Berstein's Problem for Minimal Surfaces

Natasha Diederen, supervised by Felix Schulze

Summer 2022

Contents

1 Introduction **2**

2 A Brief Introduction to Measure Theory **3**
2.1 Measures and Algebras 3
2.2 Borel σ-algebra and Radon Measures 4
2.3 Lebesgue measure 6

3 Hausdorff Measures **9**

4 Lebesgue Integration **13**

5 Rademacher's Theorem **15**
5.1 Lipschitz Functions 15
5.2 Rademacher's Theorem 16

6 Linear maps and Jacobians **16**

7 The Area Formula **19**
7.1 Applications 23

8 Submanifolds of $\mathbb{R}^{n+1}$ **23**
8.1 Differentiability properties of locally Lipschitz maps 25
8.2 Return to the Area Formula 26
8.3 Gradient and Divergence Theorem 28

9 The Second Fundamental Form **29**
9.1 The Second Fundamental Form in $\mathbb{R}^3$ 29
9.2 The Second Fundamental Form in $\mathbb{R}^{n+1}$ 31
9.3 Mean curvature 31

10 First and Second Variation Formulae **33**

11 Bernstein's Problem **35**
11.1 Minimal surfaces in $\mathbb{R}^{n+1}$ 35
11.2 Proof of Bernstein's Problem for Minimal Hypersurfaces in $\mathbb{R}^3$ 43

12 Conclusion **45**

1 Introduction

Geometric measure theory is the study of geometric objects using a measure theoretic framework. This sets it apart from the more classical differential geometry, which deals with smooth (infinitely differentiable) structures. Geometric measure theory therefore allows for greater freedom in the types of geometric objects that can be studied, as it has the capability to deal with low-regularity structures, such as surfaces with C^1 or Lipschitz-type regularity.

Geometric measure theory sprung up as an attempt to solve Plateau's problem, which deals with proving the existence of surfaces of least area among surfaces with a given boundary. In three dimensions, this mimics the soap film that forms when a closed piece of wire is dipped in soapy water and pulled out. The reason that geometric measure theory comes into play is that there is no reason why area minimisers need to be smooth. For example, Figure 1 shows the (unstable) minimal surface formed when the boundary of a cube is dipped in soapy water. The solution is smooth except at the singularities formed at the intersection between two edges. Such a surface is known as *rectifiable*.

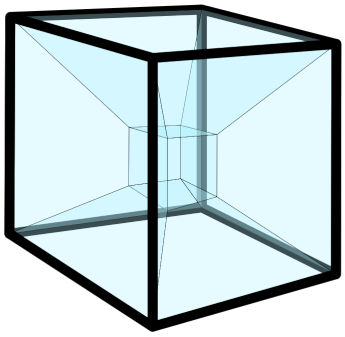


Figure 1: Soap film formed when dipping the boundary of a cube in soapy water

The main aim of this project is to discuss some properties of minimal surfaces and to explore Bernstein's problem. This asks whether minimal hypersurfaces embedded in $\mathbb{R}^n$ that are the graph of a function must be affine linear. Bernstein's problem turns out no be true for $n \leq 8$, but not for higher dimensions. We will provide a proof for the case $n = 3$.

We attempt to build up to Bernstein's problem in a way that requires no prior knowledge of geometric measure theory, or indeed measure theory. We start by exploring some key concepts of measure theory, including standard definitions, Borel sets and the Lebesgue measure. From here we move onto the Hausdorff measure which is a key element of geometric measure theory that allows us to calculate the 'volume' of lower dimensional subsets of an ambient space.

Next, there is a brief interlude in which we discuss the Lebesgue integral and how it differs from the Riemann integral. We also explore Rademacher's

Theorem which gives us a useful criterion for the differentiability of Lipschitz maps, as well as some useful concepts from linear algebra. This all builds up to an understanding of the area formula. The area formula is a very important result in geometric measure theory that allows us to calculate the Hausdorff measure of the image of a Lipschitz map via a special integral.

Following from this, we discuss some properties of submanifolds, including an extension of both Rademacher's Theorem and the area formula to locally Lipschitz maps from a submanifold. From here, we can discuss the second fundamental form, which is closely related to the curvature of a surface. In fact, the mean curvature of a surface is given by the trace of its second fundamental form. This is important since minimal surfaces are defined as having zero mean curvature everywhere. Finally, we discuss the first and second variation of a surface, which describe how a 1-parameter family of manifolds evolves over time. This sets up up nicely for a proof of Bernstein's problem.

The main resources for this project were *Measure Theory and Fine Properties of Functions* [EG15], *Introduction to Geometric Measure Theory* [Sim14] and *Minimal Surfaces and Functions of Bounded Variation* [Giu84].

2 A Brief Introduction to Measure Theory

We briefly introduce concepts in measure theory most relevant to this project. For a more detailed discussion the reader is referred to *Measure Theory and Fine Properties of Functions* [EG15, Sections 1 and 2] and *Introduction to Geometric Measure Theory* [Sim14, Section 1].

2.1 Measures and Algebras

A measure generalises the notion of 'length' or 'volume' to arbitrary sets. We begin by introducing the concepts of an outer measure, measurability and σ-algebras.

Definition 2.1 ([EG15, Section 1.1])**.** A mapping $\mu : 2^X \to [0, \infty]$ is called a **outer measure** on X provided

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

(ii) if $A \subseteq \bigcup_{k=1}^{\infty} A_k$, then $\mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k)$; this is known as subadditivity.

From Figure 2 it can be seen that the subadditivity property ties into our notion of adding together overlapping 'volumes'. The difference between a measure and an outer measure is that instead of subadditivity, a measure has the property of countable additivity:

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k)$$

for each infinite sequence $\{A_k\}$ of pairwise disjoint sets [Coh13, Section 1.2]. The reason that we use outer measures as opposed to measures is that even

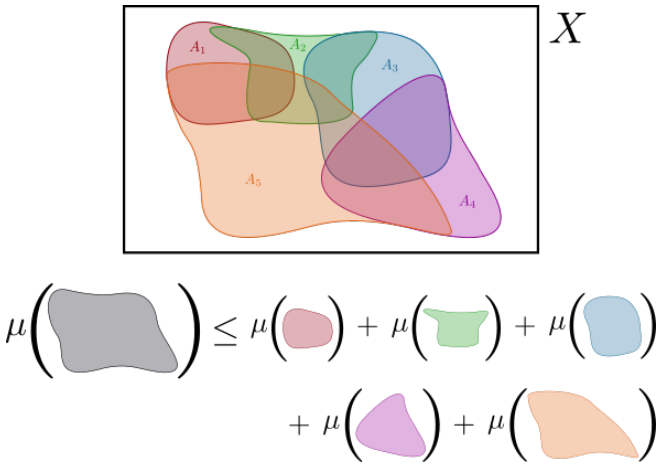


Figure 2: An illustration of the subadditivity property

non-measurable sets can be given an outer measure, which is not the case with measures.

In general, outer measures fail to be countably additive. However, the restriction of μ to the collection of μ-measurable subsets of X is indeed a measure [Coh13, Section 1.3]. We will explore this below.

Definition 2.2 ([Mor09, Chapter 2.1])**.** A set $A \subseteq X$ is **μ-measurable** if for all $E \subseteq X$, $\mu(E \cap A) + \mu(E \backslash A) = \mu(E)$.

Definition 2.3 ([Mor09, Section 2.1])**.** A **σ-algebra** is a collection $\mathcal{A} \subseteq 2^X$ of sets closed under complementation, countable union and countable intersection.

Elements of a σ-algebra are called measurable sets. By showing that μ-measurable subsets form a σ-algebra, it follows that if we just consider μ-measurable sets, then our outer measure will behave the same as a measure on them.

Theorem 2.1 ([EG15, Section 1.1])**.** *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.*

For the remainder of this essay outer measures will simply be referred to as measures. This is a common practice when no confusion is likely to arise.

2.2 Borel σ-algebra and Radon Measures

The Borel σ-algebra forms an important part of measure theory on $\mathbb{R}^n$ as it is generated by the open sets of $\mathbb{R}^n$ and thus closely related to its topology. The Borel σ-algebra contains all countable combinations of unions and intersections of open or closed sets in $\mathbb{R}^n$, but does not actually contain all power-sets of $\mathbb{R}^n$. However, it contains all the (non-pathological) sets anyone is likely to want to work with.

Definition 2.4 ([EG15, Section 1.1])**.**

(i) The **Borel σ-algebra** of $\mathbb{R}^n$ is the smallest **σ-algebra** generated by the open subsets of $\mathbb{R}^n$. Sets belonging to this σ-algebra are known as **Borel sets**.

(ii) A measure μ on $\mathbb{R}^n$ is called **Borel regular** if each Borel set is measurable and for each $A \subseteq \mathbb{R}^n$ there exists a Borel set B such that $A \subseteq B$ and $\mu(A) = \mu(B)$.

Example 2.1. Consider the measure on $\mathbb{R}$ defined by

$$\mu(A) = \inf\left\{ \sum_{i=1}^{\infty}(b_i - a_i) : A \subseteq \bigcup_{i=1}^{\infty}(a_i, b_i),\ (a_i, b_i) \in \mathbb{R} \right\}$$

we will formally define this as the Lebesgue measure later. Consider the set $A = (0, 1)$, which has measure 1. Let $B = [0, 1]$, then $A \subset B$ and

$$\mu(B) = \inf\left\{ \left(1 + \frac{1}{n} - \left(\frac{1}{n}\right)\right) \right\} = \inf\left(1 + \frac{2}{n}\right) = 1,$$

since $[0, 1]$ is covered by $\left(-\frac{1}{n}, 1 + \frac{1}{n}\right)$ for all n. So A and B have the same measure even though A is a subset of B. This is not a proof that the Lebesgue measure is regular (although it is) but serves to illustrate how two sets can have the same measure.

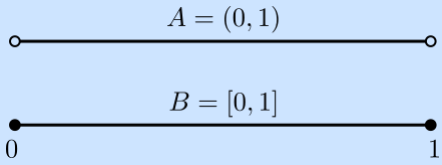


Figure 3: Two different sets with the same measure

The following theorem gives a straightforward way to check whether a measure is Borel regular or not.

Theorem 2.2 (Carathéodory's criterion [EG15, Section 1.1])**.** *Let μ be a measure on $\mathbb{R}^n$. If for all sets $A, B \subseteq \mathbb{R}^n$, we have*

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

whenever $\operatorname{dist}(A, B) > 0$*, then μ is a Borel measure.*[1]

Finally, in order to establish the compatibility of a measure with the topology of $\mathbb{R}^n$, we introduce the concept of a Radon measure.

[1] Let X be a metric space with metric ρ. For sets $A,\ B \subseteq X$, we define the **distance** between A and B by setting $\operatorname{dist}(A, B) = \inf\{\rho(a, b) : a \in A,\ b \in B\}$.

Definition 2.5 ([EG15, Section 1.1])**.** A measure μ on $\mathbb{R}^n$ is a **Radon measure** if μ is Borel regular and $\mu(K) < \infty$ for every compact set[2] $K \subset \mathbb{R}^n$.

This means that compact spaces in $\mathbb{R}^n$ cannot have a measure that blows up, which agrees with our intuitions about compact spaces which are closed and bounded on $\mathbb{R}^n$.

2.3 Lebesgue measure

We now want a way to measure subsets of $\mathbb{R}^n$ that coincides with the standard measurement of length, area and volume for $n = 1, 2, 3$ respectively, and can then be extended to higher dimensions. This is known as the Lebesgue measure.

There are many ways to construct the Lebesgue measure, all of them equivalent. Carathéodory's method constructs an outer measure on subsets of $\mathbb{R}^n$ by approximating them by countable unions of rectangles. This is the approach taken in *Lebesgue Integration on Euclidean Space* [Jon01, Chapter 3]. In general, any sets can be used as a cover, but rectangles are easier to work with.

Definition 2.6. A ***n*-dimensional rectangle** $R \subset \mathbb{R}^n$ is defined as

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n],$$

with volume

$$\mu(R) = (b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n).$$

Let $\mathcal{R}(\mathbb{R}^n)$ be the set of rectangles in $\mathbb{R}^n$.

Definition 2.7. The **Lebesgue measure** $\mathcal{L}^n(E)$ of a subset $E \subseteq \mathbb{R}^n$ is

$$\mathcal{L}^n(E) = \inf \left\{ \sum_{i=1}^{n} \mu(R_i) : E \subseteq \bigcup_{i=1}^{\infty} R_i,\ R_i \in \mathcal{R}(\mathbb{R}^n) \right\}$$

Any set can be approximated by rectangles using the method illustrated by Figure 4. In general, let E be the set we are trying to measure. We first divide $\mathbb{R}^n$ into n-dimensional cubes $Q_{1,i}$, $i \in \mathbb{N}$, with side length one. If $Q_{1,i} \subset E$, then we include it in the union. If $Q_{1,i}$ is entirely disjoint from E, then it is excluded. If $Q_{1,i}$ covers a part of E, we can bisect the sides of $Q_{1,i}$ to obtain 2^n cubes $Q_{2,j}$ of side length $\frac{1}{2}$. We can keep repeating this process until we have covered the entire set E. Let A_k be the set of indices of cubes included with the k^{th} bisection. Then

$$E = \bigcup_{k=1, i \in A_k} Q_{k,i}.$$

If we want to ensure a strict inclusion $E \subset \bigcup_{k=1, i \in A_k} Q_{k,i}$, we can take open cubes of the form

$$Q = \left(a_1 + \frac{1}{j}, b_1 + \frac{1}{j}\right) \times \left(a_2 + \frac{1}{j}, b_2 + \frac{1}{j}\right) \times \cdots \times \left(a_n + \frac{1}{j}, b_n + \frac{1}{j}\right)$$

[2] A set A is *compact* if every open cover of A has a finite subcover.

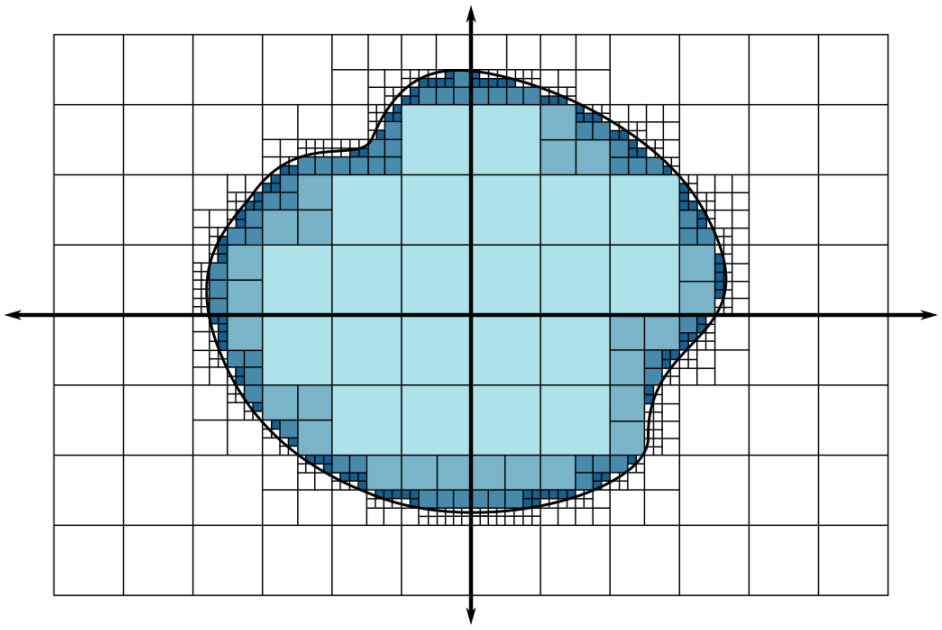


Figure 4: Approximating a set in $\mathbb{R}^2$ using countably many squares

instead and take the limit as $j \to \infty$.

