Course Plan SF3626, Analysis for PhD students.

J. Andersson

December 1, 2016

Contents

1 Lecture 1. Introduction **2**
1.1 What goes wrong with the Riemann integral. 5
1.2 Exercises: 9

2 Lecture 2: The Lebegue Measure - I. **9**

3 Lecture 3: Measure Theory II (my lecture). **18**
3.1 Reading 26

4 Lecture 4. The Lebegue integral. **27**

5 Lecture 5. Integration on product spaces and Fubini. **27**

6 Lecture 6. The L^p spaces. **27**

7 Lecture 7. Duality and functionals. **28**

8 Lecture 8: Weak compactness. **29**
8.1 Exercises: 37

9 Lecture 9. Differentiation. Distributional derivatives. $W^{k,p}$ spaces. **38**
9.1 Exercises: 47

10 Lecture 10. Differentiation and the Fundamental Theorem of Calculus, part 1. **50**
10.1 Exercices 58

11 Lecture 11. Differentiation and the Fundamental Theorem of Calculus, part 2. **58**

12 Lecture 12. Classical derivatives and Sobolev functions. **66**
12.1 Exercises: 80

13 Lecture 13. Solution to the calculus of variations problem. **80**
13.1 A more general variational problem. 82

14 Lecture 14. Difference quotients and the Euler-Lagrange equations. 86
14.1 Exercises: 97

15 Lecture 15. Harmonic measure and Riesz representation theorem. 97

1 Lecture 1. Introduction

You where given the following problem

The Problem: *Given a subset $\mathcal{D} \subset \mathbb{R}^n$ and a function $f : \partial\mathcal{D} \mapsto \mathbb{R}$ does it exist a function $u(\mathbf{x})$ such that $u(\mathbf{x}) = f(\mathbf{x})$ on $\partial\mathcal{D}$ and*

$$E(u(\mathbf{x})) \leq E(v(\mathbf{x}))\text{for all functions } v(\mathbf{x}) \text{ s.t. } v(\mathbf{x}) = f(\mathbf{x}) \text{ on } \partial\mathcal{D}?$$

Where

$$E(u(\mathbf{x})) = \int_{\mathcal{D}} |\nabla u(\mathbf{x})|^2 d\mathbf{x}, \tag{1}$$

My intention was that you shouldn't be able to solve it, so let us assume that you where not. How does one even begin to approach a, presumably, new problem? There are many possible ways. The most common way to attack a new problem is to make some simplifying assumptions and try to solve the simpler problem. But in this case we will use another strategy. We will go back to undergraduate analysis and see how we solve optimization problems there and then see if we can generalize the approach from the known case to the more difficult at hand.

The simplest relevant theorem form undergraduate analysis is the following.

Theorem 1.1. *Let $f(x)$ be lower semicontinuous*[1] *and bounded from below on the interval $[a, b]$ to $\mathbb{R}$. Then there exists an $x_0 \in [a, b]$ such that $f(x_0) \leq f(x)$ for all $x \in [a, b]$.*

Proof: We will split the proof intoseveral easy steps.

Step 1: *The greatest lower bound m of the range of f exists: $m = \inf_{x\in[a,b]}(f(x))$ exists.*

The range of f, let us denote it by $\mathcal{R}_f$, is a set of real numbers. Furthermore $\mathcal{R}_f$ is bounded form below by assumption. The completeness property of the real numbers states that every set of real numbers that is bounded below has a greatest lower bound. It follows that $\mathcal{R}_f$ has aleast lower bound that we may call m.

Step 2: *There exists a sequence of points $x_j \in [a, b]$ such that $f(x_j) \to m$.*

Since m is the greatest lower bound of $\mathcal{R}_f$ it follows that $m + \frac{1}{j} > m$ is not a lower bound of $\mathcal{R}_f$. We may conclude that there exists[2] an x_j such

[1] Remember that $f(x)$ is lower semi-continuous if $f(y) \leq \lim_{x\to y} f(x)$ for all y in the domain of f.

[2] Let me just remark here that we are using the axiom of choice in this step. As is common in mainstream mathematics today we will not bother to keep track of the usage of the axiom of choice in this course. But as part of good mathematical culture one should, at least occasionally, pay some attention of *the choice* and foundational issues.

that $f(x_j) < m + \frac{1}{j}$. Since m is a lower bound for $\mathcal{R}_f$ we can conclude that $m \leq f(x_j)$. Therefore $m \leq f(x_j) < m + \frac{1}{j}$, it follows that $f(x_j) \to m$.

Step 3: *There exists a subsequence $x_{j_k} \to x_0$ for some $x_0 \in [a, b]$.*

From the Bolzano-Weierstrass Theorem: *every bounded sequence in $\mathbb{R}$ (such as $x_j \in [a, b]$) has a convergent subsequence*; we may conclude that x_j has a convergent subsequence. Since $[a, b]$ is closed it follows that the limit $x_0 \in [a, b]$.

Step 4: *Conclusion of the Theorem.*

Since $f(x)$ is lower semi-continuous it follows that $f(x_0) \leq \lim_{k\to\infty} f(x_{j_k}) = m$. Since m is a lower bound for $f(x)$ it follows that $f(x_0) \geq m$. We may conclude that $f(x_0) = m = \inf_{x\in[a,b]} f(x)$. □

Now we have one theorem that assures the existence of a minimizer in one simple case. In order to repeat the same proof for the problem we have at hand we need to carefully analyze what properties we used in the simple theorem.

In Step 1 we only used the completeness of $\mathbb{R}$ and that f was bounded from below. Since the $E(u) \geq 0$ this part of the argument should go through without changes. But there is one complication! The range of E depends on the domain of definition of E, and we have not properly defined the domain of definition of E. Is E defined on the set of continuously differentiable functions, on the set of functions whose gradient is Riemann integrable, or what?

Question 1: *What is the natural domain of definition of E? How should we interpret the integral in the definition of E?*

We have a lot of freedom to choose a domain of definition of E, lets call the domain of definition for $\mathcal{K}_E$. But almost all of our choices will be wrong! If we choose the domain of definition $\mathcal{K}_E$ to small we might not have a minimizer; consider for instance trying to minimize continuous functions over $\mathbb{Q}$. In any case the domain of E must be be non-empty. On the other hand, if we choose it to large we might have several minimizers; or they might not behave the way we want.

If we, for the moment, assume that we have agreed on a reasonable and non empty domain of definition $\mathcal{K}_E$. Then the argument in step 1 should work and we can turn our attention to step 2. It is easy to see that step 2 should work without changes.

Step 3 is much more subtle. We are actually using two things in step 3. First we use the Bolzano-Weierstrass Theorem and then that $\mathcal{K}_E$ is closed.

Need some kind of Bolzano-Weierstrass Theorem for the domain of E.

To derive a Bolzano-Weierstrass type theorem for $\mathcal{K}_E$ we would want every bounded sequence u^j to have a convergent subsequence.[3] This leads to the next problem, how do we assure that a sequence u^j such that $E(u^j) \to m = \inf_{u\in\mathcal{K}_E} E(u)$ is bounded, and in what sense is it bounded? To choose the domain $\mathcal{K}_E$ bounded is of lesser importance. This since if $E(u^j) \to m$ then there exists a $J \in \mathbb{N}$ such that

$$\int_{\mathcal{D}} |\nabla u^j(\mathbf{x})|^2 d\mathbf{x} < 2m \text{ for } j > J. \tag{2}$$

We thus get that the sequence u^j is bounded, in the sense of (2), irregardless of whether $\mathcal{K}_E$ is bounded. We would need some kind of Bolzano-Weierstrass

[3] Later we will see that this is to strong a requirement and therefore we will have to introduce the concept of weak convergence.

Theorem that states that if a sequence u^j is bounded in the sense of (2) then there is a convergent subsequence $u^{j_k} \to u^0$. We would want to derive the following: *If u^j is a sequence such that $\int_{\mathcal{D}} |\nabla u^j|^2 d\mathbf{x} < M$ then there exists a subsequence u^{j_k} that converges to some u^0*. Notice that the concept of boundedness, $\int_{\mathcal{D}} |\nabla u^j|^2 d\mathbf{x} < M$, is given to us (or imposed on us) by the mathematics of the problem. But can we derive such a Bolzano-Weierstrass Theorem?

The natural condition for boundedness in the set $\mathcal{K}_E$ is $\int_{\mathcal{D}} |\nabla u|^2 d\mathbf{x}$.

The other property we used in step 3 was that $\mathcal{K}_E$ is a closed set. Since we are only interested in functions u such that $u(\mathbf{x}) = f(\mathbf{x})$ on $\partial\mathcal{D}$ we would need to show that if $u^j \to u^0$ and $u^j = f$ on $\partial\mathcal{D}$ then $u^0 = f$ on $\partial\mathcal{D}$. This is not entirely straightforward. We still have not really decided in what sense it is natural to consider the convergence $u^j \to u^0$ but given the appearance of $\int_{\mathcal{D}} |\nabla u^j|^2 d\mathbf{x}$ i the formulation of the tentative Bolzano-Weierstrass Theorem it is reasonable to say that $u^j \to u^0$ if $\int_{\mathcal{D}} |\nabla(u^{j_k} - u^0)|^2 d\mathbf{x} \to 0$.[4] But if the convergence only depends on the integral of $\nabla(u^j - u^0)$ it might not even be possible to talk about values of u^0 on the boundary. We will explain the problems relating to this in more detail later.

Domain of E should be closed.

The final step in the proof of Theorem 1.1 uses that f is lower semicontinuous. To emulate the proof of this step we need to show that E is lower semicontinuous with respect to the convergence $u^{j_k} \to u^0$. That is, when we derive our Bolzano-Weierstrass Theorem we need to fix a meaning for $u^{j_k} \to u^0$, say that $\int_{\mathcal{D}} |\nabla(u^{j_k} - u^0)|^2 d\mathbf{x} \to 0$, and then show that E is lower semicontinuous with respect to convergence in this sense.

In order to summarize we would need the following things in order to prove existence of minimizers of $E(u)$.

1. We need to choose the domain of definition, $\mathcal{K}_E$, of the functional $E(u)$ in such way that,
   (a) $\mathcal{K}_E$ should be, or contained in a space that is, complete with respect to some norm. The mathematics indicate that the norm should be $\|u\|^2 = \int_{\mathcal{D}} |\nabla u|^2 d\mathbf{x}$.
   (b) We need to show some sort of Bolzano-Weierstrass Theorem: If $\|u^j\| \leq M$ then there exists a function u^0 and a subsequence $u^{j_k} \to u^0$.
   (c) We need to make sure that $\mathcal{K}_E$ is non-empty.
   (d) We need to make sure that $\mathcal{K}_E$ is closed under convergence. In particular, we need to make sense of the boundary condition $u = f$ on $\partial\mathcal{D}$ and show that that is preserved under limits.
2. We need to figure out what it means for a function u to have square integrable gradient. Clearly if $u \in C^1(\overline{\mathcal{D}})$ then $|\nabla u|^2$ is integrable. But we will see that assuming that $u \in C^1(\overline{\mathcal{D}})$ is to restrictive for the theory we want to develop.
3. We need to show that the function $E(u) : \mathcal{K}_E \mapsto \mathbb{R}$ is lower semicontinuous with respect to the convergence $u^{j_k} \to u^0$.
4. We need to resolve all other problems, and there will be problems, that appear during our attempt to prove the above points.

[4] As we will see this is not what we will use later - but what we will use is even worse so the heuristics that follows is still valid.

1.1 What goes wrong with the Riemann integral.

Let us end the first lecture with a quick review of the Riemann integral. We begin by reminding ourselves of some definitions.

Definition 1.1. *We say that a function* $f : [a,b] \mapsto \mathbb{R}$ *is Riemann integrable*[5] *if there for any* $\epsilon > 0$ *exists a partition of* $[a,b]$*, that is a set of points* $a = x_0 < x_1 < x_2 < ... < x_n = b$*, such that*

$$\sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}) < \epsilon \tag{3}$$

where $M_k = \sup_{x\in(x_{k-1},x_k)} f(x)$ *and* $m_k = \inf_{x\in(x_{k-1},x_k)} f(x)$.

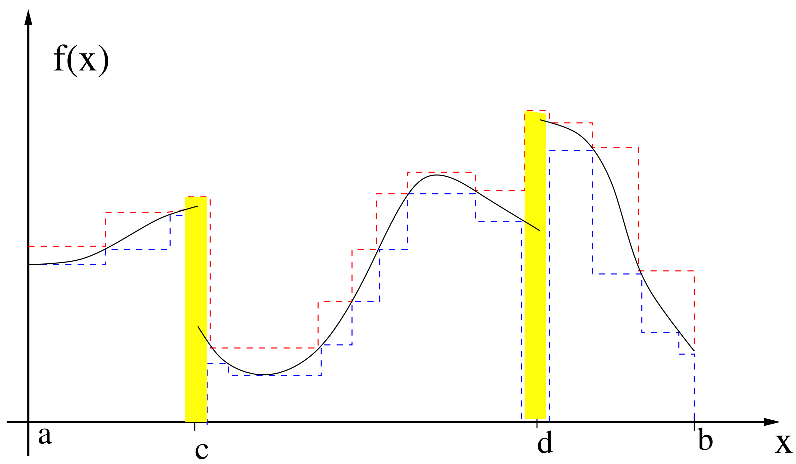
If $f(x) : [a,b] \mapsto \mathbb{R}$ *is Riemann integrable we define*

$$\int_a^b f(x)dx = \sup_P \sum_{k=1}^{n} m_k(x_k - x_{k-1}),$$

where the supremum is taken over all partitions $P:\ a = x_0 < x_1 < ... < x_n = b$ *for all* n.

This definition agrees well with our intuition with assigning a (signed) area to the region under a graph. But the definition is useless unless we actually can integrate a large class of functions. Fortunately we know from undergraduate analysis that if $f \in C([a,b])$ then u is Riemann integrable. The argument goes something like this: since f is continuous on a compact interval u is uniformly continuous and we may therefore find, for each $\epsilon > 0$, a $\delta > 0$ such that $\text{osc}_{x\in[x_k-1,x_k]} u \leq \frac{\epsilon}{b-a}$ for each $x_{k-1} < x_k$ such that $|x_k - x_{k-1}| < \delta$; we may therefore part $[a,b]$ into $n \approx \frac{2(b-a``}{\delta}$ pieces $a = x_0 < x_1 < ... < x_n = b$ such that $\text{osc}_{x\in(x_k-1,x_k)} f \leq \frac{\epsilon}{b-a}$, then (3) holds.[6]

We also know from undergraduate analysis that if $f(x)$ is bounded and piecewise continuous, that is discontinuous at finite set of points, then f is Riemann integrable.



[5] At times one calls these functions Darboux integrable. But the Riemann and Darboux concepts of integrability are equivalent.

[6] By definition the oscillation of f over $[x_{k-1}, x_k]$, in symbols $\text{osc}_{x\in(x_k-1,x_k)} f(x)$, is defined to be $\sup_{x\in(x_{k-1},x_k)} f(x) - \inf_{x\in(x_{k-1},x_k)} f(x) = M_k - m_k$.

Figure: *If $f(x)$ is discontinuous at finitely many points, say at two points c and d as in the graph, then we may capture the discontinuities in two samll rectangles (yellow in the graph) each having area less than $\epsilon/5$. The rest of the graph consits of three continuous pieces. The area under each which may be approximated from above and below by the area under step functions so that the difference of the area under the red and blue stepfunctions is less than $\epsilon/5$, this since f is uniformly continuous at each of the three pieces in the complement of the yellow rectangles. We may therefore approximate the area under the graph within an ϵ error for every $\epsilon > 0$. It follows that f is Riemann integrable.*

These undergraduate theorems gives large classes of Riemann integrable functions. But can we describe exactly when a function is Riemann integrable? A good guess would be that f should not be discontinuous in a to large set. The right way to measure largeness of the set of discontinuities is by the concept of zero measure.

Definition 1.2. *We say that a set $A \subset \mathbb{R}$ has zero measure if, for any $\epsilon > 0$, there exists a countable set of interval's (a_i, b_i) such that*

$$A \subset \cup_{i=1}^{\infty}(a_i, b_i)$$

and

$$\sum_{i=1}^{\infty}(b_i - a_i) < \epsilon.$$

Theorem 1.2. *A bounded function $f : [a, b] \mapsto \mathbb{R}$ is Riemann integrable if and only the set*

$$D = \{x \in [a, b];\ f \text{ is discontinuous at } x\}.$$

Proof: Assume that D has measure zero. We aim to prove that f is Riemann integrable. Let us write $D = \cup_{k=1}^{\infty} D_k$ where

$$D_k = \{x \in [a, b];\ \lim_{r \to 0^+} \mathrm{osc}_{(x-r,x+r)} f \geq 1/k\}.$$

It is enough to show that each D_k has measure zero. Indeed, if each D_k has measure zero then we may find a countable covering (a_i^k, b_i^k) such that

$$\sum_{i=1}^{\infty}(b_i^k - a_i^k) < \frac{\epsilon}{2^k}.$$

Then the countable covering consisting of all interval's (a_i^k, b_i^k) will satisfy

$$\sum_{k=1}^{\infty}\sum_{i=1}^{\infty}(b_i^k - a_i^k) < \sum_{k=1}^{\infty}\frac{\epsilon}{2^k} = \epsilon.$$

We now turn to finding a cover of D_k that has length less than ϵ. Since f is Riemann integrable there is a partition $a = x_0 < x_1 < ... < x_n = b$ such that

$$\sum_{j=1}^{n}(M_j - m_j)(x_j - x_{j-1}) < \frac{\epsilon}{k}. \tag{4}$$

Notice that if $D_k \cap (x_{j-1}, x_j) \neq \emptyset$ then $M_j - m_j > \frac{1}{k}$. Lets denote the indexes j such that $D_k \cap (x_{j-1}, x_j) \neq \emptyset$ by J. From (4) and the last observation we may conclude that

$$\sum_{j \in J} \frac{1}{k}(x_j - x_{j-1}) \leq \sum_{j \in J}(M_j - m_j)(x_j - x_{j-1}) \leq \sum_{j=1}^{n}(M_j - m_j)(x_j - x_{j-1}) < \frac{\epsilon}{k}.$$

It follows that $\sum_{j \in J}(x_j - x_{j-1}) < \epsilon$ and since $\cup_{j \in J}(x_j - x_{j-1})$ covers D_k, except possibly finitely many points x_i, we may conclude that D_k has measure zero.

Now assume that D has measure zero, this implies that D_k has measure zero for each k. We want to show that f is Riemann integrable. To that end we fix an $\epsilon > 0$ and choose k large enough so that $\frac{1}{k} < \frac{\epsilon}{4(b-a)}$.

Since D_k has zero measure we may find a covering, (a_i, b_i), of D_k so that $\sum(b_i - a_i) < \frac{\epsilon}{4M}$ where $M = \sup_{x \in [a,b]} f(x)$. Furthermore, for every point $x \notin D_k$ there is an $r_x > 0$ such that $\mathrm{osc}_{(x - r_x, x + r_x)} f < \frac{1}{k} < \frac{\epsilon}{4(b-a)}$.

Now D_k is closed and bounded and we may therefore find a finite subcover of (a_i, b_i) of D_k: upon relabeling the a_i and b_i we may conclude that $D_k \subset \cup_{i=1}^{I}(a_i, b_i)$ for some finite I. Also $[a,b] \setminus \cup_{i=1}^{I}(a_i, b_i)$ is closed and bounded and therefore compact so we may cover $[a,b] \setminus \cup_{i=1}^{I}(a_i, b_i)$ by a finite subset of the intervals $(x - r_x, x + r_x)$. That is there exists a finite set J such that

$$[a,b] \subset \left(\cup_{x \in J}(x - r_x, x + r_x)\right) \bigcup \left(\cup_{i=1}^{I}(a_i, b_i)\right).$$

We may therefore use all the numbers $x \pm r_x$ for $x \in J$ and a_i and b_i for $i \leq I$ to form a partition, $a = x_0 < x_1 < ... < x_n = b$, of $[a,b]$. The union of the intervals (x_{j-1}, x_j) that intersect D_k will have length less than $\cup_{i=1}^{I}(a_i, b_i)$, that is length less than $\frac{\epsilon}{4M}$ and in these intervals the oscillation $M_j - m_j \leq 2M$. The total length of the intervals (x_{j-1}, x_j) that does not intersect D_k will be less than $b - a$ and the oscillation in these intervals is $M_j - m_j < \frac{1}{k} < \frac{\epsilon}{4(b-a)}$. Therefore

$$\sum_{j=1}^{n}(M_j - m_j)(x_j - x_{j-1}) \leq 2M\frac{\epsilon}{4M} + \frac{\epsilon}{4(b-a)}(b-a) < \epsilon.$$

It follows that f is Riemann integrable. □

This theorem classifies which functions are Riemann integrable. But, as we all know, not all functions are Riemann integrable.

Example: *Let $f : [0,1] \mapsto \mathbb{R}$ be defined*

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0,1] \\ 0 & \text{if } x \notin \mathbb{Q} \cap [0,1]. \end{cases} \tag{5}$$

Then the oscillation of f will be equal to 1 in every open interval since $\mathbb{Q}$ is dense in $\mathbb{R}$. In particular, for any partition $0 = x_0 < x_1 < ... < x_n = 1$ we will have

$$\sum_{j=1}^{n}(M_j - m_j)(x_j - x_{j-1}) = 1.$$

Therefore f is not Riemann integrable.

Of course, that there is an, admittedly weird, function that is not Riemann integrable does not discredit the Riemann integral. Even the more powerful Lebesgue integral (that we will develop over the next few lectures) is not powerful enough to integrate all functions. What really is lacking in the Riemann integral is that it does not work well with convergence.

Example: *Let $x_1, x_2, ...$ be some enumeration of $\mathbb{Q} \cap [0,1]$ and define*

$$f_k(x) = \begin{cases} 1 & \text{if } x = x_j \text{ for some } j \leq k \\ 0 & \text{else.} \end{cases}$$

Then $f_k(x)$ is an increasing sequence of functions, $f_k(x) \leq f_{k+1}(x)$ for all k, and converges point-wise to f where f is defined in (5). Also $\int_0^1 f_k(x)dx = 0$ for all k. This is easy to see by choosing the partition with points 0, $x_j \pm \frac{\epsilon}{k}$ for $j = 1, 2, ..., k$ and 1.

This shows that there is a sequence of increasing functions that all have integral zero that converges to a function $f(x)$ that is not even integrable.

Clearly the above example shows that the Riemann integral has bad properties when it comes to convergence of functions. This is a very severe shortcoming of the Riemann integral.

Some heuristics for the remedy for the integral. Interestingly the function $f(x)$, defined as in (5), can be more or less integrated by hand, at least heuristically. The idea is that if $x_1, x_2, ...$ is an enumeration of $\mathbb{Q} \cap [0,1]$. Then the intervals

$$\left(x_1 - \frac{\epsilon}{2 \cdot 2^1}, x_1 + \frac{\epsilon}{2 \cdot 2^1}\right), \left(x_2 - \frac{\epsilon}{2 \cdot 2^2}, x_2 + \frac{\epsilon}{2 \cdot 2^2}\right), \left(x_3 - \frac{\epsilon}{2 \cdot 2^3}, x_3 + \frac{\epsilon}{2 \cdot 2^3}\right), ...$$

will cover $\mathbb{Q} \cap [0,1]$ and the total length of the intervals will be

$$\sum_{j=1}^{\infty} \frac{\epsilon}{2^j} = \epsilon.$$

Since $\epsilon > 0$ is arbitrary, the set where f equals to 1 has measure zero. We should be rather confident to ascribe the area under the graph of f to be zero as well.

This is not really a stringent argument since we have not defined area, but it should still be intuitively convincing. What we gain in the above reasoning that we do not have in the Riemann integral case is that we where allowed to cover the set where $f(x) = 1$ by a countable set of intervals. We seem to be getting something extra by first evaluating the "length" of the set $\{x \in [0,1];\ f(x) = 1\}$ by using a countable covering.

We may (and will) extend this idea for general functions $f(x)$. If $f(x) : [0,1] \mapsto \mathbb{R}$ and $|f(x)| \leq M$ then we may approximate $f(x)$ by the function

$$s(x) = \sum_{k=-N}^{N} \epsilon k \chi_{\{x \in [0,1]; (k-1)\epsilon < f(x) \leq k\epsilon\}}(x), \tag{6}$$

where N is chosen so that $M \leq \epsilon N$ and $\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$ is the characteristic function of the set A. Clearly

$$|s(x) - f(x)| \leq \epsilon \tag{7}$$

and the area under $s(x)$ should be

$$\sum_{k=-N}^{N} \epsilon k m\big(\{x\in[0,1];(k-1)\epsilon < f(x)\le k\epsilon\}\big),$$

where $m(A)$ somehow measures the length of the set A. Also, at least intuitively by (7),the area under $f(x)$ should differ from the area under $s(x)$ by at most ϵ. So far one might not think that we have gained anything by approximating $f(x)$ by $s(x)$ defined as in (6) instead of approximating $f(x)$ by a step function, it might even seem to be more complicated since we will have to figure out how to measure the length of a general set $\{x\in[0,1];(k-1)\epsilon < f(x)\le k\epsilon\}$. But there is a real gain, and that is, that when we measure the length of the sets $\{x\in[0,1];(k-1)\epsilon < f(x)\le k\epsilon\}$ we do that independently of the integral and we may use countable many intervals whereas we only used finitely many intervals in the partition when evaluation the Rieamann integral. It will require some work to define the length, or measure, of a general set, but it will lead to a definition of the integral that is stronger and more versatile than the Riemann integral.

1.2 Exercises:

1. Let
$$f(x)=\begin{cases}\frac{1}{q} & \text{if } x=\frac{p}{q}\in\mathbb{Q}\\ 0 & \text{if } x\in\mathbb{R}\setminus\mathbb{Q},\end{cases}$$
where, obviously, we require that p and $q>0$ does not have any common factors. Prove that f is Riemann integrable and that $\int_0^1 f(x)dx=0$.

2. In the proof of Theorem 1.2 we claim that D_k is a closed set: prove this! You may use the following steps.
   (a) Argue that is is enough to show that if $x_j\in D_k$ and $x_j\to x_0$ then $x_0\in D_k$.
   (b) Assume, in oder to get a contradiction, that there exist $x_j\to x_0\notin D_k$.
   (c) Show that, since $x_0\notin D_k$, it follows that there exists an $\epsilon>0$ and an $r_\epsilon>0$ such that $\text{osc}_{[x_0-r,x_0+r]}f<\frac{1}{k}-\epsilon$ for all $r<r_\epsilon$.
   (d) show that if $|x_j-x_0|\le r_\epsilon/2$ then $x_j\notin D_k$ and derive a contradiction. HINT: *Can you prove that the if $(a,b)\subset(c,d)$ then $osc_{(a,b)}f\le osc_{(c,d)}f$?*

2 Lecture 2: The Lebegue Measure - I.

At the end of lecture 1 we noticed that there are bounded and increasing sequences of Riemann integrable functions $f_j(x)\nearrow f(x)$ (point-wise) without $f(x)$ being Riemann integrable. That sequences that converge in rather good ways (increasing point-wise) whose limits are not integrable is a serious shortcoming of the Riemann's definition of the integral.

We also realized that the integral might be more versatile if we use the approximation

$$\sum_{j=-N}^{N} \epsilon j m\left(\{x;\ j\epsilon < f(x) \le (j+1)\epsilon\}\right) \le$$

$$\le \int_{\alpha}^{\beta} f(x)dx \le \sum_{j=-N}^{N} \epsilon(j+1) m\left(\{x;\ j\epsilon < f(x) \le (j+1)\epsilon\}\right),$$

where $m(A)$ is some measure of the length of the set A.

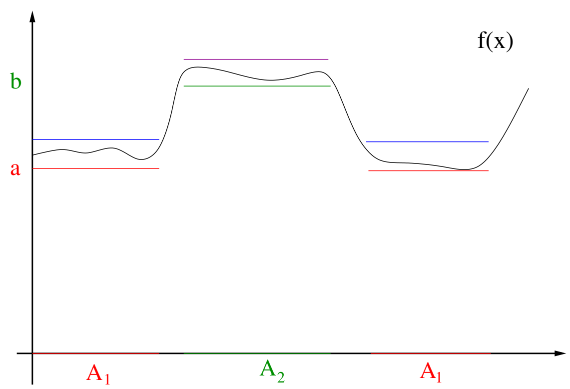


Figure: *The general idea with the Lebesgue integral is to approximate the area under the graph of f by functions that are constant where f is almost constant. In the figue we may approximate the area under the graph in on the red part of the $x-$axis by a function that has constant value a for all x in the red part of the x-axis denoted A_1, and by a constant function with value b on the green part A_2. In order to calculate the area we must find a good way to measure the length of the sets A_1 and A_2.*

The main problem is how are we supposed to define a reasonable measure of length? For simple sets, as intervals it is quite clear that the length of the interval (a, b) should be $b - a$. But the sets $\{x;\ j\epsilon < f(x) \le (j+1)\epsilon\}$ may be very complicated and have huge oscillations; imagine the set

$$\left\{x \in (0,1);\ 10^{-3} < \sin(e^{-1/x}) \le 2 \cdot 10^{-3}\right\} \tag{8}$$

which will certainly have infinitely many components - and we might come up with even worse examples[7] of sets whose length we would like to measure.

There are certain things that we would want our measure (of length) to satisfy. In particular, if we denote by $m(A)$ the measure of the set $A \subset \mathbb{R}$ then the following should hold

1. Any interval has its "natural length": $m((a,b)) = b - a$ for any interval $(a,b) \subset \mathbb{R}$.

[7]The example in (8) is not bad at all from the point of view of integration. The function $\sin(e^{-1/x})$ is Riemann integrable on $(0,1)$.

2. The measure should be countable additive: $m(\cup_j A_j) = \sum_{j=1}^{\infty} m(A_j)$ for any countable disjoint collection of sets A_j.

Later we will see that it is not possible to define any measure m, defined on all subsets of $\mathbb{R}$, in a way that satisfies these criteria.

A rather interesting fact is that open sets have length that is intuitively well defined because of the following lemma.

Lemma 2.1. *Let $U \subset \mathbb{R}$ be an open set. Then $U = \bigcup_{j=1}(a_j, b_j)$ where (a_j, b_j) are countable (or finite) and disjoint set of open intervals.*

Proof: Each connected component U_i of U is open and therefore contains a rational point $q_i \in \mathbb{Q}$, fix one such point $q_i \in \mathbb{Q}$ for each connected component of U. We may therefore define an injection from the connected components of U into a subset of $\mathbb{Q}$ that takes the connected component U_i to q_i. Therefore there is a bijection between the connected components of U and a subset of the countable set $\mathbb{Q}$. Therefore there are at most countable many connected subsets of U. Clearly each connected and open set in $\mathbb{R}$ is an interval. $\square$

Since we would want the length of an interval, (a, b), to be $b - a$ it would be natural to define the length of an open set $U = \bigcup_{j=1}(a_j, b_j) \subset \mathbb{R}$ to be $\sum_j(b_j - a_j)$. Since there are at most countable many intervals the sum is well defined, though it might be diverge to ∞. Also since all terms $b_j - a_j > 0$ the summation is independent of the order of summation. We can therefore ascribe a length of an open set in a natural way. We will use this to define an outer measure (of length).

Definition 2.1. *We define the outer Lebesgue measure m^* on subsets $A \subset \mathbb{R}$ to $[0, \infty)$ according to*

$$m^*(A) = \inf \sum_{j=0}^{\infty} (b_j - a_j), \tag{9}$$

where the infimum is taken over all countable (or finite[8]) unions of open intervals (a_j, b_j) such that $A \subset \bigcup_{j=1}^{\infty}(a_j, b_j)$.

This definition makes perfect intuitive sense. However it is not absolutely clear that $m^*(a, b) = b - a$ with this definition. We need the following lemma.

Lemma 2.2. *The Lebesgue outer measure satisfies $m^*([a, b]) = m^*((a, b)) = m^*([a, b)) = m^*((a, b]) = b - a$.*

Proof: We begin by showing that $m^*([a, b]) = b - a$. The proof has several steps.

Step 1: *It is enough to consider finite coverings.*

Since $[a, b]$ is compact every open cover $U = \cup_{j=1}^{\infty}(a_j, b_j)$ reduces to a finite sub-cover $\tilde{U} = \cup_{j=1}^{N}(a_j, b_j)$. Clearly, since $b_j - a_j > 0$, $\sum_{j=1}^{N}(b_j - a_j) < \sum_{j=1}^{\infty}(b_j - a_j)$. It is therefore enough to consider finite coverings.

Step 2: *It is enough to consider coverings consisting of one interval* (a_1, b_1).

Assume that we have a covering $\tilde{U} = \cup_{j=1}^{N}(a_j, b_j)$. Then, since $a \in U$ there exists one interval, say (a_1, b_1), so that $a \in (a_1, b_1)$. If $[a, b] \subset (a_1, b_1)$ we already

[8] In case the collection of intervals is finite then, naturally, the summation in (9) will be over the index set of the intervals and not to infinity.

have one interval that covers $[a,b]$ so lets assume that $[a,b] \not\subset (a_1,b_1)$. This means that $a < b_1 < b$ and therefore $b \in [a,b]$. There must be another interval, say (a_2,b_2), such that $B_1 \in (a_2,b_2)$. But then $(a_1,b_2),(a_3,b_3),...,(a_N,b_N)$ also cover of $[a,b]$, with only $N-1$ intervals, and furthermore

$$(b_2 - a_1) + (b_3 - a_3) + ... + (b_N - a_N) < \sum_{j=1}^{N}(b_j - a_j).$$

We have therefore shown that for any cover U, containing at least two intervals, there is another cover with one less interval and smaller sum. We may conclude that the smallest cover can be achieved with one interval.

Step 3: $m^*([a,b]) = b - a$.

Since, for any $\epsilon > 0$, $(a - \epsilon/2, b + \epsilon/2)$ is a cover of $[a,b]$ it follows that $m^*([a,b]) \leq b - a + \epsilon$ and therefore

$$m^*([a,b]) \leq b - a$$

. Also any cover (c,d) of $[a,b]$ must have $c < a \leq b < d$ which implies that $m^*([a,b]) \geq b - a$. The statement follows.

In order to show that $m^*((a,b)) = m^*([a,b)) = m^*((a,b]) = b - a$ we notice that, since $(a,b) \subset (a,b] \subset [a,b]$ and $[a,b) \subset [a,b]$, each of $m^*((a,b)) \leq b-a$, $m^*((a,b]) \leq b - a$ and $m^*([a,b)) \leq b - a$ hold. We have to prove the reverse inequalities.

To that end, let us assume that we con find a cover $U = \cup_{j=1}^{\infty}(a_j,b_j)$ of any of the intervals such that $\sum_{j=1}^{\infty}(b_j - a_j) < b - a$, say $\sum_{j=1}^{\infty}(b_j - a_j) = b - a - 5\delta$ for some $\delta > 0$. Then

$$(a - \delta, a + \delta) \bigcup (b - \delta, b + \delta) \bigcup \cup_{j=1}^{\infty}(a_j,b_j)$$

is a cover of $[a,b]$. This would imply that $m^*([a,b]) \leq b - a - \delta$ contradicting the first part of the proof. We may conclude that $m^*((a,b)) \geq b-a$, $m^*((a,b]) \geq b-a$ and $m^*([a,b)) \geq b - a$. This finishes the proof. □

The next simple lemma that we need is.

Lemma 2.3. *[Monotonicity of the measure.] If $A \subset B$ then $m^*(A) \leq m^*(B)$.*

proof: This follows from the fact that every open cover of B is also an open cover of A. Therefore

$$\inf_{A \subset U} \sum_{j=1}^{\infty}(b_j - a_j) \leq \inf_{B \subset U} \sum_{j=1}^{\infty}(b_j - a_j),$$

where the infinmum is taken over all the open sets $U = \cup_{j=1}^{\infty}(a_j,b_j)$. □

A somewhat refined estimate, the sub-additivity of the measure will be very important later. The main thing we would expect from a measure of length, besides that $m^*([a,b]) = b - a$, is that it is additive $m^*(\cup_j A_j) = \sum_j m^*(A_j)$ for disjoint sets A_j. When proving that m is additive we will repeatedly have to prove statements like $m(\cup_j A_j) = \sum_j m(A_j)$; that is $m(\cup A_j) \leq \sum m(A_j)$ and $m(\cup A_j) \geq \sum m(A_j)$. The following lemma proves one of the inequalities, and it will be referred to frequently in the next lecture.

Lemma 2.4. [SUB-ADDITIVITY OF THE OUTER MEASURE.] *Let A_j be a countable collection of sets in $\mathbb{R}$ then*

$$m^*\left(\cup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} m^*(A_j). \tag{10}$$

Proof: Using the definition of m^* to write (10) we get

$$\inf_{\cup_{j=1}^{\infty} A_j \subset U_{\cup}} \sum_{k=1}^{\infty} (d_k - c_k) \leq \sum_{J=1}^{\infty} \left(\inf_{A_j \subset U_j} \sum_{l=1}^{\infty} (b_{j,l} - a_{j,l}) \right) = \inf_{A_j \subset U_j} \sum_{J=1}^{\infty} \sum_{l=1}^{\infty} (b_{j,l} - a_{j,l})$$

where $U_{\cup} = \cup_{k=1}^{\infty}(c_k, d_k)$ and $U_j = \cup_{l=1}^{\infty}(a_{j,l}, b_{j,l})$ and all the summations and unions are countable.

The lemma follows from noticing that the countable union of the countable collections of intervals $(a_{j,l}, b_{j,l})$ is still a countable collection that we may take as (c_k, d_k). So with any choice on the right side is also a choice on the left side. This yields the lemma. □

Lemma 2.5. *If A is the disjoint union of countably many intervals (open, closed or half open) with endpoints a_j and b_j, and A is bounded,*[9] *then*

$$m^*(A) = \sum_{j=1}^{\infty} (b_j - a_j).$$

Sketch of the Proof: Let us first show the lemma for $A = \cup_{j=1}^{\infty}[a_j, b_j]$.

We claim that, for any finite N,

$$m^*(\cup_{j=1}^{N}[a_j - b_j]) = \sum_{j=1}^{N} (b_j - a_j). \tag{11}$$

Since $\cup_{j=1}^{N}[a_j - b_j]$ is compact it is enough to consider finite open coverings when calculating $m^*(\cup_{j=1}^{N}[a_j - b_j])$. But arguing as in Lemma 2.1 it is easy to see that each interval $[a_j, b_j]$ is covered by one open interval in the covering. Next one can easily show that if any open cover has a connected interval containing two adjacent intervals (adjacent is well defined since N is finite) then we may decrease the outer measure of the cover by splitting that interval into two. It follows that each $[a_j, b_j]$ there is no loss of generality in covering $\cup_{j=1}^{N}[a_j - b_j]$ by disjoint open intervals, each containing exactly one of the $[a_j, b_j]$. The equality (11) easily follows.

By monotonicity of the outer measure and (11):

$$\sum_{j=1}^{N} (b_j - a_j) = m^*(\cup_{j=1}^{N}[a_j - b_j]) \leq m^*(A).$$

Letting $N \to \infty$ implies that

$$\sum_{j=1}^{\infty} (b_j - a_j) \leq m^*(A) \leq \sum_{j=1}^{\infty} (b_j - a_j),$$

[9] Bounded is not really needed, but it makes the proof simpler.

where the last inequality follows from sub-additivity of the measure.

In case some, or all, of the intervals I_j that define $A = \cup_{j=1}^{\infty} I_j$ are open or half open the lemma still holds. Let us briefly indicate why. By sub-additivity and Lemma 2.1 it follows that

$$m^*(A) \leq \sum_{j=1}^{\infty} m^*(I_j) = \sum_{j=1}^{\infty} (b_j - a_j),$$

it is therefore enough to show that

$$m^*(A) \geq \sum_{j=1}^{\infty} (b_j - a_j).$$

Arguing by contradiction we assume that, for some $\delta > 0$

$$m^*(A) = \sum_{j=1}^{\infty} (b_j - a_j) + \delta.$$

This means that there exists an open cover $\cup_{j=1}^{\infty}(c_j, d_j)$ of A such that

$$\sum_{j=1}^{\infty} (c_j - d_j) \leq \sum_{j=1}^{\infty} (b_j - a_j) + \frac{\delta}{2}.$$

If we adjoin the intervals

$$\left(a_j - \frac{\delta}{8 \cdot 2^{-j}}, a_j + \frac{\delta}{8 \cdot 2^{-j}} \right) \text{ and } \left(b_j - \frac{\delta}{8 \cdot 2^{-j}}, b_j + \frac{\delta}{8 \cdot 2^{-j}} \right)$$

to the collection (c_j, d_j) then we get an open cover $U = \cup_{j=1}^{\infty}(e_j, f_j)$ of $\cup_{j=1}^{\infty} \overline{I}_j$ such that

$$\sum_{j=1}^{\infty} (f_j - e_j) = \sum_{j=1}^{\infty} (d_j - c_j) + 2\sum_{j=1}^{\infty} 2\frac{\delta}{8 \cdot 2^{-j}} < \sum_{j=1}^{\infty} (b_j - a_j) = m^*(\cup_{j=1}^{\infty} \overline{I}_j),$$

where we used the first part of the argument in the last equality.[10] But since $\cup_{j=1}^{\infty}(e_j, f_j)$ is an open cover of $\cup_{j=1}^{\infty} \overline{I}_j$ we would get a contradiction. □

The Lemma 2.1 shows that the outer measure m^* at least behaves the way we want on intervals. The good thing with the definition of m^* is that it also gives a well defined length of all sets $A \subset \mathbb{R}$. However, and rather amazingly it turns out that there is no measure μ whatsoever that is defined on all sets $A \subset \mathbb{R}$ that has the good properties that we would expect of a measure.

