-1})\big(B(1)\big) \subseteq B(t-\varepsilon),$$

and so, passing onto Lebesgue measures, we get

$$\mathcal{L}^n\big((S\circ T^{-1})(B(1))\big)=|\det(S\circ T^{-1})\,|\,\alpha(n)$$
$$\text{and}\quad \mathcal{L}^n(B(t-\varepsilon))=\alpha(n)(t-\varepsilon)^n$$

and so,

$$|\det(S \circ T^{-1})| \le (t - \varepsilon)^n.$$

Hence,

$$|\det S| \le (t - \varepsilon)^n |\det T|.$$

The proof of the other inequality follows in a similar way.

Now, we will "re-brand" our collection of

$$\big\{E(c, T, i) \mid c \in \mathcal{C},\ T \in \mathcal{S},\ i \in \mathbb{N}\big\}\ \ as\ \ \{E_k\}_{k=1}^{\infty}.$$

We want to show that; If $b \in B$, then $b \in \bigcup_{k=1}^{\infty} E_k$. We turn our attention once again to the Polar Decomposition Theorem; We have that $Df(b) = O \circ S$, for a symmetric map $S : \mathbb{R}^n \to \mathbb{R}^n$ and an orthogonal map $O : \mathbb{R}^n \to \mathbb{R}^m$. Since $b \in B$, we can deduce with ease that S is invertible;

Had we had S being non-invertible, we would have $|\det S| = 0$ and so $Jf(b) = 0$, which is a contardiction to the definition of the set B. Furtheremore, S is an epimorphism; Otherwise, $Im(S)$ would be a proper subspace of $\mathbb{R}^n$, therefore S would not be invertible. Consequently, $S : \mathbb{R}^n \to \mathbb{R}^n$ is a symmetric automorphism of $\mathbb{R}^n$.
From the density of $\mathcal{S}$, we can find a suitable $T \in \mathcal{S}$ such that

$$\mathrm{Lip}(T \circ S^{-1}) \le (t^{-1} + \varepsilon)^{-1} \text{ and } \mathrm{Lip}(S \circ T^{-1}) \le t - \varepsilon.$$

Indeed; Since S is a symmetric automorphism, then for any $\epsilon > 0$ there exists $T \in \mathcal{S}$ such that $\|T - S\| < \epsilon$. Thus, $\dfrac{|(T - S)(x)|}{|x|} < \epsilon, x \neq \vec{0}$, and so;

$$|Tx - Sx| < \epsilon |x|.$$

Substituting $x = S^{-1}y$ gives

$$|T(S^{-1}y) - S(S^{-1}y)| < \epsilon |S^{-1}y|,$$

thus

$$|(T \circ S^{-1})(y) - y| < \epsilon \left\|S^{-1}\right\| |y|,$$

which implies

$$\frac{|(T \circ S^{-1} - I)(y)|}{|y|} < \epsilon \left\|S^{-1}\right\|, \text{ for all } y \in \mathbb{R}^n,\ y \neq \vec{0}.$$

Therefore, we get
$$\left\|T \circ S^{-1} - I\right\| < \epsilon \left\|S^{-1}\right\|.$$
Furthermore, for any $x \in \mathbb{R}^n$ we have;
$$\begin{aligned}|(T \circ S^{-1})(x)| = |(T \circ S^{-1})(x) - x + x| &\leq |(T \circ S^{-1})(x) - x| + |x| \\ &= |(T \circ S^{-1} - I)(x)| + |x| \\ &\leq \left\|T \circ S^{-1} - I\right\| |x| + |x| \\ &< \epsilon \left\|S^{-1}\right\| |x| + |x| \\ &= (\epsilon \left\|S^{-1}\right\| + 1)|x|.\end{aligned}$$
Hence
$$\left\|T \circ S^{-1}\right\| < 1 + \epsilon \left\|S^{-1}\right\|,$$
which implies
$$\operatorname{Lip}(T \circ S^{-1}) \leq 1 + \epsilon \left\|S^{-1}\right\|.$$
We want
$$1 + \epsilon \left\|S^{-1}\right\| = \left(\frac{1}{t} + \varepsilon\right)^{-1},$$
or, equivalently
$$\left(\frac{1}{t} + \varepsilon\right)\left(1 + \epsilon \left\|S^{-1}\right\|\right) = 1.$$
Expanding on the terms and solving the equation at hand with respect to ϵ results in;
$$\epsilon = \frac{1 - \frac{1}{t} - \varepsilon}{\frac{\|S^{-1}\|}{t} + \varepsilon \left\|S^{-1}\right\| t} > 0.$$
Consequently, such a symmetric automorphism T exists, for the specific ϵ we have calculated above. For the other inequality, we simply mimic the calculations above.

Let $u \in \mathbb{R}^n$. We now have that
$$|(T \circ S^{-1})u| \leq \operatorname{Lip}(T \circ S^{-1})|u| \leq (t^{-1} + \varepsilon)^{-1}|u|,$$
and, by substituting $u = S(\tilde{u})$ and “re-naming” back to u-notation, we get
$$\left(\frac{1}{t} + \varepsilon\right)|Tu| \leq |Su| = |(O \circ S)u| = |Df(b)u|.$$
Moreover, we have that
$$|Df(b)u| = |(O \circ S)u| = |Su| = |(S \circ T^{-1})(Tu)|$$

$$= |(S \circ T^{-1})(Tu) - (S \circ T^{-1})(T\vec{0})|$$
$$\leq \text{Lip}(S \circ T^{-1})|Tu|$$
$$\leq (t - \varepsilon)|Tu|.$$

Hence, for all $u \in \mathbb{R}^n$ holds the following

$$\left(\frac{1}{t} + \varepsilon\right)|Tu| \leq |Df(b)u| \leq (t - \varepsilon)|Tu|.$$

Now, the density of $\mathcal{C}$ in B, allows us to select $c \in \mathcal{C}$, so that $|b - c| < \dfrac{1}{i}$, for i sufficiently large. At last, from the differentiability of f on b, we get that

$$\lim_{\alpha \to b} \frac{|f(\alpha) - f(b) - Df(b)(\alpha - b)|}{|\alpha - b|} = 0.$$

Hence, for $\dfrac{\varepsilon}{\text{Lip}(T^{-1})} > 0$, there exists $\delta > 0$, such that; For $|\alpha - b| < \delta$, we have

$$|f(\alpha) - f(b) - Df(b)(\alpha - b)| < \frac{\varepsilon}{\text{Lip}(T^{-1})}.$$

Thus, for any i such that $\frac{2}{i} < \delta$, we get that; For all $\alpha \in B(b, \frac{2}{i})$, holds

$$\begin{aligned}
|f(\alpha) - f(b) - Df(b)(\alpha - b)| &\leq \frac{\varepsilon}{\text{Lip}(T^{-1})}|\alpha - b| \\
&= \frac{\varepsilon}{\text{Lip}(T^{-1})}|T^{-1}(T\alpha) - T^{-1}(Tb)| \\
&\leq \frac{\varepsilon}{\text{Lip}(T^{-1})}\text{Lip}(T^{-1})|T\alpha - Tb| \\
&= \varepsilon|T(\alpha - b)|.
\end{aligned}$$

Thus $b \in E(c, T, i)$. Since this conclusion holds for all $b \in B$, we get

$$B \subseteq \bigcup_{k=1}^{\infty} E_k.$$

The reverse inclusion $\bigcup_{k=1}^{\infty} E_k \subseteq B$, is trivial, and follows directly from the definition of E_k (namely, of $E(c, T, i)$). Assertion (*1.*) is proved.

Assertion (*2.*) is trivial, considering $(\star\star\star)$.

Finally, take any set E_k, of the form $E(c, T, i)$, for some $c \in \mathcal{C}, T \in \mathcal{S}$ and $i = 1, 2, ...$ Take T_k in place of T on $(\star\star\star)$. Then, we have that

$$t^{-1}|T_k(\alpha - b)| \leq |f(\alpha) - f(b)| \leq t|T_k(\alpha - b)|$$

for all $b \in E_k$ and all $\alpha \in B(b, \frac{2}{i})$.

Notice that $E_k \subseteq B(c, \frac{1}{i})$, by definition, and that $B(c, \frac{1}{i}) \subseteq B(b, \frac{2}{i})$;
Let $z \in B(c, \frac{1}{i})$. Since $b \in E_k$, by definition, $b \in B \cap B(c, \frac{1}{i})$, hence $|b - c| < \frac{1}{i}$.
Thus; $|z - b| \leq |z - c| + |c - b| \leq \frac{1}{i} + \frac{1}{i} = \frac{2}{i}$.

Consequently $E_k \subseteq B(b, \frac{2}{i})$, and so;

$$t^{-1}|T_k(\alpha - b)| \leq |f(\alpha) - f(b)| \leq t|T_k(\alpha - b)| \tag{$\dagger$}$$

holds for all $\alpha, b \in E_k$. Letting $T_k\alpha = \widetilde{\alpha}$ and $T_k b = \widetilde{b}$, thus $\alpha = T_k^{-1}\widetilde{\alpha}$ and $b = T_k^{-1}\widetilde{b}$, gives

$$t^{-1}|\widetilde{\alpha} - \widetilde{b}| \leq |f(T_k^{-1}\widetilde{\alpha}) - f(T_k^{-1}\widetilde{b})| \leq t|\widetilde{\alpha} - \widetilde{b}|,$$

thus

$$t^{-1}|\widetilde{\alpha} - \widetilde{b}| \leq |(f \circ T_k^{-1})(\widetilde{\alpha}) - (f \circ T_k^{-1})(\widetilde{b})| \leq t|\widetilde{\alpha} - \widetilde{b}|.$$

Consequentially,

$$\text{Lip}((f|_{E_k}) \circ T_k^{-1}) \leq t.$$

In a similar way, from the left inequality of ($\dagger$), we have

$$t^{-1}|T_k(\alpha - b)| \leq |f(\alpha) - f(b)|.$$

Substituting $\widetilde{\alpha} = f(\alpha)$ and $\widetilde{b} = f(b)$, results in

$$|(T_k \circ f^{-1})(\widetilde{\alpha}) - (T_k \circ f^{-1})(\widetilde{b})| \leq t|\widetilde{\alpha} - \widetilde{b}|.$$

Hence,

$$\text{Lip}(T_k \circ (f|_{E_k})^{-1}) \leq t.$$

Finally, passing with limits on Claim, provides the estimate

$$t^{-n}|\det T_k| \leq Jf|_{E_k} \leq t^n|\det T_k|.$$

Assertion (*3.*) is proven. □

REMARK I. It is trivial to state that, we can "forge" the countable collection $\{E_k\}_{k=1}^{\infty}$ so that it consists of disjoint sets, without this affecting any one of our conclusions. Henceforward, we will impose this contention, without further justification.

REMARK II. We have demonstrated, essentially, that; For a Lipschitz map $f : \mathbb{R}^n \to \mathbb{R}^m \, (n \leq m)$ the set $\{Jf > 0\}$ can be partioned into a countable familly of Borel sets $\{E_k\}_{k=1}^{\infty}$, so that the restriction of f to each and every one of them is an injection. Furthermore, by choosing a parameter of approximation $t > 1$, we acquired an even stronger result; There exists a countable collection of linear automorphisms $T_k : \mathbb{R}^n \to \mathbb{R}^n$ such that $f|_{E_k} \circ T_k^{-1}$ is *almost* an isometry of $\mathbb{R}^n$ into $\mathbb{R}^m$. To this we own the appellation "**Linearisation**", which seemed rather "arbitrary" at first, to say the least.

REMARK III. Before proceeding any further, it is important to state a final direct consequence of the Linearisation Lemma. It is immediate that, uppon passing with limits on $(\star)$, we effectively acquire that;
For all $x \in E_k$, we have;

$$t^{-1}|Tu| \leq |Df(x)u| \leq t|Tu|$$

for all $u \in \mathbb{R}^n$. Therefore, by means of a simple substitution, we get;

$$t^{-1}|u| \leq |Df(x) \circ T^{-1}u| \leq t|u| \ \ (u \in \mathbb{R}^n).$$

Hence

$$\left\| Df(x) \circ T^{-1} \right\| \leq t.$$

At last, in the same spirit, since $x \in B$, we get that

$$\left\| T \circ Df(x)^{-1} \right\| \leq t.$$

4.2 The Area formula

In Geometric Measure Theory, the **Area formula** provides an interesting relation between the Jacobian integral (the integral of the jacobian) of a Lipschitz map over some suitable set and the **n-dimensional Hausdorff area**, namely the $\mathcal{H}^n$-integral of the multiplicity function, also referred as the $\mathcal{H}^n$-measure of the image $f(A)$ counted with multiplicity.

Theorem 4.1 (**Area formula**)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous. Then for each $\mathcal{L}^n$-measurable subset $A \subseteq \mathbb{R}^n$,*

$$\int_A Jf\ dx = \int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n(y).$$

Proof. In view of Rademacher's Theorem, we may as well assume $Df(x)$ and $Jf(x)$ exist for all $x \in A$. Also, without loss of generality, we will suppose $\mathcal{L}^n(A) < \infty$.

`Case 1:` $A \subseteq \{Jf > 0\}$.

Fix $t > 1$ and choose a collection of disjoint Borel sets $\{E_k\}_{k=1}^{\infty}$ such as in Lemma 4.4. Similarly, we define B_k as in Lemma 4.3 and consider sets

$$F_j^i = E_j \cap Q_i \cap A \quad (\,Q_i \coloneqq Q_i^k \in B_k \text{ and } i, j = 1, 2, \dots).$$

Immediately we see that the sets F_j^i are disjoint and their union decomposes A since

$$\begin{aligned}\bigcup_{i,j=1}^{\infty} F_j^i = \bigcup_{i,j=1}^{\infty} \big(E_j \cap Q_i \cap A\big) &= A \cap \left(\bigcup_{i,j=1}^{\infty} (E_j \cap Q_i)\right) \\ &= A \cap \left(\bigcup_{j=1}^{\infty} E_j \cap \bigcup_{i=1}^{\infty} Q_i\right) \\ &= A \cap \big(\{Jf > 0\} \cap \mathbb{R}^n\big) = A.\end{aligned}$$

Using Lemma 4.4 and Theorem 3.2 we deduce that

$$\begin{aligned}\mathcal{H}^n\big(f(F_j^i)\big) = \mathcal{H}^n\big(f|_{E_j} \circ T_j^{-1} \circ T_j(F_j^i)\big) &= \mathcal{H}^n\big(\big(f|_{E_j} \circ T_j^{-1}\big)\ T_j(F_j^i)\big) \\ &\le \big(\mathrm{Lip}(f|_{E_j} \circ T_j^{-1})\big)^n\, \mathcal{H}^n\big(T_j(F_j^i)\big) \\ &\le t^n \mathcal{H}^n\big(T_j(F_j^i)\big)\end{aligned}$$

and

$$\mathcal{L}^n\big(T_j(F_j^i)\big) = \mathcal{H}^n\big(T_j(F_j^i)\big) = \mathcal{H}^n\big(T_j \circ (f|_{E_j})^{-1} \circ f(F_j^i)\big) \leq t^n \mathcal{H}^n\big(f(F_j^i)\big).$$

Therefore, we get the following estimation

$$t^{-2n}\, \mathcal{H}^n\big(f(F_j^i)\big) \leq t^{-n}\mathcal{L}^n\big(T_j(F_j^i)\big) = t^{-n}|\det T_j\,|\,\mathcal{L}^n(F_j^i) \leq \int_{F_j^i} Jf\,dx$$

$$\text{and} \quad \int_{F_j^i} Jf\,dx \leq t^n|\det T_j\,|\,\mathcal{L}^n(F_j^i) = t^n\mathcal{L}^n\big(T_j(F_j^i)\big) \leq t^{2n}\,\mathcal{H}^n\big(f(F_j^i)\big),$$

i.e.

$$t^{-2n}\mathcal{H}^n\big(f(F_j^i)\big) \leq \int_{F_j^i} Jf\,dx \leq t^{2n}\mathcal{H}^n\big(f(F_j^i)\big).$$

Now, summing on i and j, and taking advantage of the decomposition of A, we get that

$$t^{-2n}\sum_{i,j=1}^{\infty} \mathcal{H}^n\big(f(F_j^i)\big) \leq \int_A Jf\,dx \leq t^{2n}\sum_{i,j=1}^{\infty} \mathcal{H}^n\big(f(F_j^i)\big). \qquad (\star)$$

`Claim 1:`

$$\lim_{k\to\infty}\sum_{i,j=1}^{\infty} \mathcal{H}^n\big(f(F_j^i)\big) = \int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\,d\mathcal{H}^n(y).$$

Proof of claim: Let us define functions $g_k : \mathbb{R}^m \to \mathbb{R}$ as

$$g_k \coloneqq \sum_{i,j=1}^{\infty} \chi_{f(F_j^i)} \ \ (k = 1, 2, ...).$$

In Lemma 4.3 we have established the $\mathcal{H}^n$-measurability of g_k. Moreover, we see that $g_k(y)$, for $y \in \mathbb{R}^m$, acts like an "enumerator", counting the number of F_j^i sets, for which $F_j^i \cap f^{-1}\{y\} \neq \varnothing$.