The Lebesgue measure on $\mathbb{R}^n$ is a measure, and assigns to each n-dimensional interval its volume [Coh13, Section 1.3]. The Lebesgue measure is also Borel regular, that is, every Borel subset of $\mathbb{R}^n$ is Lebesgue measurable [Mor09, Section 2.2]. In practice it is pretty difficult to come up with a non-measurable set, but the Vitali set is one such example [Jon01, Chapter 4A].

Example 2.2. Suppose we want to measure the set $E = [0,1] \cap \mathbb{Q}$. Since there are countably many rational numbers, for all ε we can cover each rational q_i with an interval of length $\frac{\varepsilon}{2^i}$. Then

$$\mathcal{L}^n(E) = \inf\left\{\sum_{i=1}^{\infty} \frac{\varepsilon}{2^i}\right\} = \inf \varepsilon = 0,$$

since we can take ε to be arbitrarily small.
If instead we could only cover E by a finite number of sets, the union of those sets would have to include $[0,1]$ as the rationals are dense in $[0,1]$. This illustrates why countably infinite collections of rectangles are used to define Lebesgue measure.

The Lebesgue measure has many properties that make it of particular use in studying geometric objects. These properties match our intuitions on how

volumes change under geometric transformations.

Theorem 2.3 (Geometric properties of the Lebesgue measure [Bog07, Section 1.7], [Jon01, Chapter 3]). *Let $A \subset \mathbb{R}^n$ be a measurable set.*

(i) Lebesgue measure is invariant under translation, that is

$$\mathcal{L}^n(z + A) = \mathcal{L}^n(A).$$

(ii) Lebesgue measure is scaled under dilation by a factor of t^n, that is

$$\mathcal{L}^n(tA) = t^n \mathcal{L}^n$$

(iii) Lebesgue measure is invariant under orthogonal transformation, that is

$$\mathcal{L}^n(O(A)) = \mathcal{L}^n(A),$$

for an orthogonal map O.

(iv) Lebesgue measure of a linear transformation T is scaled by the determinant of T, that is

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

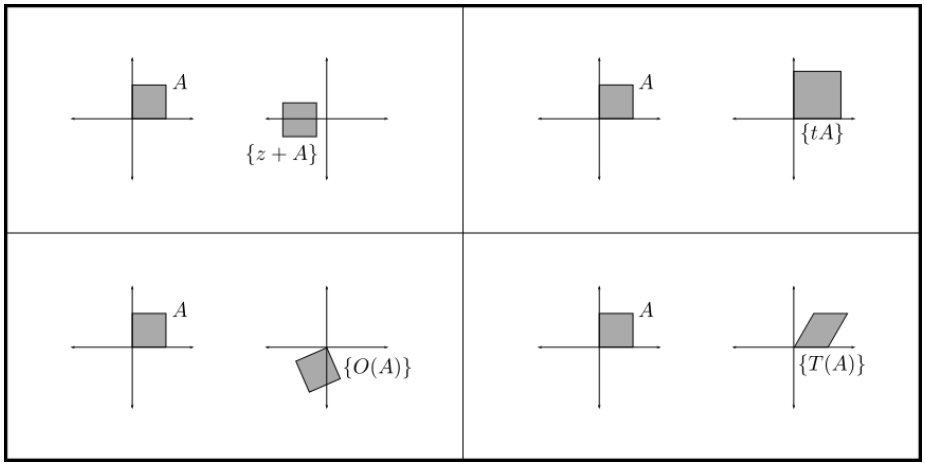


Figure 5: Illustration of geometric transformations in Theorem 2.3

Proof.

(i) If we divide the set A into rectangles, then $\{C_i + z\}$ forms a cover of $A + z$ if and only if $\{C_i\}$ forms a cover of A. Let $C_1 = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be one rectangle in our cover, then $C_1 + z = [z_1 + a_1, z_2 + b_2] \times \cdots \times [z_n + a_n, z_n + b_n]$. By definition, we see that $\mathcal{L}^n(C_1 + z) = \mathcal{L}^n(C_1)$. As this holds for all rectangles in the cover, it holds for A by countable additivity.

(ii) If we again divide the set into rectangles, then then $\{tC_i\}$ forms a cover of tA if and only if $\{C_i\}$ forms a cover of A. Similar to above, let $C_1 = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be one rectangle in our cover, then

$$tC_1 = [ta_1, tb_2] \times \cdots \times [ta_n, tb_n] = t^n\big([a_1, b_1] \times \cdots \times [a_n, b_n]\big).$$

It follows that $\mathcal{L}^n(tC_1) = t^n\mathcal{L}^n(C_1)$. As this holds for all rectangles in the cover, it holds for A by countable additivity.

(iii) An orthogonal transformation maps rectangles to rectangles with no scaling at all. Therefore if $\{C_i\}$ forms a cover of A if and only if $\{O(C_i)\}$ forms a cover of $O(A)$. Since the volume of each C_i remains unchanged, it follows that $\mathcal{L}^n(C_i) = \mathcal{L}^n(O(C_i))$. As this holds for all rectangles in the cover, the result holds for A by countable additivity.

(iv) Assume that T is non singular, otherwise its range will be a lower dimensional subspace of $\mathbb{R}^n$ and thus have Lebesgue measure zero. By polar decomposition[3] $T = O \circ S$, where S is a symmetric map such that $S^2 = T^* \circ T$ and O an orthogonal map. We can now diagonalise S as $S = P^* \circ D \circ P$, where D is a diagonal matrix and P is orthogonal. From previous result, Lebesgue measure is invariant under orthogonal maps, so

$$\mathcal{L}^n(T \circ A) = \mathcal{L}^n(D \circ A) = (\det D)\mathcal{L}^n(A).$$

Since $|\det P| = |\det O| = 1$, it follows that $\det D = \det S = |\det T|$. Hence $\mathcal{L}^n(T \circ A) = |\det T|\mathcal{L}^n(A)$. □

3 Hausdorff Measures

The Lebesgue measure works very well when we are considering a subset that is the same dimension as its ambient space. However, in the case where we have a m-dimensional subset of $\mathbb{R}^n$ with $m < n$, the Lebesgue measure will be zero and therefore not very useful. In general we want to be able to assign a meaningful measure to this m-dimensional subset, which has similar geometric properties to the Lebesgue measure (see Theorem 2.3).

In order to investigate this new measure, let start with dimension $m = 1$ and naively define a measure h^1 very similar to the Lebesgue measure on $\mathbb{R}$, but covered with sets in $\mathbb{R}^n$:

$$h^1(A) \coloneqq \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam} C_i : A \subseteq \bigcup_{i=1}^{\infty} C_i, C_i \subseteq \mathbb{R}^n\right\}.$$

To show that h^1 does not give us the information we want, consider the set $A = [0, 1] \times \{0, \varepsilon\}$, shown in Figure 6. For small ε,

$$h^1(A) \leq \operatorname{diam}(A) = \sqrt{1 + \varepsilon^2},$$

[3] refer to Theorem 6.1 for more detail

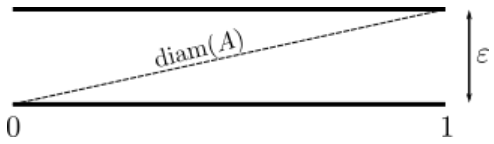


Figure 6: The set A

which is far smaller than the desired value of two [Tra94, Section 1]. This is because the diameter of the set gives a smaller measure than if the set were to be covered by small rectangles along each of the two lines. The failure of h^1 to accurately follow the geometry of a curve in $\mathbb{R}^n$ motivates the following definition, which restricts the size of sets in the cover.

Definition 3.1. Let $A \subseteq \mathbb{R}^n$, $0 \leq s < \infty$, $0 < \delta \leq \infty$. We write

$$\mathcal{H}^s_\delta(A) \coloneqq \inf\left\{ \sum_{i=1}^{\infty} \alpha(s) \left(\frac{\operatorname{diam} C_i}{2} \right)^s : A \subseteq \bigcup_{i=1}^{\infty} C_i,\ \operatorname{diam} C_i \leq \delta \right\}$$

where

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

is the volume of a s-dimensional sphere. Now for A and s as above, define

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

We call $\mathcal{H}^s$ the ***s*-dimensional Hausdorff measure** on $\mathbb{R}^n$.

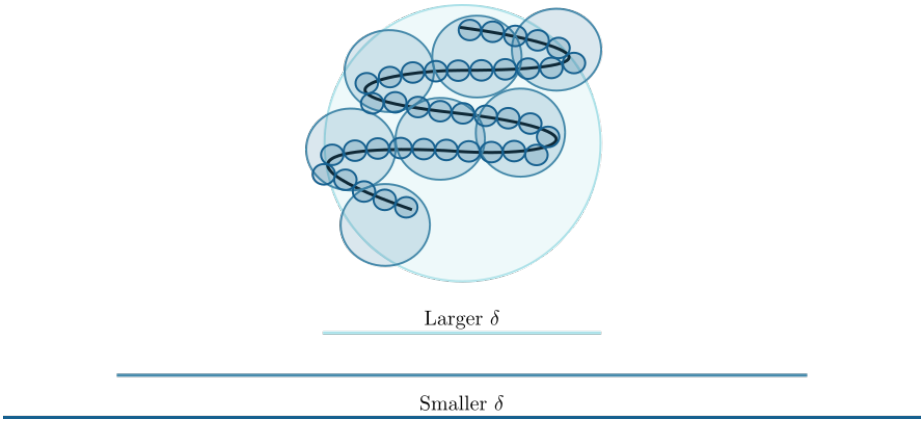


Figure 7: The effect of decreasing δ when measuring a set

Theorem 3.1 ([EG15, Section 2.1])**.** *For all $0 \leq s < \infty$, $\mathcal{H}^s$ is a Borel regular measure in $\mathbb{R}^n$.*

By construction, the Hausdorff measure has very similar properties to the Lebesgue measure, so it is similarly useful for studying geometric objects, with the added benefit that we can now also study lower-dimensional objects embedded in a higher dimensional space.

Theorem 3.2 (Properties of Hausdorff measure [EG15, Section 2.1])**.**

(i) $\mathcal{H}^0(A)$ *gives the number of points in* $A \subset \mathbb{R}^n$*.*

(ii) $\mathcal{H}^1 = \mathcal{L}^1$ *on* $\mathbb{R}$*.*

(iii) $\mathcal{H}^s \equiv 0$ *on* $\mathbb{R}^n$ *for all* $s > n$*.*

(iv) $\mathcal{H}(\lambda A) = \lambda^s \mathcal{H}^s(A)$ *for all* $\lambda > 0$*,* $A \subseteq \mathbb{R}^n$*.*

(v) $\mathcal{H}^s(L(A)) = \mathcal{H}^s(A)$ *for each affine isometry* $L : \mathbb{R}^n \to \mathbb{R}^n$*,* $A \subseteq \mathbb{R}^n$*.*

Proof. Proofs based on [EG15, Section 2.1]

(i) Since $\alpha(0) = 1$, it follows from Definition 3.1 that $\mathcal{H}^0(\{a\}) = 1$ for all $a \in A$. The result follows.

(ii) For all $A \subseteq \mathbb{R}$ and $\delta > 0$, we want to show that $\mathcal{L}^1(A) \leq \mathcal{H}^1_\delta(A)$ and $\mathcal{L}^1(A) \geq \mathcal{H}^1_\delta(A)$. Firstly, by Definition 2.7,

$$\begin{aligned}\mathcal{L}^1(A) &= \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam} C_i : A \subseteq \bigcup_{i=1}^{\infty} C_i\right\} \\ &\leq \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam} C_i : A \subseteq \bigcup_{i=1}^{\infty} C_i,\ \operatorname{diam} C_i \leq \delta\right\} \\ &= \mathcal{H}^1_\delta(A).\end{aligned}$$

This last equality comes from the fact that $\alpha(1) = 2$ (since this is the length of the unit ball in $\mathbb{R}$) and so

$$\alpha(1)\left(\frac{\operatorname{diam} C_i}{2}\right) = \operatorname{diam} C_i.$$

For the opposite inequality, set $I_k = [k\delta, (k+1)\delta]$ for $k \in \mathbb{Z}$. Then $\operatorname{diam}(C_i \cap I_k) \leq \delta$ and

$$\sum_{k=-\infty}^{\infty} \operatorname{diam}(C_i \cap I_k) \leq \operatorname{diam} C_i.$$

It follows that

$$\begin{aligned}\mathcal{L}^1(A) &= \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam} C_i : A \subseteq \bigcup_{i=1}^{\infty} C_i\right\} \\ &\geq \inf\left\{\sum_{i=1}^{\infty} \sum_{k=-\infty}^{\infty} \operatorname{diam}(C_i \cap I_k) : A \subseteq \bigcup_{i=1}^{\infty} C_i\right\} \\ &\geq \mathcal{H}^1_\delta(A).\end{aligned}$$

The final inequality holds because $\operatorname{diam}(C_i \cap I_k) \leq \delta$, so we have a sum of sets with diameter smaller than δ which leads to a larger measure than $\mathcal{H}^1_\delta$.

The two inequalities above hold for all δ, which completes the proof.

(iii) Fix an integer $m \geq 1$. Let Q be the unit cube in $\mathbb{R}^n$. We can decompose Q into m^n smaller cubes, each with side length $\frac{1}{m}$ and diameter $\frac{\sqrt{n}}{m}$. It follows from Definition 3.1 that

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

As $m \to \infty$, the last term goes to zero if $s > n$ and $\mathcal{H}^s(Q) = 0$. Hence $\mathcal{H}^s(\mathbb{R}^n) = 0$.

(iv) and (v) similar to proof of Theorem 2.3 parts (ii) and (iii). □

We now define the Hausdorff dimension, which is a number that agrees with what we usually think of as the dimension of a geometric object. For example, points have Hausdorff dimension zero and curves have Hausdorff dimension one. Fractals have non-integer Hausdorff dimension.

Definition 3.2 ([EG15, Section 2.1])**.** The **Hausdorff dimension** of a set $A \subseteq \mathbb{R}^n$ is

$$H_{\dim}(A) \coloneqq \inf\{0 \leq s < \infty : \mathcal{H}^s(A) = 0.\}$$

Proposition 3.3 ([EG15, Section 2.1])**.** *Suppose that $H_{\dim}(A) = s$. Then for all $t > s$, $\mathcal{H}^t(A) = 0$ and for all $t < s$, $\mathcal{H}^t(A) = +\infty$.*

Proof. We know that $\mathcal{H}^2(A) < \infty$, so for any $\delta > 0$, there exists $\{C_j\}_{j=1}^\infty$ such that $\operatorname{diam}(C_j) \leq \delta$ and $A \subseteq \bigcup_{j=1}^\infty C_j$. By definition,

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

We now consider

$$\mathcal{H}^t_\delta(A) = \sum_{j=1}^\infty \alpha(t)\left(\frac{\operatorname{diam} C_j}{2}\right)^t,$$

which we can rewrite as

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

As $\delta \to 0$, we obtain $\mathcal{H}^t = 0$.

Now suppose that $\mathcal{H}^s(A) > 0$ and $t < s$, we want to show that $\mathcal{H}^t(A) = \infty$. Assume for a contradiction that there exists t_1 such that $\mathcal{H}^{t_1} < \infty$. Then by the previous part, $\mathcal{H}^s(A) = 0$, which is a contradiction. □

Finally, it can be shown that $\mathcal{H}^n$ coincides with $\mathcal{L}^n$ on $\mathbb{R}^n$. Proof of this requires knowledge of the isodiametric inequality and Besicovitch covering theorem, which have not been covered here. These topics and a proof that $\mathcal{H}^n = \mathcal{L}^n$ can be found in *Measure Theory and Fine Properties of Functions* [EG15].