Proposition 2.1. [Vitali sets] *There is no non-negative function $\mu : \mathcal{P}(\mathbb{R}) \mapsto \mathbb{R}_+$, $\mu \neq 0$ and $\mu(A) \neq \infty$ for bounded sets A, such that:* [11]

1. *$\mu(A) = \mu(x_0 + A)$ for all sets $A \subset \mathbb{R}$.*[12]

[10]Here we are a little sketchy. In particular, even if I_j are disjoint it might not follow that $\overline{I}_j$ are disjoint. I am not quite sure that it is very interesting to investigate this here so I will leave it to the reader to clarify this point.

[11]We use the notation $\mathcal{P}(\mathbb{R})$ for all subsets of $\mathbb{R}$, or more generally $\mathcal{P}(S)$ is the set of all subsets of S. We also use the notation $\mathbb{R}_+ = \{x \in \mathbb{R};\ x \geq 0\}$.

[12]Here $x_0 + A$ is the translation of A by x_0; $x_0 + A = \{x_0 + x;\ x \in A\}$.

2. $\mu\left(\cup_{j=0}^{\infty}A_j\right) = \sum_{j=1}^{\infty}\mu(A_j)$ *for any countable and disjoint collection of sets* A_j*.*

Proof: Let us first show that the proposition holds on the set $\mathcal{P}([0,1))$ if we interpret all numbers modulo 1. At the end of the proof we will indicate how to treat the case stated in the proposition.

Define the equivalence relation $x \approx y$ if $x = y+q$ for some $q \in \mathbb{Q}$ (everything is calculated modulo 1). By using the axiom of choice we may form a set A consisting of one element from every equivalence class. Then each $x \in [0,1)$ may be written $x = y+q$ for $y \in A$ and $q \in \mathbb{Q}\cap[0,1)$. Since $q \in \mathbb{Q}\cap[0,1)$ is a countable set we may define the sets $A_j = q_j + A$, where $\{q_j;\ j \in \mathbb{N}\} = \mathbb{Q}\cap[0,1)$. Since every $x \in [0,1)$ can be written $x+q$ it follows that $\cup_{j=1}^{\infty}A_j = [0,1)$. By construction $A_j \cap A_k = \emptyset$ if $j \neq k$, this since A only contains one element form each equivalence class.

To summarize A_j forms a countable disjoint collection of sets such that

$$[0,1) = \cup_{j=0}^{\infty}A_j. \tag{12}$$

Furthermore, for each $j,k \in \mathbb{N}$,

$$A_j = A_k + q \quad \text{for some } q \in \mathbb{Q}\cap[0,1). \tag{13}$$

Assume, aiming for a contradiction, that there a function μ as in the proposition then, in view of assumption 1 and (13), $\mu(A_j) = \mu(A_k)$ for all $j,k \in \mathbb{N}$. Also, from assumption 2 and (12), we may conclude that

$$\mu([0,1)) = \sum_{j=0}^{\infty}\mu(A_j). \tag{14}$$

Since $\mu \neq 0$, and $\mu \geq 0$, it follows that $\mu([0,1)) > 0$ and therefor not all $\mu(A_j) \neq 0$. We conclude that $\mu(A_j) = c > 0$ for some j therefore, and because of (13), $\mu(A_j) = c > 0$ for all j. But if $\mu(A_j) = c > 0$ for all j then the series in (14) diverges which means that $\mu([0,1)) = \infty$ for a bounded set. We get a contradiction.

If we want to prove the same thing for $\mathbb{R}$ then we may argue similarly and define the set A as containing one representative from each equivalence class and then define

$$A_j = \underbrace{\{x+q_j;\ x \in A,\ x+q_j \in [0,1)\}}_{=A_j^+} \bigcup \underbrace{\{x+q_j-1;\ x \in A,\ x+q_j \in [1,2)\}}_{=A_j^-}.$$

Then, for all j, $\mu(A_j) = \mu(A)$ since $\mu(A_j) = \mu(A_j^+ \cup A_j^-) = \mu(A_j^+) + \mu(A_j^-) = \mu(A_j^+) + \mu(1 + A_j^-) = \mu(q_j + A) = \mu(A)$. We arrive at

$$\mu([0,1)) = \sum_{j=1}^{\infty}\mu(A_j). \tag{15}$$

Since the series in (15) cannot be ∞ since the left side is the measure of a finite set we can conclude that $\mu(A_j) = 0$, and therefore $\mu([0,1)) = 0$. It follows that

$$\mu(\mathbb{R}) = \mu\left(\cup_{k\in\mathbb{Z}}[k,k+1)\right) = \sum_{k\in\mathbb{Z}}\mu([0,1)) = 0,$$

where we used translation invariance (assumption 1) in the last equality. But $\mu(\mathbb{R}) = 0$ is a contradiction to $\mu \neq 0$. $\square$

Since any reasonable definition of what length is should include the assumptions 1 and 2 we need to define the measure on a smaller domain than $\mathcal{P}(\mathbb{R})$. It is not absolutely clear what domain is the right domain of definition of the measure m^*. It turns out that the right definition of the (restricted) domain of m^* are the measurable sets.

Definition 2.2. *We define the Lebesgue measure m to be equal to m^* on the sets $S \subset \mathbb{R}$ that satisfies, for all sets $X \subset \mathbb{R}$,*

$$m^*(S) = m^*(X \cap S) + m^*(X \cap S^c). \tag{16}$$

We call the sets S that satisfy (16) measurable and the collection of all measurable sets will be denoted $\mathcal{M}$.

That is, the Lebesgue measure m is just m^* with domain of definition restricted to on the collection of measurable sets $\mathcal{M}$. In order for the measure to be useful we need to show that it satisfies some basic properties. In particular we would want the measure to satisfy the countable additivity condition

$$m(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} m(A_j) \tag{17}$$

for each countable collection of measurable sets A_j. But for the countable additivity condition to be meaningful we need $\cup_{j=1}^{\infty} A_j$ to be measurable whenever the sets A_j are.

Defining the Lebesgue measure and measurable sets the way we do leads to two big questions. First: Will m satisfy the countable additivity condition (17)? Second: Which sets are measurable? If the class of measurable sets is to small then the Lebesgue measure will be useless.

We will show that all open sets are measurable and also that the measurable sets forms a $\sigma-$algebra - being a $\sigma-$algebra implies that the set of measurable sets $\mathcal{M}$ is rich and flexible enough to use for integration.

Definition 2.3. *Let R be a set and $\mathcal{S}$ be a collection of subsets of R. Then we say that $\mathcal{S}$ is a $\sigma-$algebra if*

1. *$\emptyset \in \mathcal{S}$ and $R \in \mathcal{S}$,*
2. *if $A \in \mathcal{S}$ then $A^c \in \mathcal{S}$ and*
3. *if $A_j \in \mathcal{S}$, $j \in \mathbb{N}$, then $\cup_{j=1}^{\infty} A_j \in \mathcal{S}$.*

Remark: *Notice that the third condition also implies that finite unions $\cup_{j=1}^{N} A_j \in \mathcal{S}$. This follows by choosing $A_j = \emptyset$ for $j > N$.*

One of our aims will be to show that $\mathcal{M}$ is a $\sigma-$algebra of subsets of $\mathbb{R}$. That $\emptyset \in \mathcal{M}$ and $\mathbb{R} \in \mathcal{M}$ is clear from the definition of $\mathcal{M}$. It also follows directly from the definition of $\mathcal{M}$ that if $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$.[13] We therefore only need to show that if $A_j \in \mathcal{M}$, $j \in \mathbb{N}$, then $\cup_{j=1}^{\infty} A_j \in \mathcal{M}$. The last condition is,

[13] This is clear form (16) which is symmetric in A and A^c. In particular, substituting A^c for A (and using $(A^c)^c = A$) will not change (16).

of course, what we need in order to prove the countably additivity condition for m.

It is rather easy to show that if $\mathcal{S}$ is a $\sigma-$algebra then $\mathcal{S}$ contains many more sets than what is obvious from the definition.

Proposition 2.2. *Let $\mathcal{S}$ be a $\sigma-$algebra of subsets of R then*

1. *if $A, B \in \mathcal{S}$ then $A \setminus B \in \mathcal{S}$ and*
2. *if $A_j \in \mathcal{S}$, $j \in \mathbb{N}$, then $\cap_{j=1}^{\infty} A_j \in \mathcal{S}$.*

Remark: *That finite intersections $\cap_{j=1}^{N} A_j \in \mathcal{S}$ if $A_j \in \mathcal{S}$ follows from the final condition by choosing $A_j = R$ for $j > N$.*

Proof of Proposition 2.2: The first statement follows from the second since $A \setminus B = A \cap (B^c)$ and if $B \in \mathcal{S}$ then $B^c \in \mathcal{S}$.

To show the second statement we just notice that

$$\bigcap_{j=1}^{\infty} A_j = \left(\cup_{j=1}^{\infty} A_j^c \right)^c.$$

□

We will end this lecture by arguing that the, admittedly rather technical, definition of a measurable set gives a rather natural definition of length.

Example: *Later we will prove that open and closed sets are measurable, but for this example we will assume these facts.*

Assume that $A \subset [0,1]$ and we want to find the measure of A, assume also the measure of open and closed sets is well defined. The measure of A must be greater than (or equal to) $\sup_{K \subset A} m^*(K)$, K closed, and less than (or equal to) $\inf_{A \subset U} m^*(U) = m^*(A)$, U open.[14] *The measure of A would then be well defined if*

$$\sup_{K \subset A} m^*(K) = \inf_{A \subset U} m^*(U) = m^*(A). \tag{18}$$

But if $[0,1] \setminus K$, for $K \subset A$, is an open set that contains $[0,1] \setminus A$ therefore

$$\sup_{K \subset A} m^*(K) = m^*([0,1]) - \inf_{([0,1] \setminus A) \subset U} m^*(U) = m^*([0,1]) - m^*([0,1] \setminus A) \tag{19}$$

- notice that because of Lemma 2.1 the definition of the length of open sets is rather uncomplicated and unambiguous which means that it is rather uncomplicated to define the length of closed sets.

Using (19) in (18) we get that the length of A is "well defined" only if

$$m^*([0,1]) - m^*([0,1] \setminus A) = m^*(A) \Rightarrow m^*([0,1]) = m^*([0,1] \cap A^c) + m^*([0,1] \cap A),$$

which is exactly the condition we get in the definition of measurable set with $X = [0,1]$. That allow X to be a general set instead of an interval is a matter of adjusting to the tradition of measure theory.[15] *The point is that the condition of measurable more or less states that we can ascribe a measure to a set A if the largest closed set contained in A has the same measure as the smallest open set containing A. The spirit of the definition of measurable is that the "length" of a measurable set can be sandwiched between two sets whose measure we know.*

[14] There might be a slight mystery why we want K to be closed and U to be open, but let us accept that.

[15] And that in more abstract cases, for sets different than $\mathbb{R}$, there might not be anything as natural as an interval to use.

3 Lecture 3: Measure Theory II (my lecture).

In this lecture we will prove that the measurable sets really form a $\sigma-$algebra and that the Lebesgue measure satisfies the countable additivity condition. We also need to show that the set of measurable sets is rich, in particular we will show that the measurable sets contains all open sets. The material is rather technical but, in its own way, very amazing.

Our first goal is to show that all open sets are measurable. The proof is rather long so we will begin with a lemma.

Lemma 3.1. *Assume that U is open and that $I = (a, b)$ then*

$$m^*(I) = m^*(U \cap I) + m^*(U \cap I^c).$$

Proof: By Lemma 2.4 it follows that

$$m^*(I) \leq m^*(U \cap I) + m^*(U^c \cap I).$$

Therefore we only need to show that

$$m^*(I) \geq m^*(U \cap I) + m^*(U^c \cap I). \tag{20}$$

By Lemma 2.1 we may write $U = \cup_{j=1}^{\infty}(a_j, b_j)$. We may also assume that U is bounded since we intersect U by a the bounded set I in each occurrence in (20). The argument will be split up into several steps.

Step 1: *There exists, for every $\epsilon > 0$, an N such that*

$$m^*(U) \leq \sum_{j=1}^{N}(b_j - a_j) + \epsilon.$$

There is no loss of generality to assume that each $(a_j, b_j) \cap I \neq \emptyset$ and, upon relabeling, to assume that $a_j < b_j < a_{j+1} < b_{j+1}$. Define $U_\epsilon = \cup_{j=1}^{N}(a_j, b_j)$ and notice that the closed set U_ϵ^c is the union of

$$(-\infty, a_1], [b_1, a_2], [b_2, a_3], ..., [b_{N-1}, a_N], [b_N, \infty).$$

Proof of Step 1: This is clear since $\sum_{j=1}^{\infty}(b_j - a_j)$ is convergent, by the assumption that U is bounded. From Lemma 2.5 may conclude

$$m^*(U) = \sum_{j=1}^{\infty}(b_j - a_j) \leq \sum_{j=1}^{N}(b_j - a_j) + \epsilon.$$

The final parts of step 1 are just there for some book-keeping and should be clear. Throwing out the intervals (a_j, b_j) that does not intersect I should not effect anything and that the complement of U^c is closed and have the stated form is elementary.

We have four different cases to consider:

1. $a \in [a_1, b_1]$ and $b \in [a_n, b_N]$ or
2. $a \in [a_1, b_1]$ and $b \in [b_N, \infty)$ or

3. $a \in (-\infty, a_1]$ and $b \in [a_n, b_N]$ or

4. $a \in (-\infty, a_1]$ and $b \in [b_N, \infty)$.

All cases are handled in a very similar fashion so we will assume that we are in case 2 and leave the other cases to the reader.

Step 2: *The following equality holds*

$$m^*(U_\epsilon \cap I) = (b_1 - a) + \sum_{j=2}^{N} (b_j - a_j).$$

Proof of Step 2: Notice that we may write $U_\epsilon \cap I$ as a disjoint union of intervals:

$$U_\epsilon \cap I = (a, b_1) \cup (a_2, b_2) \cup (a_3, b_3) \cup ... \cup (a_N, b_N)$$

Step 2 follows from Lemma 2.5.

Step 3: *The following equality holds*

$$m^*(U_\epsilon^c \cap I) = \sum_{j=1}^{N-1} (a_{j+1} - b_j) + (b - b_N).$$

Proof of Step 3: Similar to step 2. We may write $U_\epsilon^c \cap I$ as a disjoint union of intervals:

$$U_\epsilon \cap I = (b_1, a_2) \cup (b_2, a_3) \cup (b_3, a_4) \cup ... \cup (b_N, b)$$

Step 3 follows from Lemma 2.5.

Step 4: *For every $\epsilon > 0$ the following holds*

$$m^*(U \cap I) + m^*(U^c \cap I) \leq m^*(I) + \epsilon,$$

in particular (20) holds. This finishes the proof.

Proof of Step 4: Since $U \setminus U_\epsilon = \cup_{j=N+1}^{\infty} (a_j, b_j)$ and $\sum_{j=N+1}^{\infty} (b_j - a_j) < \epsilon$ it follows that

$$m^*(U \cap I) \leq m^*(U_\epsilon \cap I) + m^* \left((\cup_{j=N+1}^{\infty} (a_j, b_j)) \cap I \right) \leq$$

$$\leq m^*(U_\epsilon \cap I) + m^* \left((\cup_{j=N+1}^{\infty} (a_j, b_j)) \right) < m^*(U_\epsilon \cap I) + \epsilon, \tag{21}$$

where we used sub-additivity and that $(\cup_{j=N+1}^{\infty} (a_j, b_j)) \cap I \subset (\cup_{j=N+1}^{\infty} (a_j, b_j))$ (together with the monotonicity of the measure).

Also, by monotonicity of the measure and $U^c \subset U_\epsilon^c$,

$$m^*(U^c \cap I) \leq m^*(U_\epsilon^c \cap I). \tag{22}$$

From (21) and (22) we conclude that

$$m^*(U \cap I) + m^*(U^c \cap I) < m^*(U_\epsilon^c \cap I) + m^*(U_\epsilon \cap I) + \epsilon =$$

$$= (b_1 - a) + \sum_{j=2}^{N} (b_j - a_j) + \sum_{j=1}^{N-1} (a_{j+1} - b_j) + (b - b_N) + \epsilon = b - a + \epsilon,$$

where we used step 2 and 3 in the middle equality. □

Proposition 3.1. *Every open set U is measurable.*

Proof: We need to show that, for any $X \subset \mathbb{R}$,

$$m^*(X \cap U) + m^*(X \cap U^c) = m^*(X).$$

By sub-additivity it is enough to show that

$$m^*(X \cap U) + m^*(X \cap U^c) \leq m^*(X), \tag{23}$$

again we will show the last inequality with an arbitrary small ϵ error.

Let $\epsilon > 0$ and find a cover $\cup_{j=1}^{\infty} I_j$ of X, $I_j = (a_j, b_j)$, of X such that

$$\sum_{j=1}^{\infty} (b_j - a_j) < m^*(X) + \epsilon, \tag{24}$$

this we can always do by the definition of $m^*(X)$ as the infimum of all series such as the left side in (24).

By Lemma 2.1 we may also write $U = \cup_{j=1}^{\infty} J_j$, where $J_j = (c_j, d_j)$ are disjoint intervals.

Using that $X \subset \cup_j I_j$, monotonicity of the outer measure and sub-additivity of m^* we may calculate

$$m^*(U \cap X) \leq m^*\left(U \cap (\cup_j I_j)\right) \leq \sum_{j=1}^{\infty} m^*(U \cap I_j). \tag{25}$$

And similarly

$$m^*(U^c \cap X) \leq \sum_{j=1}^{\infty} m^*(U^c \cap I_j) \tag{26}$$

From (25) and (26) we may conclude that

$$m^*(U \cap X) + m^*(U^c \cap X) \leq \sum_{j=1}^{\infty} m^*(U \cap I_j) + \sum_{j=1}^{\infty} m^*(U^c \cap I_j) = \tag{27}$$

$$= \sum_{j=1}^{\infty} \left(m^*(U \cap I_j) + m^*(U^c \cap I_j) \right).$$

From Lemma 3.1 we may conclude that

$$\sum_{j=1}^{\infty} \left(m^*(U \cap I_j) + m^*(U^c \cap I_j) \right) = \sum_{j=1}^{\infty} m^*(I_j) = \sum_{j=1}^{\infty} (b_j - a_j) < m^*(X) + \epsilon, \tag{28}$$

where we also used (24) in the last inequality.

From (27) and (28) we can conclude that

$$m^*(U \cap X) + m^*(U^c \cap X) < m^*(X) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary this proves (23) and finishes the proof of the proposition.

□

Corollary 3.1. *Every closed set is measurable.*

Proof: Since closed sets are complements of open sets and the definition of measurable is symmetric w.r.t. the complement this follows directly from Proposition 3.1. □

In order to get a rich enough class of measurable sets to show that the integral has good convergence properties it is not enough to show that all the open and all the closed sets are measurable. We will need to show that null sets are measurable as well. First we need to define the concept of null sets.

Definition 3.1. *We say that a set $A \subset \mathbb{R}$ is a null set if*

$$m^*(A) = 0.$$

Proposition 3.2. *All the null sets are measurable.*

Proof: Let A be a null set then, by the monotonicity of the outer measure, $m^*(X \cap A) \leq m^*(A) = 0$ and $m^*(X \cap A^c) \leq m^*(X)$. This clearly implies that

$$m^*(X \cap A) + m^*(X \cap A^c) \leq m^*(X).$$

The reverse inequality follows from sub-additivity. □

Next we will show that m^* is countably additive on measurable sets: if A_j, $j \in \mathbb{N}$, is a countable collection of disjoint and measurable sets. Then $m^*(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} m^*(A_j)$. Again this requires some work. We begin by showing that m^* is finitely additive on disjoint and measurable sets.

Lemma 3.2. [Finite additivity] *Let A_j, $j = 1, 2, ..., N$, be a finite collection of disjoint measurable sets. Then*

$$m^*(\cup_{j=1}^{N} A_j) = \sum_{j=1}^{N} m^*(A_j).$$

Proof: By induction on N. We start with the base case, $N = 2$, then we define $X = A_1 \cup A_2$. By the definition of measurable, using that A_2 is measurable, it follows that

$$m^*(A_1 \cup A_2) = m^*(X) = m^*(\underbrace{X \cap A_2}_{=A_2}) + m^*(\underbrace{X \cap A_2^c}_{=A_1}) = m^*(A_2) + m^*(A_1),$$

the Lemma follows for $N = 2$.

Assume that the lemma holds for all measurable and disjoint collections A_j, $j = 1, 2, ..., N$, we want to show that for any collection A_j, $j = 1, 2, ..., N+1$, of disjoint and measurable sets

$$m^*(\cup_{j=1}^{N+1} A_j) = \sum_{j=1}^{N+1} m^*(A_j). \tag{29}$$

We argue as in the base case and define $X = \left(\cup_{j=1}^{N} A_j\right) \cup A_{N+1}$. Then, since A_{N+1} is measurable,

$$m^*\left(\left(\cup_{j=1}^{N} A_j\right) \cup A_{N+1}\right) = m^*(X) = m^*(\underbrace{X \cap A_{N+1}}_{=A_{N+1}}) + m^*(\underbrace{X \cap A_{N+1}^c}_{=\cup_{j=1}^{N} A_j}) =$$

$$= m^*(A_{N+1}) + m^*\left(\cup_{j=1}^N A_j\right) = \sum_{j=1}^{N+1} m^*(A_j),$$

where we used that $m^*\left(\cup_{j=1}^N A_j\right) = \sum_{j=1}^N m^*(A_j)$ by the induction hypothesis in the last equality. The lemma follows by induction. $\square$

We are now ready to prove countable additivity.

Proposition 3.3. [Countable Additivity.] *Let A_j, $j = 1, 2, 3, ...,$ be a countable collection of disjoint measurable sets. Then*

$$m^*(\cup_{j=1}^\infty A_j) = \sum_{j=1}^\infty m^*(A_j).$$

Proof: Again by sub-additivity of the measure it is enough to prove that

$$\sum_{j=1}^\infty m^*(A_j) \le m^*\left(\cup_{j=1}^\infty A_j\right). \tag{30}$$

By monotonicity of the measure it follows that

$$m^*\left(\cup_{j=1}^\infty A_j\right) \ge m^*\left(\cup_{j=1}^N A_j\right) = \sum_{j=1}^N m^*(A_j). \tag{31}$$

Passing to the limit $N \to \infty$ in (31) gives (30). $\square$

We also need to show that the measurable sets $\mathcal{M}$ forms a $\sigma-$algebra. By the definition of measurable sets, in particular by the symmetry in taking complements it is clear that $A \in \mathcal{M} \Rightarrow A^c \in \mathcal{M}$. Also that $\emptyset \in \mathcal{M}$ (since $\emptyset$ is a null set) and that $\mathbb{R} \in \mathcal{M}$ (since $\mathbb{R} = \emptyset^c$) is clear. We will however have to show that if $A_j \in \mathcal{M}$, $j \in \mathbb{N}$, then $\cup_{j=1}^\infty A_j \in \mathcal{M}$. This will require some work. We begin with a lemma.

Lemma 3.3. *If $A_j \in \mathcal{M}$, $j = 1, 2, ..., N$, then $\cup_{j=1}^N A_j \in \mathcal{M}$.*

Proof: We will prove this by induction on N. We begin with the base case $N = 2$.

To no ones surprise (I hope) it is, by sub-additivity, enough to show the inequality

$$m^*\left([A_1 \cup A_2] \cap X\right) + m^*\left([A_1 \cup A_2]^c \cap X\right) \le m^*(X),$$

for any set $X \subset \mathbb{R}$.

In order to show this we use that A_1 is measurable and calculate

$$m^*(X) = m^*(X \cap A_1) + m^*(X \cap A_1^c) =$$

$$= m^*(X \cap A_1) + m^*\left([X \cap A_1^c] \cap A_2\right) + m^*\left([X \cap A_1^c] \cap A_2^c\right), \tag{32}$$

where we used that A_2 is measurable in the last equality, which implies that

$$m^*(X \cap A_1^c) = m^*\left([X \cap A_1^c] \cap A_2\right) + m^*\left([X \cap A_1^c] \cap A_2^c\right).$$

In order to continue we use sub-additivity on the first two terms on the right in (32)

$$m^*(X \cap A_1) + m^*\left([X \cap A_1^c] \cap A_2\right) \ge$$

$$\geq m^*\left([X \cap A_1] \cup [(X \cap A_1^c) \cap A_2]\right) = m^*\left(X \cap [A_1 \cup A_2]\right), \tag{33}$$

where we used the equality

$$[X \cap A_1] \cup [(X \cap A_1^c) \cap A_2] = X \cap [A_1 \cup A_2]$$

in the last step of the calculation. We also rewrite the last term in (32) as

$$m^*\left([X \cap A_1^c] \cap A_2^c\right) = m^*\left(X \cap [A_1 \cup A_2]^c\right). \tag{34}$$

Using (33) and (34) in (32) we can conclude that for any set $X \in \mathbb{R}$

$$m^*(X) \geq m^*\left(X \cap [A_1 \cup A_2]\right) + m^*\left(X \cap [A_1 \cup A_2]^c\right),$$

which implies that $A_1 \cup A_2$ is measurable.

The induction step is easy. Given sets A_j, $j =, 2, ..., N+1$, and assuming that then lemma holds for all collections consisting of at most N sets we can conclude that

$$\cup_{j=1}^{N} A_j = \underbrace{\left(\cup_{j=1}^{N} A_j\right)}_{\in \mathcal{M}} \bigcup A_{N+1} \in \mathcal{M},$$

since the first union to the right is measurable by the induction hypothesis. □

We can now prove that $\mathcal{M}$ is closed under countable unions.

Proposition 3.4. *Let $A_j \in \mathcal{M}$, $j = 1, 2, 3, ...,$ then $A = \cup_{j=1}^{\infty} A_j \in \mathcal{M}$.*

Proof: As always we only need to show that

$$m^*(X) \geq m^*(X \cap A) + m^*(X \cap A^c).$$

The trick is to write A as a countable disjoint union (in order to use additivity of the measure). To that end we recursively define $T_1 = A_1$ and $T_n = A_n \setminus \cup_{j=1}^{n-1} T_j$. Then each T_j is measurable since A_n is and, by Lemma 3.3, $\cup_{j=1}^{n-1} T_j$ is. We also define $L_n = \cup_{j=1}^{n} T_j = \cup_{j=1}^{n} A_j$, which is also measurable by Lemma 3.3. Therefore

$$m^*(X) = m^*(X \cap L_n) + m^*(X \cap L_n^c) \geq m^*(X \cap L_n) + m^*(X \cap A^c), \tag{35}$$

by the monotonicity of the measure since $X \cap A^c \subset X \cap L_n^c$. Using the definition of L_n in (35) we conclude that

$$m^*(X) \geq m^*\left(X \cap (\cup_{j=1}^{n} T_j)\right) + m^*(X \cap A^c). \tag{36}$$

Since T_n is measurable we may calculate

$$m^*\left(X \cap (\cup_{j=1}^{n} T_j)\right) =$$

$$= m^*\Big(\underbrace{\left[X \cap (\cup_{j=1}^{n} T_j)\right] \cap T_n}_{= X \cap T_n}\Big) + m^*\Big(\underbrace{\left[X \cap (\cup_{j=1}^{n} T_j)\right] \cap T_n^c}_{= X \cap (\cup_{j=1}^{n-1} T_j)}\Big) =$$

$$= m^*\big(X \cap T_n\big) + m^*\big(X \cap (\cup_{j=1}^{n-1} T_j)\big) = \{\text{repeat}\} = \sum_{j=1}^{n} m^*(X \cap T_j),$$

where the first two equalities shows how to "move out " T_n from the measure and the "repeat" just indicates that we do the same argument again to "move out" T_{n-1} and then T_{n-1} et.c. Using the last equality in (36) we can conclude that

$$m^*(X) \geq \sum_{j=1}^{n} m^*(X \cap T_j) + m^*(X \cap A^c). \tag{37}$$

Letting $n \to \infty$ in (37) we get

$$m^*(X) \geq \sum_{j=1}^{\infty} m^*(X \cap T_j) + m^*(X \cap A^c). \tag{38}$$

But $A = \cup_{j=1}^{\infty} A_j = \cup_{j=1}^{\infty} T_j$ and thus by sub-additivity

$$m^*(X \cap A) = m^*\left(X \cap \left(\cup_{j=1}^{\infty} T_j\right)\right) =$$

$$= m^*\left(\cup_{j=1}^{\infty}\left(X \cap T_j\right)\right) \leq \sum_{j=1}^{\infty} m^*(X \cap T_j). \tag{39}$$

Inserting (39) in (38) we can conclude that

$$m^*(X) \geq m^*(X \cap A) + m^*(X \cap A^c).$$

The proposition follows. □

Let us formulate the main results we have proven this far in a theorem.

Theorem 3.1. *The Lebesgue measure m, that is the outer Lebesgue measure m^* restricted to the collection of measurable sets $\mathcal{M}$, is a non-negative countable additive function.*

Furthermore, the collection of measurable sets $\mathcal{M}$ forms a $\sigma-$algebra.

This is terribly abstract and at this point it might be difficult to see if it was worth it to go through many very technical proofs in order to derive a this theorem. It is hardly the kind of theorem whose statement makes its applications obvious. In the next lecture we will see that this theorem is exactly what we need in order to define a versatile integral that behaves well under limits. And it is only when proving the theorems for the integral we will be able to see that the theory developed this far is right.

Before we define the integral it will be good to gain a little better understanding of what a measurable set is and how general or bad it can be. We will also need a "continuity" result for measures. We will begin to show that any measurable set can be written as the intersection of countable many open sets and a null set.

Proposition 3.5. *Let $A \subset \mathbb{R}$ be any set, then there exists a set $B = \bigcap_{j=1}^{\infty} U_j$, where U_j are open sets and $A \subset B$, such that*

$$m^*(A) = m^*(B). \tag{40}$$

If A is measurable then

$$m^*(B \setminus A) = 0,$$

in particular any measurable set differs from the intersection of countably many open sets by a null set.

Proof: For a general set A it follows directly from the definition of the outer measure m^* that there exists an open set U_j such that $A \subset U_j$

$$m^*(U_j) \leq m^*(A) + \frac{1}{j}, \tag{41}$$

We may assume that

$$... \subset U_j \subset U_{j-1} \subset ... \subset U_1. \tag{42}$$

If not we may define $\tilde{U}_1 = U_1$ and inductively $\tilde{U}_j = U_j \cap \tilde{U}_{j-1}$, then $\tilde{U}_j \subset \tilde{U}_{j-1}$ will satisfy (42). The sets $\tilde{U}_j$ will also satisfy (41), by the monotonicity of the measure and from $A \subset \tilde{U}_j \subset U_j$.

By monotonicity of the measure, (42) and (41) it follows that $B = \cap_{j=1}^{\infty} U_j$ satisfies

$$m^*(A) \leq m^*(B) \leq m^*(U_j) \leq m^*(A) + \frac{1}{j}$$

for any $j \in \mathbb{N}$. This implies (40).

If A is measurable then it follows from the definition of measurable, $A \subset B$ and (40) that

$$m(A) = m(B) = m(B \cap A) + m(B \setminus A) = m(A) + m(B \setminus A),$$

subtracting $m(A)$ from both sides gives $m(B \setminus A) = 0$. □

The decomposition in the previous proposition is very powerful, and exact, but it is based on an possibly infinite intersection. At times it might be beneficial to have a finitary approximation of a measurable set. We provide such a finite approximation in the next proposition.

Proposition 3.6. *Assume that A is bounded and measurable. Then, for any $\epsilon > 0$, there exists a finite union of open intervals $U = \cup_{j=1}^{N}(a_j, b_j)$ such that*

$$m(A \Delta U) < \epsilon,$$

here Δ is denotes the symmetric difference $A \Delta U = (A \setminus U) \bigcup (U \setminus A)$.

Proof: If A is measurable then there exists an open set $\hat{U} = \cup_{j=1}^{\infty}(a_j, b_j)$ such that $A \subset \hat{U}$ and

$$m(\hat{U}) < m(A) + \frac{\epsilon}{2}.$$

Choose N large enough so that

$$m(\hat{U}) < \sum_{j=1}^{N}(b_j - a_j) + \frac{\epsilon}{2}$$

and define $U = \cup_{j=1}^{N}(a_j, b_j)$. Then using sub-additivity and monotonicity

$$m(A \Delta U) \leq m(\underbrace{A \setminus U}_{\subset \hat{U} \setminus U}) + m(\underbrace{U \setminus A}_{\subset \hat{U} \setminus A}) \leq m(\hat{U} \setminus U) + m(\hat{U} \setminus A) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

The final result of this lecture is

Proposition 3.7. [Continuity of the Measure] *If A_j is a countable collection of measurable sets then*

1. *if $A_j \subset A_{j+1}$ and $A = \cup_{j=1}^{\infty} A_j$ then*

$$m(A) = \lim_{j\to\infty} m(S_j). \tag{43}$$

2. *if $A_{j+1} \subset A_j$, $m(A_1) < \infty$, and $A = \cap_{j=1}^{\infty} A_j$ then*

$$m(A) = \lim_{j\to\infty} m(S_j). \tag{44}$$

Proof: We begin with the first statement and assume that $A_j \subset A_{j+1}$. We construct the disjoint sets $B_j = A_j \setminus A_{j-1}$ (taking $A_0 = \emptyset$) then B_j forms a disjoint measurable collection and $A = \cup_{k=1}^{\infty} B_k$ and $A_j = \cup_{k=1}^{j} B_k$.

By countable additivity for the measure we may calculate

$$\lim_{j\to\infty} m(A_j) = \lim_{j\to\infty} m(\cup_{k=1}^{j} B_k) = \lim_{j\to\infty} \sum_{k=1}^{j} m(B_k) =$$

$$= \sum_{k=1}^{\infty} m(B_k) = m(\cup_{k=1}^{\infty} B_k) = m(A).$$

This proves the first statement.

The second part follows from the first part. Notice that the sets $C_j = A_1 \setminus A_j$ forms an increasing sequence of measurable sets and $\cup_{j=1}^{\infty} C_j = A_1 \setminus A$. Using that all sets are measurable and (43) (with C_j in place of A_j) at the indicated place we can derive that

$$m(A_1) - m(A) = m(A_1 \setminus A) = \{(43)\} = \lim_{j\to\infty} m(C_j) =$$

$$= \lim_{j\to\infty} (m(A_1) - m(A_j)) = m(A_1) - \lim_{j\to\infty} m(A_j).$$

The statement (44) follows by canceling $m(A_1)$ and multiplying by -1. □

3.1 Reading

The above material is what I covered during the lecture. If you want to read it is a carefully edited text see:

W. Rudin, Real and complex analysis: Chapter 1-2 (not the Riesz representation Thm): Most important theorems: Thm 1.26 and 1.34

Recommended exercises: *1.3, 1.4, 1.5, 1.6, 1.7, 1.9, 1.12, 2.5, 2.20, 2.25*

H.L. Royden P.M Fitzpatrick, Real Analysis: Chapter 2.1-4.5 (Skip 2.7 for now) Most important: Bdd Conv. Thp p.78 and Mon. Conv. Thm p.83

Recommended exercises: *2.5, 2.6, 2.17, 2.20, 3.1, 3.3, 3.7, 3.12, 3.18, 3.22, 3.31, 4.10, 4.13, 2.25, 4.28*

E. Stein R. Shakarchi, Real Analysis Chapter 1-2.2. Most important: Thm 1.4 and Cor 1.9

4 Lecture 4. The Lebegue integral.

We will continue with the material from the previous section (no new reading) and try to understand the convergence properties of sequences of Lebesgue integrable functions. We will also try to understand how different a Lebesgue integrable function is from a continuous function.

5 Lecture 5. Integration on product spaces and Fubini.

In lecture 2-3 we will primarily focus on functions of one variable. In this lecture we will try to understand integration in several variables.

Reading:

W. Rudin, Real and complex analysis: Chapter 8 pp.160-170

RECOMMENDED EXERCISES: *8.2, 8.3, 8.6, 8.12*

H.L. Royden P.M Fitzpatrick, Real Analysis: Chapter 20.1 (skim through 20.2)

RECOMMENDED EXERCISES: *20.6, 20.9, 20.10, 20.16*

E. Stein R. Shakarchi, Real Analysis Chapter 2.3

6 Lecture 6. The L^p spaces.

One of the main problems in showing existence of minimizers in the calculus of variations problem was that we needed the domain $\mathcal{K}_E$ of the functional E to be complete. So far we have only defined an integral (in several dimensions) that has good convergence properties. In this lecture we will define a space $L^p(D)$, for $1 \leq p \leq \infty$, of all measurable functions $u : D \mapsto \mathbb{R}$ such that

$$\|u\|_{L^p(D)} := \left(\int_{\mathcal{D}} |u(\mathbf{x})|^p d\mathbf{x} \right)^{1/p} < \infty.$$

We will prove several good properties of this space. Later we will consider a subset of the functions $u(\mathbf{x})$ such that $\frac{\partial u(\mathbf{x})}{\partial x_i} \in L^p(\mathcal{D})$ as the domain $\mathcal{K}_E$ of E. But that will require some more work - since we do not even know what it means of a derivative to be in $L^p(\mathcal{D})$.

Reading:

W. Rudin, Real and complex analysis: Chapter 3 and pages 95-97

RECOMMENDED EXERCISES: *3.11, 3.12, 3.13, 3.18*

H.L. Royden P.M Fitzpatrick, Real Analysis: Chapter 7.1-7.3

RECOMMENDED EXERCISES: *7.1, 7.10, 7.13, 7.19, 7.24, 7.25*

E. Stein R. Shakarchi, Real Analysis **Not covered**

7 Lecture 7. Duality and functionals.

So far we have seen that L^p is a complete space: that is if $u^j \in L^p$ forms a Cauchy sequence (for all $\epsilon > 0$ there exists an $N_\epsilon \in \mathbb{N}$ such that $\|u^j - u^k\|_{L^p} < \epsilon$ for all $j, k > N_\epsilon$) then there exists a function $u^0 \in L^2$ such that $\|u^j - u^0\|_{L^p} \to 0$. However, what we are interested in is to show that if $\|\nabla u^j\|_{L^p} \leq M$ then there exits a sub-sequence u^{j_k} and a function u^0 such that $\|\nabla u^{j_k} - \nabla u^0\|_{L^p} \to 0$ (that is some sort of Bolzano-Weierstrass Theorem). Since we have not discussed derivatives yet we will, in this and in particular in the next section, consider whether every sequence u^j that is bounded in L^p norm has a convergent sub-sequence. The answer is subtle and we will have to define the concept of weak convergence in order to answer the question affirmatively.

Let us begin by remind ourself about the theorem in $\mathbb{R}^n$.

Theorem 7.1. [BOLZANO-WEIERSTRASS THEOREM IN $\mathbb{R}^n$] *Let $K \subset \mathbb{R}^n$ be a closed and bounded set and $\{\boldsymbol{x}^j\}_{j=1}^\infty$ be a sequence such that $\boldsymbol{x}^j \in K$. Then there exists a sub-sequence $\{\boldsymbol{x}^{j_k}\}_{k=1}^\infty$ of $\{\boldsymbol{x}^j\}_{j=1}^\infty$ such that $\lim_{k\to\infty} \boldsymbol{x}^{j_k}$ exists.*

Sketch of the Proof: Let $\mathbf{x}^j = (x_1^j, x_2^j, ..., x_n^j)$ then by the Bolzano-Weierstrass Theorem in $\mathbb{R}$ we may find a sub-sequence, which we may denote $\{x_1^{j_{k,1}}\}_{k=1}^\infty$, of $\{x_1^j\}_{j=1}^\infty$ such that $\{x_1^{j_{k,1}}\}_{k=1}^\infty$ converges.

Again by the one dimensional Bolzano-Weierstrass Theorem we may find a subsequence of $j_{k,1}$, lets denote it $j_{k,2}$, such that $\{x_2^{j_{k,2}}\}_{k=1}^\infty$ converges. We may then find a subsequence of $j_{k,2}$, lets denote it $j_{k,3}$, such that $\{x_3^{j_{k,3}}\}_{k=1}^\infty$ converges et.c.

In the end we find a sequence $j_k = j_{k,n}$ such that $\{x_n^{j_k}\}_{k=1}^\infty$ converges. But since j_k, by construction, is a sub-sequence of each sequence $\{j_{k,l}\}_{k=1}^\infty$, $l = 1, 2, ..., n$ it follows that

$$\lim_{k\to\infty} x_l^{j_k} = x_l^0 \quad \text{for} \quad l = 1, 2, ..., n.$$

Next we notice that, by the triangle inequality,

$$\lim_{k\to\infty} |\mathbf{x}^{j_k} - \mathbf{x}^0| \leq \lim_{k\to\infty} \sum_{l=1}^n |x_l^{j_k} - x_l^0| = 0$$

This finishes the proof. □

If we want to emulate this argument[16] for L^p we need to find a concept that corresponds to the coordinates in $\mathbb{R}^n$.

