Version: 3/18/07

Contents

1. Lim sup, Lim inf

Throughout these notes "$\limsup, \liminf$" appear. We realized, as the course went on, that understanding and use of these elementary notions needed some beefing up, and we do so retrospectively here, right at the top. If $\{a_j\}_{j=1}^{\infty}$ is any sequence of extended real numbers (elements of $[-\infty, \infty]$) then $\limsup_{j\to\infty} a_j$ and $\liminf_{j\to\infty} a_j$ are defined. This is in contrast with the plain old "limit," which may or may not exist, and this flexibility saves a lot of fussing around. Moreover, recall that the sequence $\{a_j\}_{j=1}^{\infty}$ converges iff

$$\limsup_{j\to\infty} a_j \leq \liminf_{j\to\infty} a_j$$

and then the common value is the limit.

There are various equivalent ways to define $\limsup, \liminf$.

Way 1:

$$\begin{aligned} &\text{(i)}\ \limsup_{j\to\infty} a_j = \lim_{m\to\infty} \sup\{a_j : j \geq m\} \\ &\text{(ii)}\ \liminf_{j\to\infty} a_j = \lim_{m\to\infty} \inf\{a_j : j \geq m\} \end{aligned} \tag{1.1}$$

The existence of the limits on the right hand sides above is due to monotonicity. For example, $\sup\{a_j : j \geq m\} \leq \sup\{a_j : j \geq m+1\}$.

Way 2: let us call $L \in [-\infty, \infty]$ a "subsequential limit" of $\{a_j\}_{j=1}^{\infty}$ if there exists a subsequence $\{a_{j_l}\}_{l=1}^{\infty}$ for which $a_{j_l} \to L$ as $l \to \infty$.

$$\begin{aligned} &\text{(i)}\ \limsup_{j\to\infty} a_j = \text{ largest subsequentional limit of } \{a_j\}_{j=1}^{\infty}, \\ &\text{(ii)}\ \liminf_{j\to\infty} a_j = \text{ smallest subsequentional limit of } \{a_j\}_{j=1}^{\infty}. \end{aligned} \tag{1.2}$$

You should convince yourself that there are largest and smallest subsequential limits; one way to do this is to verify that (1.1) and (1.2) define the same notions.

There are other definitions. For example, restricting ourselves to $\limsup$, $\limsup_{j\to\infty} a_j = L \in (-\infty, \infty]$ iff L is the largest element of $(-\infty, \infty]$ such that for every $M < L$

$$\{j : M \leq a_j\} \quad \text{is infinite.}$$

This way requires modification to handle $L = -\infty$, but we won't use it anyway.

Given two sequences $\{a_j\}_{j=1}^{\infty}, \{b_j\}_{j=1}^{\infty}$, we ask about the validity of the statements

$$\begin{aligned} &\text{(a)}\ \liminf_{j\to\infty}(a_j + b_j) \leq \liminf_{j\to\infty} a_j + \liminf_{j\to\infty} b_j, \\ &\text{(b)}\ \liminf_{j\to\infty} a_j + \liminf_{j\to\infty} b_j \leq \liminf_{j\to\infty}(a_j + b_j), \\ &\text{(c)}\ \limsup_{j\to\infty}(a_j + b_j) \leq \limsup_{j\to\infty} a_j + \limsup_{j\to\infty} b_j, \\ &\text{(d)}\ \limsup_{j\to\infty} a_j + \limsup_{j\to\infty} b_j \leq \liminf_{j\to\infty}(a_j + b_j), \\ &\text{(e)}\ \liminf_{j\to\infty}(a_j + b_j) \leq \liminf_{j\to\infty} a_j + \limsup_{j\to\infty} b_j, \\ &\text{(f)}\ \liminf_{j\to\infty}(-a_j) = -\limsup_{j\to\infty} a_j, \\ &\text{(g)}\ \liminf_{j\to\infty}(a_j - b_j) \leq \liminf_{j\to\infty} a_j - \limsup_{j\to\infty} b_j, \end{aligned} \tag{1.3}$$

as well as others like these.

This kind of issue, when you meet it, is generally not something you resolve while reading material in which it appears via "remembering" what is true and what is not. You figure it out on the spot, based on some experience. I generally use Way 2 myself to settle the issue in real time. For example, for (a) I think of a subsequence of the $a_j + b_j$ which approaches the $\liminf$ on the left - is there any reason the a_j and b_j should separately get small enough along this subsequence to bound them by the $\liminf$'s on the right? Nope, and counterexamples are easy to come by. On the other hand, re (b), if $a_{j_l} + b_{j_l} \to \liminf_{j\to\infty}(a_j + b_j)$, then we may pass to a further subsequence along which the a_{j_l} AND the b_{j_l} converge, making (b) obvious. Well, there is a problem here in that none of the statements make sense if the expression $\infty - \infty$ appears anywhere, so we'd have to rule that out. That is, (b) is true if the $a_j + b_j$ and the left hand side do not involve $\infty - \infty$. For another example, consider (e). I would choose a subsequence of the a_j's so that $a_{j_l} \to \liminf_{j\to\infty} a_j$, and then a further subsequence along which b_{j_l} converges. Than $a_{j_l} + b_{j_l}$ converges to something not more than the right hand side of (e) and not less than the left-hand side of (e) (again, we need to rule out $\infty - \infty$).

Questions for you: decide which of (a)-(f) are always true, and which are true if one of the sequences $\{a_j\}, \{b_j\}$ actually converges (in all cases, if there are no $\infty - \infty$'s).

2. Conventions and Notations

- If X is a set, then $\mathcal{P}(X)$ is the set of subsets of X.
- An *outer measure* on a set X is a monotone and countably subadditive mapping $\mu^* : \mathcal{P}(X) \to [0, \infty]$ which satisfies $\mu^*(\emptyset) = 0$. "Monotone" means that $\mu^*(A) \le \mu^*(B)$ whenever $A \subset B$. *Note that we always use the "countable" versions of things! Also note that μ^* can take the value $+\infty$.*
- A *σ-algebra* $\mathcal{M}$ of subsets of a set X is a subset of $\mathcal{P}(X)$ with $X \in \mathcal{M}$, which is closed under complements, intersections and countable unions.
- A *measurable space* is a pair $(X, \mathcal{M})$ where $\mathcal{M}$ is a σ-algebra of subsets of X.
- If X is a metric space, $(X, \mathcal{B})$ is the measurable space in which $\mathcal{B}$ is the σ-algebra of Borel subsets of X.
- If X is a set, $E \subset X$, and $\mathcal{S} \subset \mathcal{P}(X)$, then $\mathcal{S}$ is a *cover* of E if $E \subset \bigcup_{S\in\mathcal{S}} S$.
- A *measure* μ on a measurable space $(X, \mathcal{M})$ is a countably additive function $\mu : \mathcal{M} \to [0, \infty]$ which satisfies $\mu(\emptyset) = 0$. Note that μ may take the value ∞.
- A *measure space* is a triple $(X, \mathcal{M}, \mu)$ where $(X, \mathcal{M})$ is a measurable space and μ is a measure on $(X, \mathcal{M})$.
- "SS" means "Stein and Sakarchi," KF means Kolmogorov and Fomin, "EG" means "Evans and Gariepy," authors of "Measure Theory and Fine Properties of Functions" (available on reserve in the library), and Rudin indicates the book "Real and Complex Analysis" by Rudin (also on reserve).
- $Z := W$ means Z is defined to be W.

3. Some Review, Some New

3.1. **Construction of Outer Measures and Measures.** Most of the material below is found in SS Chapter 6, with a different twist, here or there.

Let X be a set and $\mathcal{S} \subset \mathcal{P}(X)$ such that X is covered by a countable subset of S and $\emptyset \in S$. Let $\nu : \mathcal{S} \to [0, \infty]$. Think of $A \in \mathcal{S}$ as a basic set whose "measure" $\nu(A)$ we know, eg, $\mathcal{S}$ could be the set of cubes in $\mathbb{R}^n$. Assume that

$$\nu(\emptyset) = 0. \tag{3.1}$$

Then for any $A \subset X$ we define

$$\nu^*(A) := \inf \left\{ \sum_{k=1}^{\infty} \nu(S_k) : S_k \in \mathcal{S}, A \subset \cup_{k=1}^{\infty} S_k \right\}. \tag{3.2}$$

By taking all but a finite number of the S_k above to be the empty set, we see that the infimum above includes the case of finite covers of A as well as countable covers. It is quite possible that $\nu^*(A) = \infty$ for some A's.

Proposition 3.1. *Under the above assumptions, ν^* is an outer measure on X.*

Proof. All the properties of an outer measure are obvious, except, perhaps, the subadditivity. Let $\{A_j\} \subset \mathcal{P}(X)$. For $\epsilon > 0$ and any $j = 1, 2, \ldots$, there is, by definition, a sequence of sets $S_{j,k} \in \mathcal{S}$ such that

$$A_j \subset \cup_{k=1}^{\infty} S_{j,k} \quad \text{and} \quad \nu^*(A_j) + \frac{\epsilon}{2^j} \geq \sum_{k=1}^{\infty} \nu(S_{j,k}).$$

Using this, we have

$$\begin{aligned} \cup_{j=1}^{\infty} A_j \subset \cup_{j,k=1}^{\infty} S_{j,k} \implies \nu^*\left(\cup_{j=1}^{\infty} A_j\right) &\leq \sum_{j,k=1}^{\infty} \nu(S_{j,k}) = \sum_{j=1}^{\infty} \left(\sum_{k=1}^{\infty} \nu(S_{j,k}) \right) \\ &\leq \sum_{j=1}^{\infty} \left(\nu^*(A_j) + \frac{\epsilon}{2^j} \right) = \sum_{j=1}^{\infty} \nu^*(A_j) + \epsilon. \end{aligned}$$

$\square$

Remark 3.2. We call ν^* the "outer measure" generated or defined by ν.

KF (Kolmogorov and Fomin) and SS (Stein and Shakarchi) define Lebesgue outer measure as the outer measure defined by taking ν to be volume defined on the set of all cubes (SS) or rectangles (KF); these yield the same outer measure (SS Example 4, pg 12). KF focuses on the algebraic structure of a "semi-ring," and therefore cannot restrict attention to cubes. KF also rules out $\nu^*(X) = \infty$, an important case.

We now think of starting with an "outer measure" ν^*, no matter where it came from.

If $\nu^*(X) < \infty$, one familiar (KF top of pg 32) way to try to define "measurable sets" associated with ν^* is via the equality of "inner" and "outer" measure, where inner-measure ν_* might be given by

$$\nu_*(E) = \nu^*(X) - \nu^*(X \setminus E). \tag{3.3}$$

Then measurability of $E \subset X$ in the sense $\nu^*(E) = \nu_*(E)$ just says, if $\nu^*(X) < \infty$,

$$\nu^*(X) = \nu^*(E) + \nu^*(X \setminus E) = \nu^*(X \cap E) + \nu^*(X \setminus E). \tag{3.4}$$

This is Definition 3', pg 32, of KF with X in place of E and E in place of A (as is consistent with the notation of SS). However, (3.3) is not defined when $\nu^*(X) = \nu^*(X \setminus E) = \infty$, and when $\nu^*(X) = \infty$, (3.4) is always satisfied.

Carathéodory discovered that a simple change makes everything sing. Instead of requiring (3.4) only for X, we ask that it hold for every subset A of X. In the case in which X is the square and ν^* is Lebesgue outer measure, there is no difference (KF Exercise 4, pg 15, and SS Exercise 3, pg 312).

Theorem 3.3. *Let ν^* be an outer measure on X. Then the set $\mathcal{M}$ of subsets of E of X which satisfy*

$$\nu^*(A) = \nu^*(A \cap E) + \nu^*(A \setminus E) \quad \textit{for all subsets } A \textit{ of } X \tag{3.5}$$

is a σ-algebra and ν^* restricted to $\mathcal{M}$ is a countably additive measure.

Definition 3.4. We will call the sets satisfying (3.5) the ν^* *measurable sets.*

The short, sweet proof is found in SS, pg 265. It is too much like other things you have done and are bored with to give in class.

Since all the business about semi-rings and such has disappeared, what was it good for? Well, for one thing, that discussion contained proofs that a "measure" (measures of any kind are always countably additive for us) originally given on a semi-ring extended to the sigma-ring generated by the original ring, and hence to all Borel sets (in the Lebesgue context). Let us think about this relative to Theorem 3.3.

Suppose ν is in fact a measure on a ring $\mathcal{S} = \mathcal{R} \subset \mathcal{P}(X)$. It follows from KF Theorem 2, pg 29 (0r SS Lemma 6.1.4) that for $E \in \mathcal{R}$ we have $\nu^*(E) \geq \nu(E)$. Since $\nu^*(E) \leq \nu(E)$, (E is a cover of E), we have

$$\nu^*(E) = \nu(E) \quad \text{for} \quad E \in \mathcal{R}. \tag{3.6}$$

Moreover, if $A \in \mathcal{P}(X)$, $E \in \mathcal{R}$, and $\{A_j\} \subset \mathcal{R}$ is a cover of A such that $\nu^*(A) + \epsilon \geq \sum_{j=1}^{\infty} \nu(A_j)$, we have

$$\begin{aligned} \nu^*(A) + \epsilon \geq \sum_{j=1}^{\infty} \nu(A_j) &= \sum_{j=1}^{\infty} (\nu(A_j \cap E) + \nu(A_j \setminus E))) \\ &\geq \nu^*(A \cap E) + \nu^*(A \setminus E). \end{aligned}$$

The inequality $\nu^*(A) \le \nu^*(A \cap E) + \nu^*(A \setminus E)$ always holds, so we conclude that $\mathcal{R}$ is contained in the set of ν^*-measurable sets. From this and (3.6), we have that the restriction of ν^* to the ν^*-measurable sets is an extension of ν on $\mathcal{R}$.

The following theorem relates measures given by the Carathéodory theorem to topology, if there is topology to consider.

Theorem 3.5. *Let X, d be a metric space and ν^* be an outer measure on the subsets of X. Further, assume that ν^* is a "metric outer measure," that is, if $A, B \subset X$ and*

$$\operatorname{dist}(A, B) := \inf\{d(x, y) : x \in A, y \in B\} > 0, \tag{3.7}$$

then

$$\nu^*(A \cup B) = \nu^*(A) + \nu^*(B). \tag{3.8}$$

Then the Borel sets of X are ν^* measurable.

See SS Thm 1.2 pg 267. Lebesgue outer measure is a metric outer measure.

3.2. **Hausdorff Measures.** Let X, d be a metric space. If $A \subset X$, set

$$\operatorname{diam}(A) = \sup\{d(x, y) : x, y \in A\}; \tag{3.9}$$

$\operatorname{diam}(A)$ is the "diameter of A." We put $\operatorname{diam}(\emptyset) = 0$. Let D^δ be the set of subsets of X whose diameter is at most $\delta > 0$. Assume that there is a countable subset of D^δ which covers X. Let $\alpha \ge 0$ and take $\mathcal{S} = D^\delta$ and $\nu(A) = \operatorname{diam}(A)^\alpha$ in Proposition 3.1. The result is an outer measure $\mathcal{H}^\delta_\alpha$. Since D^δ includes fewer sets as δ gets smaller, $\delta \to \mathcal{H}^\delta_\alpha$ increases as δ decreases and the limit

$$\mathcal{H}^*_\alpha(A) = \lim_{\delta \downarrow 0} \mathcal{H}^\delta_\alpha(A) \tag{3.10}$$

exists (in $[0, \infty]$) for each $A \subset X$. It follows from the next result that $\mathcal{H}^*_\alpha$ is an outer measure (take $\nu^*_j = \mathcal{H}^{1/j}_\alpha$.)

Proposition 3.6. *Let ν^*_j, $j = 1, 2, 3, \ldots$, be an increasing sequence of outer measures on a set X; that is, $\nu^*_j(A) \le \nu^*_{j+1}(A)$, for $A \subset X$. Then $\nu^*(A) := \lim_{j\to\infty} \nu^*_j(A)$ defines an outer measure ν^* on X.*

Proof. It is immediate that ν^* is monotone and $\nu^*(\emptyset) = 0$. Next suppose that S_k, $k = 1, 2, \ldots$, is a countable cover of $A \subset X$. Then $\nu^*_j \le \nu^*$ implies

$$\nu^*_j(A) \le \sum_{k=1}^\infty \nu^*_j(S_k) \le \sum_{k=1}^\infty \nu^*(S_k).$$

Letting $j \to \infty$ on the left yields the subadditivity. $\square$

Proposition 3.7. *The outer measure $\mathcal{H}^*_\alpha(A)$ given by (3.10) is a metric outer measure.*

Proof. Let $A, B \subset X$ and $\operatorname{dist}(A,B) = \gamma > 0$. Let $0 < \delta < \gamma$. Then any set $S \in D^\delta$ cannot meet both A and B, so any cover of $A \cup B$ by sets in D^δ can be reduced to a cover of B and a disjoint cover of A by throwing out the sets of the cover which do not meet $A \cup B$. It follows that

$$\mathcal{H}^\delta_\alpha(A \cup B) \geq \mathcal{H}^\delta_\alpha(A) + \mathcal{H}^\delta_\alpha(B),$$

and then, letting $\delta \downarrow 0$,

$$\mathcal{H}^*_\alpha(A \cup B) \geq \mathcal{H}^*_\alpha(A) + \mathcal{H}^*_\alpha(B).$$

The opposite inequality follows from subadditivity. □

It follows from Theorem (3.5) that $\mathcal{H}^*_\alpha$ restricted to the Borel sets $\mathcal{B}$ of X is a measure on $(X, \mathcal{B})$.