Since $f|_{E_j}$ is one-to-one, from Lemma 4.4, this holds true for $f|_{F_j^i}$ as well. Hence $f(F_j^i) = f(E_j \cap Q_i \cap A) = f(E_j) \cap f(Q_i) \cap f(A)$. Moreover, we notice that $A \subseteq \{Jf(x) > 0\}$ implies that $f(A) \subseteq \bigcup_{j=1}^{\infty} f(E_j)$, also a consequence of Lemma 4.4.

Finally, a keen observer notices that our definition of g_k closely resembles the one in Lemma 4.3. Therefore, if we mimick our previous work, we obtain that; As $k \to \infty$,

$$g_k(y) \to \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big).$$

Now, from the Monotone Convergence Theorem follows that

$$\begin{aligned}
\lim_{k\to\infty} \sum_{i,j=1}^{\infty} \mathcal{H}^n\big(f(F_j^i)\big) &= \lim_{k\to\infty} \int \sum_{i,j=1}^{\infty} \chi_{f(F_j^i)} \, d\mathcal{H}^n \\
&= \lim_{k\to\infty} \int g_k \, d\mathcal{H}^n \\
&= \int \lim_{k\to\infty} g_k \, d\mathcal{H}^n \\
&= \int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big) \, d\mathcal{H}^n(y).
\end{aligned}$$

Taking limits as $k \to \infty$ in $(\star)$ and making use of Claim 1, we get that

$$t^{-2n}\int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big) \, d\mathcal{H}^n(y) \le \int_A Jf \, dx \le t^{2n}\int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big) \, d\mathcal{H}^n(y)$$

Sending $t \to 1^+$ concludes the proof of Case 1.

`Case 2:` $A \subseteq \{Jf = 0\}$.

Fix $0 < \varepsilon \le 1$. We make use of the following expression for our function f:

$$f = p \circ g,$$

where $g : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^n$ is the mapping

$$g(x) \coloneqq (f(x), \varepsilon x),$$

and, $p : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ is the projection in the first argument, i.e.

$$p(y, z) = y.$$

`Claim 2:` There exists a constant C such that

$$0 < Jg(x) \le C\varepsilon$$

for $x \in A$.

Proof of claim:
Writing down g analytically, we get $g = (f^1, ..., f^m, \varepsilon x_1, ..., \varepsilon x_n)$. Hence

$$Dg(x) = \begin{pmatrix} Df(x) \\ \varepsilon I_n \end{pmatrix}_{(n+m)\times n}$$

Since $Jg(x)^2$ equals the sum of the squares of the $(n \times n)$-subdeterminants of $Dg(x)$, due to the Binet-Cauchy formula, we have that

$$Jg(x)^2 \geq \varepsilon^{2n} > 0.$$

For the upper estimate we will need a little more effort;

First, we notice that the first m rows of the $Dg(x)$ matrix are simply $\nabla f^i(x)$. Hence, we get that

$$\|\nabla f^i(x)\| = \|Df^i(x)\| \leq \sqrt{n}\,\mathrm{Lip}(f^i) \leq \sqrt{n}\,\mathrm{Lip}(f) \coloneqq \boldsymbol{\vartheta}.$$

Furthermore, using again the Binet-Cauchy formula, we compute that

$$Jg(x)^2 = Jf(x)^2 + \left\{\begin{matrix} \text{sum of squares of n}-\text{dimensional sub}-\text{determinants,} \\ \text{of matrices having at}-\text{least one row in }\varepsilon\mathrm{I_n} \end{matrix}\right\}$$

Since $0 < \varepsilon \leq 1$ and the rows of $Df(x)$ are bounded in norm by $\boldsymbol{\vartheta}$, each minor i.e. $(n \times n)$-subdeterminant of the latter type is bounded by $\varepsilon \cdot \max(1, \boldsymbol{\vartheta}^{n-1})$, via Hadamard's inequality (Theorem 3.11). Upon careful consideration, since we have already taken into account all those minors forming the $Jf(x)^2$, we are left with $\binom{n+m}{n} - \binom{m}{n}$ summands.
Hence, for each $x \in A \subseteq \{Jf = 0\}$, we get

$$Jg(x)^2 \leq Jf(x)^2 + \left(\binom{n+m}{n} - \binom{m}{n}\right)\varepsilon^2 \cdot \max(1, \boldsymbol{\vartheta}^{n-1})^2.$$

Therefore, we end up with $Jg(x) \leq C\varepsilon$, where

$$C = \sqrt{\binom{n+m}{n} - \binom{m}{n}}\max(1, \boldsymbol{\vartheta}^{n-1}).$$

This concludes the proof of our claim.

Now, recall that $p : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ is a projection. Thus, we may employ what we obtained in Case 1 above, in order to get

$$\begin{aligned}\mathcal{H}^n(f(A)) &\le \mathcal{H}^n(g(A)) \\ &= \int_{g(A)} d\mathcal{H}^n(y,z) \\ &\le \int_{\mathbb{R}^{n+m}} \mathcal{H}^0\big(A \cap g^{-1}\{y,z\}\big)\, d\mathcal{H}^n(y,z) \\ &= \int_A Jg(x)\, dx \\ &\le \varepsilon C \mathcal{L}^n(A) < \infty.\end{aligned}$$

Letting $\varepsilon \to 0$, we obtain that $\mathcal{H}^n(f(A)) = 0$.
Moreover, since $supp\big\{\, \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\big\} \subseteq f(A)$, we conclude that

$$\int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n(y) = 0.$$

Consequentially,

$$\int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n(y) = 0 = \int_A Jf\; dx.$$

`Case 3:` $A \subseteq \{Jf \ge 0\}$ for every $x \in A$.

In the general case, we write $A = A_1 \cup A_2$, with $A_1 \subseteq \{Jf > 0\}$ and $A_1 \subseteq \{Jf = 0\}$ and employ Cases 1 and 2 as above. □

The role of the Multiplicity function

Although we have gone through a detailed and analytical proof of the Area Formula, and we have established its validity, a question might still linger on the exact purpose of the Multiplicity function as an integrand. We target this question with the following example.

Definition. A Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called a **local isometry**, provided that $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ is an orthogonal map for a.e. $x \in \mathbb{R}^n$.

Remark. This definition is in full accord with the "classical" bibliographic definition of locality using differentials, since the differential of a linear map is its induced matrix.

Thus, for a local isometry, we calculate that

$$Jf(x) = \llbracket Df(x) \rrbracket = \det\left(Df(x)^* \circ Df(x)\right) = \left(\det Df(x)\right)^2 = 1^2 = 1.$$

In this case, the left-hand side of the Area Formula (Theorem 4.1) is simply $\mathcal{L}^n(A)$. Now, if we make the assumption that our local isometry is also *injective*, then we get that $\mathcal{H}^0\big(A \cap f^{-1}\{y\}\big) = 1$ on the image of f and zero elsewhere, and so,

$$\mathcal{L}^n(A) = \int_{\mathbb{R}^m} 1 \, d\mathcal{H}^n(y) = \mathcal{H}^n(f(A)).$$

Therefore, for an injective local isometry, we ened up with;

$$\mathcal{H}^n(f(A)) = \mathcal{L}^n(A).$$

Note that, local isometries are not injective in general. Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$, defined by

$$f(x_1, x_2) = \begin{cases} (x_1, x_2) & ,\text{if } x_1 > 0 \\ (-x_1, x_2) & ,\text{if } x_1 \leq 0. \end{cases}$$

It is immediate that f is a local isometry. Therefore, by taking the open cube $Q = (-1, 1)^2$ to be our "test-subset", we get that $\mathcal{L}^2(Q) = 4$, yet,

$$\mathcal{H}^2(f(Q)) = 2 \neq 4 = \mathcal{L}^2(Q) = \int_Q Jf.$$

This is the case, evidently, because f, as a map, folds $\mathbb{R}^2$ onto $\{x_1 \geq 0\}$, having $\mathcal{H}^0\big(f^{-1}(\{x\})\big) = 2$ for all $x = (x_1, x_2) \in \mathbb{R}^2$ for which $x_1 > 0$.

Conclusion. Therefore, to answer our question, Multiplicity function "emerges" in a natural way, and it is there so as to compensate for "overlap effects" in the image of our function.

Theorem 4.2 (Change of Variables). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz function. Then for each $\mathcal{L}^n$-summable function $g : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\int_{\mathbb{R}^n} g(x)Jf(x)\,dx = \int_{\mathbb{R}^m} \left[\sum_{x\in f^{-1}\{y\}} g(x) \right] d\mathcal{H}^n(y).$$

Proof. We will proceed in steps.

`Case 1:` $g \geq 0$. We recall that for such a function g, from Theorem 1.10 stems the following expression

$$g = \sum_{i=1}^{\infty} \frac{1}{i}\chi_{A_i}$$

for appropriate $\mathcal{L}^n$-measurable sets $\{A_i\}_{i=1}^{\infty}$. Employing the Monotone Convergence Theorem and the Area formula, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} g(x)Jf(x)\,dx &= \int_{\mathbb{R}^n} \left(\sum_{i=1}^{\infty} \frac{1}{i}\chi_{Ai}(x) \right) Jf(x)\,dx = \\
&= \sum_{i=1}^{\infty} \frac{1}{i} \int_{\mathbb{R}^n} \chi_{A_i} Jf\,dx \\
&= \sum_{i=1}^{\infty} \frac{1}{i} \int_{A_i} Jf\,dx \\
&= \sum_{i=1}^{\infty} \frac{1}{i} \int_{\mathbb{R}^m} \mathcal{H}^0\big(A_i \cap f^{-1}\{y\}\big)\,d\mathcal{H}^n(y) \\
&= \int_{\mathbb{R}^m} \sum_{i=1}^{\infty} \frac{1}{i} \sum_{x\in f^{-1}\{y\}} \chi_{A_i}(x)\,d\mathcal{H}^n(y) \\
&= \int_{\mathbb{R}^m} \sum_{x\in f^{-1}\{y\}} \left(\sum_{i=1}^{\infty} \frac{1}{i}\chi_{A_i}(x) \right) d\mathcal{H}^n(y) \\
&= \int_{\mathbb{R}^m} \left[\sum_{x\in f^{-1}\{y\}} g(x) \right] d\mathcal{H}^n(y).
\end{aligned}$$

`Case 2:` Let now, in favor of generality, g be any $\mathcal{L}^n$-summable function.

Simply, we write $g = g^+ - g^-$ and apply Case 1 on g^+ and g^-. □

4.3 Applications

For reasons of simplicity and elegance, we restate the **Area formula** in a more "practical" way, in the form that we will need in the Applications;

AREA FORMULA. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ $(n \leq m)$ be Lipschitz continuous. Then for each $\mathcal{L}^n$-measurable subset $A \subseteq \mathbb{R}^n$,

$$\int_A Jf\, dx = \int_{f(A)} \mathcal{H}^0\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n(y).$$

It is clear that for any $y \notin f(A)$, we get $\mathcal{H}^0\big(A \cap f^{-1}\{y\}\big) = 0$, which does not contribute anything to the integral.

A. Length of a curve. Assume $f : \mathbb{R} \to \mathbb{R}^m$ $(m \geq 1)$ is Lipschitz and 1-1. Let us denote $f = (f^1, ..., f^m)$ and so $Df = \left(\dfrac{df^1}{dt}, \cdots, \dfrac{df^m}{dt}\right)$ Therefore

$$Jf = \sqrt{(Df)\cdot(Df)^T} = \sqrt{\sum_{i=1}^{m}\left(\frac{df^i}{dt}\right)^2} = \|Df\| = \left\|\frac{df}{dt}\right\|.$$

Consider $-\infty < \alpha < b < \infty$ and define the curve $C \coloneqq f([\alpha, b]) \subseteq \mathbb{R}^m$.
Since f is injective, for any $y \in C$ there exists a unique $x \in [\alpha, b]$ such that $f(x) = y$. Hence, in this case $\mathcal{H}^0\big([\alpha, b] \cap f^{-1}\{y\}\big) = 1$.
Consequentially, by the Area formula we get that Then

$$\int_\alpha^b Jf(t)\, d\mathcal{L}^n(t) = \int_{f([\alpha,b])} 1\, d\mathcal{H}^1(y) = \int_C d\mathcal{H}^1 = \mathcal{H}^1(C).$$

This, effectively, proves that

$$\mathcal{H}^1(C) = \text{length of } C = \int_\alpha^b \left\|\frac{df}{dt}\right\| dt.$$

B. Surface area of a graph. Assume $g : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz function. We define $f : \mathbb{R}^n \to \mathbb{R}^{n+1}$ as

$$f(x) \coloneqq (x, g(x)).$$

Hence

$$Df = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ g_{x_1} & \cdots & g_{x_n} \end{pmatrix}_{(n+1)\times n}$$

and

$$\begin{aligned} (Jf)^2 &= \text{sum of squares of } (n \times n) - \text{subdeterminants} \\ &= 1 + \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x_i} \right)^2 \\ &= 1 + \|Dg\|^2 . \end{aligned}$$

Now, for each open set $U \subseteq \mathbb{R}^n$, we define the graph of g over U as

$$G = G(g; U) \coloneqq \big\{ (x, g(x)) \mid x \in U \big\} = U \times g(U) \subseteq \mathbb{R}^{n+1}.$$

It is easy to notice that f is one-to-one, hence, as we saw previously, for any $y \in G = f(U)$ we get that $\mathcal{H}^0\big(U \cap f^{-1}\{y\}\big) = 1$. Consequently,

$$\int_U Jf(x)\, d\mathcal{L}^n(x) = \int_G 1\, d\mathcal{H}^n(y) = \mathcal{H}^n(G).$$