4 Lebesgue Integration

We now introduce the Lebesgue integral, which allows us to integrate a larger class of functions compared to the Riemann integral. The key difference between the two methods is that Riemann integration partitions the domain into arbitrary sub-intervals, whereas Lebesgue integration partitions the domain via the inverse images of intervals in the range.

We first introduce some terminology that allows us give an entire set a property even if it does not hold for a small number of points in the set.

Definition 4.1 ([KP08, Section 1.3])**.** The term **μ-almost** can serve as an adjective or adverb in the following ways:

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

$$\mu\big(\{x \in X : \mathcal{P}(x) \text{ is false}\}\big) = 0.$$

If X is understood from context, then we say that $\mathcal{P}(x)$ holds **μ-almost everywhere**.

(ii) Two sets $A, B \subseteq X$ are **μ-almost equal** if

$$\mu((A \backslash B) \cup (B \backslash A)) = 0.$$

(iii) Two functions f and g defined for μ-almost every $x \in X$ are said to be **μ-almost equal** if $f(x) = g(x)$ holds for μ-almost every $x \in X$.

As with sets, it is also possible to define μ-measurability for functions.

Definition 4.2 ([EG15, Section 1.3])**.** Let Y be a topological space. A **μ-measurable Y-valued function** is a Y-valued function f defined for μ-almost every $x \in X$ such that

(i) $f : D \subseteq X \to Y$,

(ii) $\mu(X \backslash D) = 0$,

(iii) $f^{-1}(U)$ is μ-measurable whenever $U \subseteq Y$ is open.

Definition 4.3 ([EG15, Section 1.3])**.** For a function $f : X \to \mathbb{R}$, the **positive part** of f is the function $f^+ : X \to [0, \infty]$ defined by $f^+ = \max(f, 0)$. The **negative part** of f is the function $f^- \to [0, \infty]$ defined by $f^- = \max(-f, 0)$.

It follows from this definition that we can write any function f as $f = f^+ - f^-$. This means that we only ever have to deal with non-negative functions. We now want to approximate our function f using functions with countable image.

Definition 4.4 ([EG15, Section 1.3])**.** A function $g : X \to [-\infty, \infty]$ is a **simple function** if the image of g is countable. This is equivalent to saying that g can be written in the form

$$g = \sum_{i=1}^{n} a_i \chi_{A_i}$$

Since there our countably many values that g can take, this means that we are able to take a sum over these values to obtain an integral as follows.

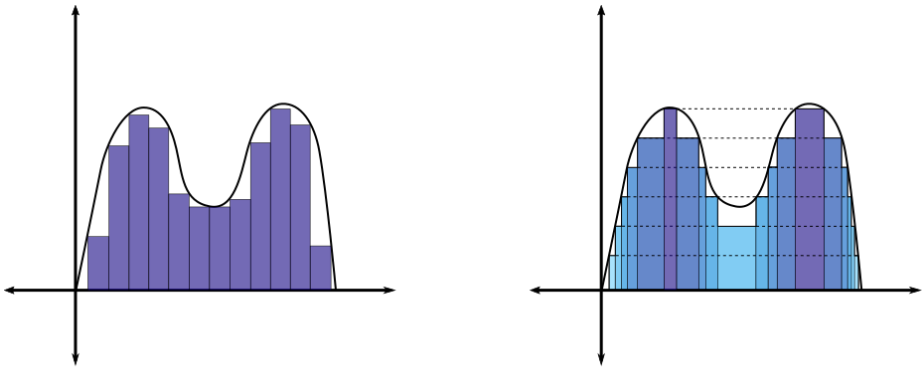
Definition 4.5 ([EG15, Section 1.3])**.**

(i) If g is a non-negative, simple, μ-measurable function, we define its **integral**

$$\int g \, d\mu = \sum_{0 \le y \le \infty} y\mu(g^{-1}\{y\}).$$

(ii) If g is a simple μ-measurable function and either $\int g^+ \, d\mu < \infty$ or $\int g^- \, d\mu < \infty$, we call g a **μ-integrable simple function** and define its **integral**

$$\int g \, d\mu = \int g^+ \, d\mu - \int g^- \, d\mu.$$



(a) A lower Riemann integral

(b) A lower Lebesgue integral

Figure 8: The difference between a Riemann and Lebesgue integral

Definition 4.6 ([EG15, Section 1.3])**.**

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

$$\int^* f \, d\mu = \inf \left\{ \int g \, d\mu : g \text{ is } \mu\text{-integrable, simpele and } g \geq f \ \mu\text{-a.e.} \right\}$$

and the **lower integral**

$$\int_* f \, d\mu = \inf \left\{ \int g \, d\mu : g \text{ is } \mu\text{-integrable, simpele and } g \leq f \ \mu\text{-a.e.} \right\}.$$

(ii) A μ-measurable function $f : X \to [-\infty, \infty]$ is called **μ-integrable** if $\int^* f \, d\mu = \int_* f \, d\mu$, in which case we write

$$\int f \, d\mu = \int^* f \, d\mu = \int_* f \, d\mu.$$

All integrals in the remaining document will be assumed to be Lebesgue integrals.

5 Rademacher's Theorem

Lipschitz continuity is a stronger condition than continuity, but weaker than differentiability. However it can be shown that Lipschitz functions are differentiable almost everywhere with respect to the Lebesgue measure, which is a very powerful result.

5.1 Lipschitz Functions

We first introduce Lipschitz functions.

Definition 5.1 ([EG15, Section 3.1])**.**

(i) Let $A \subseteq \mathbb{R}^n$. A function $f : A \to \mathbb{R}^m$ is called **Lipschitz continuous** provided

$$|f(x) - f(y)| \leq C|x - y| \tag{5.1}$$

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

(ii) The smallest constant C such that (5.1) holds for all x, y is denoted

$$\mathrm{Lip}(f) \coloneqq \sup \left\{ \frac{|f(x) - f(y)|}{|x - y|} : x, y \in A, \ x \neq y \right\}.$$

Thus

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

(iii) A function $f : A \to \mathbb{R}^m$ is called **locally Lipschitz continuous** if for each compact $K \subseteq A$, there exists a constant C_K

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

for all $x, y \in K$.

Example 5.1. The function $f(x) = |x|^{\frac{1}{n}}$ is continuous but not Lipschitz continuous at 0. To show this, suppose for a contradiction that there exists C such that

$$\left||x|^{\frac{1}{n}} - |y|^{\frac{1}{n}}\right| \leq C|x-y|.$$

Set $x = \frac{1}{k^n}$ for $k \in \mathbb{N}$ and $y = 0$. Then we see that

$$\frac{1}{k} \leq \frac{C}{k^n},$$

so $k^{n-1} \leq C$ for all $k \in \mathbb{N}$. This contradicts the fact that C is finite.

5.2 Rademacher's Theorem

Before stating Rademacher's Theorem, we define what it means for a function to be differentiable.

Definition 5.2 ([EG15, Section 3.1])**.** The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at $x \in \mathbb{R}^n$ if there exists a linear mapping

$$L : \mathbb{R}^n \to \mathbb{R}^m$$

such that

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

If such a linear mapping L exists, then it is unique and we can write $Df(x)$ for L. We call $Df(x)$ the **derivative** of f at x.

Theorem 5.1 (Rademacher's Theorem [EG15, Section 3.1])**.** *Assume that $f : \mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitz continuous function. Then f is differentiable $\mathcal{L}^n$-a.e.*

Rademacher's Theorem means that the points on a Lipschitz function that are not differentiable form a set of Lebesgue measure zero. Proof can be found in [EG15, Section 3.1]. See Figure 6 for an example.

6 Linear maps and Jacobians

This section contains a review of some concepts in linear algebra, which will be useful for subsequent sections. We start with a few definitions.

Definition 6.1 ([EG15, Section 3.2])**.**

(i) A linear map $O : \mathbb{R}^n \to \mathbb{R}^m$ is **orthogonal** if

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

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

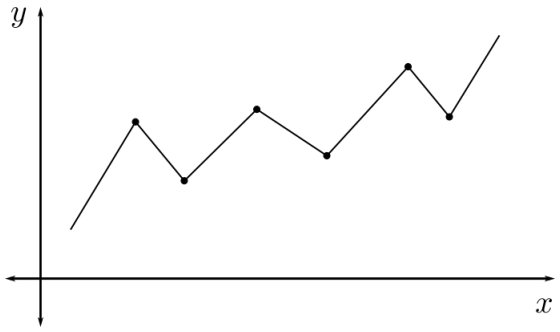


Figure 9: Points with no derivative (marked) form a set of Lebesgue measure zero

(ii) A linear map $S : \mathbb{R}^n \to \mathbb{R}^n$ is **symmetric** if

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

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

(iii) 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

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

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

It is possible to write any linear map as a composition of symmetric and orthogonal maps, as shown by the following theorem. This result was used earlier to prove Theorem 2.3(iv).

Theorem 6.1 (Polar decomposition [EG15, Section 3.2])**.** *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping.*

(i) 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 $L = O \circ S$.

(ii) 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 $L = S \circ O^*$.

Polar decomposition characterises the map in terms of a rotation or reflection given by the orthogonal map and a scaling given by the symmetric map. There are also nice properties of orthogonal and symmetric matrices which can be exploited when L is written in this form.

Definition 6.2 ([EG15, Section 3.2])**.** Assume $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear.

(i) If $n \leq m$, we write $L = O \circ S$ as above, and we define the **Jacobian** of L to be

$$[\![L]\!] = |\det S|.$$

(ii) If $n \geq m$, we write $L = S \circ O^*$ as above, and we define the **Jacobian** of L to be

$$[\![L]\!] = |\det S|.$$

It would be more useful to characterise $[\![L]\!]$ in terms of the linear map L so the polar decomposition does not need to be explicitly calculated. The following result allows us to do so.

Theorem 6.2 ([EG15, Section 3.2])**.**

(i) If $n \leq m$, $[\![L]\!]^2 = \det(L^* \circ L)$.

(ii) If $n \geq m$, $[\![L]\!]^2 = \det(L \circ L^*)$.

Proof.

(i) If $n \leq m$, then by Theorem 6.1 we write $L = O \circ S$ and $L^* = S \circ O^*$. Now

$$\det(L^* \circ L) = \det(S \circ O \circ O^* \circ S) = \det(S^2) = |\det(S)|^2 = [\![L]\!]^2.$$

(ii) Similar to (i).

□

Alternatively, in the case where $n \leq m$, it is possible to calculate $[\![L]\!]^2$ by computing the sum of the squares of the determinants of the $(n \times n)$-submatrices of the $(m \times n)$-matrix representing L. This is known as the Binet-Cauchy formula.

Theorem 6.3 (Binet-Cauchy formula [EG15, Section 3.2])**.** *Assume that $n \leq m$ and $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear. Then*

$$[\![L]\!]^2 = \sum_{\lambda \in \Lambda(m,n)} (\det(P_\lambda \circ L))^2,$$

where $\Lambda(m,n) = \{\lambda : \{1,\dots,n\} \to \{1,\dots,m\} : \lambda \text{ is increasing}\}$ and $P_\lambda : \mathbb{R}^m \to \mathbb{R}^n$ is defined by $P_\lambda(x_1,\dots,x_m) \coloneqq \big(x_{\lambda(1)},\dots,x_{\lambda(n)}\big)$.

So far we have only discussed the Jacobian of a linear map, however we also want to be able to compute the Jacobian of a general (Lipschitz) function. Suppose we have a Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$. By Rademacher's Theorem this function is differentiable $\mathcal{L}^n$-a.e., so $Df(x)$ exists for $\mathcal{L}^n$-a.e. $x \in \mathbb{R}^n$ [EG15, Section 3.2].

The differential Df provides us with a good linear approximation to f, which leads to the following definition.

Definition 6.3 ([EG15, Section 3.2])**.** If $f : \mathbb{R}^n \to \mathbb{R}^m$, $f = (f^1,\dots,f^m)$, we write the gradient matrix

$$Df = \begin{pmatrix} f^1_{x_1} & \cdots & f^1_{x_n} \\ \vdots & \ddots & \vdots \\ f^m_{x_1} & \cdots & f^m_{x_n} \end{pmatrix}$$

at each point where Df exists.

Definition 6.4 ([EG15, Section 3.2])**.** For $\mathcal{L}^n$-a.e. point x, we define the **Jacobian** of f to be

$$Jf(x) \coloneqq [\![Df(x)]\!].$$

This is equivalent to writing

$$Jf(x) = \sqrt{\mathcal{J}(x)}$$

where $\mathcal{J}(x)$ is the $n \times n$ matrix with $(D_i f(x) \cdot D_j f(x))$ in the i^{th} row and j^{th} column.

7 The Area Formula

Suppose we have a hypersurface S embedded in $\mathbb{R}^3$ that can be written as $z = f(x, y)$ for f smooth and $(x, y) \in \Omega$. Then using the tools of differential geometry, we can calculate the surface area of S as

$$\iint_\Omega \left| \frac{\partial f}{\partial x} \times \frac{\partial f}{\partial y} \right| dxdy = \iint_\Omega \sqrt{1 + |Df|^2} dxdy.$$

The area formula is a generalisation of this concept that allows us to calculate the Hausdorff measure of the image $f(A)$ of any Lipschitz map $f : \mathbb{R}^n \to \mathbb{R}^m$, provided that $n \leq m$.

Theorem 7.1 (Area formula [EG15, Section 3.3])**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous, $n \leq m$. Then for each $\mathcal{L}^n$-measurable subset $A \subset \mathbb{R}^n$,*

$$\int_A Jf \; d\mathcal{L}^n = \int_{R^m} \mathcal{H}^0(A \cap f^{-1}\{y\}) \; d\mathcal{H}^n(y).$$

Here, the map $\mathcal{H}^0(A \cap f^{-1}\{y\})$ is the **multiplicity function**, which counts how many points in A get mapped to $y \in \mathbb{R}^m$ by f. Before we provide a proof of the above Theorem (for injective maps only), we state and prove the following result.

Lemma 7.2 ([Sim14, Section 2.3])**.** *Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear, $n \leq m$. Then*

$$\mathcal{H}^n(L(A)) = [\![L]\!] \mathcal{L}^n(A)$$

for all $A \subseteq \mathbb{R}^n$.