It is not absolutely clear how to do this. Let us consider the coordinate x_i of $\mathbf{x} = (x_1, x_2, ..., x_n)$ as the value of the linear function $\mathbf{e}_i : \mathbb{R}^n \mapsto \mathbb{R}$ defined by $\mathbf{e}_i\mathbf{x} = x_i$, or expressed in familiar notation $\mathbf{e}_i \cdot \mathbf{x} = x_i$ where $\mathbf{e}_1 = (1, 0, ..., 0)$, $e_2 = (0, 1, 0, ..., 0)$ et.c. Then we may consider the coordinates of $\mathbf{x}$ as the values of linear functions acting on $\mathbf{x}$. We will use this view of coordinates when we consider the L^p spaces.

Definition 7.1. *We say that $T : L^p(\mathcal{D}) \mapsto \mathbb{R}$ is a linear functional on $L^p(\mathcal{D})$ if*

$$T(af + bg) = aT(f) + bT(g) \text{ for all } f, g \in L^p(\mathcal{D}) \text{ and } a, b \in \mathbb{R}.$$

[16] It might be in place to warn you here that we will not be able to prove that bounded sequences has convergent subsequences in L^p-spaces.

We are in particular interested in **bounded functionals**.

Definition 7.2. *We say that a linear functional on $L^p(\mathcal{D})$ is bounded if there exists a constant $M \geq 0$ such that*

$$|T(u)| \leq M\|u\|_{L^p(\mathcal{D})} \text{ for all } u \in L^p(\mathcal{D}).$$

We will denote the least such constant M by $\|T\|_{(L^p(\mathcal{D}))^*}$.

The reason for our interest in bounded functionals is that $\|T\|_{(L^p(\mathcal{D}))^*}$ defines a norm on the set of bounded linear functionals. We will, as already indicated, denote the set of all bounded linear functionals on $L^p(\mathcal{D})$ by $(L^p(\mathcal{D}))^*$.

Lemma 7.1. $\|T\|_{(L^p(\mathcal{D}))^*}$ *defines a norm on* $(L^p(\mathcal{D}))^*$.

Proof: We leave the proof as an exercise (see exercise 7.1). □

In order to get the idea of using functionals $T \in (L^p(\mathcal{D}))^*$ as coordinates (that is, we hope that $u \in L^p(\mathcal{D})$ should be characterized by the numbers $T(u)$ for $T \in (L^p(\mathcal{D}))^*$) we would like to better understand what a bounded linear functional on $L^p(\mathcal{D})$ is. For this we need the following Theorem.

Theorem 7.2. *Let $1 \leq p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then $(L^p(\mathcal{D}))^* = L^q(\mathcal{D})$ in the following sense:*

1. *For each $T \in (L^p(\mathcal{D}))^*$ there exists a function $g \in L^q(\mathcal{D})$ such that*

$$Tf = \int_{\mathcal{D}} f(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}. \tag{45}$$

2. *Each $g \in L^q(\mathcal{D})$ defines an element $T \in (L^p(\mathcal{D}))^*$ defined by (45).*

Furthermore, if T and g are as in (45) then

$$\|T\|_{(L^p(\mathcal{D}))^*} = \|g\|_{L^q(\mathcal{D})}.$$

The proof is rather long and involves many ideas see (and read) for instance

W. Rudin, Real and complex analysis: pp. 76-82 and pp 116-129

Recommended exercises: *Chapter 4: 1,5,6; Chapter 6: 4,9*

H.L. Royden P.M Fitzpatrick, Real Analysis: Chapter 8.1 (1d case) and chapter 19.2 (general case).

Recommended exercises: *Chapter 8: 3, 5; Chapter 19.2: 7, 8*

E. Stein R. Shakarchi, Real Analysis Not covered

8 Lecture 8: Weak compactness.

From the last lecture we have a complete and compact description of what a bounded linear functional on $L^p(\mathcal{D})$ is. But the question remains if we can use bounded linear functionals as "coordinates". For a given $u \in L^p(\mathcal{D})$ will the values Tu, for $T \in (L^p(\mathcal{D}))^*$, uniquely determine u? We need the following simple lemma.

Lemma 8.1. *Let $u, v \in L^p(\mathcal{D})$ and $u \neq v$ then there exists a functional $T \in (L^p(\mathcal{D}))^*$ such that $Tu \neq Tv$. In particular if $Tu = Tv$ for all $T \in (L^p(\mathcal{D}))^*$ then $u = v$.*

Proof: By linearity it is enough to find a $T \in (L^p(\mathcal{D}))^*$ such that $T(u-v) = Tu - Tv \neq 0$. If we denote $w = u - v \neq 0$ it is enough to find a functional $T \in (L^p(\mathcal{D}))^*$ such that $Tw \neq 0$. Let $g = \text{sgn}(w)|w|^{\frac{p}{q}}$, where $\frac{1}{p} + \frac{1}{q} = 1$ and

$$\text{sgn}(w) = \begin{cases} 1 & \text{if } w > 0 \\ 0 & \text{if } w = 0 \\ -1 & \text{if } w < 0. \end{cases}$$

Then $g \in L^q(\mathcal{D})$ and we may define the functional $T \in (L^p(\mathcal{D}))^*$ as in (45). This leads to

$$Tw = \int_{\mathcal{D}} w(\mathbf{x})g(\mathbf{x})d\mathbf{x} = \int_{\mathcal{D}} |w|^{\frac{p}{q}+1}d\mathbf{x} = \int_{\mathcal{D}} |w|^p d\mathbf{x} = \|w\|^p_{L^p(\mathcal{D})} \neq 0,$$

where we used that $\frac{p}{q} + 1 = p$ and the definition of $\|w\|^p_{L^p(\mathcal{D})}$ as well as the assumption that $w \neq 0$. This finishes the proof. $\square$

We are now in a good position to replicate the proof of Theorem 7.1, but we are not quite there yet. There are still two problems that we need to consider.

First, if we have a bounded sequence $u^j \in L^p(\mathcal{D})$ then, for each $T \in (L^p(\mathcal{D}))^*$ we may find a subsequence u^{j_k} such that the numbers Tu^{j_k} converge. We may therefore for every, countable, set $\{T_1, T_2, T_3, ...\} \subset (L^p(\mathcal{D}))^*$ find a sub-sequence u^{j_l} of u^j such that $\lim_{l\to\infty} T_m u^{j_l}$ exists (for details, see the proof below). But there are clearly uncountable many functionals in $(L^p(\mathcal{D}))^*$.

Second, even if we manage to make sure that Tu^{j_l} converges for all $T \in (L^p(\mathcal{D}))^*$ we need to be able to identify a limit $u^0 \in L^p(\mathcal{D})$ from the values $\lim_{l\to\infty} Tu^{j_l}$ - it is not obvious that the values of the limits gives one function $u^0 \in L^p(\mathcal{D})$ such that $\lim_{l\to\infty} Tu^{j_l} = Tu^0$.

We begin to address the first problem. Even though $(L^p(\mathcal{D}))^*$, or equivalently by Theorem 7.2 $L^q(\mathcal{D})$, is an uncountable set we may approximate each function $g \in L^q(\mathcal{D})$ by a function from a countable set.

Definition 8.1. *We say that a Banach space (for instance $L^p(\mathcal{D})$) is separable if it contains a dense and countable set.*

Proposition 8.1. *$L^p(\mathcal{D})$ is separable.*

Proof: The proof is rather long; it is also a rather standard proof in mathematical analysis. We need to approximate an arbitrary function $u \in L^p(\mathcal{D})$ by a function f contained in a countable set $\mathcal{A} \subset L^p(\mathcal{D})$ within an $\epsilon > 0$ error. The construction is straightforward we begin by approximate u by a simple function, $s(\mathbf{x}) = \sum_{k=1}^M a_k \chi_{A_k}$, with an arbitrary precision, then we approximate the simple function by a simple function $s_r(\mathbf{x}) = \sum_{k=1}^M \hat{a}_k \chi_{A_k}$ where $s_r(\mathbf{x})$ only attains rational values (that is values in a countable set). Then we approximate the measurable sets $A_k = \{s_r(\mathbf{x}) = a_k \in \mathbb{Q}\}$ by measurable sets, $\hat{A}_k$, constructed by the finite union of cubes with rational coordinates (such cubes forms a countable set); this gives a new simple function $\sum_{k=1}^M \hat{a}_k \chi_{\hat{A}_k}$. This last simple function will be contained in a countable set $\mathcal{A} \subset L^p(\mathcal{D})$. Let us sketch some details (since the details will also remind us of the constructions of measures - and also show how flexible the construction is).

Step 1: *For any* $\epsilon_1 > 0$ *we may approximate* $u \in L^p(\mathcal{D})$ *by a simple function* $s(\mathbf{x}) = \sum_{k=1}^{M} a_k \chi_{A_k}$ *so that* $\|u - s\|_{L^p(\mathcal{D})} < \epsilon_1$.

Proof of Step 1: Remember that the continuous functions with compact support are dense in $L^p(\mathcal{D})$. It is therefore enough to show that for every continuous function $v \in C_c(\mathcal{D})$ may find a simple function s such that $\|v - s\|_{L^p(\mathcal{D})} < \epsilon_1/2$. Indeed, then by the triangle inequality we have

$$\|u - s\|_{l^p(\mathcal{D})} \leq \|u - v\|_{l^p(\mathcal{D})} + \|v - s\|_{l^p(\mathcal{D})},$$

and by choosing $v \in C_c(\mathcal{D})$ such that $\|u - v\|_{L^p} < \epsilon_1/2$ it follows that $\|u - s\|_{l^p(\mathcal{D})} < \epsilon_1$.

If $v \in C_c(\mathcal{D})$ then v is bounded, say $|v| \leq S$. We define the sets

$$A_k = \{\mathbf{x} \in \mathcal{D}; \delta(k-1) < v(\mathbf{x}) \leq k\delta\}, \text{ for } k \in \mathbb{Z} \text{ and } |k| \leq S/\delta,$$

where $\delta > 0$ will be chosen later. Moreover we set $a_k = \delta k$.

Then $0 \leq v(\mathbf{x}) - s(\mathbf{x}) \leq \delta$, where $s(\mathbf{x}) = \sum_{k=1}^{M} a_k \chi_{A_k}$. Therefore

$$\int_{\mathcal{D}} |v(\mathbf{x}) - s(\mathbf{x})|^p d\mathbf{x} \leq \delta^p |\mathrm{spt}(v)| < \left(\frac{\epsilon_1}{2}\right)^p$$

if $\delta < \epsilon_1/2|\mathrm{spt}(v)|^{1/p}$. This proves step 1.

Step 2: *For any* $\epsilon_1 > 0$ *we may approximate* $s(\mathbf{x})$ *(from step 1) by a function* $s_r(\mathbf{x})$ *that only attains rational values and* $\|s - s_r\|_{L^p(\mathcal{D})} < \epsilon$

Proof of Step 2: This is just a matter of choosing $a_k \in \mathbb{Q} \cap \big((k-1)\delta, k\delta\big]$ in the previous step; or we could just choose $\delta \in \mathbb{Q}$.

Step 3: *For any* $\epsilon_2 > 0$ *we may approximate any bounded and measurable set* A_k *by a set* $\hat{A}_k$ *such that* $\hat{A}_k = \sum_{j=1}^{S} Q^j$*, where* Q^j *are cubes with rational coordinates, and* $m(A_k \Delta \hat{A}_k) < \epsilon_2$*. Here* m *is the Lebesgue measure and* $A_k \Delta \hat{A}_k = \{\mathbf{x};\ \mathbf{x} \in A_k \setminus \hat{A}_k$ *or* $\mathbf{x} \in \hat{A}_k \setminus A_k\}$ *is the symmetric difference between* A_k *and* $\hat{A}_k$*. In particular, for any* $\epsilon > 0$ *we may approximate* $s_r(\mathbf{x})$ *by a function*

$$\hat{s}(\mathbf{x}) = \sum_{k=1}^{M} \hat{a}_k \chi_{\hat{A}_k}, \tag{46}$$

where $\hat{a}_k \in \mathbb{Q}$ *and* $\hat{A}_k$ *is the union of a finite set of cubes with rational coordinates and for any* $\epsilon_3 > 0$

$$\|s_r - \hat{s}\|_{L^p(\mathcal{D})} < \epsilon_3. \tag{47}$$

Proof of Step 3: Since the Lebesgue measure is regular there exists a compact set K and an open set U such that $K \subset A_k \subset U$ and $m(U \setminus K) < \epsilon$. Since K is closed and U open the distance $d = \mathrm{dist}(K, U^c) > 0$ is well defined and we can choose $\delta > 0$ to be a rational number such that $\delta < d/2\sqrt{n}$.

Next we let Q^k be all the cubes with side length δ and having the corners in points whose coordinates are integer multiples of δ. Next we cover K by the cubes Q^j such that $K \cup \overline{Q^j} \neq \emptyset$. Since K is compact we only need finitely many of the cubes $\overline{Q^j}$ to cover K, by relabeling the cubes we may assume that $\overline{Q^1}, \overline{Q^2}, ..., \overline{Q^S}$ covers K.

Define $\hat{A}_k = \cup_{j=1}^S Q^j$ then

$$K \subset \cup_{j=1}^S \overline{Q^j}$$

and, since $\mathrm{dist}(K, U^c) > 2\sqrt{n}\delta$, it follows that $\cup_{j=1}^S \overline{Q}^j \subset U$. We may conclude that

$$A_k \Delta \overline{\hat{A}_k} \subset U \setminus K.$$

In particular,

$$m(A_k \Delta \overline{\hat{A}_k}) \leq m(U \setminus K) < \epsilon_2.$$

Next we observe that

$$\overline{A_k} = \cup_{j=1}^S \overline{Q^j} = \left(\cup_{j=1}^S Q^j\right) \bigcup \left(\cup_{j=1}^S \overline{Q^j} \setminus Q^j\right),$$

it follows that

$$m(A_k \Delta \hat{A}_k) \leq m(A_k \Delta \overline{\hat{A}_k}) + m\left(\cup_{j=1}^S \overline{Q^j} \setminus Q^j\right) < \epsilon_2,$$

where we used that $\overline{Q^j} \setminus Q^j = \partial Q^j$ is a null-set in the last inequality.

It remains to prove (47). To that end we chose $\epsilon_2 < \frac{\epsilon_3^p}{S(2\sup(s_r))^p}$ then

$$\int_D |s_r - \hat{s}|^p d\mathbf{x} = \int_{(\cup A_k)\Delta(\cup \hat{A}_k)} |s_r - \hat{s}|^p \leq$$

$$\leq (2\sup(s_r))^p m((\cup A_k)\Delta(\cup \hat{A}_k)) \leq (2\sup(s_r))^p S\epsilon_2 < \epsilon_3^p,$$

taking pth roots on both sides gives the desired estimate.

Step 4: *The set*

$$\mathcal{A} = \left\{ \sum_{k=1}^M \hat{a}_k \chi_{\hat{A}_k};\ \hat{a}_k \in \mathbb{Q},\ \hat{A}_k \text{ as in step 3},\ M \in \mathbb{N} \right\}$$

is a countable set.

Proof of Step 4: Let us denote

$$\mathcal{A}^S = \left\{ \sum_{k=1}^M \hat{a}_k \chi_{\hat{A}_k};\ \hat{a}_k \in \mathbb{Q},\ \hat{A}_k \text{ as in step 3},\ M \leq S \right\},$$

for $S \in \mathbb{N}$. Clearly $\mathcal{A}^S$ is countable for any S; indeed every cube is determined by its center and side-length and therefore by a vector in $\mathbb{Q}^{n+1}$ and therefore $\mathcal{A}^S$ does not have more elements than $\mathbb{Q}^{(n+2)S}$.[17]

But $\mathcal{A} = \cup_{S=1}^\infty \mathcal{A}^S$ and countable unions of countable sets are countable.

Step 5: *Finnish up the proof.*

Proof of Step 5: To finish the proof we use the triangle inequality to estimate

$$\|u - \hat{s}\|_{L^p(\mathcal{D})} = \left\|u - s_r + s_r - \hat{s}\right\|_{L^p(\mathcal{D})} \leq \|u - s_r\|_{L^p(\mathcal{D})} + \|s_r - \hat{s}\|_{L^p(\mathcal{D})} < \epsilon_1 + \epsilon_3,$$

for any $\epsilon_1, \epsilon_3 > 0$. Choosing $\epsilon_1 = \epsilon_3 = \epsilon/2$ implies that there for any $\epsilon > 0$ exists a $\hat{s}$ in the countable set $\mathcal{A}$ such that $\|u - \hat{s}\|_{L^p(\mathcal{D})} < \epsilon$. This finishes the proof. □

[17] The "2" in $(n+2)S$ comes form the fact that we also have to choose one $a_k \in \mathbb{Q}$ for each $k = 1, 2, ..., S$.

Proposition 8.2. *Assume that T_k is a sequence of functionals in $(L^p(\mathcal{D}))^*$, T_k is bounded by M ($\|T_k\|_{(L^p(\mathcal{D}))^*} \leq M$ for all k) and that $\lim_{k\to\infty} T_k u^j$ exists for all u^j where u^j is a dense sub-set of $L^p(\mathcal{D})$. Then $\lim_{k\to\infty} T_k u$ exists for all $u \in L^p(\mathcal{D})$.*

Proof: We need to define Tu for an arbitrary $u \in L^p(\mathcal{D})$. Since u^j is dense in $L^p(\mathcal{D})$ we may find an u^j such that $\|u - u^j\|_{L^p} < \epsilon$, where $\epsilon > 0$ is arbitrary. We want to use this to show that $T_k u$ is a Cauchy sequence. To this end we fix an $\epsilon > 0$ and a j such that $\|u - u^j\|_{L^p} < \epsilon$.

Since $T_k u^j$ converges as $k \to \infty$ it follows that $T_k u^j$ is a Cauchy sequence in k. Thus there exists a $K > 0$ such that $|T_k u^j - T_l u^j| < \epsilon$ for all $k, l > K$. But this implies that

$$|T_k u - T_l u| = |T_k(u - u^j + u^j) - T_l(u - u^j + u^j)| \leq$$

$$\leq \underbrace{|T_k u^j - T_l u^j|}_{<\epsilon} + \underbrace{|T_k(u - u^j)|}_{\leq M\|u-u^j\|_{L^p}} + \underbrace{|T_l(u - u^j)|}_{\leq M\|u-u^j\|_{L^p}} < \epsilon + 2M\|u - u^j\|_{L^p} =$$

$$= (2M + 1)\epsilon,$$

for all $k, l > K$. Since $\epsilon > 0$ is arbitrary it follows that $T_k u$ is a Cauchy sequence and thus convergent. □

Since $L^q(\mathcal{D}) = (L^p(\mathcal{D}))^*$ is separable there exists a countable dense subset $\{T_1, T_2, ...\} \subset (L^p(\mathcal{D}))^*$. This implies that for any bounded sequence $u^j \in L^p(\mathcal{D})$ we can extract a subsequence $u^{j_{1,k}}$ such that $T_1 u^{j_{1,k}}$ converges, and a subsequence $u^{j_{2,k}}$ of $u^{j_{1,k}}$ such that both $T_1 u^{j_{2,k}}$ and $T_2 u^{j_{2,k}}$ converges et.c. This implies that the diagonal sequence $u^{j_l} = u^{j_{l,l}}$ will be such that $T_k u^{j_l}$ converges for all k and by Proposition 8.2 Tu^{j_l} will converge for all $T \in (L^p(\mathcal{D}))^*$.

However, this does not allow us to identify the limit $\lim_{l\to\infty} u^{j_l}$ directly. It only gives a strong indication that if u^{j_l} converges, in some sense, to u^0 as $l \to \infty$ then the value Tu^0, for $T \in (L^p(\mathcal{D}))^*$, should be the unique value of the limit $\lim_{l\to\infty} Tu^{j_l}$.

In order to identify u^0 we need another beautiful trick. By Theorem 7.2, for every $T \in (L^p(\mathcal{D}))^*$ there exists a function $g \in L^q(\mathcal{D})$ such that (45) holds. If we identify g with T then we have a unique value $\lim_{l\to\infty} Tu^{j_l}$ for each $g \in L^q(\mathcal{D})$ we should therefore be able to define the functional on $L^q(\mathcal{D})$

$$g \mapsto \lim_{l\to\infty} Tu^{j_l} = \lim_{l\to\infty} \int_{\mathcal{D}} g(\mathbf{x}) u^{j_l}(\mathbf{x}) d\mathbf{x}. \tag{48}$$

But **if** this functional is linear and bounded then this functional, again by Theorem 7.2, should correspond to a unique element $u^0 \in L^p(\mathcal{D})$. We may therefore identify the limit $\lim_{l\to\infty} u^{j_l}$ with u^0 in the sense that

$$\lim_{l\to\infty} \int_{\mathcal{D}} g(\mathbf{x}) u^{j_l}(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{D}} g(\mathbf{x}) u^0(\mathbf{x}) d\mathbf{x} \tag{49}$$

for every $g \in L^q(\mathcal{D})$. We still need to show that the above argument works. First of all we need to show that the mapping defined by (48) is indeed linear and bounded. But before we do that we will define the new concept of convergence introduced in (49).

Definition 8.2. *We say that a sequence* $u^j \in L^p(\mathcal{D})$, $1 \leq p < \infty$, *converges weakly to* u^0 *if*

$$\int_{\mathcal{D}} g(\boldsymbol{x})u^j(\boldsymbol{x})d\boldsymbol{x} \to \int_{\mathcal{D}} g(\boldsymbol{x})u^0(\boldsymbol{x})d\boldsymbol{x}$$

for all $g \in L^q(\mathcal{D})$ *and* $\frac{1}{p} + \frac{1}{q} = 1$. *If* u^j *converges to* u^0 *weakly we write* $u^j \rightharpoonup u^0$.

We are now ready to prove that if we view $u^j \in L^p(\mathcal{D})$ as a sequence of functionals on $L^q(\mathcal{D})$ then the mapping defined in (48) is indeed linear and bounded.

Lemma 8.2. *Let* $S_j \in (L^q(\mathcal{D}))$ *be a sequence of linear functionals such that* $|S_j g| \leq M\|g\|_{L^q(\mathcal{D})}$ *for all* $g \in L^q(\mathcal{D})$, *or equivalently* $u^j \in L^p(\mathcal{D})$ *be sequence of functions* $\|u^j\|_{L^p(\mathcal{D})} \leq M$, *such that* $\lim_{j\to\infty} S_j g$ *exists for every* $g \in L^q(\mathcal{D})$. *Then the functional* $S \in (L^q(\mathcal{D}))^*$

$$Sg = \lim_{j\to\infty} S_j g$$

is linear and bounded.

Proof: To show that S is linear we need to show that

$$S(af + bg) = \lim_{j\to\infty} \int_{\mathcal{D}} (af(\mathbf{x}) + bg(\mathbf{x}))u^j(\mathbf{x})d\mathbf{x} =$$

$$= \lim_{j\to\infty} \left(a\int_{\mathcal{D}} f(\mathbf{x})d\mathbf{x} + b\int_{\mathcal{D}} g(\mathbf{x}))u^j(\mathbf{x})d\mathbf{x} \right) = aSf + bSg,$$

by the linearity of the integral.

That S is bounded follows from continuity of the absolute value:

$$|Sg| = \lim_{j\to\infty} |S_j g| \leq M\|g\|_{L^q(\mathcal{D})}.$$

□

We are now ready to prove the weak compactness theorem for L^p-spaces.

Theorem 8.1. *Let* $u^j \in L^p(\mathcal{D})$, $1 < p < \infty$, *be a bounded sequence of functions, that is* $\|u^j\|_{L^p(\mathcal{D})} \leq M$ *for* M *independent of* j. *Then* u^j *has a weakly convergent subsequence* $u^{j_l} \to u^0$.

Proof: We begin by identifying each u^j with a functional $S_j \in (L^q(\mathcal{D}))^*$. Then, since L^q is separable, we may find a countable and dense set $g_k \in L^q(\mathcal{D})$. By a diagonalization argument we may inductively define subsequences $\{u^{j_{l,k+1}}\}_{l=1}^{\infty}$ of the sequence $\{u^{j_{l,k}}\}_{l=1}^{\infty}$ such that

$$S_{j_{l,k+1}} g_{k+1} = \int_{\mathcal{D}} u^{j_{l,k+1}}(\mathbf{x}) g_{k+1}(\mathbf{x})d\mathbf{x}$$

converges. Then the sequences $S_{j_l} g_k = S_{j_{l,l}} g_k$ converges for every k, that is on a dense subset of $L^q(\mathcal{D})$. It follows, from Proposition 8.2, that $S_{j_l} g$ converges for all $g \in L^q(\mathcal{D})$. Furthermore the functional S defined by

$$Sg = \lim_{l\to\infty} S_{j_l} g \text{ for all } g \in L^q(\mathcal{D})$$

is linear and bounded. It follows from Theorem 7.2 that there exists a unique u^0 such that

$$\int_{\mathcal{D}} g(\mathbf{x})u^0(\mathbf{x})d\mathbf{x} = Sg = \lim_{l\to\infty}\int_{\mathcal{D}} g(\mathbf{x})u^{j_l}(\mathbf{x})d\mathbf{x}.$$

This proves the theorem. $\square$

This shows that we have some version of the Bolzano-Weierstrass Theorem for L^p-spaces: every bounded sequence u^j has a convergent sub-sequence *if we interpret convergence in the weak sense as in Definition 8.2.* But we need to understand how the weak convergence relates to convergence in norm: $\lim_{j\to\infty}\|u^j - u^0\|_{L^p} = 0$. Are the two concepts the same? Is one stronger than the other?

It is very easy to see that convergence in norm (at times called strong convergence) is a stronger concept than weak convergence.

Proposition 8.3. *Assume that $u^j \to u^0$ in norm: $\lim_{j\to\infty}\|u^j - u^0\|_{L^p(\mathcal{D})} = 0$. Then $u^j \rightharpoonup u^0$ weakly.*

Proof: For any bounded linear functional $T \in (L^p(\mathcal{D}))^*$ we have the following inequality

$$|Tu^j - Tu^0| = |T(u^j - u^0)| \le \|T\|_{(L^p(\mathcal{D}))^*}\|u^j - u^0\|_{L^p(\mathcal{D})}.$$

Therefore $|Tu^j - Tu^0| \to 0$ as $j \to \infty$ if $u^j \to u^0$ in norm. $\square$

So strong convergence (that is convergence in norm) implies weak convergence. Does the other implication hold?

Example (concentration phenomena): We claim that a weakly convergent sequence $u^j \in L^p(\mathcal{D})$ does not necessary converge strongly, we will assume that $1 < p < \infty$. In order to see this we assume that we have a weakly convergent sequence $u^j \rightharpoonup u^0$. By considering $u^j - u^0$ in place for u^j we may assume that $u^0 = 0$. We thus assume that we have a sequence $u^j \rightharpoonup 0$ and aim to show that this does not imply that $\|u^j\|_{L^p(\mathcal{D})} \to 0$. How can this happen? Well the norm of u^j has to be somewhat large, say $\|u^j\|_{L^p(\mathcal{D})} = 1$ at the same time as u^j tends to zero in some sense. Let us therefore try to find a function that converges point-wise to zero at the same time as the integral is large, this must mean that u^j is large on a small set.

To fix ideas, let us choose $\mathcal{D} = (0,1)$ and try to find a sequence of functions such that $u^j \to 0$ pointwise but $\|u^j\|_{L^p(0,1)} = 1$. Say that $u^j(x) = 0$ for $x > 1/j$ then $u^j(x) \to 0$ for all $x \in (0,1)$. Let us choose $u^j(x) = \textit{constant}$ for $x \le 1/j$ such that $\|u^j\|_{L^p(0,1)} = 1$. That is we define

$$u^j(x) = \begin{cases} j^{1/p} & \text{for } x \in (0, 1/j] \\ 0 & \text{for } x \in (1/j, 1). \end{cases}$$

Then $\|u^j\|_{L^p(0,1)} = 1$ and $u^j \to 0$ pointwise. That $u^j(x) \to 0$ pointwise strongly indicates that $u^j \rightharpoonup 0$, but we need to prove this.

Let us pick an arbitrary $g \in L^q(0,1)$ and show that $\int_0^1 g(x)u^j(x)dx \to 0$. To that end we choose an $\epsilon > 0$ and a simple function $s(x)$ such that $\|s-g\|_{L^q(0,1)} < \epsilon$. Since s only attains finitely many values there exists a constant C such that $|s(x)| \le C$. We can therefore calculate

$$\left|\int_0^1 g(x)u^j(x)dx\right| \le \left|\int_0^1 (g(x)-s(x))u^j(x)dx\right| + \left|\int_0^1 s(x)u^j(x)dx\right| \le$$

$$\leq \|g-s\|_{L^q}\|u^j\|_{L^p} + \left|\int_0^{1/j} s(x)j^{1/p}dx\right| < \epsilon + \frac{C}{j^{p-1}p} < 2\epsilon$$

if $j > \left(\frac{C}{\epsilon}\right)^{p/p-1}$. It follows that $\int_0^1 g(x)u^j(x)dx \to 0$.

We have thus shown that weak convergence does not imply strong convergence in general. In this example the failure of the strong convergence is due to concentration - that there exists a set U of arbitrarily small measure such that $\lim_{j\to\infty}\|u^j\|_{L^p(U)} \not\to 0$.

Example (oscillation phenomena): There is another way that weak compactness may fail to be strong, if the sequence u^j shows infinite oscillations. Again we illustrate the phenomena with a one dimensional example. Let $u^j = \sin(jx) \in L^p((0,\pi))$ for $1 < p < \infty$. We claim that $u^j \rightharpoonup 0$. We need to show that for an arbitrary $g \in L^q(0,\pi)$ and we want to show that $\int_0^\pi g(x)u^j(x)dx \to 0$. To see this we approximate g by a continuous function $g_\epsilon \in C([0,\pi])$ such that $\|g-g_\epsilon\|_{L^q} < \epsilon$ then

$$\left|\int_0^\pi g(x)u^j(x)dx\right| \leq \left|\int_0^\pi (g(x)-g_\epsilon(x))u^j(x)dx\right| + \left|\int_0^\pi g_\epsilon(x)u^j(x)dx\right| \leq$$

$$\leq \|g-g_\epsilon\|_{L^q(0,\pi)}\|u^j\|_{L^p(0,\pi)} + \left|\int_0^\pi g_\epsilon(x)u^j(x)dx\right| \leq (\pi)^{1/p}\epsilon + \left|\int_0^\pi g_\epsilon(x)u^j(x)dx\right|, \tag{50}$$

where we used that $|u^j| \leq 1$. To estimate the last integral in (50) we split the interval $(0,\pi)$ into $N = \pi/\delta$ where δ is chosen so small that $\sup_{x\in[k\delta,(k+1)\delta]}|g_\epsilon(x) - g\epsilon(k\delta)| < \epsilon$, this is possible since g is uniformly continuous on the compact interval $[0,\pi]$. We may therefore estimate the integral

$$\left|\int_0^\pi g_\epsilon(x)u^j(x)dx\right| = \left|\sum_{k=0}^{N-1}\int_{k\delta}^{(k+1)\delta} g_\epsilon(k\delta)u^j(x)dx\right| + \pi\epsilon = \tag{51}$$

$$= \left|\sum_{k=0}^{N-1} g_\epsilon(k\delta)\frac{\cos(jk\delta)-\cos(j(k+1)\delta)}{j}\right| + \pi\epsilon \leq \frac{2}{j}\sup_{x\in[0,\pi]}|g_\epsilon| + \pi\epsilon,$$

where we used that $u^j(x) = \sin(jx)$ in order to evaluate the integral. Using the estimate (51) in (50) we may derive that

$$\lim_{j\to\infty}\left|\int_0^\pi g(x)u^j(x)dx\right| \leq \left((\pi)^{1/p}+\pi\right)\epsilon.$$

Since $\epsilon > 0$ is arbitrary it follows that $\lim_{j\to\infty}\left|\int_0^\pi g(x)u^j(x)dx\right| = 0$ and $u^j \rightharpoonup 0$. However, $\|u^j\|_{L^q(0,\pi)} \not\to 0$ and therefore $u^j \not\to 0$ in the strong sense.

The two examples above shows that in general weak convergence is not the same as strong convergence. The examples also shows that any strong form of the Bolzano-Weierstrass Theorem is not true in L^p. If we could find a subsequence u^{j_k} of u^j such that $u^{j_k} \to u^0$ with u^j as in either of the examples then by Proposition 8.3 it would follow that $u^{j_k} \rightharpoonup u^0$ and since the weak limit is unique (see exercise 6 below) it would follow that $u^0 = 0$ but this would contradict $u^{j_k} \to u^0$.

It is natural to ask how the weak convergence relates to the strong convergence. We will only prove the following easy proposition.

Proposition 8.4. *Assume that* $u^j \rightharpoonup u^0$ *in* $L^p(\mathcal{D})$, $1 < p < \infty$, *and that* $\|u^j\|_{L^p(\mathcal{D})} \le M$. *Then*

$$\|u^0\|_{L^p(\mathcal{D})} \le \liminf_{j\to\infty} \|u^j\|_{L^p(\mathcal{D})}.$$

Proof: Let u^{j_k} be a sub-sequence such that $\lim_{k\to\infty} \|u^{j_k}\|_{L^p(\mathcal{D})} = \liminf_{j\to\infty} \|u^j\|_{L^p(\mathcal{D})}$ and choose $g(\mathbf{x}) = \mathrm{sgn}(u^0(\mathbf{x}))|u^0(\mathbf{x})|^{p/q}$, for $\frac{1}{p} + \frac{1}{q} = 1$. Then by Hölder's inequality

$$\lim_{k\to\infty} \int_{\mathcal{D}} g(\mathbf{x})u^{j_k}(\mathbf{x})d\mathbf{x} \le \lim_{k\to\infty} \|g\|_{L^q}\|u^{j_k}\|_{L^p} = \lim_{k\to\infty} \|u^0\|_{L^p}^{p/q}\|u^{j_k}\|_{L^p} \tag{52}$$

and by weak convergence

$$\lim_{k\to\infty} \int_{\mathcal{D}} g(\mathbf{x})u^{j_k}(\mathbf{x})d\mathbf{x} = \int_{\mathcal{D}} g(\mathbf{x})u^0(\mathbf{x})d\mathbf{x} = \int_{\mathcal{D}} |u^0(\mathbf{x})|^p d\mathbf{x} = \|u^0\|_{L^p(\mathcal{D})}^p. \tag{53}$$

From (52) and (53) we can conclude that

$$\|u^0\|_{L^p(\mathcal{D})}^p \le \lim_{k\to\infty} \|u^0\|_{L^p(\mathcal{D})}^{p/q}\|u^{j_k}\|_{L^p(\mathcal{D})} = \|u^0\|_{L^p(\mathcal{D})}^{p/q} \lim_{k\to\infty} \|u^{j_k}\|_{L^p(\mathcal{D})},$$

dividing both sides by $\|u^0\|_{L^q}^{p/q}$ and using that $p - p/q = p(1-1/q) = 1$

$$\|u^0\|_{L^p(\mathcal{D})} \le \lim_{k\to\infty} \|u^{j_k}\|_{L^p(\mathcal{D})}.$$

□

Summary of the calculus of variation problem. Let us take the opportunity to remind ourselves what we are trying to achieve here. We are aiming to show that there exist minimizers to the energy $E(u) = \int_{\mathcal{D}} |\nabla u|^2 d\mathbf{x}$. So far we have made significant progress. If we consider a sequence u^j such that $E(u^j) \to \inf_{\mathcal{K}_E} E(u^j) = m$ then $E(u^j) = \|\nabla u^j\|_{L^2(\mathcal{D})}^2$ is bounded and thus we may extract a weakly convergent subsequence $\nabla u^{j_k} \to \nabla u^0$ and by Proposition 8.4 it follows that $\|\nabla u^0\|_{L^2(\mathcal{D})} \le \liminf_{k\to\infty} \|\nabla u^j\|_{L^2(\mathcal{D})}^2 = m$ thus $E(u^0) \le m$. But since m is the infimum of all possible energies we expect $E(u^0) \ge m$ - but that is of course dependent on whether $u^0 \in \mathcal{K}_E$ or not. We have however reached the integration theory needed for the solution of the minimization problem.

Next we are going to investigate the problem of derivatives. In particular we need to understand what it means for $\nabla u \in L^p(\mathcal{D})$. This is actually a rather subtle thing that will take several lectures to investigate.

8.1 Exercises:

1. Give an example of a function $f(x) : [0,1] \mapsto \mathbb{R}$ that is point-wise bounded but not bounded.

2. Prove Lemma 7.1.

3. Does weak compactness hold in $L^1([0,1])$?

   Hint: *Try to find a sequence* $u^j \in L^1$ *such that the support of* u^j *converges to one point.*

4. What goes wrong in our proof of weak compactness if we try to apply it to L^∞?

5. Prove that if $u^j \rightharpoonup u^0$ in $L^p(\mathcal{D})$ then u^j converges to u^0 in average; that is for each bounded measurable set $\Sigma \subset \mathcal{D}$
$$\int_\Sigma u^j(\mathbf{x})d\mathbf{x} \to \int_\Sigma u^0(\mathbf{x})d\mathbf{x}.$$
Prove that also the converse holds: if u^j converges to u^0 in average then $u^j \rightharpoonup u^0$.

6. Prove that if $u^j \rightharpoonup v$ and $u^j \rightharpoonup w$ in $L^p(\mathcal{D})$ then $v = w$.

7. Assume that $u^j(\mathbf{x}) \to u^0(\mathbf{x})$ a.e. and that $\|u^j\|_{L^p(\mathcal{D})} < M$. Show that $u^j \to u^0$ in $L^q(\mathcal{D})$ for every $q < p$.

9 Lecture 9. Differentiation. Distributional derivatives. $W^{k,p}$ spaces.

We need to interpret what it means for a function $u(\mathbf{x})$ to have derivatives $\nabla u(\mathbf{x}) = (\partial_1 u(\mathbf{x}), \partial_2 u(\mathbf{x}), ..., \partial_n u(\mathbf{x})) \in L^p(\mathcal{D})$. The problem is rather subtle.

Clearly if $u(\mathbf{x}) \in C^1(\mathcal{D})$ then the gradient $\nabla u \in C(\mathcal{D})$ and therefore $|\nabla u|$ is bounded (at least if $\overline{\mathcal{D}}$ is compact) - we can therefore find many functions u such that $\nabla u \in L^p(\mathcal{D})$ (at least if $\overline{\mathcal{D}}$ is compact). However, we are interested in convergence of sequences of functions u^j such that $\nabla u^j \in L^p(\mathcal{D})$ and if we only consider C^1-functions we loose the completeness of the space under the L^p-norm.

Example 1: *Let* $u^\epsilon(x) = \frac{x^2}{|x|+\epsilon} \in C^1([-1,1])$, *for* $\epsilon > 0$, *be a family of functions defined on* $[-1,1]$. *Then* $u^\epsilon(x) \to |x|$ *and*
$$\frac{\partial u^\epsilon}{\partial x} = \frac{x|x| + 2\epsilon x}{(|x|+\epsilon)^2} \to \mathrm{sgn}(x) = \begin{cases} 1 & \text{if } x > 1 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0, \end{cases}$$
as $\epsilon \to 0^+$. *It is also easy to see that* $\|u^\epsilon - |x|\|_{L^p} + \|\partial u^\epsilon - sgn(x)\|_{L^p} \to 0$ *as* $\epsilon \to 0^+$. *Therefore* u^ϵ *should converge to* $|x|$ *and* ∂u^ϵ *to* $sgn(x)$ *as* $\epsilon \to 0^+$. *However, the function* $|x|$ *is not differentiable at* $x = 0$ *- and certainly not* C^1.

One might think that the example is stupid since if we consider ∂u^ϵ as a function in $L^p([-1,1])$ then the value of ∂u^ϵ on a null-set is irrelevant, in particular the value (or lack of value) at $x = 0$ should not matter. There is some truth to this criticism, but a more sophisticated example should convince you that assuming that u is differentiable a.e. is not enough in order to have a well defined derivative.