Thus, we get that

$$\mathcal{H}^n(G) = \text{surface area of G} = \int_U \big(1 + \|Dg\|^2\big)^{\frac{1}{2}} dx.$$

C. Surface area of a parametric hypersurface. Assume $f : \mathbb{R}^n \to \mathbb{R}^{n+1}$ is Lipschitz function and 1-1. Denote f as $f = (f^1, ..., f^{n+1})$. Thus

$$Df = \begin{pmatrix} f^1_{x_1} & \cdots & f^1_{x_n} \\ \vdots & \ddots & \vdots \\ f^{n+1}_{x_1} & \cdots & f^{n+1}_{x_n} \end{pmatrix}_{(n+1)\times n}$$

Therefore

$$\begin{aligned}(Jf)^2 &= \text{sum of squares of } (n \times n) - \text{subdeterminants} \\ &= \sum_{k=1}^{n+1} \left[\frac{\partial(f^1, ..., f^{k-1}, f^{k+1}, ..., f^{n+1})}{\partial(x_1, ..., x_n)} \right]^2 .\end{aligned}$$

Now, for each open set $U \subseteq \mathbb{R}^n$, we write $S \coloneqq f(U) \subseteq \mathbb{R}^{n+1}$ Hence,

$$\begin{aligned}\mathcal{H}^n(S) &= \text{n} - \text{dimensional surface area of S} \\ &= \int_U \left(\sum_{k=1}^{n+1} \left[\frac{\partial(f^1, ..., f^{k-1}, f^{k+1}, ..., f^{n+1})}{\partial(x_1, ..., x_n)} \right]^2 \right)^{\frac{1}{2}} dx.\end{aligned}$$

D. Submanifolds. Let $M \subseteq \mathbb{R}^m$ be a n-dimensional embedded Lipschitzian submanifold. Suppose that $U \subseteq \mathbb{R}^n$ and $f : U \to M$ a chart for M. Let $A \subseteq f(U)$ a Borel subset and set $B \coloneqq f^{-1}(A)$. We denote

$$\frac{\partial f}{\partial x_i} \coloneqq \left(\frac{\partial f^1}{\partial x_i}, \cdots, \frac{\partial f^n}{\partial x_i}\right).$$

Define

$$g_{ij} = \frac{\partial f}{\partial x_i} \cdot \frac{\partial f}{\partial x_j}, \ \ (i, j = 1, ..., n).$$

Then, the metric g induces the following matrix

$$G = (g_{ij}) = (Df)^* \circ Df$$

and so

$$Jf = \sqrt{\det G} \ .$$

Therefore, by applying the Area Formula, we get that

$$\begin{aligned}
\int_B \sqrt{\det G}\, dx &= \int_{f(B)} \mathcal{H}^0\big(B \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n(y) \\
&= \int_A \mathcal{H}^0\big(f^{-1}(A) \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n(y) \\
&\overset{f:1-1}{=} \int_A 1\, d\mathcal{H}^n(y) \\
&= \mathcal{H}^n(A).
\end{aligned}$$

Hence

$$\mathcal{H}^n(A) = \text{volume of A in M} = \int_B \sqrt{\det G}\, dx.$$

CHAPTER 5

The Coarea formula

In this Chapter, we will present the so-called Coarea formula, which is the other side of the problem we are studying, involving Lipschitz continuous mappings of the form

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

for $n \geq m$, this time.

We start by proving some introductory lemmas, and then proceed to the formula. We conclude by presenting some important applications, showcasing the vast spectrum of results, stemming from both Formulæ.

For a detailed listing of the Bibliographic sources used in the present Chapter, we direct to the *References and notes* paragraph on p.140.

Throughout this Chapter, we assume

$$n \geq m.$$

5.1 Preliminaries

Lemma 5.1. *Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, and $A \subseteq \mathbb{R}^n$ is a $\mathcal{L}^n$-measurable set. Then*

1. *The mapping $y \mapsto \mathcal{H}^{n-m}\big(A \cap L^{-1}\{y\}\big)$ is $\mathcal{L}^m$-measurable.*
2. $$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap L^{-1}\{y\}\big)\, dy = [\![L]\!]\, \mathcal{L}^n(A).$$

Proof. We will proceed by examining the different cases.

`Case 1:` $\dim L(\mathbb{R}^n) < m$.

Then for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$, we have that $A \cap L^{-1}\{y\} = \varnothing$, hence

$$\mathcal{H}^{n-m}\big(A \cap L^{-1}\{y\}\big) = 0.$$

This concludes the measurability of the map $y \mapsto \mathcal{H}^{n-m}\big(A \cap L^{-1}\{y\}\big)$.

Moreover, by Polar Decomposition Theorem we have that $L = S \circ O^*$, for a symmetric map $S : \mathbb{R}^m \to \mathbb{R}^m$ and an orthogonal map $O : \mathbb{R}^m \to \mathbb{R}^n$. Hence, $O^* : \mathbb{R}^n \to \mathbb{R}^m$ and so $L(\mathbb{R}^n) = S(O^*(\mathbb{R}^n)) = S(\mathbb{R}^m)$. Thus $\dim S(\mathbb{R}^m) < m$ and $[\![L]\!] = |\det S| = 0$. Assertion (*2.*) is proven trivially.

`Case 2:` $L = P =$ orthogonal projection of $\mathbb{R}^n$ onto $\mathbb{R}^m$.

Then for each $y \in \mathbb{R}^m$, the inverse image $P^{-1}\{y\}$ is an $(n-m)$-dimensional affine subspace of $\mathbb{R}^n$ and a translate of $P^{-1}\{0\}$. Indeed, via elementary calculations, we can see that for a fixed $y \in \mathbb{R}^m$, we get;

$$\begin{aligned}
P^{-1}\{y\} &= \{x \in \mathbb{R}^n \mid P(x) = y\} \\
&= \{x = (z,w) \in \mathbb{R}^m \times \mathbb{R}^{n-m} \mid P(z,w) = y\} \\
&= \{x = (z,w) \in \mathbb{R}^m \times \mathbb{R}^{n-m} \mid z = y\} \\
&= \{(y,w) \mid w \in \mathbb{R}^{n-m}\} \\
&= \{(y,0) + (0,w) \mid w \in \mathbb{R}^{n-m}\} \\
&= (y,0) + \{(0,w) \mid w \in \mathbb{R}^{n-m}\} \\
&= (y,0) + \{x = (0,w) \in \mathbb{R}^n \mid w \in \mathbb{R}^{n-m} \,\&\, P(x) = 0\} \\
&= (y,0) + P^{-1}\{0\}.
\end{aligned}$$

Then Fubini's Theorem implies

$$y \mapsto \mathcal{H}^{n-m}\big(A \cap P^{-1}\{y\}\big) \ \text{ is } \ \mathcal{L}^m - \text{measurable}$$

and

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap P^{-1}\{y\}\big)\, dy = \mathcal{L}^n(A). \tag{$\star$}$$

Indeed, simply let $A_y \coloneqq \{z \in \mathbb{R}^{n-m} \mid (y,z) \in A\}$. Then $\chi_{A_y}(z) = \chi_A(y,z)$. Hence, we compute as follows

$$\begin{aligned}
\mathcal{L}^n(A) = \int_{\mathbb{R}^n} \chi_A(y,z)\, d\mathcal{L}^n(y,z) &= \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^{n-m}} \chi_{A_y}(z)\, d\mathcal{L}^{n-m}(z) \right) d\mathcal{L}^m(y) \\
&= \int_{\mathbb{R}^m} \mathcal{L}^{n-m}(A_y)\, d\mathcal{L}^m(y)
\end{aligned}$$

$$= \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A_y)\, d\mathcal{L}^m(y)$$
$$= \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap P^{-1}\{y\}\big)\, d\mathcal{L}^m(y).$$

`Case 3:` $\dim L(\mathbb{R}^n) = m$.

Again, by Polar Decomposition we get the $L = S \circ O^*$ expression, for a symmetric map $S : \mathbb{R}^m \to \mathbb{R}^m$ and an orthogonal map $O : \mathbb{R}^m \to \mathbb{R}^n$. This time, since S has full rank, we have that $[\![L]\!] = |\det S| > 0$.

`Claim:` We contend that; There exists an orthogonal map $Q : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$O^* = P \circ Q$$

where P is the orthogonal projection of $\mathbb{R}^n$ onto $\mathbb{R}^m$.

Proof of claim: We will construct the map Q in steps; Let $\{e_1, ..., e_m\}$ the canonical base of $\mathbb{R}^m$. Since $O : \mathbb{R}^m \to \mathbb{R}^n$ is an orthogonal map, we set $v_i = O(e_i)$ for $1 \leq i \leq m$, and define

$$\begin{cases} Q(v_1) = (e_1, \vec{0}) \in \mathbb{R}^n \\ \vdots \\ Q(v_m) = (e_m, \vec{0}) \in \mathbb{R}^n \end{cases}$$

We extend the set $\{v_1, ..., v_m\}$ to an orthonormal base of $\mathbb{R}^n$, let us denote it as $\{v_1, ..., v_m, v_{m+1}, ..., v_n\}$, and we set

$$Q(v_i) = w_i \in \mathbb{R}^n \;\; (i = m+1, ..., n),$$

where the choice of w_i is such that

$$\big\{(e_i, \vec{0}) \;:\; i = 1, ..., m\big\} \cup \big\{w_{m+1}, ..., w_n\big\}$$

is an orthonormal base of $\mathbb{R}^n$. For ease of our notation, we will denote by $w_i \coloneqq (e_i, \vec{0})$ for $i = 1, ..., m$. And so;

$$Q(v_i) = w_i, \;\; (i = 1, ..., m, m+1, ..., n),$$

where both the sets $\{v_1, ..., v_m, v_{m+1}, ..., v_n\}$ and $\{w_1, ..., w_m, w_{m+1}, ..., w_n\}$ are orthonormal bases of of $\mathbb{R}^n$. Therefore, the map $Q : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal.

We now turn our attention to its adjoint, $Q^* : \mathbb{R}^n \to \mathbb{R}^n$. From the defining property, for any $\bar{x} = (x_1, ..., x_m, 0, ..., 0) \in \mathbb{R}^n$ we get that;

$$\bar{x} \cdot Q(y) = Q^*(\bar{x}) \cdot y \;\text{ for all } y \in \mathbb{R}^n.$$

Observe that;

$$Q(y) = \sum_{i=1}^{n} (y \cdot v_i) Q(v_i) = \sum_{i=1}^{n} (y \cdot v_i) w_i = \sum_{i=1}^{m} \lambda_i (e_i, \vec{0}) + \sum_{i=m+1}^{n} \lambda_i w_i,$$

where $\lambda_i = (y \cdot v_i)$. Moreover, from the orthogonality between the vectors of the $\{w_i : i = 1, ..., n\}$, we get that

$$w_j \cdot (e_i, \vec{0}) = 0 \ \text{ for all } j = m+1, ..., n \text{ and all } i = 1, ..., m,$$

which results in the first m coordinates of $\{w_i : i = m+1, ..., n\}$ being equal to zero. Hence, we get the following expression;

$$w_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \star_1 \\ \star_2 \\ \vdots \\ \star_{n-m} \end{pmatrix} \quad \text{for all } i = m+1, ..., n.$$

Therefore, we have that;

$$Q^*(\bar{x}) \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdot \left(\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \star_1 \\ \vdots \\ \star_{n-m} \end{pmatrix} \right) = \lambda_1 x_1 + \cdots + \lambda_n x_n = \sum_{i=1}^{m} \lambda_i x_i$$

$$= \sum_{i=1}^{m} (y \cdot v_i) x_i = \sum_{i=1}^{m} x_i \big(O(e_i) \cdot y \big) = O\left(\sum_{i=1}^{m} x_i e_i \right) \cdot y = O(\bar{x}) \cdot y$$

for all $y \in \mathbb{R}^n$. Hence, we end up with the following equality

$$Q^*(x_1, ..., x_m, 0, ..., 0) = O(x_1, ..., x_m)$$

for all $x = (x_1, ..., x_m) \in \mathbb{R}^m$. Moreover, from the defining property of the Adjoint, after "feeding" it with the canonical basis vectors of $\mathbb{R}^n$ and performing the necessary calculations, we get that;

$$P^*(x_1, ..., x_m) = (x_1, ..., x_m, 0, ..., 0) \in \mathbb{R}^n.$$

Consequently,

$$O = Q^* \circ P^*,$$

which gives;

$$O^* = (Q^* \circ P^*)^* = (P^*)^* \circ (Q^*)^* = P \circ Q,$$

which concludes the proof of our contention.

Returning to our proof; It is easy to deduce that $L^{-1}\{0\}$ is an $(n-m)$-dimensional subspace of $\mathbb{R}^n$ and $L^{-1}\{y\}$ is a translate of $L^{-1}\{0\}$ for all $y \in \mathbb{R}^m$.

Hence Fubini's Theorem implies the $\mathcal{L}^m$-measurability of the map $y \mapsto \mathcal{H}^{n-m}\big(A \cap L^{-1}\{y\}\big)$.

Now, we employ Case 2 from above, applying it on the projection map P, and we make use of our Claim, in order to calculate that

$$\begin{aligned} \mathcal{L}^n(A) &= \mathcal{L}^n(Q(A)) \\ &\overset{(\star)}{=} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(Q(A) \cap P^{-1}\{y\}\big)\, dy \\ &= \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap (Q^{-1} \circ P^{-1}\{y\})\big)\, dy. \end{aligned}$$

We perform a simple "re-branding" of our variable, employing the help of our symmetric map S, by setting $z = Sy$. Thus

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap (Q^{-1} \circ P^{-1} \circ S^{-1}\{z\})\big)\, \frac{dz}{|\det S\,|} = \mathcal{L}^n(A). \qquad (\star\star)$$

Observe, now, that $L = S \circ O^* = S \circ P \circ Q$ and what $(\star\star)$ essentially gives us is that

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap (L^{-1}\{z\})\big)\, dz = |\det S\,|\, \mathcal{L}^n(A) = [\![L]\!]\, \mathcal{L}^n(A).$$

The proof of the Lemma is now complete. □

REMARK. For the first case of our Lemma, we made a delicate contention, we would like to adress here, namely that; For an Orthogonal map $O : \mathbb{R}^m \to \mathbb{R}^n$ $(m \leq n)$, we have that

$$O^*(\mathbb{R}^n) = \mathbb{R}^m$$

namely, that O^* is onto, i.e. an epimorphism to its image.

We adress this contention, via the following well-known **Proposition** of Linear Algebra;

Proposition 5.1. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map and $T^* : \mathbb{R}^m \to \mathbb{R}^n$ be its adjoint. Then*

$$(Ker\ T)^{\perp} = Im(T^*),$$

where by $V^{\perp}$ we denote the orthogonal complement of a subset V of a vector space X, namely the set;

$$V^{\perp} := \{x \in X \mid v \cdot x = 0, \text{ for all } v \in V\}.$$

Proof. We prove the two set-inclusions seperately;
Claim 1: $Im(T^*) \subseteq (Ker\ T)^{\perp}$.