Theorem 3.8. *Suppose $A \in \mathcal{P}(X)$ and $0 \leq \alpha < \beta$. If $\mathcal{H}^*_\beta(A) > 0$, then $\mathcal{H}^*_\alpha(A) = \infty$.*

Proof. The assumption $\mathcal{H}^*_\beta(A) > 0$ and the definitions imply that there are positive numbers $\delta_0 < 1$ and κ such that

$$\sum_{j=1} \operatorname{diam}(A_j)^\beta \geq \kappa \quad \text{whenever} \quad \{A_j\} \subset D^{\delta_0} \quad \text{is a cover of} \quad A.$$

But then

$$\sum_{j=1} \operatorname{diam}(A_j)^\alpha = \sum_{j=1} \operatorname{diam}(A_j)^\beta \operatorname{diam}(A_j)^{\alpha-\beta} \geq \sum_{j=1} \operatorname{diam}(A_j)^\beta \delta^{\alpha-\beta} \geq \frac{\kappa}{\delta^{\beta-\alpha}}$$

whenever $\{A_j\} \subset D^\delta$ is a cover of A and $0 < \delta < \delta_0$. The result follows. □

Corollary 3.9. *Let $A \in \mathcal{P}(X)$ and $\hat\alpha = \sup\{\alpha \geq 0 : \mathcal{H}^*_\alpha(A) = \infty\}$. Then*

$$\mathcal{H}^*_\alpha(A) = \begin{cases} \infty & \text{for} \quad 0 \leq \alpha < \hat\alpha, \\ 0 & \text{for} \quad \hat\alpha < \alpha. \end{cases} \tag{3.11}$$

Proof. Suppose $0 \leq \alpha < \hat\alpha$. By the definition of $\hat\alpha$, there is an α', $\alpha < \alpha' \leq \hat\alpha$, such that $\mathcal{H}^*_{\alpha'}(A) = \infty$. $\mathcal{H}^*_\alpha(A) = \infty$ now follows from Theorem 3.8. Similarly, if $\hat\alpha < \alpha$, then $\mathcal{H}^*_\alpha(A) > 0$ and Theorem 3.8 contradict the definition of $\hat\alpha$. Hence $\mathcal{H}^*_\alpha(A) = 0$. □

Definition 3.10. Let X, d be a metric space and $A \in \mathcal{P}(X)$. Then the *Hausdorff dimension* of A is the number $\hat\alpha$ of Corollary 3.9.

Corollary 3.11. *Let $X = I\!R^n$ with the Euclidean metric. If $0 \leq \alpha < n$, then $(I\!R^n, \mathcal{B}, \mathcal{H}^*_\alpha|_\mathcal{B})$ is not σ-finite.*

3.3. **Integration Over Sets of Infinite Measure.** If $(X, \mathcal{M}, \mu)$ is a measure space and $\mu(X) = \infty$, what definition of

$$\int_X f \, d\mu \tag{3.12}$$

shall we take? The use of "uniform convergence" in defining (3.12) as in KF goes out the window if $\mu(X) = \infty$. Eg, with Lebesgue measure on the line, the functions $f_k(x) = \frac{1}{k}\chi_{[0,k]}$ converge uniformly to 0, but $\int_R f_k \, dx = 1$ for all k.

We assume that $f : X \to [0, \infty)$ is measurable. In SS (pgs 27, 274), a simple function g is a *finite* sum of the form

$$g(x) = \sum_{j=1}^{N} a_k \chi_{E_k}(x) \tag{3.13}$$

where the $a_k \in \mathbb{R}$ and the E_k are measurable sets of *finite* measure. Clearly we want $\int_X g\,d\mu = \sum_{j=1}^{N} a_k \mu(E_k)$. Then the integral (3.12) is defined by SS to be

$$\int_X f\,d\mu = \sup \left\{ \int_X g\,d\mu : g \quad \text{simple}, \quad g \le f \right\}. \tag{3.14}$$

Now one possibility is that $\mathcal{M} = \{\emptyset, X\}$ and $\mu(X) = \infty$. Then the *only* simple functions in the SS sense are of the form $a\chi_{\{\emptyset\}}$! Moreover, the only measurable functions are constants, and the definition (3.14) implies

$$\int_X 1\,d\mu = 0 \quad \text{since } \chi_{\{\emptyset\}}(x) = 0\ \forall x, \tag{3.15}$$

while clearly we would prefer $\int_X 1\,d\mu = \mu(X) = \infty$.

Thus one needs either a technical condition on $(X, \mathcal{M}, \mu)$ to guarantee against this sort of anomally or to modify the definition (3.14). SS uses a condition called "σ-finite" which reads as follows:

Definition 3.12. The measure space $(X, \mathcal{M}, \mu)$ is σ-finite if there exists an increasing sequence of measurable sets F_k of finite measure such that $X = \cup_{k=1}^{\infty} F_k$.

For example, $(\mathbb{R}^n, \mathcal{B}, \mathcal{L}^n)$ is σ-finite, where $\mathcal{B}$ is the Borel σ-algebra (we could use the Lebesgue σ-algebra as well, but don't have enough symbols at the moment) and $\mathcal{L}^n$ is Lebesgue measure.

We have met important measure spaces which are not σ-finite, namely $(\mathbb{R}^n, \mathcal{B}, H^*_\alpha\big|_{\mathcal{B}})$ for $0 \le \alpha < n$. The quickest and slickest presentation of integration in the general case is given in Rudin (early pages). Rudin notes "Throughout integration theory, one inevitably encounters ∞." For example, the Lebesgue measure of $\mathbb{R}^n$ is ∞, and $\limsup_{j\to\infty} f_j$ might take ∞ as a value even if the individual f_j do not. So let's talk a bit about $\pm\infty$.

The fundamental thing is to define the integral (3.12), or, a bit more generally,

$$\int_E f\,d\mu \tag{3.16}$$

when E is a measurable set and $f : X \to [0, \infty]$ is measurable. "measurable" will always mean that $f^{-1}(\mathcal{O})$ is measurable when $\mathcal{O}$ is an open subset of the range of f. The open sets $\mathcal{O}$ of $[0, \infty]$ are those for which $\mathcal{O} \cap [0, \infty)$ is open in $[0, \infty)$ and, if $\infty \in \mathcal{O}$, then it must contain a basic neighborhood of ∞, which is a set of the form $(a, \infty]$.

Suppose $f = a\chi_E$ for some measurable E and $a \in [0, \infty]$. Clearly we want

$$\int_X f\,d\mu = a\mu(E),$$

and we need to interpret this in cases in which $a = \infty$ and/or $\mu(E) = \infty$. There is no problem with $\infty(\infty) = \infty$, we all agree with that, and there is no problem with $a(\infty) = \infty$ or $\infty\mu(E) = \infty$ if $a > 0$ and $\mu(E) > 0$, we all agree with that. What about $0(\infty)$ and $\infty(0)$? Since we all agree that $\int_R 0\,dx = 0$, we should take $0(\infty) = 0$. Assigning the right value to $\int_{\{0\}} \infty\,dx$ is less obvious, but what turns out to work is $\infty(0) = 0$, which we need to make our "multiplication" commute anyway. In another view, if we want to evaluate $\infty\chi_E$ at a point $x \notin E$, we get $\infty(0)$, and clearly want this to be 0. Algebra in $[-\infty, \infty]$ is obvious once we made the decision that $0(\pm\infty) = \pm\infty 0 = 0$, except we don't know what to do with $\infty - \infty$ yet - and we never will. We rule out dealing with $\infty - \infty$, it remains undefined.

For us, a simple function s is any function with a representation as finite sum

$$s = \sum_{j=1}^{N} a_j \chi_{E_j} \tag{3.17}$$

where the E_j are measurable sets, $a_j \in [-\infty, \infty]$ and evaluating $s(x)$ does not yield expressions of the form $\infty - \infty$. Equivalently, $s : X \to [-\infty, \infty]$ is measurable and $s(X)$ is a finite set - then we have (3.17) with the a_j the distinct values of s and the pairwise disjoint sets $E_j = s^{-1}(\{a_j\})$. For such a simple function s, let us assume the a_j are the distinct values of s, and put

$$\int_E s\,d\mu = \sum_{j=1}^{N} a_j \mu(E_j \cap E) \tag{3.18}$$

provided that not both ∞ and $-\infty$ appear in the sum with our algebraic conventions. When this holds, we say that "$\int_E s\,d\mu$ is defined."

Rudin's definition of $\int_E f\,d\mu$ for $f : X \to [0, \infty]$ and $E \in \mathcal{M}$ amounts to

$$\int_E f\,d\mu = \sup\left\{\int_E s\,d\mu : s \quad \text{is simple and} \quad 0 \le s \le f\right\}. \tag{3.19}$$

Let Us Agree: All functions are hereafter assumed to be measurable in whatever context they appear, unless otherwise said. Similarly, unless otherwise said, all sets which appear are to be measurable.

Actually, Rudin requires s above to have finite values. A few clunky exercises to warm you up to "∞" and (3.19) are given in (4), (5), (6).

For general, not necessarily nonnegative, f Rudin defines

$$\int_X f\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu \tag{3.20}$$

where $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ are the positive and negative parts of f, PROVIDED that at least one of $\int_X f^\pm\,d\mu$ is finite. Thus $\int_X f\,d\mu$ might be $+\infty$ or $-\infty$.

We also want to integrate complex valued functions. This is done as follows: If $u, v : X \to \mathbb{R}$, and

$$f(x) = u(x) + iv(x)$$

then we define

$$\int_E f \; d\mu = \int_E u \, d\mu + i \int_E v \, d\mu \tag{3.21}$$

provided $\int_E |u| \, d\mu$ and $\int_E |v| \, d\mu$ are both finite. Note that $\max(|u|, |v|) \leq |f| \leq |u| + |v|$ implies

$$\int_X |u| \, d\mu \; \& \int_X |v| \, d\mu < \infty \Longleftrightarrow \int_X |f| \, d\mu < \infty.$$

Here is a sequence of results concerning the definition (3.19), in the order Rudin establishes them. In the background is a measure space $(X, \mathcal{M}, \mu)$.

- Let $f : X \to [0, \infty]$. Then there is a sequence of simple functions $s_j : X \to [0, \infty)$ such that $0 \leq s_1 \leq s_2, \ldots$ and $f(x) = \lim_{j \to \infty} s_j(x)$. (pg 15)
- If $0 \leq f \leq g$, then $\int_E f \; d\mu \leq \int_E g \, d\mu$. (pg 19)
- If $A \subset B$ and $f \geq 0$, then $\int_A f \; d\mu \leq \int_B f \, d\mu$. (pg 19)
- If $f \geq 0$ and $0 \leq c < \infty$, then $c \int_E f \, d\mu = \int_E cf \, d\mu$. (pg 20)
- Let $\{f_k\}$ be a nondecreasing sequence of nonnegative functions on X such that $f_k(x) \to f(x)$ for $x \in X$. Then $\int_X f_k \; d\mu \to \int_X f \, d\mu$ (Lebegue's Monotone Convergence Theorem). (pg 21)
- Let $f_k : X \to [0, \infty]$, $k = 1, 2, \ldots$, and

  $$f(x) = \sum_{k=1}^{\infty} f_k(x).$$

  Then

  $$\int_X f \; d\mu = \sum_{k=1}^{\infty} \int_X f_k \, d\mu.$$

  (Pg 22)
- If $f_k : X \to [0, \infty]$ for $k = 1, 2, \ldots$, then

  $$\int_X \left(\liminf_{k \to \infty} f_k \right) \, d\mu \leq \liminf_{k \to \infty} \int_X f_k \, d\mu$$

  (Fatou's Lemma). (pg 23)
- If $f, g : X \to [0, \infty]$, then

  $$\phi(E) := \int_E f \; d\mu \quad (E \in \mathcal{M})$$

  defines a measure ϕ on $\mathcal{M}$ and

  $$\int_X g \, d\phi = \int_X gf \, d\mu.$$

(pg 23)

- Let $f : X \to \mathbb{C}$ and $\int_X |f|\,d\mu < \infty$. Then
$$|\int_X f\,d\mu| \le \int_X |f|\,d\mu.$$
(Page 26)

- (Lebesgue's Dominated Convergence Theorem) Let $f_j, f : X \to \mathbb{C}$, $g_j, g : X \to [0,\infty)$, $|f_j| \le g_j$ a.e., $g_j \to g$ a.e. and
$$\int_X g_j\,d\mu \to \int_X g\,d\mu < \infty.$$
Then $\int_X |f - f_j|\,d\mu \to 0$. This is a small variant of Rudin pg 26, and it has the same proof. Applying Fatou's Lemma to the sequence $g_j + g - |f - f_j|$, which is nonnegative, yields
$$\begin{aligned}\int_X 2g\,d\mu &= \int_X \liminf_{j\to\infty}(g_j + g - |f - f_j|)\,d\mu \\ &\le \liminf_{j\to\infty}\left(\int_X (g_j+g)\,d\mu - \int_X (|f - f_j|\right)\,d\mu \\ &= \int_X 2g\,d\mu - \limsup_{j\to\infty}\int_X |f - f_j|\,d\mu,\end{aligned}$$
which implies $\limsup_{j\to\infty}\int_X |f - f_j|\,d\mu = 0$.

4. The $L^p(\mu)$ Spaces

Let $(X, \mathcal{M}, \mu)$ be a measure space. If $f : X \to [-\infty,\infty]$, then $\int_X f\,d\mu$ is defined ((3.20)) if one of $\int_X f^{\pm}\,d\mu$ is finite. If they are both finite, then $\int_X f\,d\mu \in \mathbb{R}$ and $\int_X |f|\,d\mu < \infty$. The "space" $L^1(\mu)$ is the set of all extended real valued functions with this property:

$$L^1(\mu) = \left\{ f : X \to [-\infty,\infty] : \int_X |f|\,d\mu < \infty \right\}. \tag{4.1}$$

Recall that, in the background, we are assuming the modifier "measurable" without writing it; f above is to be measurable. If $f \in L^1(\mu)$, then the set $\{x : f(x) = \pm\infty\}$ has measure 0 and the function $\tilde{f}$ which agrees with f on $\{f \ne 0\}$ and for which $\tilde{f}(x) = 0$ if $f(x) = \pm\infty$ agrees with f almost everywhere. Moreover, for every set $E \in \mathcal{M}$, $\int_E \tilde{f}\,d\mu = \int_E f\,d\mu$. Sets of measure 0 don't matter for integration theory (partly because $0\infty = \infty 0 = 0$). We assume you know this. In particular, if $N \in \mathcal{M}$ is a "null set" (ie, $\mu(N) = 0$), and $f : X \setminus N \to [0,\infty]$, the integrals $\int_E f\,d\mu$ can be defined as $\int_E \tilde{f}\,d\mu$ where $\tilde{f}$ is given any value on N. If $(X, \mathcal{M}, \mu)$ is *complete*, ie, all subsets of null sets belong to $\mathcal{M}$, then $\tilde{f}$ can be defined on N in any way at all.

The set $L^1(\mu)$ just defined is the "real" version. Similarly, we can consider complex valued functions, and the spaces $L^p(\mu)$, $1 \le p < \infty$, real or complex, are defined by

$$L^p(\mu) := \left\{ f : \int_X |f|^p \, d\mu < \infty \right\}. \tag{4.2}$$

In general, we put

$$\|f\|_{L^p(\mu)} := \left(\int_X |f|^p \, d\mu \right)^{1/p}, \tag{4.3}$$

whether or not $\|f\|_{L^p(\mu)}$ is finite ($(\infty)^{1/p} = \infty$). We'll see why the definition (4.2) is a smart idea later. Thus

$$L^p(\mu) = \left\{ f : \|f\|_{L^p(\mu)} < \infty \right\}.$$

The space $L^\infty(\mu)$ is not defined by integration. The quantity (4.3) is replaced by

$$\|f\|_{L^\infty(\mu)} = \inf \left\{ 0 \le M : |f| \le M \text{ a.e} \right\}. \tag{4.4}$$

Lemma 4.1. $|f| \le \|f\|_{L^\infty(\mu)}$ *a.e.*

Proof. By definition, $|f| \le \|f\|_{L^\infty(\mu)} + \frac{1}{k}$ except on a null set N_k. But then $|f| \le \|f\|_{L^\infty(\mu)}$ except on the null set $\cup_{k=1}^\infty N_k$. □

Theorem 4.2. *For $1 \le p \le \infty$, the function $d_p : L^p(\mu) \times L^p(\mu) \to [0, \infty)$ given by*

$$d_p(f, g) := \|f - g\|_{L^p(\mu)}$$

satisfies $d_p(f, g) = d_p(g, f)$ (symmetry) and the triangle inequality

$$d_p(f, g) \le d_p(f, h) + d_p(h, g) \quad (f, g, h \in L^p(\mu)). \tag{4.5}$$

Moreover, if $\{f_k\}$ is a sequence in $L^p(\mu)$ and

$$\lim_{j,k\to\infty} d_p(f_k, f_j) = 0, \tag{4.6}$$

then there exists $f \in L^p(\mu)$ such that

$$\lim_{k\to\infty} d_p(f_k, f) = 0. \tag{4.7}$$

We will come back to the proof of this theorem later. Right now we don't even know that $d_p(f, g) < \infty$, but we will shortly see this is so.

The function d_p has all the properties of a metric, except one. It is possible that $d_p(f, g) = 0$ even if $f \ne g$. However, $d_p(f, g) = 0$ only if $\{f \ne g\}$ is a null set. Suppose X is any set and $d : X \times X \to [0, \infty)$ is symmetric and satisfies the triangle inequality, but it is not necessarily true that $d(x, y) = 0$ only if $x = y$. Let us call such a thing a "pseudo metric." Then we can form a metric space in the following way: let $x \sim y$ iff $d(x, y) = 0$. Then "$\sim$" is an equivalence relation and the space

$$X/\sim = \{[x] : x \in X\} \quad \text{where} \quad [x] = \{y \in X : x \sim y\} \tag{4.8}$$

can be equipped with the metric $\tilde{d}$ given by

$$\tilde{d}([x],[y]) = d(x,y).$$

This metric space is *complete* iff

$$\lim_{j,k\to\infty} d(x_j,x_k) = 0 \implies \exists x \in X \ni \lim_{j\to\infty} d(x_j,x) = 0.$$

Why do we call $L^p(\mu)$ a "space?" First, it is clear that if c is a scalar, then

$$\|cf\|_{L^p(\mu)} = |c|\|f\|_{L^p(\mu)}.$$

so $f \in L^p(\mu) \implies cf \in L^p(\mu)$. Next, if $1 \le p < \infty$,

$$|f+g|^p \le (2\max(|f|,|g|))^p \le 2^p(|f|^p + |g|^p), \tag{4.9}$$

so

$$f,g \in L^p(\mu) \implies f+g \in L^p(\mu).$$

Altogether then, $L^p(\mu)$ is a vector space if $1 < p < \infty$. It is also a vector space if $p=\infty$, because $|f+g| \le |f| + |g|$.