Example 2: [Devil's Staircase] *We define* $f_0(x) = x$ *on* $[0,1]$ *and inductively*
$$f_{k+1}(x) = \begin{cases} \frac{1}{2} f_k(3x) & \text{if } 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \text{if } \frac{1}{3} < x < \frac{2}{3} \\ \frac{1}{2} + \frac{1}{2} f_k(3(x-2/3)) & \text{if } \frac{2}{3} \le x \le 1. \end{cases}$$

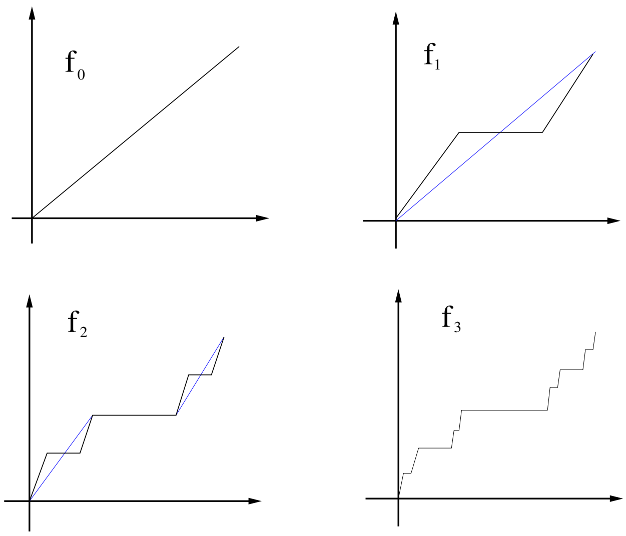


Figure 1: *The functions f_0, f_1, f_2 and f_3.*

Clearly the set where $f_k'(x) \neq 0$ will be of measure $\left(\frac{2}{3}\right)^k \to 0$. Also, by construction,

$$\sup_{x\in[0,1]} |f_{k+1}(x) - f_k(x)| = \frac{1}{2} \sup_{x\in[0,1]} |f_k(x) - f_{k-1}(x)|$$

which in particular implies that the sequence of continuous functions f_k converges uniformly to some function $f(x)$ which must also be continuous.[18] *Also $f'(x) = 0$ a.e. However, $f(x)$ is not constant.*

This is not a contradiction to anything and we have to live with the fact that functions may have zero derivatives a.e. without being constant. But the main reason that the derivative is an interesting concept is that the derivative relates to the change of a function. We would certainly want to be able to prove some version of the fundamental theorem of calculus, $\int_a^b f'(x)dx = f(b) - f(a)$, for differentiable functions. For that we need to find "the right" definition of derivative for functions that are not differentiable at every point.

In order to define the derivative of a function that might be discontinuous, in particular we want to be able to define derivatives in such way that it makes sense for them to belong to the space $L^p(\mathcal{D})$, we need to find some formula that captures the essence of the derivative without using the normal point-wise difference quotient. If $\nabla u \in L^p(\mathcal{D})$ then it makes sense to integrate ∇u - but not to talk about ∇u point-wise - therefore we need a good formula for the derivative that only uses integration and not point-wise properties. This leads us to define the derivative by means of the integration by parts formula.

Definition 9.1. *Let u be integrable in some domain $\mathcal{D} \subset \mathbb{R}^n$. Then if there exists an integrable function $w(x)$ such that*

$$\int_{\mathcal{D}} \frac{\partial v(x)}{\partial x_i} u(x)dx = -\int_{\mathcal{D}} v(x)w(x)dx \text{ for every } v(x) \in C_c^1(\mathcal{D}) \qquad (54)$$

[18] Remember that uniform convergence preserves continuity.

then we say that $u(x)$ is weakly differentiable in x_i and that $w(x)$ is the weak x_i-derivative of $u(x)$ and write $\frac{\partial u(x)}{\partial x_i} = w(x)$.

Remark: *Notice that the definition is made so as the partial integration formula works. In particular, if $u(x)$ is weakly differentiable in x_i then (54) become the normal integration by parts formula*

$$\int_{\mathcal{D}} \frac{\partial v(x)}{\partial x_i} u(x) dx = -\int_{\mathcal{D}} v(x) \frac{\partial u(x)}{\partial x_i} dx.$$

It follows directly that every continuously differentiable function is weakly differentiable.

We need to make sure that the definition of weak derivative preserves enough properties of the derivative in order to be a useful tool in analysis. But first we will define Sobolev spaces, spaces of functions whose weak gradient are L^p functions.

Definition 9.2. *Let $u(x) \in L^p(\mathcal{D})$ be weakly differentiable in every direction x_i, $i = 1, 2, ..., n$ and the weak derivatives $\frac{\partial u}{\partial x_i} \in L^p(\mathcal{D})$ for all $i = 1, 2, ..., n$. Then we say that $u \in W^{1,p}(\mathcal{D})$. We call the space of all such functions equipped with the norm*

$$\|u\|_{W^{1,p}(\mathcal{D})} = \left(\int_{\mathcal{D}} |u(x)|^p dx + \int_{\mathcal{D}} |\nabla u(x)|^p dx \right)^{1/p}, \tag{55}$$

where $\nabla u(x) = \left(\frac{\partial u(x)}{\partial x_1}, \frac{\partial u(x)}{\partial x_2}, ..., \frac{\partial u(x)}{\partial x_n} \right)$.

We will call the space $W^{1,p}(\mathcal{D})$ for a Sobolev space.

We will also write $W^{k,p}(\mathcal{D})$ for all functions defined on $\mathcal{D}$ such that all weak derivatives up to order k exists and

$$\|u\|_{W^{k,2}(\mathcal{D})} = \left(\sum_{|\alpha| \le k} \int_{\mathcal{D}} |D^\alpha u|^p \right)^{1/p} < \infty,$$

where the summation is over all mutiindexes α of length $|\alpha| \le k$.

The first thing we need to verify is that the space $W^{1,p}(\mathcal{D})$ is complete. That follows easily from the fact that $L^p(\mathcal{D})$ is complete.

Lemma 9.1. *The space $W^{1,p}(\mathcal{D})$ with norm (55) is a complete space.*

Furthermore every bounded sequence of functions $u^j \in W^{1,p}(\mathcal{D})$ has a subsequence u^{j_k} so that $u^{j_k} \rightharpoonup u^0$ and $\frac{\partial u^{j_k}}{\partial x_i} \rightharpoonup \frac{\partial u^0}{\partial x_i}$ in $L^p(\mathcal{D})$ for some $u^0 \in W^{1,p}(\mathcal{D})$.

Remark: *We say that the subsequence u^{j_k} converges weakly to u^0 in $W^{1,p}(\mathcal{D})$, written $u^{j_k} \rightharpoonup u^0$ in $W^{1,p}(\mathcal{D})$.*

Proof: That $W^{1,p}(\mathcal{D})$ is complete follows from the completeness of $L^p(\mathcal{D})$.

By Theorem 8.1 we can extract a subsequence such that u^{j_k} and $\frac{\partial u^{j_k}}{\partial x_i}$, for all $i = 1, 2, ..., n$, converges weakly. We need to show that the limit of the partial derivatives converges to the partial derivatives of the limit $u^{j_k} \rightharpoonup u^0$. That is easy. For any $\phi \in C^1(\mathcal{D})$ we have

$$\int_{\mathcal{D}} \frac{\partial \phi(x)}{\partial x_i} u^0(x) dx \leftarrow \int_{\mathcal{D}} \frac{\partial \phi(x)}{\partial x_i} u^{j_k}(x) dx = \tag{56}$$

$$-\int_{\mathcal{D}} \phi(x)\frac{\partial u^{j_k}(x)}{\partial x_i}dx \to -\int_{\mathcal{D}} \phi(x)\lim_{j_k\to\infty}\frac{\partial u^{j_k}(x)}{\partial x_i}dx,$$

Since (56) holds for every ϕ it follows that $\frac{\partial u^{j_k}(x)}{\partial x_i} \rightharpoonup \frac{\partial u^0(x)}{\partial x_i}$. This finishes the proof. □

We are mistaken if we try to evaluate a mathematical concept, such as weak derivatives, only in view of one result. The preceding lemma shows that weak derivatives, and in particular Sobolev spaces, have the good completeness properties that we desire. In order to convince ourselves that the definitions of weak derivatives and Sobolev spaces and weak derivatives are "right" we will have to work much harder. We would like to gain an understanding of how Sobolev functions relates to the classical functions, we would like to derive some version of the fundamental theorem of calculus and we would like to gain a better understanding of how the weak derivative relates to the classical derivative. Only after we have gained a thorough understanding of these issues can we be sure that the definition is right.

We begin by showing that the space $C_c^\infty(\mathcal{D})$ is dense in $W^{1,p}(\mathcal{D})$, this shows that the space $W^{1,p}(\mathcal{D})$ as defined is the only possible way of making C_c^∞ complete under the norm $\|\cdot\|_{W^{1,p}}$. The first step is to show that if $u \in W^{1,p}(\mathcal{D})$ then, for every compact set, we may approximate u by a $C^\infty-$function.

Proposition 9.1. *Let $\phi(\boldsymbol{x})$ be the standard mollifier:*

$$\phi(\boldsymbol{x}) = \begin{cases} ce^{\frac{1}{|\boldsymbol{x}|^2-1}} & \text{if } |\boldsymbol{x}| < 1 \\ 0 & \text{if } |\boldsymbol{x}| \geq 1, \end{cases}$$

where c is chosen so that $\int_{\mathbb{R}^n}\phi(\boldsymbol{x})d\boldsymbol{x} = 1$, also define $\phi_\epsilon(\boldsymbol{x}) = \frac{1}{\epsilon^n}\phi(\boldsymbol{x}/\epsilon)$.[19] *Furthermore let $u \in L^p(\mathcal{D})$, for some open set $\mathcal{D}$, and define*

$$u_\epsilon(\boldsymbol{x}) = \phi_\epsilon * u(\boldsymbol{x}) = \int_{\mathcal{D}} \phi_\epsilon(\boldsymbol{x}-\boldsymbol{y})u(\boldsymbol{y})d\boldsymbol{y},$$

on for $\boldsymbol{x} \in \mathcal{D}_\epsilon = \mathcal{D} \cap \{\boldsymbol{y} \in \mathcal{D};\ dist(\boldsymbol{y}, \mathcal{D}^c) > \epsilon\}$.

Then

1. *$u_\epsilon \in C^\infty(\mathcal{D}_\epsilon)$.*
2. *If $u \in C(\mathcal{D})$ then, for every $\boldsymbol{x} \in \mathcal{D}$, $u_\epsilon(\boldsymbol{x}) \to u(\boldsymbol{x})$ as $\epsilon \to 0^+$.*
3. *For $1 \leq p < \infty$ and any compact set $K \subset \mathcal{D}$ it follows that*
$$\lim_{\epsilon\to 0^+}\|u - u_\epsilon\|_{L^p(K)} = 0.$$
4. *If $u \in W^{1,p}(\mathcal{D})$ then*
$$\frac{\partial u_\epsilon}{\partial x_i} = \phi_\epsilon * \frac{\partial u}{\partial x_i}.$$
5. *for any compact set $K \subset \mathcal{D}$*
$$\lim_{\epsilon\to 0^+}\|u - u_\epsilon\|_{W^{1,p}(K)} = 0.$$

[19] Notice that $\text{spt}(\phi_\epsilon) = B_\epsilon(0)$ and that $\int_{\mathbb{R}^n}\phi_\epsilon(\mathbf{x})d\mathbf{x} = 1$ for all $\epsilon > 0$.

Proof: We first remark that $\phi_\epsilon(\mathbf{x}) \in C^\infty(\mathbb{R}^n)$ for any $\epsilon > 0$. That $\phi_\epsilon(\mathbf{x}) \in C^\infty(B_\epsilon(0))$ and that $\phi_\epsilon(\mathbf{x}) \in C^\infty(\mathbb{R}^n \setminus B_\epsilon(0))$ is obvious. In order to see that $\phi_\epsilon(\mathbf{x}) \in C^\infty(\mathbb{R}^n)$ we only need to verify that the derivatives are continuous across $\partial B_\epsilon(0)$. This follows from small calculation using that $\lim_{|\mathbf{x}|\to 1^-} \frac{e^{\frac{1}{|\mathbf{x}|^2-1}}}{(1-|\mathbf{x}|^2)^k} = 0$ for all $k \in \mathbb{N}$, we leave the details to the reader.

Let us prove the four statements in the proposition.

1. This follows more or less directly from the definition of the derivative. By the definition of u_ϵ, for $\mathbf{x} \in \mathcal{D}_\epsilon$, $\mathbf{v} \in \mathbb{R}^n$ and $h > 0$ small enough,

$$\frac{u_\epsilon(\mathbf{x}+h\mathbf{v}) - u_\epsilon(\mathbf{x})}{h} = \int_{\mathcal{D}_\epsilon} \left(\frac{\phi_\epsilon(\mathbf{x}+h\mathbf{v}-\mathbf{y}) - \phi_\epsilon(\mathbf{x}-\mathbf{y})}{h} \right) u(\mathbf{y}) d\mathbf{y}. \quad (57)$$

We want to pass to the limit $h \to 0^+$ in (57) for that we use the Dominated Convergence Theorem, which we may apply since

$$\left| \left(\frac{\phi_\epsilon(\mathbf{x}-\mathbf{y}) - \phi_\epsilon(\mathbf{x}+h\mathbf{v}-\mathbf{y})}{h} \right) u(\mathbf{y}) \right| \leq \|\nabla\phi_\epsilon \cdot \mathbf{v}\|_{L^\infty} |u| \in L^1.$$

In particular

$$\nabla u_\epsilon(\mathbf{x}) \cdot \mathbf{v} = \lim_{h\to 0^+} \frac{u_\epsilon(\mathbf{x}+h\mathbf{e}_i) - u_\epsilon(\mathbf{x})}{h} = \int_{\mathcal{D}_\epsilon} \left(\nabla\phi_\epsilon(\mathbf{x}-\mathbf{y}) \cdot \mathbf{v}\right) u(\mathbf{y}) d\mathbf{y}.$$

It follows that u_ϵ is differentiable. One may argue similarly to show that higher differentials exist. It follows that $u_\epsilon \in C^\infty$.

2. Fix any $\delta > 0$ then, since u is continuous, there exists an $\epsilon_\delta > 0$ such that $\sup_{\mathbf{y}\in B_{\epsilon_\delta}(\mathbf{x})} |u(\mathbf{x}) - u(\mathbf{y})| < \delta$. This implies that, for any $0 < \epsilon < \epsilon_\delta$,

$$|u(\mathbf{x}) - u_\epsilon(\mathbf{x})| = \left| \int_{\mathcal{D}} \phi_\epsilon(\mathbf{x}-\mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} \right| =$$

$$= \left| \int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{x}-\mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} \right| \leq$$

$$\leq \int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{x}-\mathbf{y})|u(\mathbf{x}) - u(\mathbf{y})| d\mathbf{y} \leq \delta \int_{B_\epsilon(\mathbf{x})} \phi_\epsilon d\mathbf{y} = \delta,$$

where we used the definition of u_ϵ and that $\int \phi_\epsilon(\mathbf{x}) d\mathbf{x} = 1$ in the first equality[20], that $\phi_\epsilon(\mathbf{x}-\mathbf{y}) = 0$ outside $B_\epsilon(\mathbf{x})$ in the second and that $|u(\mathbf{x}) - u(\mathbf{y})| < \delta$ in the first inequality. Since $\delta > 0$ is arbitrary the statement follows.

3. We can find a continuous function $v(\mathbf{x})$ such that $\|u - v\|_{L^p(\mathcal{D})} < \delta$ and $m(\{\mathbf{x};\ u(\mathbf{x}) \neq v(\mathbf{x})\}) < \delta$, see exercise 10.

Now

$$\|u - u_\epsilon\|_{L^p(K)}^p = \|(u - v) - (u_\epsilon - v_\epsilon) + (v - v_\epsilon)\|_{L^p(K)}^p \leq \quad (58)$$

$$\leq \delta^p + \|u_\epsilon - v_\epsilon\|_{L^p(K)}^p + \|v - v_\epsilon\|_{L^p(K)}^p.$$

[20] This implies in particular that $u(\mathbf{x}) = u(\mathbf{x}) \int \phi_\epsilon(\mathbf{x}) d\mathbf{x} = \int u(\mathbf{x})\phi_\epsilon(\mathbf{x}) d\mathbf{x}$.

If ϵ is small enough then the last term in (58) will be less than δ^p since $v_\epsilon \to v$ point-wise. Therefore we only need to estimate the term $\|u_\epsilon - v_\epsilon\|^p_{L^p(K)}$. Since $\|u - v\|^p_{L^p(\mathcal{D})} < \delta^p$ it is enough to how that $\|w_\epsilon\|_{L^p(K)} \leq \|w\|_{L^p(\mathcal{D})}$ for any $w \in L^p(\mathcal{D})$.

Let $w \in L^p(\mathcal{D})$ then for any $\mathbf{x} \in \mathcal{D}_\epsilon$

$$|w_\epsilon(\mathbf{x})| = \left| \int_{B_\epsilon(\mathbf{x})} \phi_\epsilon^{1-1/p}(\mathbf{x}-\mathbf{y}) \phi_\epsilon^{1/p}(\mathbf{x}-\mathbf{y}) w(\mathbf{y}) d\mathbf{y} \right| \leq$$

$$\leq \left(\int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{x}-\mathbf{y}) d\mathbf{y} \right)^{1-1/p} \left(\int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{x}-\mathbf{y}) |w(\mathbf{y})|^p d\mathbf{y} \right)^{1/p} =$$

$$= \left(\int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{x}-\mathbf{y}) |w(\mathbf{y})|^p d\mathbf{y} \right)^{1/p},$$

where we used that $\int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{x}-\mathbf{y}) d\mathbf{y} = 1$ in the last equality. Raising both sides to the power p and integrating over $\mathcal{D}_\epsilon$ results in

$$\int_{\mathcal{D}_\epsilon} |w_\epsilon(\mathbf{x})|^p d\mathbf{x} \leq \int_{\mathcal{D}_\epsilon} \left(\int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{x}-\mathbf{y}) |w(\mathbf{y})|^p d\mathbf{y} \right) d\mathbf{x} =$$

$$= \int_{\mathcal{D}_\epsilon} \left(\int_{B_\epsilon(0)} \phi_\epsilon(-\mathbf{y}) |w(\mathbf{y}+\mathbf{x})|^p d\mathbf{y} \right) d\mathbf{x} =$$

$$= \int_{B_\epsilon(0)} \phi_\epsilon(-\mathbf{y}) \left(\int_{\mathcal{D}_\epsilon} |w(\mathbf{y}+\mathbf{x})|^p d\mathbf{x} \right) d\mathbf{y} \leq$$

$$\leq \int_{B_\epsilon(0)} \phi_\epsilon(-\mathbf{y}) \underbrace{\left(\int_{\mathcal{D}} |w(\mathbf{x})|^p d\mathbf{x} \right)}_{=\|w\|^p_{L^p(\mathcal{D})}} d\mathbf{y} \leq \|w\|^p_{L^p(\mathcal{D})}.$$

It follows that

$$\|u_\epsilon - v_\epsilon\|^p_{L^p(\mathcal{D}_\epsilon)} \leq \|u - v\|^p_{L^p(\mathcal{D})} < \delta^p.$$

We may conclude from (58) that, for any $\delta > 0$,

$$\|u - u_\epsilon\|^p_{L^p(K)} < 3\delta^p$$

if ϵ is small enough. The statement follows.

4. This statement follows from the integration by parts formula

$$\frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i} = \int_{\mathcal{D}} \frac{\partial \phi_\epsilon(\mathbf{x}-\mathbf{y})}{\partial x_i} u_\epsilon(\mathbf{y}) d\mathbf{y} =$$

$$= -\int_{\mathcal{D}} \frac{\partial \phi_\epsilon(\mathbf{x}-\mathbf{y})}{\partial y_i} u_\epsilon(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{D}} \phi_\epsilon(\mathbf{x}-\mathbf{y}) \frac{\partial u_\epsilon(\mathbf{y})}{\partial y_i} d\mathbf{y} =$$

$$= \phi_\epsilon * \frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i}.$$

5. This follows directly from the two previous parts of the proposition.

□

The previous proposition shows that we may approximate $W^{1,p}(\mathcal{D})$ functions on compact sets by $C^\infty-$functions. At times it is convenient to approximate $W^{1,p}(\mathcal{D})$ on the entire set $\mathcal{D}$. The way to do that is to write $u \in W^{1,p}(\mathcal{D})$ as a series of functions with compact support and then approximate these functions individually. In order to write u as a series of functions with compact support we need to introduce the concept of partition of unity.

Proposition 9.2. *Given an open set $\mathcal{D}$ and an open countable covering U_k of $\mathcal{D}$, by an open covering we mean countable many sets U_k such that $\mathcal{D} \subset \cup_{k=1}^\infty U_k$. Furthermore we assume that each point $\boldsymbol{x} \in \mathcal{D}$ there is only finitely many k such that $\boldsymbol{x} \in U_k$.*

Then there exists functions $\psi_k \in C^\infty(\cup_k U_k)$ such that

1. *$spt(\psi_k) \subset U_k$,*
2. *$0 \le \psi_k \le 1$ and*
3. *$\sum_{k=1}^\infty \psi_k(\boldsymbol{x}) = 1$ on $\mathcal{D}$.*

We call ϕ_k a partition of unity of $\mathcal{D}$ with respect to the cover U_k.

Proof: Let

$$\eta_j(\mathbf{x}) = \begin{cases} \mathrm{dist}(\mathbf{x}, \partial U_j \cap \mathcal{D}) & \text{if } \mathbf{x} \in U_j \\ 0 & \text{if } \mathbf{x} \notin U_j \end{cases}$$

and define the sets

$$K_j = \{\mathbf{x} \in U_j; \eta_j(\mathbf{x}) \ge \sup_{k \ne j} \eta_k(\mathbf{x})\}.$$

We claim that K_j are closed in $\mathcal{D}$. To show this we will show that any Cauchy sequence $x_n \in K_j$ converges to a point in $\mathbf{x}^0 \in K_j$. Since at least one of the functions $\eta_k \ne 0$ at every point and $\eta_j(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial U_j$ it follows that $\mathbf{x}^0 \notin \partial U_j$. If $\mathbf{x}_n \to \mathbf{x}^0 \in U_j$ it follows that $\eta_j(\mathbf{x}^0) = \delta$ for $\delta = \mathrm{dist}(\mathbf{x}^0, U_j^c) > 0$. And since $\mathbf{x}^0$ is contained in only finitely many of the U_k only finitely many $\eta_k(\mathbf{x}^0) > 0$. Since $|\nabla \eta_k| \le 1$ we may conclude that only finitely many η_k have values greater than $\delta/2$ in $B_{\delta/2}(\mathbf{x}^0)$, say that only $\eta_1, \eta_2, ..., \eta_N$ have values greater than $\delta/2$ in $B_{\delta/2}(\mathbf{x}^0)$. Also, since $\eta_j(\mathbf{x}^0) = \delta$ and $|\nabla \eta_j| \le 1$ we may conclude that $\eta_j(\mathbf{x}) \ge \delta/2$ in $B_{\delta/2}(\mathbf{x}^0)$ Therefore

$$K_j \cap B_{\delta/2}(\mathbf{x}^0) = \{\mathbf{x} \in U_j; \eta_j(\mathbf{x}) \ge \sup_{k=1,2,...,N} \eta_k(\mathbf{x})\} \cap B_{\delta/2}(\mathbf{x}^0).$$

Since η_j and $\sup_{k=1,2,...,N} \eta_k(\mathbf{x})$ are a continuous functions it follows that K_j is relatively closed in $B_{\delta/2}(\mathbf{x}^0)$. Therefore $\mathbf{x}_n \in K_j$ and $\mathbf{x}_n \to \mathbf{x}^0$ will imply that $\mathbf{x}^0 \in K_j$. It follows that K_j is closed.

Also notice that $\cup_j K_j = \mathcal{D}$ since each $\mathbf{x} \in \mathcal{D}$ will be contained in at least on eK_j. Furthermore, since K_j is relatively closed in $\mathcal{D}$ and U_j is open, $\mathrm{dist}(K_j, \mathcal{D} \setminus U_j) = 2\epsilon_j > 0$.

We may define

$$\hat{\psi}_j(\mathbf{x}) = \phi_{\epsilon_j} * \chi_{K_j}(\mathbf{x}),$$

where ϕ_ϵ is the standard mollifier as in Proposition 9.1 and $\chi_{U_{K_j}}$ is the characteristic function of K_j:

$$\chi_{K_j}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in K_j \\ 0 & \text{if } \mathbf{x} \notin K_j. \end{cases}$$

Then, by Proposition 9.1, $\hat{\psi}_j(\mathbf{x}) \in C^\infty$ and $\hat{\psi}_j$ has compact support in U_j.

Since $\mathcal{D} = \cup_j K_j$ it follows that $\sum_k \hat{\psi}_j > 0$ in $\mathcal{D}$. Therefore

$$\psi_j(\mathbf{x}) = \frac{\hat{\psi}_j(\mathbf{x})}{\sum_{k=1}^\infty \hat{\psi}_k(\mathbf{x})}$$

is well defined. Clearly $\text{spt}(\psi_j) \subset U_j$ and $\sum_{k=1}^\infty \psi_k(\mathbf{x}) = 1$ in $\mathcal{D}$. This finishes the proof. □

Theorem 9.1. *$C^\infty(\mathcal{D})$ is dense in $W^{1,p}(\mathcal{D})$, in particular if $\epsilon > 0$ and $u \in W^{1,p}(\mathcal{D})$ then there exists a function $u_\epsilon \in C^\infty(\mathcal{D})$ such that $\|u - u_\epsilon\|_{W^{1,p}(\mathcal{D})} < \epsilon$.*

Proof: We begin by proving the theorem under the assumption that $\mathcal{D}$ is bounded.

Let us cover $\mathcal{D}$ by the sets $U_k = \{\mathbf{x} \in \mathcal{D}; \frac{1}{2k} < \text{dist}(\mathbf{x}, \mathcal{D}^c) < \frac{2}{k}\}$, for $k = 1, 2, ...,$ and $U_0 = \{\mathbf{x} \in \mathcal{D}; \text{dist}(\mathbf{x}, \mathcal{D}^c) > 1\}$. Then, by Proposition 9.2, there exists a partition of unity ϕ_k with respect to the cover U_k. We define the functions $u_k(\mathbf{x}) = \psi_k(\mathbf{x})u(\mathbf{x})$, then $u = \sum_{k=0}^\infty u_k(\mathbf{x})$ and each u_k has compact support in $\mathcal{D}$. We may therefore, by Proposition 9.1, approximate u_k by a function $v_k \in C^\infty(\mathcal{D})$ so that $\|u_k - v_k\|_{W^{1,p}(\mathcal{D})} < \frac{\epsilon}{2^{k+1}}$. We define $v = \sum_{k=0}^\infty v_k$, then $v \in C^\infty(\mathcal{D})$ since only finitely many v_k are non-zero at each point $\mathbf{x}$.

By the triangle inequality it follows that

$$\|u - v\|_{W^{1,p}(\mathcal{D})} = \left\| \sum_{k=0}^\infty (u_k - v_k) \right\|_{W^{1,p}(\mathcal{D})} \leq$$

$$\leq \sum_{k=0}^\infty \|(u_k - v_k)\|_{W^{1,p}(\mathcal{D})} \leq \sum_{k=0}^\infty \frac{\epsilon}{2^{k+1}} = \epsilon.$$

As for the general case when the support of u is not compact we may first approximate u by $u(\mathbf{x})\eta_N(\mathbf{x})$ where

$$\eta_N(\mathbf{x}) = \begin{cases} 1 & \text{if } |\mathbf{x}| \leq N \\ 0 & \text{if } |\mathbf{x}| > N+1 \end{cases}$$

and $\eta_N \in C^\infty(\mathbb{R}^n)$. By choosing N large enough we can assure that $\|u - u(\mathbf{x})\eta_N(\mathbf{x})\|_{W^{1,p}(\mathcal{D})} < \epsilon$. Since $u(\mathbf{x})\eta_N(\mathbf{x})$ has compact support in $\mathcal{D}$ the first part of the proof assures that there exists a $v \in C^\infty(\mathcal{D})$ such that $\|u(\mathbf{x})\eta_N(\mathbf{x}) - v\|_{W^{1,p}(\mathcal{D})} < \epsilon$. By the triangle inequality we can conclude that

$$\|u - v\|_{W^{1,p}(\mathcal{D})} \leq \|u(\mathbf{x})\eta_N(\mathbf{x}) - v\|_{W^{1,p}(\mathcal{D})} + \|u - u(\mathbf{x})\eta_N(\mathbf{x})\|_{W^{1,p}(\mathcal{D})} < 2\epsilon.$$

□

That we may approximate $W^{1,p}$ functions by C^∞ functions will help us to prove some classical theorems for $W^{1,p}$ functions.

Proposition 9.3. [PRODUCT AND CHAIN RULES FOR $W^{1,p}$ FUNCTIONS.]

1. [PRODUCT RULE] *If $u, v \in W^{1,p}(\mathcal{D}) \cap L^\infty(\mathcal{D})$ then $uv \in W^{1,p}(\mathcal{D}) \cap L^\infty(\mathcal{D})$ and*
$$\frac{\partial u(\boldsymbol{x})v(\boldsymbol{x})}{\partial x_i} = u(\boldsymbol{x})\frac{\partial v(\boldsymbol{x})}{\partial x_i} + v(\boldsymbol{x})\frac{\partial u(\boldsymbol{x})}{\partial x_i}.$$

2. [CHAIN RULE] *Assume that $f \in C^1(\mathbb{R})$, with bounded derivative and $f(0) = 0$, assume furthermore that $u \in W^{1,p}(\mathcal{D})$. Then $f(u(\boldsymbol{x})) \in W^{1,p}(\mathcal{D})$ and*
$$\frac{\partial f(u(\boldsymbol{x}))}{\partial x_i} = f'(u(\boldsymbol{x}))\frac{\partial u(\boldsymbol{x})}{\partial x_i}.$$

Proof: The proof is based on approximation. We begin by proving the product rule. Let $\phi \in C_c^1(\mathcal{D})$ then and u_ϵ, v_ϵ be the C^∞ approximations of u and v from Theorem 9.1, by the triangle inequality,

$$\left|\int_{\mathcal{D}} u(\mathbf{x})v(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial x_i}d\mathbf{x} + \int_{\mathcal{D}}\left(u(\mathbf{x})\frac{\partial v(\mathbf{x})}{\partial x_i} + v(\mathbf{x})\frac{\partial u(\mathbf{x})}{\partial x_i}\right)\phi(\mathbf{x})d\mathbf{x}\right| \leq \tag{59}$$

$$\leq \left|\int_{\mathcal{D}} u_\epsilon(\mathbf{x})v_\epsilon(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial x_i}d\mathbf{x} + \int_{\mathcal{D}}\left(u_\epsilon(\mathbf{x})\frac{\partial v_\epsilon(\mathbf{x})}{\partial x_i} + v_\epsilon(\mathbf{x})\frac{\partial u\epsilon(\mathbf{x})}{\partial x_i}\right)\phi(\mathbf{x})d\mathbf{x}\right| +$$
$$+\left|\int_{\mathcal{D}}(uv - u_\epsilon v_\epsilon)\frac{\partial\phi(\mathbf{x})}{\partial x_i}d\mathbf{x}\right| + \left|\int_{\mathcal{D}}\left(u(\mathbf{x})\frac{\partial v(\mathbf{x})}{\partial x_i} - u_\epsilon(\mathbf{x})\frac{\partial v_\epsilon(\mathbf{x})}{\partial x_i}\right)\phi(\mathbf{x})d\mathbf{x}\right| +$$
$$+\left|\int_{\mathcal{D}}\left(v(\mathbf{x})\frac{\partial u(\mathbf{x})}{\partial x_i} - v_\epsilon(\mathbf{x})\frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i}\right)\phi(\mathbf{x})d\mathbf{x}\right| = I_1 + I_2 + I_3 + I_4.$$

It follows from partial integration that $I_1 = 0$. We claim that $I_2, I_3, I_4 \to 0$ as $\epsilon \to 0^+$. We will only prove this for I_2. We estimate

$$I_2 \leq \left\|\frac{\partial\phi}{\partial x_i}\right\|_{L^\infty(\mathcal{D})}\int_{\mathcal{D}}|uv - u_\epsilon v_\epsilon|d\mathbf{x} \leq \left\|\frac{\partial\phi}{\partial x_i}\right\|_{L^\infty(\mathcal{D})}\int_{\mathcal{D}}|u(v - v_\epsilon)|d\mathbf{x} +$$
$$+\left\|\frac{\partial\phi}{\partial x_i}\right\|_{L^\infty(\mathcal{D})}\int_{\mathcal{D}}|v_\epsilon(u - u_\epsilon)| \leq \left\|\frac{\partial\phi}{\partial x_i}\right\|_{L^\infty(\mathcal{D})}\|u\|_{L^\infty}\int_{\mathcal{D}}|(v - v_\epsilon)|d\mathbf{x} +$$
$$+\left\|\frac{\partial\phi}{\partial x_i}\right\|_{L^\infty(\mathcal{D})}\|v_\epsilon\|_{L^\infty}\int_{\mathcal{D}}|(u - u_\epsilon)|d\mathbf{x} \to 0,$$

since $u_\epsilon \to u$ and $v_\epsilon \to v$ in L^1. The proofs that $I_3, I_4 \to 0$ are similar. Inserting that $I_1, ..., I_4 \to 0$ in (59) proves the product rule.

Next we prove the chain rule. Notice that since f' is bounded, say $|f'| < C$, it follows that
$$|f(u(\mathbf{x})) - f(u_\epsilon(\mathbf{x}))| \leq C|u(\mathbf{x}) - u_\epsilon(\mathbf{x})|.$$

Therefore
$$\lim_{\epsilon\to 0^+}\int_{\mathcal{D}} f(u_\epsilon(\mathbf{x}))\frac{\partial\phi}{\partial x_i}d\mathbf{x} = \int_{\mathcal{D}} f(u(\mathbf{x}))\frac{\partial\phi(\mathbf{x})}{\partial x_i}d\mathbf{x}. \tag{60}$$

Making an integration by parts in the integral on the left in (60) we see that

$$\int_{\mathcal{D}} f(u_\epsilon(\mathbf{x}))\frac{\partial\phi}{\partial x_i}d\mathbf{x} = -\int_{\mathcal{D}} f'(u_\epsilon(\mathbf{x}))\frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i}\phi d\mathbf{x} =$$

$$
= -\int_{\mathcal{D}} f'(u(\mathbf{x}))\frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i}\phi d\mathbf{x} - \int_{\mathcal{D}} \left(f'(u_\epsilon) - f'(u)\right)\left(\frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i} - \frac{\partial u(\mathbf{x})}{\partial x_i}\right)\phi d\mathbf{x} -
$$

$$
- \int_{\mathcal{D}} \left(f'(u_\epsilon) - f'(u)\right)\frac{\partial u(\mathbf{x})}{\partial x_i}\phi d\mathbf{x} = I_1 + I_2 + I_3. \tag{61}
$$

Since $\frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i} \to \frac{\partial u(\mathbf{x})}{\partial x_i}$ it follows that

$$
I_1 \to -\int_{\mathcal{D}} f'(u(\mathbf{x}))\frac{\partial u(\mathbf{x})}{\partial x_i}\phi d\mathbf{x} \quad \text{as } \epsilon \to 0^+.
$$

We may estimate I_2 according to

$$
|I_2| \le \left\|f'(u_\epsilon) - f'(u)\right\|_{L^\infty}\left\|\frac{\partial u_\epsilon(\mathbf{x})}{\partial x_i} - \frac{\partial u(\mathbf{x})}{\partial x_i}\right\|_{L^1(\mathcal{D})} \to 0.
$$

Finally, in order to estimate I_3 we notice that $f'(u_\epsilon) \rightharpoonup f'(u)$. This weak convergence follows from the assumption that f' is bounded and therefore uniformly bounded in L^1 which is weakly compact. Therefore

$$
I_3 \to 0 \quad \text{as } \epsilon \to 0^+.
$$

Using the limits of I_1, I_2 and I_3 in (60) and (61) the theorem follows. □

So far we have spent some time to show that functions in $W^{1,p}$ behaves like classically differentiable functions. But the space $W^{1,p}$ is more flexible that the space C^1. In order to show this we state the following proposition.

Proposition 9.4. *If $u \in W^{1,p}(\mathcal{D})$ then $u^+ = \max(u, 0) \in W^{1,p}(\mathcal{D})$ and $\frac{\partial u^+(\boldsymbol{x})}{\partial x_i} = \frac{\partial u^+(\boldsymbol{x})}{\partial x_i}\chi_{\{u>0\}}(\boldsymbol{x})$ where $\chi_{\{u>0\}}$ is the characteristic function of the set $\{u > 0\}$.*

Proof: Let us define

$$
f_\epsilon(t) = \begin{cases} \sqrt{t^2 + \epsilon^2} - \epsilon & \text{if } t \ge 0 \\ 0 & \text{if } t < 0. \end{cases}
$$

Then $f_\epsilon \in C^1(\mathbb{R})$ and therefore, by the chain rule, for any $\phi \in C_c^1(\mathcal{D})$

$$
\int_{\mathcal{D}} f_\epsilon(u(\mathbf{x}))\frac{\partial \phi(\mathbf{x})}{\partial x_i} d\mathbf{x} = -\int_{\mathcal{D}} f_\epsilon'(u(\mathbf{x}))\frac{\partial u(\mathbf{x})}{\partial x_i}\phi(\mathbf{x}) d\mathbf{x}.
$$

Passing to the limit $\epsilon \to 0^+$ proves the proposition. □

9.1 Exercises:

1. Let f_k be a sequence of continuous functions on $[0, 1]$ such that $f_k(x) \to f(x)$ uniformly on $[0, 1]$. Prove that $f(x)$ is continuous.
2. Provide all the details in example 2 (the devils staircase). In particular show that
   (a) Each f_k is continuous.
   (b) That $f'(x) = 0$ almost everywhere.

3. Let $\mathcal{D}$ be a bounded domain and $u(x) \in C^1(\mathcal{D})$. Show that $u(x) \in W^{1,2}(\mathcal{D})$.

4. Show that the weak derivatives of the following functions, $f(x)$, either exist or does not exist. Then calculate the weak derivative if it exists.

   (a) $f(x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x < 0 \end{cases}$ where $f(x)$ is defined on $[-1, 1] \subset \mathbb{R}$.
   Does weak derivatives have to be continuous?

   (b) $f(x) = \begin{cases} \frac{1}{x^{1/4}} \text{ if } x > 0 \\ 0 \text{ if } x < 0 \end{cases}$ where $f(x)$ is defined on $[-1, 1] \subset \mathbb{R}$.

   (c) $f(x) = \begin{cases} x^{3/4} \text{ if } x > 0 \\ 0 \text{ if } x < 0 \end{cases}$ where $f(x)$ is defined on $[-1, 1] \subset \mathbb{R}$.
   Does weak derivatives have to be bounded?

   (d) $f(x) = \begin{cases} \frac{x_1}{|x|} \text{ if } |x| > 0 \\ 0 \text{ if } x = 0 \end{cases}$ where $f(x)$ is defined on $B_1 \subset \mathbb{R}^3$.

5. Consider the function $u(x) = \ln(1/|x|)$ defined in $B_1(0) \subset \mathbb{R}^3$.

   (a) Show that $u(x) \in W^{1,2}(B_1(0))$.

   (b) Conclude that there are discontinuous, and even unbounded, functions $u(x) \in W^{1,2}(B_1(0))$.

6. Show that $\lim_{|\mathbf{x}| \to 1^-} \frac{e^{\frac{1}{|\mathbf{x}|^2-1}}}{(1-|\mathbf{x}|^2)^k} = 0$ and use this to argue that

$$\phi(\mathbf{x}) = \begin{cases} ce^{\frac{1}{|\mathbf{x}|^2-1}} & \text{if } |\mathbf{x}| < 1 \\ 0 & \text{if } |\mathbf{x}| \geq 1, \end{cases}$$

   belongs to $C^\infty(\mathbb{R}^n)$.

7. Let $\mathcal{D}$ be a bounded domain and $u(x) \in C^1(\mathcal{D})$.

   (a) If $\mathcal{D} = B_1(0)$ and $u(x) = 0$ on $\partial B_1(0)$ show that

$$u(0) = \frac{1}{\omega_n} \int_{B_1(0)} \frac{y \cdot \nabla u(y)}{|y|^n} dy.$$

   Hint: *By the fundamental theorem of calculus*

$$u(0) = -\int_0^1 y \cdot \nabla u(ty) dt$$

   for any y such that $|y| = 1$. Integrate this over the unit sphere $\partial B_1(0) = \{y;\ |y = 1|\}$.

   (b) Show that for all $x \in B_1(0)$.

$$u(x) = \frac{1}{\omega_n} \int_{B_1(0)} \frac{(y - x) \cdot \nabla u(y)}{|y - x|^n} dy.$$

(c) Use the following inequality, known as Hölder's inequality,

$$\int_{B_1} f(x)g(x)dx \le \left(\int_{B_1} |f(x)|^p dx\right)^{1/p} \left(\int_{B_1} |f(x)|^q dx\right)^{1/q}$$

for $\frac{1}{p} + \frac{1}{q} = 1$, to show that for any $\epsilon > 0$ there exists a constant C_ϵ such that

$$\sup_{B_1(0)} |u(x)| \le C_\epsilon \left(\int_{B_1(0)} |\nabla u(x)|^{n+\epsilon}\right)^{\frac{1}{n+\epsilon}}.$$

8. Verify that the spaces $W^{k,p}(\mathcal{D})$ are complete.

9. Verify that $\text{spt}(\phi_\epsilon) = B_\epsilon(0)$ and that $\int_{\mathbb{R}^n} \phi_\epsilon(\mathbf{x})d\mathbf{x} = 1$ for all $\epsilon > 0$.

10. Prove that for any function $u \in L^p(\mathcal{D})$ there exists a function $v \in C(\mathcal{D})$ such that $\|u - v\|_{L^p(\mathcal{D})} < \delta$ and $m(\{\mathbf{x};\ u(\mathbf{x}) \neq v(\mathbf{x})\}) < \delta$. You may use the following steps.

   (a) Define $u_M(\mathbf{x}) = \begin{cases} M & \text{if } u(\mathbf{x}) \ge M \\ u(\mathbf{x}) & \text{if } |u(\mathbf{x})| < M \\ -M & \text{if } u(\mathbf{x}) \le -M \end{cases}$ and show that if M is large enough then $\|u - u_M\|_{L^p(\mathcal{D})} < \delta_1$ and $m(\{\mathbf{x};\ u(\mathbf{x}) \neq u_M(\mathbf{x})\}) < \delta_1$.