Let $x \in Im(T^*)$. Then there exists a $y \in \mathbb{R}^m$ such that $x = T^*y$. Now, for all $u \in Ker\ T$, we get that;

$$x \cdot u = T^*y \cdot u = y \cdot Tu = 0.$$

Hence, by definition, we get that $x \in (Ker\ T)^{\perp}$.
Claim 2: $(Ker\ T)^{\perp} \subseteq Im(T^*)$.

For this part of the proof we will work in a clever way, and we will show instead that;

$$Im(T^*)^{\perp} \subseteq Ker\ T.$$

This is possible, since we work in finite dimensional spaces, due to the following well-established properties;

$$A \subseteq B \Rightarrow B^{\perp} \subseteq A^{\perp} \ \ \& \ \ (A)^{\perp\perp} = A.$$

Indeed, let $x \in Im(T^*)^{\perp}$. Then, for all $v \in Im(T^*)$, we get that $v \cdot x = 0$. Since v belongs in the image of T^*, and the above holds for the whole of $Im(T^*)$, we deduce that;

$$T^*y \cdot x = 0, \text{ for all } y \in \mathbb{R}^m.$$

Hence, by the defining property of the Adjoint, we get that

$$y \cdot Tx = 0, \text{ for all } y \in \mathbb{R}^m.$$

Therefore, we must have that $Tx = 0$, ergo $x \in Ker\ T$, which concludes our proof.

Now, for the matter at hand; Since $O : \mathbb{R}^m \to \mathbb{R}^n$ is an Orthogonal map, therefore an isometry, we get that $Ker\, O = \{\vec{0}\}$. Hence, it is trivial to observe that;

$$O^*(\mathbb{R}^n) = Im(O^*) = (Ker\, O)^{\perp} = \{\vec{0}\}^{\perp} = \mathbb{R}^m,$$

which concludes the proof for our contention. □

Lemma 5.2. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz continuous map and $A \subseteq \mathbb{R}^n$ be a $\mathcal{L}^n$-measurable set. Then*

1. *$A \cap f^{-1}\{y\}$ is $\mathcal{H}^{n-m}$-measurable for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$.*
2. *The mapping $y \mapsto \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)$ is $\mathcal{L}^m$-measurable.*
3. $$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)\, dy \le \frac{\alpha(n-m)\alpha(m)}{\alpha(n)}(\mathrm{Lip}(f))^m \mathcal{L}^n(A).$$

Proof. We shall proceed in a retrograde motion.
From the measurability of set A, we deduce that: For each $j = 1, 2, ...$ there exist closed balls $\{B_i^j\}_{i=1}^\infty$ such that

$$A \subseteq \bigcup_{i=1}^\infty B_i^j \text{ with } \operatorname{diam} B_i^j \le \frac{1}{j} \text{ for which} \sum_{i=1}^\infty \mathcal{L}^n(B_i^j) \le \mathcal{L}^n(A) + \frac{1}{j}.$$

We define functions $g_i^j : \mathbb{R}^m \to [0, \infty)$ as

$$g_i^j \coloneqq \alpha(n-m)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n-m} \chi_{f(B_i^j)}.$$

Observe that g_i^j are $\mathcal{L}^m$-measurable.

Moreover, note that for all $y \in \mathbb{R}^m$, we have that

$$\mathcal{H}^{n-m}_{\frac{1}{j}}\big(A \cap f^{-1}\{y\}\big) \le \sum_{i=1}^\infty g_i^j(y),$$

since

$$A \cap f^{-1}\{y\} \subseteq \left(\bigcup_{i=1}^\infty B_i^j\right) \cap f^{-1}\{y\} = \bigcup_{i=1}^\infty \big(B_i^j \cap f^{-1}\{y\}\big) \quad \text{with}$$
$$\operatorname{diam}\big(B_i^j \cap f^{-1}\{y\}\big) \le \operatorname{diam} B_i^j \le \frac{1}{j}$$

and thus

$$\mathcal{H}^{n-m}_{\frac{1}{j}}\big(A \cap f^{-1}\{y\}\big) \le \sum_{i=1}^\infty \alpha(n-m)\left(\frac{\operatorname{diam}\big(B_i^j \cap f^{-1}\{y\}\big)}{2}\right)^{n-m}$$

$$\leq \sum_{i=1}^{\infty} \alpha(n-m)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n-m} \chi_{f(B_i^j)} = \sum_{i=1}^{\infty} g_i^j.$$

In order to proceed further, since the measurability of the map $y \mapsto \mathcal{H}^{n-m}\big(A\cap f^{-1}\{y\}\big)$ is yet to be proved, we will employ the upper integral we defined in 1.11 for the Lebesgue measure. With this and also with Fatou's Lemma and the Isodiametric Inequality, we compute

$$\begin{aligned}
&\int_{\mathbb{R}^m}^{\star} \mathcal{H}^{n-m}\big(A\cap f^{-1}\{y\}\big)\, dy \\
&\qquad = \int_{\mathbb{R}^m}^{\star} \lim_{j\to\infty} \mathcal{H}^{n-m}_{\frac{1}{j}}\big(A\cap f^{-1}\{y\}\big)\, dy \\
&\qquad \leq \int_{\mathbb{R}^m} \liminf_{j\to\infty} \sum_{i=1}^{\infty} g_i^j \, dy \\
&\qquad \leq \liminf_{j\to\infty} \sum_{i=1}^{\infty} \int_{\mathbb{R}^m} g_i^j \, dy \\
&\qquad = \liminf_{j\to\infty} \sum_{i=1}^{\infty} \int_{\mathbb{R}^m} \alpha(n-m)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n-m} \chi_{f(B_i^j)}\, dy \\
&\qquad = \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(n-m)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n-m} \int_{\mathbb{R}^m} \chi_{f(B_i^j)}\, dy \\
&\qquad = \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(n-m)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n-m} \mathcal{L}^m\big(f(B_i^j)\big) \\
&\qquad \leq \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(n-m)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n-m} \alpha(m)\left(\frac{\operatorname{diam} f(B_i^j)}{2}\right)^{m} \\
&\qquad \leq \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(n-m)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n-m} \alpha(m)(\operatorname{Lip}(f))^m\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{m} \\
&\qquad = \frac{\alpha(n-m)\alpha(m)}{\alpha(n)}(\operatorname{Lip}(f))^m \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(n)\left(\frac{\operatorname{diam} B_i^j}{2}\right)^{n} \\
&\qquad = \frac{\alpha(n-m)\alpha(m)}{\alpha(n)}(\operatorname{Lip}(f))^m \liminf_{j\to\infty} \sum_{i=1}^{\infty} \mathcal{L}^n(B_i^j) \\
&\qquad \leq \frac{\alpha(n-m)\alpha(m)}{\alpha(n)}(\operatorname{Lip}(f))^m \mathcal{L}^n(A).
\end{aligned}$$

Thus

$$\int_{\mathbb{R}^m}^{\star} \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)\, dy \leq \frac{\alpha(n-m)\alpha(m)}{\alpha(n)}(\mathrm{Lip}(f))^m \mathcal{L}^n(A). \qquad (\star)$$

Recall that (*3.*) will stem from ($\star$) once we establish (*2.*).

`Case 1:` A compact.

Fix $t \geq 0$ and for each positive integer i, we define U_i as the set of points $y \in \mathbb{R}^m$, for which there exist finitely many open sets $S_1, ..., S_\ell$ such that

$$\begin{cases} A \cap f^{-1}\{y\} \subseteq \bigcup_{j=1}^{\ell} S_j \\ \operatorname{diam} S_j \ \leq \dfrac{1}{i} \\ \displaystyle\sum_{j=1}^{\ell} \alpha(n-m)\left(\frac{\operatorname{diam} S_j}{2}\right)^{n-m} \leq t + \frac{1}{i} \end{cases} \qquad (\star\star)$$

`Claim 1:` U_i is open.

Proof of claim: Assume $y \in U_i$ (for some $i = 1, 2, ...$). Then there exist sets $S_1, ..., S_\ell$ such that ($\star\star$) hold. We contend that; There exists $r > 0$ such that

$$A \cap f^{-1}\big(\mathcal{N}(y, r)\big) \subseteq \bigcup_{j=1}^{\ell} S_j,$$

where $\mathcal{N}(\bullet, r)$ is used to denote the open ball with radius $r > 0$.

Let us suppose that there is no such $r > 0$. Then, we can locate a sequence $(y_N)_{N \in \mathbb{N}}$ in $\mathbb{R}^m$ converging to the point y, such that;

$$\text{For every } N \in \mathbb{N}, \text{ there exists } \ x_N \in f^{-1}\{y_N\} \cap A \setminus \bigcup_{j=1}^{\ell} S_j.$$

Since A is taken to be compact and S_j $(1 \leq j \leq \ell)$ are open, then $A \setminus \bigcup_{j=1}^{\ell} S_j$ is also a compact set. From the Sequential compactness we deduce that the sequence $(x_N)_{N \in \mathbb{N}} \subseteq A \setminus \bigcup_{j=1}^{\ell} S_j$ has a convergent sub-sequence, which we will denote the same way, "abusing" slightly our notation, namely;

$$x_N \to x \in A \setminus \bigcup_{j=1}^{\ell} S_j.$$

Now, the continuity of f implies that

$$f(x) = \lim f(x_N) = \lim y_N = y.$$

Hence $x \in f^{-1}\{y\} \cap A \setminus \bigcup_{j=1}^{\ell} S_j$. We have reached a contradiction.

This, essentially, concludes the proof of our first Claim, since the preceding contention implies that $\mathcal{N}(y, r) \subseteq U_i$.

`Claim 2:` $\big\{y \mid \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \le t\big\} = \bigcap_{i=1}^{\infty} U_i.$

Proof of claim: We will prove the two inclusions.
Let $y \in \big\{y \mid \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \le t\big\}$. Then, since $\mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \le t$, we get that for all $\delta > 0$,

$$\mathcal{H}_\delta^{n-m}\big(A \cap f^{-1}\{y\}\big) \le t.$$

Now, fix an index i. We will choose $\delta \in \left(0, \frac{1}{i}\right)$. The definition of $\mathcal{H}_\delta^{n-m}$ measure, implies the existence of a cover $\{S_j\}_{j=1}^{\infty}$ for which

$$\begin{cases} A \cap f^{-1}\{y\} \subseteq \bigcup_{j=1}^{\infty} S_j \\ \operatorname{diam} S_j \le \delta < \frac{1}{i} \\ \sum_{j=1}^{\infty} \alpha(n-m) \left(\frac{\operatorname{diam} S_j}{2}\right)^{n-m} < t + \frac{1}{i} \end{cases}$$

We may as well assume that S_j are open; Recall that in the Remark following Theorem 2.1, we made a similar contention, by taking a closed cover. The justification in the present contention is analogous. Now, since $A \cap f^{-1}\{y\}$ is compact, there exists a finite subcollection $\{S_1, ..., S_\ell\}$ covering $A \cap f^{-1}\{y\}$. Consequentially, $y \in U_i$. Finally, seeing that i was arbitrary, we get that;

$$\big\{y \mid \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \le t\big\} \subseteq \bigcap_{i=1}^{\infty} U_i.$$

On the other hand, if $y \in \bigcap_{i=1}^{\infty} U_i$, then conditions $(\star\star)$ hold, resulting in

$$\mathcal{H}^{n-m}_{\frac{1}{i}}\big(A \cap f^{-1}\{y\}\big) \leq t + \frac{1}{i} \quad \text{(for each } i\text{)}$$

and so

$$\mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \leq t.$$

Hence

$$\bigcap_{i=1}^{\infty} U_i \subseteq \big\{y \mid \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \leq t\big\},$$

completing the proof of our second Claim.

Consequently, the set $\big\{y \mid \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \leq t\big\}$ is Borel. Hence, for a compact set A, the mapping $y \mapsto \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)$ is a Borel map.

`Case 2:` A is open.

We can "exhaust" A by compact sets, i.e. there exist compact sets $K_1 \subseteq K_2 \subseteq \cdots \subseteq A$, such that

$$A = \bigcup_{i=1}^{\infty} K_i.$$

Hence for each $y \in \mathbb{R}^m$, from regularity of the $\mathcal{H}^{n-m}$ measure, we get that

$$\begin{aligned}\mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) &= \mathcal{H}^{n-m}\left(\bigcup_{i=1}^{\infty} K_i \ \cap \ f^{-1}\{y\}\right) \\ &= \mathcal{H}^{n-m}\left(\bigcup_{i=1}^{\infty} \big(K_i \cap f^{-1}\{y\}\big)\right) \\ &= \lim_{i\to\infty} \mathcal{H}^{n-m}\big(K_i \cap f^{-1}\{y\}\big).\end{aligned}$$

Therefore, the mapping $y \mapsto \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)$ is Borel measurable.

`Case 3:` $\mathcal{L}^n(A) < \infty$.

There exists a countable family of open sets $V_1 \supseteq V_2 \supseteq \cdots \supseteq A$, such that $\mathcal{L}^n(V_1) < \infty$ and

$$\lim_{i\to\infty} \mathcal{L}^n(V_i \setminus A) = \mathcal{L}^n\left(\bigcap_{i=1}^{\infty} (V_i \setminus A)\right) = \mathcal{L}^n\left(\left(\bigcap_{i=1}^{\infty} V_i\right) \setminus A\right) = 0.$$

Moreover, observe that $V_i \subseteq A \cup (V_i \setminus A)$, hence

$$\mathcal{H}^{n-m}\big(V_i \cap f^{-1}\{y\}\big) \leq \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) + \mathcal{H}^{n-m}\big((V_i \setminus A) \cap f^{-1}\{y\}\big).$$

Thus, we now get that

$$\begin{aligned}\limsup_{i\to\infty} \int_{\mathbb{R}^m}^{\star} &\big|\mathcal{H}^{n-m}\big(V_i \cap f^{-1}\{y\}\big) - \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)\big| \, dy \\ &\leq \limsup_{i\to\infty} \int_{\mathbb{R}^m}^{\star} \mathcal{H}^{n-m}\big((V_i \setminus A) \cap f^{-1}\{y\}\big) \, dy \\ &\overset{(\star)}{\leq} \limsup_{i\to\infty} \frac{\alpha(n-m)\alpha(m)}{\alpha(n)} (\mathrm{Lip}(f))^m \mathcal{L}^n(V_i \setminus A) = 0.\end{aligned}$$

Consequently, by employing Lemma 1.2, we obtain that;

$$\mathcal{H}^{n-m}\big(V_i \cap f^{-1}\{y\}\big) \to \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big) \ \ (\mathcal{L}^m - a.e.).$$

Since V_i are open sets, we may employ Case 2, and conclude that

$$y \mapsto \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)$$

is $\mathcal{L}^m$-measurable,as the limit of measurable maps $y \mapsto \mathcal{H}^{n-m}\big(V_i \cap f^{-1}\{y\}\big)$.

Moreover, we deduce that $\mathcal{H}^{n-m}\big((V_i \setminus A) \cap f^{-1}\{y\}\big) \to 0$ $(\mathcal{L}^m - a.e)$ and so $A \cap f^{-1}\{y\}$ is $\mathcal{H}^{n-m}$-measurable for $\mathcal{L}^m - a.e\ y$, since we can "decompose" it as

$$A \cap f^{-1}\{y\} = \left(\bigcap_{i=1}^{\infty} V_i\right) \cap f^{-1}\{y\} \cap \left(\bigcap_{i=1}^{\infty} \Big((V_i \setminus A) \cap f^{-1}\{y\}\Big)\right).$$

`Case 4:` $\mathcal{L}^n(A) = \infty$.