The inequality (4.9) is crude. The sharp result is *Minkowski's Inequality*, which states

$$\|f+g\|_{L^p(\mu)} \le \|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)} \tag{4.10}$$

for $1 \le p \le \infty$. The cases $p = 1, \infty$ are obvious. The cases $1 < p < \infty$ follow from the *Hölder inequality,* which states that if $1 < p < \infty$ and q is defined by

$$\frac{1}{p} + \frac{1}{q} = 1, \quad \text{equivalently,} \quad q = \frac{p}{p-1}, \tag{4.11}$$

then

$$f \in L^p(\mu), g \in L^q(\mu) \implies fg \in L^1(\mu) \quad \& \quad \int_X |fg|\,d\mu \le \|f\|_{L^p(\mu)}\|g\|_{L^q(\mu)}. \tag{4.12}$$

The number q is called the *Hölder conjugate* of p and one writes $q = p'$. The term "conjugate" has to do with the fact that $(p')' = p$. One sets $1' = \infty$ and $\infty' = 1$, which is consistent with (4.9).

In the case $p = 1, q = \infty$, (4.12) is immediate from

$$|fg| \le |f|\|g\|_{L^\infty(\mu)} \quad \text{a.e.}$$

To prove (4.12) for $1 < p < \infty$, we recall that if $a, b \ge 0$, then

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}. \tag{4.13}$$

This is clear if $ab = 0$. If $a, b > 0$, divide both sides by a^p and note that $(ab/a^p)^q = b^q/a^p$ because of (4.11), to obtain the equivalent form

$$\xi - \frac{\xi^q}{q} \le \frac{1}{p} \quad (\xi = b/a^{p-1} \ge 0).$$

The maximum of $\xi - \xi^q/q$ (which is positive for small $\xi > 0$ and negative for large ξ, so it has a max) occurs when the derivative vanishes, or $1 = \xi^{q-1}$, or $\xi = 1$, where equality holds above. Since equality holds only at $\xi = 1$, the inequality (4.13) is strict except when $b^q = a^p$. This Math 3A proof of (4.13) is the one I can find whenever I want it, but it isn't elegant. See Rudin pg 64 eqn (5) for a slick proof. Writing

$$ab = \frac{a}{r}(rb) \quad \text{for } r > 0,$$

we also have

$$ab \leq \frac{a^p}{r^p p} + \frac{r^q b^q}{q}. \tag{4.14}$$

To prove (4.13), we may assume that $f, g \geq 0$. Using $a = f(x), b = g(x)$ and integrating yields

$$\int_X fg\, d\mu \leq \frac{\|f\|^p_{L^p(\mu)}}{r^p p} + \frac{\|g\|^q_{L^q(\mu)} r^q}{q}.$$

We may assume $\|f\|_{L^p(\mu)} > 0, \|g\|_{L^q(\mu)} > 0$ (otw, $f = 0$ or $g = 0$ ae and then $fg = 0$ ae). Minimizing the right hand side over $r \geq 0$, we find, using $p + q = pq$,

$$r = \frac{\|f\|^{1/q}_{L^p(\mu)}}{\|g\|^{1/p}_{L^q(\mu)}},$$

and, for this value of r,

$$\frac{\|f\|^p_{L^p(\mu)}}{r^p} = \|g\|^q_{L^q(\mu)} r^q = \|f\|_{L^p(\mu)} \|g\|_{L^q(\mu)},$$

whence the result.

Since $|f + g| \leq |f| + |g|$, to prove the Minkowski inequality (4.10), we may assume $f, g \geq 0$. The Minkowski inequality then follows from

$$\begin{aligned}
\int_X |f+g|^p\, d\mu &= \int_X (f+g)(f+g)^{p-1}\, d\mu \\
&= \int_X f(f+g)^{p-1}\, d\mu + \int_X g(f+g)^{p-1}\, d\mu \\
&\leq \|f\|_{L^p(\mu)} \|(f+g)^{p-1}\|_{L^q(\mu)} + \|g\|_{L^p(\mu)} \|(f+g)^{p-1}\|_{L^q(\mu)}
\end{aligned}$$

and $(f+g)^{(p-1)q} = (f+g)^p$. Thus the above amounts to

$$\|f+g\|^p_{L^p(\mu)} \leq (\|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)}) \|f+g\|^{p/q}_{L^p(\mu)}.$$

Since $p/q = p - 1$, we are done.

Proof of Theorem 4.2. The (pseudo) metric d_p was defined by $d_p(f,g) = \|f - g\|_{L^p(\mu)}$. On the other hand, the Minkowski inequality tells us that

$$\|f - g\|_{L^p(\mu)} = \|f - h + (h - g)\|_{L^p(\mu)} \leq \|f - h\|_{L^p(\mu)} + \|h - g\|_{L^p(\mu)} = d_p(f,h) + d_p(h,g),$$

which is the triangle inequality for d_p.

In general, a (pseudo) *norm* $\| \ \|$ on a vector space V is a mapping

$$\| \ \| : V \to [0,\infty)$$

such that

$$\begin{aligned} &\|0\| = 0, \\ (4.15) \qquad &\|x+y\| \le \|x\| + \|y\| \quad (x,y \in V) \\ &\|cx\| = |c|\|x\| \quad (c \text{ a scalar, } x \in V). \end{aligned}$$

The "pseudo" refers to the possibility that $\|x\| = 0$ even if x is not the zero vector 0 of V. If also $\|x\| = 0 \implies x = 0$, then $\| \ \|$ is a *norm* on V and $(V, \| \ \|)$ is a *normed vector space.* What we have just seen is that if $\| \ \|$ is a pseudo norm on V, then $d(x,y) = \|x-y\|$ is a pseudo metric on V. Then $x \sim y$ iff $\|x-y\| = 0$ is an equivalence relation on $V/\sim$ (see Exercise 7) and $V/\sim$ is a vector space with the addition $[x]+[y] = [x+y]$. Moreover, $\|[x]\|^{\sim} = \|x\|$ is a *norm* on $V/\sim$.

In sum, $L^p(\mu)$ is a normed vector space, provided one regards its elements as the "cosets"

$$[f] = \{g : X \setminus N \to \mathbb{R} : N \quad \text{is a null set and} \quad f = g \text{ a.e.}\}.$$

where $f \in L^p(\mu)$. One does this, and then forgets about it, and acts as if the elements of $L^p(\mu)$ are functions, without further pedantry.

We turn to the issue of completeness. Let us recall "convergence in measure." Consider a sequence of functions $\{f_j\}$ and another function f. One says that $f_j \to f$ in measure if

$$(4.16) \qquad \forall \epsilon > 0 \quad \lim_{j\to\infty} \mu\left(\{x : |f_j(x) - f(x)| > \epsilon\}\right) = 0$$

and that $\{f_j\}$ is *Cauchy in measure* if

$$(4.17) \qquad \forall \epsilon > 0 \quad \lim_{j,k\to\infty} \mu\left(\{x : |f_j(x) - f_k(x)| > \epsilon\}\right) = 0.$$

Lemma 4.3. *Let $\{f_j\}$ be Cauchy in measure. Then there exists a function f and a subsequence $\{f_{j_l}\}$ of $\{f_j\}$ such that $f_{j_l} \to f$ a.e.*

Proof. Choose N_m so that $\mu\left(\left\{x : |f_j(x) - f_k(x)| > \frac{1}{2^m}\right\}\right) < \frac{1}{2^m}$ for $j,k > N_m$ and then a sequence of integers j_m such that $j_m < j_{m+1}$ and $j_m > N_m$. Then

$$\mu\left(\left\{x : |f_{j_{l+1}}(x) - f_{j_l}(x)| > \frac{1}{2^m}\right\}\right) < \frac{1}{2^m} \quad \text{for} \quad l \ge m.$$

Thus

$$|f_{j_{l+r}}(x) - f_{j_l}(x)| \le \sum_{i=1}^{r} |f_{j_{l+i}}(x) - f_{j_{l+i-1}}(x)| \le \sum_{i=1}^{r} \frac{1}{2^{l+i-1}} \le \frac{1}{2^{l-1}}$$

except on

$$\bigcup_{i=1}^{r} \left\{x : |f_{j_{l+i}}(x) - f_{j_{l+i-1}}(x)| > \frac{1}{2^{l+i-1}}\right\} \subset \bigcup_{i=1}^{\infty} \left\{x : |f_{j_{l+i}}(x) - f_{j_{l+i-1}}(x)| > \frac{1}{2^{l+i-1}}\right\}$$

which has measure at most

$$\mu\left(\bigcup_{i=1}^{\infty}\left\{x : |f_{j_{l+i}}(x) - f_{j_{l+i-1}}(x)| > \frac{1}{2^{l+i-1}}\right\}\right) \leq \sum_{i=1}^{\infty}\mu\left(\left\{x : |f_{j_{l+i}}(x) - f_{j_{l+i-1}}(x)| > \frac{1}{2^{l+i-1}}\right\}\right)$$
$$\leq \sum_{i=1}^{\infty}\frac{1}{2^{l+i-1}} = \frac{1}{2^{l-1}}$$

Thus

$$\limsup_{l,r\to\infty}|f_{j_{l+r}}(x) - f_{j_l}(x)| = 0 \tag{4.18}$$

unless

$$x \in E_l := \bigcup_{i=1}^{\infty}\left\{x : |f_{j_{l+i}}(x) - f_{j_{l+i-1}}(x)| > \frac{1}{2^{l+i-1}}\right\} \text{ io},$$

where "io" means "infinitely often," or for infinitely many l's. This is the same as saying

$$x \in \cup_{l=M}^{\infty}E_l \quad \text{for all} \quad M \quad \text{or} \quad x \in \cap_{M=1}^{\infty}\cup_{l=M}^{\infty}E_l.$$

Now $\cup_{l=M}^{\infty}E_l$ decreases as M increases and

$$\mu\left(\cup_{l=M}^{\infty}E_l\right) \leq \sum_{l=M}^{\infty}\mu(E_l) \leq \sum_{l=M}^{\infty}\frac{1}{2^l} = \frac{1}{2^{M-1}}$$

so

$$\mu\left(\cap_{M=1}^{\infty}\cup_{l=M}^{\infty}E_l\right) = \lim_{M\to\infty}\mu\left(\cup_{l=M}^{\infty}E_l\right) = 0.$$

Hence (4.18) holds except on a null set, and $\{f_{j_l}\}$ is Cauchy except on a set of measure 0. Let $f(x) = \lim_{l\to\infty}f_{j_l}(x)$. □ □

How is this related to $L^p(\mu)$ convergence? Well, for any $\epsilon > 0$,

$$\epsilon^p\mu(\{|f-g| > \epsilon\}) \leq \int_X |f-g|^p\,d\mu$$

or

$$\mu(\{|f-g| > \epsilon\}) \leq \frac{\|f-g\|_{L^p(\mu)}^p}{\epsilon^p}.$$

Hence a Cauchy sequence in $L^p(\mu)$ is also Cauchy in measure and has a subsequence which converges a.e. Suppose

$$\lim_{j,k\to\infty}\int_X |f_j - f_k|^p\,d\mu = 0$$

while $f_{j_l}(x) \to f(x)$ a.e. as $l \to \infty$. Suppose

$$\int_X |f_j - f_k|^p\,d\mu < \epsilon \quad \text{if} \quad j,k > N_\epsilon.$$

Then

$$\int_X |f - f_k|^p\,d\mu = \int_X \liminf_{l\to\infty}|f_{j_l} - f_k|^p\,d\mu \leq \liminf_{l\to\infty}\int_X |f_{j_l} - f_k|^p\,d\mu \leq \epsilon \quad \text{if} \quad k \geq N_\epsilon$$

and therefore $f_k \to f$ in $L^p(\mu)$.

We have shown that $L^p(\mu)$ is a normed vector space which is complete in the metric induced by the norm. Such a thing is called a *Banach space.*

Let us note the following special cases: a function

$$x : \{1, 2, \dots, n\} \to \mathcal{F} \tag{4.19}$$

or, with $\mathbb{N} = \{1, 2, \dots\}$ (the set of positive integers),

$$x : \mathbb{N} \to \mathcal{F}, \tag{4.20}$$

or, $\mathbb{Z} = \{\dots - 2, -1, 0, 1, 2, \dots\}$ (the set of integers),

$$x : \mathbb{Z} \to \mathcal{F} \tag{4.21}$$

were $\mathcal{F}$ is $\mathbb{R}$ or $\mathbb{C}$, depending on our mood, gives a scalar $x(i)$ for each i in the domain of x. In all cases we will write x_i rather than $x(i)$. In this way, the set of functions like (4.19) is identified with the set of sequences $(x_1, x_2, \dots, x_n)$, that is, $\mathbb{R}^n$. The set of functions like (4.20) is identified with the set of sequences $\{x_j\}_{i=1}^{\infty}$, which we think of as the "infinite dimensional vector" $(x_1, x_2, \dots,)$. The set of functions like (4.21) is identified with the set of doubly infinite sequences $\{x_j\}_{i=-\infty}^{\infty}$, which we think of as the "doubly infinite vector" $(\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots,)$.

Let μ^c be counting measure on $\{1, 2, \dots, n\}$ or $\mathbb{N}$ or $\mathbb{Z}$. The measurable sets are all subsets of the space. Any point in $\mathbb{R}^n$ (respectively, any infinite sequence or doubly infinite sequence) is measurable and has a $L^p(\mu^c)$ norm.

Both all cases, we will denote $L^p(\mu^c)$ norm by $|x|_p$. In the case of $\mathbb{R}^n$, this is

$$|x|_p = \begin{cases} (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p} & \text{if} \quad 1 \le p < \infty \\ \max\{|x_1|, |x_2|, \dots, |x_n|\} & \text{if} \quad p = \infty. \end{cases} \tag{4.22}$$

In the case of $\mathbb{N}$, this is

$$|x|_p = \begin{cases} \left(\displaystyle\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} & \text{if} \quad 1 \le p < \infty \\ \sup\{|x_i| : i \in \mathbb{N}\} & \text{if} \quad p = \infty. \end{cases} \tag{4.23}$$

In the case of $\mathbb{Z}$, this is

$$|x|_p = \begin{cases} \left(\displaystyle\sum_{i=-\infty}^{\infty} |x_i|^p \right)^{1/p} & \text{if} \quad 1 \le p < \infty \\ \sup\{|x_i| : i \in \mathbb{Z}\} & \text{if} \quad p = \infty. \end{cases} \tag{4.24}$$

Definition 4.4. The Banach space $(\mathbb{R}^n, |\cdot|_p)$ is denoted l_n^p and the Banach space consisting of all sequences for which (4.23) is finite is denoted by l^p and the Banach space consisting of all doubly infinite sequences for which (4.24) is finite is denoted by l_d^p.

Before we turn to more abstract things, we prove the following results (see SS Theorem 2.4 pg 71). The point is that to check some statement concerning a metric space, it can be enough to demonstrate it's validity on a dense subspace, and then it is nice to have subspaces with good properties. We will see how this goes later. Right now, we verify that certain subspaces of $L^p(\mathbb{R}^n)$ are dense.

Theorem 4.5. *The following families of functions are dense in $L^p(\mathbb{R}^n)$, $1 \le p < \infty$.*

(i) *Simple functions of the form $\sum_{j=1}^N a_j \chi_{E_j}$ where the $a_j \in \mathcal{F}$ and $\mathcal{L}^n(E_j) < \infty$.*
(ii) *Step functions; that is simple functions of the form (i) in which the E_j are (n-dimensional) rectangles.*
(iii) *Continuous functions of compact support.*

Proof. We first take $p = 1$. We assume that $f = u + iv = u^+ - u^- + i(v^+ - v^-)$ is complex valued. It suffices to approximate each of the nonnegative functions u^+, u^-, v^+, v^- by functions of the desired form, so we may assume that $f \ge 0$. For (i), let $k > 0$ be an integer and define

$$\phi_k(r) = \begin{cases} \dfrac{j-1}{2^k} & \text{if} \quad \dfrac{j-1}{2^k} \le r < \dfrac{j}{2^k} \quad \text{and} \quad 1 \le j < 2^{2k} \\ 2^k & \text{if} \quad 2^k \le r. \end{cases}$$

Boardwork shows that $\phi_k(r) \uparrow r$ as $k \uparrow \infty$. The function $x \mapsto \phi_k(f(x))$ is of the form (i) and $0 \le \phi_k(f(x)) \le f(x)$. By the monotone convergence theorem (or the dominated convergence theorem)

$$\|f - \phi_k(f)\|_{L^1(\mathbb{R}^n)} = \int_{\mathbb{R}^n} (f - \phi_k(f))\, dx \to 0$$

as $k \to \infty$.

To prove (ii), it suffices to approximate each summand in (i) by a step function as in (ii). Let us approximate $a_1\chi_{E_1}$, assuming, as we may, that $a_1 \neq 0$. Since $\mathcal{L}^n(E_1) < \infty$ (and E_1 is measurable), for $\kappa > 0$ there is a collection of almost (which means "up to boundaries") disjoint rectangles $R_1, R_2, \ldots, R_m$ such that

$$\mathcal{L}^n\left(E_1 \Delta \cup_{j=1}^m R_j\right) < \kappa.$$

(See proof of SS, Theorem 4.2, Chapter 2, as necessary.) Then

$$\int_{\mathbb{R}^n} |a_1\chi_{E_1} - a_1 \sum_{j=1}^n \chi_{R_j}|\, dx \le |a_1|\mathcal{L}^n\left(E_1 \Delta \cup_{j=1}^m R_j\right) \le |a_1|\kappa$$

which can be made as small as desired.