   (b) Use Lusin's Theorem to show that we may approximate u_M by a continuous function v such that $m(\{\mathbf{x};\ v(\mathbf{x}) \neq u_M(\mathbf{x})\}) < \delta_2$.

   (c) Show the following inequalities

   $$\|u - v\|_{L^p(\mathcal{D})} \le \|u - u_M\|_{L^p(\mathcal{D})} + \|u_M - v\|_{L^p(\mathcal{D})} \le \delta_1 + \delta_2^{1/p} M$$

   and choose δ_1 and δ_2 appropriately to conclude. [Hint:] *You need Hölder's inequality in the final estimate.*

11. [A really bad function.] In this exercise we will construct a really bad function - in mathematical analysis we love bad functions as examples.

   (a) Show that $u(x) = \begin{cases} \frac{x_1}{|x|^{4/3}} \text{ if } |x| > 0 \\ 0 \text{ if } x = 0 \end{cases}$ satisfies $u(x) \in W^{1,2}(B_2(0))$ when the space dimension $n \ge 3$. Also show that $u(x)$ is not bounded in any neighborhood of $x = 0$.

   (b) Since $\mathbb{Q}^3$ is countable we may define a sequence $\{q_j\}_{j=1}^\infty$ such that $\cup_{j=1}^\infty \{q_j\} = \mathbb{Q}^3 \cap B_1(0)$. Define $w(x) = \sum_{j=1}^\infty 2^{-j} u(x - q_j)$ and show that $w(x) \in W^{1,2}(B_1^+(0))$ and that $w(x)$ is not bounded, neither from above nor from below, on any open set of $B_1^+(0)$.

   [Hint:] *In order to show that $w(x) \in W^{1,2}(B_1^+(0))$ it might be helpful to use the following triangle inequality* $\left\|\sum_j f_j(x)\right\|_{W^{1,2}} \le \sum_j \|f_j(x)\|_{W^{1,2}}$.

   (c) What is $\limsup_{x \to x^0} w(x)$ and $\liminf_{x \to x^0} w(x)$ for $x^0 \in B_1(0)$?

10 Lecture 10. Differentiation and the Fundamental Theorem of Calculus, part 1.

In the last lecture we introduced Sobolev spaces and showed that Sobolev functions have some good properties. We will continue our investigation by trying to better understand differentiability properties of functions in Sobolev and Lebesgue spaces, we would in particular want to derive a fundamental theorem of calculus that is as general as possible.

Reading:

W. Rudin, Real and complex analysis: Chapter 7 and pages 135-156

There are rather many pages to read in Rudin this week. The must important theory is the fundamental theorem of calculus (contained in the pages 144-150) focus on that and make sure that you understand that.

RECOMMENDED EXERCISES: *7.1, 7.5, 7.9, 7.10, 7.12*

H.L. Royden P.M Fitzpatrick, Real Analysis: Chapter 6.1-6.5 (also very many pages...)

RECOMMENDED EXERCISES: *6.3, 6.7, 6.18, 6.21, 6.30, 6.33, 6.38, 6.44*

The lectures was based on the following material.

The undergraduate version of the fundamental theorem of calculus states that if f is continuous on the interval $[a, b] \subset \mathbb{R}$ and if

$$F(x) = \int_a^x f(t)dt \qquad \text{then} \qquad F'(x) = f(x). \tag{62}$$

At times one sees the following version of the fundamental theorem: if $f(x)$ is differentiable on $[a, b]$ then

$$f(x) - f(a) = \int_a^x f'(t)dt. \tag{63}$$

The standard proof in undergraduate calculus books uses the intermediate value theorem; and thus on the continuity of f in (62) and on f' in (63).

If we would like to, and indeed we do like to, derive a version of the fundamental theorem of calculus that is applicable without the continuity assumption on f we need to find a different way to investigate the derivative of an integral.

Does it even hold for arbitrary integrable functions f? If not to what extent does the fundamental theorem hold. Are the versions of the fundamental theorem in (62) and (63) the same? How do we even approach the problem?

In general the derivative

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

might not be well defined at a point. In order to get something well defined to work with when we differentiate a function we will define the Daboux derivatives.

Definition 10.1. *Let f be defined in a neighbourhood of the point $x \in \mathbb{R}$ then we define the following Darboux derivatives:*

$$D^+ f(x) = \limsup_{h \to 0^+} \frac{f(x+h) - f(x)}{h}, \qquad D_+ f(x) = \liminf_{h \to 0^+} \frac{f(x+h) - f(x)}{h},$$

$$D^- f(x) = \limsup_{h \to 0^-} \frac{f(x+h) - f(x)}{h}, \text{ and } D_- f(x) = \liminf_{h \to 0^-} \frac{f(x+h) - f(x)}{h}.$$

Note that all the Darboux derivatives are well defined in the extended real numbers. We may therefore, for any given function $f : [a, b] \mapsto \mathbb{R}$, define the function $D^+ f(x) : [a, b] \mapsto \mathbb{R} \cup \{\infty\}$, $D_+ f(x) : [a, b] \mapsto \mathbb{R} \cup \{-\infty\}$ et.c.[21]

Theorem 10.1. *Assume that f is non-decreasing on $[a, b]$ then $D^+ f(x)$, $D_+ f(x)$, $D^- f(x)$ and $D_- f(x)$ are all measurable on $[a, b]$.*

Proof: We will only show that $D^+ f(x)$ is measurable, the proofs for $D_+ f(x)$, $D^- f(x)$ and $D_- f(x)$ are similar. In this proof we will extend f to $[a, \infty)$ by setting $f(x) = f(b)$ for $x > b$.

In order to show that $D^+ f(x)$ is measurable we define the rational Darboux derivative

$$D^+_{\mathbb{Q}} f(x) = \limsup_{h \in \mathbb{Q},\, h \to 0^+} \frac{f(x+h) - f(x)}{h}.$$

Since f is non-decreasing it is measurable and therefore, for any $q \in \mathbb{Q}$,

$$f_q(x) = \frac{f(x+q) - f(x)}{q}$$

is measurable. But since the supremum of a countable set of measurable functions is measurable it follows that

$$f^h(x) = \sup_{q \in \mathbb{Q},\, 0 < q < h} f_q(x)$$

is measurable.

But, since the infinum of a countable set of measurable functions is measurable it follows that

$$D^+_{\mathbb{Q}} f(x) = \inf_{n \in \mathbb{N}} f^{1/n}(x)$$

is measurable.

In order to complete the proof we need to show that $D^+_{\mathbb{Q}} f(x) = D^+ f(x)$. We begin by showing this when $D^+ f(x)$ is finite. It is enough to show that, for any $\epsilon > 0$,

$$D^+ f(x) - \epsilon < D^+_{\mathbb{Q}} f(x) < D^+ f(x) + \epsilon. \tag{64}$$

The right inequality in (64) is obvious since we take the supremum over a larger set in D^+ than in $D^+_{\mathbb{Q}}$. In order to show the left inequality in (64) we pick an $h > 0$ such that

$$D^+ f(x) - \frac{\epsilon}{2} < \frac{f(x+h) - f(x)}{h}. \tag{65}$$

[21] Technically, $D^+ f(x)$ is not defined at $x = b$, but as a measurable function we do not care about the value of f at single points and will therefore assign an arbitrary value to $D^+ f(b)$

Then, for $q \in \mathbb{Q}$, and $q > h$ we get

$$\frac{f(x+q)-f(x)}{q} \geq \frac{f(x+h)-f(x)}{h}\frac{h}{q} > \left(D^+f(x) - \frac{\epsilon}{2}\right)\frac{h}{q}, \tag{66}$$

where we used that f is non-decreasing and $q > h$ in the first inequality and (65) in the second. By choosing q close enough to h in (66) the left inequality in (64) follows.

It remains to show that $D^+_{\mathbb{Q}}f(x) = \infty$ if $D^+f(x) = \infty$. The argument is very similar as when $D^+f(x)$ is finite. If $D^+f(x) = \infty$ then for any $N \in \mathbb{N}$ there exists an $h > 0$ such that

$$N + 1 < \frac{f(x+h)-f(x)}{h}.$$

By arguing as in (66) it follows that $D^+_{\mathbb{Q}}f(x) > N$ for any $N \in \mathbb{N}$. That $D^+_{\mathbb{Q}}f(x) = D^+f(x) = \infty$ follows. □

We also need to be able to extract some global information from the Darboux derivative $D^+f(x)$. This is rather subtle. By definition $D^+f(x)$ only depend on the values of f in arbitrarily small neighborhoods of x. But we would want information about the derivative, such as $D^+f(x) \geq 0$, to say something about the derivative in an entire interval. We will do that by using a covering theorem, but first we need to define what we mean by a covering.

Definition 10.2. *Let S be a subset of the real numbers. Then we say that a collection of closed intervals, $\mathcal{F}$, is a Vitali covering of S if for every for every $x \in S$ and any $h > 0$ there is an interval $[a,b] \in \mathcal{F}$ such that $x \in [a,b]$ and $|b-a| < h$. We explicitly exclude the possibility that an interval consists of a single point in the Vitali covering.*

Theorem 10.2. [The Vitali Covering Theorem]. *Let $S \subset [a,b]$ and $\mathcal{F}$ be a Vitali covering of S. Then there exists sequence of intervals $I_n \in \mathcal{F}$ such that $S \setminus \bigcup_{n=1}^{\infty} I_n$ is a null set.*[22] *Furthermore $(I_j \cap I_k) \cap [a,b] = \emptyset$ for $j \neq k$, that is the intervals I_n are disjoint in $[a,b]$.*

Proof: It will be expedient to consider the intervals $I \in \mathcal{F}$ to be contained in $[a,b]$. Let us therefore consider the Vitali cover $\mathcal{F}' = \{I \cap [a,b];\ I \in \mathcal{F}\}$ of S. Clearly if we can construct can find a sequence of intervals $I'_n \in \mathcal{F}'$ such as in the Theorem then the corresponding sequence $I_n \in \mathcal{F}$ will also satisfy the conditions of the theorem. We will therefore assume that all intervals $I \in \mathcal{F}$ is contained in $[a,b]$.

If $I \in \mathcal{F}$ then $I \subset [a,b]$, by the assumption in the last paragraph, and therefore $m(I) \leq b - a$. Therefore the following is well defined

$$m_1 = \sup_{I \in \mathcal{F}, S \cap I \neq \emptyset} m(I).$$

We may therefore choose an interval $I_1 \in \mathcal{F}$ such that $I_1 \cap S \neq \emptyset$ and $m(I_1) > m_1/2$.

[22] We do not exclude the possibility that there are only finitely many intervals I_n, in that case we use the convention that I_n is the "empty interval" for n large enough.

Let us inductively define the interval I_{n+1} as follows. If $S \subset \cup_{k=1}^n I_k$ then the collection will satisfy the conditions of the Theorem and we are done. If $S \not\subset \cup_{k=1}^n I_k$ then we set

$$m_{n+1} = \sup_I m(I), \tag{67}$$

where the supremum is over all $I \in \mathcal{F}$ such that $I \cap S \neq \emptyset$ and $I \cap \cup_{k=1}^n I_k = \emptyset$. We then choose an interval I_{n+1}, among all intervals $I \in \mathcal{F}$ such that $I \cap S \neq \emptyset$ and $I \cap \cup_{k=1}^n I_k = \emptyset$, satisfying

$$m(I_{n+1}) > \frac{m_{n+1}}{2}, \tag{68}$$

among all intervals $I \in \mathcal{F}$ such that $I \cap S \neq \emptyset$ and $I \cap \cup_{k=1}^n I_k = \emptyset$.

We need to show that there are always intervals $I \in \mathcal{F}$ that satisfies the criteria. If $S \not\subset \cup_{k=1}^n I_k$ then there exists at least one point $\mathbf{x} \in S \setminus \cup_{k=1}^n I_k$. But since each I_k is closed the union $\cup_{k=1}^n I_k$ is closed and the distance form a point to a closed set is always strictly positive (unless the point is in the set): thus $\text{dist}(\mathbf{x}, \cup_{k=1}^n I_k) = \delta > 0$. But then, by the definition of Vitali cover, it follows that there is an interval $[\alpha, \beta] = I \in \mathcal{F}$ such that $\beta - \alpha < \delta/2$ and $\mathbf{x} \in [\alpha, \beta]$. It follows that the set of intervals we take the supremum over in (67) is non-empty and the supremum is therefore well defined.

Unless the procedure ends after a finite number of steps, in which case we have a collection as in the Theorem, we will get an infinite sequence $I_1, I_2, ...$ We claim that this sequence of intervals satisfy the conditions of the theorem.

We will argue by contradiction from now on and assume that

$$m(S \setminus \cup_{n=1}^\infty I_n) = \delta > 0. \tag{69}$$

Since the intervals I_n are disjoint and $\cup_{n=1}^\infty I_n \subset [a, b]$ it follows, from monotonicity of the measure, that

$$b - a = m([a, b]) \geq m(\cup_{n=1}^\infty) = \sum_{n=1}^\infty m(I_n), \tag{70}$$

where we also used countable additivity of the measure in the last equality. Notice that (70) implies that $m(I_n) \to 0$ as $n \to \infty$ and therefore since $0 \leq m_n < 2m(I_n)$

$$m_n \to 0. \tag{71}$$

We will denote by $J_n = [c_n, d_n]$ the interval that has the same center as $I_n = [a_n, b_n]$ but five times the length: $m(J_n) = 5m(I_n)$. Therefore

$$\sum_{n=1}^\infty m(J_n) = 5 \sum_{n=1}^\infty m(I_n) \leq 5(b - a),$$

which implies that if N is large enough then

$$m\left(\cup_{n=N+1}^\infty J_n\right) = \sum_{n=N+1}^\infty m(J_n) < \delta = m(S \setminus \cup_{n=1}^\infty I_n) \tag{72}$$

The inequalities (72) and (69) implies that there must exist an

$$\mathbf{x} \in S \setminus \bigcup_{n=1}^\infty I_n \text{ such that } \mathbf{x} \notin \bigcup_{n=N+1}^\infty J_n. \tag{73}$$

We fix one of these $\mathbf{x}$. It follows from (73) that

$$\mathbf{x} \notin \bigcup_{n=N+1}^{\infty} I_n.$$

Since $\bigcup_{n=1}^{N} I_n$ is closed there is a fixed distance $\epsilon > 0$ between $\mathbf{x}$ and $\bigcup_{n=1}^{N} I_n$:

$$\operatorname{dist}\left(\mathbf{x}, \bigcup_{n=1}^{N} I_n\right) = \epsilon > 0.$$

And since $\mathcal{F}$ is a Vitali cover there exists an $I_{\mathbf{x}} \in \mathcal{F}$ such that $\mathbf{x} \in I_{\mathbf{x}}$ and $m(I_{\mathbf{x}}) < \epsilon$, in particular

$$I_{\mathbf{x}} \bigcap \left(\cup_{n=1}^{N} I_n\right) = \emptyset. \tag{74}$$

Notice that by construction $m(I_{\mathbf{x}}) \leq m_N$.

Since $m_n \to 0$ by (71) it follows that for some large enough $q \in \mathbb{N}$ that

$$m(I_{\mathbf{x}}) > m_q = \sup_I m(I),$$

where the supremum was taken over all $I \in \mathcal{F}$ such that $I \cap S \neq \emptyset$ and $I \bigcap \cup_{n=1}^{q-1} I_n = \emptyset$ it follows that

$$I_{\mathbf{x}} \bigcap \left(\bigcap_{n=1}^{q-1} I_n\right) \neq \emptyset.$$

We may choose $p \in \mathbb{N}$ as the smallest integer such that

$$I_x \cap I_p \neq \emptyset. \tag{75}$$

Since p was the smallest integer this implies that

$$I_{\mathbf{x}} \bigcap \left(\bigcup_{n=1}^{p-1} I_n\right) = \emptyset \tag{76}$$

and therefore, by (71),

$$m(I_{\mathbf{x}}) \leq m_{p-1} \leq 2m(I_p) \tag{77}$$

where we used the choice (68) in the last inequality.

We are now ready to derive a contradiction to the assumption (69). Since $I_{\mathbf{x}} \bigcap \cup_{n=1}^{N} I_n = \emptyset$, from (74) and $I_{\mathbf{x}} \cap I_p \neq \emptyset$, by (75) it follows that $p \geq N+1$. And since $\mathbf{x} \notin \cup_{n=N+1}^{\infty} J_n$, by (73) we can conclude that $\mathbf{x} \notin J_p$.

But if $\mathbf{x} \notin J_p = [c_p, d_p]$ will imply that

$$\operatorname{dist}(\mathbf{x}, I_p) > 2m(I_p), \tag{78}$$

since the interval J_p had five times the length of I_p.

But we also have that $I_{\mathbf{x}} \cap I_p \neq \emptyset$, by (75) again, which implies that

$$\operatorname{dist}(\mathbf{x}, I_p) \leq \frac{m(I_{\mathbf{x}})}{2} < m(I_p), \tag{79}$$

where we used (77) in the last equality.

Putting (78) and (79) together gives

$$2m(I_p) < \operatorname{dist}(\mathbf{x}, I_p) < m(I_p)$$

which is clearly a contradiction. □

Corollary 10.1. *Let $\mathcal{F}$ be a Vitali covering of $S \subset [a,b]$. Then, for every, $\epsilon > 0$, there exists a finite set $I_1, I_2, ..., I_N \in \mathcal{F}$ of disjoint intervals (that are disjoint in $[a,b]$) such that*

$$m\left(S \setminus \cup_{n=1}^{N} I_n\right) < \epsilon.$$

Proof: Let I_n, $n = 1, 2, 3, ...$, be the covering from the Theorem, we also assume that $I_n \subset [a,b]$ as we did in the proof of the theorem. Then

$$m\left(\cup_{n=1}^{\infty} I_n\right) = \sum_{n=1}^{\infty} m(I_n) \leq b - a,$$

which implies that $\sum_{n=1}^{\infty} m(I_n)$ is (absolutely) convergent.

We may therefore choose N so that

$$\sum_{n=N+1}^{\infty} m(I_n) < \epsilon.$$

Since

$$S \setminus \bigcup_{n=1}^{N} I_n \subset \left(S \setminus \bigcup_{n=1}^{\infty} I_n\right) \bigcup \left(\bigcup_{n=N+1}^{\infty} I_n\right)$$

it follows that

$$m\left(S \setminus \bigcup_{n=1}^{N} I_n\right) = m\left(S \setminus \bigcup_{n=1}^{\infty} I_n\right) + m\left(\bigcup_{n=N+1}^{\infty} I_n\right) < 0 + \epsilon,$$

since where we used the Theorem to conclude that $S \setminus \cup_{n=1}^{\infty} I_n$ is a null set. $\Box$

We are now ready to show that if f is non-decreasing then $f'(x)$ exists a.e.

Theorem 10.3. *Let $f(x)$ be non-decreasing on $[a,b]$ then $f'(x)$ exists for a.e. $x \in [a,b]$.*

Proof: The derivative exists if $D^+ f(x) = D_+ f(x) = D^- f(x) = D_- f(x)$. We therefore need to show that $D^+ f(x) = D_+ f(x)$, $D^+ f(x) = D^- f(x)$ and $D^- f(x) = D_- f(x)$ for a.e. $x \in [a,b]$. Since the proof of these statements are very similar we will only prove that $D^+ f(x) = D_+ f(x)$ a.e. Since $D^+ f$ is defined by a limsup and $D_+ f$ by a liminf it follows directly that $D^+ f(x) \geq D_+ f(x)$, therefore we only need to prove that $D^+ f(x) \leq D_+ f(x)$ a.e.

The idea of the proof is that if $D^+ f(x) > D_+ f(x)$ then there exists two rational numbers $p, q \in \mathbb{Q}$ such that

$$D_+ f(x) < p < q < D^+ f(x). \tag{80}$$

We will denote the set where (80) is satisfied by

$$E_{pq} = \{x \in [a,b];\ D_+ f(x) < p < q < D^+ f(x)\}.$$

Since f is non-decreasing $D^+ f$ and $D_+ f$ are measurable, by Theorem 10.1, and therefore the set $E_{p,q}$ is measurable. We will show that $m(E_{pq}) = 0$ for every $p, q \in \mathbb{Q}$. Since

$$E = \{x;\ D_+ f(x) < D^+ f(x)\} = \bigcup_{p,q \in \mathbb{Q}} E_{pq},$$

and the union is countable, it follows that $D^+f = D_+f$ a.e. if E_{pq} is null for each $p, q \in \mathbb{Q}$.

Arguing by contradiction we may assume that there exists $p, q \in \mathbb{Q}$ such that

$$m(E_{pq}) = \alpha > 0 \text{ for some } p, q \in \mathbb{Q}.$$

Since E_{pq} is measurable there exists, for every $\epsilon > 0$, an open set U such that $E_{pq} \subset U$ and

$$\alpha = m(E_{pq}) \leq m(U) < m(E_{pq}) + \epsilon = \alpha + \epsilon. \tag{81}$$

If $x \in E_{pq}$ there exists arbitrary small intervals $[x, y]$ such that

$$\frac{f(y) - f(x)}{y - x} < p. \tag{82}$$

Since the intervals may be chosen arbitrarily small it follows that the intervals $[x, y]$ such that $x \in E_{pq}$ and (82) holds forms a Vitali cover $\mathcal{F}$ of E_{pq}. Since $E_{pq} \subset U$ and U is open we may arrange it so that $I \subset U$ for every $I \in \mathcal{F}$.

From Corollary 10.1 it follows that there is a finite set of intervals

$$[x_1, y_1], [x_2, y_2], ...[x_N, y_N] \in \mathcal{F}$$

such that

$$m\left(E_{p,q} \setminus \cup_{i=1}^N [x_i, y_i]\right) < \epsilon \tag{83}$$

and therefore

$$\sum_{i=1}^N (f(y_i) - f(x_i)) < p \sum_{i=1}^N (y_i - x_i) < pm(U) < p(\alpha + \epsilon), \tag{84}$$

where we have used that $[x_i, y_i] \subset U$ and (81).

Equation (84) gives a good estimate on the growth f from above. We need an estimate from below in order to derive a contradiction. To that end we consider $\hat{E}_{pq} = E_{pq} \cap \left(\cup_{i=1}^N (x_i, y_i)\right)$. For any point $u \in \hat{E} \subset E_{pq}$ we may find an arbitrary small interval $[u, v]$ and $[u, v] \subset \cup_{i=1}^N (x_i, y_i)$ such that

$$\frac{f(v) - f(u)}{v - u} > q.$$

Again these intervals forms a Vitali cover of $\hat{E}_{pq}$ and we may find a finite set $[u_1, v_1], [u_2, v_2], ..., [u_M, v_M]$ such that

$$m\left(\hat{E}_{pq} \setminus \cup_{i=1}^M [u_i, v_i]\right) < \epsilon \tag{85}$$

and

$$\sum_{i=1}^M (f(v_i) - f(u_i)) > q \sum_{i=1}^M (v_i - u_i). \tag{86}$$

Now assume that $[u_1, v_1], [u_2, v_2], ..., [u_{M_1}, u_{M_1}] \subset (x_1, y_1)$ then, since f is non-decreasing,

$$\sum_{i=1}^{M_1} (f(v_i) - f(u_i)) = -f(u_1) + \underbrace{f(v_1) - f(u_2)}_{\leq 0} + \underbrace{f(v_2) - f(u_3)}_{\leq 0} + ... + f(v_{M_1}) \leq$$

$$\leq f(v_{M_1}) - f(u_1) \leq f(y_1) - f(x_1), \tag{87}$$

similar estimates holds for the intervals

$$[u_{M_1+1}, v_{M_1+1}], [u_{M_1+2}, v_{M_1+2}], ..., [u_{M_1}, u_{M_2}] \subset (x_2, y_2).$$

From (84), (86) and (87) it follows that

$$q\sum_{i=1}^{M}(v_i - u_i) \leq \sum_{i=1}^{M}(f(v_i) - f(u_i)) \leq \sum_{i=1}^{N}(f(y_i) - f(x_i)) \leq p(\alpha + \epsilon). \tag{88}$$

But also

$$\sum_{i=1}^{M}(v_i - u_i) = m\left(\cup_{i=1}^{M}(u_i, v_i)\right) > m\left(\hat{E}_{pq}\right) - \epsilon > \alpha - 2\epsilon \tag{89}$$

where we used (85) and (83).

Equations (88) and (89) together implies that

$$q(\alpha - 2\epsilon) \leq p(\alpha + \epsilon).$$

Since $\alpha > 0$ and $\epsilon > 0$ is arbitrary this contradicts $p < q$. □

We are ready to prove the first version of the fundamental theorem of calculus, a version of the form (63) - but with an inequality.

Theorem 10.4. *Assume that f is bounded and non-decreasing on $[a, b]$ then f' is integrable*[23] *and*

$$\int_a^b f'(x)dx \leq f(b) - f(a)$$

Proof: In this proof we extend $f(x)$ to $[a, \infty)$ by $f(b)$; that is, $f(x) = f(b)$ for $x > b$.

We let f_n be defined

$$f_n(x) = \frac{f(x + 1/n) - f(x)}{1/n} \geq 0,$$

since f is non-decreasing. Then, since f' exists a.e. (Theorem 10.3)

$$\lim_{n\to\infty} f_n(x) = f'(x)$$

point-wise for a.e. $x \in [a, b]$. Also f' is measurable by Theroem 10.1 since $f'(x) = D^+f(x)$ a.e.

By Fatou's Lemma

$$\liminf_{n\to\infty} \int_a^b f_n(x)dx \geq \int_a^b \liminf_{n\to\infty} f_n(x)dx = \int_a^b f'(x)dx. \tag{90}$$

It follows (from (90)) that it is enough to show that

$$\liminf_{n\to\infty} \int_a^b f_n(x)dx \leq f(b) - f(a). \tag{91}$$

[23] We do not assume that f' is pointwise defined or finite, however f' will be defined and finite a.e. by Theorem 10.3.

We calculate

$$\liminf_{n\to\infty}\int_a^b \frac{f(x+1/n)-f(x)}{1/n}dx = \liminf_{n\to\infty} n\left[\int_a^b f(x+1/n)dx - \int_a^b f(x)dx\right] =$$

$$= \liminf_{n\to\infty} n\Bigg[\int_b^{b+1/n} \underbrace{f(x)}_{=f(b)}\, dx - \int_a^{a+1/n} \underbrace{f(x)}_{\leq f(a)}\, dx\Bigg] \leq \tag{92}$$

$$\leq \liminf_{n\to\infty} n\Bigg[\int_b^{b+1/n} f(b)dx - \int_a^{a+1/n} f(a)dx\Bigg] = f(b)-f(a),$$

where we used that $f(x)=f(b)$ for $x \geq b$ and that $f(x)$ is non-decreasing. We ahve already concluded, in (91), that (92) was enough to prove the theorem. □

A rather annoying feature of Theorem 10.4 is that we get an inequality in the conclusion. This is however not an artifact of a bad proof or a flaw in the theory. The following examples shows that the inequality is necessary.

Example 1: Let $f(x) = \begin{cases} 0 & \text{if } x \in [-1,0] \\ 1 & \text{if } x \in (0,1]. \end{cases}$ Then $f'(x) = 0$ a.e. and therefore

$$\int_{-1}^1 f'(x)dx = 0 < f(1)-f(0) = 1.$$

Example 2: It is not the discontinuity that makes fundamental theorem of calculus break down in the previous example. The cantor function also has derivative equal to zero a.e., and it is continuous, but the cantor function is still not constant.

10.1 Exercices

1. Prove that if $f(x)$ is non-decreasing then $f(x)$ is continuous except possibly on a countable set.
2. Prove that every non-decreasing function $f : [a,b] \mapsto \mathbb{R}$ is measurable.

11 Lecture 11. Differentiation and the Fundamental Theorem of Calculus, part 2.

We know that, even for monotone and thus a.e. differentiable functions, we do not get an equality in

$$\int_a^b f(x)dx \leq f(b)-f(a).$$

One would hope that if $f \in L^1(a,b)$ then

$$F(x) = \int_a^x f(t)dt \Rightarrow F'(x) = f(x) \text{ for } a.e.\, x \in [a,b].$$

This is indeed true and it will be our next aim to prove this. First we show that integrals define continuous functions.

Lemma 11.1. *Let $f \in L^1([a,b])$ and $F(x) = \int_a^x f(t)dt$. Then $F(x)$ is continuous.*

Proof: We know that for every $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that if Σ is measurable and $m(\Sigma) < \delta_\epsilon$ then

$$\int_\Sigma |f(t)|dt < \epsilon.$$

It follows that if $|x-y| < \delta_\epsilon$ then

$$|F(x) - F(y)| = \left|\int_x^y f(t)dt\right| < \epsilon,$$

therefore $F(x)$ is continuous.[24] □

We need the following proposition.

Proposition 11.1. *Let $f \in L^1(a,b)$ and assume that*

$$F(x) = \int_a^x f(t)dt = 0$$

for $x \in [a,b]$. Then $f(x) = 0$ for a.e. $x \in (a,b)$.

Proof: We will show that $f(x) \leq 0$ a.e., a similar argument shows that $f(x) \geq 0$ a.e. as well and the proposition follows.

It is enough to show that

$$S_n = \left\{x;\ f(x) \geq \frac{1}{n}\right\}$$

has measure zero for every $n \in \mathbb{N}$ since the countable family of sets S_n covers the set where $f(x) > 0$.

We argue by contradiction and assume that $m(S_n) = \delta > 0$ for some $n \in \mathbb{N}$ and δ. S_n is measurable (since f is) and therefore there exists a closed set $K \subset S_n$ such that $m(K) > \delta/2$.

Consider the open set $U = (a,b) \setminus K$. Since U is open we may cover U by a countable union of disjoint intervals (a_i, b_i). It follows that

$$0 = \int_a^b f(t)dt = \int_K f(t)dt + \int_U f(t)dt = \int_K f(t)dt + \sum_{i=1}^\infty \int_{a_i}^{b_i} f(t)dt =$$

$$= \int_K f(t)dt + \sum_{i=1}^\infty \underbrace{\left[\int_a^{b_i} f(t)dt - \int_a^{a_i} f(t)dt\right]}_{=0} = \int_K f(t)dt \geq \frac{1}{n}\int_K dt > 0.$$

This is clearly a contradiction. □

[24] This even shows that F is uniformly continuous. A minor modification of the argument will give that F is absolutely continuous - but that is a concept of continuity that we will wait a few pages before we define.

Theorem 11.1. *Assume that* $f \in L^1([a,b])$ *and*

$$F(x) = \int_a^x f(t)dt.$$

Then $F'(x) = f(x)$ *for a.e.* $x \in [a,b]$*. Furthermore*

$$F(x) - F(a) = \int_a^x F'(t)dt. \tag{93}$$

Proof: We will prove the theorem in several steps (under various simplifying assumptions).

Step 1: *The Theorem holds under the extra assumption that* f *is non-negative and bounded:* $0 \leq f(x) \leq M$.

Proof of Step 1: If $f(x) \geq 0$ then $F(x)$ is clearly non-decreasing and therefore differentiable (Theorem 10.3) at a.e. point.

We will argue as in Theorem 10.4 and denote

$$F_n(x) = \frac{F(x+1/n) - F(x)}{1/n} = n \int_x^{x+1/n} f(t)dt \leq M,$$

where we used the extra assumption that $f(x) \leq M$ in the last inequality, we also extend $f(x)$ so that $f(x) = f(b)$ for $x > b$. We also note that $F_n \geq 0$ (since F is non decreasing) and that $F_n(x) \to F'(x)$ at every point where F is differentiable.

By the dominated convergence theorem[25] (in the second inequality) we may conclude that

$$\int_a^x F'(t)dt = \int_a^x \lim_{n\to\infty} F_n(t)dt = \lim_{n\to\infty} \int_a^x F_n(t)dt =$$

$$= \lim_{n\to\infty} n \left[\int_a^x F(t+1/n)dt - \int_a^x F_n(t)dt \right] =$$

$$= \lim_{n\to\infty} n \left[\int_x^{x+1/n} F(t+1/n)dt - \int_a^{a+1/n} F_n(t)dt \right] =$$

$$= \lim_{n\to\infty} \left[F(x+\xi_x/n) - F(a+\xi_a/n) \right] = F(x) - F(a),$$

where we used the mean value theorem for the integral (valid since $F \in C([a,b])$ by Lemma 11.1) and $\xi_x, \xi_a \in (0,1)$ (and may depend on n) in the next to last equality and continuity of F in the last.

We may conclude that, for every $x \in [a,b]$,

$$\int_a^x \left[f(t) - F'(t) \right] dt = 0.$$

It follows, from Proposition 11.1, that $F'(x) = f(x)$ a.e.

[25]Notice that the argument is almost exactly the same as in Theorem 10.4 but we use the dominated convergence theorem instead of Fatou's Lemma which gives us an equality instead of an inequality.

Step 2: *The Theorem holds under the extra assumption that f is non-negative: $0 \leq f(x)$.*

Proof of Step 2: We may approximate f by $f_s(x) = \min(f(x), s)$, for any $s \geq 0$, then f_s satisfies the assumptions of Step 1. Also $\lim_{s\to\infty} f_s(x) = f(x)$. Therefore, by the dominated convergence theorem,

$$F(x) = \int_a^x f(t)dt = \int_a^x \lim_{s\to\infty} f_s(t)dt = \lim_{s\to\infty} \int_a^x f_s(t)dt. \tag{94}$$

If we write

$$F_s(x) = \int_a^x f_s(t)dt$$

then $F(x) - F_s(x) \geq 0$ and for $y > x$

$$F(x) - F_s(x) = \int_a^x \underbrace{f(t) - f_s(t)}_{\geq 0} dt \leq \int_a^y [f(t) - f_s(t)]\, dt = F(y) - F_s(y)$$

which implies that $F(x) - F_s(x)$ is non-decreasing.

From Theorem 10.4 it follows that (where we have used that $F(a) = F_s(a) = 0$)

$$F(x) - F_s(x) \geq \int_a^x [F'(t) - F_s'(t)]\, dt \geq 0, \tag{95}$$

where we used that $F(x) - F_s(x)$ is non-decreasing, and therefore have non-negative derivative, in the last inequality. Passing to the limit $s \to \infty$, and using that $F_s'(t) = f_s(t) \to f(t)$, in (95) we conclude that

$$0 \geq \int_a^x [F'(t) - f(t)]\, dt \geq 0, \tag{96}$$

for any $x \in [a, b]$. It follows from Proposition 11.1 that $F'(x) = f(x)$ a.e. The conclusion (93) is also a direct consequence of (96) and the definition of $F(x)$.

Step 3: *The Theorem holds without extra assumptions.*

Proof of Step 3: We may write $F(x) = F^+(x) - F^-(x)$ where

$$F^{\pm}(x) = \int_a^x \max(\pm f(t), 0)dt.$$

Then Step 2 applies to both F^+ and F^-. It follows that

$$F'(x) = DF^+(x) - DF^-F(x) = \max(f(t), 0) - \max(-f(x), 0) = f(x)$$

at a.e. $x \in [a, b]$. □

We have now shown the first version of the fundamental theorem of calculus that if $f \in L^1$ then $\frac{\partial}{\partial x} \int_a^x f(t)dt = f(x)$ a.e. But we would like to find conditions on f that also assures that

$$f(b) - f(a) = \int_a^b f'(t)dt. \tag{97}$$

We know that (97) does not hold in general, even if $f'(x)$ is defined a.e. Therefore we need to make some extra assumption on $f(x)$ in order for (97) to hold for f. The idea is that using the Vitali covering Theorem we can control the variation (total change) of an a.e. differentiable function f on a set $\cup_{i=1}^{N} I_n$ where the measure of the complement of $\cup_{i=1}^{N} I_n$ is small: $m\left([a,b]\setminus \cup_{i=1}^{N} I_n\right)$. However, in general, as with the Cantor function, if the derivative is infinite on a set of zero measure the formula in (97) might break down. Therefore, we need make some extra assumption that assures that f does not oscillate to much on small sets.

The correct assumption we need to derive (97) is absolute continuity. But it will require very little extra work to first define the weaker concept of bounded variation.

Definition 11.1. *Let f be defined on $[a,b]$ then we say that f is of bounded variation if*

$$V(f,[a,b]) = \sup_P \sum_{i=1}^{N} |f(a_i) - f(f(a_{i-1})| < \infty,$$

where the supremum is taken over all finite partitions

$$P = \{a = a_0 < a_1 < a_2 < .... < a_N = b\}$$

of the interval $[a,b]$ (the supremum is also over all $N \in \mathbb{N}$).

Proposition 11.2. *A function f is of bounded variation on $[a,b]$ if and only if there exists two finite and non-decreasing functions $g,h : [a,b] \mapsto \mathbb{R}$ such that $f(x) = g(x) - h(x)$.*

Proof: If $f(x) = g(x) - h(x)$, with g,h bounded and non decreasing then for each partition P

$$\sum_{i=1}^{N} |f(a_i) - f(a_{i-1})| = \sum_{i=1}^{N} |g(a_i) - g(a_{i-1}) - h(a_i) + h(a_{i-1})| \leq$$

$$\leq \sum_{i=1}^{N} |g(a_i) - g(a_{i-1})| + \sum_{i=1}^{N} |h(a_i) - h(a_{i-1})| =$$

$$= \sum_{i=1}^{N} g(a_i) - g(a_{i-1}) + \sum_{i=1}^{N} h(a_i) - h(a_{i-1}) =$$

$$= g(b) - g(a) + h(b) - h(a) \leq 2\sup_{[a,b]} |g(x)| + 2\sup_{[a,b]} |h(x)|,$$

where we used that f,g are non-decreasing in the second equality. Since g,h are bounded it follows that the variation $V(f,[a,b])$ is bounded.

To show the converse we assume that f is of bounded variation and define

$$g(x) = V(f,[a,x]).$$

Clearly g is then increasing on $[a,b]$, $g(0) = 0$ and $g(b) = V(f,[a,b]) < \infty$ so g is bounded.

Next we define $h(x) = g(x) - f(x)$ then, for $y \geq x$,

$$h(y) - h(x) = V(f,[a,y]) - f(y) - V(f,[a,x]) + f(x) = \tag{98}$$

$$= V(f,[x,y]) - f(y) + f(x) \geq |f(y) - f(x)| - f(y) + f(x) \geq 0,$$

where we used that $V(f,[a,y]) - V(f,[a,x]) = V(f,[x,y])$ which follows directly from the definition of partition[26] and that $V(f,[x,y]) \geq |f(y) - f(x)|$ which is easily seen by choosing the trivial partition $P = \{x = a_0 < a_1 = b\}$ in the supremum in Definition 11.1. It follows from (98) that h is non-decreasing. That h is bounded follows from the fact that f and g are (note that $|f(x)| \leq f(a) + V(f,[a,b])$).

Therefore both g and h are non-decreasing and bounded and

$$g(x) - h(x) = V(f,[a,x]) - (V(f,[a,x]) - f(x)) = f(x).$$

This finishes the proof. □

Corollary 11.1. *If $f(x)$ is of bounded variation on $[a,b]$ then $f'(x)$ exists a.e. and $f'(x)$ is integrable on $[a,b]$.*

Proof: By the previous proposition we may write $f(x) = g(x) - h(x)$ where $g(x)$ and $h(x)$ are non-decreasing bounded functions. It follows, from Theorem 10.4, that g' and h' are both integrable. The corollary follows from the linearity of the derivative and that the sum of two integrable functions is integrable. □

Remark: *We may not derive any information of the relation between f' and f, as in Theorem 10.4, for functions of bounded variation.*

We are now ready to define the concept, absolute continuity, that will assure that the formula (97) holds.

Definition 11.2. *We say that a function $f : [a,b] \mapsto \mathbb{R}$ is absolutely continuous if, for every $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that for any disjoint set of intervals*

$$[a_1,b_1],[a_2,b_2],....,[a_N,b_N] \subset [a,b]$$

such that

$$\sum_{i=1}^{N} |b_i - a_i| < \delta_\epsilon$$

it follows that

$$\sum_{i=1}^{N} |f(b_i) - f(a_i)| < \epsilon.$$

The concept of absolute continuity assures that the variation of f on small sets is small. This in particular excludes Cantor function like behaviors; that functions may change by positive quantities on sets of measure zero. Also, integrals of $L^1([a,b])$ functions are absolutely continuous.

Proposition 11.3. *Let $f \in L^1([a,b])$ then*

$$F(x) = \int_a^x f(t)dt$$

is absolutely continuous.

[26] For any two partitions $P_{[a,x]}$ and $P_{[x,y]}$ of $[a,x]$ and $[x,y]$ we may form a partition of $[a,y]$ that includes all the points in $P_{[a,x]}$ and $P_{[x,y]}$ plugging this into the definition of bounded variation will imply that $V(f,[a,x]) + V(f,[x,y]) \leq V(f,[a,y])$. And for any partition P of $[a,y]$ we may insert the point x in P and then split it into two partitions for $[a,x]$ and $[x,y]$ respectively, using these partitions in the definition of V it also follows that $V(f,[a,x]) + V(f,[x,y]) \geq V(f,[a,y])$. We leave the details to the reader.