We express A as a union of an increasing sequence of bounded $\mathcal{L}^n$-measurable sets and apply Case 3 in order to prove the $\mathcal{H}^{n-m}$-measurability of $A \cap f^{-1}\{y\}$ for $\mathcal{L}^m - a.e.y$ and the $\mathcal{L}^m$-measurability of the map

$$y \mapsto \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big).$$

This concludes the proof of *(1.)* and *(2.)*. Then *(3.)* follows directly from $(\star)$. □

REMARK. We will offer here a "replacement" for Assertion (*3.*) above. Mimicking essentially the preceding proof, we will demonstrate that;

$$\int_{\mathbb{R}^\ell}^{\star} \mathcal{H}^k\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^\ell(y) \leq \frac{\alpha(k)\alpha(\ell)}{\alpha(k+\ell)} (\mathrm{Lip}(f))^\ell \mathcal{H}^{k+\ell}(A)$$

for a Lipschitz map $f : \mathbb{R}^n \to \mathbb{R}^\ell$ and an arbitrary $k \in [0, \infty)$, for **all** $\boldsymbol{A} \subseteq \mathbb{R}^{\boldsymbol{n}}$, i.e. without the additional assumption for $\mathcal{L}^n$-measurability of the set A.

Proof: From the definition of $\mathcal{H}^{k+\ell}_{\frac{1}{j}}$ measure,we get that; For each $j = 1, 2, ...$ There exists a cover $\{C_i^j\}_{i=1}^\infty$ consisting of closed sets, such that $A \subseteq \bigcup_{i=1}^{\infty} C_i^j$ with $\operatorname{diam} C_i^j \leq \frac{1}{j}$, for which

$$\sum_{i=1}^{\infty} \alpha(k+\ell) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^{k+\ell} \leq \mathcal{H}^{k+\ell}_{\frac{1}{j}}(A) + \frac{1}{j}.$$

We define functions $g_i^j : \mathbb{R}^\ell \to [0, \infty)$ as

$$g_i^j \coloneqq \alpha(k) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^k \chi_{f(C_i^j)}.$$

Clearly g_i^j are $\mathcal{L}^m$-measurable. Moreover, observe (in a similar way as above) that for all $y \in \mathbb{R}^\ell$, we get $\mathcal{H}^k_{\frac{1}{j}}\big(A \cap f^{-1}\{y\}\big) \leq \sum_{i=1}^{\infty} g_i^j(y)$.

Once again, we make use of Fatou's Lemma and the Isodiametric Inequality, and compute that

$$\begin{aligned}
&\int_{\mathbb{R}^\ell}^{\star} \mathcal{H}^k\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^\ell \\
&\qquad = \int_{\mathbb{R}^\ell}^{\star} \lim_{j\to\infty} \mathcal{H}^k_{\frac{1}{j}}\big(A \cap f^{-1}\{y\}\big)\, d\mathcal{H}^\ell \\
&\qquad \leq \int_{\mathbb{R}^\ell} \liminf_{j\to\infty} \sum_{i=1}^{\infty} g_i^j \, d\mathcal{H}^\ell \\
&\qquad \leq \liminf_{j\to\infty} \sum_{i=1}^{\infty} \int_{\mathbb{R}^\ell} g_i^j \, d\mathcal{H}^\ell \\
&\qquad = \liminf_{j\to\infty} \sum_{i=1}^{\infty} \int_{\mathbb{R}^\ell} \alpha(k) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^k \chi_{f(C_i^j)} \, d\mathcal{H}^\ell
\end{aligned}$$

$$
\begin{aligned}
&= \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(k) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^k \int_{\mathbb{R}^\ell} \chi_{f(C_i^j)} \, d\mathcal{H}^\ell \\
&= \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(k) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^k \mathcal{H}^\ell\big(f(C_i^j)\big) \\
&= \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(k) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^k \mathcal{L}^\ell\big(f(C_i^j)\big) \\
&\leq \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(k) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^k \alpha(\ell) \left(\frac{\operatorname{diam} f(C_i^j)}{2} \right)^\ell \\
&\leq \frac{\alpha(k)\alpha(\ell)}{\alpha(k+\ell)} (\operatorname{Lip}(f))^\ell \liminf_{j\to\infty} \sum_{i=1}^{\infty} \alpha(k+\ell) \left(\frac{\operatorname{diam} C_i^j}{2} \right)^{k+\ell} \\
&\leq \frac{\alpha(k)\alpha(\ell)}{\alpha(k+\ell)} (\operatorname{Lip}(f))^\ell \liminf_{j\to\infty} \left(\mathcal{H}^{k+\ell}_{\frac{1}{j}}(A) + \frac{1}{j} \right) \\
&= \frac{\alpha(k)\alpha(\ell)}{\alpha(k+\ell)} (\operatorname{Lip}(f))^\ell \, \mathcal{H}^{k+\ell}(A).
\end{aligned}
$$

REMARK. The preceding inequality in **Assertion *(3.)*** of Lemma 5.2 and its **variant** in the Remark above is known as **Eilenberg's Coarea Inequality** or simply "the Coarea inequality". It is considered to be a tool of great importance in Geometric Measure Theory, playing a key-role in the proof of the Coarea formula.

The inequality essentially says that the average size of "fibers" of f, "captured" by the integrand $\mathcal{H}^\bullet(f^{-1}\{y\} \cap A)$, is bounded by a term based on the Lipschitz constant, the dimensions and the original size of the set we are interested in, namely $\mathcal{H}^\bullet(A)$.

Eilenberg's Coarea Inequality's historical "journey" showcases the collaborative and evolving nature of mathematical research; Proved first by Eilenberg in 1938, for the case when the function was the distance to a fixed point in a metric space, it was later generalized by Eilenberg and Harold, in 1943, to the case of any real-valued Lipschitz function on a metric space, with the burden of some extra assumptions. In the next years, Federer sought a proof which would get rid of those additional assumptions, being convinced that they were unnecessary. He achieved a partial result in 1954, but a complete proof remained elusive. Only in 1984, R.O. Davies' work on Hausdorff measures provided the insights necessary, so that the inequality could finally be proved, they way it was predicted.

Lemma 5.3. *Let $t > 1$ and assume $h : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz function. We define set*

$$B \coloneqq \{x \mid Dh(x) : \text{exists}, \ Jh(x) > 0\}$$

Then there exists a countable collection $\{D_k\}_{k=1}^{\infty}$ consisting of Borel subsets of $\mathbb{R}^n$ such that

1. $\mathcal{L}^n\left(B \setminus \bigcup_{k=1}^{\infty} D_k\right) = 0,$

2. $h|_{D_k}$ *is one-to-one* $(k = 1, 2, ...)$*, and*

3. *For each $k = 1, 2, ...,$ there exists a symmetric automorphism $S_k : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$\operatorname{Lip}(S_k^{-1} \circ (h|_{D_k})) \leq t, \ \operatorname{Lip}((h|_{D_k})^{-1} \circ S_k) \leq t,$$
$$t^{-n}|\det S_k| \leq Jh|_{D_k} \leq t^n|\det S_k|.$$

Proof. We will proceed in a "constructive" way.

First, we will employ Lemma 4.4 on h, in place of f, in order to get disjoint Borel sets $\{E_k\}_{k=1}^{\infty}$ and symmetric automorphisms $T_k : \mathbb{R}^n \to \mathbb{R}^n \, (k = 1, 2, ...)$ such that

i. $B = \bigcup_{k=1}^{\infty} E_k,$

ii. $h|_{E_k}$ is one-to-one $(k = 1, 2, ...),$

iii. $\operatorname{Lip}((h|_{E_k}) \circ T_k^{-1}) \leq t, \ \operatorname{Lip}(T_k \circ (h|_{E_k})^{-1}) \leq t,$ and
$t^{-n}|\det T_k| \leq Jh|_{E_k} \leq t^n|\det T_k|.$

`Claim 1:` $(h|_{E_k})^{-1}$ is a Lipschitz continuous map.

Proof of claim: Since T_k^{-1} and $T_k \circ h|_{E_k}^{-1}$ are both Lipscitz maps, then their composition

$$h|_{E_k}^{-1} = T_k^{-1} \circ \left(T_k \circ h|_{E_k}^{-1}\right)$$

is also a Lipschitz map, with constant $\operatorname{Lip}(h|_{E_k}^{-1}) \leq t \left\|T_k^{-1}\right\|$.

Thus, the Extension Theorem (Thm. 3.1) provides us with a Lipschitz continuous mapping $h_k : \mathbb{R}^n \to \mathbb{R}^n$ such that $h_k = h|_{E_k}^{-1}$ on $h(E_k)$.

`Claim 2:` $Jh_k > 0$ $\mathcal{L}^n - a.e.$ on $h(E_k)$.

Proof of claim: Since $h_k \circ h(x) = x$ for all $x \in E_k$, Theorem 3.5 implies that

$$Dh_k(h(x)) \circ Dh(x) = I \quad \mathcal{L}^n - a.e \text{ on } E_k,$$

hence

$$Jh_k(h(x))Jh(x) = 1 \quad \mathcal{L}^n - a.e \text{ on } E_k.$$

Again, employing the (iii.) from above, we get that $Jh_k(h(x)) > 0$ for $\mathcal{L}^n$- a.e. $x \in E_k$. Now, since h is Lipschitz continuous, it is immediate that $Jh_k > 0$ $\mathcal{L}^n - a.e.$ on $h(E_k)$.

Once again, we will employ Lemma 4.4 to each and every h_k $(k = 1, 2, ...)$; There exists a collection of disjoint Borel sets $\{F_j^k\}_{j=1}^{\infty}$ and symmetric automorphisms $R_j^k : \mathbb{R}^n \to \mathbb{R}^n$ such that

iv. $\mathcal{L}^n\left(h(E_k) \setminus \bigcup_{k=1}^{\infty} F_j^k\right) = 0,$

v. $h_k|_{F_j^k}$ is one-to-one $(k = 1, 2, ...)$,

vi. $\mathrm{Lip}((h_k|_{F_j^k}) \circ (R_j^k)^{-1}) \leq t,\ \mathrm{Lip}(R_j^k \circ (h_k|_{F_j^k})^{-1}) \leq t,$ and
$$t^{-n}|\det R_j^k| \leq Jh_k|_{F_j^k} \leq t^n|\det R_j^k|.$$

Now, we define

$$D_j^k \coloneqq E_k \cap h^{-1}(F_j^k) \ \text{ and } \ S_j^k \coloneqq (R_j^k)^{-1} \quad (k = 1, 2, ...).$$

`Claim 3:` $\mathcal{L}^n\left(B \setminus \bigcup_{k,j=1}^{\infty} D_j^k\right) = 0.$

Proof of claim: We have that

$$h_k\left(h(E_k) \setminus \bigcup_{j=1}^{\infty} F_j^k\right) = h^{-1}\left(h(E_k) \setminus \bigcup_{j=1}^{\infty} F_j^k\right)$$
$$= h^{-1}\big(h(E_k)\big) \setminus h^{-1}\left(\bigcup_{j=1}^{\infty} F_j^k\right) = E_k \setminus \bigcup_{j=1}^{\infty} h^{-1}(F_j^k).$$

Moreover,

$$\begin{aligned}
E_k \setminus \bigcup_{j=1}^{\infty} D_j^k &= E_k \cap \left(\bigcup_{j=1}^{\infty} D_j^k\right)^c = E_k \cap \left(\bigcap_{j=1}^{\infty} (D_j^k)^c\right) \\
&= E_k \cap \left(\bigcap_{j=1}^{\infty} \Big(E_k \cap h^{-1}(F_j^k)\Big)^c\right) = E_k \cap \bigcap_{j=1}^{\infty} \Big(E_k^c \cup \big(h^{-1}(F_j^k)\big)^c\Big) \\
&= E_k \cap \left(E_k^c \cup \left(\bigcap_{j=1}^{\infty} \big(h^{-1}(F_j^k)\big)^c\right)\right) = E_k \cap \left(\bigcap_{j=1}^{\infty} \big(h^{-1}(F_j^k)\big)^c\right) \\
&= E_k \cap \left(\bigcup_{j=1}^{\infty} h^{-1}(F_j^k)\right)^c = E_k \setminus \bigcup_{j=1}^{\infty} h^{-1}(F_j^k),
\end{aligned}$$

where we denoted by $(\bullet)^c$ the complement of a set. We have demonstrated that;

$$h_k\left(h(E_k) \setminus \bigcup_{j=1}^{\infty} F_j^k\right) = E_k \setminus \bigcup_{j=1}^{\infty} D_j^k.$$

Therefore, from (iv.) follows

$$\begin{aligned}
\mathcal{L}^n\left(E_k \setminus \bigcup_{j=1}^{\infty} D_j^k\right) &= \mathcal{L}^n\left(h_k\left(h(E_k) \setminus \bigcup_{j=1}^{\infty} F_j^k\right)\right) \\
&\leq \big(\mathrm{Lip}(h_k)\big)^n \,\mathcal{L}^n\left(h(E_k) \setminus \bigcup_{k=1}^{\infty} F_j^k\right) = 0.
\end{aligned}$$

This, essentially, concludes the proof of Claim 3;

$$\begin{aligned}
\mathcal{L}^n\left(B \setminus \bigcup_{k,j=1}^{\infty} D_j^k\right) &\overset{\text{(i)}}{=} \mathcal{L}^n\left(\bigcup_{k=1}^{\infty} E_k \setminus \bigcup_{k,j=1}^{\infty} D_j^k\right) = \mathcal{L}^n\left(\bigcup_{k=1}^{\infty}\left(E_k \setminus \bigcup_{j=1}^{\infty} D_j^k\right)\right) \\
&\leq \sum_{k=1}^{\infty} \mathcal{L}^n\left(E_k \setminus \bigcup_{j=1}^{\infty} D_j^k\right) = 0.
\end{aligned}$$

Furthermore, it is easy to see that; Since $h|_{E_k}$ is one-to-one and $D_j^k \subseteq E_k$, for all $k = 1, 2, ...$ the map $h|_{D_j^k}$ is one-to-one.