Proof. From Theorem 2.3, we have that for a linear map $\lambda : \mathbb{R}^n \to \mathbb{R}^n$,

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

Here, we want to consider a more general linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, with $m = n + k$. Suppose that $L(\mathbb{R}^n) \subset F$, where F is a n-dimensional subspace of $\mathbb{R}^{n+k}$. Consider an orthogonal transformation $Q : \mathbb{R}^{n+k} \to \mathbb{R}^{n+k}$, chosen

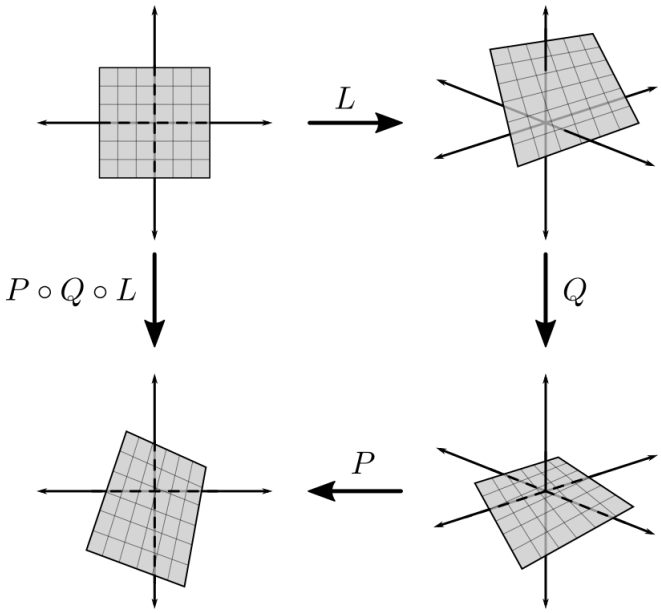


Figure 10: An illustration of the maps L, Q and P used in the proof

such that $Q \circ F = \mathbb{R}^n \times \{0\}$. Choose another map $P : \mathbb{R}^{n+k} \to \mathbb{R}^n$ such that $P(x, y) = x$ for $(x, y) \in \mathbb{R}^n \times \mathbb{R}^k$. Then the composition $P \circ Q \circ L$ is a map from $\mathbb{R}^n$ to $\mathbb{R}^n$. See Figure 10 for reference.

By Theorem 2.3, we have

$$\mathcal{L}^n((P \circ Q \circ L)(A)) = |\det(P \circ Q \circ L)|\mathcal{L}^n(A).$$

Now consider $(P \circ Q \circ L)^* \circ (P \circ Q \circ L) = L^* \circ Q^* \circ (P^* \circ P) \circ Q \circ L$. Since $P^* \circ P$ is the identity on $\mathbb{R}^n \times \{0\}$ and $Q \circ L(A)$ is a subspace of $\mathbb{R}^n \times \{0\}$, it follows that

$$L^* \circ Q^* \circ (P^* \circ P) \circ Q \circ L = L^* \circ (Q^* \circ Q) \circ L = L^* \circ L.$$

Hence, $\mathcal{L}^n((P \circ Q \circ L)(A)) = [\![L]\!]\mathcal{L}^n(A)$. So far we have only dealt with Lebesgue measure, but ideally we want the n-dimensional Hausdorff measure of $L(A)$ embedded in $\mathbb{R}^{n+k}$. If we can show that $\mathcal{H}^n(L(A)) = \mathcal{H}^n((P \circ Q \circ L)(A))$, then we are done.

To show this, consider any covering of $L(A)$. Under the isometry Q, the sets in this cover maintain the same Hausdorff measure as per Theorem 3.2. Since Q maps onto $\mathbb{R}^n \times \{0\}$, it follows that $(Q \circ L)(A)$ has the same cover as $(P \circ Q \circ L)(A)$ with the last k coordinates removed (which are 0 anyway).

In conclusion,

$$\mathcal{H}^n(L(A)) = \mathcal{H}^n((P \circ Q \circ L)(A)) = \mathcal{L}^n((P \circ Q \circ L)(A)) = [\![L]\!]\mathcal{L}^n(A). \quad \square$$

We now provide a proof of the area formula in the case where f is injective. This case follows the same reasoning as the general proof, but is less technical. We will apply the following two lemmas without proof.

Lemma 7.3 ([Tra94, Section 2]). *Let f and g map a set A such that*

$$|f(x) - f(y)| \leq c|g(x) - g(y)|,$$

for all $x, y \in A$. Then

$$\mathcal{H}^n(f(A)) \leq c^n \mathcal{H}^n(g(A)).$$

Lemma 7.4 ([Tra94, Section 2]). *Let $f : V \subset \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable, V and open set and $\lambda > 1$. Then $V_0 \coloneqq \{x \in V : Df \text{ is injective}\}$ is a disjoint union of a sequence of Borel sets B_i for which there exists T_i such that for all $x, y \in B_i$*

$$\frac{1}{\lambda}|T_i z| \leq |Df(x)z| \leq \lambda |Tz|, \quad \text{for all } z \in \mathbb{R}^n \tag{7.1}$$

and

$$\frac{1}{\lambda}|f(x) - f(y)| \leq |Tx - Ty| \leq \lambda |f(x) - f(y)|. \tag{7.2}$$

Proof of area formula for injective maps. By Rademacher's Theorem, we know that that $Df(x)$ and $Jf(x)$ exist $\mathcal{H}^n$-a.e. If $Df(x)$ has full rank, then $Jf(x) > 0$, otherwise $Jf(x) = 0$. So we spilt our analysis into two cases.

We first consider the case where $A \subseteq \{x : Jf(x) > 0\}$. Fix $\lambda > 1$ and choose B_i and T_i as in Lemma 7.4.

Let $A_i = A \cap B_i$. By inequality (7.1) and Lemma 7.3,

$$\frac{1}{\lambda^n}\mathcal{H}^n(T_i(Q)) \leq \mathcal{H}^n(Df(x)(Q)) \leq \lambda^n \mathcal{H}^n(T_i(Q)).$$

By applying Lemma 7.2 and cancelling the common factor of $\mathcal{H}^n(Q)$,

$$\frac{1}{\lambda^n}[\![T_i]\!] \leq Jf(x) \leq \lambda^n [\![T_i]\!].$$

If we integrate over A_i, we get that

$$\frac{1}{\lambda^n}\mathcal{H}^n(T_i(A_i)) \leq \int_{A_i} Jf \, d\mathcal{H}^n \leq \lambda^n \mathcal{H}^n(T_i(A_i)), \tag{7.3}$$

since $[\![T_i]\!] \int_{A_i} d\mathcal{H}^n = \mathcal{H}^n(T_i(A_i))$ by Lemma 7.2. Looking at the right hand side of the equation, we can apply inequality (7.2) and Lemma 7.3, to obtain the inequality

$$\mathcal{H}^n(T_i(A_i)) \leq \lambda^n \mathcal{H}^n(f(A_i)).$$

Combining this with the right hand side of 7.3, we obtain

$$\int_{A_i} Jf \, d\mathcal{H}^n \leq \lambda^{2n} \mathcal{H}^n(f(A_i)),$$

which we can sum over i to get

$$\int_A Jf \; d\mathcal{H}^n \leq \lambda^{2n} \mathcal{H}^n(f(A))$$

as we chose the A_i to be pairwise disjoint. Performing similar operations on the left hand side of (7.3) and taking the limit $\lambda \to 1$, it follows that

$$\int_A Jf \; d\mathcal{H}^n = \mathcal{H}^n(f(A)).$$

We now consider the case where $A \subseteq \{x : Jf(x) = 0\}$. In this case $\int_A Jf \; d\mathcal{H}^n = 0$, so we want to show that $\mathcal{H}^n(f(A)) = 0$. Fix $0 < \varepsilon \leq 1$. Consider the decomposition of f as $f = P \circ g$, where $g : \mathbb{R}^n \to \mathbb{R}^{m+n}$ is defined by

$$g(x) = (f(x), \epsilon x),$$

and $P : \mathbb{R}^{m+n} \to \mathbb{R}^m$ is the projection

$$P(y, z) = y.$$

It can be seen that

$$Dg(x) = \begin{pmatrix} Df(x) \\ \epsilon I \end{pmatrix}_{(n+m)\times n}$$

By the Binet-Cauchy formula $Jg(x)^2$ is equal to the sum of the squares of the $(n \times n)$-subdeterminants of Dg, so under the assumption that $Jf(x) = 0$,

$$Jg(x)^2 = Jf(x)^2 + p(\varepsilon) = p(\varepsilon) \tag{7.4}$$

where $p(\varepsilon)$ is a polynomial in which term involves at least one ϵ. By the asymptotic behaviour of polynomials at 0, we can bound (7.4) above by $c\epsilon^2$ for some c, so there exists C such that

$$Jg(x) \leq C\varepsilon.$$

It also follows from the Binet-Cauchy formula that

$$Jg(x)^2 \geq \varepsilon^{2n} > 0.$$

Since the map P is a projection, it follows that

$$\mathcal{H}^n(f(A)) \leq \mathcal{H}^n(g(A)) = \int_A Jg(x) \; \mathcal{H}^n \leq C\varepsilon \mathcal{H}^n(A).$$

Taking the limit $\epsilon \to 0$, we conclude that $\mathcal{H}^n(f(A)) = 0$. □

Finally, we state the following closely related Theorem without proof.

Theorem 7.5 (Change of variables [EG15, Section 3.3]). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous, $n \leq m$. Then for each $\mathcal{L}^n$-summable function $g : \mathbb{R}^n \to \mathbb{R}$,*

$$\int_{\mathbb{R}^n} g(x) Jf(x) \; dx = \int_{\mathbb{R}^m} \left[\sum_{x \in f^{-1}\{y\} g(x)} \right] d\mathcal{H}^n(y).$$

7.1 Applications

We show that the area formula agrees with the pre-existing measurement of length of a curve and surface area of a graph.

Example 7.1 (Length of curve). Suppose that $f : \mathbb{R} \to \mathbb{R}^m$ is Lipschitz continuous and 1:1. Write $f = (f^1, f^2, \ldots, f^m)$, then $Df = (\dot{f}^1, \ldots, \dot{f}^m)$ and

$$Jf = |Df| = |\dot{f}|.$$

Define the curve C as $C \coloneqq f([a,b]) \subset \mathbb{R}^m$, then by the area formula

$$\mathcal{H}^1(C) = \int_a^b Jf \, dx = \int_a^b |\dot{f}| \, dt.$$

Example 7.2 (Surface area of a graph). Let $u : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous. The graph M of u over $\mathbb{R}^n$ is given by

$$M = \{(x, u(x)) : x \in \mathbb{R}^n\}.$$

Define $f : \mathbb{R}^n \to \mathbb{R}^{n+1}$ by $f(x) = (x, u(x))$, then

$$Df = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ u_{x_1} & \cdots & u_{x_n} \end{pmatrix}.$$

The Binet-Cauchy formula (Theorem 6.3) tells us that $(Jf)^2$ is the sum of squares of the $n \times n$ sub-determinants of the matrix Df. It follows that

$$(Jf)^2 = 1 + |Du|^2.$$

By the area formula

$$\mathcal{H}^n(M) = \int_{\mathbb{R}^n} Jf \, d\mathcal{H}^n = \int_{\mathbb{R}^n} \sqrt{1 + |Du|^2} \, d\mathcal{H}^n.$$

8 Submanifolds of $\mathbb{R}^{n+l}$

In this section we will introduce the concepts of submanifold and tangent space, then discuss differentiability properties of locally Lipschitz maps. We will also extend the area formula to maps where the domain is a submanifold, and briefly discuss the Divergence Theorem.

Definition 8.1 ([Sim14, Section 2.4]). A ***n*-dimensional C^r submanifold** M of $\mathbb{R}^{n+l}$ is a subset of $\mathbb{R}^{n+l}$ such that for each $p \in M$, there exist open sets

$V \subset \mathbb{R}^n$, $W \subset \mathbb{R}^{n+l}$ with $p \in W$, and a 1:1 C^r map $\psi : V \to W$ with

$$\psi(V) = W \cap M.$$

We additionally require that ψ is proper, that is if $K \subset W$ is compact, then $\psi^{-1}(K)$ is compact in V. It is also important that $D\psi(x)$ has rank n at each point $x \in V$.

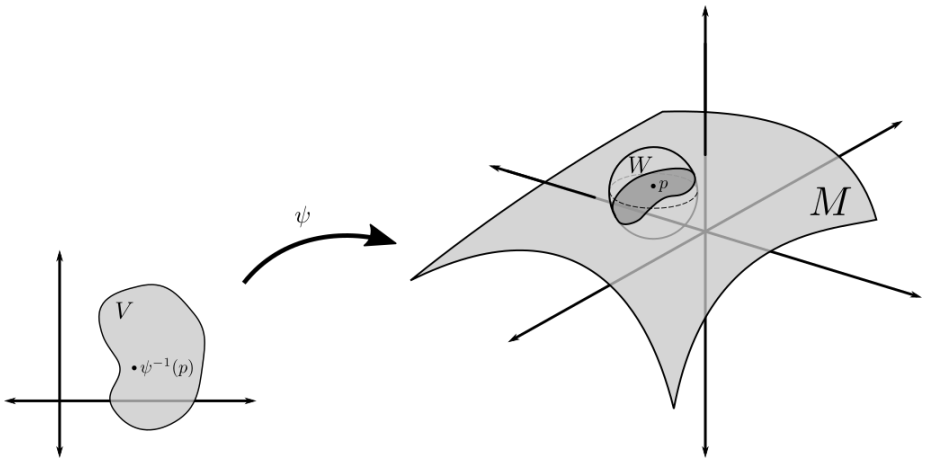


Figure 11: A 2-dimensional submanifold of $\mathbb{R}^3$

Definition 8.2 ([Sim14, Section 2.4])**.** The **tangent space** T_yM of M at y is the subspace of $\mathbb{R}^{n+l}$ consisting of those $\tau \in \mathbb{R}^{n+l}$ such that $\tau = \dot{\gamma}(0)$ for some C^1 curve $\gamma : (-1, 1) \to \mathbb{R}^{n+l}$. Where $\gamma(-1, 1) \subset M$, $\gamma(0) = y$.

We can equivalently think of the tangent space as a linear subspace spanned by the partial derivatives of ψ, as shown by the following Proposition.

Proposition 8.1. *T_yM is a linear subspace of $\mathbb{R}^{n+l}$ with basis $D_1\psi(x), \ldots, D_n\psi(x)$.*

Proof. Let $\gamma : (-1, 1) \to \mathbb{R}^{n+l}$ be a curve as in Definition 8.2 and $\psi : \mathbb{R}^n \to \mathbb{R}^{n+l}$ a local mapping onto the submanifold M as in Definition 8.1. Then we can write

$$\gamma(t) = \psi(x_1(t), x_2(t), \ldots, x_n(t)).$$

By the chain rule,

$$\frac{d\gamma}{dt} = \frac{\partial\psi}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial\psi}{\partial x_n}\frac{dx_n}{dt},$$

where $\frac{\partial\psi}{\partial x_i} = D_i\psi = \left(\frac{\partial\psi_1}{\partial x_i}, \ldots, \frac{\partial\psi_{n+l}}{\partial x_i}\right)^T$ for $i = 1, \ldots, n$. □

8.1 Differentiability properties of locally Lipschitz maps

We now want to investigate the differentiability properties for locally Lipschitz maps $f : M \to \mathbb{R}^P$ with $P \geq 1$, and ultimately obtain a similar result to Rademacher's Theorem. Unlike the previous Lipschitz maps discussed, these have an n-dimensional submanifold as their domain, rather than n-dimensional Euclidean space.

Consider $f : M \to \mathbb{R}^P$, for each $x \in M$ we assume that there are $\rho, L > 0$ with

$$|f(y) - f(z)| \leq L|y - z|, \quad y, z \in M \cap \mathbb{B}_\rho(x), \tag{8.1}$$

where $\mathbb{B}_\rho(x)$ is the closed ball with centre x and radius ρ. This map is **locally Lipschitz** in the region $M \cap \mathbb{B}_\rho(x)$.

Definition 8.3 ([Sim14, Section 2.4])**.** For a given $\tau \in T_y M$, the **directional derivative** is defined as

$$D_\tau f = \frac{d}{dt} f(\gamma(t))\bigg|_{t=0}$$

for any given C^1 curve $\gamma : (-1, 1) \to M$ with $\gamma(0) = y$ and $\dot{\gamma}(0) = \tau$, wherever this derivative exists.