Finally, to prove (iii), we now use (ii) to claim that it is only necessary to approximate χ_R in $L^1(\mathbb{R}^n)$ by a continuous function of compact support when R is a rectangle. Since $R = \prod_{j=1}^n [a_j, b_j]$ is the product of intervals, we let

$$g_j \quad \text{be the piecewise linear function which is 1 on} \quad [a_j, b_j]$$

and runs linearly down to 0 at $a_j - \kappa, b_j + \kappa$ and is 0 thereafter. Let

$$R_s = [a_1 - s, b_1 + s] \times [a_2 - s, b_2 + s] \times \cdots [a_n - s, b_n + s]$$

which has Lebesgue measure

$$\mathcal{L}^n(R_s) = \prod_{j=1}^{n} (b_j - a_j + 2s).$$

The integrand in

$$\int_{\mathbb{R}^n} |\chi_E(x) - \prod_{j=1}^{n} g_j(x_j)|\, dx \leq 2^n \kappa^n,$$

is vanishes ae off of $R_\kappa \setminus R_0$, and the integrand is at most one on this set. Thus the integral is at most

$$\mathcal{L}^n(R_\kappa) - \mathcal{L}^n(R_0),$$

which is a continuous function of κ (a polynomial) which vanishes at $\kappa = 0$. Thus it can be made as small as desired by choosing κ small.

We turn to $L^p(\mathbb{R}^n)$. For $f \in L^p(\mathbb{R}^n)$ and $R > 0$, define

$$f_R(x) = \begin{cases} f(x) & \text{if} \quad |x| \leq R \,\&\, |f(x)| \leq R, \\ 0 \quad \text{if} \quad |f(x)| > R \quad \text{or} \quad |x| > R. \end{cases} \tag{4.25}$$

Clearly $|f_R| \leq R$. Moreover, since $|f - f_R| \leq |f|$,

$$|f - f_R|^p \leq |f|^p.$$

Since also $f - f_R \to 0$ a.e. as $R \to \infty$, the dominated convergence theorem implies that

$$\|f - f_R\|_{L^p(\mathbb{R}^n)} \to 0 \quad \text{as} \quad R \to \infty.$$

Thus it suffices to approximate bounded functions f which vanish off some ball $B_R = \{x \in \mathbb{R}^n : |x| \leq R.\}$ This follows from the L^1 results, since, if $|f|, |g| \leq M$ and both vanish off the ball of radius $2R$,

$$\int_{\mathbb{R}^n} |f - g|^p\, dx \leq \int_{\mathbb{R}^n} |f - g||f - g|^{p-1}\, dx \leq (2M)^{p-1} \int_{B_R} |f - g|\, dx.$$

□

Remark 4.6. Approximation in $L^p(E)$ where $E \subset \mathbb{R}^n$ follows upon extending $f \in L^p(E)$ to all of $\mathbb{R}^n$ by letting the extension (call it $\tilde{f}$) vanish on $\mathbb{R}^n \setminus E$. Then

$$\int_E |f - g|^p\, dx \leq \int_{\mathbb{R}^n} |\tilde{f} - g|^p\, dx.$$

5. Normed Spaces, Banach Spaces, Linear Operators

Let us review some of the definitions embedded in the above. Below, the field $\mathcal{F}$ can be either $\mathbb{R}$ or $\mathbb{C}$.

Definition 5.1. A norm on a vector space X over $\mathcal{F}$ is a mapping $\| \, \| : V \to [0, \infty)$ with the properties

$$\begin{aligned} &\|x+y\| \leq \|x\| + \|y\| \quad (x, y \in X) \\ &\|cx\| = |c|\|x\| \qquad\quad (c \in \mathcal{F}, x \in X) \\ &\|x\| = 0 \implies x = 0. \end{aligned}$$

If $\| \, \|$ is a norm on X, we call $(X, \| \, \|)$ a normed vector space. If the normed vector space $(X, \| \, \|)$ is complete with the metric $d(x, y) = \|x - y\|$, then $(X, \| \, \|)$ is a Banach space. If $\mathcal{F} = \mathbb{R}$, we speak of a "real" Banach space and if $\mathcal{F} = \mathbb{C}$, we speak of a "complex" Banach space.

We have shown that $L^p(\mu)$ is a Banach space, real or complex, depending on whether we take its elements to be real or complex valued. Of course, there is the "coset" business here.

Remark 5.2. A complex normed vector space $(X, \| \, \|)$ is a real normed vector space! This seemingly silly remark is to dispel the notion that "complex" is more general than "real" in this context. Many - but not all - results about real vector spaces apply at once to complex vector spaces. Examples are Theorems 5.6 and 5.7 below.

Recall that if (X, d) and (Y, ρ) are two metric spaces and $f : X \to Y$, then f is *Lipschitz continuous* if there is a constant L such that

$$\rho(f(x), f(y)) \leq Ld(x, y) \quad (x, y \in X). \tag{5.1}$$

When this holds, we say that L is a Lipschitz constant for f or f is Lipschitz with constant L. The least such constant,

$$L_0 = \inf \{L : \rho(f(x), f(y)) \leq Ld(x, y) \text{ for } x, y \in X\}$$

is called the least Lipschitz constant for f and it is denoted by $\mathrm{Lip}\,(f)$.

In any normed space $(X, \| \, \|)$, one has

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|$$

which implies

$$|\|x\| - \|y\|| \leq \|x - y\|. \tag{5.2}$$

In particular, $\| \, \|$ is Lipschitz with constant 1.

Proposition 5.3. *Let $(X, \| \|_X)$ and $(Y, \| \|_Y)$ be two normed vector spaces (and therefore metric spaces) over $I\!R$ and $\Lambda : X \to Y$ be a linear mapping. Then Λ is continuous iff and only if it is Lipschitz continuous and then*

$$\begin{aligned} \operatorname{Lip}(\Lambda) &= \sup\{\|\Lambda x\|_Y : x \in X \text{ and } \|x\|_X \le 1\} \\ &= \sup\{\|\Lambda x\|_Y : x \in X \text{ and } \|x\|_X = 1\} \\ &= \frac{1}{r}\sup\{\|\Lambda x\|_Y : x \in X \text{ and } \|x\|_X \le r\} \quad (r > 0) \\ &= \frac{1}{r}\sup\{\|\Lambda x\|_Y : x \in X \text{ and } \|x\|_X = r\} \quad (r > 0). \end{aligned} \tag{5.3}$$

Proof. If $\operatorname{Lip}(\Lambda) < \infty$, then

$$\|\Lambda x\|_Y = \|\Lambda x - \Lambda 0\|_Y \le \operatorname{Lip}(\Lambda)\|x - 0\|_X = \operatorname{Lip}(\Lambda)\|x\|.$$

It follows that $\operatorname{Lip}(\Lambda)$ is at least as big as the displayed quantities on the right of (5.3). OTOH, if $L = \frac{1}{r}\sup\{\|\Lambda w\|_Y : w \in X \text{ and } \|w\|_X = r\}$, then for $x, \hat{x} \in X$, $x \ne \hat{x}$,

$$\|\Lambda x - \Lambda\hat{x}\|_Y = \|\Lambda(x - \hat{x})\|_Y = \|x - \hat{x}\|_X \frac{1}{r}\|\Lambda\left(\frac{r(x-\hat{x})}{\|x-\hat{x}\|_X}\right)\|_Y \le L\|x - \hat{x}\|_X;$$

That is, $\operatorname{Lip}(\Lambda) \le L$. All the equalities now follow. Finally, if Λ is continuous, then it is continuous at 0, and for r sufficently small, the third quantity displayed on the right-hand side of (5.3) is finite, by continuity. □

Let $B_r = \{x \in X : \|x\|_X \le r\}$. By the above, the image of B_r under Λ is bounded iff Λ is Lipschitz continuous. Thus, in this linear case, rather than say "Λ is Lipschitz continuous" one says "Λ is bounded" or "Λ is continuous." Let

$$\mathcal{L}(X, Y) = \{\text{bounded linear operators } \Lambda : X \to Y\}. \tag{5.4}$$

Define $\| \| : \mathcal{L}(X, Y) \to [0, \infty)$ by

$$\|\Lambda\| = \operatorname{Lip}(\Lambda) = \sup\{\|\Lambda x\|_Y : x \in X \text{ and } \|x\|_X \le 1\}. \tag{5.5}$$

Equivalently, $\|\Lambda\|$ is the least number for which

$$\|\Lambda x\|_Y \le \|\Lambda\|\|x\|_X \quad (x \in X). \tag{5.6}$$

It is straightforward to check that $(\mathcal{L}(X, Y), \| \|)$ is a normed vector space, which is complete (and hence a Banach space) if Y is a Banach space.

Theorem 5.4. *Let $(X, \| \|_X)$ and $(Y, \| \|_Y)$ be normed vector spaces. Then $\mathcal{L}(X, Y)$ is a normed vector space when equipped with the norm (5.5), (5.6). If $(Y, \| \|_Y)$ is a Banach space, then so is $\mathcal{L}(X, Y)$.*

Proof. The set of linear maps from X to Y is a linear space. If $\Lambda, \hat{\Lambda} \in \mathcal{L}(X, Y)$ and $c \in \mathcal{F}, c \ne 0$, we clearly have

$$\begin{aligned} \sup\{\|c\Lambda x\|_Y : \|x\|_X \le 1\} &= \sup\{\|\Lambda(cx)\|_Y : \|x\| \le 1\} \\ &= \sup\{\|\Lambda(cx)\|_Y : \|cx\|_X \le |c|\} = |c|\|\Lambda\| \end{aligned}$$

and

$$\|(\Lambda+\hat{\Lambda})x\|_Y = \|\Lambda x + \hat{\Lambda}x\|_Y \le \|\Lambda x\|_Y + \|\hat{\Lambda}x\|_Y \le (\|\Lambda\|_Y + \|\hat{\Lambda}\|_Y)\|x\|_X.$$

The first of these relations establishes $\|c\Lambda\| = |c|\|\Lambda\|$ and the second shows that

$$\|\Lambda+\hat{\Lambda}\| \le \|\Lambda\| + \|\hat{\Lambda}\|_Y;$$

in particular, $c\Lambda, \Lambda+\hat{\Lambda} \in \mathcal{L}(X,Y)$ and $\| \;\|$ is indeed a norm.

Suppose now that Y is complete and $\{\Lambda_j\}$ is a Cauchy sequence in $\mathcal{L}(X,Y)$. Then for $x \in X$

$$\|\Lambda_j x - \Lambda_k x\|_Y \le \|\Lambda_j - \Lambda_k\|\|x\|_X$$

shows that $\Lambda_j x$ is Cauchy in Y. Define

$$\Lambda x := \lim_{j\to\infty} \Lambda_j x.$$

Clearly $\Lambda : X \to Y$ is linear. Now, if

$$\|\Lambda_j - \Lambda_k\| < \epsilon \quad \text{for} \quad j,k \ge N_\epsilon,$$

then, by the continuity of norms (see (5.2)), we have

$$\|\Lambda_j x - \Lambda x\|_Y = \lim_{k\to\infty} \|\Lambda_j x - \Lambda_k x\|_Y \le \limsup_{k\to\infty} \|\Lambda_j - \Lambda_k\|\|x\|_X \le \epsilon\|x\|_X$$

for $j \ge N_\epsilon$. It follows that $\|\Lambda_j - \Lambda\| \to 0$, and $\mathcal{L}(X,Y)$ is complete. □

Remark 5.5. An important special case arise by taking $(Y, \| \;\|_Y)$ to be $(\mathcal{F}, |\;|)$, in which case we are discussing the *bounded linear functionals* mapping X into its scalar field; somewhat in contrast to the case of general linear operators, these are more commonly called the *continuous linear functionals.* $\mathcal{L}(X,\mathcal{F})$ is called the *dual space* of X. It is complete, by the above. One often writes X^* for the dual space of X (with $\| \;\|_X$ now being "understood.")

Here follow a few basic theorems we will prove, in the fullness of time.

Theorem 5.6 (Banach Steinhauss Theorem, aka the Uniform Boundedness Principle)**.** *Let $(X, \| \;\|_X)$ be a Banach space and $(Y, \| \;\|_Y)$ be a normed vector space. Let $\mathcal{A}$ be an index set and $T_\alpha \in \mathcal{L}(X,Y)$ for $\alpha \in \mathcal{A}$. Then either*

$$\sup\{\|T_\alpha x\|_Y : \alpha \in \mathcal{A}\} = \infty$$

for all x in a dense $G_\delta \subset X$ or there exists $M < \infty$ such that

$$\|T_\alpha\| \le M \quad (\alpha \in \mathcal{A}). \tag{5.7}$$

Proof. We follow Rudin, pg 98 (with less language). Put

$$V_n = \bigcup_{\alpha\in\mathcal{A}} \{x : \|T_\alpha x\|_Y > n\} = \left\{x : \sup_{\alpha\in\mathcal{A}} \|T_\alpha x\|_Y > n\right\} \quad (n = 1,2,\dots) \tag{5.8}$$

By the continuity of $x \mapsto \|T_\alpha x\|_Y$, V_n is open in X. If one of the V_n, say V_N, fails to be dense, then there is an $x_0 \in X$ and $r > 0$ such that $x_0 + x \notin V_N$ for $\|x\|_X \le r$. But then for $\|x\|_X \le r$,

$$\|T_\alpha x\|_Y \le \|T_\alpha(x_0 + x)\|_Y + \|T_\alpha x_0\|_Y \le 2N \quad (\alpha \in \mathcal{A}),$$

and $\|T_\alpha\| \le 2N/r$ follows.

On the other hand, if V_n is dense for every n, and we now know that this must be the case if $\sup_{\alpha\in\mathcal{A}} \|T_\alpha\| = \infty$, then by Baire's Theorem $\cap_{n=1}^\infty V_n$ is a dense G_δ in X. Since $\sup_{\alpha\in\mathcal{A}} \|T_\alpha x\|_Y = \infty$ for $x \in \cap_{n=1}^\infty V_n$, we are done. $\square$

Theorem 5.7 (Open Mapping Theorem)**.** *Let $(X, \| \|_X)$ and $(Y, \| \|_Y)$ be Banach spaces. Let $T \in \mathcal{L}(X, Y)$ and $TX = Y$ (i.e., T is onto Y.) Let $U := \{x \in X : \|x\|_X < 1\}$, $V := \{y \in Y : \|y\|_Y < 1\}$ be the open unit balls of X and Y. Then there exists a $\delta > 0$ such that*

$$TU \supset \delta V.$$

In consequence, for every open subset $\mathcal{O}$ of X, $T\mathcal{O}$ is open in Y. Finally, if $TX = Y$ and T is 1-1, then T^{-1} is bounded.

Proof. X is the union of the sets kU, $k = 1, 2, \dots,$

$$Y = \bigcup_{k=1}^\infty T(kU).$$

Since Y is complete, by Baire's Theorem, there is a nonempty open set W which is contained in the closure of some $T(kU)$.

To simplify writing, let us notice some things. First, we may take $k = 1$. This is because $T(kU) = kTU$ and so $\overline{T(kU)} = k\overline{TU}$, where the overline indicates "closure" (see Exercise 15 for this and other remarks we use here). Thus the interior of $\overline{TU}$, $\overline{TU}^\circ$, is $\frac{1}{k}\overline{T(kU)}^\circ$, which is not empty.

U is convex, that is, if $x, \hat{x} \in U$, then the line segment $[x, \hat{x}]$ joining x and $\hat{x}$, namely

$$[x, \hat{x}] := \{(1-t)x + t\hat{x} : 0 \le t \le 1\}, \tag{5.9}$$

also lies in U. Since the line segment joining Tx and $T\hat{x}$ is, by linearity, $[Tx, T\hat{x}] = T[x, \hat{x}]$, TU is also convex. Moreover, TU is symmetric about the origin, that is, $-TU = TU$, since $-U = U$. Hence if $\overline{TU}$ contains an open set W, it also contains

$$\frac{1}{2}W - \frac{1}{2}W,$$

which is open (Exercise 15) and contains 0. Thus $\overline{TU}$ contains a ball about the origin of Y. Let us say

$$\overline{TU} \supset \delta V \quad \text{where } \delta > 0. \tag{5.10}$$

The for any $r > 0$

$$r\overline{TU} = \overline{T(rU)} \supset r\delta V. \tag{5.11}$$

That is, if $\|y\|_Y < r\delta$ and $\epsilon > 0$, there is $x \in X$ satisfying

$$\|y - Tx\|_Y < \epsilon \quad \text{and} \quad \|x\|_X < r. \tag{5.12}$$

Now we iterate: let $\{\epsilon_j\}$ be any sequence of positive numbers. Start with $y \in \delta V$ and choose x_1 such that

$$\|y - Tx_1\|_Y < \epsilon_1 \quad \text{and} \quad \|x_1\|_X < 1. \tag{5.13}$$

Then choose x_2 according to

$$\|y - (Tx_2 + Tx_1)\|_Y = \|(y - Tx_1) - Tx_2\|_Y < \epsilon_2 \quad \text{and} \quad \|x_2\|_X < \frac{\epsilon_1}{\delta}; \tag{5.14}$$

and, in general,

$$\|y - (Tx_1 + Tx_2 + \cdots + Tx_n)\|_Y < \epsilon_n \quad \text{and} \quad \|x_n\|_X < \frac{\epsilon_n}{\delta} \tag{5.15}$$

for $n = 1, 2, \ldots$.