`Claim 4:` For $k, j = 1, 2, ...$ we have

$$\mathrm{Lip}((S_j^k)^{-1} \circ (h|_{D_j^k})) \leq t, \ \ \mathrm{Lip}((h|_{D_j^k})^{-1} \circ S_j^k) \leq t \ \text{ and}$$
$$t^{-n}|\det S_j^k| \leq Jh|_{D_j^k} \leq t^n |\det S_j^k|.$$

Proof of claim: Observe that

$$\begin{aligned}\mathrm{Lip}((S_j^k)^{-1} \circ (h|_{D_j^k})) &= \mathrm{Lip}(R_j^k \circ (h|_{D_j^k})) \\ &\overset{\mathrm{h|_{E_k} = h_k^{-1}}}{\leq} \mathrm{Lip}(R_j^k \circ (h_k|_{F_j^k})^{-1}) \overset{(vi.)}{\leq} t.\end{aligned}$$

Also

$$\begin{aligned}\mathrm{Lip}((h|_{D_j^k})^{-1} \circ S_j^k) &= \mathrm{Lip}((h|_{D_j^k})^{-1} \circ (R_j^k)^{-1}) \\ &\overset{\mathrm{h|_{E_k}^{-1}} = h_k}{\leq} \mathrm{Lip}((h_k|F_j^k) \circ (R_j^k)^{-1}) \overset{(vi.)}{\leq} t.\end{aligned}$$

And recall that

$$Jh_k(h(x))Jh(x) = 1 \quad \mathcal{L}^n - a.e \text{ on } D_j^k.$$

Therefore

$$\begin{aligned}t^{-n}|\det S_j^k| = t^{-n}|\det (R_j^k)^{-1}| &= \frac{1}{t^n|\det R_j^k|} \\ &\overset{(vi.)}{\leq} \frac{1}{Jh_k|_{F_j^k}} = J(h_k|_{F_j^k})^{-1} \leq Jh|_{D_j^k} \\ &\overset{(iii.)}{\leq} \frac{t^n}{|\det R_j^k|} = t^n|\det S_j^k|.\end{aligned}$$

Finally, the $\{D_k\}_{k=1}^{\infty}$of the Lemma arise from a much-needed "re-branding" of $\{D_k^j\}_{k,j=1}^{\infty}$ following after the removal of those "few" points which do not fall into the last estimation. □

REMARK I. A keen observant would immediately notice the striking resemblance of this Lemma and Lemma 4.4. This is no coincidence, as Lemma 5.3 is also a Linearisation Lemma, in the sense we described in the previous chapter. Therefore, we turn our attention to *Remark III*, corollary of Lemma 4.4. It is natural to expect a similar estimate to hold here, as well.

Indeed; Since we invoked Lemma 4.4 on each h_k map ($k = 1, 2, ...$), we immediately get that; For all $x \in E_k$, we have

$$t^{-1}|R_j^k u| \leq |Dh_k(h(x))u| \leq t|R_j^k u| \ \ (u \in \mathbb{R}^n).$$

Thus

$$t^{-1}|(S_j^k)^{-1}u| \leq |Dh_k(h(x))u| \leq t|(S_j^k)^{-1}u|.$$

Consequently

$$t^{-1}|u| \leq |Dh_k(h(x)) \circ S_j^k u| \leq t|u| \ \ (u \in \mathbb{R}^n).$$

Therefore, in the laguage of the Operator norm, we get

$$\left\| Dh_k(h(x)) \circ S_j^k \right\| \leq t$$

and similarly;

$$\left\| (S_j^k)^{-1} \circ Dh_k(h(x))^{-1} \right\| \leq t.$$

However, we shall not forget that

$$Dh_k(h(x)) \circ Dh(x) = I, \ \ \mathcal{L}^n - a.e \text{ on } E_k.$$

Hence, for $\mathcal{L}^n$-a.e. $x \in D_j^k$, we have;

$$Dh(x) = Dh_k(h(x))^{-1}.$$

Consequently, we end up with

$$\left\| Dh(x)^{-1} \circ S_j^k \right\| \leq t$$

and

$$\left\| (S_j^k)^{-1} \circ Dh(x) \right\| \leq t.$$

REMARK II. Finally, after the "filtration" we performed on our notation in the end of the Lemma, the endgame of Remark I can be re-stated as;

For all $x \in D_k$ ($k = 1, 2, ...$), we have that

$$\left\| Dh(x)^{-1} \circ S_k \right\| \leq t$$

and

$$\left\| S_k^{-1} \circ Dh(x) \right\| \leq t.$$

5.2 The Coarea formula

Theorem 5.1 (**Coarea formula**)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous. Then for each $\mathcal{L}^n$-measurable subset $A \subseteq \mathbb{R}^n$,*

$$\int_A Jf\ dx = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)\, dy.$$

REMARK. It is obvious that the Coarea formula coincides with the Area formula for n=m.

Proof. In view of Lemma 5.3, we may as well assume $Df(x)$ and $Jf(x)$ exist for all $x \in A$. Also, without loss of generality, we will suppose $\mathcal{L}^n(A) < \infty$. We will proceed in steps.

`Case 1:` $A \subseteq \{Jf > 0\}$.

We define the following set of indicatrices;

$$\Lambda(n, n-m) \coloneqq \big\{\lambda : \{1, ..., n-m\} \to \{1, ..., n\} \mid \lambda : \text{strictly increasing}\big\}$$

and for each $\lambda \in \Lambda(n, n-m)$ the indexed projection $P_\lambda : \mathbb{R}^n \to \mathbb{R}^{n-m}$ as

$$P_\lambda(x_1, ..., x_n) = (x_{\lambda(1)}, ..., x_{\lambda(n-m)}).$$

The main idea here is quite interesting; We want to approximate f by its derivative. There, we will employ the Polar Decomposition theorem. The Orthogonal part, as we will see, does not contribute much to what is taking place. We will target the Symmetric part and we will try to extract it, using the following trick; For each $\lambda \in \Lambda(n, n-m)$ we "decompose" f as

$$f = q \circ h_\lambda,$$

where $h_\lambda : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^{n-m}$ and $q : \mathbb{R}^m \times \mathbb{R}^{n-m} \to \mathbb{R}^m$ are defined as

$$h_\lambda(x) \coloneqq \big(f(x), P_\lambda(x)\big) \quad (\, x \in \mathbb{R}^n\,)$$

and

$$q(y, z) \coloneqq y \quad (\, y \in \mathbb{R}^m \text{ and } z \in \mathbb{R}^{n-m}\,)$$

respectively. We denote

$$A_\lambda \coloneqq \{x \in A \mid \det Dh_\lambda \neq 0\}.$$

Expanding on h_λ we see that

$$h_\lambda(x) = \big(f(x), P_\lambda(x)\big) = \big(f_1(x), ..., f_m(x), x_{\lambda(1)}, ..., x_{\lambda(n-m)}\big).$$

Notice that;

$$A = \bigcup_{\lambda \in \Lambda(n,n-m)} A_\lambda.$$

Finally, we observe that the indicator set $\Lambda(n, n-m)$ is finite. This leaves us with a great advantage; We can simplify the framework of the problem, by demanding that A is a-priori contained in some set A_λ, namely that $A \subseteq A_\lambda$ for some $\lambda \in \Lambda(n, n-m)$.

Fix $t > 1$. By applying Lemma 5.3 to $h = h_\lambda$, we obtain Borel sets $\{D_k\}_{k=1}^{\infty}$, which we assume to be disjoint, and symmetric automorphisms $S_k : \mathbb{R}^n \to \mathbb{R}^n$ for which Assertions (*1.*)-(*2.*)-(*3.*) of that Lemma hold true.

Set $G_k \coloneqq A \cap D_k$.

`Claim 1:`

$$t^{-n} \left[\!\left[q \circ S_k \right]\!\right] \leq Jf|_{G_k} \leq t^n \left[\!\left[q \circ S_k \right]\!\right].$$

Proof of claim: Our previous "decomposition" of f, implies that; For $\mathcal{L}^n$-a.e we get

$$\begin{aligned} Df = D(q \circ h) &= q \circ Dh \\ &= q \circ S_k \circ S_k^{-1} \circ Dh \\ &= q \circ S_k \circ C. \end{aligned}$$

where $C \coloneqq S_k^{-1} \circ Dh$.

From Remark II of Lemma 5.3, we deduce that

$$\left\| C^{-1} \right\| \leq t \ \text{ and } \ \|C\| \leq t$$

on D_k, therefore on G_k as well. Interpreting the Operator norm, we obtain that

$$t^{-1}|u| \leq |Cu| \leq t|u| \text{ on } G_k \ (u \in \mathbb{R}^n). \tag{$\star$}$$

Employing the Polar Decomposition Theorem for Df, $q \circ S_k : \mathbb{R}^n \to \mathbb{R}^m$, we get

$$Df = S \circ O^* \ \text{ and } \ q \circ S_k = T \circ P^*.$$

where $S, T : \mathbb{R}^m \to \mathbb{R}^m$ are symmetric and $O, P : \mathbb{R}^m \to \mathbb{R}^n$ orthogonal maps. Consequently,

$$S \circ O^* = T \circ P^* \circ C, \tag{$\star\star$}$$

hence
$$S = T \circ P^* \circ C \circ O.$$
Since $G_k \subseteq A \subseteq \{Jf > 0\}$ we have that $\det S \neq 0$, thus $\det T \neq 0$. Therefore, for $u \in \mathbb{R}^m$, we get that
$$\begin{aligned} |\, T^{-1} \circ Su \,| &= |\, P^* \circ C \circ Ou \,| \\ &\overset{3.9}{\leq} |\, C \circ Ou \,| \\ &\overset{(\star)}{\leq} t|Ou| \\ &= t|u|. \end{aligned}$$
This implies
$$(T^{-1} \circ S)(B(1)) \subseteq B(t),$$
and so, passing with Lebesgue measures on the inequality, we get
$$|\det T^{-1} \circ S| \leq t^n.$$
Moreover, it is easy to see that
$$\begin{aligned} \left[\!\left[q \circ S_k \right]\!\right]^2 = \left[\!\left[T \circ P^* \right]\!\right]^2 &= \det T \circ P^* \circ (T \circ P^*)^* = \det T \circ P^* \circ P \circ T^* \\ &= \det T \circ I_m \circ T^* = \det T^2. \end{aligned}$$
Consequently
$$\begin{aligned} Jf = |\det S\,| = |\det T \circ (T^{-1} \circ S)\,| &= |\det T|\,|\det T^{-1} \circ S| \\ &\leq t^n |\det T\,| = t^n \left[\!\left[q \circ S_k \right]\!\right]. \end{aligned}$$
Similarly, we have that
$$\begin{aligned} S^{-1} = \big(T \circ P^* \circ C \circ O\big)^{-1} &= O^{-1} \circ C^{-1} \circ (P^*)^{-1} \circ T^{-1} \\ &= O^* \circ C^{-1} \circ P \circ T^{-1}, \end{aligned}$$
thus, for $u \in \mathbb{R}^m$, we have that
$$\begin{aligned} |\, S^{-1} \circ Tu \,| = |\, O^* \circ C^{-1} \circ Pu \,| &\overset{3.9}{\leq} |\, C^{-1} \circ Pu \,| \\ &\overset{(\star)}{\leq} t|Pu| \\ &= t|u|. \end{aligned}$$
Hence, by mimicking the calculations above, we end up with the estimate
$$\left[\!\left[q \circ S_k \right]\!\right] = |\det T\,| \leq t^n |\det S\,| = t^n\, Jf.$$
Thus, completing the proof of our Claim.

We continue with some calculations.

$$
\begin{aligned}
&t^{-3n+m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(G_k \cap f^{-1}\{y\}\big)\ dy \\
&= t^{-3n+m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(G_k \cap (q \circ h)^{-1}\{y\}\big)\ dy \\
&= t^{-3n+m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(G_k \cap h^{-1} \circ q^{-1}\{y\}\big)\ dy \\
&= t^{-3n+m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(h^{-1}(h(G_k) \cap q^{-1}\{y\})\big)\ dy \\
&= t^{-3n+m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big((h^{-1} \circ S_k)\big(S_k^{-1}(h(G_k) \cap q^{-1}\{y\})\big)\big)\ dy \\
&\underset{\text{Theorem 3.2}}{\overset{\text{Lemma 5.3}}{\leq}} t^{-3n+m}\, t^{n-m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(S_k^{-1}(h(G_k) \cap q^{-1}\{y\})\big)\ dy \\
&= t^{-2n} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big((S_k^{-1} \circ h)(G_k) \cap (S_k^{-1} \circ q^{-1})\{y\}\big)\ dy \\
&= t^{-2n} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big((S_k^{-1} \circ h)(G_k) \cap (q \circ S_k)^{-1}\{y\}\big)\ dy \\
&\overset{\text{Lemma 5.1}}{=} t^{-2n} \left[\!\left[q \circ S_k \right]\!\right] \mathcal{L}^n\big((S_k^{-1} \circ h)(G_k)\big) \\
&\underset{\text{Theorem 3.2}}{\overset{\text{Lemma 5.3}}{\leq}} t^{-2n}\, t^n \left[\!\left[q \circ S_k \right]\!\right] \mathcal{L}^n(G_k) \quad \underset{\text{we have that } \mathcal{H}^n = \mathcal{L}^n}{\text{since for } G_k \subseteq \mathbb{R}^n} \\
&= t^{-n} \left[\!\left[q \circ S_k \right]\!\right] \mathcal{L}^n(G_k) \\
&\overset{\texttt{Claim 1}}{\leq} \int_{G_k} Jf\ dx.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\int_{G_k} Jf\ dx \\
&\overset{\texttt{Claim 1}}{\leq} \int_{G_k} t^n \left[\!\left[q \circ S_k \right]\!\right]\ dx \\
&= t^n \left[\!\left[q \circ S_k \right]\!\right] \mathcal{L}^n(G_k) \\
&= t^n \left[\!\left[q \circ S_k \right]\!\right] \mathcal{L}^n\big((S_k^{-1} \circ h)^{-1}\big((S_k^{-1} \circ h)(G_k)\big)\big) \\
&= t^n \left[\!\left[q \circ S_k \right]\!\right] \mathcal{L}^n\big((h^{-1} \circ S_k)\big((S_k^{-1} \circ h)(G_k)\big)\big) \\
&\overset{\text{Theorem 3.2}}{\leq} t^{2n} \left[\!\left[q \circ S_k \right]\!\right] \mathcal{L}^n\big((S_k^{-1} \circ h)(G_k)\big) \\
&\overset{\text{Lemma 5.1}}{=} t^{2n} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big((S_k^{-1} \circ h)(G_k) \cap (q \circ S_k)^{-1}\{y\}\big)\ dy
\end{aligned}
$$

$$
\begin{aligned}
&= t^{2n}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\Big((S_k^{-1}\circ h)\big(G_k\cap(h^{-1}\circ q^{-1})\{y\}\big)\Big)\,dy\\
&\underset{\text{Lemma 5.3}}{\overset{\text{Theorem 3.2}}{\leq}} t^{2n}\,t^{n-m}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(G_k\cap(q\circ h)^{-1}\{y\}\big)\,dy\\
&= t^{3n-m}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(G_k\cap f^{-1}\{y\}\big)\,dy.
\end{aligned}
$$

Eventually, we have derived the following estimate

$$
\begin{aligned}
t^{-3n+m}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(G_k\cap f^{-1}\{y\}\big)\,dy &\leq \int_{G_k} Jf\,dx\\
&\leq t^{3n-m}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(G_k\cap f^{-1}\{y\}\big)\,dy.
\end{aligned}
$$

Now, taking into account that

$$
\mathcal{L}^n\left(A\setminus\bigcup_{k=1}^{\infty}G_k\right)=0.
$$

which stems from the initial invocation of Lemma 5.3, and that the sets $\{G_k\}_{k=1}^{\infty}$ are constructed to be disjoint, we can sum on k and, finally, let $t\to 1^+$. Thus, we conclude that

$$
\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\left(\bigcup_{k=1}^{\infty}G_k\cap f^{-1}\{y\}\right)dy=\int_{\bigcup_{k=1}^{\infty}G_k}Jf\,dx.
$$

Moreover, employing Eilenberg's Inequality (Lemma 5.3) again, gives us;

$$
\begin{aligned}
&\int_{\mathbb{R}^n}\mathcal{H}^{n-m}\left(\left(A\setminus\bigcup_{k=1}^{\infty}G_k\right)\cap f^{-1}\{y\}\right)dy\\
&\qquad\leq\frac{\alpha(n-m)\alpha(m)}{\alpha(n)}(\mathrm{Lip}(f))^m\mathcal{L}^n\left(A\setminus\bigcup_{k=1}^{\infty}G_k\right)=0.
\end{aligned}
$$

Hence, we conclude that

$$
\begin{aligned}
\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(A\cap f^{-1}\{y\}\big)\,dy &= \int_{\mathbb{R}^n}\mathcal{H}^{n-m}\left(\left(A\setminus\bigcup_{k=1}^{\infty}G_k\right)\cap f^{-1}\{y\}\right)dy\\
&\quad+\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\left(\bigcup_{k=1}^{\infty}G_k\cap f^{-1}\{y\}\right)dy
\end{aligned}
$$

$$
\begin{aligned}
&= 0 + \int_{\bigcup_{k=1}^{\infty} G_k} Jf \; dx \\
&= \int_{A \setminus \bigcup_{k=1}^{\infty} G_k} Jf \; dx + \int_{\bigcup_{k=1}^{\infty} G_k} Jf \; dx \\
&= \int_A Jf \; dx.
\end{aligned}
$$

`Case 2:` $A \subseteq \{Jf = 0\}$.
We fix $0 < \varepsilon \leq 1$ and define maps $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and $p : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ as

$$g(x, y) \coloneqq f(x) + \varepsilon y \;\text{ and }\; p(x, y) \coloneqq y.$$

Then

$$Dg = \big(Df, \varepsilon I\big)_{m \times (n+m)}$$

and we have the following estimate

$$\varepsilon^m \leq Jg = \llbracket Dg \rrbracket = \llbracket Dg^* \rrbracket \leq C\varepsilon,$$

where C is a constant, analogous to the one we calculated in **Claim 2** of the proof of the Area formula (Theorem 4.5).