In fact, since f is Lipschitz, the existence and value of the directional derivative are independent of the chosen curve γ, so long as the initial conditions remain the same [Sim14, Section 2.4]. To demonstrate this, for two different curves γ and $\tilde{\gamma}$, it follows from inequality (8.1) that

$$\lim_{t\to 0} \frac{|f(\gamma(t)) - f(\tilde{\gamma}(t))|}{t} \leq L \lim_{t\to 0} \frac{|\gamma(t) - \tilde{\gamma}(t)|}{t} = 0,$$

since $\gamma(0) = \tilde{\gamma}(0)$.

The following theorem in a local version of Rademacher's Theorem that allows us to take directional derivatives on a submanifold.

Proposition 8.2 ([Sim14, Section 2.4])**.** *There exists a set E of $\mathcal{H}^n-$measure zero such that $\forall x \in M\backslash E$, $D_\tau f(x)$ exists and the map $\tau \mapsto D_\tau f(x)$ is a linear map $T_x M \to \mathbb{R}^P$.*

Proof. Suppose $U \subset \mathbb{R}^{n+l}$, and let $\phi : U \cap M \to V \subset \mathbb{R}^n$ be a local coordinate chart. The inverse $\psi : V \to \mathbb{R}^{n+l}$ is C^1 with $\psi(V) = M \cap U$.

Let $f : M \to \mathbb{R}^P$ be a locally Lipschitz map. Since ψ is also Lipschitz as a C^1 function, it follows that $f \circ \psi : V \to \mathbb{R}^P$ is Lipschitz. By Rademacher's Theorem (Theorem 5.1) there is $E \subset V$ with Hausdorff measure zero such that $f \circ \psi$ is differentiable at every point of $V\backslash E$. In particular, for every $\eta \in \mathbb{R}^n$ and $x \in V\backslash E$ the directional derivative

$$D_\eta(f \circ \psi) = \frac{d}{dt} f(\psi(x + t\eta))\bigg|_{t=0}.$$

exists and is linear in η.

This gives us a curve $\gamma(t) = \psi(x + t\eta)$ with $\gamma(0) = x$ and $\dot{\gamma}(0) = \eta$. As per Definition 8.3, if we write $\tau = \sum_{i=1}^{n} \eta_i D_i \psi$, then for $y = \psi(x) \in M$

$$D_\eta(f \circ \psi)(x) = D_{\sum_{i=1}^{n} \eta_i D_i \psi(x)} f(y). \tag{8.2}$$

This tells us that the directional derivative does in fact exist and is linear in η for all $x \in V \backslash E_0$ and $y = \psi(x) \in U \cap M \backslash \psi(E)$.

Since $D_1\psi(x), \ldots, D_n\psi(x)$ form a basis of $T_{\psi(x)}M$, this means that the theorem does hold at points of $U \cap M \backslash \psi(E)$ with $\psi(E)$ being a set of $\mathcal{H}^n$-measure zero since ψ is locally Lipschitz on V. □

8.2 Return to the Area Formula

In Section 7, we discussed the area formula for Lipschitz maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ with $n \leq m$. Here we will show that the same area formula holds for locally Lipschitz maps from M to $\mathbb{R}^P$ where M is a n-dimensional submanifold and $P \geq n$.

As before, let M be a n-dimensional submanifold and let $\phi : U \cap M \to V$ be a local coordinate chart with $U \subset \mathbb{R}^{n+l}$ and $V \subset \mathbb{R}^n$. This chart has a C^1 inverse $\psi : V \to \mathbb{R}^{n+1}$, with $\psi(V) = M \cap U$. Let $f : M \to \mathbb{R}^P$ be locally Lipschitz. See Figure 12 for an illustration

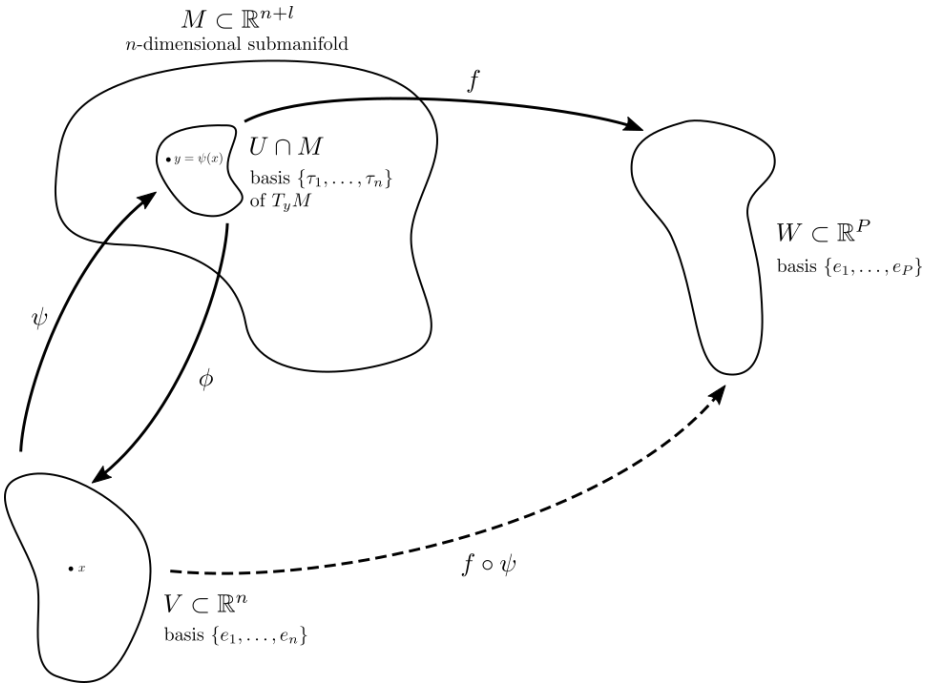


Figure 12: The maps and bases being considered in Section 8.2

From the proof of Proposition 8.2, consider the case in where $\eta = e_i$ in equation (8.2). In other words, we take a directional derivative in the direction of a coordinate vector. For an orthonormal basis $\tau_1, \ldots, \tau_n$ of T_yM, we can write

$$D_i\psi(x) = \sum_{l=1}^{n} (D_i\psi(x) \cdot \tau_l)\tau_l.$$

So by equation (8.2) with $y = \psi(x) \in U \cap M \backslash \psi(E)$,

$$\begin{aligned} D_i(f \circ \psi)(x) &= D_{D_i\psi(x)} f(y) \\ &= D_{\sum_{l=1}^{n} (D_i\psi(x)\cdot\tau_l)\tau_l} f(y) \\ &= \sum_{l=1}^{n} D_{\tau_l} f(y)(D_i\psi(x) \cdot \tau_l) \end{aligned}$$

with the last equality following from the linearity of the derivative.

Therefore,

$$D_i(f \circ \psi)(x) \cdot D_j(f \circ \psi)(x) = \sum_{k,m=1}^{n} (D_i\psi(x) \cdot \tau_k)(D_j\psi(x) \cdot \tau_m) D_{\tau_k} f(y) \cdot D_{\tau_m} f(y).$$

We can equivalently think of the right hand side as the product of three $n \times n$ matrices

$$\begin{pmatrix} D_1\psi \cdot \tau_1 & \cdots & D_1\psi \cdot \tau_n \\ \vdots & \ddots & \vdots \\ D_n\psi \cdot \tau_1 & \cdots & D_n\psi \cdot \tau_n \end{pmatrix} \begin{pmatrix} D_{\tau_1} f \cdot D_{\tau_1} f & \cdots & D_{\tau_n} f \cdot D_{\tau_1} f \\ \vdots & \ddots & \vdots \\ D_{\tau_n} f \cdot D_{\tau_1} f & \cdots & D_{\tau_n} f \cdot D_{\tau_n} f \end{pmatrix} \begin{pmatrix} D_1\psi \cdot \tau_1 & \cdots & D_n\psi \cdot \tau_1 \\ \vdots & \ddots & \vdots \\ D_1\psi \cdot \tau_n & \cdots & D_n\psi \cdot \tau_n \end{pmatrix},$$

or more succinctly,

$$(D_i\psi \cdot \tau_j)^* (D_{\tau_k} f \cdot D_{\tau_m} f)(D_i\psi \cdot \tau_j)$$

where i, j, k, m all range from 1 to n.

Since the determinant of square matrices is multiplicative, it follows that

$$\det\big(D_i(f \circ \psi) \cdot D_j(f \circ \psi)\big) = \det\big(D_{\tau_k} f \cdot D_{\tau_m}\big) \det\big((D_i\psi \cdot \tau_j)^* (D_i\psi \cdot \tau_j)\big). \quad (8.3)$$

Now, the $(i,j)^{\text{th}}$ entry of $(D_i\psi \cdot \tau_j)^*(D_i\psi \cdot \tau_j)$ is

$$\begin{aligned} \sum_{k=1}^{n} (D_i\psi \cdot \tau_k)(D_i\psi \cdot \tau_k) &= \sum_{k=1}^{n} \left(\sum_{m=1}^{n} (\alpha_m^i \tau_m) \cdot \tau_k \right) \left(\sum_{m=1} (\alpha_m^j)(\tau_m) \cdot \tau_k \right) \\ &= \sum_{k=1}^{n} \alpha_k^i \alpha_k^j \\ &= D_i\psi \cdot D_j\psi, \end{aligned}$$

since $D_i\psi \in \operatorname{span}\{\tau_1, \cdots, \tau_n\}$ and $\{\tau_1, \cdots, \tau_n\}$ forms an orthonormal basis, so $\tau_m \cdot \tau_k = \delta_{mk}$.

Therefore, we can rewrite equation (8.3) as

$$\det\left(D_i(f\circ\psi)(x)\cdot D_j(f\circ\psi(x))\right) \\ = \det\left((D_i\psi(x)\cdot D_j\psi(x))\right)\det\left(D_{\tau_k}f(y)\cdot D_{\tau_m}f(y)\right)$$

or taking square roots of both sides, and by Definition 6.4,

$$J(f\circ\psi)(x) = (J\psi(x))(Jf(y)) \tag{8.4}$$

In the above equation, we have $J\psi(x)$ with respect to an orthonormal basis $\{e_1,\dots,e_n\}$ of $V\subset\mathbb{R}^n$, whereas $Jf(y)$ is with respect of an orthonormal basis $\{\tau_1,\dots,\tau_n\}$ of T_yM as this is how we define a basis on a submanifold.

Now suppose that $P\geq n$ and f is injective, then by the area formula (Theorem 7.1)

$$\mathcal{H}^n(f(\psi(V))) = \int_V J(f\circ\psi)(x)\ d\mathcal{H}^n(x).$$

By equation (8.4),

$$\mathcal{H}^n(f(\psi(V))) = \int_V (J\psi(x))(Jf(\psi(x)))\ d\mathcal{H}^n(x).$$

If we now apply Theorem 7.5 (change of variables), then

$$\int_V (J\psi(x))(Jf(\psi(x)))\ d\mathcal{H}^n(x) = \int_{\psi(V)}\left[\sum_{x\in\psi^{-1}\{y\}} Jf(\psi(x))\right]d\mathcal{H}^n(y)$$

So it follows that

$$\mathcal{H}^n(f(A)) = \int_A Jf\ d\mathcal{H}^n, \tag{8.5}$$

where $A=\psi(V)\subset M$. So we can conclude that the area formula holds for $f:M\to\mathbb{R}^P$ Lipschitz.

8.3 Gradient and Divergence Theorem

From now on, we will restrict our attention to n-dimensional submanifolds of $\mathbb{R}^{n+1}$. These are known as hypersurfaces or surfaces of codimension one. A more general discussion of n-dimensional submanifolds of $\mathbb{R}^{n+l}$ can be found in Simon's *Introduction to Geometric Theory* [Sim14, Section 2.4].

Definition 8.4 ([Sim14, Section 2.4])**.** In the case that $P=1$ (so f is real-valued), we define the **gradient** $\nabla^M f$ of f by

$$\nabla^M f(y) = \sum_{j=1}^n (D_{\tau_j}f(y))\tau_j, \quad y\in M,$$

where $\tau_1,\dots,\tau_n$ is any orthonormal basis for T_yM. Note that $\nabla^M f(y)$ is a $(n+1)\times 1$ vector in the tangent space. Alternatively, if we let $\nabla^M_j f\equiv e_j\cdot\nabla^M f$ then we can write $\nabla^M f(y)=\sum_{j=1}^{n+1}\nabla^M_j f(y)e_j$.

Definition 8.5 ([Sim14, Section 2.4]). Given a vector field $X = (X^1, \dots, X^{n+1}) : M \to \mathbb{R}^{n+1}$ with $X^j \in C^1(M)$, $j = 1, \dots, n+1$, we define the **divergence**

$$\operatorname{div}_M X = \sum_{j=1}^{n+1} \nabla_j^M X^j.$$

Remark. It is not required that $X(y) \in T_y M$.

We now want to write the divergence in a form that includes directional derivatives. By Definition 8.5

$$\begin{aligned} \operatorname{div}_M X &= \sum_{j=1}^{n+1} e_j \cdot \left(\nabla^M X^j\right) \\ &= \sum_{j=1}^{n+1} e_j \cdot \left(\sum_{i=1}^{n} \left(D_{\tau_i} X^j\right) \tau_i \right). \end{aligned}$$

Since $X = \sum_{j=1}^{n+1} X^j e_j$, it follows that

$$\operatorname{div}_M X = \sum_{i=1}^{n} \tau_i \cdot D_{\tau_i} X. \tag{8.6}$$

We now state the following important theorem.

Theorem 8.3 (Divergence Theorem [Sim14, Section 2.4]). *If the closure $\overline{M}$ of M is a smooth compact manifold with boundary $\partial M = \overline{M} \backslash M$, and if $X(y) \in T_y M \ \forall y \in M$, then*

$$\int_M \operatorname{div}_M X \ d\mathcal{H}^n = -\int_{\partial M} X \cdot \eta \ d\mathcal{H}^{n-1} \tag{8.7}$$

where η is the inward pointing unit co-normal of ∂M. That is, $|\eta| = 1$, η is normal to ∂M, tangent to M and points into M at each point of ∂M.

9 The Second Fundamental Form

The second fundamental form is a very important concept in differential geometry that is closely related to the curvature of a surface. We first explore the second fundamental form of a hypersurface in $\mathbb{R}^3$ before giving a more general definition for hypersurfaces in $\mathbb{R}^{n+1}$.

9.1 The Second Fundamental Form in $\mathbb{R}^3$

Let $M \subset \mathbb{R}^3$ be a smooth surface and let $y \in M$. For a neighbourhood $V \subset M$ of y we can choose a smooth unit normal vector field $\nu : V \to \mathbb{S}^2$. This assigns to each $y \in M$ a unit normal vector, meaning that $\nu(y)$ is perpendicular to $T_y M$

for all $x \in V$. The map ν is called the **Gauss map** [dC76, Section 3.1]. The Gauss map is differentiable, with differential

$$T_y\nu : T_yM \to T_{\nu(y)}\mathbb{S}^2$$

known as the **shape operator**. Since T_pM and $T_{\nu(y)}\mathbb{S}^2$ are parallel planes, we can equally consider

$$T_y\nu : T_yM \to T_yM.$$

The shape operator measures the rate of chance of a normal vector relative to a curve on the surface and is therefore a natural candidate for the definition of curvature on a surface, which measures how much a surface deviates from being a plane.