It follows that (more remarks in class) that if $\sum_{j=1}^{\infty} \epsilon_j < \infty$, then

$$x = \sum_{j=1}^{\infty} x_j = \lim_{n\to\infty} \sum_{j=1}^{n} x_j \quad \text{exists}$$

and, by continuity of T and (5.15)

$$Tx = \lim_{n\to\infty} T \sum_{j=1}^{n} x_j = \lim_{n\to\infty} \sum_{j=1}^{n} Tx_j$$

$$\|x\|_X \le \|x_1\|_X + \sum_{j=1}^{\infty} \|x_j\|_X < \|x_1\|_X + \frac{1}{\delta} \sum_{j=1}^{\infty} \epsilon_j. \tag{5.16}$$

If $\sum_{j=1}^{\infty} \epsilon_j < \delta(1 - \|x_1\|_X)$, and we can always arrange that, then $\|x\|_X < 1$ and $Tx = y$. Thus $TU \supset \delta V$.

To see that $T\mathcal{O}$ is open if $\mathcal{O}$ is open in X, note that if $x_0 \in \mathcal{O}$, then $x_0 + \mu U \subset \mathcal{O}$ for some $\mu > 0$ and

$$T\mathcal{O} \supset Tx_0 + \mu TU \supset Tx_0 + \delta\mu V;$$

that is, $T\mathcal{O}$ contains a ball about any of its points, so it is open.

If T is also 1-1, then for any open set $\mathcal{O}$ in X, the inverse image of $\mathcal{O}$ under T^{-1} is $(T^{-1})^{-1}\mathcal{O} = T\mathcal{O}$, which we just verified is open. Hence T^{-1} is continuous. □

Theorem 5.8 (Hahn Banach Theorem). *Let $(X, \|\ \|)$ be a normed vector space (real or complex). Let M be a subspace of X and $f \in M^*$. Then there exists $F \in X^*$ satisfying $F|_W = f$ and $\|F\| = \|f\|$.*

We'll probably modify this one later, using some notes of Stephen Simons. NOTE: This has been done, see Section 8 below.

6. Fourier Series

We will apply Theorems 5.6 and 5.7 to problems arising in the theory of Fourier series. Partially to warm you up for next quarter, consider the system of ordinary differential equations

$$\frac{d}{dt}X + AX = 0 \tag{6.1}$$

coupled with the initial conditions

$$X(0) = X_0 \tag{6.2}$$

where now

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

is a vector in $\mathbb{R}^n$ (or we could take $\mathbb{C}^n$) and $A = [a_{ij}]$ is an $n \times n$ real (or we could take complex) matrix and $X_0 \in \mathbb{R}^n$ (respectively, $\mathbb{C}^n$). $X(t)$ is the "state" of the system at time t which is X_0 at $t = 0$, and which evolves according to the "dynamics" (6.1).

If (λ, v) is an "eigenpair" of A, that is, $v \neq 0$ and

$$Av = \lambda v, \tag{6.3}$$

then

$$\begin{aligned}\frac{d}{dt}(e^{-\lambda t}v) + A(e^{-\lambda t}v) &= e^{-\lambda t}\,(-\lambda v + Av) \\ &= e^{-\lambda t}(-\lambda v + \lambda v) = 0.\end{aligned}$$

That is, $e^{-\lambda t}v$ is a solution of the system (6.1). The ode's are linear and homogeneous, so if (λ_j, v^j) is an eigenpair for $j = 1, 2, \ldots, k$, then

$$X(t) = \sum_{j=1}^{k} a_j e^{-\lambda_j t} v^j \tag{6.4}$$

is also a solution of (6.1) for any scalars a_j. In order that this X satisfy the initial conditions (6.2) we need

$$X_0 = \sum_{j=1}^{k} a_j v^j. \tag{6.5}$$

If $j = n$ and the v^j are linearly independent, then (6.5) does hold for exactly one choice of the a_j's.

The best situation arises if A is *self-adjoint.* That is,

$$\langle x, Ay\rangle = \langle x, Ay\rangle \quad \text{for all } x, y \in \mathbb{C}^n, \tag{6.6}$$

where

$$\langle x, y\rangle = \sum_{j=1}^{n} x_j \overline{y_j} \quad \text{for } x, y \in \mathbb{C}^n \tag{6.7}$$

is the "inner-product." Note that

$$\begin{aligned} \langle x, y\rangle &= \overline{\langle y, x\rangle}, \\ \langle x, x\rangle &= \sum_{j=1}^{n} |x_j|^2 =: |x|^2, \\ \langle \lambda x, y\rangle &= \lambda\langle x, y\rangle. \end{aligned} \tag{6.8}$$

The condition (6.6) amounts to $a_{i,j} = \overline{a_{j,i}}$ for $i, j = 1, \ldots, n$.

If (λ, v) is an eigenpair for A, taking $x = y = v$ in (6.6) yields

$$\lambda|v|^2 = \langle \lambda v, v\rangle = \langle Av, v\rangle = \langle v, Av\rangle = \langle v, \lambda v\rangle = \bar{\lambda}|v|^2. \tag{6.9}$$

That is, the eigenvalues of A are real. Moreover, if (λ, v), $(\hat{\lambda}, \hat{v})$ are two eigenpairs for A and $\lambda \neq \hat{\lambda}$, we have

$$\lambda\langle v, \hat{v}\rangle = \langle Av, \hat{v}\rangle = \langle v, A\hat{v}\rangle = \langle v, \hat{\lambda}\hat{v}\rangle = \hat{\lambda}\langle v, \hat{v}\rangle, \tag{6.10}$$

which implies that

$$\langle v, \hat{v}\rangle = 0. \tag{6.11}$$

Thus eigenvectors of A for different eigenvalues are *orthogonal*.

Ok, this is all part of the proof that if A is self-adjoint, then there are n-eigenpairs (λ_j, v^j), $j = 1, 2, \ldots, n$, with the v^j mutually orthogonal, that is,

$$\langle v^j, v^k\rangle = 0 \quad \text{for} \quad j \neq k. \tag{6.12}$$

Once this "spectral" information has been found, the solution of (6.1), (6.2) is determined by writing (see (6.5))

$$X_0 = \sum_{j=1}^{n} a_j v^j \tag{6.13}$$

where the coefficients a_j are determined (using (6.12)) by

$$\langle X_0, v^k\rangle = \langle \sum_{j=1}^{n} a_j v^j, v^k\rangle = a_k|v^k|^2; \text{ equivalently, } a_k = \frac{\langle X_0, v^k\rangle}{|v^k|^2} \tag{6.14}$$

and the solution of (6.1), (6.2) is

$$X(t) = \sum_{j=1}^{n} e^{-\lambda_j t}\frac{\langle X_0, v^j\rangle}{|v^j|^2} v^j. \tag{6.15}$$

Suppose now that we seek the temperature in a circular piece of wire. We can identify the position of points on the wire in the form e^{ix}, $-\pi \le x \le \pi$. Let the temperature at time t and position e^{ix} be $u(t,x)$. Suppose that u satisfies the *heat equation*,

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0. \tag{6.16}$$

We couple this with the initial condition

$$u(0,x) = f(x) \quad \text{for} \quad -\pi < x < \pi. \tag{6.17}$$

Moreover, since $x = \pi$ and $x = -\pi$ correspond to the same points of the wire, u has to be "doing" the same thing at $x = \pi$ as it is doing at $x = -\pi$, which corresponds to the *periodic boundary conditions*

$$u(t,-\pi) = u(t,\pi), \quad \frac{\partial u}{\partial x}(t,-\pi) = \frac{\partial u}{\partial x}(t,\pi) \quad \text{for} \quad t > 0. \tag{6.18}$$

Now we change our point of view a bit. The "state" of our wire at time t is the function $x \mapsto u(t,x)$, which gives the temperature at each time t. Let this function be called $U(t)$; that is, $U(t)(x) = u(t,x)$. For each t, $U(t)$ is a function. Then we think of $\frac{\partial u}{\partial t}$ as the time derivative of $U(t)$, and we think of $-\frac{\partial^2 u}{\partial x^2}$ (which is doing something to the state at time t) as $AU(t)$, where A is the "operation" of taking the second derivative. That is, $AU(t) = -U(t)''$, where the primes refer to "x derivatives." Then, formally (meaning "without rigor"), our equation (6.16) writes up as

$$\frac{d}{dt} U + AU = 0, \tag{6.19}$$

which looks just like (6.1). The initial condition is

$$U(0) = f, \tag{6.20}$$

and the boundary conditions we will "put into A", that is, A has some domain consisting of functions of x which have two derivatives and satisfy the periodic boundary conditions. That is, $v \in D(A)$ (the domain of A) requires

$$v(-\pi) = v(\pi) \quad \text{and} \quad v'(-\pi) = v'(\pi). \tag{6.21}$$

Now just as in (6.1), if (λ, v) is an eigenpair for A, then $u(t,x) = e^{-\lambda t} v(x)$, equivalently, $U(t)(x) = e^{-\lambda t} v(x)$, satisfies

$$\begin{aligned} \frac{\partial}{\partial t}\left(e^{-\lambda t} v(x)\right) - \frac{\partial^2}{\partial x^2}\left(e^{-\lambda t} v(x)\right) &= \frac{d}{dt} e^{-\lambda t} v + A e^{-\lambda t} v \\ &= -\lambda e^{-\lambda t} v + e^{-\lambda t} A v \\ &= e^{-\lambda t}\left(-\lambda v + \lambda v\right) = 0. \end{aligned} \tag{6.22}$$

We will see that A is (formally) self-adjoint. Our inner-product is now

$$\langle v, w \rangle = \int_{-\pi}^{\pi} v(x) \bar{w}(x)\, dx. \tag{6.23}$$

If v, w are in the domain of A, we have, integrating by parts,

$$\begin{aligned}
\langle Av, w\rangle &= -\int_{-\pi}^{\pi} v''(x)\bar{w}(x)\,dx \\
&= -v'(\pi)\bar{w}(\pi) + v'(-\pi)\bar{w}(-\pi) + \int_{-\pi}^{\pi} v'(x)\bar{w}'(x)\,dx \\
&= \int_{-\pi}^{\pi} v'(x)\bar{w}'(x)\,dx \\
&= v(\pi)\bar{w}'(\pi) - v(-\pi)\bar{w}'(-\pi) - \int_{-\pi}^{\pi} v(x)\bar{w}''(x)\,dx \\
&= \langle v, Aw\rangle.
\end{aligned} \tag{6.24}$$

The boundary terms disappear at each stage, due to the bundary conditions satisfied by v, w. It follows that the eigenvalues of A are real, and that eigenfunctions for different eigenvalues are orthogonal in the inner-product (6.23).

We can see more. Letting $v = w$ in (6.24), and $Av = \lambda v$, we find from the first integration by parts that

$$\lambda\langle v, v\rangle = \langle Av, v\rangle = \int_{-\pi}^{\pi} |v'(x)|^2\,dx = \langle v', v'\rangle.$$

Hence, unless $\langle v', v'\rangle = 0$, $\lambda > 0$. But $\langle v', v'\rangle = 0$ makes v constant, and constants are indeed eigenfunctions, they satisfy the boundary conditions and $A1 = 01 = 0$. Thus the eigenvalues of A are real, and positive, except for the eigenvalue 0, which has 1 as an eigenfunction.

The other eigenfuctions are easy to come by. Write $\lambda = \kappa^2$ for some $\kappa > 0$. Then

$$Av = -v'' = \kappa^2 v \implies v = a\cos(\kappa x) + b\sin(\kappa x) \tag{6.25}$$

for some a, b. Imposing the boundary conditions on this form of function, we find that the boundary conditions imply $\kappa \in \mathbb{Z}$, $\kappa = \pm n$ for some integer n and then the eigenvalue is $\lambda = \kappa^2 = n^2$.

For elegance sake, we choose the eigenfunctions e^{inx} where $n = 0, \pm 1, \pm 2, \ldots$. Recall that

$$\cos(nx) = \frac{e^{inx} + e^{-inx}}{2}, \quad \sin(nx) = \frac{e^{inx} - e^{-inx}}{2i}.$$

These are 2π periodic functions, they satisfy the periodic boundary conditions, and

$$Ae^{inx} = -\frac{d^2}{dx^2}e^{inx} = n^2 e^{inx}.$$

Hence any finite sum

$$u(t, x) = \sum_{n=-N}^{N} a_n e^{-n^2 t} e^{inx} \tag{6.26}$$

is a solution of the heat equation which satisfies the periodic boundary conditions. Don't be troubled by the complex valued looking expression; if $a_n = \overline{a_{-n}}$ then u is real-valued. By the derivation, e^{inx} and e^{imx} are orthogonal if $n^2 \neq m^2$, as they are eigenfunctions for different eigenvalues and A is (formally) self-adjoint. In fact, they are pairwise orthogonal:

$$\langle e^{imx}, e^{inx} \rangle = \int_{-\pi}^{\pi} e^{imx} e^{-inx}\, dx = \frac{1}{i(m-n)} \int_{-\pi}^{\pi} \frac{d}{dx} \left(e^{i(m-n)x} \right) dx = 0 \tag{6.27}$$

if $m \neq n$.

If we seek to satisfy the initial condition (6.17) with the expression (6.26) we must have

$$u(0,x) = f(x) = \sum_{n=-N}^{N} a_n e^{inx}. \tag{6.28}$$

The coefficients a_n can be computed from f just as before:

$$\langle f, e^{ikx} \rangle = \langle \sum_{n=-N}^{N} a_n e^{inx}, e^{imx} \rangle = a_k \langle e^{ikx}, e^{ikx} \rangle = a_k \int_{-\pi}^{\pi} 1\, dx = a_k 2\pi \tag{6.29}$$

or

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx}\, dx. \tag{6.30}$$

Initial conditions of the form (6.28), those finite linear combinations, are called "trigonometric polynomials;" they are very special. Most functions f are not trigonometric polynomials.

The formulas (6.29) do not depend on N, the degree of the trig poly. One then asks if it is possible that fairly general f's can be written as *infinite sums*

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx} \tag{6.31}$$

with the coefficients given by (6.29). This is a bold question, with many aspects. Of course, there are also questions about the $u(t,x)$ given by (6.26) under the same assumptions.

However, at this point we leave the heat equation and focus on (6.31) with (6.30). Moreover, we will now replace x by t (recall, we have dumped the heat equation setting). Fourier series, the objects above, are even more important in signal processing, and then there is no "space variable." The formulas (6.29) make sense if $f \in L^1(-\pi,\pi)$. We now define, for any $f \in L^1(-\pi,\pi)$,

$$f \sim \sum_{n=-\infty}^{\infty} a_n e^{int} \tag{6.32}$$

to mean

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt \quad \text{for} \quad n \in \mathbb{Z}. \tag{6.33}$$

A better notation is this: for $f \in L^1(-\pi,\pi)$ and $n \in \mathbb{Z}$, define

$$\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt \quad \text{for} \quad n \in \mathbb{Z}. \tag{6.34}$$

Notice that

$$|\hat{f}(n)| \le \Big|\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt\Big| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(t)|\,dt = \frac{1}{2\pi}\|f\|_{L^1(-\pi,\pi)}.$$

That is, $f \mapsto \hat{f}$ is a bounded linear operator from $L^1(-\pi,\pi)$ into the Banach space of doubly infinite bounded sequences $\{x_n\}_{n=-\infty}^{\infty}$ with the supremum norm (see (4.24)); in fact, we showed that

$$\|\left\{\hat{f}(n)\right\}\|_\infty \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |f|\,dt. \tag{6.35}$$

Let us define $\mathcal{F}f = \hat{f}$ to give this map a name.

One obvious questions is this: is it true that $\mathcal{F}L^1(-\pi,\pi) = l_d^\infty$? Is every bounded doubly infinite sequence the set of Fourier coefficients for some function in $L^1(-\pi,\pi)$?

The answer is "no."

Lemma 6.1 (Riemann-Lebesgue)**.** *Let $f \in L^1(-\pi,\pi)$. Then*

$$\lim_{n\to\pm\infty} \hat{f}(n) = 0. \tag{6.36}$$

Proof. This result follows from an approximation lemma.

Lemma 6.2. *Trigonometric polynomials are dense in the space $C_p([-\pi,\pi])$ of continuous functions $g : [-\pi,\pi] \to \mathbb{C}$ which satisfy $g(-\pi) = g(\pi)$ (which carries the norm $\|g\| = \max_{[-\pi,\pi]}|g|$). In consequence, the trigonometric polynomials are dense in $L^p(-\pi,\pi)$ for $1 \le p < \infty$.*

We defer the proof of this until the next section. Since a trig poly

$$p(t) = \sum_{n=-N}^{N} a_n e^{inx}$$

satisfies $\hat{p}(m) = 0$ for $m > n$, the claim of Lemma 6.1 holds for trig polys. Since

$$|\hat{f}(m)| \le |\widehat{(f-p)}(m)| + |\hat{p}(m)| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |f-p|\,dt + |\hat{p}(m)|$$

we can choose a poly p such that the first term on the right is at most $\epsilon > 0$ and then m so large that $\hat{p}(m) = 0$. The conclusion is that $|\hat{f}(m)| \le \epsilon$ as soon as m is large enough, and Lemma 6.1 is proved. $\square$

For the rest of this section, we denote by c_0 the space

$$c_0 = \left\{ x \in l_d^\infty : \lim_{n\to\pm\infty} |x_n| = 0. \right\} \tag{6.37}$$

We won't ask for it to be turned in, but you should do the exercise of showing that c_0 is a closed subpace of l_d^∞, hence it is complete, hence it is a Banach space.

Theorem 6.3. *The mapping $\mathcal{F}f = \hat{f}$ is a 1-1 bounded linear transformation of $L^1(-\pi,\pi)$ into, but not onto, c_0.*

Proof. We already showed that $\mathcal{F}$ is a bounded linear transformation as claimed.