Proof: The proof is almost the same as the proof of Lemma 11.1. We know that for every $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that if Σ is measurable and $m(\Sigma) < \delta_\epsilon$ then

$$\int_\Sigma |f(t)|dt < \epsilon.$$

It follows that if

$$\sum_{i=1}^{N} |b_i - a_i| < \delta_\epsilon$$

then, with $\sigma = \cup_{i=1}^{N}[a_i, b_i]$,

$$\sum_{i=1}^{N} |f(b_i) - f(a_i)| = \left|\sum_{i=1}^{N} \int_{a_i}^{b_i} f(t)dt\right| = \left|\int_\Sigma f(t)dt\right| < \epsilon,$$

therefore $F(x)$ is absolutely continuous. □

Also absolutely continuous functions are of bounded variation.

Proposition 11.4. *If $f : [a, b] \mapsto \mathbb{R}$ is absolutely continuous then f is of bounded variation. In particular, $f'(x)$ exists a.e. and $f'(x)$ is integrable on $[a, b]$.*

Proof: Let $P = \{a = a_0 < a_1 < .... < a_n = b\}$ be any partition and let $\delta_1 > 0$ be the δ_ϵ corresponding to $\epsilon = 1$ in Definition 11.2. We may refine the partition P by adding the points $a+\delta_1, a+2\delta_1, a+3\delta_1, ..., a+M\delta_1$, where M is the smallest integer such that $M \geq \frac{b-a}{\delta_1}$ (notice that M is a fixed constant depending only on f and $[a, b]$), to get a new partition $\hat{P} = \{a = b_0 < b_1 < ...b_N = b\}$.

It follows that

$$\sum_{i=1}^{n} |f(a_i) - f(a_{i-1})| \leq \sum_{i=1}^{N} |f(b_i) - f(b_{i-1})|. \tag{99}$$

We may split the last sum into M peaces each of the form $\sum_{i=k}^{l} |f(b_i) - f(b_{i-1})|$ where $b_{k-1} = m\delta_1$ and $b_l = (m + 1)\delta_1$. Using the absolute continuity of f and the choice of δ_1 as well as the continuity of f we may conclude that each of the M peaces of the sum satisfies

$$\sum_{i=k}^{l} |f(b_i) - f(b_{i-1})| \leq 1.$$

Using this in (99) it follows that

$$\sum_{i=1}^{n} |f(a_i) - f(a_{i-1})| \leq M.$$

But since the original partition P was arbitrary it follows that f is of bounded variation.

That f' exists a.e. and is integrable follows from Corollary 11.1. □

Next we show that absolutely continuous functions actually do not have Cantor function like behavior. The Proposition is very similar to Proposition 11.1.

Proposition 11.5. *Let $f : [a,b] \mapsto \mathbb{R}$ be absolutely continuous and $f'(x) = 0$ a.e. Then $f(x)$ is constant.*

Proof: We need to show that $f(c) = f(a)$ for every $c \in [a,b]$. Let $\epsilon > 0$ be arbitrary and $\delta_\epsilon > 0$ be as in Definition 11.2. Also let $Z_c = [a,c] \cap \{x \in [a,b];\ f'(x) = 0\}$ be the set of where the derivative is defined and zero.

Since $f'(x) = 0$ on Z_c there exists an arbitrarily small interval $[x,y]$ s.t.

$$\left|\frac{f(y)-f(x)}{y-x}\right| < \epsilon,$$

that is

$$|f(y)-f(x)| < \epsilon|y-x|. \tag{100}$$

Since we may choose $y > x$ arbitrarily close to x the intervals $[x,y]$, for $x \in Z_c$, forms a Vital coveri $\mathcal{F}$ of Z_c. We may therefore find disjoint intervals

$$[x_1,y_1],[x_2,y_2],[x_3,y_3],...,[x_N,y_N] \in \mathcal{F}$$

such that

$$m\left(Z_c \setminus \cup_{i=1}^N [x_i,y_i]\right) < \delta_\epsilon. \tag{101}$$

We may use these intervals to form a partition of $[a,c]$:[27]

$$P = \{a < x_1 < y_1 < x_2 < y_2 < x_3 < ... < x_n < y_N < c\}.$$

Using this partition we may estimate

$$|f(c)-f(a)| \leq |f(x_1)-f(a)| + \sum_{i=1}^N |f(y_i)-f(x_i)| + \tag{102}$$

$$+\sum_{i=1}^{N-1} |f(x_{i+1})-f(y_i)| + |f(c)-f(y_N)|.$$

From (100) we can conclude that

$$\sum_{i=1}^N |f(y_i)-f(x_i)| < \epsilon \sum_{i=1}^N |y_i - x_i| \leq \epsilon(c-a), \tag{103}$$

and from (101) we may conclude that

$$|x_1 - a| + \sum_{i=1}^{N-1} |x_{i+1}-y_i| + |c-y_N| < \delta_\epsilon \tag{104}$$

which implies that we may use the absolute continuity of f to deduce that

$$|f(x_1)-f(a)| + \sum_{i=1}^{N-1} |f(x_{i+1})-f(y_i)| + |f(c)-f(y_N)| < \epsilon. \tag{105}$$

Using the estimates (103) and (105) in (102) we may conclude that

$$|f(c)-f(a)| < \epsilon(c-a+1).$$

Since $\epsilon > 0$ is arbitrary it follows that $f(c) = f(a)$ but $c \in [a,b]$ was arbitrary so it follows that $f(x) = f(a)$ for all $x \in [a,b]$. □

[27] We have, for definiteness assumed that $a < x_1$ and $y_N < c$, in case one or both of these are equalities the argument is very similar.

Theorem 11.2. *If $f : [a, b] \mapsto \mathbb{R}$ is absolutely continuous then $f'(x)$ is defined for a.e. $x \in [a, b]$, $f'(x)$ is integrable and*

$$f(x) - f(a) = \int_a^x f'(t)dt. \tag{106}$$

Proof: That f' is defined a.e. and integrable follows from Proposition 11.4. Therefore we only need to prove (106). We define the function

$$g(x) = \int_a^x f'(t)dt - f(x).$$

Since f' is integrable the function $g(x)$ is well defined. By Proposition 11.3 it follows that $g(x)$ is absolutely continuous.[28]

By Theorem 11.1,

$$g'(x) = f'(x) - f'(x) = 0 \quad \text{for a.e. } x \in [a, b]. \tag{107}$$

Proposition 11.5 together with (11.1) and the absolute continuity of g implies that

$$g(x) = \int_a^x f'(t)dt - f(x) = \text{ constant } = g(a) = f(a).$$

The theorem follows from the last equalities. □

12 Lecture 12. Classical derivatives and Sobolev functions.

In this lecture we will investigate to what extent Sobolev functions have classical derivatives. The first step consists in proving that Sobolev functions are absolutely continuous on a.e. line. This would imply that the fundamental theorem of calculus is applicable on a.e. line. We begin with a simple lemma in the one dimensional case.

Lemma 12.1. *Assume that $f_j(x) \in C^\infty(a, b)$, that $\|f_j'\|_{L^p(a,b)} \leq C$ for some constant C (independent of j) and $p \geq 1$ and that $f_j \to f_0$ in $L^1(a, b)$. Then $f_0 \in W^{1,p}(a, b)$ and f_0 is absolutely continuous.*

Proof: Since $f_j \in C^\infty$ the fundamental theorem of calculus holds and we may conclude that

$$|f_j(x_1) - f_j(x_0)| = \left| \int_{x_0}^{x_1} f_j'(x)dx \right| \leq$$

$$\leq |x_1 - x_0|^{1-1/p} \|f_j'\|_{L^p(x_0,x_1)} \leq C|x_1 - x_0|^{1-1/p},$$

where we used Hölders inequality. It follows that the sequence f_j is equicontinuous and thus, by Arzela-Ascoli's Theorem we may extract a subsequence $f_{j_k} \to f_0$ uniformly where f_0 is a continuous function on (a, b).

[28] We also use that the sum of two absolutely continuous functions is absolutely continuous here. But that is a direct consequence of the triangle inequality and the definition of absolute continuity.

It remains to show that f_0 is absolutely continuous. This follows from the fact that each f_j is absolutely continuous. In particular, for any set of disjoint intervals (α_i, β_i)

$$\sum_{i=1}^{N} |f_j(\beta_i) - f_j(\alpha_i)| = \sum_{i=1}^{N} \left| \int_{\alpha_i}^{\beta_i} f_j'(x)dx \right| \leq$$

$$\leq \int_{\cup_{i=1}^N (\alpha_i,\beta_i)} |f_j'(x)|dx \leq \left|\cup_{i=1}^N (\alpha_i, \beta_i)\right|^{1-1/p} \|f_j'\|_{L^p(a,b)} \leq$$

$$\leq C \left|\cup_{i=1}^N (\alpha_i, \beta_i)\right|^{1-1/p}.$$

Therefore, for any $\epsilon > 0$, if $\sum_{i=1}^N (\beta_i - \alpha_i) < \left(\frac{\epsilon}{C}\right)^{p/(p-1)}$ then

$$\sum_{i=1}^{N} |f_j(\beta_i) - f_j(\alpha_i)| < \epsilon.$$

It follows that each f_j is absolutely continuous. Since $f_j \to f_0$ uniformly it follows that f_0 is absolutely continuous. □

Corollary 12.1. *If $f \in W^{1,p}(a,b)$ then f is absolutely continuous.*

Proof: If $f_j(x) = \phi_{1/j} * f(x)$, where $\phi_{1/j}$ is the standard mollifier, then f_j and f satisfies the conditions of Lemma 12.1. □

Theorem 12.1. *Let $u \in W^{1,p}(\mathcal{D})$ where $\mathcal{D} \subset \mathbb{R}^n$. Then for every $k = 1, 2, ..., n$ and for a.e. $(x_1, x_2, ..., x_{k-1}, x_{k+1}, ..., x_n)$ (with respect to the $(n-1)-$ dimensional Lebesgue measure) the function $u(x_1, x_2, ..., x_{k-1}, t, x_{k+1}, ..., x_n)$ is absolutely continuous in t.*

Proof: Let us fix a k, say $k = 1$. Consider $u_\epsilon(\mathbf{x})$ defined as in as in Theorem 9.1 then $u_\epsilon \to u$ as $\epsilon \to 0^+$. This implies that for any $\delta > 0$ the sequence $\hat{u}_j = u_{\delta^{2/p}/2^{j/p}}$ converges to u in $W^{1,p}$, we even have that $\|u - \hat{u}_j\|^p_{W^{1,p}(\mathcal{D})} \leq \delta^2/2^j$.

By Fubini's theorem

$$\int \left(\int |\nabla(u(\mathbf{x}) - \hat{u}_j(\mathbf{x}))|^p dx_1 \right) d\mathbf{x}' = \|\nabla(u - \hat{u}_j)\|^p_{L^p(\mathcal{D})} < \delta^2/2^j, \tag{108}$$

where the integration to the left is also over $\mathcal{D}$ and $\mathbf{x}' = (x_2, x_3, ..., x_n)$. From (108) we can conclude that the set

$$A_{\delta,j} = \left\{ \mathbf{x}';\ \int_{(x_1,\mathbf{x}')\in\mathcal{D}} \left| \frac{\partial u(x_1, \mathbf{x}') - \hat{u}_j(x_1, \mathbf{x}')}{\partial x_1} dx_1 \right|^p dx_1 \geq \delta \right\}$$

has measure $(n-1)-$dimensional Lebesgue measure less than $\delta/2^j$. This implies that the set of x_1 where

$$\int_{(x_1,\mathbf{x}')\in\mathcal{D}} \left| \frac{\partial u(x_1, \mathbf{x}') - \hat{u}_j(x_1, \mathbf{x}')}{\partial x_1} dx_1 \right|^p dx_1 \geq \delta$$

for some j must have measure less than δ:

$$\left|\cup_{j=1}^\infty A_{\delta,j}\right| \leq \sum_{j=1}^{\infty} \frac{\delta}{2^j} \leq \delta.$$

To exploit this we fix some decreasing sequence $\delta_j \to 0^+$ and define $u_j(\mathbf{x}) = u_{\delta_j^{2/p}/2^{j/p}}$. By the above it follows that for any $\epsilon > 0$ there exists a $j_\epsilon > 0$ such that the compliment of the set of points $(x_2, x_3, ..., x_n)$ where

$$\int_{(x_1,\mathbf{x}')\in\mathcal{D}} \left|\frac{\partial u(x_1,\mathbf{x}') - u_j(x_1,\mathbf{x}')}{\partial x_1} dx_1\right|^p dx_1 \geq \epsilon$$

for some $j > j_\epsilon$ has measure less than ϵ; just pick j_ϵ as the smallest j such that $\delta_j < \epsilon$.

But this implies that

$$\int_{(x_1,\mathbf{x}')\in\mathcal{D}} \left|\frac{\partial u(x_1,\mathbf{x}') - u_j(x_1,\mathbf{x}')}{\partial x_1} dx_1\right|^p dx_1 \to 0$$

except on a set of zero measure. By Lemma 12.1 it follows that $u(t, x_2, ..., x_n)$ is absolutely continuous for a.e. $(x_2, ..., x_n)$. □

One final way we will indicate that the definition of weak derivative makes sense is to show that we may actually define the weak derivative by means of difference quotients. However, since $W^{1,p}-$functions may be discontinuous, we cannot take the difference quotients point-wise. Instead we will prove that if the difference quotients exists, in an integral sense, for a function u then the function is in a Sobolev space.

Lemma 12.2. *Let $\mathcal{C} \subset \mathcal{D}$ be a compact set such that $\tilde{\mathcal{C}}_\delta = \{x;\ dist(\boldsymbol{x}, \mathcal{C}) < \delta\} \subset \mathcal{D}$.*

Furthermore assume that $u(\boldsymbol{x}) \in L^p(\mathcal{D})$, $1 < p < \infty$, and that there exists a constant C such that

$$\int_{\tilde{\mathcal{C}}_\delta} \left|\frac{u(\boldsymbol{x} + e_i h) - u(\boldsymbol{x})}{h}\right|^p dx \leq C, \tag{109}$$

for all $|h| < \delta$.

Then the weak derivative $\frac{\partial u}{\partial x_i}$ exists in $\mathcal{C}$ and

$$\int_{\mathcal{C}} \left|\frac{\partial u(\boldsymbol{x})}{\partial x_i}\right|^p dx \leq C.$$

In particular, if (109) holds for every $i = 1, 2, ..., n$ then $u \in W^{1,p}(\mathcal{C})$.

Proof: Notice that (109) just states that for any sequence $h_j \to 0$ the functions $\frac{u(\mathbf{x}+e_i h_j) - u(\mathbf{x})}{h_j}$ are bounded in $L^p(\tilde{\mathcal{C}}_\delta)$. Thus, by the weak compactness theorem for L^p-functions, Theorem 8.1, there exists a sub-sequence, still denoted h_j, such that

$$\frac{u(\mathbf{x} + e_i h_j) - u(\mathbf{x})}{h_j} \rightharpoonup g_i(\mathbf{x}) \in L^p(\tilde{\mathcal{C}}_\delta).$$

By Proposition 8.4 it follows that $\|g_i\|_{L^p(\tilde{\mathcal{C}}_\delta)} \leq C$.

We claim that $g_i(\mathbf{x})$ is the weak x_i-derivative of $u(\mathbf{x})$. To see this we calculate, for any $\phi \in C^1_c(\mathcal{D})$,

$$-\int_{\mathcal{C}} \frac{\partial\phi(\mathbf{x})}{\partial x_i} u(\mathbf{x}) dx = \lim_{h_j\to 0} -\int_{\mathcal{C}} \frac{\phi(\mathbf{x} + h_j e_i) - \phi(\mathbf{x})}{h_j} u(\mathbf{x}) dx =$$

$$= \left\{ \begin{array}{l} \text{Change of var.} \\ \mathbf{x} + h_j e_i \to \mathbf{x} \\ \text{in } \phi(\mathbf{x} + h_j e_i) \end{array} \right\} = \lim_{h_j \to 0} \int_{\mathbf{x} - e_i h_j \in \mathcal{C}} \phi(x) \frac{u(\mathbf{x}) - u(x - e_i h_j)}{h_j} dx \to$$

$$\to \int_{\mathcal{C}} \phi(\mathbf{x}) g_i(\mathbf{x}) dx,$$

by the weak convergence of $\frac{u(\mathbf{x})-u(\mathbf{x}-e_i h_j)}{h_j} \rightharpoonup g_i$.

This proves that $g_i(\mathbf{x}) = \frac{\partial u(\mathbf{x})}{dx_i}$. The final statement of the lemma follows from the definition of the Sobolev space. □

We are now in a very good position to prove existence of minimizers for the minimization problem stated in the first lecture. We have indeed constructed a good space, $W^{1,2}(\mathcal{D})$, that has the right compactness properties and also is dependent on a flexible enough definition of the derivative. There is one more issue that we need to resolve: the existence of boundary values. We can be quite sure that the domain $\mathcal{K}_E$ of the functional $E(u) = \int_{\mathcal{D}} |\nabla u|^2 d\mathbf{x}$ should be a subset of the Sobolev space $W^{1,2}(\mathcal{D})$. But we required the minimizer u to satisfy the boundary condition $u(\mathbf{x}) = f(\mathbf{x})$ on $\partial\mathcal{D}$ for some given function $f(\mathbf{x})$. Since functions $u \in W^{1,2}(\mathcal{D})$ are only defined up to sets of measure zero and the boundary $\mathcal{D}$ of most domains $\mathcal{D}$ has measure zero[29] it is not entirely clear how to prescribe boundary values to functions $u \in W^{1,p}(\mathcal{D})$. Let us consider some examples to better appreciate the difficulty to assign boundary values.

Example 1: *Let us begin with the most elementary example we can imagine. Let* $u(x) = \cos(1/x) \in C(0,1)$*. Then, even though* u *is continuous, there is no natural way to assign a value to* u *at the boundary point* $x = 0$*. There is certainly no continuous extension of* $u(x)$ *to* $[0,1)$*.*

But do we have the same problems for Sobolev functions?

Example 2: *Let* $u(\mathbf{x}) = \cos(\ln(|\mathbf{x}|))$ *in the punctured domain* $\mathcal{D} = B_1(0) \setminus \{0\}$ *in* $\mathbb{R}^3$*. Then* $|\nabla u(\mathbf{x})| = \frac{-\sin(\ln(|\mathbf{x}|))}{|\mathbf{x}|}$ *and therefore we may intergrate in polar coordinates*

$$\int_{\mathcal{D}} |\nabla u(\mathbf{x})|^2 d\mathbf{x} = 4\pi \int_0^1 \sin^2(\ln(r)) dr \leq 4\pi.$$

This implies that $u \in W^{1,2}(\mathcal{D})$*. But there is no reasonable way to ascribe a boundary value to the point* $0 \in \partial\mathcal{D}$ *since* u *will oscillate between* -1 *and* 1 *in every neighborhood of* 0*.*

Example 3: *Let us consider a final example* $u(\mathbf{x}) = \ln(|\mathbf{x}|)$ *in the punctured domain* $\mathcal{D} = B_1(0) \setminus \{0\}$ *in* $\mathbb{R}^3$*. Then again* $u \in W^{1,2}(\mathcal{D})$ *but* $\lim_{|\mathbf{x}| \to 0} u(\mathbf{x}) = \infty$ *which would impliy that the only natural candidate for a boundary value at* $0 \in \partial\mathcal{D}$ *would be* ∞*. Again we have to conclude that we are not able to assign boundary values to* $W^{1,2}-$*functions.*

In order to get around these problems we notice that if $u \in W^{1,p}(\mathcal{D})$ does not attain any boundary values at a point $\mathbf{x}^0 \in \mathcal{D}$ then: either u oscillates fast around $\mathbf{x}^0$ (as in Example 2) or $u \to \infty$ as $\mathbf{x} \to \mathbf{x}^0$ (as in Example 3). Of course, we cannot claim that this is the only thing that might happen just because it is the only examples we have, but since these are the only examples we know of we

[29]There are specially constructed domains $\mathcal{D}$ such that $\partial\mathcal{D}$ has positive measure. But these are very special domains that are mostly irrelevant in the calculus of variations.

may use them as a guide for our intuition. In any case, both fastly oscillating functions and functions tending to infinity must have large derivatives. So if a function is in $W^{1,p}$ we might conjecture that around most points the function does neither oscillate fast nor tend to infinity. Maybe we can use this intuition to prove that boundary values exists at almost every point, we will actually be able to define boundary values in an L^p-space.

One subtle thing relating to the existence of boundary data is that the existence of boundary data is related to the behavior of the boundary of the domain. In order to prove the next theorem we need the following definition.

Definition 12.1. *We say that a domain $\mathcal{D}$ has C^1 boundary if we can cover $\partial\mathcal{D}$ by finitely many balls $B_{r_j}(\boldsymbol{x}^j)$, $\boldsymbol{x}^j \in \mathcal{D}$, such that $\partial\mathcal{D} \cap B_{2r_j}(\boldsymbol{x}^j)$ is the graph of a C^1 function f^j in some coordinate system.*

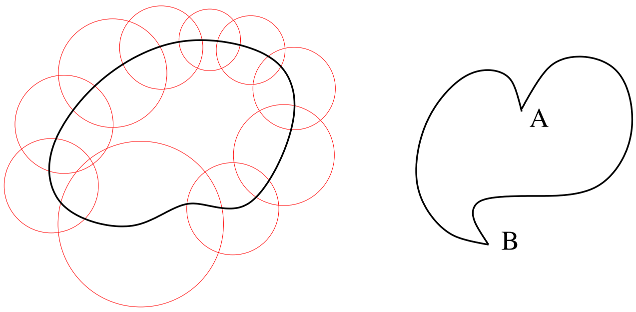


Figure. *The left domain is a C^1 domain since we may cover its boundary by balls (in red) such that the boundary is the graph if a C^1 function in each ball. The right domain, however, is not C^1 since there is no neighbourhood around the points A and B where the boundary is the graph of a C^1 function.*

Theorem 12.2. [Trace Theorem.] *Let $\mathcal{D}$ be a bounded domain with continuously differentiable boundary. Then there exists an operator*

$$T : W^{1,p}(\mathcal{D}) \mapsto L^p(\partial\mathcal{D})$$

that assigns boundary values (in the trace sense) of $u \in W^{1,2}(\mathcal{D})$ onto the boundary $\partial\mathcal{D}$. The operator T satisfies the following estimate

$$\|Tu\|_{L^2(\partial\mathcal{D})} \leq C\|u\|_{W^{1,p}(\mathcal{D})}, \tag{110}$$

where the constant depends on $\mathcal{D}$ and on p but not on u.

Furthermore $Tu = u\lfloor_{\partial\mathcal{D}}$ for all functions $u \in C(\overline{\mathcal{D}}) \cap W^{1,p}(\mathcal{D})$.

Proof:

Step 1 [Straightening of the boundary.]: *It is enough to prove that the operator $T : W^{1,p}(B_1^+(0)) \mapsto L^p(B_{3/4}(0) \cap \{x_n = 0\})$, where $B_1^+(0) = B_1(0) \cap \{x_n > 0\}$* exists.

Proof of Step 1: By definition a domain has C^1 boundary if we can cover $\partial\mathcal{D}$ by finitely many balls $B_{r_j}(x^j)$ such that $\partial\mathcal{D} \cap B_{2r_j}(x^j)$ is the graph of a C^1 function f^j in some coordinate system.

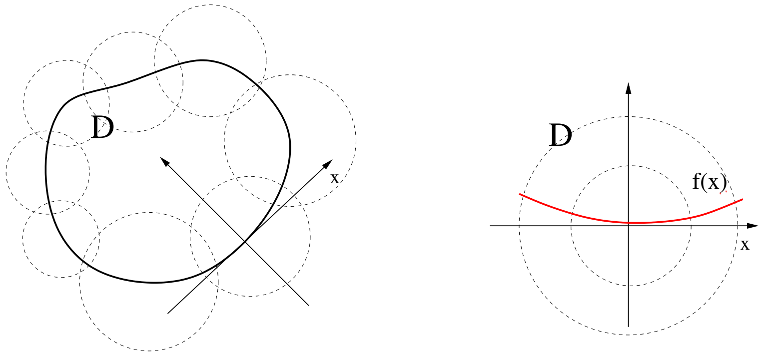


Figure: *The left figure shows a domain $\mathcal{D}$ with C^1 boundary. This means that we may cover $\partial\mathcal{D}$ by a finite number of balls $B_{r_j}(x^j)$ such that for each ball there is a coordinate system so that $\partial\mathcal{D}\cap B_{2r_j}(x^j)$ is a graph in the coordinate system. The right picture shows the same coordinate system rotated and the boundary portion $\partial\mathcal{D}\cap B_{2r_j}(x^j)$ (in red) which is clearly the graph of some function $f(x)$. We will change coordinates to straighten the red part of the boundary.*

The idea of the proof is that we may *"straighten the boundary"* in $B_{2r_j}(x)$ by defining the new coordinates $\hat{x}$ so that

$$(x_1,x_2,...,x_n)=(\hat{x}_1,\hat{x}_2,...,\hat{x}_n+f^j(\hat{x}'))$$
$$\Leftrightarrow$$
$$(\hat{x}_1,\hat{x}_2,...,\hat{x}_n)=(x_1,x_2,...,x_n-f^j(x')),$$

where $x'=\hat{x}'=(x_1,...,x_{n-1})$. Then the part of the boundary $x_n=f^j(x')$ will be mapped to the hyperplane $\hat{x}_n=0$ in the $\hat{x}$ coordinates. We may write the function $u(x)$ in these coordinates as $\hat{u}(\hat{x})$.

By the chain rule we get that

$$\frac{\partial\hat{u}(\hat{x})}{\partial\hat{x}_i}=\sum_{k=1}^n\frac{\partial x_k}{\partial\hat{x}_i}\frac{\partial u(x)}{\partial x_k}=\sum_{k=1}^n\left(\frac{\partial u(x)}{\partial x_k}+\frac{\partial f^j(x')}{\partial x_k}\frac{\partial u}{\partial x_n}\right).$$

In particular, $|\nabla\hat{u}(\hat{x})|$ will be comparable in size with $|\nabla u(x)|$.

Since $f(x)$ is continuously differentiable we can conclude that

$$\int_{\mathcal{D}\cap B_{2r}(x^0)}|\nabla\hat{u}(\hat{x})|^p d\hat{x}\le C\int_{\mathcal{D}\cap B_{2r}(x^0)}|\nabla u(x)|^p dx,$$

where the constant C only depend on the maximum value of $|\nabla' f(x')|$.

Notice that $\hat{u}(\hat{x})$ is defined in a set where part of the boundary is straight (in the $\hat{x}$ coordinates). If we can define boundary values for $\hat{u}$ on the straight part of the boundary then we can define boundary values of $u(x)$ on the portion of the boundary that lays in $B_{r_j}(x^j)$ by the equality $u(x)=\hat{u}(x',0)$. But the entire boundary $\partial\mathcal{D}$ can be covered by finitely many balls $R_{r_j}(x^j)$ so we can define boundary values for $u(x)$ on the entire boundary $\partial\mathcal{D}$.

Step 2: *Let* $u(x)\in W^{1,p}(B_2^+(0))$ *then*

$$\int_{B_1'(0)}|u(x',t)-u(x',s)|^p dx'\le|s-t|^{p-1}\|\nabla u(x)\|^p_{L^p(B_2^+)}$$

where $B_1'(0) = \{x' \in \mathbb{R}^{n-1}; |x'| \leq 1\}$ *is the unit ball in the* x' *coordinates.*

Proof of step 2: Using the fundamental theorem of calculus, which we may apply on $\mathbf{x}'$ almost every line, since $u(\mathbf{x}', t)$ is absolutely continuous on a.e. line (by Theorem 12.1),

$$\int_{B_1'(0)} |u(x',t) - u(x',s)|^p dx' = \int_{B_1'(0)} \left| \int_s^t \frac{\partial u(x)}{\partial x_n} dx_n \right|^p dx' \leq$$

$$\leq |s-t|^{p-1} \int_{B_1'(0)} \int_s^t \left| \frac{\partial u(x)}{\partial x_n} \right|^p dx' dx_n \leq |s-t|^{p-1} \|\nabla u(x)\|^p_{L^p(B_2^+)}$$

where we used the Hölder's inequality[30] in the first inequality.

Step 3: *Let* $u(x) \in W^{1,p}(B_2^+(0))$ *then the limit* $\lim_{t\to 0'} u(x',t) = u^0(x',0)$ *exists (and is therefore unique) in* $L^p(B_1'(0))$ *and the function* $u^0(x',0) \in L^p(B_1'(0))$*, where* $B_1'(0) = \{\mathbf{x}' \in \mathbb{R}^{n-1};\ |\mathbf{x}'| < 1\}$*, and satsifies the estimate*

$$\|u^0\|_{L^p(B_1'(0))} \leq C \|u\|_{W^{1,p}(B_2^+(0))} \tag{111}$$

where the constant C *does not depend on* u*.*

Proof of Step 3: Since $u(x) \in W^{1,p}(B_2^+(0))$ it follows that

$$\int_0^{1/4} \int_{B_1'(0)} |u(x)|^p dx' dx_n \leq \int_{B_2^+(0)} |u(x)|^p dx < \infty, \tag{112}$$

where we used $B_1'(0) \times (0, 1/4) \subset B_2^+(0)$ in the first inequality and the definition of $W^{1,p}(B_2^+(0))$ (see Definition 9.2) in the second inequality.

From (112) we can conclude that there exists an $s \in (0, 1/4)$ such that

$$\int_{B_1'(0)} |u(x',s)|^p dx' \leq 4 \int_{B_2^+(0)} |u(x)|^p dx.$$

This implies that $u(x',s) \in L^p(B_1'(0))$ and therefore, from step 2, that $u(x',t) \in L^p(B_1'(0))$.

By Step 2 the sequence of functions $u(x', s/j)$ will form a Cauchy sequence and is therefore convergent, in $L^p(B_1'(0))$ to some function $u^0(x',0) \in L^p(B_1'(0))$. Also, by step 2,

$$\|u(x', s/j) - u^0(x',0)\|_{L^p(B_1'(0))} \leq \left(\frac{s}{j}\right)^{1-1/p} \|\nabla u(x)\|_{L^p(B_2^+)}.$$

We only need to assure that $\lim_{t\to 0^+} u(x',t) = u^0(x',0)$. But that follows from step 2 and the triangle inequality:

$$\|u(x',t) - u^0(x',0)\|_{L^p(B_1'(0))} \leq \tag{113}$$

$$\leq \|u(x',t) - u^0(x', s/j)\|_{L^p(B_1'(0))} + \|u(x', s/j) - u^0(x',0)\|_{L^p(B_1'(0))} \leq \tag{114}$$

$$\left(\left(t - \frac{s}{j}\right)^{1-1/p} + \left(\frac{s}{j}\right)^{1-1/p} \right) \|\nabla u(x)\|_{L^p(B_2^+)}, \tag{115}$$

[30]With $g(x) = \frac{\partial u(x)}{\partial x_n}$ and $h(x) = 1$.

where we choose j so large that $s/j \leq t$. It clearly follows from (113)-(115) that

$$\|u(x',t) - u^0(x',0)\|_{L^p(B_1'(0))} \leq 2t^{1-1/p}\|\nabla u(x)\|_{L^p(B_2^+)}. \tag{116}$$

It follows that $\lim_{t\to 0'} u(x',t) = u^0(x',0)$ in $L^p(B_1'(0))$.

Step 4: *The trace operator satisfies the estimate (110).*

Proof of Step 4: The estimate (111) was derived for the function with straightened boundary. But changing back to the original domain, by $(x_1, x_2, ..., x_n) = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_n + f^j(\hat{x}'))$ in the notation of step 1, will directly imply that

$$\|u^0\|_{L^p(\partial\mathcal{D}\cap B_1(\mathbf{x}^j))} \leq C\|u\|_{W^{1,p}(\mathcal{D}\cap B_2(\mathbf{x}^j))},$$

where C only depend on the constant in (111) and on $\mathcal{D}$ (in particular on f^j through the change of variables formula) Since $\partial\mathcal{D}$ is covered by the balls $B_1(\mathbf{x}^j)$, for $j = 1, 2, ..., J$ where J is a finite number, we may estimate

$$\|u^0\|^p_{L^p(\partial\mathcal{D})} \leq \sum_{j=1}^{J} \|u^0\|^p_{L^p(\partial\mathcal{D}\cap B_1(\mathbf{x}^j))} \leq \sum_{j=1}^{J} C\|u\|^p_{W^{1,p}(\mathcal{D}\cap B_2(\mathbf{x}^j))} \leq JC\|u\|^p_{W^{1,p}(\mathcal{D})}.$$

Since J and C only depend on the domain $\mathcal{D}$ the inequality (110) follows.

We also claim that $Tu = u\lfloor_{\partial\mathcal{D}}$ if u is continuous up to the boundary, but this should be clear since $u(\mathbf{x}', t) \to Tu$ but if u is continuous up to the boundary then $u(\mathbf{x}', t) \to u\lfloor_{\partial\mathcal{D}}$. □

The previous theorem states that we may define values of $f \in W^{1,p}(\mathcal{D})$ on the boundary $\partial\mathcal{D}$, at least if $\mathcal{D}$ is C^1. However, that does not imply that the boundary values are preserved under weak limits, something we need in order to solve the variational problem. This leads us to some questions relating to the completeness and compactness properties of the Sobolev spaces $W^{1,p}$.

The next theorem is known as the Gagliardo-Nirenberg-Sobolev inequality.

Theorem 12.3. *Assume that $f \in W^{1,p}(\mathbb{R}^n)$, $1 \leq p < n$, then there exists a constant $C(n,p)$ such that*

$$\|u\|_{L^{p^*}(\mathbb{R}^n)} \leq C\|\nabla u\|_{L^p(\mathbb{R}^n)}$$

where $p^* = \frac{np}{n-p}$.

Proof:

Claim 1: *We may assume that $u \in C_c^\infty(\mathbb{R}^n)$.*

Proof of claim 1: If we manage to prove the theorem for $\eta_R(\mathbf{x})u_\epsilon(\mathbf{x})$ where $\eta_R(\mathbf{x})$ is a cut off function (that is $\eta \in C^\infty$, $\eta = 1$ in B_R and $\eta = 0$ outside $B_{2R}(0)$) and u_ϵ is the regularized version of u. Then, since $u_\epsilon \to u$ and $\nabla u_\epsilon \to 0$ the general result follows from first sending $R \to \infty$ and then $\epsilon \to 0$.

Claim 2: *The theorem holds for $p = 1$.*

Proof of claim 2: The proof follows from the fundamental theorem of calculus. In particular,

$$u(\mathbf{x}) = \int_{-\infty}^{x_i} \frac{\partial u(x_1, ..., x_{i-1}, t_i, x_{x+1}, ..., x_n)}{\partial x_i} dt_i. \tag{117}$$

Taking absolute values on both sides in (117) we may deduce that

$$
\begin{aligned}
|u(\mathbf{x})| &= \left|\int_{-\infty}^{x_i} \frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i} dt_i\right| \le \\
&\le \int_{-\infty}^{x_i} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i \le \\
&\le \int_{-\infty}^{\infty} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i.
\end{aligned}
\tag{118}
$$

Multiplying (118) for $i = 1,..,n$ and raising the resulting inequality to the power $\frac{1}{n-1}$ gives

$$
|u(\mathbf{x})|^{\frac{n}{n-1}} \le \prod_{i=1}^{n} \left(\int_{-\infty}^{\infty} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i\right)^{\frac{1}{n-1}}. \tag{119}
$$

The equation (119) gives good point-wise estimates of u but we need to integrate away the dependence of $\mathbf{x}$. So integrate both sides with respect to x_1, this is a well defined integral since u has compact support by claim 1,

$$
\begin{aligned}
\int_{\mathbb{R}} |u(\mathbf{x})|^{\frac{n}{n-1}} dx_1 &\le \int_{\mathbb{R}} \prod_{i=2}^{n} \left(\int_{-\infty}^{\infty} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i\right)^{\frac{1}{n-1}} dx_1 = \\
&= \left(\int_{\mathbb{R}} \left|\frac{\partial u(t_1,x_2,...,x_n)}{\partial x_1}\right| dt_1\right)^{\frac{1}{n-1}} \times \\
&\times \int_{\mathbb{R}} \prod_{i=2}^{n} \left(\int_{-\infty}^{\infty} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i\right)^{\frac{1}{n-1}} dx_1,
\end{aligned}
\tag{120}
$$

where we used that the first integral is independent of x_1. Notice that

$$
\begin{aligned}
&\int_{\mathbb{R}} \prod_{i=2}^{n} \left(\int_{-\infty}^{\infty} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i\right)^{\frac{1}{n-1}} dx_1 \le \\
&\le \left(\prod_{i=2}^{n} \int_{\mathbb{R}} \int_{-\infty}^{\infty} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i dx_1\right)^{\frac{1}{n-1}},
\end{aligned}
\tag{121}
$$

by the generalized Hölder inequality.[31] From (120) and (121) we may conclude that

$$
\begin{aligned}
\int_{\mathbb{R}} |u(\mathbf{x})|^{\frac{n}{n-1}} dx_1 &\le \left(\int_{\mathbb{R}} \left|\frac{\partial u(t_1,x_2,...,x_n)}{\partial x_1}\right| dt_1\right)^{\frac{1}{n-1}} \times \\
&\times \left(\prod_{i=2}^{n} \int_{\mathbb{R}} \int_{-\infty}^{\infty} \left|\frac{\partial u(x_1,...,x_{i-1},t_i,x_{x+1},...,x_n)}{\partial x_i}\right| dt_i dx_1\right)^{\frac{1}{n-1}}.
\end{aligned}
\tag{122}
$$

[31] The generalized Hölder inequality, see exercise 4, states that if $\sum_{j=1}^{n} \frac{1}{p_j} = 1$ then

$$
\int_{\mathcal{D}} u_1(\mathbf{x})u_2(\mathbf{x})\dots u_n(\mathbf{x})dx \le \prod_{j=1}^{n} \|u_j\|_{L^{p_j}(\mathcal{D})}.
$$

Integrating the (122) over x_2 gives

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |u(\mathbf{x})|^{\frac{n}{n-1}} dx_1 dx_2 \leq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left|\frac{\partial u(t_1, x_2, ..., x_n)}{\partial x_1}\right| dt_1\right)^{\frac{1}{n-1}} \times$$

$$\times \left(\prod_{i=2}^{n} \int_{\mathbb{R}} \int_{-\infty}^{\infty} \left|\frac{\partial u(x_1, ..., x_{i-1}, t_i, x_{x+1}, ..., x_n)}{\partial x_i}\right| dt_i dx_1\right)^{\frac{1}{n-1}} dx_2.$$

This time the integral

$$\left(\int_{-\infty}^{\infty} \left|\frac{\partial u(x_1, t_2, ..., x_n)}{\partial x_2}\right| dt_2\right)^{\frac{1}{n-1}}$$

is independent of the outer integral and we may break out that integral and then use the generalized Hölder inequality:

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |u(\mathbf{x})|^{\frac{n}{n-1}} dx_1 dx_2 \leq$$

$$\left(\int_{\mathbb{R}}\int_{-\infty}^{\infty} \left|\frac{\partial u}{\partial x_2}\right| dt_2 dx_1\right)^{\frac{1}{n-1}} \times$$

$$\times \left(\left(\int_{\mathbb{R}}\int_{\mathbb{R}} \left|\frac{\partial u}{\partial x_1}\right| dt_1\right)^{\frac{1}{n-1}} \left(\prod_{i=3}^{n} \int_{\mathbb{R}} \int_{-\infty}^{\infty} \left|\frac{\partial u}{\partial x_i}\right| dt_i dx_1\right)^{\frac{1}{n-1}}\right) dx_2 \leq$$

$$\leq \left(\int_{\mathbb{R}}\int_{-\infty}^{\infty} \left|\frac{\partial u}{\partial x_2}\right| dt_2 dx_1\right)^{\frac{1}{n-1}} \times$$

$$\times \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \left|\frac{\partial u}{\partial x_1}\right| dt_1 dx_2\right)^{\frac{1}{n-1}} \left(\prod_{i=3}^{n} \int_{\mathbb{R}}\int_{\mathbb{R}} \left|\frac{\partial u}{\partial x_i}\right| dx_1 dx_2 dt_i\right)^{\frac{1}{n-1}}.$$

Inductively, integrating over x_k and breaking out integrals and using the generalized Hölder inequality we reach

$$\int_{\mathbb{R}^n} |u(\mathbf{x})|^{\frac{n}{n-1}} d\mathbf{x} \leq \prod_{i=1}^{n} \left(\int_{\mathbb{R}^n} \left|\frac{\partial u}{\partial x_2}\right| dx_1 dx_2 ... dt_i ... dx_n\right)^{\frac{1}{n-1}} \leq$$

$$\leq \|\nabla u\|_{L^1(\mathbb{R}^n)}^{\frac{n}{n-1}}.$$

Noticing that the left side in the last inequality is just $\|u\|_{L^{p^*}(\mathbb{R}^n)}^{\frac{n}{n-1}}$ the claim follows.