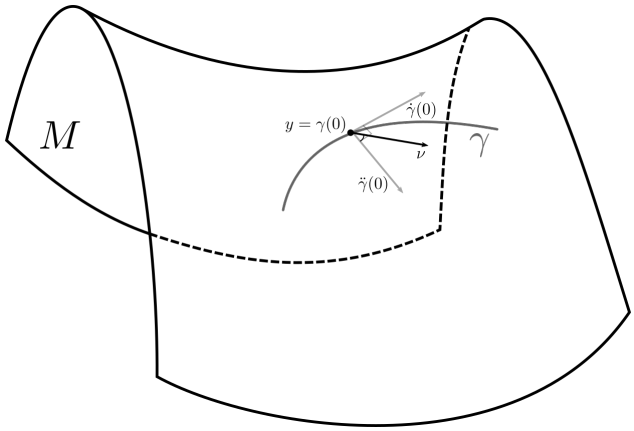


Figure 13: The set up to calculate normal curvature

Let $\gamma : (-1, 1)$ be a curve with $\gamma(0) = y$ and $\dot{\gamma}(0) = \tau$. The **normal curvature** of the surface is defined as $\ddot{\gamma} \cdot \nu$, or equivalently the component of the acceleration of the curve γ that is normal to the surface. A theorem by Meusnier states that normal curvature is dependent only on the tangent vector τ and not the curve γ [dC76, Section 3.2].

A beautiful result by Euler shows that eigenvalues of $-T_y\nu$ give us the maximum and minimum curvatures at a point; these are known as the **principal curvatures**. A more thorough discussion can be found in *A Comprehensive Introduction to Differential Geometry* [Spi99].

A concept closely related to the shape operator is the second fundamental form.

Definition 9.1 ([Spi99, Chapter 3B])**.** Let $M \subset \mathbb{R}^3$ be a smooth surface. The bilinear form $\mathrm{I\!I}_y : T_yM \times T_yM \to \mathbb{R}$ defined for $v, w \in T_yM$ by

$$\mathrm{I\!I}_y(v, w) = -T_y\nu(v) \cdot w$$

is called the **second fundamental form** of M.

If we choose an orthonormal basis of T_yM, then the second fundamental form is equivalent to the negative of the shape operator.

9.2 The Second Fundamental Form in $\mathbb{R}^{n+1}$

Now that we have the visual intuition, we can extend this definition to a n-dimensional submanifold M of $\mathbb{R}^{n+1}$.

Definition 9.2 ([Sim14, Section 2.4])**.** Let M be a submanifold that is at least C^2, we define the **second fundamental form** of M at y to be the bilinear form $B_y : T_yM \times T_yM \to (T_yM)^\perp$ such that

$$B_y(\tau, \eta) = -(\eta \cdot D_\tau \nu)\nu\Big|_y, \quad \tau, \eta \in T_yM,$$

where ν is a locally defined vector field with $\nu(z) \in (T_zM)^\perp$ for every z in some neighbourhood of y.

Remark. Definitions 9.1 and 9.2 are similar aside from the fact that B_y is a vector belonging to T_yM as opposed to $\mathrm{I\!I}_y$ which is real valued. The significance of this is that in higher codimensions, for example when we have a n-dimensional submanifold of $\mathbb{R}^{n+l}$, we have more than one 'normal vector', so the second fundamental form will be the vector sum with contributions from all of them. See *Introduction to Geometric Measure Theory* [Sim14, Section 2.4] for more details.

The second fundamental form measures the normal curvature of M in the direction τ. Suppose that $\tau \in T_yM$ with $|\tau| = 1$. Let $\gamma : (-1, 1) \to \mathbb{R}^{n+1}$ be a C^2 curve with $\gamma(0) = y$ and $\dot\gamma(0) = \tau$. Since the tangent of a curve is perpendicular to its normal, it follows that $\nu(\gamma(t)) \cdot \dot\gamma(t) = 0$. Differentiating yields

$$\begin{aligned} \nu(\gamma(0)) \cdot \ddot\gamma(0) &= -(D_\tau \nu) \cdot \dot\gamma(0) \\ \nu(y) \cdot \ddot\gamma(0) &= -(D_\tau \nu) \cdot \tau \end{aligned}$$

If we now multiply by ν, we get

$$(\ddot\gamma(0))^\perp = -(\tau \cdot D_\tau \nu)\nu = B_y(\tau, \tau). \tag{9.1}$$

where $(\ddot\gamma(0))^\perp = (\ddot\gamma \cdot \nu)\nu$ is the acceleration of the curve in the normal direction, and thus the normal curvature.

9.3 Mean curvature

The mean curvature of a surface is a measure of curvature important to the study of minimal surfaces which are defined as having zero mean curvature everywhere. It takes the mean average (or sometimes sum) of the curvatures along each basis element of the tangent space. By equation (9.1), the curvature in a direction τ is given by the second fundamental form B_y.

Definition 9.3 ([Sim14, Section 2.4])**.** The **mean curvature** vector $\underline{H}$ of M at y is $\operatorname{trace} B_y$, so

$$\underline{H}(y) = \sum_{i=1}^{n} B_y(\tau_i, \tau_i) \in (T_y M)^\perp,$$

where $\tau_1, \ldots, \tau_n$ is an orthonormal basis of $T_y M$.

Remark. The definition above (from *Introduction to Geometric Measure Theory* [Sim14, Section 2.4]) takes a sum rather than mean average. We will follow this definition.

By Definition 9.2 and equation (8.6),

$$\underline{H}(y) = -\sum_{i=1}^{n} (\tau_i \cdot D_{\tau_i} \nu) \nu(y) = -(\operatorname{div}_M \nu)\nu \tag{9.2}$$

near y.

In the case where the condition $X(y) \in T_y M$ is dropped, we can no longer apply the Divergence Theorem (Theorem 8.3). However, it is still possible to compute $\int_M \operatorname{div}_M X$.

We first use orthogonal decomposition to decompose X into tangent and normal parts

$$X = X^\top + X^\perp.$$

Since we are dealing with surfaces of codimension one, locally

$$X^\perp = (\nu \cdot X)\nu,$$

so near y,

$$\operatorname{div}_M X^\perp = (\nu \cdot X) \operatorname{div}_M \nu.$$

Taking the scalar product of X with both sides of equation (9.2),

$$\operatorname{div}_M X^\perp = -X \cdot \underline{H} \tag{9.3}$$

at each point of M. By equation (8.7) (the Divergence Theorem),

$$\int_M \operatorname{div}_M X^\top \, d\mathcal{H}^n = -\int_{\partial M} X \cdot \eta \, d\mathcal{H}^{n-1}. \tag{9.4}$$

Combining equations (9.3) and (9.4) with $\operatorname{div}_M X = \operatorname{div}_M X^\top + \operatorname{div}_M X^\perp$, we have

$$\int_M \operatorname{div}_M X \, d\mathcal{H}^n = -\int_M X \cdot \underline{H} \, d\mathcal{H}^n - \int_{\partial M} X \cdot \eta \, d\mathcal{H}^{n-1} \tag{9.5}$$

as a more general version of the Divergence Theorem.

10 First and Second Variation Formulae

We now discuss the first and second variation formula, which allow us to analyse the perturbations of a family of manifolds over time.

As before, we assume M is a n-dimensional C^1 submanifold of $\mathbb{R}^{n+1}$ such that $U \cap M \neq \varnothing$ for any open subset $U \in \mathbb{R}^{n+1}$. In addition, $\mathcal{H}^n(K \cap M) < \infty$ for each compact $K \subset U$.

We want to start by considering a 1-parameter family of diffeomorphisms $\{\phi_t\}_{-1 \leq t \leq 1} : U \to U$ such that for compact $K \subset U$

$$\begin{cases} \phi_t(x) = \phi(t,x) \text{ is a } C^2 \text{ map of } (t,x) \in (-1,1) \times U \to U \\ \phi_0(x) = x, \quad \forall x \in U \\ \phi_t(x) = x, \quad \forall (t,x) \in (-1,1) \times U \backslash K \end{cases}$$

[Sim14, Section 2.5].

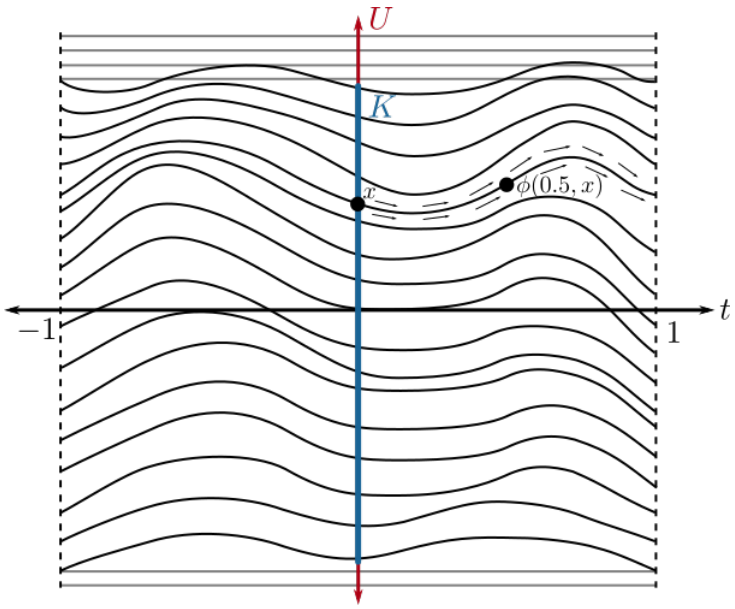


Figure 14: Example of a diffeomorphism where U is the real line

Let the initial velocity vector for ϕ_t be denoted

$$X_x = \left.\frac{\partial \phi(t,x)}{\partial t}\right|_{t=0}$$

and let the initial acceleration vector for ϕ_t be denoted

$$Z_x = \left.\frac{\partial^2 \phi(t,x)}{\partial t^2}\right|_{t=0}.$$

Then X and Z both have compact support. Using a Taylor expansion around $t = 0$, we derive $\phi_t(x) = x + tX_x + \frac{t^2}{2}Z_x + O(t^3)$.

We want to do a similar thing for a submanifold M. Let M_t be a 1-parameter family of manifolds with the following properties:

$$\begin{cases} M_t = \phi_t(M \cap K) \text{ for compact } K \subset U, \\ M_0 = M \cap K \text{ for compact } K \subset U, \\ M_t = M \text{ outside some compact subset of } U. \end{cases}$$

We now want to compute the first and second variation of M, defined as follows.

Definition 10.1 ([Sim14, Section 2.5])**.** The **first variation** of M is defined as

$$\frac{d}{dt}\mathcal{H}^n(M_t)\bigg|_{t=0}.$$

The **second variation** of M is defined as

$$\frac{d^2}{dt^2}\mathcal{H}^n(M_t)\bigg|_{t=0}.$$

The area formula for submanifolds (equation (8.5)), can be used to find the area $\mathcal{H}^n(\psi_t(M \cap K))$ as an integral of $J\psi_t$ where $\psi_t = \phi_t|_{M\cap U}$:

$$\mathcal{H}^n(\psi_t(M \cap K)) = \int_{M\cap K} J\psi_t \, d\mathcal{H}^n. \tag{10.1}$$

To compute this integral, we want to derive a Taylor series expansion for $J\psi_t$. This is a tedious calculation and details can be found in *Introduction to Geometric Measure Theory* [Sim14]. The final result is

$$\begin{aligned} J\psi_t &= 1 + t\,\mathrm{div}_M X \\ &\quad + \frac{t^2}{2}\left(\mathrm{div}_M Z + (\mathrm{div}_M X)^2 + \sum_{i=1}^{n} |(D_{\tau_i}X)^\perp|^2 - \sum_{i,j=1}^{n} (\tau_j \cdot D_{\tau_j}X)(\tau_j \cdot D_{\tau_i}X)\right) \\ &\quad + O(t^3). \end{aligned}$$

Using the equation (10.1) above, the first variation is

$$\frac{d}{dt}\mathcal{H}^n(M_t)\bigg|_{t=0} = \int_{M\cap K} \frac{d}{dt} J\psi_t \, d\mathcal{H}^n\bigg|_{t=0} = \int_M \mathrm{div}_M X \, d\mathcal{H}^n. \tag{10.2}$$

Similarly, the second variation is

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{H}^n(M_t)\bigg|_{t=0} &= \int_{M\cap K} \frac{d^2}{dt^2} J\psi_t \, d\mathcal{H}^n\bigg|_{t=0} \\ &= \int_M \mathrm{div}_M Z + (\mathrm{div}_M X)^2 + \sum_{i=1}^{n} |(D_{\tau_i}X)^\perp|^2 \\ &\qquad - \sum_{i,j=1}^{n} (\tau_j \cdot D_{\tau_j}X)(\tau_j \cdot D_{\tau_i}X) \, d\mathcal{H}^n. \end{aligned} \tag{10.3}$$

Definition 10.2 ([Sim14, Section 2.5])**.** We say that M is **stationary** in U if $\mathcal{H}^n(M \cap K) < \infty$ for each compact $K \subset U$ and if $\frac{d}{dt}\mathcal{H}^n(M_t)\big|_{t=0} = 0$.

By the first variation (10.2), this is equivalent to

$$\int_M \operatorname{div}_M X \, d\mathcal{H}^n = 0$$

when X is C^1 with compact support. Recall the generalised Divergence Theorem, equation (9.5):

$$\int_M \operatorname{div}_M X \, d\mathcal{H}^n = -\int_M X \cdot \underline{H} \, d\mathcal{H}^n - \int_{\partial M} X \cdot \eta \, d\mathcal{H}^{n-1}.$$

Using this we have the following lemma.

Lemma 10.1 ([Sim14, Section 2.5])**.** *Suppose M is a C^2 submanifold of $\mathbb{R}^{n+1}$ with mean curvature $\underline{H}$ and $U \subset \mathbb{R}^{n+1}$ is open. Then*

(i) If $\overline{M}$ is a C^2 submanifold with smooth $(n-1)$-dimensional boundary $\partial M = \overline{M} \backslash M$, then M is stationary in U if and only if $\underline{H} \equiv 0$ on $M \cap U$ and $\partial M \cap U = \varnothing$.

(ii) If $\overline{U} \cap M$ is a compact subset of M, then M is stationary in U if and only if $\underline{H} = 0$ on $M \cap U$.

11 Bernstein's Problem

Bernstein's Problem is a problem in minimal surface theory which asks: given that the graph of $f : \mathbb{R}^n \to \mathbb{R}$ is a minimal surface in $\mathbb{R}^{n+1}$, is the function affine linear? This turns out to be true for n up to 7, and false for higher dimensions. Here we will provide a proof of the case $n = 2$.

Theorem 11.1 (Bernstein's Problem)**.** *Let $u : \mathbb{R}^2 \to \mathbb{R}$ be a function such that $M = \operatorname{graph}(u)$ is a minimal surface, then u is affine linear.*

11.1 Minimal surfaces in $\mathbb{R}^{n+1}$

We first take a step back and prove some general results about graphs that are minimal surfaces in $\mathbb{R}^{n+1}$. By Lemma 10.1, we know that for a function $u : \mathbb{R}^n \to \mathbb{R}$, its graph $M = \operatorname{graph}(u) \subset \mathbb{R}^{n+1}$ is a minimal surface if and only if $\underline{H}(\operatorname{graph}(u)) = 0$.