To show that $\mathcal{F}$ is 1-1, we need only show that $\hat{f}(n) = 0$ for all n implies that $f = 0$. If $\hat{f}(n) = 0$ for all n, then

$$\int_{-\pi}^{\pi} f(t)p(t)\,dt = 0$$

for every trig poly p, and then, via Lemma 6.2,

$$\int_{-\pi}^{\pi} f(t)g(t)\,dt = 0$$

for $g \in C_p([-\pi,\pi])$. Given a closed set $E \subset (-\pi,\pi)$, put

$$g(t) = g_m(t) := \left(1 - \frac{\text{dist}\,(t,E)}{2\pi}\right)^m$$

Clearly $|g_m| \le 1$ and $g_m(t) \to 1$ if $t \in E$, and $g_m(t) \to 0$ if $t \notin E$. By LDC, we conclude that

$$\int_{-\pi}^{\pi} f(t)\chi_E(t)\,dt = 0$$

for every closed $E \subset (-\pi,\pi)$. In general, if $F \subset [-\pi,\pi]$, then there is an increasing sequence of closed sets $E_j \subset E_{j+1} \subset F \cap (-\pi,\pi)$, such that $\mathcal{L}(F \setminus \cup_{j=1}^\infty E_j) = 0$. Putting $E = E_j$ above and sending $j \to \infty$, LDC yields

$$\int_{-\pi}^{\pi} f(t)\chi_F(t)\,dt = 0.$$

It follows that $f = 0$ a.e.

Since $\mathcal{F}$ is 1-1, by the Open Mapping Theorem, if it is onto, then $\mathcal{F}^{-1}$ is bounded. This implies

$$\|f\|_{L^1(\pi,\pi)} = \|\mathcal{F}^{-1}\mathcal{F}f\|_{L^1(\pi,\pi)} \le \|\mathcal{F}^{-1}\|\|\mathcal{F}f\|_{c_0}. \tag{6.38}$$

Let

$$D_N(t) = \sum_{k=-N}^{N} e^{ikt} \quad (N = 0,1,2,\ldots). \tag{6.39}$$

For a 2π periodic function f, the N^{th} partial sum of its Fourier series is

$$
\begin{aligned}
s_N(f,t) &= \sum_{n=-N}^{N} \hat{f}(n)e^{int} \\
&= \sum_{n=-N}^{N} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)e^{-ins}\,ds\,e^{int} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\sum_{n=-N}^{N} e^{in(t-s)}\,ds \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)D_N(t-s)\,ds.
\end{aligned}
\tag{6.40}
$$

The D_N are called the *Dirichlet kernels*. We have

$$
\begin{aligned}
2i\sin\left(\frac{t}{2}\right)D_N(t) &= e^{it/2}D_N(t) - e^{-it/2}D_N(t) \\
&= \sum_{k=-N}^{N} e^{i(k+\frac{1}{2})t} - \sum_{k=-N}^{N} e^{i(k-\frac{1}{2})t} \\
&= \sum_{k=-N}^{N} e^{i(k+\frac{1}{2})t} - \sum_{j=-N-1}^{N-1} e^{i(j+\frac{1}{2})t} \\
&= e^{i(N+\frac{1}{2})t} - e^{-i(N+\frac{1}{2})t} = 2i\sin\left(\left(N+\frac{1}{2}\right)t\right),
\end{aligned}
$$

so

$$
D_N(t) = \frac{\sin\left(\left(N+\frac{1}{2}\right)t\right)}{\sin\left(\frac{t}{2}\right)}. \tag{6.41}
$$

According to Exercise 16

$$
\int_{-\pi}^{\pi} |D_N(t)|\,dt \to \infty \quad \text{as} \quad n\to\infty. \tag{6.42}
$$

On the other hand, clearly, $\hat{D}_N(n)$ is either 1 or 0 for all n. Hence no estimate of the form (6.38) can hold for all the choices $f = D_N$. It follows from the Open Mapping Theorem that $\mathcal{F}$ is not onto c_0. $\square$

Another question is this: suppose we strengthen the assumption on f and require $f \in C_p([-\pi,\pi])$ (see Lemma 6.2). Is it then true that

$$
f(t) = \lim_{N\to\infty} s_N(f,t) = \lim_{N\to\infty}\sum_{n=-N}^{N}\hat{f}(n)e^{int} = \lim_{N\to\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)D_N(t-s)\,ds \tag{6.43}
$$

for all $t \in [-\pi, \pi]$? Does the Fourier series of a continuous (periodic) function converge everywhere to the function?

The mapping $f \to s_N(f,t)$ is a linear mapping from $C_p([-\pi,\pi])$ into $\mathbb{C}$ for fixed t. If (6.43) holds, then for fixed f, t, $\{|s_N(f,t)|\}_{N=1}^{\infty}$ is bounded, and then by the Uniform Boundedness Principle, there will be a constant M such that

$$|s_N(f,t)| \le M \max_{s\in[-\pi,\pi]} |f(s)| = M\|f\|_{C_p([-\pi,\pi])} \quad (N = 1, 2, \ldots). \tag{6.44}$$

However, this does not hold. For example,

$$s_N(f,0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)D_N(s)\,ds.$$

It follows that the norm of the linear functional $T_N f := s_N(f,0)$ is at most $\int_{-\pi}^{\pi} |D_N(s)|\,ds/2\pi$ because

$$|T_N f| \le \frac{1}{2\pi}\Big|\int_{-\pi}^{\pi} f(s)D_N(s)\,ds\Big| \le \|f\|_{C_p([-\pi,\pi])}\int_{-\pi}^{\pi} |D_N(s)|\,ds. \tag{6.45}$$

In fact,

$$\|T_N\| = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_N(s)|\,ds.$$

To see this, choose a sequence $g_j \in C_p([-\pi,\pi])$ such that $|g_j| \le 1$ and $g_j(s)D_N(s) \to |D_N(s)|$ a.e. (construction explained in class). Then

$$|s_N(g_j,0)| = \frac{1}{2\pi}\Big|\int_{-\pi}^{\pi} g_j(s)D_N(s)\,ds\Big| \le \|T_N\|\|g_j\|_{C_b([-\pi,\pi])} \le \|T_N\| \quad \text{and}$$

$$\Big|\int_{-\pi}^{\pi} g_j(s)D_N(s)\,ds\Big| \to \int_{-\pi}^{\pi} |D_N(s)|\,ds$$

establish this claim. Since we have (6.42), it follows that $\{s_N(f,0)\}$ is unbounded for a dense G_δ set of f's in $C_p([-\pi,\pi])$, and for such an f, $s_N(f,0)$ does not converge to $f(0)$. See Rudin pg 102 for further implications of this argument.

The above results are negative, showing that properties one might hope for do not hold. Here is a positive result:

Theorem 6.4. *Let $f \in L^2(-\pi,\pi)$. Then*

$$\lim_{N\to\infty} \|f - s_N(f)\|_{L^2(-\pi,\pi)} = 0 \tag{6.46}$$

where $s_N(f)$ denotes the function $t \mapsto s_N(f,t)$. That is, the Fourier series of $f \in L^2(-\pi,\pi)$ converges to f in $L^2(-\pi,\pi)$. Moreover, the mapping

$$f \mapsto \left\{\sqrt{2\pi}\hat{f}(n)\right\}_{n=-\infty}^{\infty}, \tag{6.47}$$

is an isometry of $L^2(-\pi,\pi)$ onto l_d^2 (see (4.24) and below).

Proof. Hereafter we write $\| \ \|_2$ instead of $\| \ \|_{L^2(-\pi,\pi)}$. Using the inner-product (6.23) (see the Appendix, Section 12), we have

$$\langle f, f\rangle = \int_{-\pi}^{\pi} |f(t)|^2 \, dt,$$

so $L^2(-\pi,\pi)$ is a Hilbert space when equipped with this inner-product (that is, the norm is just the norm coming from the inner-product). Now $\{e^{int}\}_{n-\infty}^{\infty}$ is a pairwise orthogonal system in this Hilbert space, so the Bessel inequality of (12.13) yields

$$\|f\|_2^2 \geq \|s_N(f)\|_2^2 \quad \text{for} \quad N = 1, 2, \dots. \tag{6.48}$$

It follows that $\|s_N\| \leq 1$. Moreover, for any trigonometric polynomial p, $s_N(p) = p$ for N large enough. Hence for every trigonometric polynomial

$$\begin{aligned}
\|f - s_N(f)\|_2 &= \|f - p + p - s_N(p) + s_n(p) - s_N(f)\|_2 \\
&\leq \|f - p\|_2 + \|p - s_N(p)\|_2 + \|s_N(p) - s_N(f)\|_2 \\
&\leq \|f - p\|_2 + \|p - s_N(p)\|_2 + \|s_N\|\|p - f\|_2 \\
&\leq 2\|f - p\|_2 + \|p - s_N(p)\|_2.
\end{aligned}$$

Using Lemma (6.2), if $\epsilon > 0$, we can choose p so that the first term is at most ϵ, and then N so large that the second term is 0. The claim (6.46) follows.

In view of $\|e^{int}\|_2^2 = 2\pi$, and (6.46), it follows from Proposition 12.4 that

$$\|f\|_2^2 = 2\pi \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2 = \| \left\{ \sqrt{2\pi}\hat{f}(n) \right\}_{n=-\infty}^{\infty} \|_{l_d^2}^2. \tag{6.49}$$

That is, the mapping (6.47) is indeed an isometry.

It remains to see that this mapping is onto. However, we see from Proposition 12.5 that

$$\sum_{j=-\infty}^{\infty} x_n e^{int} \quad \text{converges in } L^2(-\pi,\pi) \tag{6.50}$$

iff and only if $x = (\cdots, x_{-1}, x_0, x_1, \cdots) \in l_d^2$. Thus, when $x \in l_d^2$, we have

$$f := \sum_{j=-\infty}^{\infty} x_n e^{int} \in L^2(-\pi,\pi) \quad \text{and} \quad \hat{f}(n) = x_n, \ (n \in \mathbb{Z}).$$

Thus the mapping is onto.

□

7. Convolutions and Approximations

Below, we discuss three types of mappings $T : \mathbb{R}^n \to \mathbb{R}^n$, translations, dilations and reflections. These are given by

$$
Tx = \begin{cases} \tau_h x := x + h \quad (\text{translation by } h \in \mathbb{R}^n), \\ D_\delta x := \delta x \quad (\text{dilation by } \delta > 0), \\ Rx = -x \quad (\text{reflection}). \end{cases} \tag{7.1}
$$

Notice that, in general, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and invertible (1-1 and onto), then

$$
\chi_{TE}(x) = 1 \iff x \in TE \iff T^{-1}x \in E \iff \chi_E(T^{-1}x) = 1. \tag{7.2}
$$

Using this, if $\mathcal{L}^n(TE) = \lambda \mathcal{L}^n(E)$ for some constant λ and all E, the following then holds for all characteristic functions, hence all simple functions, and hence all $f \in L^1(\mathbb{R}^n)$:

$$
\int_{\mathbb{R}^n} f(T^{-1}x)\, dx = \lambda \int_{\mathbb{R}^n} f(x)\, dx. \tag{7.3}
$$

This includes the claim that if $f \in L^1(\mathbb{R}^n)$, then so is $x \mapsto f(T^{-1}x)$.

Since $\lambda = 1$ for translations and reflections, while $\lambda = \delta^n$ for D_δ, we conclude that

$$
\begin{aligned}
\int_{\mathbb{R}^n} f(x-h)\, dx &= \int_{\mathbb{R}^n} f(x)\, dx, \\
\int_{\mathbb{R}^n} f(x/\delta)\, dx &= \delta^n \int_{\mathbb{R}^n} f(x)\, dx, \\
\int_{\mathbb{R}^n} f(-x)\, dx &= \int_{\mathbb{R}^n} f(x)\, dx.
\end{aligned} \tag{7.4}
$$

Let us also note the following about dilations: for $r > 0$

$$
\begin{aligned}
\int_{\{r \le |x|\}} f\left(\frac{x}{\delta}\right)\, dx &= \int_{\mathbb{R}^n} \chi_{\{y : r \le |y|\}}(x) f\left(\frac{x}{\delta}\right)\, dx \\
&= \int_{\mathbb{R}^n} \chi_{\{y : r/\delta \le |y|\}}\left(\frac{x}{\delta}\right) f\left(\frac{x}{\delta}\right)\, dx \\
&= \delta^n \int_{\mathbb{R}^n} \chi_{\{y : r/\delta \le |y|\}}(x)\, f(x)\, dx \\
&= \delta^n \int_{\{r/\delta \le |x|\}} f(x)\, dx.
\end{aligned} \tag{7.5}
$$

The *convolution* $f * g$ of two functions f, g on $\mathbb{R}^n$ is the function defined by

$$
f * g(x) = \int_{\mathbb{R}^n} f(x-y) g(y)\, dy \tag{7.6}
$$

for those x's for which $y \mapsto f(x-y)g(y)$ is in $L^1(\mathbb{R}^n)$.

Using (7.4) several times, we have

$$
\begin{aligned}
\int_{\mathbb{R}^n} f(x-y)g(y)\,dy &= \int_{\mathbb{R}^n} f(x-(y+x))g(y+x)\,dy \\
&= \int_{\mathbb{R}^n} f(-y)g(y+x)\,dy \\
&\int_{\mathbb{R}^n} f(y)g(x-y)\,dy.
\end{aligned} \tag{7.7}
$$

That is, $f * g = g * f$; the convolution "product" is commutative.

Theorem 7.1. *Let $f \in L^1(\mathbb{R}^n)$ and $g \in L^p(\mathbb{R}^n)$ where $1 \leq p \leq \infty$. Then $f * g$ is defined ae, measurable, and*

$$
\|f * g\|_{L^p(\mathbb{R}^n)} \leq \|f\|_{L^1(\mathbb{R}^n)} \|g\|_{L^p(\mathbb{R}^n)}. \tag{7.8}
$$

Proof. It suffices to assume that $f, g \geq 0$. The integral on the right of (7.6) is then defined for every x, although it may be infinite. The first issue is the measurability of the right-hand side of (7.6). However, there are Borel measurable functions f_0, g_0 such that $f = f_0$, $g = g_0$ a.e. (See Exercise 21.) The maps

$$
\begin{aligned}
\mathbb{R}^n \times \mathbb{R}^n \ni (x,y) \to x - y \in \mathbb{R}^n \\
\mathbb{R}^n \times \mathbb{R}^n \ni (x,y) \to y \in \mathbb{R}^n
\end{aligned}
$$

are continuous and hence Borel Measurable, so $(x,y) \to f_0(x-y)$ and $(x,y) \to g_0(y)$ are Borel measurable on $\mathbb{R}^n \times \mathbb{R}^n$. By the proof of Fubini,

$$
x \mapsto \int_{\mathbb{R}^n} f_0(x-y)g_0(y)\,dy = \int_{\mathbb{R}^n} f(x-y)g(y)\,dy
$$

is Borel measurable.

The case $p = \infty$ is immediate. For $p = 1$ we have, via Fubini,

$$
\begin{aligned}
\|f * g\|_{L^1(\mathbb{R}^n)} &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y)\,dy \right) dx \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y)\,dx \right) dy \\
&= \|f\|_{L^1(\mathbb{R}^n)} \int_{\mathbb{R}^n} g(y)\,dy \\
&= \|f\|_{L^1(\mathbb{R}^n)} \|g\|_{L^1(\mathbb{R}^n)}.
\end{aligned} \tag{7.9}
$$

Assume that $1 < p < \infty$. Now, using Hölder, $p' = p/(p-1)$, Fubini, and the identity above,

$$\begin{aligned}\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(x-y)g(y)\,dy\right)^p dy &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(x-y)^{\frac{p-1}{p}} f(x-y)^{\frac{1}{p}} g(y)\,dy\right)^p dx \\ &\le \int_{\mathbb{R}^n}\left(\left(\int_{\mathbb{R}^n} f(x-y)\,dy\right)^{p-1}\int_{\mathbb{R}^n} f(x-y)g(y)^p\,dy\right) dx \\ &= \|f\|^{p-1}_{L^1(\mathbb{R}^n)}\|f * g^p\|_{L^1(\mathbb{R}^n)} = \|f\|^p_{L^1(\mathbb{R}^n)}\|g\|^p_{L^p(\mathbb{R}^n)}.\end{aligned}$$

$\square$

7.1. **Approximate Identities.** The function

$$H_t(x) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/4t} \tag{7.10}$$

solves the heat equation

$$\frac{\partial u}{\partial t} - \Delta u = \frac{\partial u}{\partial t} - \left(\frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}\right) = 0 \tag{7.11}$$

on $\mathbb{R}^n \times (0,\infty)$. A solution u of the heat equation which satisfies the initial condition $u(0,x) = f(x)$ is given by

$$u(t,x) = \int_{\mathbb{R}^n} H_t(x-y)f(y)\,dy; \quad \text{equivalently,} \quad u(t,x) = (H_t * f)(x). \tag{7.12}$$

H_t is called the *heat kernel.*

Clearly $H_t(\cdot)$ belongs to every $L^p(\mathbb{R}^n)$, $1 \le p \le \infty$. Exercise 24 indicates why this u solves the heat equation for, for example, $f \in L^1(\mathbb{R}^n)$. Derivatives may be calculated under the integral sign.

We examine a sense in which u satisfies the initial condition $u(0,x) = f(x)$. We do this more generally. H_t is a special case of the following situation. Take a function $\rho \in L^1(\mathbb{R}^n)$, $\rho \ge 0$, which satisfies

$$\int_{\mathbb{R}^n} \rho\,dx = 1. \tag{7.13}$$

For $\epsilon > 0$ define

$$\rho_\epsilon(x) = \frac{1}{\epsilon^n}\rho\left(\frac{x}{\epsilon}\right) \tag{7.14}$$

Note that, using (7.5),

$$\int_{\mathbb{R}^n} \rho_\epsilon(x)\,dx = \frac{1}{\epsilon^n}\int_{\mathbb{R}^n} \rho\left(\frac{x}{\epsilon}\right)\,dx = 1. \tag{7.15}$$

With this notation, and a bit of notational abuse, if $\rho(x) = H_1(x)$, then $\rho_{\sqrt{t}} = H_t$ and (7.13) is satisfied:

$$\int_{\mathbb{R}^n} \rho\, dx = \int_{\mathbb{R}^n} H_1(x)\, dx = 1; \tag{7.16}$$

we assume that you know the last equality - it follows from $\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$, a Math 5 calculation, and Fubini. It then follows from the next theorem that if $f \in L^p(\mathbb{R}^n)$, $1 \le p < \infty$, then $H_t * f \to f$ in $L^p(\mathbb{R}^n)$ as $t \downarrow 0$. Later we will see that this convergence also holds almost everywhere.