`Claim 2:` For $y, w \in \mathbb{R}^m$, we define a $B \coloneqq A \times B(1) \subseteq \mathbb{R}^{n+m}$. We have that

$$B \cap g^{-1}\{y\} \cap p^{-1}\{w\} = \begin{cases} \varnothing & if \; w \notin B(1) \\ \big(A \cap f^{-1}\{y - \varepsilon w\}\big) \times \{w\} & if \; w \in B(1). \end{cases}$$

Proof of claim: We have $(x, z) \in B \cap g^{-1}\{y\} \cap p^{-1}\{w\}$ if and only if

$$(x, z) \in B \;\text{ and }\; g(x, z) = y \;\text{ and }\; p(x, z) = w,$$

which implies

$$x \in A, \; z \in B(1), \; f(x) + \varepsilon z = y \;\text{ and }\; z = w,$$

and so

$$x \in A, \; z = w \in B(1) \;\text{ and }\; f(x) = y - \varepsilon w,$$

thus

$$w \in B(1) \;\text{ and }\; (x, z) \in \big(A \cap f^{-1}\{y - \varepsilon w\}\big) \times \{w\}.$$

Consequently, for all $(y,w)\in\mathbb{R}^m\times\mathbb{R}^m$, we get the following;

$$\chi_{B(0,1)}(w)\cdot\mathcal{H}^{n-m}\Big(\big(A\cap f^{-1}\{y-\varepsilon w\}\big)\times\{w\}\Big)=\mathcal{H}^{n-m}\big(B\cap g^{-1}\{y\}\cap p^{-1}\{w\}\big)$$

Now, we are able to compute that

$$\begin{aligned}
&\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(A\cap f^{-1}\{y\}\big)\,dy\\
&=\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(A\cap f^{-1}\{y-\varepsilon w\}\big)\,dy\quad\text{for all } w\in\mathbb{R}^m\\
&=\frac{1}{\alpha(m)}\int_{B(0,1)}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(A\cap f^{-1}\{y-\varepsilon w\}\big)\,dy\,dw\\
&=\frac{1}{\alpha(m)}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\chi_{B(0,1)}(w)\cdot\mathcal{H}^{n-m}\big(A\cap f^{-1}\{y-\varepsilon w\}\big)\,dy\,dw\\
&\overset{\text{Fubini}}{=}\frac{1}{\alpha(m)}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\chi_{B(0,1)}(w)\cdot\mathcal{H}^{n-m}\big(A\cap f^{-1}\{y-\varepsilon w\}\big)\,dw\,dy.
\end{aligned}$$

We continue our calculations;

$$\begin{aligned}
&\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(A\cap f^{-1}\{y\}\big)\,dy\\
&\quad=\frac{1}{\alpha(m)}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\chi_{B(0,1)}(w)\cdot\mathcal{H}^{n-m}\Big(\big(A\cap f^{-1}\{y-\varepsilon w\}\big)\times\{w\}\Big)\,dw\,dy\\
&\quad\overset{\text{Claim 2}}{=}\frac{1}{\alpha(m)}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(B\cap g^{-1}\{y\}\cap p^{-1}\{w\}\big)\,dw\,dy\\
&\quad\underset{\text{inequality}}{\overset{\text{Eilenberg's}}{\leq}}\frac{1}{\alpha(m)}\frac{\alpha(n-m)\alpha(m)}{\alpha(n)}\int_{\mathbb{R}^m}\mathcal{H}^{n}\big(B\cap g^{-1}\{y\}\big)\,dy\\
&\quad=\frac{\alpha(n-m)}{\alpha(n)}\int_{\mathbb{R}^m}\mathcal{H}^{n}\big(B\cap g^{-1}\{y\}\big)\,dy\\
&\quad\overset{\texttt{Case 1}}{=}\frac{\alpha(n-m)}{\alpha(n)}\int_B Jg\,dx\,dz\\
&\quad\leq\frac{\alpha(n-m)}{\alpha(n)}\int_B\sup_B Jg\,dx\,dz\\
&\quad=\frac{\alpha(n-m)}{\alpha(n)}\mathcal{L}^{n+m}(B)\sup_B Jg\\
&\quad=\frac{\alpha(n-m)}{\alpha(n)}\mathcal{L}^{n}(A)\alpha(m)\sup_B Jg
\end{aligned}$$

$$= \frac{\alpha(n-m)\alpha(m)}{\alpha(n)} \mathcal{L}^n(A) \sup_B Jg$$
$$\leq \widetilde{C}\varepsilon.$$

where $\widetilde{C} = \dfrac{\alpha(n-m)\alpha(m)C\mathcal{L}^n(A)}{\alpha(n)}$ is constant.
Letting $\varepsilon \to 0$, gives us

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(A \cap f^{-1}\{y\}\big)\, dy = 0 = \int_A Jf\, dx.$$

`Case 3:` $A \subseteq \{Jf \geq 0\}$ for every $x \in A$.

In the general case, we write $A = A_1 \cup A_2$, with $A_1 \subseteq \{Jf > 0\}$ and $A_1 \subseteq \{Jf = 0\}$ and employ Cases 1 and 2 as above. □

Fubini-Tonelli's analogue in Curvilinear Coordinates

Theorem 5.2 (Change of Variables). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz function. Then for each $\mathcal{L}^n$-summable function $g : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\int_{\mathbb{R}^n} g(x)Jf(x)\,dx = \int_{\mathbb{R}^m}\left[\int_{x\in f^{-1}\{y\}} g\,d\mathcal{H}^{n-m}\right] dy.$$

Proof. We will proceed in steps.
`Case 1:` $g \geq 0$. We recall that for such a function g, by Theorem 1.10 we get the following expression

$$g = \sum_{i=1}^{\infty} \frac{1}{i}\chi_{A_i}$$

for appropriate $\mathcal{L}^n$-measurable sets $\{A_i\}_{i=1}^{\infty}$. Then we employ the Monotone Convergence Theorem and the Coarea formula, and thus we get

$$\begin{aligned}
\int_{\mathbb{R}^n} g(x)Jf(x)\,dx &= \int_{\mathbb{R}^n}\left(\sum_{i=1}^{\infty}\frac{1}{i}\chi_{Ai}(x)\right)Jf(x)\,dx = \\
&= \sum_{i=1}^{\infty}\frac{1}{i}\int_{\mathbb{R}^n}\chi_{A_i}Jf\,dx \\
&= \sum_{i=1}^{\infty}\frac{1}{i}\int_{A_i}Jf\,dx \\
&= \sum_{i=1}^{\infty}\frac{1}{i}\int_{\mathbb{R}^m}\mathcal{H}^{n-m}\big(A_i\cap f^{-1}\{y\}\big)\,dy \\
&= \int_{\mathbb{R}^m}\sum_{i=1}^{\infty}\frac{1}{i}\,\mathcal{H}^{n-m}\big(A_i\cap f^{-1}\{y\}\big)\,dy \\
&\overset{\text{Lemma 5.2}}{=} \int_{\mathbb{R}^m}\int g\,\chi_{f^{-1}\{y\}}\,d\mathcal{H}^{n-m}(x)\,dy \quad \text{since } A_i\cap f^{-1}\{y\} \text{ is } \mathcal{H}^{n-m}\text{-measurable} \\
&= \int_{\mathbb{R}^m}\left[\int_{x\in f^{-1}\{y\}} g\,d\mathcal{H}^{n-m}\right] dy.
\end{aligned}$$

`Case 2:` Let now, in favor of generality, g be any $\mathcal{L}^n$-summable function.

Simply, we write $g = g^+ - g^-$ and apply Case 1 on g^+ and g^-. □

5.3 Applications

THEOREM A. (Polar coordinates) Assume $g : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{L}^n$-summable. Then

$$\int_{\mathbb{R}^n} g\, dx = \int_0^{\infty} \left(\int_{\partial B(r)} g\, d\mathcal{H}^{n-1} \right) dr.$$

More specifically,

$$\frac{d}{dr}\left(\int_{B(r)} g\, dx \right) = \int_{\partial B(r)} g\, d\mathcal{H}^{n-1}$$

for $\mathcal{L}^1$-a.e. $r > 0$

Proof: Let $f : \mathbb{R}^n \to \mathbb{R}$ be $f(x) = \|x\|$; Then, for $x \neq 0$, we have that

$$Df(x) = \frac{x}{\|x\|}.$$

Therefore

$$Jf = \sqrt{(Df) \cdot (Df)^T} = \sqrt{\sum_{i=1}^{n} \left(\frac{x_i^2}{\|x\|} \right)} = 1.$$

Thus, Theorem 5.2 gives us

$$\begin{aligned}
\int_{\mathbb{R}^n} g(x) \cdot 1\, dx &= \int_{\mathbb{R}} \left[\int_{f(x)=y} g\, d\mathcal{H}^{n-1} \right] dy \\
&= \int_{\mathbb{R}} \left[\int_{\|x\|=y} g\, d\mathcal{H}^{n-1} \right] dy \\
&= \int_0^{\infty} \left[\int_{x \in \partial B(y)} g\, d\mathcal{H}^{n-1} \right] dy.
\end{aligned}$$

Taking $f|_{B(r)} : B(r) \subseteq \mathbb{R}^n \to [0, r]$, proves the second Assertion.

THEOREM B. (Integration over level sets.) Assume $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous. Then

1. $$\int_{\mathbb{R}^n} |Df|\,dx = \int_{-\infty}^{\infty} \mathcal{H}^{n-1}\big(\{f = t\}\big)\,dt.$$

2. Assume also $\operatorname{ess\,inf} |Df| > 0$, and take a function $g : \mathbb{R}^n \to \mathbb{R}$ to be $\mathcal{L}^n$-summable. Then
$$\int_{\{f>t\}} g\,dx = \int_t^{\infty} \left(\int_{\{f=s\}} \frac{g}{|Df|}\,d\mathcal{H}^{n-1} \right) ds.$$

3. Moreover,
$$\frac{d}{dt}\left(\int_{\{f>t\}} g\,dx \right) = -\int_{\{f=t\}} \frac{g}{|Df|}\,d\mathcal{H}^{n-1}$$
for $\mathcal{L}^1$-a.e. $t \in \mathbb{R}$.

Proof: (*1.*) Since $Jf = |Df|$, Coarea formula implies immediately that
$$\int_{\mathbb{R}^n} |Df|\,dx = \int_{\mathbb{R}} \mathcal{H}^{n-1}\big(\mathbb{R}^n \cap f^{-1}\{y\}\big)\,dy = \int_{-\infty}^{\infty} \mathcal{H}^{n-1}\big(\{f = t\}\big)\,dt.$$

(*2.*) We consider sets $E_t \coloneqq \{f > t\}$ and we employ Theorem 5.2 to get that
$$\begin{aligned}
\int_{\{f>t\}} g\,dx &= \int_{\mathbb{R}^n} \chi_{E_t} \frac{g}{|Df|} |Df|\,dx \\
&= \int_{\mathbb{R}^n} \left(\chi_{E_t} \frac{g}{|Df|} \right) Jf\,dx \\
&= \int_{-\infty}^{\infty} \left(\int_{\partial E_s} \frac{g}{|Df|} \chi_{E_t}\,d\mathcal{H}^{n-1} \right) ds \\
&= \int_t^{\infty} \left(\int_{\partial E_s} \frac{g}{|Df|}\,d\mathcal{H}^{n-1} \right) ds.
\end{aligned}$$

(*3.*) We simply differentiate both sides.

THEOREM C. (Level sets of distance functions.) Let $K \subseteq \mathbb{R}^n$ be a (non-empty) compact set. As usual, we denote by

$$d(x) \coloneqq \operatorname{dist}(x, K)$$

the distance function of a point $x \in \mathbb{R}^n$ from K. Then for each $0 < \alpha < b$ we have

$$\int_\alpha^b \mathcal{H}^{n-1}\big(\{d = t\}\big)\, dt = \mathcal{L}^n\big(\{x \mid \alpha \leq d(x) \leq b\}\big).$$

Proof: For a given $x \in \mathbb{R}^n$, since K is a compact subset of $\mathbb{R}^n$, hence closed and bounded, we denote by c the element from K for which the distance is attained, i.e.

$$d(x) = \operatorname{dist}(x, K) = |x - c|.$$

Thus, for any other point $y \in \mathbb{R}^n$, we get that

$$\begin{aligned} d(y) - d(x) = \operatorname{dist}(y, K) - |x - c| &= \inf_{k \in K}\{|y - k|\} - |x - c| \\ &\leq |y - c| - |x - c| \leq |(y - c) - (x - c)| = |y - x|. \end{aligned}$$

Working in a symmetrical way, interchanging the roles of x and y, we get, eventually, that

$$|d(y) - d(x)| \leq |y - x|.$$

Consequently,

$$\operatorname{Lip}(d) \leq 1,$$

and so, from Rademacher's Theorem, it follows that the distance function is $\mathcal{L}^n$-a.e. differentiable.

Observe that, for any point x outside of K at which $Dd(x)$ exists, we get from the definition of the derivative, that $|Dd(x)| \leq 1$. Moreover

$$d(tx + (1 - t)c) = |tx + (1 - t)c - c| = t|x - c|$$

for all $t \in [0, 1]$ and $c \in K$ as above. Now, from the differentiability of d, we have that

$$d(x) = d(c) + Dd(x) \cdot |x - c| + o(|x - c|),$$

and so

$$|x - c| = Dd(x) \cdot |x - c| \overset{C-S}{\leq} |Dd(x)||x - c|.$$

Thus

$$|Dd(x)| \geq 1.$$

Hence,

$$|Dd(x)| = 1 \quad \mathcal{L}^n - a.e. \text{ in } \mathbb{R}^n \setminus K.$$

Finally, we employ Theorem B. from above (Integration over level sets) and the results follows immediately once we restrict ourselves on the domain where $0 < \alpha \leq \text{dist}(\bullet, K) \leq b$.