It can easily be checked that the downward normal to M is the $(n+1)$-dimensional vector

$$\nu = \frac{1}{\sqrt{1 + |Du|^2}}(Du, -1).$$

Proposition 11.2. *Suppose $M = \operatorname{graph}(u)$ is a minimal surface, then*

$$\operatorname{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\right) = 0.$$

Proof. In Section 7.1, we showed that by the area formula,

$$\mathcal{H}^n(M) = \int_{\mathbb{R}^n} \sqrt{1 + |Du|^2}\, d\mathcal{H}^n.$$

We now consider diffeomorphisms of the form $\phi_t = u + t\eta$, where $\eta : \mathbb{R}^n \to \mathbb{R}$ has compact support. This corresponds to perturbations of the surface M as we change the value in the $(n+1)^{\text{th}}$ dimension. The first variation is given by

$$\frac{d}{dt}\mathcal{H}^n(M_t) = \frac{d}{dt}\mathcal{H}^n(\phi_t(M \cap K)).$$

It follows that

$$\frac{d}{dt}\int_{\mathbb{R}^n} \sqrt{1 + |D(u + t\eta)|^2}\, d\mathcal{H}^n \bigg|_{t=0} = \int_{\mathbb{R}^n} \frac{Du \cdot D\eta}{\sqrt{1 + |Du|^2}}\, d\mathcal{H}^n. \tag{11.1}$$

Now consider

$$\begin{aligned}
\operatorname{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\eta\right) &= \sum_{i=1}^n \frac{\partial}{\partial x_i}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\eta\right) \\
&= \sum_{i=1}^n \eta_{x_i} \frac{u_{x_i}}{\sqrt{1 + |Du|^2}} + \eta \sum_{i=1}^n \frac{\partial}{\partial x_i}\frac{Du}{\sqrt{1 + |Du|^2}} \\
&= D\eta \cdot \frac{Du}{\sqrt{1 + |Du|^2}} + \eta \operatorname{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\right).
\end{aligned}$$

Hence, we can rewrite the right hand side of equation (11.1) as

$$\int_{\mathbb{R}^n} \operatorname{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\eta\right) d\mathcal{H}^n - \int_{\mathbb{R}^n} \operatorname{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\right)\eta\, d\mathcal{H}^n. \tag{11.2}$$

Since η has compact support, it takes the value zero outside of a compact set K, so choose $U \supset K$, then by the divergence theorem

$$\int_U \operatorname{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\eta\right) d\mathcal{H}^n = -\int_{\partial U} \frac{Du}{\sqrt{1 + |Du|^2}}\eta \cdot n_+\, d\mathcal{H}^{n-1} = 0,$$

since $\eta = 0$ on ∂U. Therefore the first integral in equation (11.2) is equal to zero, so we have

$$\frac{d}{dt}\mathcal{H}^n(M_t) = -\int_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1 + |Du|^2}}\eta\, d\mathcal{H}^n\right) \tag{11.3}$$

Alternatively, we can instead define the first variation as per equation (10.2), and the generalised Divergence Theorem (equation (9.5)) as

$$\frac{d}{dt}\mathcal{H}^n(M_t) = \int_M \operatorname{div}_M(X)\, d\mathcal{H}^n = -\int_M X \cdot \underline{H}\, d\mathcal{H}^n - \int_{\partial M} X \cdot n_+\, d\mathcal{H}^n,$$

where $X = \eta\ e_{n+1}$ and $\underline{H} = -H\nu = -H\frac{(Du,-1)}{\sqrt{1+|Du|^2}}$. The second term on the right hand side vanishes as before because η has compact support. By the change of variables formula (Theorem 7.5),

$$-\int_M X \cdot \underline{H}\ d\mathcal{H}^n = -\int_{\mathbb{R}^n} (Ju)\ X \cdot \underline{H}\ d\mathcal{H}^n.$$

Now, $Ju = \sqrt{1+|Du|^2}$ and $X \cdot \underline{H} = H\frac{1}{\sqrt{1+|Du|^2}}\eta$. Therefore

$$-\int_{\mathbb{R}^n} (Ju)\ X \cdot \underline{H}\ d\mathcal{H}^n = -\int_{\mathbb{R}^n} H\eta\ d\mathcal{H}^n. \tag{11.4}$$

Equations (11.3) and (11.4) are both representations of the first variation of area, so combining them we get

$$\int_{\mathbb{R}^n} \mathrm{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1+|Du|^2}}\right)\eta\ d\mathcal{H}^n = \int_{\mathbb{R}^n} H\eta\ d\mathcal{H}^n.$$

It follows that since M is a minimal surface, $0 = H = \mathrm{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1+|Du|^2}}\right)$. □

Define $\hat{\nu}$ on $\mathbb{R}^{n+1}$ by $\hat{\nu}(x, x_{n+1}) = \nu(\tilde{x})$ where $x \in \mathbb{R}^n$ and $\tilde{x} = (x, u(x))$ belongs to $\mathbb{R}^{n+1}$.

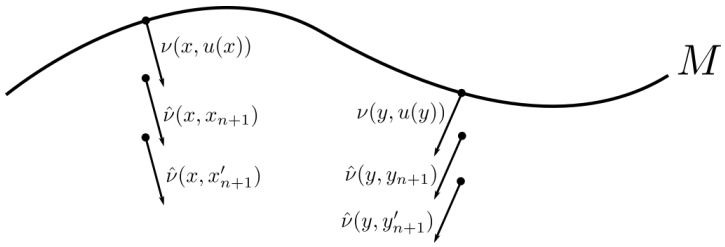


Figure 15: Illustration of the normal field $\hat{\nu}$

Then

$$\begin{aligned}\mathrm{div}_{\mathbb{R}^{n+1}}(\hat{v}) &= \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\frac{1}{\sqrt{1+|Du|^2}}u_{x_i} - \frac{\partial}{\partial x_{n+1}}\frac{1}{\sqrt{1+|Du|^2}} \\ &= \mathrm{div}_{\mathbb{R}^n}\left(\frac{Du}{\sqrt{1+|Du|^2}}\right) = 0\end{aligned}$$

by Proposition 11.2. The term $\frac{\partial}{\partial x_{n+1}}\frac{1}{\sqrt{1+|Du|^2}}$ vanishes because it is only a function of x, not x_{n+1}.

Proposition 11.3. *Let $u : \mathbb{R}^n \to \mathbb{R}$ be a function with graph $M = \mathrm{graph}(u)$ a minimal surface. Suppose that M' is any other graph such that $M' = M$ outside of a compact set $K \subset \mathbb{R}^{n+1}$. Then*

$$\mathcal{H}^n(M' \cap K) \geq \mathcal{H}^n(M \cap K).$$

Proof. Let Ω be the region in compact $K \subset \mathbb{R}^{n+1}$ enclosed between $M \cap K$ and $M' \cap K$. Then $\partial\Omega = (M \cup M') \cap K$. On $\mathbb{R}^{n+1} \backslash K$, $M = M'$, so there is no region between them.

By the Divergence Theorem (8.3),

$$0 = \int_\Omega \operatorname{div}_{\mathbb{R}^{n+1}}(\hat{\nu})\, d\mathcal{H}^{n+1} = -\int_{M\cap K} \hat{\nu}\cdot\eta\, d\mathcal{H}^n - \int_{M'\cap K} \hat{\nu}\cdot\eta\, d\mathcal{H}^n,$$

where η is the inward facing normal for $\partial\Omega$.

Because $\hat{v}$ is a downwards facing normal, but the Divergence Theorem uses an inward pointing normal vector to $\partial\Omega$, our calculations chance depending on whether u is above or below v. Therefore, we want to divide up K into regions where $u < v$, $u > v$ and $u = v$, call them K^+, K^- and K^0.

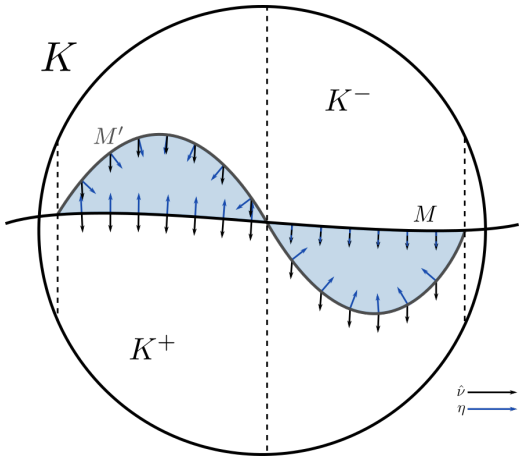


Figure 16: Relative directions of $\hat{\nu}$ and η

Case 1

In the case where $u < v$, $\hat{\nu}$ and η point in opposite directions on $M \cap K$ and in the same direction on $M' \cap K$. Therefore $\hat{\nu}\cdot\eta = -\hat{\nu}\cdot\hat{\nu} = -1$, and

$$\begin{aligned}
\mathcal{H}^n(M\cap K) &= -\int_{M\cap K} \hat{\nu}\cdot\eta\, d\mathcal{H}^n \\
&= \int_{M'\cap K} \hat{\nu}\cdot\eta\, d\mathcal{H}^n \\
&\le \int_{M'\cap K} |\hat{\nu}\cdot\eta|\, d\mathcal{H}^n \\
&\le \int_{M'\cap K} |\hat{\nu}|\,|\eta|\, d\mathcal{H}^n \\
&= \int_{M'\cap K} d\mathcal{H}^n = \mathcal{H}^n(M'\cap K).
\end{aligned}$$

Case 2

In the case where $u > v$, $\hat{\nu}$ and η point in the same direction on $M \cap K$ and in opposite directions on $M' \cap K$. Here, $\hat{\nu} \cdot \eta = 1$, and

$$\mathcal{H}^n(M \cap K) = \int_{M \cap K} \hat{\nu} \cdot \eta \; d\mathcal{H}^n = -\int_{M' \cap K} \hat{\nu} \cdot \eta \; d\mathcal{H}^n.$$

The result follows as in the first case, as $-\hat{\nu} \cdot \eta \leq |\hat{\nu} \cdot \eta|$.

Case 3

In the case where $u = v$, there is no region between $M \cap K$ and $M' \cap K$, so $\mathcal{H}^n(M \cap K) = \mathcal{H}^n(M' \cap K)$.

Now for any arbitrary set K, since $\mathcal{H}^n(M \cap K) \leq \mathcal{H}^n(M' \cap K)$ on K^+, K^- and K^0, it follows that the property holds on all of K. □

From the above, we conclude that M locally minimises area. We now want to find a bound for the intersection of a minimal surface with a n-dimensional ball.

Proposition 11.4. *Let $u : \mathbb{R}^n \to \mathbb{R}$ be a function with $M = \text{graph}(u)$ a minimal surface, then for $x \in M$*

$$\mathcal{H}^n(M \cap \mathbb{B}^{n+1}(x, r)) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x, r)),$$

where $\mathbb{B}^{n+1}(x, r) = \{y \in \mathbb{R}^{n+1} : |y - x| \leq r\}$ is the $(n+1)$-dimensional ball of radius r and centre x.

Proof. Suppose we allow the submanifold M to intersect $\mathbb{B}^{n+1}(x, r)$ in any number of places. Let M_{in} be the subset of M that lies inside the ball and M_{out} be the subset of M that lies outside the ball. Let A be the subset of $\partial\mathbb{B}^{n+1}(x, r)$ that would result from perturbing M_{in} to the boundary of the ball. Let B be the subset $\partial\mathbb{B}^{n+1}(x, r) \backslash A$.

With this division, $\partial A = \partial B = M \cap \partial\mathbb{B}^{n+1}(x, r)$. Moreover, we know that $\mathcal{H}^n(A) + \mathcal{H}^n(B) = \mathcal{H}^n(\partial\mathbb{B}^{n+1}(x, r))$, so this results in two cases:

(i) $\mathcal{H}^n(A) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x, r))$, or

(ii) $\mathcal{H}^n(B) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x, r))$.

Case 1

We first consider the case where $\mathcal{H}^n(A) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x, r))$. Let Ω be the region between A and M_{in}. By the Divergence Theorem (8.3),

$$0 = \int_\Omega \text{div}_{\mathbb{R}^{n+1}}(\hat{\nu}) \; d\mathcal{H}^n = -\int_{M \cap \mathbb{B}^{n+1}(x,r)} \hat{\nu} \cdot \eta \; d\mathcal{H}^n - \int_A \hat{\nu} \cdot \eta \; d\mathcal{H}^n, \qquad (11.5)$$

where η is the inward facing normal for $\partial\Omega$.

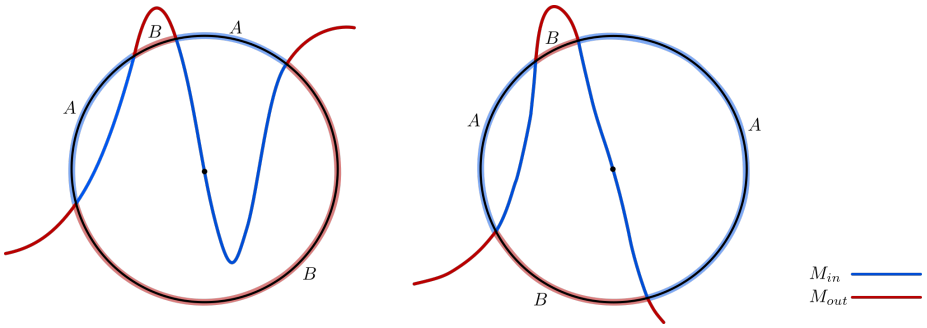


Figure 17: An example of the regions A and B in $\mathbb{R}^2$

On $M \cap \mathbb{B}^{n+1}(x,r)$, we have that $\eta = -\hat{\nu}$, so

$$-\int_{M\cap\mathbb{B}^{n+1}(x,r)} \hat{\nu}\cdot\eta \; d\mathcal{H}^n = \mathcal{H}^n(M\cap\mathbb{B}^{n+1}(x,r)).$$

On A, we are on the boundary of a ball, so $\eta = -\frac{y}{|y|}$ for $y \in A$.

Therefore by equation (11.5)

$$\begin{aligned}\mathcal{H}^n(M\cap\mathbb{B}^{n+1}(x,r)) &= \int_A \hat{\nu}\cdot\frac{y}{|y|}\; d\mathcal{H}^n\{y\}\\ &\leq \int_A |\hat{\nu}|\frac{|y|}{|y|}\; d\mathcal{H}^n\\ &= \mathcal{H}^n(A) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x,r)).\end{aligned}$$

Case 2

Now we consider the case where $\mathcal{H}^n(B) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x,r))$. Again, let Ω be the region enclosed by B and M_{in}. We apply the Divergence Theorem (8.3) again, to show that

$$0 = \int_\Omega \mathrm{div}_{\mathbb{R}^{n+1}}(\hat{\nu})\; d\mathcal{H}^n = -\int_{M\cap\mathbb{B}^{n+1}(x,r)} \hat{\nu}\cdot\eta\; d\mathcal{H}^n - \int_B \hat{\nu}\cdot\eta\; d\mathcal{H}^n, \qquad (11.6)$$

where η is the inward facing normal for $\partial\Omega$.

In this case we have $\eta = \hat{\nu}$ on $M \cap \mathbb{B}^{n+1}(x,r)$, so

$$-\int_{M\cap\mathbb{B}^{n+1}(x,r)} \hat{\nu}\cdot\eta\; d\mathcal{H}^n = -\mathcal{H}^n(M\cap\mathbb{B}^{n+1}(x,r)).$$

Moreover, for $y \in B$ we have $\eta = -\frac{y}{|y|}$ as before.

Therefore, by equation (11.6)

$$\begin{aligned}\mathcal{H}^n(M \cap \mathbb{B}^{n+1}(x,r)) &= \int_B \hat{\nu} \cdot \frac{-y}{|y|}\, d\mathcal{H}^n \\ &\leq \int_B \left| \hat{\nu} \cdot \frac{y}{|y|} \right| d\mathcal{H}^n \\ &= \int_B |\hat{\nu}| \frac{|y|}{|y|}\, d\mathcal{H}^n \\ &= \mathcal{H}^n(B) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x,r)).\end{aligned}$$

From these cases, we conclude that if the graph M is a minimal surface, then $\mathcal{H}^n(M \cap \mathbb{B}^{n+1}(x,r)) \leq \frac{1}{2}\mathcal{H}^n(\partial\mathbb{B}^{n+1}(x,r))$. □

Since the n-dimensional volume of a ball of radius r is proportional to r^n, it follows that

$$\mathcal{H}^n(M \cap \mathbb{B}^{n+1}(x,r)) \leq cr^n \tag{11.7}$$

for some constant c.