Theorem 7.2. *Let $\rho \in L^1(\mathbb{R}^n)$, $\rho \ge 0$, satisfy (7.13). Let $f \in L^p(\mathbb{R}^n)$ and $1 \le p < \infty$. Then*

$$\lim_{\epsilon \downarrow 0} \|f - \rho_\epsilon * f\|_{L^p(\mathbb{R}^n)} = 0. \tag{7.17}$$

Proof. Since $\rho \in L^1(\mathbb{R}^n)$, for every $\delta > 0$ there exists R_δ such that

$$\int_{\{|x|>R_\delta\}} \rho(x)\, dx < \delta \quad \text{if} \quad R \ge R_\delta. \tag{7.18}$$

Hence

$$\int_{\{|x|>R\}} \rho_\epsilon(x)\, dx = \frac{1}{\epsilon^n}\int_{\{|x|>R\}} \rho\left(\frac{x}{\epsilon}\right)\, dx = \int_{\{|y|>R/\epsilon\}} \rho(y)\, dy < \delta \quad \text{if} \quad \frac{R}{\epsilon} \ge R_\delta. \tag{7.19}$$

Hence

$$\lim_{\epsilon \downarrow 0} \int_{\{R \le |x|\}} \rho_\epsilon(x)\, dx = 0 \quad \text{for} \quad R > 0. \tag{7.20}$$

Let $f \in C_c(\mathbb{R}^n)$ (the continuous functions on $\mathbb{R}^n$ with compact support). We first claim that

$$\lim_{\epsilon \downarrow 0} \rho_\epsilon * f(x) = f(x) \quad \text{unformly for} \quad x \in \mathbb{R}^n. \tag{7.21}$$

To see this, note that

$$\begin{aligned} \rho_\epsilon * f(x) - f(x) &= \frac{1}{\epsilon^n}\int_{\mathbb{R}^n} \rho\left(\frac{x-y}{\epsilon}\right)(f(y) - f(x))\, dy \\ &= \frac{1}{\epsilon^n}\int_{\{|x-y|\le R\}} \rho\left(\frac{x-y}{\epsilon}\right)(f((y) - f(x))\, dy \\ &\quad + \frac{1}{\epsilon^n}\int_{\{|x-y|\ge R\}} \rho\left(\frac{x-y}{\epsilon}\right)(f(y) - f(x))\, dy. \end{aligned} \tag{7.22}$$

The first term of the rightmost expression is estimated by

$$
\begin{aligned}
&\Big|\frac{1}{\epsilon^n}\int_{\{|x-y|\le R\}}\rho\left(\frac{x-y}{\epsilon}\right)(f(y)-f(x))\,dy\Big| \\
&\qquad\le \frac{1}{\epsilon^n}\int_{\{|x-y|\le R\}}\rho\left(\frac{x-y}{\epsilon}\right)\,dy\max_{|x-y|\le R}|f(y)-f(x)| \\
&\qquad\le \max_{\{|x-y|\le R\}}|f(y)-f(x)|.
\end{aligned}
\tag{7.23}
$$

Since f is uniformly continuous, given $\kappa>0$, we may choose R sufficiently small (even if "R" does not look small) to guarantee that this quantity is at most κ. Fix this κ and then R.

The second term of the rightmost expression is estimated by

$$
\begin{aligned}
&\Big|\frac{1}{\epsilon^n}\int_{\{|x-y|\ge R\}}\rho\left(\frac{x-y}{\epsilon}\right)(f(y)-f(x))\,dy\Big| \\
&\qquad\le \left(\frac{1}{\epsilon^n}\int_{\{|x-y|\ge R\}}\rho\left(\frac{x-y}{\epsilon}\right)\,dy\right)2\max_{\mathbb{R}^n}|f|.
\end{aligned}
\tag{7.24}
$$

By (7.18), for all ϵ sufficiently small this at most κ; therefore $|\rho_\epsilon * f - f|\le 2\kappa$ as soon as ϵ is sufficiently small, and (7.21) is proved.

Still taking $f\in C_c(\mathbb{R}^n)$, we claim that $\rho_\epsilon * f\to f$ in $L^1(\mathbb{R}^n)$. Indeed, as $\epsilon\downarrow 0$,

$$|\rho_\epsilon * f|\le \rho_\epsilon * |f|\to |f| \quad \text{everywhere}$$

while, using Fubini, (we did this before, (7.9)),

$$\int_{\mathbb{R}^n}\rho_\epsilon * |f|\,dx=\int_{\mathbb{R}^n}\left(\frac{1}{\epsilon^n}\int_{\mathbb{R}^n}\rho\left(\frac{x-y}{\epsilon}\right)|f(y)|\,dy\right)dx=\int_{\mathbb{R}^n}|f(y)|\,dy.$$

It follows from the LDC theorem, in the form we have it on page 8 (or thereabouts), that $\rho_\epsilon * f\to f$ in $L^1\left(\mathbb{R}^n\right)$. From this and

$$|\rho_\epsilon * f(x)-f(x)|\le 2\max_{\mathbb{R}^n}|f|$$

we deduce, for $1\le p<\infty$, that

$$\int_{\mathbb{R}^n}|\rho_\epsilon * f(x)-f(x)|^p\,dx\le\left(2\max_{\mathbb{R}^n}|f|\right)^{p-1}\|\rho_\epsilon * f-f\|_{L^1(\mathbb{R}^n)}\to 0 \tag{7.25}$$

as $\epsilon\downarrow 0$. Since $C_c(\mathbb{R}^n)$ is dense in $L^p\left(\mathbb{R}^n\right)$, and the norm of the linear mapping $f\mapsto\rho_\epsilon * f$ as an element of $\mathcal{L}(L^p\left(\mathbb{R}^n\right),L^p\left(\mathbb{R}^n\right))$ is at most 1 (Lemma 7.1), we are done (Exercise 18). $\square$

7.2. **Proof of Theorem 6.2.** At last, we indicate the proof of Lemma 6.2. Set

$$Q_k(t) = c_k \left(\frac{1+\cos t}{2}\right)^k \quad (k = 1, 2, \ldots) \tag{7.26}$$

where c_k is choosen so that

$$\int_{-\pi}^{\pi} Q_k(t)\,dt = 1. \tag{7.27}$$

Let $f \in C_p([-\pi, \pi])$, and put

$$f_k(t) = \int_{-\pi}^{\pi} Q_k(t-s) f(s)\,ds; \tag{7.28}$$

we then have

$$f_k(t) - f(t) = \int_{-\pi}^{\pi} Q_k(t-s)(f(s) - f(t))\,ds. \tag{7.29}$$

Clearly f_k is a trigonometric polynomial. The key facts about Q_k are related to those for the ρ_ϵ above: $Q_k \geq 0$, (7.27), and

$$\lim_{k\to\infty} \max_{\delta \leq |t| \leq \pi} Q_k(t) = 0 \quad \text{for } \delta > 0. \tag{7.30}$$

To see this, we use that $Q_k(t)$ is even in t and (7.27) to conclude that

$$1 = 2c_k \int_0^{\pi} \left(\frac{1+\cos t}{2}\right)^k dt \geq 2c_k \int_0^{\pi} \left(\frac{1+\cos t}{2}\right)^k \sin t\,dt = \frac{4c_k}{(k+1)},$$

so $c_k \leq (k+1)/4$, and

$$\max_{\delta \leq |t| \leq \pi} Q_k(t) \leq \frac{k+1}{4} \left(\frac{1+\cos\delta}{2}\right)^k \to 0 \quad \text{as} \quad k \to \infty.$$

Now break up the integral in (7.29) into a the pieces where $|t-s| < \delta$ and where $|t-s| \geq \delta$ and mimic (7.23) and (7.24), using (7.30) in place of (7.20); it follows that $f_k - f \to 0$ uniformly.

Finally, to show the density of trig polynomials in $L^p(-\pi, \pi)$, $1 \leq p < \infty$, we use the density of continuous functions on $[-\pi, \pi]$ in these spaces (Theorem 4.5) and a bit of hand waving in class to show that this guarantees the density of $C_p([-\pi, \pi])$, and once we know that, for $f \in L^p(-\pi, \pi)$ and $g \in C_p([-\pi, \pi])$, we have, for every trig poly $P(t)$,

$$\|f - P\|_p \leq \|f - g\|_p + \|g - P\|_p \leq \|f - g\|_p + (2\pi)^{1/p} \|g - P\|_{C_p([-\pi,\pi])},$$

whence the result (choose g to make the first term on the right as small as desired, then P to make the second term as small as desired).

Remark 7.3. It matters how one adds things up. We showed that the Fourier series of a continuous periodic function f does not neccesarily converge at 0 to $f(0)$, and this was so for a dense G_δ bunch of functions. You showed that the Fourier series of an integrable function does not converge to the function in L^1 (Exercise 17). However, if one adds things up differently, these irritating facts go away. That is, it is true that

$$f = \lim_{N\to\infty} \frac{s_0(f) + s_1(f) + \cdots + s_N(f)}{N+1},$$

with uniform convergence if $f \in C_p([-\pi,\pi])$, and with $L^p(-\pi,\pi)$ convergence if $f \in L^p(-\pi,\pi)$, $1 \le p < \infty$. This is because the kernel

$$K_N := \frac{D_0 + D_1 + \cdots + D_N}{N+1}$$

has properties like the ones we used in this section. This method of summing a possibly divergent series, using the sequence of averages, is called the method of Cesáro means. The kernels K_N are called the Fejér kernels. Read about Fourier series on Wikipedia for more info. The subject is infinite.

8. Mazur-Orlicz, Hahn-Banach, Minimax, and —

The presentation of the material in this section is adapted to our taste from slides for seminars given this year by Professor Stephen Simons. These slides, now modified, are available at

www.math.ucsb.edu/˜ simons/SOAFA.html

together with a preprint of a related paper, The Hahn-Banach-Lagrange Theorem, found via Simons main web page (click through available via the above site), which just appeared in the journal *Optimization*.

As a source for the results in the section title, I think the current notes, which uses ideas of Simons, are as slick as it gets.

8.1. **Preliminaries.** We need a few concepts and elementary facts about them. Let E be a real vector space. A function $P : E \to \mathbb{R}$ is *sublinear* if

$$\begin{aligned} &\text{(i) } P(\lambda x) = \lambda P(x) \quad (x \in E, 0 \le \lambda), \\ &\text{(ii) } P(x+y) \le P(x) + P(y) \quad (x, y \in E). \end{aligned} \tag{8.1}$$

When P satisfies (8.1) (i), it is called *positively homogeneous*, and when P satisfies (8.1) (ii) it is called *subadditive*. The conditions (8.1) look familiar, they are satisfied by any norm on E. However, a norm has the additional property $\|-x\| = \|x\|$, which is not satisfied by general subadditive functions; for example, linear functionals are sublinear. Conversely, if P is sublinear and $P(x) = P(-x)$, then it is what we once called a "pseudo norm" ("semi-norm" is more common); it then has all the properties of a norm except that $P(x) = 0$ only for $x = 0$ does not necessarily hold.

The example $P : \mathbb{R}^n \to \mathbb{R}$ given by

$$P(x) = \bigvee_{j=1}^{n} x_j := \max\{x_j : j = 1, \ldots, n\} \tag{8.2}$$

is sublinear and will pop up later. It is clear that this P is positively homogeneous. It is also clear that it is subadditive, or

$$\bigvee_{j=1}^{n}(x_j + y_j) \leq \bigvee_{j=1}^{n} x_j + \bigvee_{j=1}^{n} y_j,$$

since the biggest $x_j + y_j$ is certainly at most the biggest of the x_j plus the biggest of the y_j. A similar observation is:

Finally, we note that if P is sublinear, then

$$P(0) = 0, \text{ and } P(0) = P(x - x) \leq P(x) + P(-x) \implies -P(-x) \leq P(x). \tag{8.3}$$

8.2. **The Mazur-Orlicz Lemma.** Everywhere below, E is a nonzero real vector space and P is a sublinear functional on E.

Theorem 8.1 (Mazur-Orlicz Lemma). *Let D be a convex subset of E. Then there is a linear functional L on E such that*

$$L \leq P \quad \textit{on} \quad E \quad \textit{and} \quad \inf_D L = \inf_D P. \tag{8.4}$$

Remark 8.2. Simons calls Theorem 8.1 with $D = \{0\}$ the *Hahn-Banach Theorem*. Other people think of Theorem 8.5 below as the Hahn-Banach Theorem. Simons' slides first present Theorem 8.1 with $D = \{0\}$ (when the result reduces to $L \leq P$) and then derive the general case. We do the same, with the $D = \{0\}$ case being Lemma 8.4 - we just don't give it a name.

The following lemma is a key ingredient in the proof of Theorem 8.1.

Lemma 8.3. *Let D be a convex subset of E and $\kappa := \inf_D P > -\infty$. Then*

$$P_D(x) := \inf_{y \in D,\ 0 < \lambda} (P(x + \lambda y) - \lambda\kappa) \tag{8.5}$$

is a sublinear functional on E which satisfies $P_D \leq P$.

Proof. Clearly $P_D \leq P$. If $y \in D$, we have

$$\lambda\kappa \leq P(\lambda y) = P(x + \lambda y - x) \leq P(x + \lambda y) + P(-x)$$

and therefore

$$-P(-x) \leq P(x + \lambda y) - \lambda\kappa \implies -P(-x) \leq P_D(x) \tag{8.6}$$

the import being that $-\infty < P_D$ everywhere.

Moreover, if $x_1, x_2 \in E$, $0 < \lambda_1, \lambda_2$, and $y_1, y_2 \in D$, then

$$\begin{aligned} P_D(x_1 + x_2) &\leq P\left(x_1 + x_2 + (\lambda_1 + \lambda_2)\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} y_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} y_2\right)\right) - (\lambda_1 + \lambda_2)\kappa \\ &= P(x_1 + \lambda_1 y_1 + x_2 + \lambda_2 y_2) - (\lambda_1 + \lambda_2)\kappa \\ &\leq P(x_1 + \lambda_1 y_1) - \lambda_1\kappa + P(x_2 + \lambda_2 y_2) - \lambda_2\kappa, \end{aligned}$$

so P_D is subadditive; it is also trivially positive homogeneous. $\square$

Lemma 8.4. *There is a sublinear functional Q on E such that $Q \leq P$ and if $\hat{Q}$ is another sublinear functional satisfying $\hat{Q} \leq P \wedge Q$, then $Q = \hat{Q}$. Moreover, any such Q is linear.*

Proof. Consider the set

$$\mathcal{Q} = \{T : E \to \mathbb{R};\ T \ \text{ is sublinear and } T \leq P\}$$

of sublinear functions which are less or equal to P. It is naturally partially ordered by "$\leq$ ", that is $\leq$ defines a reflexive, antisymmetric, transitive relation on $\mathcal{Q}$. Here "antisymmetric" means that $T \leq \hat{T}$ and $\hat{T} \leq T$ only if $T = \hat{T}$. A "chain" $\mathcal{C}$ in $\mathcal{Q}$ is a nonempty "totally ordered subset" of $\mathcal{Q}$. This means that if $T, \hat{T} \in \mathcal{C}$, then either $T \leq \hat{T}$ or $\hat{T} \leq T$. According to *Zorn's Lemma,* which is equivalent to the axiom of choice, if we can show that for any chain $\mathcal{C}$ in $\mathcal{Q}$ there is a $Q \in \mathcal{Q}$ such that $Q \leq T$ for $T \in \mathcal{C}$, (i.e., Q is a *lower bound* for $\mathcal{C}$), then there is a minimal element Q of $\mathcal{Q}$. Here Q is "minimal" means that if $\hat{Q} \in \mathcal{Q}$ and $\hat{Q} \leq Q$, then $\hat{Q} = Q$.

The proof proceeds by showing that any chain has a lower bound, and then that any minimal element of $\mathcal{Q}$ is necessarily linear.

To see that a chain $\mathcal{C}$ has a lower bound, define

$$Q(x) = \inf_{T \in \mathcal{C}} T(x).$$

If we show that Q is sublinear, we have produced a lower bound for $\mathcal{C}$. First we note that $-P \leq -T$ implies that $-P(-x) \leq -T(-x) \leq T(x)$ by (8.3). Hence $-P(-x) \leq Q(x)$, and Q is finite everywhere. Now suppose that $\hat{T}, T \in \mathcal{C}$. As $\hat{T} \wedge T := \min\left\{\hat{T}, T\right\} \geq Q$, and $\hat{T} \wedge T$ is either $\hat{T}$ or T (since $\mathcal{Q}$ is a chain) we have, for $\hat{x}, x \in E$,

$$\begin{aligned} Q(\hat{x} + x) &\leq \hat{T}(\hat{x} + x) \wedge T(\hat{x} + x) \\ &\leq (\hat{T}(\hat{x}) + \hat{T}(x)) \wedge (T(\hat{x}) + T(x)) \leq \hat{T}(\hat{x}) + T(x), \end{aligned}$$

so

$$Q(\hat{x} + x) \leq \hat{T}(\hat{x}) + T(x).$$

Infing over $\hat{T}$ and then T, we conclude that Q is subadditive. The infimum of positively homogeneous functions is clearly positively homogeneous, and $\mathcal{Q}$ thus contains a minimal element Q.