Claim 3: *The general case.*

Proof of claim 3: Let $\alpha = \frac{(n-1)p}{n-p}$ and apply claim 2 to the function $|u|^\alpha$:

$$\left(\int_{\mathbb{R}^n} |u|^{\frac{\alpha n}{n-1}} d\mathbf{x}\right)^{\frac{n-1}{n}} \leq \int_{\mathbb{R}^n} |\nabla |u|^\alpha| d\mathbf{x} = \alpha \int_{\mathbb{R}^n} |u|^{\alpha-1} |\nabla u| d\mathbf{x} \leq$$

$$\leq \alpha \left(\int_{\mathbb{R}^n} |u|^{\frac{p(\alpha-1)}{n-p}} d\mathbf{x}\right)^{\frac{p-1}{p}} \left(\int_{\mathbb{R}^n} |\nabla u|^p d\mathbf{x}\right)^{\frac{1}{p}},$$

where we used Hölder's inequality in the last step. By our definition of α it follows that this is the same as

$$\left(\int_{\mathbb{R}^n} |u|^{\frac{np}{n-p}} d\mathbf{x}\right)^{\frac{n-1}{n}} \leq \alpha \left(\int_{\mathbb{R}^n} |u|^{\frac{np}{n-p}} d\mathbf{x}\right)^{\frac{p-1}{p}} \left(\int_{\mathbb{R}^n} |\nabla u|^p d\mathbf{x}\right)^{\frac{1}{p}},$$

dividing both sides by $\left(\int_{\mathbb{R}^n} |u|^{\frac{np}{n-p}} d\mathbf{x}\right)^{\frac{p-1}{p}}$ gives the Theorem. □

Corollary 12.2. *Let $\mathcal{D}$ be a bounded C^1 domain and $u \in W^{1,p}(\mathcal{D})$ and $u = f$ on $\partial\mathcal{D}$ in the sense of traces. Then*

$$\|u\|_{L^p(\mathcal{D})} \leq C \left(\|\nabla u\|_{L^p(\mathcal{D})} + \|f\|_{L^p(\partial\mathcal{D})}\right), \tag{123}$$

where the constant C does not depend on u (but it will depend on $\mathcal{D}$).

Sketch of the Proof: If we can extend u by zero to $\hat{u}$,

$$\hat{u}(\mathbf{x}) = \begin{cases} u(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{D} \\ 0 & \text{if } \mathbf{x} \notin \mathcal{D}, \end{cases} \tag{124}$$

so that $\hat{u} \in W^{1,p}(\mathbb{R}^n)$. Then, by Theorem 12.3,

$$\|u\|_{L^p(\mathcal{D})} = \|\hat{u}\|_{L^p(\mathcal{D})} \leq m(\mathcal{D})^{1/n}\|\hat{u}\|_{L^{p^*}(\mathcal{D})} = m(\mathcal{D})^{1/n}\|\hat{u}\|_{L^{p^*}(\mathbb{R}^n)} \leq$$

$$\leq |m(\mathcal{D})|^{1/n}\|\nabla\hat{u}\|_{L^p(\mathcal{D})} = m(\mathcal{D})^{1/n}\|\nabla u\|_{L^p(\mathbb{R}^n)},$$

where we used Hölder's inequality and m is the Lebesgue measure as usual. This proves the theorem in the case when we can extend u by zero to a function in $W^{1,p}(\mathbb{R}^n)$.

For the general case we may define a cut-off function $\psi(x) \in C_c^\infty(\mathcal{D})$ such that $0 \leq \psi(x) \leq 1$ and $\psi(x) = 1$ for $\text{dist}(x, \partial\mathcal{D}) > r_0/2$ where r_0 is the smallest radius in the proof of the Trace Theorem. We then split up $u(x) = (1 - \psi(x))u(x) + \psi(x)u(x)$. Then $\psi(x)u(x)$ can be extended by zero to a function in $W^{1,p}(\mathbb{R}^n)$ and the argument in the previous paragraph applies. The function $(1 - \psi(x))u(x)$ can be estimated in terms of the boundary values and the norm $\|\nabla(1-\psi)u\|_{L^2}$ as in (116). We leave the details to the reader. □

One can also show that the space $W^{1,p}(\mathcal{D})$ is compactly embedded in $L^q(\mathcal{D})$ for $q < \frac{np}{n-p}$. We begin with a lemma.

Lemma 12.3. *Assume that $u^j \in W^{1,p}(\mathbb{R}^n)$, $\|u^j\|_{W^{1,p}(\mathbb{R}^n)} \leq M$ for some constant M and that all the functions u^j has support contained in some compact set K. Assume furthermore that $\nabla u^j \rightharpoonup 0$ weakly in $L^p(\mathbb{R}^n)$. Then $\|u^j\|_{L^p(\mathbb{R}^n)} \to 0$.*

Proof: Consider the regularized sequence u_ϵ^j, for some small $\epsilon > 0$. Then u_ϵ^j and ∇u_ϵ^j are equicontinuous and we may therefore extract a subsequence $u_\epsilon^{j_k} \to u^0$ uniformly and $\nabla u_\epsilon^{j_k} \to \nabla u_\epsilon^0$. Since $\nabla u^{j_k} \rightharpoonup 0$ it follows that, for each $x \in \mathbb{R}^n$,

$$\nabla u_\epsilon^{j_k}(\mathbf{x}) = \int_{\mathbb{R}^n} \nabla_{\mathbf{x}}\phi_\epsilon(\mathbf{x}-\mathbf{y})u^{j_k}(\mathbf{y})dy = \tag{125}$$

$$-\int_{\mathbb{R}^n} \nabla_{\mathbf{y}}\phi_\epsilon(\mathbf{x}-\mathbf{y})u^{j_k}(\mathbf{y})dy = \int_{\mathbb{R}^n} \phi_\epsilon(\mathbf{x}-\mathbf{y})\nabla_{\mathbf{y}}u^{j_k}(\mathbf{y})dy \to 0.$$

Estimating $u(\mathbf{x})$ as in (117) it follows from (125) that

$$u_\epsilon^{j_k}(\mathbf{x}) \to 0 \quad \text{as } k \to \infty. \tag{126}$$

Since $\text{spt}(u^{j_k}) \subset K$,

$$\int_{\mathbb{R}^n} |u_\epsilon^{j_k}(\mathbf{x})|^p d\mathbf{x} = \int_K |u_\epsilon^{j_k}(\mathbf{x})|^p d\mathbf{x} = \tag{127}$$

$$= \underbrace{\int_{K\cap\{|u_\epsilon^{j_k}(\mathbf{x})|<\delta\}} |u_\epsilon^{j_k}(\mathbf{x})|^p d\mathbf{x}}_{<\delta^p m(K)} + \int_{K\cap\{|u_\epsilon^{j_k}(\mathbf{x})|>\delta\}} |u_\epsilon^{j_k}(\mathbf{x})|^p d\mathbf{x}.$$

From (126) we can conclude that for any $\delta > 0$ if k is large enough then the measure $m(K \cap \{|u_\epsilon^{j_k}(\mathbf{x})| > \delta\}) < \delta$ which in particular implies, by Hölder's inequality, that

$$\int_{K\cap\{|u_\epsilon^{j_k}(\mathbf{x})|>\delta\}} |u_\epsilon^{j_k}(\mathbf{x})|^p d\mathbf{x} \leq \tag{128}$$

$$\leq m(K \cap \{|u_\epsilon^{j_k}(\mathbf{x})| > \delta\})^{p/n} \left(\int_{K\cap\{|u_\epsilon^{j_k}(\mathbf{x})|>\delta\}} |u_\epsilon^{j_k}|^{\frac{np}{n-p}} \right)^{\frac{n-p}{n}} \leq C\delta^{p/n} M^p,$$

where we used Theorem 12.3 and that $\|\nabla u_\epsilon^{j_k}\|_{L^p} \leq \|u_\epsilon^{j_k}\|_{W^{1,p}} \leq M$ in the last inequality.

Using (128) in (127) we may conclude that

$$\int_{\mathbb{R}^n} |u_\epsilon^{j_k}(\mathbf{x})|^p d\mathbf{x} < m(K)\delta^p + C\delta^{p/n} M^p.$$

Since $\delta > 0$ is arbitrary we may conclude that $u_\epsilon^{j_k} \to 0$ in $L^p(\mathbb{R}^n)$. This does not imply that $u^{j_k} \to 0$ in L^p.

In order to complete the proof we need to sow that $u_\epsilon^{j_k}$ approximates u^{j_k} in L^p-norm. To that end we calculate

$$|u_\epsilon^{j_k}(\mathbf{x}) - u^{j_k}(\mathbf{x})| = \left| \int_{B_\epsilon(\mathbf{x})} \phi_\epsilon(\mathbf{y}-\mathbf{x}) u^{j_k}(\mathbf{y}) d\mathbf{y} - u^{j_k}(\mathbf{x}) \right| =$$

$$= \left\{ \begin{array}{l} \text{Change of var.} \\ \mathbf{z} = \frac{\mathbf{y}-\mathbf{z}}{\epsilon} \end{array} \right\} = \left| \int_{B_1(0)} \phi(\mathbf{z}) u^{j_k}(\mathbf{x} - \epsilon\mathbf{z}) d\mathbf{z} - u^{j_k}(\mathbf{x}) \right| \leq \tag{129}$$

$$\leq \int_{B_1(0)} \phi(\mathbf{z}) \left| u^{j_k}(\mathbf{x} - \epsilon\mathbf{z}) - u^{j_k}(\mathbf{x}) \right| d\mathbf{z},$$

where we used the definition that $\phi_\epsilon(\mathbf{y}-\mathbf{x}) = \frac{1}{\epsilon^n}\phi\left(\frac{\mathbf{y}-\mathbf{x}}{\epsilon}\right)$ and that the standard mollifier $\phi(\mathbf{x})$ has integral equal to one.

By the fundamental theorem of calculus we know that on a.e. line, and similarly for a.e. $\mathbf{z}$

$$u^{j_k}(\mathbf{x} - \epsilon\mathbf{z}) - u^{j_k}(\mathbf{x}) = \epsilon \int_0^1 \frac{\partial u^{j_k}(\mathbf{x} - \epsilon t\mathbf{z})}{\partial t} dt, \tag{130}$$

which inserted in (129) implies

$$|u_\epsilon^{j_k}(\mathbf{x}) - u^{j_k}(\mathbf{x})| \le \epsilon \int_{B_1(0)} \phi(\mathbf{z}) \left| \int_0^1 \frac{\partial u^{j_k}(\mathbf{x} - \epsilon t \mathbf{z})}{\partial t} dt \right| d\mathbf{z}. \tag{131}$$

Taking both sides of (131) to the power p and then integrating over $\mathbb{R}^n$ gives

$$\int_{\mathbb{R}^n} |u_\epsilon^{j_k}(\mathbf{x}) - u^{j_k}(\mathbf{x})|^p d\mathbf{x} \le \epsilon^p \int_{\mathbb{R}^n} \int_{B_1(0)} \left(\phi(\mathbf{z}) \left| \int_0^1 \frac{\partial u^{j_k}(\mathbf{x} - \epsilon t \mathbf{z})}{\partial t} dt \right| d\mathbf{z} \right)^p d\mathbf{x} \le$$

$$\le \epsilon^p \int_{B_1(0)} \phi(\mathbf{z}) \int_0^1 \int_{\mathbb{R}^n} \left| \frac{\partial u^{j_k}(\mathbf{x} - \epsilon t \mathbf{z})}{\partial t} d\mathbf{x} \right|^p dt d\mathbf{z} \le \tag{132}$$

$$\le \epsilon^p \|\nabla u^{j_k}\|^p_{L^p(\mathbb{R}^n)} \le \epsilon^p M^p,$$

where we used Hölder's inequality in the second inequality and that $\left|\frac{\partial u^{j_k}}{\partial t}\right| \le |\nabla u^{j_k}|$ in the last.

From (128) and (132) we can conclude that, for any $\delta > 0$ and $\epsilon > 0$, it follows for j_k large enough that

$$\|u^{j_k}\|_{L^p(\mathbb{R}^n)} \le \|u_\epsilon^{j_k} - u^{j_k}\|_{L^p(\mathbb{R}^n)} + \|u_\epsilon^{j_k}\|_{L^p(\mathbb{R}^n)} \le \epsilon M + C\delta^{1/n} M,$$

it follows that $u^{j_k} \to 0$ in $L^p(\mathbb{R}^n)$. □

Theorem 12.4. *Assume that $u^j \in W^{1,p}(\mathbb{R}^n)$, $\|u^j\|_{W^{1,p}(\mathbb{R}^n)} \le M$ for some constant M and that all the functions u^j has support contained in some compact set K. Assume furthermore that $\nabla u^j \rightharpoonup u^0$ weakly in $L^p(\mathbb{R}^n)$. Then $u^j \to u^0$ strongly in $L^q(\mathbb{R}^n)$ for any $q < \frac{np}{n-p} = p^*$.*

Proof: The sequence $\nabla(u^j - u^0) \rightharpoonup 0$ which by Lemma 12.3 implies that $\|u^j - u^0\|_{L^p(\mathbb{R}^n)} \to 0$. We claim that there exists an $s \in (0,1)$ for any $p < q < \frac{np}{n-p}$ such that

$$\|u^j - u^0\|_{L^q(\mathbb{R}^n)} \le \underbrace{\|u^j - u^0\|^s_{L^{\frac{np}{n-p}}(\mathbb{R}^n)}}_{\le CM \text{ by Thm 12.3}} \underbrace{\|u^j - u^0\|^{(1-s)q}_{L^p(\mathbb{R}^n)}}_{\to 0 \text{ by Lem 12.3}} \to 0. \tag{133}$$

To prove (133) we need to choose constants appropriately in the Hölder inequality. We may calculate, for $0 < s < 1$ and $w = u^j - u^0$,

$$\int_{\mathbb{R}^n} |w(\mathbf{x})|^q d\mathbf{x} = \int_{\mathbb{R}^n} \left(|w(\mathbf{x})|^{qs}\right) \left(|w(\mathbf{x})|^{(1-s)q}\right) d\mathbf{x} \le$$

$$\le \left(\int_{\mathbb{R}^n} |w|^p d\mathbf{x} \right)^{\frac{s}{p}} \left(\int_{\mathbb{R}^n} |w|^{\frac{(1-s)pq}{p-qs}} d\mathbf{x} \right)^{\frac{p-qs}{p}} =$$

$$= \|w\|^s_{L^p} \|w\|^{(1-s)q}_{L^{p^*}},$$

if we choose s so that $\frac{(1-s)pq}{p-qs} = p^*$, it is a matter of elementary algebra to verify that $0 < s < 1$ for $p < q < p^*$.[32] This proves (133).

[32]One may calculate that $s = \frac{n(p-q)+pq}{pq} = \frac{np-(n-p)q}{pq}$ that $p < q < p^*$ implies that $0 < s < 1$ easily follows.

For $q \leq p$ the theorem easily follows; since by the Hölder inequality

$$\|u^j - u^0\|_{L^q(\mathbb{R}^n)} = \|u^j - u^0\|_{L^q(K)} \leq (m(K))^{\frac{p-q}{p}} \|u^j - u^0\|_{L^p(K)} \to 0.$$

□

The next Corollary states that traces are preserved under week limits in $W^{1,p}(\mathcal{D})$.

Corollary 12.3. *Let $u^j \rightharpoonup u^0$ in $W^{1,p}(\mathcal{D})$ where $\mathcal{D}$ is a bounded C^1 domain. Assume furthermore that $u^j = f$ on $\partial\mathcal{D}$ in the trace sense. Then $u^0 = f$ on $\partial\mathcal{D}$.*

Sketch of the proof: Since $\mathcal{D}$ is a C^1 domain we may cover its boundary by a finite number $B_{r_j}(\mathbf{x}^j)$ of balls such that the boundary $\partial\mathcal{D}$ is a graph in each ball. It is clearly enough to show that $u^j \to f$ on $\partial\mathcal{D} \cap B_{r_j}(\mathbf{x}^j)$ for each ball $B_{r_j}(x^j)$.

In any given ball $B_{r_j}(\mathbf{x}^j)$ we may straighten the boundary as in Theorem 12.2. It is therefore enough to show that if $u^j \rightharpoonup u^0$ in $W^{1,p}(B_2^+(0))$[33] and $u^j(\mathbf{x}) = f(\mathbf{x})$ on $B_2(0) \cap \{x_n = 0\}$ in the sense of traces then $u^0 = f$ on $B_1(0) \cap \{x_n = 0\}$ in the sense of traces. Here we also use the property from the definition of C^1-domain that the balls overlap and it is therefore enough to show that $u^0 = f$ on $B_1(0) \cap \{x_n = 0\}$ instead of $B_2(0) \cap \{x_n = 0\}$.

In order to do that we use that any $u \in W^{1,p}(B_2^+(0))$ is absolutely continuous in the x_n variable for a.e. $\mathbf{x}' = (x_1, x_2, ..., x_{n-1}, t)$. Therefore

$$f(\mathbf{x}') = \int_0^s \frac{\partial u(\mathbf{x}', t)}{\partial x_n} dt - u(\mathbf{x}', s). \tag{134}$$

Integrating both sides from $s = 0$ to $s = 1/2$ say gives

$$f(\mathbf{x}') = 2\int_0^{1/2} \int_0^s \frac{\partial u(\mathbf{x}', t)}{\partial x_n} dt ds - 2\int_0^{1/2} u(\mathbf{x}', s) ds =$$

$$= \int_0^{1/2} \left(\frac{1}{4} - t^2\right) \frac{\partial u(\mathbf{x}', t)}{\partial x_n} dt - 2\int_0^{1/2} u(\mathbf{x}', s) ds.$$

Next multiply both sides by $g(\mathbf{x}') \in L^q(B_1(0) \cap \{x_n = 0\})$, where $\frac{1}{p} + \frac{1}{q} = 1$, and integrate over $B_1(0) \cap \{x_n = 0\}$ to deduce

$$\int_{B_1(0)\cap\{x_n=0\}} f(\mathbf{x}')g(\mathbf{x}') d\mathbf{x}' = \tag{135}$$

$$= \int_{\mathcal{S}} \left(\frac{1}{4} - x_n^2\right) g(\mathbf{x}') \frac{\partial u(\mathbf{x})}{\partial x_n} d\mathbf{x} - 2\int_{\mathcal{S}} g(x')u(\mathbf{x}) d\mathbf{x}',$$

where we changed the names of some variables (x_n for t and s) and

$$\mathcal{S} = (B_1(0) \cap \{x_n = 0\}) \times (0, 1/2).$$

The identity (135) is valid for all $u \in W^{1,p}(B_2^+(0))$ where f is the trace of u. In particular we may apply it on $u^0 - u^j \in W^{1,p}(B_2^+(0))$. If we use the notation f^0 for the trace of u^0 we have shown that

$$\int_{B_1(0)\cap\{x_n=0\}} (f^0(\mathbf{x}') - f(\mathbf{x}'))g(\mathbf{x}') d\mathbf{x}' =$$

[33]Here we use the standard notation $B_r^+(0) = B_r(0) \cap \{\mathbf{x};\ x_n > 0\}$ for the upper half ball.

$$= \int_{\mathcal{S}} \left(\frac{1}{4} - x_n^2\right) g(\mathbf{x}') \frac{\partial u^0 - u^j}{\partial x_n} d\mathbf{x} - 2 \int_{\mathcal{S}} g(x')(u^0 - u^j) d\mathbf{x}' \to 0,$$

where the right integral converges to zero since $u^j \rightharpoonup u^0$. We can conclude that

$$\int_{B_1(0)\cap\{x_n=0\}} (f^0(\mathbf{x}') - f(\mathbf{x}'))g(\mathbf{x}')d\mathbf{x}' = 0$$

for all $g \in L^q(B_1(0) \cap \{x_n = 0\})$. We may conclude, from Lemma 8.1, that $f^0 = f$. □

12.1 Exercises:

1. Let $u \in W^{1,p}(B_1(0))$ be of the form $u(\mathbf{x}) = \frac{1}{|\mathbf{x}|^\alpha}$. Calculate the least upper bound of α (the least upper bound will depend on the dimension n and on p.) Use this to calculate the values q such that $u \in L^q(B_1(0))$.

   COMMENT: *If your memory is, like mine, more similar to a goldfish's than to an elephant's then this is a simple way to calculate the constants in the Sobolev embedding theorem.*

2. Why is not the following argument valid for Theorem 12.1: By Fubini's Theorem it follows that $\frac{\partial u}{\partial x_k}$ is integrable for almost every $(x_1, x_2, ..., x_{k-1}, x_{k+1}, ..., x_n)$. It follows, from Lemma 12.1, that u is absolutely continuous on a.e. line.

3. Provide an argument for the final statement in the proof of Lemma 12.1.

4. Prove the generalized Hölder inequality: $\sum_{j=1}^n \frac{1}{p_j} = 1$ then

   $$\int_{\mathcal{D}} u_1(\mathbf{x})u_2(\mathbf{x})\dots u_n(\mathbf{x})dx \le \prod_{j=1}^n \|u_j\|_{L^{p_j}(\mathcal{D})}.$$

   You may use the following steps:

   (a)

13 Lecture 13. Solution to the calculus of variations problem.

We have now reached the level of sophistication so that we can "solve" the calculus of variations problem introduced in lecture 1. We put "solve" in quotation marks since the initial problem was not clearly defined and can therefore not have a solution. Part of the problem has to be to find the right setting for its solution.

We want to find a $u \in \mathcal{K}_E$ such that for all $v \in \mathcal{K}_E$

$$E(u) = \int_{\mathcal{D}} |\nabla u(\mathbf{x})|^2 d\mathbf{x} \le \int_{\mathcal{D}} |\nabla u(\mathbf{x})|^2 d\mathbf{x}.$$

Part of the definition of $\mathcal{K}_E$ must be that if $v \in \mathcal{K}_E$ then $v(\mathbf{x}) = f(\mathbf{x})$ on $\partial\mathcal{D}$.

The natural set $\mathcal{K}_E$ is the set

$$\mathcal{K}_E = \{v \in W^{1,2}(\mathcal{D}); \text{ such that } v = f \text{ on } \partial\mathcal{D} \text{ in the sense of traces}\}.$$

For the trace operator to be defined $T : W^{1,2}(\mathcal{D}) \mapsto \partial\mathcal{D}$ then we must assume that $\mathcal{D}$ is a bounded C^1-domain, see Theorem 12.2. We also need to make sure that $\mathcal{K}_E$ is non-empty. This means that we need to assume that $f(\mathbf{x})$ is the trace of some function $w \in W^{1,2}(\mathcal{D})$. This is necessary since so far we have only proved that, for a bounded C^1-domain $\mathcal{D}$, every function $w \in W^{1,2}(\mathcal{D})$ gives rise to a trace $f \in L^2(\partial\mathcal{D})$. We do have not shown that for every $f \in L^2(\mathcal{D})$ there is a $w \in W^{1,2}(\mathcal{D})$ such that f is the trace of w.[34]

We are now ready to prove the main theorem.

Theorem 13.1. *Assume that $\mathcal{D}$ is a bounced C^1-domain and that $f \in L^2(\partial\mathcal{D})$ is the trace of $w \in W^{1,2}(\mathcal{D})$. If we denote*

$$\mathcal{K}_E = \big\{v \in W^{1,2}(\mathcal{D}); \text{ such that } v = f \text{ on } \partial\mathcal{D} \text{ in the sense of traces}\big\}$$

then there exists a unique $u \in \mathcal{K}_E$ such that

$$E(u) := \int_{\mathcal{D}} |\nabla u(\boldsymbol{x})|^2 d\boldsymbol{x} \leq \int_{\mathcal{D}} |\nabla v(\boldsymbol{x})|^2 d\boldsymbol{x}$$

for all $v \in \mathcal{K}_E$.

Proof:

Step 1: *There exists an m such that* $\inf_{v \in \mathcal{K}_E} E(v) = m$.

Proof of Step 1: Define the set of real numbers

$$V = \left\{ \int_{\mathcal{D}} |\nabla v(\mathbf{x})| d\mathbf{x};\ v \in \mathcal{K}_E \right\}.$$

Since $w \in \mathcal{K}_E$ it follows that $V \neq \emptyset$. Furthermore, since the integrand $|\nabla v|^2 \geq 0$ it follows that the set V is bounded from below by 0. By the completeness of the real numbers we may conclude that the greatest lower bound of V $(g.l.b.(V))$ exists; define $m = g.l.b.(V)$.

Step 2: *There exist a sequence $u^j \in \mathcal{K}_E$ such that $E(u^j) \to m$.*

Proof of Step 2: Since m is the greatest lower bound of V it follows that $m + \frac{1}{j}$ is not a lower bound. We can therefore find a function $u^j \in \mathcal{K}_E$ such that $m \leq E(u^j) < m + \frac{1}{j}$. In particular, $E(u^j) \to m$ as $j \to \infty$.

Step 3: *The sub-sequence $u^{j_k} \rightharpoonup u^0$ where $u^0 \in \mathcal{K}_E$.*

Proof of Step 3: This is the heart of the proof. From Lemma 9.1 it follows that the subsequence $u^{j_k} \rightharpoonup u^0$ exists if $\|u^j\|_{W^{1,2}}(\mathcal{D}) \leq M$ for some M.[35]

Since $E(u^j) = \|\nabla u^j\|^2_{L^2(\mathcal{D})} \to m$ it follows that

$$\|\nabla u^j\|^2_{L^2(\mathcal{D})} \leq m + 1 \quad \text{for } j \text{ large enough.}$$

[34] As a matter of fact this is not true. If f is very discontinuous then it is not the trace of any $w \in W^{1,2}(\mathcal{D})$. We will not discuss this in the next lecture.

[35] It is worth to remark here that the last statement is only true since we interpret the integral in the Lebesgue sense. It is not even clear that the the function u^0 is integrable in the Riemann sense even if all the u^j are.

It is therefore enough to show that $\|u^j\|_{L^2(\mathcal{D})}$ is bounded. From Theorem 12.2 (in particular from the estimate (110)) we may conclude that

$$\|f\|_{L^2(\partial\mathcal{D})} \le C\|w\|_{W^{1,2}(\mathcal{D})}. \tag{136}$$

And from Corollary 12.2 and (136) it follows that

$$\|u^j\|_{L^2(\mathcal{D})} \le C\left(\|\nabla u^j\|_{L^2(\mathcal{D})} + \|f\|_{L^2(\partial\mathcal{D})}\right) \le C\left(m + 1 + \|w\|_{W^{1,2}(\mathcal{D})}\right) =: M,$$

where M depend on w, m, and on $\mathcal{D}$ but is independent of j. It follows that u^j is bounded in $W^{1,2}(\mathcal{D})$ and thus, by Lemma 9.1, there is a convergent subsequence $u^{j_k} \rightharpoonup u^0 \in W^{1,2}(\mathcal{D})$.

We need to show that the trace of u^0 equals f in order to conclude that $u^0 \in \mathcal{K}_E$. But this follows from Corollary 12.3. We may conclude that $u^0 \in \mathcal{K}_E$.

Step 4: *Conclusion of the theorem:* $E(u^0) = \inf_{v\in\mathcal{K}_E} E(v)$.

Proof of Step 4: Since $\nabla u^{j_k} \rightharpoonup \nabla u^0$ it follows from Proposition 8.4 that

$$m = \lim_{k\to\infty} \|\nabla u^{j_k}\|^2_{L^2(\mathcal{D})} \ge \|\nabla u^0\|^2_{L^2(\mathcal{D})} = E(u^0) \ge m, \tag{137}$$

where we used that $E(u) = \|\nabla u\|^2_{L^2(\mathcal{D})}$ and that $m = \inf_{v\in\mathcal{K}_E} E(v)$ in respectively the first and last inequalities. It follows from (137) that

$$E(u^0) = m = \inf_{v\in\mathcal{K}_E} E(v).$$

□

13.1 A more general variational problem.

The minimization problem above has a certain structure, just as the simple minimization problem in Theorem 1.1. Also, the only substantial difficulties in proving Theorem 13.1 was to develop the analysis. We should therefore be able to easily extend Theorem 13.1 to more general cases that fits the structure of the proof. We only need to look at the proof axiomatically and see what we use and then make assumptions that assures that all assumptions are satisfied.

So what do we use of the functional $\int_{\mathcal{D}} |\nabla u|^2 d\mathbf{x}$. We use the following:

1. In step 1 we use that $|\nabla u|^2 \ge -C$ in order to assure that m exists.
2. In Step 2 we use nothing relating to the functional.
3. In step 3 we use that $E(u^j) \to m$ implies that $\|\nabla u^j\|_{L^2(\mathcal{D})} \le M$.
4. In step 4 we use that the functional E is lower semi-continuous with respect to weak convergence in $W^{1,2}(\mathcal{D})$.

We should therefore be able to use the same proof to show the existence of minimizers for any functional

$$\int_{\mathcal{D}} F(\nabla u(\mathbf{x}))d\mathbf{x},$$

as long as

1. $F \geq -C_0$, for some constant C_0.
2. We somehow need to assure that if $\int_{\mathcal{D}} F(\nabla u)d\mathbf{x}$ is bounded then $\|\nabla u\|_{L^p(\mathcal{D})}$ is. It is enough to assume that, for some constant C_1,
$$F(\nabla u(\mathbf{x})) \geq |\nabla u(\mathbf{x})|^p - C_1. \tag{138}$$
3. We need to assure that $\int_{\mathcal{D}} F(\nabla u)d\mathbf{x}$ is lower semi-continuous with respect to weak convergence in $W^{1,p}$.

The only thing we need to investigate is under what conditions does the lower semi-continuity hold?

We could, of course, just assume that F is such that the energy is lower semi-continuous. But in general, it is not meaningful to have theorems if we cannot verify when the assumptions are satisfied. It would therefore be much more reassuring if we could find some criteria that implies lower semi-continuity for the functional. This is what we will do next. As so often in mathematics we will try to understand a complicated situation by constructing an example easy enough for us to explicitly calculate it. In non-linear analysis that usually means construction a one dimensional example; since the power of one dimensional calculus allows us to do most calculations explicitly in one dimension.

Example: Consider the one-dimensional minimizing problem

$$\text{minimize } J_F(f(x)) = \int_0^1 F(f'(x))dx \tag{139}$$

in the set
$$K = \left\{ f \in W^{1,p}(0,1);\ f(0) = 0 \text{ and } f(1) = 1 \right\}.$$

We need to choose our function $F(\cdot)$ which we choose quite randomly to be the function with the graph

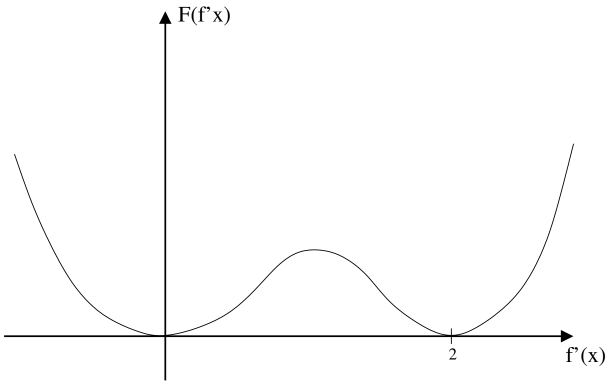


Figure 2: *The graph of the function $F(\cdot)$.*

Since $F(\cdot) \geq 0$ we can conclude that $J_F(f) \geq 0$ for all functions $f \in K$. But if
$$f'(x) = \begin{cases} 0 \text{ if } x \in A \\ 2 \text{ if } x \notin A \end{cases} \tag{140}$$

for some set A then the energy $J_F(f'(x)) = 0$ since $F(0) = F(2) = 0$. Thus any function $f(x)$ of the form (140) will be a minimizer to (139). Notice that such a minimizer can arbitrarily well approximate (in $C^0([0,1])-$norm) any function $g(x)$ satisfying $0 \leq g' \leq 2$. This can be clearly seen in the following picture:

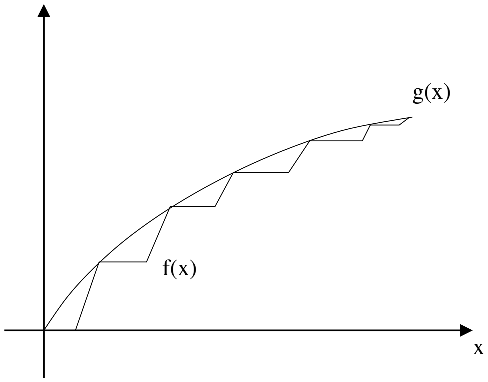


Figure 3: *Graphic representation of how a function $f(x)$ whose derivative takes the values* 0 *and* 2 *approximates an arbitrary function $g(x)$ with derivative* $0 \leq g'(x) \leq 2$.

This implies that for any function $g(x) \in K$ such that $0 \leq g'(x) \leq 2$ we can find a sequence $f^j \in K$ such that $f^j \to g$ uniformly and $J(f^j(x)) = 0$. But $J(g(x))$ may very well be strictly positive, for instance if $g(x) = x$. Thus the functional $J_F(f)$ defined in (139) is not lower semi-continuous.[36]

The question we need to ask is: *Is the problem that the function F is zero at two different points?* A simple example shows that that is not the case.

Consider for instance the one dimensional minimization problem

$$\text{minimize } J_G(f(x)) = \int_0^1 G(f'(x))dx$$

in the set

$$K = \left\{ f \in W^{1,p}(0,1);\ f(0) = 0 \text{ and } f(1) = 1 \right\},$$

where the function G is given by the graph:

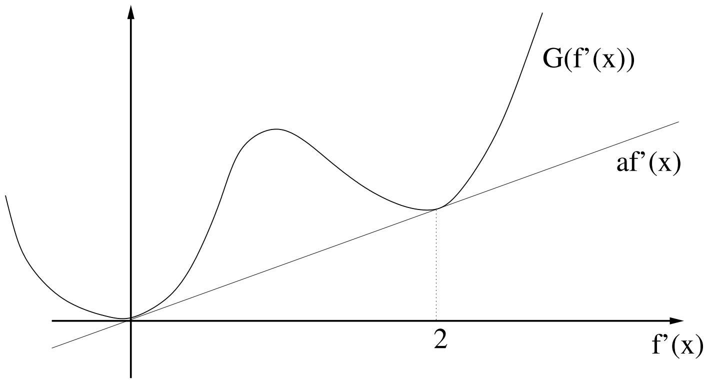


Figure 4: *The graph of the function $G(f'(x))$ and $af'(x)$.*

[36]Notice that the lower semi-continuity of J_F is not really related to the continuity of F. We may very well, as in the example, have that F is a continuous function but J_F is not continuous on the space $W^{1,2}(\mathcal{D})$.

If we subtract the linear function $af'(x)$ from $G(f'(x))$ we will get a function with graph looking like the one in Figure 2; we may even assume that $G(f'(x)) = F(f'(x)) + af'(x)$. This leads to

$$J_G(f'(x)) = \int_0^1 G(f'(x))dx = \int_0^1 F(f'(x))dx + a\int_0^1 f'(x)dx =$$
$$= J_F(f'(x)) + af(1) - af(0),$$

where we used an integration by parts in the last equality. Since $af(1) - af(0) = a$ for all $f \in K$ we can conclude that

$$J_G(f'(x)) = J_F(f'(x)) + a \text{ for all } f \in K.$$

And since J_F and J_G only differ by a constant we can conclude that J_G cannot have a minimizer since J_F does not have a minimizer. □

What conclusion can we draw from this example? The reason that there a minimizer to $\int_0^1 G(f'(x))dx$ does not exist was that we could touch the graph of G from below, at two different points, by a linear function. That is: G is not convex. Clearly convexity is a necessary condition for a minimizer to exist, at least for minimization in $\mathbb{R}$. Is convexity enough for us to conclude that $\int_{\mathcal{D}} F(\nabla u)d\mathbf{x}$ is lower semi-continuous with respect to weak convergence in $W^{1,p}$? The only way to know is to investigate.

First we need to remind ourselves of what it means for a function to be convex. In order to simplify somewhat we will assume that F is continuously differentiable. Then F is convex if and only if

$$F(\mathbf{q}) \geq F(\mathbf{p}) + F'(\mathbf{p})(\mathbf{q} - \mathbf{p}) \quad \text{ for any } \mathbf{p}, \mathbf{q} \in \mathbb{R}^n.$$

Let us try to use the convexity condition in the variational integral

$$\int_{\mathcal{D}} F(\nabla u^j(\mathbf{x}))d\mathbf{x} \geq \int_{\mathcal{D}} \left(F(\nabla u_0) + F'(\nabla u^0)(\nabla u^j - \nabla u^0)\right) d\mathbf{x}. \tag{141}$$

Observe that if $u^j \rightharpoonup u^0$ in $W^{1,p}$ then

$$\int_{\mathcal{D}} G(\mathbf{x})(\nabla u^j - \nabla u^0)d\mathbf{x} \to 0,$$

for any $G(\mathbf{x}) \in L^q(\mathcal{D})$, where $\frac{1}{p} + \frac{1}{q} = 1$. This means that if

$$F'(\nabla u^0) \in L^q(\mathcal{D}) \tag{142}$$

then

$$\int_{\mathcal{D}} F(\nabla u^j)d\mathbf{x} \geq \int_{\mathcal{D}} \big(F(\nabla u_0) + \underbrace{F'(\nabla u^0)(\nabla u^j - \nabla u^0)}_{\to 0}\big)d\mathbf{x} \to \int_{\mathcal{D}} F(\nabla u_0)d\mathbf{x},$$

which implies that $\int_{\mathcal{D}} F d\mathbf{x}$ is lower semi-continuous.

The important condition is (142). But this is clearly implied by

$$|F'(\mathbf{p})| \leq |\mathbf{p}|^{\frac{p}{q}} + C_2.$$

We may are now ready to formulate a generalized version of Theorem 13.1.

Theorem 13.2. *Let C_0, C_1 and C_2 be three constants and assume that $F(\nabla u(\boldsymbol{x}))$ satisfies, for some $1 < p < \infty$,*

1. $F \geq -C_0$.
2. $$F(\nabla u(\boldsymbol{x})) \geq |\nabla u(\boldsymbol{x})|^p - C_1. \tag{143}$$
3. *F is continuously differentiable and*
$$|F'(\boldsymbol{p})| \leq |\boldsymbol{p}|^{\frac{p}{q}} + C_2.$$
4. *F is convex*
$$F(\boldsymbol{q}) \geq F(\boldsymbol{p}) + F'(\boldsymbol{p})(\boldsymbol{q} - \boldsymbol{p}) \quad \text{for any } \boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n.$$

Furthermore assume that $\mathcal{D}$ is a bounded C^1 domain and that $f = Tw$ on $\partial\mathcal{D}$ for some $w \in W^{1,p}(\mathcal{D})$ in the sense of traces. Then there exist a unique minimizer to
$$E(v) = \int_{\mathcal{D}} F(\nabla v(\boldsymbol{x}))d\boldsymbol{x}$$
in the set
$$\mathcal{K}_E = \left\{v \in W^{1,p}(\mathcal{D});\ v = f \text{ on } \partial\mathcal{D} \text{ in the sense of traces.}\right\}.$$

Instead of Exercises: We have now reached the main theorem of this course. But, in this course, we are not primarily interested in the theorem. Rather, we are interested in all the analysis we need in order to prove the theorem. We are therefore at a good vantage point from which we can overview the theory. Your only exercise this week is therefore to repeat the previous lectures. Try to see how the main theorems in lecture 2-10 are used in the proof of Theorem 13.1. Some main theorems in the course are not mentioned in the proof of Theorem 13.1. For instance, we do not mention that absolutely continuous functions satisfy the fundamental theorem of calculus directly in the proof - but we use the trace theorem that uses that theorem in its proof and we also use that Sobolev functions are differentiable on a.e. line. I would therefore urge you to backtrack and repeat previous proofs and try to see how the entire course is interconnected and used in Theorem 13.1. It might be just one exercise, but it is rather substantial.

14 Lecture 14. Difference quotients and the Euler-Lagrange equations.

In the previous section we achieved an amazing theorem. We now know that solutions to the variational problem introduced in lecture 1 exists. But existence does not tell us much about the solution. How does the minimizer behave? Can we say anything about it more than it is a minimizer? Can analysis help us in investigating the solution further?

In this section we will say some things abut these questions. Our main interest is however classical analysis and not the calculus of variations. We will therefore view the material in these lectures as a way to see that the theory is rightly conceived in the sense that it helps us to prove difficult results. When applying analysis to some difficult problems we will also see how it works in practice.

We begin by proving the most basic formula in the calculus of variations.

Theorem 14.1. *Let u be the unique minimizer, in $W^{1,2}(\mathcal{D})$, of the Dirichlet energy (under the assumptions of Theorem 13.1)*

$$\int_{\mathcal{D}} |\nabla u(\boldsymbol{x})|^2 d\boldsymbol{x}.$$

Then

$$\int_{\mathcal{D}} \nabla u(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x}) d\boldsymbol{x} = 0, \tag{144}$$

for any $v \in W^{1,2}(\mathcal{D})$ such that $v = 0$ on $\partial\mathcal{D}$ (in the sense of traces).