Proposition 11.5. *For a locally minimising graph $M = \mathrm{graph}(u)$ for $u : \mathbb{R}^n \to \mathbb{R}$ as defined above, we have the following inequality*

$$\int_M |\nabla^M \zeta|^2 \, d\mathcal{H}^n \geq \int_M \zeta^2 |B|^2 \, d\mathcal{H}^n \quad \forall \zeta \in C_c^{0,1}(M).$$

where

$$|B|^2 \coloneqq \sum_{i,j=1}^{n} |B(\tau_i, \tau_j)|^2.$$

Proof. By equation (10.3) in Section 10, the second variation formula is given by

$$\begin{aligned}\frac{d^2}{dt^2}\mathcal{H}^n(M_t)\bigg|_{t=0} = \int_M \mathrm{div}_M\, Z + (\mathrm{div}_M\, X)^2 + \sum_{i=1}^{n} |(D_{\tau_i} X)^\perp|^2 \\ - \sum_{i,j=1}^{n} (\tau_i \cdot D_{\tau_j} X)(\tau_j \cdot D_{\tau_i} X)\, d\mathcal{H}^n,\end{aligned}$$

where X is the initial velocity vector and Z is the initial acceleration vector. Since we are considering minimal surfaces without boundary, we can apply the generalised Divergence Theorem, (9.5), to show that

$$\int_M \mathrm{div}_M\, Z \, d\mathcal{H}^n = -\int_M Z \cdot \underline{H} \, d\mathcal{H}^n - \int_{\partial M} Z \cdot \eta \, d\mathcal{H}^{n-1} = 0,$$

since $\underline{H} \equiv 0$ on M and $\partial M = \varnothing$. Moreover, because $X \in (T_y M)^\perp$, by equation (9.3), we have $\operatorname{div}_M X = -X \cdot \underline{H} = 0$. Therefore for a minimal graph M,

$$\frac{d^2}{dt^2}\mathcal{H}^n(M_t)\bigg|_{t=0} = \int_M \sum_{i=1}^n |(D_{\tau_i} X)^\perp|^2 - \sum_{i,j=1}^n (\tau_i \cdot D_{\tau_j} X)(\tau_j \cdot D_{\tau_i} X)\, d\mathcal{H}^n.$$

By the definition of the second fundamental form, we can rewrite $\tau_i \cdot D_{\tau_j} X$ as $-X \cdot B(\tau_i, \tau_j)$, so it follows that

$$\frac{d^2}{dt^2}\mathcal{H}^n(M_t)\bigg|_{t=0} = \int_M \sum_{i=1}^n |(D_{\tau_i} X)^\perp|^2 - \sum_{i,j=1}^n (X \cdot B(\tau_i, \tau_j))^n\, d\mathcal{H}^n. \tag{11.8}$$

Now suppose that $X = \zeta\nu$ for scalar function ζ with compact support. Then

$$(D_{\tau_i} X)^\perp = (D_{\tau_i}(\zeta\nu))^\perp = \nu D_{\tau_i}\zeta$$

since $D_{\tau_i}\nu \in T_y M$, so we only care about $D_{\tau_i}(\zeta)$ in the direction ν. As a result,

$$\sum_{i=1}^n |(D_{\tau_i} X)^\perp|^2 = \sum_{i=1}^n (\nu D_{\tau_i}\zeta) \cdot (\nu D_{\tau_i}\zeta) = \sum_{i=1}^n |D_{\tau_i}\zeta|^2. \tag{11.9}$$

We also have that $\nabla^M \zeta = \sum_{j=1}^n (D_{\tau_i}\zeta)\tau_i$, so

$$|\nabla^M \zeta|^2 = \sum_{i=1}^n |D_{\tau_i}\zeta|^2, \tag{11.10}$$

since the τ_is form an orthonormal basis.

Combining equations (11.9) and (11.10), we obtain

$$\sum_{i=1}^n |(D_{\tau_i} X)^\perp|^2 = |\nabla^M \zeta|^2. \tag{11.11}$$

In addition, considering the second term in equation (11.8), we have

$$\sum_{i,j=1}^n (X \cdot B(\tau_i, \tau_j))^2 = \sum_{i,j=1}^n (\zeta\nu \cdot B(\tau_i, \tau_j))^2 = \zeta^2 \sum_{i,j=1}^n |B(\tau_i, \tau_j)|^2 = \zeta^2 |B|^2, \tag{11.12}$$

since $|\nu|^2 = 1$.

Finally, combining equations (11.11) and (11.12) with equation (11.8), we have the following simplified integral for the second variation

$$\frac{d^2}{dt^2}\mathcal{H}^n(M_t)\bigg|_{t=0} = \int_M (|\nabla^M \zeta|^2 - \zeta^2 |B|^2)\, d\mathcal{H}^n.$$

From Proposition 11.3, we know that M is a local minimum, so it follows that $\frac{d^2}{dt^2}\mathcal{H}^n(M_t)\Big|_{t=0} \geq 0$ and

$$\int_M |\nabla^M \zeta|^2\, d\mathcal{H}^n \geq \int_M \zeta^2 |B|^2\, d\mathcal{H}^n \quad \forall \zeta \in C_c^{0,1}(M). \qquad \square$$

11.2 Proof of Bernstein's Problem for Minimal Hypersurfaces in $\mathbb{R}^3$

We are now ready to prove Theorem 11.1. The proof given here is based on the one given in *Measure Theory and Functions of Bounded Variation* [Giu84, Section 17], and is very different from the proof given by Bernstein in 1915, which is more topological. In this proof the function $u : \mathbb{R}^n \to \mathbb{R}$ is assumed to be C^2, otherwise the second fundamental form is not defined.

Proof of Theorem 11.1. We now assume that $n = 2$, so we have a 2-dimensional minimal graph embedded in $\mathbb{R}^3$. From Proposition 11.5, we know that

$$\int_M \zeta^2 |B|^2 \, d\mathcal{H}^n \leq \int_M |\nabla^M \zeta|^2 \, d\mathcal{H}^n \tag{11.13}$$

for all ζ with compact support.

The following uses the log cut-off trick from *Minimal Surfaces and Functions of Bounded Variation* [Giu84, Section 17] We now consider the function,

$$\zeta_j = \begin{cases} 1 & |x| \leq j \\ 2 - \frac{\log |x|}{\log j} & j < |x| < j^2 \\ 0 & |x| \geq j^2 \end{cases}$$

for all $j \in \mathbb{N}$.

We define

$$\nabla^M \zeta(x) = \sum_{j=1}^{n} (D_{\tau_j} \zeta(x)) \tau_j, \quad x \in M,$$

where $\tau_1, \ldots, \tau_n$ is any orthonormal basis for $T_x M$. It follows that

$$|\nabla^M \zeta(x)|^2 \leq |D\zeta|^2 \tag{11.14}$$

since we lose the normal component of each derivative on the left hand side when we project to the tangent vector.

By equation (11.13) and inequality (11.14), it follows that for $\mathbb{B}(j)$, a 3-dimensional ball of radius j,

$$\int_{M \cap \mathbb{B}(j)} |B|^2 \, d\mathcal{H}^2 \leq \int_M |\nabla^M \zeta_j|^2 \, d\mathcal{H}^2 \leq \int_M |D\zeta_j|^2 \, d\mathcal{H}^2. \tag{11.15}$$

Let $\Sigma_j = M \cap (\mathbb{B}(j^2) - \mathbb{B}(j))$, then

$$\begin{aligned} \int_M |D\zeta_j|^2 \, d\mathcal{H}^2 &= \frac{1}{(\log j)^2} \int_{\Sigma_j} \frac{1}{|x|^2} \, d\mathcal{H}^n \\ &= \frac{1}{(\log j)^2} \int_0^\infty \mathcal{H}^2 \left\{ x \in \Sigma_j : \frac{1}{|x|^2} > t \right\} dt. \end{aligned} \tag{11.16}$$

Equivalently, we are choosing $x \in \Sigma_j$ such that $|x| < t^{-\frac{1}{2}}$. In Σ_j, $j < |x| < j^2$, we consider two bounds for t: $j < |x| < t^{-\frac{1}{2}} < j^2$:

(i) $j < |x| < t^{-\frac{1}{2}} < j^2$, and

(ii) $j < |x| < j^2 < t^{-\frac{1}{2}}$.

In the first case, $\frac{1}{j^4} < t < \frac{1}{j^2}$, and

$$\left\{ x \in \Sigma_j : \frac{1}{|x|^2} > t,\ \frac{1}{j^4} < t < \frac{1}{j^2} \right\} = M \cap \mathbb{B}\Big(t^{-\frac{1}{2}}\Big).$$

In the second case, $t < \frac{1}{j^4}$, and

$$\left\{ x \in \Sigma_j : \frac{1}{|x|^2} > t,\ t < \frac{1}{j^4} \right\} \subseteq \Sigma_j.$$

So overall, equation (11.16) has an upper bound of

$$\begin{aligned}
&\frac{1}{(\log j)^2}\bigg(\int_{j^{-4}}^{j^{-2}} \mathcal{H}^2\Big(M \cap \mathbb{B}\Big(t^{-\frac{1}{2}}\Big)\Big)\ dt + \int_0^{j^{-4}} \mathcal{H}^2(\Sigma_j)\ dt \bigg) \\
&\qquad = \frac{1}{(\log j)^2}\bigg(\int_{j^{-4}}^{j^{-2}} \mathcal{H}^2\Big(M \cap \mathbb{B}\Big(t^{-\frac{1}{2}}\Big)\Big)\ dt + \frac{1}{j^4}\mathcal{H}^2(\Sigma_j) \bigg).
\end{aligned}$$

We know that $\mathcal{H}^2(M \cap \mathbb{B}(r)) \leq cr^2$. So $\mathcal{H}^2(M \cap \mathbb{B}(t^{-\frac{1}{2}})) \leq \frac{c}{t}$. Moreover, since $\mathcal{H}^2(M \cap \mathbb{B}(j^2)) \leq Cj^4$, it follows that $\mathcal{H}^2(\Sigma_j) \leq \mathcal{H}^2(M \cap \mathbb{B}(j^2)) \leq Cj^4$.

Therefore,

$$\begin{aligned}
&\frac{1}{(\log j)^2}\bigg(\int_{j^{-4}}^{j^{-2}} \mathcal{H}^2\Big(M \cap \mathbb{B}\Big(t^{-\frac{1}{2}}\Big)\Big)\ dt + \frac{1}{j^4}\mathcal{H}^2(\Sigma_j) \bigg) \\
&\qquad \leq \frac{1}{(\log j)^2}\bigg(\int_{j^{-4}}^{j^{-2}} \frac{c}{t}\ dt + \frac{1}{j^4} Cj^4 \bigg) \\
&\qquad = \frac{1}{(\log j)^2}\Big(2c\log(j) + C \Big) \\
&\qquad = \frac{2c}{\log(j)} + \frac{C}{(\log j)^2}.
\end{aligned}$$

Thus, $\int_M |D\zeta_j|^2\ d\mathcal{H}^2 < \frac{A}{\log j}$ for some constant $A = 2c + \frac{C}{\log j}$. By inequality (11.15), as j approaches ∞, $\int_{M \cap \mathbb{B}(j)} |B|^2 = 0$ and hence $|B|^2 = 0$ (since the integrand is non-negative).

It remains to show that M is a plane. We know that $|B|^2 = \sum_{i,j=1}^2 |B(\tau_i, \tau_j)|^2$, where τ_1 and τ_2 form an orthonormal basis of the tangent space T_yM. By the previous step, it follows that $|B(\tau_i, \tau_j)|^2 = 0$ for all $i, j \in \{1, 2\}$. Moreover,

$$0 = |B(\tau_i, \tau_j)|^2 = (\tau_i \cdot D_{\tau_j}\nu)^2,$$

so $\tau_i \cdot D_{\tau_j}\nu = 0$.

In particular $D_{\tau_i}\nu = 0$ for all τ_i. So ν is constant. In particular, suppose $\gamma : (-1, 1) \to \mathbb{R}^3$ is a curve on M such that $\gamma(0) = y$ and $\dot{\gamma}(0) = \tau_i$. Now consider $(\gamma(t) - \gamma(0)) \cdot \nu$, which has derivative $\dot{\gamma}(t) \cdot \nu$. Since $\dot{\gamma}(t)$ is perpendicular to ν, it follows that $\dot{\gamma}(t) \cdot \nu = 0$, so $(\gamma(t) - \gamma(0)) \cdot \nu =$constant. This is the equation of a plane. $\square$

12 Conclusion

In this document we explored the basics of geometric measure theory leading to a proof of Bernstein's problem. Although it may seem trivial that Bernstein's problem holds for a minimal 2-dimensional graph embedded in $\mathbb{R}^3$, what is perhaps more surprising is that it fails to hold for minimal graphs of 7-dimensions and above (of codimension one). The Bernstein problem was finally solved in its entirety in 1969, after being posed in 1915. This marks over 50 years of work on the topic in which many new ideas were developed.

The proof for graphs of dimensions three to seven and the subsequent counterexample for dimensions eight and above rely on a very different method of proof to the one given above. In particular, the proof for higher dimensions relies on a knowledge of minimal cones, which were not addressed. More precisely, Bernstein's problem for a minimal graph embedded in $\mathbb{R}^{n+1}$ is equivalent to showing that there are no minimal cones in $\mathbb{R}^n$. More details are provided in *Minimal Surfaces and Functions of Bounded Variation* [Giu84].

Following on from the proof of the Bernstein problem, many analogous problems were posed. One of these is the spherical Bernstein problem posed by Chern, which asks whether the equators in $\mathbb{S}^{n+1}$ are the only smooth embedded minimal hypersurfaces which are topological spheres. Hsiang has proved many cases of this theorem [Hsi83a, Hsi83b, HS86], but the complete solution remains an open problem. There are also many formulations of Bernstein's problem for minimal graphs of higher codimension.

References

[Bog07] Vladimir Bogachev. *Measure Theory*. Springer, Berlin, 2007.

[Coh13] Donald Cohn. *Measure Theory*. Birkhäuser, Boston, 2 edition, 2013.

[dC76] Manfredo P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, New Jersey, 1976.

[EG15] Lawrence Evans and Ronald Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Florida, 2015.

[Giu84] Enrico Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Springer Science + Business Media, New York, 1984.

[HS86] Wu-Yi Hsiang and Ivan Sterling. Minimal cones and the spherical Bernstein problem, III. *Inventiones Mathematicae*, 85(2):223–247, 1986.

[Hsi83a] Wu-Yi Hsiang. Minimal cones and the spherical Bernstein problem, I. *Annals of Mathematics*, 118(1):61–73, 1983.

[Hsi83b] Wu-Yi Hsiang. Minimal cones and the spherical Bernstein problem, II. *Inventiones Mathematicae*, 74(3):351–369, 1983.

[Jon01] Frank Jones. *Lebesgue Integration on Euclidean Space*. Jones and Bartlett Publishers, Massachusetts, 2001.

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

[Mor09] Frank Morgan. *Geometric Measure Theory: A Beginner's Guide*. Academic Press, London, 4 edition, 2009.

[Sim14] Leon Simon. *Introduction to Geometric Measure Theory*. Stanford University, 2014.

[Spi99] Michael Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 2. Publish or Perish, Houston, 3 edition, 1999.

[Tra94] Tim Traynor. Change of variables for Hausdorff measure (from the beginning). 1994.