5.3.1 Crofton's formula

Let $\mathbb{O}^*(n,m)$ denote the set of orthogonal projections P of $\mathbb{R}^n$ onto m-dimensional subspaces. For topological reasons[6], there exists a unique probability measure γ on $\mathbb{O}^*(n,m)$ which is invariant under Euclidean motions.

For any Borel set B,we define the so-called **integral-geometric measure** as

$$\mathcal{I}^m(B) \coloneqq \frac{1}{\beta(n,m)} \int_{P \in \mathbb{O}^*(n,m)} \int_{y \in Image(P) \equiv \mathbb{R}^m} \mathcal{H}^0\big(B \cap P^{-1}\{y\}\big)\, d\mathcal{L}^m(y)\, d\gamma(P)$$

where $\beta(n,m)$ is a normalising constant defined as

$$\beta(n,m) = \frac{\Gamma\left(\dfrac{m+1}{2}\right)\Gamma\left(\dfrac{n-m+1}{2}\right)}{\Gamma\left(\dfrac{n+1}{2}\right)\sqrt{\pi}}.$$

Furthermore, a set $E \subseteq \mathbb{R}^n$ will be called **m-dimensional rectifiable**, if there is a countable family of Lipschitz maps $f_i : \mathbb{R}^m \to \mathbb{R}^n$ for which $\mathcal{H}^m(E) < \infty$ and

$$\mathcal{H}^m\left(E \setminus \bigcup_{i=1}^{\infty} f(\mathbb{R}^m)\right) = 0.$$

THEOREM D1. (Crofton's formula)

For an m-dimensional rectifiable set A, its integral-geometric measure is equal to its $\mathcal{H}^m$-measure, namely

$$\mathcal{I}^m(A) = \mathcal{H}^m(A).$$

We will now proceed a step further, into some more general settings. Again, the results we state spring from the Coarea formula, yet a solid substantiation requires highly advanced tools of Algebraic and Geometric nature, such as the double fibration technique, as well as some "heavy" notions from Integral Geometry and Integral Calculus on Manifolds, hence reaching far beyond the scopes of the present thesis.

[6]See [15] for a detailed explanation.

Denote by $\mathbf{Graff}^1(\mathbb{R}^n)$ the set of all affine hyperplanes in $\mathbb{R}^n$ and by $\mathbf{Graff}^{n-1}(\mathbb{R}^n)$ the set of affine lines in $\mathbb{R}^n$. Then we get the following;

THEOREM D2. (Crofton's formula for curves)
Let H be an affine hyperplane $H \in \mathbf{Graff}^1(\mathbb{R}^n)$ and take C a simple closed C^2-differentiable curve, parameterised by arclength. Then, the function

$$\mathbf{Graff}^1(\mathbb{R}^n) \ni H \mapsto \mathcal{H}^0\big(H \cap C\big)$$

is measurable and

$$\int_{\mathbf{Graff}^1(\mathbb{R}^n)} \mathcal{H}^0\big(H \cap C\big)\, |dV_g(H)| = \alpha(n-1)\,\text{length of } C = \frac{\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n+1}{2})}\mathcal{H}^1(C)$$

where $|dV_g|$ is the volume density associated with a suitable metric g on $\mathbf{Graff}^1(\mathbb{R}^n)$.

THEOREM D3. (Crofton's formula for sub-manifolds)
Let L an affine line $L \in \mathbf{Graff}^{n-1}(\mathbb{R}^n)$ and M a (n-1)-dimensional sub-manifold of $\mathbb{R}^n$. Then, the function

$$\mathbf{Graff}^{n-1}(\mathbb{R}^n) \ni L \mapsto \mathcal{H}^0\big(L \cap M\big)$$

is measurable and

$$\int_{\mathbf{Graff}^{n-1}(\mathbb{R}^n)} \mathcal{H}^0\big(L \cap M\big)\, |d\hat{\mu}| = \frac{\pi^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n+1}{2}\right)}\, Vol_{n-1}(M)$$

where $|d\hat{\mu}|$ is defined appropriately, in order to coincide with the density on $\mathbf{Graff}^{n-1}(\mathbb{R}^n)$.

Notice that, for $n = 2$, as curve in $\mathbb{R}^2$ can be regarded as a co-dimension 1 submanifold of $\mathbb{R}^2$, thus "bringing together" the preceding two Theorems. Hence, we get the following result

COROLLARY. (Crofton's formula in $\mathbb{R}^2$)
Let C be a curve of $\mathbb{R}^2$. Then

$$\text{length of } C = \frac{1}{2}\int_{\mathrm{L} \in \mathbf{Graff}^1(\mathbb{R}^2)} \mathcal{H}^0\big(L \cap C\big)$$

This essentially means that we can relate the length of a curve to the expected number of times a "random" line intersects it.

5.3.2 Sard-type Corollaries

THEOREM. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz function.
`i.` If $n \leq m$, applying the **Area formula** to the set $E := \big\{x \in \mathbb{R}^n \mid Jf(x) = 0\big\} = \{Jf = 0\}$, results in

$$\int_{\mathbb{R}^m} \mathcal{H}^0\big(E \cap f^{-1}\{y\}\big)\, d\mathcal{H}^n(y) = 0.$$

This implies $\mathcal{H}^0\big(E \cap f^{-1}\{y\}\big) = 0$, therefore $f(E) \cap \{y\} = \varnothing$, for $\mathcal{H}^n$-a.e. $y \in \mathbb{R}^m$. Consequentially, $\mathcal{H}^n\big(f(E)\big) = 0$ and thus $Jf > 0$ on $f^{-1}\{y\}$ for $\mathcal{H}^n$-a.e. $y \in \mathbb{R}^m$.
`ii.` If $n \geq m$, then the **Coarea formula** applied on $E = \{Jf = 0\}$ implies that

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}\big(E \cap f^{-1}\{y\}\big)\, dy = 0.$$

Consequently, $\mathcal{H}^{n-m}\big(E \cap f^{-1}\{y\}\big) = 0$ for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$. Hence, $Jf > 0$ $\mathcal{H}^{n-m}$-a.e. on $f^{-1}\{y\}$ for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$.

The above theorem is a weak variant of the **Morse-Sard Theorem**, which we will state right away, after establishing some preliminary definitions;

DEFINITIONS. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be an arbitrary function. A point $x \in \mathbb{R}^n$ is said to be a **critical point**, if $Df(x)$ is not of maximum rank. Equivalently, when $Jf(x) = 0$. A point $y = f(x)$ is said to be a **critical value**, when x is a critical point of f.

The "classical" **Morse-Sard Theorem** states the following;

THEOREM. (Morse-Sard) Let $f : \mathbb{R}^n \to \mathbb{R}^m$. We distinguish two cases;
`i.` If $n < m$ and f is of class C^1, then the set of critical values has $\mathcal{L}^m$-measure zero.
`ii.` If $n \geq m$ and f is of class at-least C^{n-m+1} or higher, then the set of critical values is a set of $\mathcal{L}^m$-measure zero.

REMARK. Let it be noted that the condition in (ii.) cannot be weakened, as it is possible to construct functions not smooth enough, that hold a set of critical values of positive measure. This highlights the importance of the weakened variant stated above, because the only requirement for f is to be a Lipschitz function.

5.3.3 An Application in sample distribution theory

PRELIMINARIES. Let (Ω, Σ, p) be a probability space, consisted of a **sample space** Ω, a σ-algebra $\Sigma \subseteq 2^{\Omega}$ called **events** and a countably additive probability measure p. Take X to be a (vector valued) **random variable**, i.e. a Σ-measurable map $X : \Omega \to \mathbb{R}^n$. X is sometimes called the **data**.

Let Y be any measurable function of the data X, namely, Y is a random variable, defined as $Y = \phi(X)$ for some function $\phi : \mathbb{R}^n \to \mathbb{R}^m$. Now, Y is often called a **statistic**. One problem somebody addresses in Sample Distribution Theory is finding the probability distribution of the statistic Y knowing the distribution of X.

THEOREM. Let (Ω, Σ, p) be a probability space and $n, m \in \mathbb{N}$ with $n > m$. Consider a random variable $X : \Omega \to \mathbb{R}^n$, which is absolutely continuous to the Lebesgue measure, i.e. if p_X is the distribution of X, then $p_X << \mathcal{L}^n$, having a probability density function f_X. Take $\phi : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz map with a differential $D\phi$ of maximum rank a.e..

Then the statistic $Y = \phi(X)$ is again an absolutely continuous to the $\mathcal{L}^m$-measure random variable, having probability density function f_Y given as

$$f_Y(y) = \int_{\phi^{-1}\{y\}} f_X(x) \frac{1}{J\phi(x)} \, d\mathcal{H}^{m-n}(x) \text{ for } \mathcal{L}^m - a.e. \; y \in \phi(\mathbb{R}^n)$$

and is 0 elsewhere.

Proof: Let $A \subseteq \mathbb{R}^m$. Since ϕ is Lipschitz mapping and its differential has maximum rank, we get, as we saw in the Sard-type Corollary earlier, that $J\phi > 0$ $\mathcal{H}^{n-m}$-a.e. on $f^{-1}\{y\}$, for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$. Hence

$$\begin{aligned} p_Y(A) = p\big(Y^{-1}(A)\big) &= p(X^{-1}(\phi^{-1}(A))) \\ &= \int_{\phi^{-1}(A)} f_X(x) \, dx \\ &= \int_{\phi^{-1}(A)} \frac{f_X(x)}{J\phi(x)} J\phi(x) \, dx, \\ &\qquad \text{where by employing the Coarea formula we get} \\ &= \int_{\phi(\phi^{-1}(A))} \int_{x \in \phi^{-1}\{y\}} \frac{f_X(x)}{J\phi(x)} \, d\mathcal{H}^{n-m}(x) \, dy \\ &= \int_A \int_{x \in \phi^{-1}\{y\}} \frac{f_X(x)}{J\phi(x)} \, d\mathcal{H}^{n-m}(x) \, dy. \end{aligned}$$

□

References and notes

The primary source for this Chapter has been the book of Evans & Gariepy [8, 7]. Our goal throughout this Thesis was to shed plenty of light on those fine concepts of all the techniques and ideas we employ in our journey, in a way that the material could be comprehended in-depth by our readers. Therefore, we shall also point-out, once more, to [12] and [9].

In our brief paragraph of Crofton's Formula and some of it's generalised results, we have consulted [16] and [23], alongside with [15]. The definition of *m-dimensional Rectifiable set* is differentiated slightly from the "original" one, given by Federer in [10], and resembles more the one found in [21] or [20], corresponding to what Federer would call a countably $(\mathcal{H}^m, \mathbb{R}^n)$-rectifiable set.

For the Sard-type Corollaries, we have consulted D.W.M. van Dijk's exposition in [29]. At last, and in order to demonstrate the vast spectrum of the Applications of Coarea Formula, we have included a result from Sample Distribution Theory, presented in [22].

Bibliography

[1] Apostol, T. M. *Calculus: One-variable calculus, with an introduction to linear algebra.* Calculus. Wiley, 1967.

[2] Axler, S. *Linear algebra done right.* Springer Nature, 2015.

[3] Bretscher, O. *Linear algebra with applications.* Prentice Hall Eaglewood Cliffs, NJ, 2018.

[4] Briggs, J., and Tyree, T. Hausdorff measure, 2016. (Report).

[5] Cannarsa, P., and D'Aprile, T. Lecture notes on measure theory and functional analysis, 2006/07.

[6] Carothers, N. L. *Real analysis.* Cambridge University Press, 2000.

[7] Evans, L. C., and Gariepy, R. F. *Measure Theory and Fine Properties of Functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, Florida, 1992.

[8] Evans, L. C., and Gariepy, R. F. *Measure Theory and Fine Properties of Functions - Revised Edition.* Studies in Advanced Mathematics. CRC Press, Boca Raton, Florida, 2015.

[9] Fanghua, L., and Xiaoping, Y. *Geometric measure theory: an introduction.* Science Press, Beijing, China, 2002. International Press, Cambridge, U.S.A., 2002.

[10] Federer, H. *Geometric measure theory.* Springer-Verlag, 1969.

[11] Filipponi, A., and Martucci, D. La formula di coarea e alcune sue applicazioni. (Lecture Notes).

[12] Giaquinta, M., and Modica, G. *Mathematical Analysis: Foundations and Advanced Techniques for Functions of Several Variables*. Birkhäuser, 2011. (translated and revised from: M. Giaquinta, G. Modica Analisi Matematica, V. Funzioni di pi'u variabili: ulteriori sviluppi, Pitagora Ed., Bologna, 2005.).

[13] Hardt, R. An introduction to geometric measure theory. Lectures notes. Melbourne University, 1979.

[14] Holopainen, I. Geometric measure theory. Lectures held by Ilkka Holopainen in the Academic Year 2016. Last modified: 8/5/2021.

[15] Inversi, M. Lecture notes of geometric measure theory. Free reworking of the lectures held by Professor Giovanni Alberti in the Academic Year 2019/2020, Università di Piza, Dipartimento di Matematica. Last modified: 27/12/2020.

[16] Jr., P. S. *Crofton Formulæ*. Bachelor's thesis, University of Notre Dame, Department of Mathematics. Written under prof. Liviu Nicolaescu, 2018.

[17] Krantz, S., and Parks, H. *Geometric Integration Theory*. Cornerstones. Birkhäuser Boston, 2008.

[18] Maggi, F. *Sets of finite perimeter and geometric variational problems: an introduction to Geometric Measure Theory*. Cambridge University Press, 2012.

[19] Makarov, B., and Podkorytov, A. *Real analysis: measures, integrals and applications*. Springer Science & Business Media, 2013.

[20] Mattila, P. *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[21] Morgan, F. *Geometric Measure Theory: A Beginner's Guide*. Elsevier Science, 2000.

[22] Negro, L. Sample distribution theory using coarea formula. *Communications in Statistics - Theory and Methods* (01 Sep 2022). DOI: 10.1080/03610926.2022.2116284.

[23] Nicolaescu, L. The crofton formula. (Lecture Notes) University of Notre Dame, Department of Mathematics.

[24] ΝΕΓΡΕΠΟΝΤΗΣ Γ., and ΚΟΥΜΟΥΛΛΗΣ Σ. *Θεωρια Μετρου*. εκδ. Συμμετρία, 2005.

[25] Simon, L. Introduction to geometric measure theory. Lectures notes. Last modified: February 2014.

[26] Spivak, M. *Calculus On Manifolds: A Modern Approach To Classical Theorems of Advanced Calculus*. Avalon Publishing, 1965.

[27] Theochari-Apostolidi, T., Vavatsoulas, C., and Charalampous, C. *Linear Algebra*. (The book is in Greek) Tziola Publications, 2018.

[28] Treibergs, A. Steiner symmetrization and applications. Lecture slides, University of Utah, Utah, 2008.

[29] van Dijk, D. *On the sets of critical points and critical values*. Bachelor's Thesis, Mathematical Institute, Leiden University, 2014.