Proof: Since u is a minimizer and $u + tv \in W^{1,2}(\mathcal{D})$ with trace f it follows that

$$\int_{\mathcal{D}} |\nabla u(\mathbf{x})|^2 d\mathbf{x} \leq \int_{\mathcal{D}} |\nabla(u(\mathbf{x}) + tv(\mathbf{x})|^2 d\mathbf{x} =$$

$$= \int_{\mathcal{D}} \left(|\nabla u(\mathbf{x})|^2 + 2t\nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x}) + t^2|\nabla v(\mathbf{x})|^2\right) d\mathbf{x}.$$

Subtracting $\int_{\mathcal{D}} |\nabla u(\mathbf{x})|^2 d\mathbf{x}$ and dividing by $|t|$ gives

$$0 \leq \int_{\mathcal{D}} \left(2\frac{t}{|t|}\nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x}) + |t||\nabla v(\mathbf{x})|^2\right) d\mathbf{x}.$$

Sending $t \to 0^{\pm}$ gives

$$0 \leq \pm \int_{\mathcal{D}} 2\nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x}) d\mathbf{x},$$

where we get "plus" if $t \to 0^+$ and "minus" if $t \to 0^-$. This implies (144). □

Some heuristics: *Seeing (144) one becomes very tempted to do an integration by parts and conclude that*[37]

$$\int_{\mathcal{D}} \nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{D}} v(\mathbf{x}) \Delta u(\mathbf{x}) d\mathbf{x} = 0, \tag{145}$$

for all $v(\mathbf{x}) \in W^{1,2}(\mathcal{D})$ with vanishing trace. From (145) one would like to conclude directly that $\Delta u(\mathbf{x}) = 0$. This would imply that minimizers of the Dirichlet energy are harmonic functions. The problem with this argument is that we do not know whether $u(\mathbf{x})$ has well defined second derivatives, say $u \in W^{2,2}(\mathcal{D})$; and therefore we do not know that $\Delta u(\mathbf{x})$ even makes sense or that we can make an integration by parts in (145). But deriving that any

[37]Remember that $\Delta u(\mathbf{x}) = \mathrm{div}(\nabla u(\mathbf{x})) = \sum_{i=1}^{n} \frac{\partial^2 u(\mathbf{x})}{\partial x_i^2}$. Also remember that we call a function u harmonic if $\Delta u(\mathbf{x}) = 0$.

minimizer of the Dirichlet energy is harmonic is such a nice result that we will try to derive that minimizers have second derivatives.

Since (145) makes perfect sense for any function $u \in W^{1,2}(\mathcal{D})$ we will prove the slightly stronger result that any function satisfying (145) has second derivatives. We begin by a definition.

Definition 14.1. *We say that a function $u \in W^{1,2}(\mathcal{D})$ is a weak solution to*

$$\Delta u(\boldsymbol{x}) = 0,$$

or that u is weakly harmonic, if for any $v \in W^{1,2}(\mathcal{D})$ with compact support

$$\int_{\mathcal{D}} \nabla u(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x}) d\boldsymbol{x} = \int_{\mathcal{D}} v(\boldsymbol{x}) \Delta u(\boldsymbol{x}) d\boldsymbol{x} = 0. \tag{146}$$

Our aim is to prove that any weakly harmonic function is actually harmonic; that would (in view of Theorem 14.1) imply that any minimizer of the Dirichlet energy is harmonic. We begin with a simple "warm up" result that is not strictly needed; but it will give us a main idea of the methodology we are going to use.

Proposition 14.1. *Assume that u is weakly harmonic. Then, for any compact set $K \subset \mathcal{D}$,*

$$\|\nabla u\|_{L^2(K)} \leq \frac{C}{\delta} \|u\|_{L^2(B_R(0))}, \tag{147}$$

where $\delta > 0$ is the distance from K to ∂D.

Proof: There is only one thing that we know about u; that u satisfies (146). Therefore the only thing we can do in order to prove (147), or anything else about u for that matter, is to choose v in a smart way.

Let us choose $v = \psi^2(\mathbf{x})u(\mathbf{x})$ for some $\psi \in C_c^\infty(\mathcal{D})$ such that $\psi(\mathbf{x}) = 1$ in K and $|\nabla \psi(\mathbf{x})| \leq \frac{4}{\delta}$, we can for instance choose ψ as a mollification of the characteristic function of the set K: $\psi(\mathbf{x}) = \phi_{\delta/2} * \chi_K(\mathbf{x})$. Since ψ has compact support in $\mathcal{D}$ it follows that it has zero trace and since $u \in W^{1,2}(\mathcal{D})$ (by the assumption of being a weak solution) it follows from the product rule (Theorem 9.3) that $v \in W^{1,2}(\mathcal{D})$. We may therefore use $v = \psi^2(\mathbf{x})u(\mathbf{x})$ in the definition of weak solution and derive that

$$0 = \int_{B_R(0)} \nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x}) d\mathbf{x} = \int_{B_R(0)} \left(\psi^2 |\nabla u|^2 + 2u\psi \nabla \psi \cdot \nabla u\right) d\mathbf{x}, \tag{148}$$

where we used the definition of $v(\mathbf{x}) = \psi^2(\mathbf{x})u(\mathbf{x})$ and the product rule in the last equality.

Rearranging terms in (148) we see that

$$\int_{B_R(0)} \psi^2(\mathbf{x}) |\nabla u|^2 d\mathbf{x} = -\int_{B_R(0)} \left(2u\psi \nabla \psi \cdot \nabla u\right) d\mathbf{x} \leq \tag{149}$$

$$\leq \left(\int_{B_R(0)} \psi^2(\mathbf{x}) |\nabla u|^2 d\mathbf{x}\right)^{1/2} \left(\int_{B_R(0)} u(\mathbf{x}) |\nabla \psi|^2 d\mathbf{x}\right)^{1/2},$$

where we used the Cauchy-Schwartz inequality in the last inequality.

If we divide both sides of (149) by $\left(\int_{B_R}\psi^2|\nabla u|^2 d\mathbf{x}\right)^{1/2}$ and then square we get

$$\int_{B_R(0)}\psi^2(\mathbf{x})|\nabla u|^2 \le \int_{B_R(0)} u(\mathbf{x})|\nabla\psi|^2 d\mathbf{x}.$$

The conclusion of the proposition follows if we notice that $\psi = 1$ in $B_{R_0}(0)$ and $|\nabla\psi| \le \frac{2}{\delta}$ and therefore

$$\int_{B_{R_0}(0)}|\nabla u|^2 \le \int_{B_R(0)}\psi^2(\mathbf{x})|\nabla u|^2 \le$$

$$\le \int_{B_R(0)} u(\mathbf{x})|\nabla\psi|^2 d\mathbf{x} \le \frac{4}{\delta^2}\int_{B_R(0)} u(\mathbf{x})d\mathbf{x}.$$

□

Next we will show that any weakly harmonic function have second derivatives in L^2. The method of proof is very similar to the method of proof in the previous proposition. We begin with a *difference quotient argument*, this is a standard argument in PDE theory and the calculus of variations. In the proof we use the notation $e_1 = (1,0,0,..,0), \cdots, e_i = (0,..,0,1,0,..)$ for the standard unit vectors. We want to estimate the difference quotients $\frac{\nabla(u(\mathbf{x}+e_i h)-u(\mathbf{x}))}{h}$, if these can be estimates by a quantity independent of h then the weak second derivatives will exist by Lemma 12.2.

Lemma 14.1. *Let $u(\boldsymbol{x})$ be a weakly harmonic function in a domain $\mathcal{D}$. Then for each compact set $\mathcal{C} \subset \mathcal{D}$ there exists a constant C only depending on dist$(\mathcal{C},\mathcal{D}^c)$ such that*

$$\int_{\mathcal{C}}\left|\frac{\nabla(u(\boldsymbol{x}+e_i h)-u(\boldsymbol{x}))}{h}\right|^2 d\boldsymbol{x} \le C\int_{\mathcal{D}}\left|\frac{u(\boldsymbol{x}+e_i h)-u(\boldsymbol{x})}{h}\right|^2 d\boldsymbol{x} \qquad (150)$$

for any $h \in \mathbb{R}$ satisfying $|h| < $ dist$(\mathcal{C},\mathcal{D}^c)/2$.

Proof: We know that

$$0 = \int_{\mathcal{D}}\nabla u(\mathbf{x})\cdot\nabla\phi(\mathbf{x})d\mathbf{x} \qquad (151)$$

for any ϕ with compact support such that $u + t\phi \in \mathcal{K}_E$.

Now we choose $\phi(\mathbf{x}) = \psi(\mathbf{x})^2(u(\mathbf{x}+e_i h) - u(\mathbf{x}))$ for some $\psi \in C_c^\infty(\mathcal{D})$ that satisfies

1. $0 \le \psi(\mathbf{x}) \le 1$ for all $\mathbf{x} \in \mathcal{D}$,
2. $\psi(\mathbf{x}) = 1$ for $\mathbf{x} \in \mathcal{C}$,
3. $\psi(\mathbf{x}) = 0$ for all $\mathbf{x}$ such that $\text{dist}(\mathbf{x},\mathcal{C}) > \frac{\text{dist}(\mathcal{C},\mathcal{D}^c)}{2}$ and
4. $|\nabla\psi(\mathbf{x})| < \frac{4}{\text{dist}(\mathcal{C},\mathcal{D}^c)}$.

Notice that $u(\mathbf{x}) + t\psi^2(\mathbf{x})(u(\mathbf{x}+e_ih) - u(\mathbf{x})) = f(\mathbf{x})$ on $\partial\mathcal{D}$ since $\psi(\mathbf{x}) = 0$ on $\partial\mathcal{D}$ and $u(\mathbf{x}) = f(\mathbf{x})$ on $\partial\mathcal{D}$.[38]

With this choice of $\phi(\mathbf{x})$ in (151) we arrive at

$$\int_{\mathcal{D}} \Big(\nabla u(\mathbf{x}) \cdot \nabla(\psi(\mathbf{x})^2(u(\mathbf{x}+e_ih) - u(\mathbf{x})))\Big)d\mathbf{x} = 0. \tag{152}$$

Next we notice that $u(\mathbf{x}+he_i)$ is a weak solution in a slightly shifted domain with boundary values $f(\mathbf{x}+he_i)$. Arguing similarly as above we arrive at (with $\phi(\mathbf{x}) = \psi(\mathbf{x})^2(u(\mathbf{x}) - u(\mathbf{x}+he_i))$

$$\int_{\mathcal{D}} \Big(\nabla u(\mathbf{x}+he_i) \cdot \nabla(\psi(\mathbf{x})^2(u(\mathbf{x}) - u(\mathbf{x}+he_i)))\Big)d\mathbf{x} = 0. \tag{153}$$

If we add (152) and (153) and rearrange the terms we arrive at

$$0 = \int_{\mathcal{D}} \left(\nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\right) \cdot \nabla\left(\psi(\mathbf{x})^2(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\right) d\mathbf{x} =$$

$$= \int_{\mathcal{D}} \left(\psi(\mathbf{x})^2 \left|\nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\right|^2\right) d\mathbf{x} +$$

$$+ \int_{\mathcal{D}} \left(2\psi(\mathbf{x})(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\nabla\psi(\mathbf{x}) \cdot \nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\right) d\mathbf{x}$$

That is

$$\int_{\mathcal{D}} \psi(\mathbf{x})^2 \left|\nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\right|^2 d\mathbf{x} = \tag{154}$$

$$= -\int_{\mathcal{D}} 2\psi(\mathbf{x})(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\nabla\psi(\mathbf{x}) \cdot \nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x}))d\mathbf{x}.$$

In order to continue we use that for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we have the following inequality $2\mathbf{v}\cdot\mathbf{w} \le 2|\mathbf{v}|^2 + \frac{1}{2}|\mathbf{w}|^2$ which implies that

$$2\psi(\mathbf{x})(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\nabla\psi(\mathbf{x}) \cdot \nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x})) =$$

$$= \underbrace{(2(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\nabla\psi(\mathbf{x}))}_{=\mathbf{v}} \cdot \underbrace{(\psi(\mathbf{x})\nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x})))}_{=\mathbf{w}} \le \tag{155}$$

$$\le 8|\nabla\psi(\mathbf{x})|^2(u(\mathbf{x}+e_ih) - u(\mathbf{x}))^2 + \frac{1}{2}|\psi(\mathbf{x})|^2 \left|\nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\right|^2 - .$$

Using (155) in (154) we can deduce that

$$\int_{\mathcal{D}} \psi(\mathbf{x})^2 \left|\nabla(u(\mathbf{x}+e_ih) - u(\mathbf{x}))\right|^2 d\mathbf{x} \le \tag{156}$$

$$\le 16\int_{\mathcal{D}} |\nabla\psi(\mathbf{x})|^2(u(\mathbf{x}+e_ih) - u(\mathbf{x}))^2 d\mathbf{x}.$$

[38] There is a slight technical detail that should be mentioned here. Since $u(\mathbf{x})$ is defined on $\mathcal{D}$ it follows that $u(\mathbf{x}+he_i)$ is defined on the set $\mathcal{D}_{-h} = \{\mathbf{x};\ \mathbf{x}+he_i \in \mathcal{D}\} \neq \mathcal{D}$. In particular, the function $u(\mathbf{x}) + t\psi(\mathbf{x})^2(u(\mathbf{x}+e_ih) - u(\mathbf{x}))$ is only defined on $\mathcal{D}_{-h} \cap \mathcal{D}$ which is a strictly smaller set than $\mathcal{D}$. But since $\psi(\mathbf{x}) = 0$ on $\mathcal{D} \setminus (\mathcal{D}_{-h} \cap \mathcal{D})$ for $|h| < \mathrm{dist}(\mathcal{C}, \mathcal{D}^c)/2$ we may consider the function that equals $u(\mathbf{x}) + t\psi(\mathbf{x})^2(u(\mathbf{x}+e_ih) - u(\mathbf{x}))$ in $\mathcal{D}_{-h} \cap \mathcal{D}$ and equals zero in $\mathcal{D} \setminus (\mathcal{D}_{-h} \cap \mathcal{D})$ for $|h| < \mathrm{dist}(\mathcal{C}, \mathcal{D}^c)/2$. That function is well defined and all the calculations goes through for that function. It is not uncommon that one uses the simplified convention that an undefined function times zero is zero - it simplifies things.

Since $\psi(\mathbf{x}) = 1$ in $\mathcal{C}$ we can estimate the left side of (156) according to

$$\int_{\mathcal{C}} |\nabla(u(\mathbf{x}+e_i h) - u(\mathbf{x}))|^2 \, d\mathbf{x} \leq \int_{\mathcal{D}} \psi(\mathbf{x})^2 \, |\nabla(u(\mathbf{x}+e_i h) - u(\mathbf{x}))|^2 \, d\mathbf{x} \quad (157)$$

and using that $|\nabla\psi| \leq \frac{4}{\text{dist}(\mathcal{C},\mathcal{D}^c)}$ we can estimate the right side of (156) according to

$$16 \int_{\mathcal{D}} |\nabla\psi(\mathbf{x})|^2 (u(\mathbf{x}+e_i h) - u(\mathbf{x}))^2 d\mathbf{x} \leq \quad (158)$$

$$\leq \frac{256}{\text{dist}(\mathcal{C},\mathcal{D}^c)^2} \int_{\mathcal{D}} (u(\mathbf{x}+e_i h) - u(\mathbf{x}))^2 d\mathbf{x}.$$

Putting (156), (157) and (158) and dividing by h^2 we arrive at (150). □

Lemma 14.1 provides an integral estimate for the difference quotient of the derivatives. But unless we can also show that the right side in (150) is uniformly bounded in h the Lemma would not be very useful. We therefore need the following integral version of the mean value property for the derivatives.

Lemma 14.2. *Assume that $u \in W^{1,2}(\mathcal{D})$ and $\mathcal{C} \subset \mathcal{D}$ is a compact set. Then there exists a constant C depending only on the dimension such that for any $|h| \leq dist(\mathcal{C},\mathcal{D}^c)$*

$$\int_{\mathcal{C}} \left| \frac{u(\boldsymbol{x}+e_i h) - u(\boldsymbol{x})}{h} \right|^2 d\boldsymbol{x} \leq C \int_{\mathcal{D}} \left| \frac{\partial u(\boldsymbol{x})}{\partial x_i} \right|^2 d\boldsymbol{x}. \quad (159)$$

Proof: We will use the following simple version of the Cauchy-Schwartz inequality: *Let $f \in L^2(\Sigma)$ and $|\Sigma|$ denote the area of Σ then*

$$\left| \int_{\Sigma} |f(\mathbf{x})| d\mathbf{x} \right|^2 \leq |\Sigma| \int_{\Sigma} |f(\mathbf{x})|^2 d\mathbf{x}. \quad (160)$$

From the Fundamental Theorem of calculus, which applies on a.e. line by Theorem 12.1, we see that on a.e. line

$$\frac{u(\mathbf{x}+e_i h) - u(\mathbf{x})}{h} = \frac{1}{h} \int_0^h \frac{\partial u(\mathbf{x}+se_i)}{\partial x_i} ds.$$

Thus

$$\int_{\mathcal{C}} \left| \frac{u(\mathbf{x}+e_i h) - u(\mathbf{x})}{h} \right|^2 d\mathbf{x} = \int_{\mathcal{C}} \left| \frac{1}{h} \int_0^h \frac{\partial u(\mathbf{x}+se_i)}{\partial x_i} ds \right|^2 d\mathbf{x} \leq \quad (161)$$

$$\leq \int_{\mathcal{C}} \frac{1}{h} \int_0^h \left| \frac{\partial u(\mathbf{x}+se_i)}{\partial x_i} \right|^2 ds d\mathbf{x}, \quad (162)$$

where we used (160), with $\Sigma = (0,h)$, in the last inequality. We may continue to estimate (162) by using the Fubini Theorem

$$\int_{\mathcal{C}} \frac{1}{h} \int_0^h \left| \frac{\partial u(\mathbf{x}+se_i)}{\partial x_i} \right|^2 ds d\mathbf{x} = \frac{1}{h} \int_0^h \int_{\mathcal{C}} \left| \frac{\partial u(\mathbf{x}+se_i)}{\partial x_i} \right|^2 d\mathbf{x} ds \leq \quad (163)$$

$$\leq \frac{1}{h} \int_0^h \left(\int_{\mathcal{D}} \left| \frac{\partial u(\mathbf{x})}{\partial x_i} \right|^2 d\mathbf{x} \right) ds \leq \int_{\mathcal{D}} \left| \frac{\partial u(\mathbf{x})}{\partial x_i} \right|^2 d\mathbf{x}. \quad (164)$$

Putting (161), (162), (163) and (164) together gives the Lemma. □

Putting the above results together we may deduce the weak solutions have second derivatives in L^2.

Theorem 14.2. *Let $u(\boldsymbol{x})$ be weakly harmonic. Then $u(\boldsymbol{x})$ has weak derivatives of second order on any compact subset $\mathcal{C}$ of $\mathcal{D}$ and there exists a constant $C_{\mathcal{C}}$ (depending on $\mathcal{C}$) such that for any i*

$$\int_{\mathcal{C}} \left|\nabla \frac{\partial u(\boldsymbol{x})}{\partial x_i}\right|^2 d\boldsymbol{x} \leq C_{\mathcal{C}} \int_{\mathcal{D}} \left|\frac{\partial u(\boldsymbol{x})}{\partial x_i}\right|^2 d\boldsymbol{x}, \tag{165}$$

this implies that

$$\int_{\mathcal{C}} \left|D^2 u(\boldsymbol{x})\right|^2 d\boldsymbol{x} \leq C_{\mathcal{C}} \int_{\mathcal{D}} |u(\boldsymbol{x})|^2 d\boldsymbol{x}, \tag{166}$$

where the constant $C_{\mathcal{C}}$ is independent of u.

Proof: From (150) and (159) we see that

$$\int_{\tilde{\mathcal{C}}_\delta} \left|\frac{\nabla(u(\mathbf{x}+e_i h)-u(\mathbf{x}))}{h}\right|^2 d\mathbf{x} \leq C \int_{\tilde{\mathcal{C}}_{\delta/2}} \left|\frac{\partial u(\mathbf{x})}{\partial x_i}\right|^2 d\mathbf{x}, \tag{167}$$

where we have used the notation $\tilde{\mathcal{C}_\delta} = \{\mathbf{x};\ \text{dist}(\mathbf{x},\mathcal{C}) < \delta\} \subset \mathcal{D}$ introduced in Lemma 12.2 and chosen $\delta > 0$ small enough so that $\tilde{\mathcal{C}_\delta} \subset \mathcal{D}$.

From Lemma 12.2 and (167) we can conclude that $\nabla u(\mathbf{x})$ is weakly differentiable in x_i and

$$\int_{\mathcal{C}} \left|\nabla \frac{\partial u(\mathbf{x})}{\partial x_i}\right|^2 d\mathbf{x} \leq C_{\mathcal{C}} \int_{\tilde{\mathcal{C}}_{\delta/2}} \left|\frac{\partial u(\mathbf{x})}{\partial x_i}\right|^2 d\mathbf{x}, \tag{168}$$

this proves (165).

From Proposition 14.1 we may refine (168) to

$$\int_{\mathcal{C}} \left|\nabla \frac{\partial u(\mathbf{x})}{\partial x_i}\right|^2 d\mathbf{x} \leq C_{\mathcal{C}} \int_{\mathcal{D}} \left|\frac{\partial u(\mathbf{x})}{\partial x_i}\right|^2 d\mathbf{x}, \tag{169}$$

maybe for a different constant $C_{\mathcal{C}}$.[39]

If we sum (169) over $i = 1, 2, ..., n$ the estimate (166) and the theorem follows. □

We are now ready to prove that minimizers are actually solutions to $\Delta u(\mathbf{x}) = 0$.

Theorem 14.3. *Assume that u is a minimizer of the energy functional as in Theorem 13.1. Then $\Delta u(\boldsymbol{x}) = 0$ a.e. in $\mathcal{D}$.*

Proof: By Theorem 14.1 we conclude that u is weakly harmonic. From Theorem 14.2 it follows that $D^2 u \in L^2(K)$ for any compact set $K \subset \mathcal{D}$. We may therefore make an integration by parts in the definition of weak solution

$$0 = \int_{\mathcal{D}} \nabla v(\mathbf{x}) \cdot \nabla u(\mathbf{x}) d\mathbf{x} = -\int_{\mathcal{D}} v(\mathbf{x}) \Delta u(\mathbf{x}) d\mathbf{x}. \tag{170}$$

We need to show that this implies that $\Delta u = 0$ a.e.

In order to do that we let $w = \Delta u(\mathbf{x}) \chi_{\mathcal{C}}(\mathbf{x})$ for some compact set $\mathcal{C} \subset \mathcal{D}$. Then w is an L^2 function and we may consider the regularization $w_\epsilon(\mathbf{x}) \in$

[39] At times it is important to keep track of the size of the constants $C_{\mathcal{C}}$. In this case $C_{\mathcal{C}} \approx \frac{1}{\delta^4}$. But we will not care about such matters in this course.

$C_c^\infty(\mathcal{D})$, by Proposition 9.1, if $\epsilon > 0$ is chosen $\epsilon < \text{dist}(\mathcal{C}, \partial\mathcal{D})$. Furthermore, also by Proposition 9.1, $w_\epsilon \to \Delta u(\mathbf{x})\chi_{\mathcal{C}}(\mathbf{x})$ in L^2 as $\epsilon \to 0^+$. Using w_ϵ for v in (170) we derive that

$$0 = -\int_{\mathcal{D}} w_\epsilon(\mathbf{x})\Delta u(\mathbf{x})d\mathbf{x}. \tag{171}$$

Sending $\epsilon \to 0^+$ in (171) we may conclude that

$$0 = -\int_{\mathcal{C}} |\Delta u(\mathbf{x})|^2 d\mathbf{x}$$

for any compact set $\mathcal{C} \subset \mathcal{D}$. The Theorem follows. □

It is worth to remark that we use an argument that is often convenient in the previous proof. It is worth formalizing that argument into a Lemma.

Lemma 14.3. *If $u \in W^{1,2}(\mathcal{D})$ and if*

$$\int_{\mathcal{D}} \nabla v(\boldsymbol{x}) \cdot \nabla u(\boldsymbol{x})d\boldsymbol{x} = 0 \tag{172}$$

for all $v(\boldsymbol{x}) \in C_c^\infty(\mathcal{D})$ then $u(\boldsymbol{x})$ is a weak solution to $\Delta u(\boldsymbol{x}) = 0$.

Proof: We will argue by contradiction. If $u(\mathbf{x})$ is not weakly harmonic then there exists a $v \in W^{1,2}(\mathcal{D})$ with compact support such that

$$\int_{\mathcal{D}} \nabla v(\mathbf{x}) \cdot \nabla u(\mathbf{x})d\mathbf{x} = \delta \neq 0.$$

Arguing as in the previous proof we may define $w_\epsilon(\mathbf{x})$ to be a $C_c^\infty(\mathcal{D})$ approximation of v, as in Theorem 9.1. Since v has compact support it will follow that w_ϵ has compact support if ϵ is chosen small enough. Using $w_\epsilon \in C_c^\infty(\mathcal{D})$ in the equality (172) and sending $\epsilon \to 0^+$ will lead to a contradiction. □

The last theorem ties the calculus of variations to the theory of partial differential equations. This allows us to use variational methods to analyze solutions to partial differential equations and the other way around. We will not dig very deep into this theory since we are primarily interested i analysis. We will however indicate a few more results in order to show the power of the theory we have developed. The first one uses that Sobolev spaces are much more flexible which helps us to prove difficult results with little effort.

Proposition 14.2. *[The maximum principle.] Let u be a weakly harmonic and assume that $|u(\boldsymbol{x})| \le M$ on $\partial\mathcal{D}$. Then $|u(\boldsymbol{x})| \le M$ in $\overline{\mathcal{D}}$.*

Proof: From Proposition 9.4 it follows that $v(\mathbf{x}) = (u(\mathbf{x}) - M)^+ \in W^{1,2}(\mathcal{D})$. We may use $v(\mathbf{x})$ as a test-function in the definition of weakly harmonic and derive that

$$0 = \int_{\mathcal{D}} \nabla v(\mathbf{x}) \cdot \nabla u(\mathbf{x})d\mathbf{x} = \int_{\{u>M\}\cap\mathcal{D}} |\nabla v(\mathbf{x})|^2 d\mathbf{x},$$

where we used Proposition 9.4 to conclude that $\nabla v(\mathbf{x}) = \nabla u(\mathbf{x})\chi_{\{u>M\}\cap\mathcal{D}}(\mathbf{x})$ in the last equality.

From Theorem 12.3 we can conclude that $\|v\|_{L^2} = 0$, thus $v(\mathbf{x}) = (u(\mathbf{x}) - M)^+$ is zero a.e.

We may argue similarly to show that $(M - u(\mathbf{x}))^+ = 0$ a.e. $\square$

Remark: *One should notice that the above argument would not work if we try to prove the existence of minimizers in the space $C^1(\mathcal{D})$; this since $(u(\mathbf{x}) - M)^+ \notin C^1(\mathcal{D})$. We gain a something by the powerful flexibility of the Sobolev spaces.*

One thing that is rather disappointing with our existence theorem for minimizers, Theorem 13.1. We assume that the boundary data $f(\mathbf{x})$ are the trace of some function $w \in W^{1,2}(\mathcal{D})$. We know that

$$f \text{ is the trace of } w \in W^{1,2}(\mathcal{D}) \Rightarrow f \in L^2(\mathcal{D}).$$

But does the implication in the other direction hold? The answer is: No!

Example 1: *Let the following function be defined on $[-1, 1]$*

$$f(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Then $f(x)$ is not the trace of any function $u \in W^{1,2}(\mathcal{D})$ where $\mathcal{D} \subset \mathbb{R}^2$ is a domain such that $[-1, 1] \subset \partial\mathcal{D}$ and $Q_1^+(0) = \{(x, y);\ |x| < 1,\ 0 < y < 1\} \subset \mathcal{D}$.

In order to see this we notice that if we calculate the line integral along

$$P_r = \{\text{the path } (r, 0) \to (r, r) \to (-r, r) \to (-r, 0)\}$$

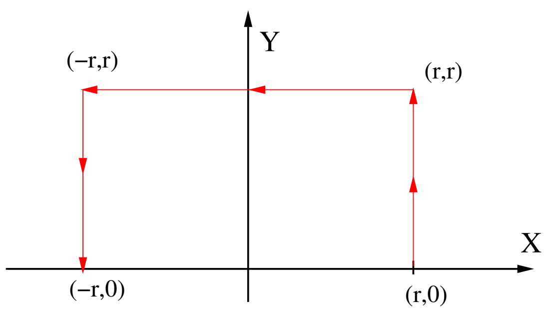


Figure: *The integration path P_r.*

then, for a.e. $r \in (0, 1)$ the following integral holds,

$$\int_{P_r} \frac{du}{d\eta} dt = u(-r, 0) - u(r, 0) = -2, \tag{173}$$

where $\frac{du}{d\eta}$ denotes the derivative in the direction of the path. From Hölders inequality we can conclude that

$$2 \leq \int_{P_r} \left|\frac{du}{d\eta}\right| dt \leq \left(\int_{P_r} dt\right)^{1/2} \left(\int_{P_r} \left|\frac{du}{d\eta}\right|^2 dt\right)^{1/2} = (4r)^{1/2} \left(\int_{P_r} \left|\frac{du}{d\eta}\right|^2 dt\right)^{1/2}$$

By rearranging terms and squaring we arrive at

$$\frac{1}{r} \leq \int_{P_r} \left|\frac{du}{d\eta}\right|^2 dt \leq \int_{P_r} |\nabla u|^2 \, dt. \tag{174}$$

Noticing that if we integrate the right side over $0 < r < 1$ we get $\|\nabla u\|^2_{L^2(Q_1^+(0))}$ whereas the left side diverges. We can conclude that $u \notin W^{1,2}(Q_1^+(0))$.

The above example can be refined to the following.

Example 2: *There are functions $f \in C([-1,1])$ that are not the trace of any function in $W^{1,2}(\mathcal{D})$ for any domain $Q_1^+(0) \subset \mathcal{D}$.*

To see this we just repeat the above argument, using that $u(-r,0)-u(r,0) = f(-r) - f(r)$, which leads to the corresponding analogue to (174)

$$\int_0^1 \frac{f(-r) - f(r)}{2r} dr \leq \int_{Q_1^+} |\nabla u(\mathbf{x})|^2 d\mathbf{x}.$$

If, for instance, f is the following continuous function

$$f(x) = \begin{cases} \frac{1}{|\ln(|x|)|} & \text{for } 0 < |x| < 1 \\ 0 & \text{for } x = 0, \end{cases}$$

then

$$\int_0^1 \frac{f(-r) - f(r)}{2r} dr = \int_0^1 \frac{1}{r|\ln(r)|} dr = \infty.$$

After the admittedly substantial work we have done in this course we are not able, due to Example 2, to prove that solutions exists for the following problem:

$$\begin{array}{ll} \Delta u(\mathbf{x}) = 0 & \text{in } \mathcal{D} \\ u(\mathbf{x}) = f(\mathbf{x}) & \text{on } \partial\mathcal{D}, \end{array}$$

where $f \in C(\partial\mathcal{D})$ and $\mathcal{D}$ is a C^1 domain. This is rather disappointing, and we will not prove that such solutions exists either. We will however sketch as much of the proof of such a theorem that we can do without getting into to much theory of partial differential equations. Partly we will do this in order to discuss what analysis is and partly because we would want to have some application of the main theorem in the next section.

Theorem 14.4. *Let $\mathcal{D}$ be a C^1 domain and $f \in C(\partial\mathcal{D})$. Then there exists a unique solution $u \in C^2(\mathcal{D}) \cap C(\overline{\mathcal{D}})$ to the following problem:*

$$\begin{array}{ll} \Delta u(\boldsymbol{x}) = 0 & \text{in } \mathcal{D} \\ u(\boldsymbol{x}) = f(\boldsymbol{x}) & \text{on } \partial\mathcal{D}. \end{array}$$

Sketch of a proof with big gaps: When we wanted to ascribe a *"length"* to a general measurable set $S \subset \mathbb{R}$ we approximated it by an open set U, such that $S \subset U$. Since open sets $U \subset \mathbb{R}$ are the countable union of open sets it was easy to assign a length (or measure) to U, taking the infimum of the measure of the open sets U gave us the measure of S. When trying to integrate a measurable function, f, we approximated f by simple functions whose integral we could calculate, using a limit argument or taking the supremum gave the integral of f. When showing that $\int_{\mathcal{D}} f_j(\mathbf{x})g(\mathbf{x})d\mathbf{x} \to \int_{\mathcal{D}} f_0(\mathbf{x})g(\mathbf{x})d\mathbf{x}$ in relation to weak convergence, $f_j \rightharpoonup f_0$ we approximated g by a dense subset of L^q and then passed to the limit (see Proposition 8.2). When we wanted to understand the derivatives of $u \in W^{1,p}(\mathcal{D})$ we approximated u by C^∞ functions and passed to the limit (see for instance Theorem 12.1). There is a pattern to all this: in order to do something we cannot do we approximate it with a problem that we can solve and then we pass to the limit.

In proving this Theorem we need to approximate f by functions for which we can solve the Dirichlet problem, that is with functions that are traces. In order to do that we start by noticing that we may write

$$f(\mathbf{x}) = \sum_{j=1}^{N} \psi_j(\mathbf{x}) f(\mathbf{x}),$$

where ψ_j is a partition of unity with respect to the cover $B_{2r_j}(\mathbf{x}^j)$ that exists in virtue of $\mathcal{D}$ being a C^1 domain (see Definition 12.1 for explanation of the balls $B_{2r_j}(\mathbf{x}^j)$). It is therefore enough to show that each function $\psi_j(\mathbf{x})f(\mathbf{x})$ can be approximated by the trace of some $w_j \in W^{1,2}(\mathcal{D})$.

Furthermore, we may straighten the boundary in $B_{2r_j}(\mathbf{x}^j)$ as in Step 1 in the proof of Theorem 12.2. It is therefore enough to show that a continuous function with support on a straight part of the boundary can be approximated by the trace some a $W^{1,2}$ function, we may therefore assume (w.l.o.g.) that $\psi_j(\mathbf{x})f(\mathbf{x})$ was already defined on a straight part of the boundary say on a part of the boundary where $x_n = 0$. In order to approximate $\psi_j(\mathbf{x})f(\mathbf{x})$ we use mollification, which is well defined on the compact straight part of the boundary, and define $g_\epsilon(\mathbf{x}) = \phi_\epsilon * (\psi_j f)(\mathbf{x})$. Then $g_\epsilon \in C_c^\infty$ on the boundary $\mathcal{D}$ and we may extend $g_\epsilon(\mathbf{x})$ to a C^∞ function in the entire domain $\mathcal{D}$. Since $g_\epsilon \in C^\infty$ on a compact set it follows that its trace is well defined and since g_ϵ is continuous the trace agrees with the values of g_ϵ on the boundary.

By Theorem 13.1 there exists a solution u_ϵ to the minimization problem with boundary data g_ϵ and by Theorem 14.3 it follows that $\Delta u_\epsilon(\mathbf{x}) = 0$ a.e.

Also, by point 2 in Proposition 9.1 $g_\epsilon(\mathbf{x}) \to f(\mathbf{x})$ uniformly on $\partial\mathcal{D}$. Which means that g_ϵ forms a Cauchy sequence in the sumremum norm:

$$\forall \kappa > 0,\ \exists \delta_\kappa > 0 \text{ s.t. } \epsilon_1, \epsilon_2 < \delta_\kappa \Rightarrow \sup_{\mathbf{x}\in\partial\mathcal{D}} |g_{\epsilon_1}(\mathbf{x}) - g_{\epsilon_2}(\mathbf{x})| < \kappa.$$

By the maximum principle, Proposition 14.2, it follows that

$$\sup_{\mathcal{D}} |u_{\epsilon_1}(\mathbf{x}) - u_{\epsilon_2}(\mathbf{x})| \le \sup_{\mathbf{x}\in\partial\mathcal{D}} |g_{\epsilon_1}(\mathbf{x}) - g_{\epsilon_2}(\mathbf{x})| < \kappa$$

and therefore that u_ϵ also forms a Cauchy sequence in the supremum norm.

So far this proof is rather complete, although without details. However in order to complete the proof we would have to prove that

$$u_\epsilon \in C^2(\mathcal{D}) \cap C(\overline{\mathcal{D}}). \tag{175}$$

One can indeed prove that (175) holds,[40] but that would take us to far afield from pure analysis so we will not prove that here.

In any case, (175) together with uniform convergence implies that $u_\epsilon \to u \in C(\overline{\mathcal{D}})$ (uniform limits of sequences of continuous functions converges to

[40]To be precise one can prove that $u_\epsilon \in C(\overline{\mathcal{D}})$ and that $u_\epsilon \in C^2(\mathcal{C})$ for any compact set $\mathcal{C} \subset \mathcal{D}$ and provide estimates on the norm $\|u_\epsilon\|_{C^2(\mathcal{C})}$ that are independent of ϵ. This proves that u_ϵ is two times continuously differentiable in $\mathcal{D}$. Often texts, but we will not, reserve the notation that $u \in C^2(\mathcal{D})$ if the second derivatives of u are bounded in the entire set $\mathcal{D}$, this does not hold in general under our assumptions.

continuous functions). Also, each u_ϵ is a weak solution to $\Delta u = 0$ and therefore

$$\int_{\mathcal{D}} \nabla u_\epsilon(\mathbf{x}) \cdot \nabla v(\mathbf{x}) d\mathbf{x} = 0, \tag{176}$$

for any $v \in C_c^\infty(\mathcal{D})$. By proposition 14.1

$$\|\nabla u_\epsilon\|_{L^2(\mathrm{spt}(v))} \leq C\|u_\epsilon\|_{L^2(\mathcal{D})} \leq C\|u_\epsilon\|_{L^\infty(\mathcal{D})} \leq C\left(\|f\|_{L\infty(\partial\mathcal{D})} + 1\right) \leq C,$$

where we also used Proposition 14.2 to estimate $\|u\|_{L^\infty(\mathcal{D})} \leq C\left(\|f\|_{L\infty(\partial\mathcal{D})} + 1\right)$ and we let the constant C differ in each occurrence. Since ∇u_ϵ is bounded in L^2 we may find a weakly convergent sub-sequence to u such that, by (176),

$$\int_{\mathcal{D}} \nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x}) d\mathbf{x} = 0,$$

it follows that u is weakly harmonic and, from Theorem 14.3, that $\Delta u = 0$. □

Remark: *You are not supposed to really understand the above proof. It felt worthwhile to include it in order to indicate a way ahead and how analysis is used to prove difficult theorems. It will also be our entrance point to the main theorem in the next section.*

14.1 Exercises:

1. Let $g \in W^{1,2}(\mathcal{D})$ and assume that $u(\mathbf{x})$ minimizes

$$\int_{\mathcal{D}} \left(|\nabla u(\mathbf{x})|^2 + 2u(x)g(x)\right) d\mathbf{x},$$

   in the set $K = \{u \in W^{1,2}(\mathcal{D});\ u = f \text{ on } \partial\mathcal{D}\}$ where $f(\mathbf{x})$ is some given function. Show that $u \in W^{2,2}(\mathcal{C})$ for any compact $\mathcal{C} \subset \mathcal{D}$.

   REMARK: *The same is true for $g \in L^2(\mathcal{D})$, can you prove it?*

2. Show that the function

$$u(x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \leq 0 \end{cases}$$

   is not a function in $W^{2,2}(-1,1)$.

3. Use the methodology of example 2 to show that, for any $1 < p \leq \infty$, there is a continuous functions $f \in C([-1,1])$ that is not the trace of any function in $W^{1,p}$.

15 Lecture 15. Harmonic measure and Riesz representation theorem.

If $\mathcal{D}$ is a C^1 domain then, by Theorem 14.4, there exists a harmonic function $u \in C(\overline{\mathcal{D}})$ for each continuous function $f \in C(\partial\mathcal{D})$. But can we calculate the values of $u(\mathbf{x})$? The answer is in general "no", but we can still try to say something about $u(\mathbf{x})$.

Given an $\mathbf{x} \in \mathcal{D}$ we can define a function $F : C(\partial\mathcal{D}) \mapsto \mathbb{R}$ according to so that $F(f) = u(\mathbf{x})$ where u is the harmonic function with boundary data f. This is a well defined function, but as far as we know it might be terrible complicated. But there is a very nice Theorem, *The Riesz Representation Theorem*, that states that any linear and bounded function $F : C(\partial\mathcal{D}) \mapsto \mathbb{R}$ may be represented as a measure. One might conclude the following Theorem.

Theorem 15.1. *Let $\mathcal{D}$ be a C^1 domain and $\boldsymbol{x} \in \mathcal{D}$. Then there exists a measure $\omega_{\boldsymbol{x}}$ defined on $\partial\mathcal{D}$ such that if u be the solution to*

$$\begin{array}{rl} \Delta u(\boldsymbol{x}) = 0 & \text{in } \mathcal{D} \\ u(\boldsymbol{x}) = f(\boldsymbol{x}) & \text{on } \partial\mathcal{D} \end{array}$$

and if $f \in C(\partial\mathcal{D})$ then

$$u(\boldsymbol{x}) = \int_{\partial\mathcal{D}} f(\boldsymbol{y}) d\omega_{\boldsymbol{x}}(\boldsymbol{y}).$$

The Riesz Representation Theorem has many other applications and it is considered to be a standard tool in mathematical analysis. Since classification theorems are so important in mathematics, I want to end the course with a classification theorem from analysis. Your reading for this lecture is:

Reading:

W. Rudin, Real and complex analysis: Pages 40-47

Recommended exercises: *2.4, 2.19*

H.L. Royden P.M Fitzpatrick, Real Analysis: Chapter 21.5

Recommended exercises: *21.46, 21.47, 21.48*

E. Stein R. Shakarchi, Real Analysis **Not covered**