Now fix $y \in E$ and put $D = \{y\}$ and $P = Q$ in Lemma 8.3, producing $Q_{\{y\}} \leq Q$. By minimality of Q, $Q = Q_{\{y\}}$; in particular

$$Q(x) = Q_{\{y\}}(x) \leq Q(x + \lambda y) - \lambda Q(y)$$

for $0 < \lambda$, $x \in E$ and $y \in E$. Since also $Q(x + \lambda y) \leq Q(x) + \lambda Q(y)$, we have equality,

$$Q(x + \lambda y) = Q(x) + \lambda Q(y).$$

Therefore linearity of Q will follow if we show that $Q(-z) = -Q(z)$ for $z \in E$. Put $x = -2z$, $y = z$ and $\lambda = 1$ above to find

$$Q(-z) = Q(-2z) + Q(z) = 2Q(-z) + Q(z) \implies Q(-z) = -Q(z).$$

□

Proof of Theorem 8.1 First, we may assume that $\kappa := \inf_D P > -\infty$. Indeed, if $\inf_D P = -\infty$, we may replace D by $\{0\}$, and then the inequality of (8.4) implies the equality in (8.4) with the original D. Apply Lemma 8.4 to P_D in place of P, obtaining a minimal subadditive function L, which will be linear, satisfying $L \le P_D$. For $y \in D$,

$$-L(y) = L(-y) \le P_D(-y) \le P(-y + \lambda\hat{y}) - \lambda\kappa$$

for $\hat{y} \in D$, $\lambda > 0$. Choosing $\lambda = 1, \hat{y} = y$ yields $-L(y) \le -\kappa$. Thus $\inf_D L \ge \kappa = \inf_D P$; the opposite inequality follows from $L \le P_D \le P$. □

8.3. **Consequences of the Mazur-Orlicz Lemma.** In the results immediately below, we sometimes use the notation $\langle x^*, x\rangle$ to denote the value of a continuous linear functional x^* at a vector x in its domain and E^*, F^* are the spaces of continuous linear functionals on E and F, respectively.

Theorem 8.5 (Hahn-Banach Extension Theorem)**.** *Let E be a real normed space, F be a subspace of E and $y^* \in F^*$. Then there exists $x^* \in E^*$ such that $x^*|_F = y^*$ and $\|x^*\|_{E^*} = \|y^*\|_{F^*}$.*

Proof. Define $P : E \times F \to \mathbb{R}$ by

$$P(x, y) = \|y^*\|_{F^*}\|x\| - \langle y^*, y\rangle.$$

Let $D = \{(y, y) : y \in F\}$. Notice that $P(y, y) \ge 0$ for $y \in F$ and $P(0, 0) = 0$ to conclude that

$$\inf_D P = \inf_{y\in F} P(y, y) = 0.$$

Apply Theorem 8.1 to conclude that there exists a linear $L : E \times F \to \mathbb{R}$ such that

(i) $L(x, y) \le \|y^*\|_{F^*}\|x\| - \langle y^*, y\rangle \quad \text{for} \quad x \in E,\ y \in F.$

(ii) $\inf_{y\in F} L(y, y) = 0.$

Choosing $x = 0$ in (i) yields $L(0, y) = -\langle y^*, y\rangle$ (if two linear functionals are ordered, they are equal, or, as they say in Spain, a plane below a plane, is the plane :)). Choosing $y = 0$ in (i) we see that $L(x, 0) \le \|y^*\|_{F^*}\|x\|$, which implies (upon replacing x by $-x$) that $|L(x, 0)| \le \|y^*\|_{F^*}\|x\|$. In particular, the map $x \mapsto L(x, 0)$ is continuous on E; let us call it $x^* \in E^*$, and $\|x^*\|_{E^*} \le \|y^*\|$. In view of

$$L(y, y) = L(y, 0) + L(0, y) = \langle x^*, y\rangle - \langle y^*, y\rangle \quad (y \in F),$$

(ii) becomes $\langle x^*, y\rangle \ge \langle y^*, y\rangle$ for $y \in F$, which forces $x^*\big|_F = y^*$. □

Corollary 8.6 (complex Hahn-Banach)**.** *Let E be a complex normed space, F be a subspace of E and $y^* \in F^*$. Then there exists $x^* \in E^*$ such that $x^*|_F = y^*$ and $\|x^*\|_{E^*} = \|y^*\|_{F^*}$.*

Proof. Given $y^* \in F^*$, consider the real linear functional z^* on F defined by

$$\langle z^*, y\rangle := \Re\langle y^*, y\rangle.$$

Let w^* be a real norm-preserving extension of z^* to E as provided by Theorem 8.5. Then

$$\langle x^*, x\rangle := \langle w^*, x\rangle - i\langle w^*, ix\rangle$$

is a complex linear extension of y^* to E. Clearly x^* is real linear, and the full complex linearity follows from $\langle x^*, ix\rangle = i\langle x^*, x\rangle$, a consequence of the definition of x^*. Then one checks that the real and imaginary parts of x^* agree with the real and imaginary parts of y^* on F. To verify the norm preserving property, $\|x^*\|_{E^*} = \|y^*\|_{F^*}$, given $x \in E$, choose $\alpha \in \mathbb{C}$, $|\alpha| = 1$, such that $\alpha\langle x^*, x\rangle = |\langle x^*, x\rangle|$ and then

$$|\langle x^*, x\rangle| = |\langle x^*, \alpha x\rangle| = \langle w^*, \alpha x\rangle \le \|w^*\|_{(\text{real})E^*}\|x\| = \|z^*\|_{(\text{real})F^*}\|x\| \le \|y^*\|_{F^*}\|x\|.$$

□

Remark 8.7. While the above deduction of Corollary 8.6 from Theorem 8.5 seems inevitable - what else could one do? - and transparent, very good mathematicians, including Banach, could not find it for a time after Theorem 8.5 was proved. See the comments about this in Rudin, page 105. Question: What is the moral of this story?

Corollary 8.8. *Let F be a closed linear subspace of the normed space E and $x_0 \notin F$. Then there exists $x^* \in E^*$ such that $\|x^*\| \le 1$, and $\langle x^*, x_0\rangle = \operatorname{dist}(x_0, F)$ and $\langle x^*, F\rangle = \{0\}$. In consequence, if F is not necessarily closed, then*

$$\bar{F} = \{x \in E : x^* \in E^* \text{ and } \langle x^*, F\rangle = \{0\} \implies \langle x^*, x\rangle = 0\}.$$

Proof. By assumption, $\operatorname{dist}(x_0, F) > 0$. Define y^* on the linear span of F and x_0 by

$$\langle y^*, y + \lambda x_0\rangle = \lambda \operatorname{dist}(x_0, F).$$

Since

$$\|y + \lambda x_0\| = |\lambda| \left\| \frac{y}{\lambda} + x_0 \right\| \ge |\lambda| \operatorname{dist}(x_0, F) \quad (\lambda \ne 0, y \in F),$$

we have $\|y^*\|_{F^*} \le 1$. We are done upon applying Corollary 8.6. □

Remark 8.9. Taking $F = \{0\}$, we learn that if $x_0 \ne 0$, then there exists $x^* \in E^*$, $\|x^*\| \le 1$, such that $\langle x^*, x_0\rangle = \|x_0\|$ (and so $\|x^*\| = 1$). See the remarks in Rudin, page 108.

For convenience, we will call a vector $\lambda = (\lambda_1, \dots, \lambda_m) \in R^m$ a *probability vector* if

$$0 \le \lambda_1, \dots, \lambda_m \quad \text{and} \quad \lambda_1 + \cdots + \lambda_m = 1. \tag{8.7}$$

Lemma 8.10 (Lemma on m Convex Functions)**.** *Let C be a nonempty convex subset of E and $f_1, \dots, f_m$ be m convex real-valued functions on C. Then there is a probability vector $\lambda \in I\!R^m$ such that*

$$\inf_C(\lambda_1 f_1 + \cdots + \lambda_m f_m) = \inf_C[f_1 \vee \cdots \vee f_m]. \tag{8.8}$$

Proof. Define $P : \mathbb{R}^m \to \mathbb{R}$ and $D \subset \mathbb{R}^m$ by

$$P(x_1, \dots, x_m) = \bigvee_{j=1}^{m} x_j := x_1 \vee \dots \vee x_m, \; x = (x_1, \dots, x_m) \in \mathbb{R}^m,$$
$$D = \{(d_1, \dots, d_m) : \exists\, c \in C \text{ such that } f_j(c) \le d_j, j = 1, \dots, m\}\,.$$

P is sublinear and D is convex, so by Theorem 8.1 there exists $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^m$ such that

$$\text{(8.9)} \qquad \begin{aligned} &\text{(i) } \langle \lambda, x\rangle = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_m x_m \le \bigvee_{j=1}^{m} x_j \quad \text{for} \quad x \in \mathbb{R}^m, \\ &\text{(ii) } \inf_D \langle \lambda, x\rangle = \inf_D \left(\bigvee_{j=1}^{m} x_j \right). \end{aligned}$$

Putting $x = (-1, 0, \dots, 0)$ in (8.9) (i) we find

$$-\lambda_1 \le 0 \implies 0 \le \lambda_1; \quad \text{similarly,} \quad 0 \le \lambda_j, j = 1, \dots, m.$$

Putting $x = (-1, -1, \dots, -1)$ and then $x = (1, 1, \dots, -1)$ in (8.9) (i) yields

$$1 \le \lambda_1 + \dots + \lambda_m \le 1 \quad \text{or} \quad \lambda_1 + \dots + \lambda_m = 1.$$

In view of the definition of D, (8.9) (ii) just amounts to the second relation of (8.8). $\square$

If C is a convex subset of a real vector space E, then $k : C \to]-\infty, \infty]$ is *convex* if

$$\text{(8.10)} \qquad k((1-t)x + ty) \le (1-t)k(x) + tk(y) \quad (0 \le t \le 1, x, y \in C).$$

Similarly, k is *concave* if $-k$ is convex, or

$$\text{(8.11)} \qquad (1-t)k(x) + tk(y) \le k((1-t)x + ty) \quad (0 \le t \le 1, x, y \in C).$$

Theorem 8.11 (Minimax)**.** *Let A be a nonempty convex subset of a vector space, B be a nonempty convex subset of a (possibly different) vector space and B also be a compact space. Let $h : A \times B \to \mathbb{R}$ be concave on A and convex and lower-semicontinuous on B. Then*

$$\text{(8.12)} \qquad \sup_{a\in A} \min_{b \in B} h(a, b) = \min_{b\in B} \sup_{a \in A} h(a, b).$$

Remark 8.12. If A, B are any sets and $h : A \times B \to \mathbb{R}$, then

$$\sup_{a\in A} \inf_{b \in B} h(a, b) \le \inf_{b\in B} \sup_{a \in A} h(a, b).$$

The inequality can be strict. For example, take $A = B = \{0, 1\}$ and define $h(0, 1) = h(1, 0) = 0$ and $h(0, 0) = h(1, 1) = 1$.

Remark 8.13. (von Neumann Minimax Theorem) The first result of this kind was due to von Neumann; it is important in a number of fields, in particular, it is a keystone of game theory. In his original work, von Neumann considered the case in which A is the set of probability vectors in $\mathbb{R}^n$, B is the set of probability vectors in $\mathbb{R}^m$ and

$$h(a,b) = \langle b, Ma\rangle$$

where M is an $m \times n$ real matrix. As h is linear in each argument, this is a very special case of Theorem 8.11.

Proof. Let

$$\beta := \sup_{a\in A} \min_{b\in B} h(a,b). \tag{8.13}$$

Assuming that $\beta < \min_{b\in B} \sup_{a\in A} h(a,b)$, we derive a contradiction, proving the theorem. With this assumption, we have

$$\cup_{a\in A} \{b \in B : h(a,b) > \beta\} = B.$$

Since h is lower-semicontinuous on B, the sets $\{b \in B : h(a,b) > \beta\}$ are open, and since B is compact, it is covered by finitely many of these sets. Let $a_1, a_2, \ldots, a_m \in A$ be such that

$$\{b \in B : h(a_1,b) > \beta\} \cup \cdots \cup \{b \in B : h(a_m,b) > \beta\} = B.$$

That is,

$$\min_{b\in B} \left(\bigvee_{j=1}^{m} h(a_j,b) \right) > \beta.$$

From the Lemma on m convex functions, with $f_j(b) = h(a_j,b)$, there exists

$$0 \le \lambda_j,\ \ \lambda_1 + \cdots + \lambda_m = 1,$$

such that

$$\min_{b\in B} (\lambda_1 h(a_1,b) + \cdots + \lambda_m h(a_m,b)) \ge \beta.$$

By concavity of h on A, this implies

$$h(\lambda_1 a_1 + \cdots + \lambda_m a_m, b) > \beta,$$

which contradicts the definition (8.13) of β. $\square$

Theorem 8.14 (Hahn-Banach-Lagrange-Simons)**.** *Let C be a nonempty convex subset of some vector space F, $k : C \to]-\infty,\infty]$ be convex and $k \not\equiv \infty$, and $j : C \to E$ satisfy*

$$P(j(\alpha_1 y_1 + \alpha_2 y_2) - (\alpha_1 j(y_1) + \alpha_2 j(y_2))) \le 0 \qquad \textit{for} \quad y_1, y_2 \in C, 0 < \alpha_1, \alpha_2, \alpha_1 + \alpha_2 = 1. \tag{8.14}$$

Then there exists a linear functional L on E such that

$$L \le P \quad \textit{on} \quad E \quad \textit{and} \quad \inf_{y\in C}(L(j(y)) + k(y)) = \inf_{y\in C}(P(j(y)) + k(y)). \tag{8.15}$$

Proof. Define the sublinear functional $\hat{P} \to E \times \mathrm{I\!R}$ by

$$\hat{P}(x, \lambda) = P(x) + \lambda$$

and $D \subset E \times \mathrm{I\!R}$ by

$$D := \{(z, \lambda) \in E \times \mathrm{I\!R} : \exists\, y \in C \text{ such that } P(j(y) - z) \leq 0, k(y) \leq \lambda\}\,. \tag{8.16}$$

Note that

$$(j(\hat{y}), k(\hat{y})) \in D \quad \text{for } \hat{y} \in C \tag{8.17}$$

because we may choose $y = \hat{y}, \lambda = k(\hat{y})$ in (8.16). We show that D is convex; this is where (8.14) is needed. Suppose that $(z_j, \lambda_j) \in D$ for $j = 1, 2$ and $0 < \alpha_1, \alpha_2, \alpha_1 + \alpha_2 = 1$. We want to show that

$$(\alpha_1 z_1 + \alpha_2 z_2, \alpha_1 \lambda_1 + \alpha_2 \lambda_2) \in D. \tag{8.18}$$

By the definition of D, there exists $y_j \in C$ such that

$$P(j(y_j) - z_j) \leq 0,\ k(y_j) \leq \lambda_j,\ j = 1, 2. \tag{8.19}$$

Then, by the convexity of k,

$$k(\alpha_1 y_1 + \alpha_2 y_2) \leq \alpha_1 k(y_1) + \alpha_2 k(y_2) \leq \alpha_1 \lambda_1 + \alpha_2 \lambda_2.$$

Moreover, using (8.14) and then subadditivity and (8.19),

$$\begin{aligned}
&P(j(\alpha_1 y_1 + \alpha_2 y_2) - (\alpha_1 z_1 + \alpha_2 z_2) \\
&\quad \leq P(j(\alpha_1 y_1 + \alpha_2 y_2) - (\alpha_1 j(y_1) + \alpha_2 j(y_2))) + P(\alpha_1 j(y_1) + \alpha_2 j(y_2) - (\alpha_1 z_1 + \alpha_2 z_2)) \\
&\quad \leq P(\alpha_1 j(y_1) + \alpha_2 j(y_2) - (\alpha_1 z_1 + \alpha_2 z_2)) \\
&\quad \leq \alpha_1 P(j(y_1) - z_1) + \alpha_2 P(j(y_2) - z_2) \leq 0.
\end{aligned}$$

Hence D is convex.

Let $\hat{L}$ be a linear functional (provided by Theorem 8.1) which satisfies

$$\hat{L} \leq \hat{P} \quad \text{on} \quad E \times \mathrm{I\!R} \quad \text{and} \quad \inf_{(z,\lambda) \in D} \hat{L}(z, \lambda) = \inf_{(z,\lambda) \in D} (P(z) + \lambda).$$

Writing $\hat{L}(x, \lambda) = \langle x^*, x \rangle + \gamma\lambda$ for some $x^* \in E^*, \gamma \in \mathrm{I\!R}$, this translates into

$$\langle x^*, x \rangle + \gamma\lambda \leq P(x) + \lambda \quad (x \in E, \lambda \in \mathrm{I\!R}) \tag{8.20}$$

and

$$\inf_{(z,\lambda) \in D} (\langle x^*, z \rangle + \gamma\lambda) = \inf_{(z,\lambda) \in D} (P(z) + \lambda). \tag{8.21}$$

From (8.20) with $x = 0$, we have $\gamma = 1$ and $\langle x^*, x \rangle \leq P(x)$. Now, using (8.17),

$$\inf_{(z,\lambda) \in D} (\langle x^*, z \rangle + \lambda) \leq \langle x^*, j(y) \rangle + k(y) \quad \text{for} \quad y \in C. \tag{8.22}$$

Conversely, if $(z, \lambda) \in D$ and $y \in C$, $P(j(y) - z) \leq 0, k(y) \leq \lambda$, then $\langle x^*, j(y) - z \rangle \leq P(j(y) - z) \leq 0$, so if $(z, \lambda) \in D$,

$$\exists\, y \in C \quad \text{such that} \quad \langle x^*, j(y) \rangle + k(y) \leq \langle x^*, z \rangle + \lambda; \tag{8.23}$$

from these last two relations we see that

$$\inf_{(z,\lambda)\in D}(\langle x^*, z\rangle + \lambda) = \inf_{y\in C}(\langle x^*, j(y)\rangle + k(y)).$$

Just as in (8.22),

$$\inf_{(z,\lambda)\in D}(P(z)+\lambda) \le P(y)+k(y) \quad \text{for} \quad y\in C$$

and, this time using that $P(j(y)-z)\le 0$ implies

$$P(j(y)) \le P(z) + P(j(y)-z) \le P(z),$$

we have the analogue of (8.23): if $(z,\lambda)\in D$, then

$$\exists\, y\in C \quad \text{such that} \quad P(j(y)) + k(y) \le P(z)+\lambda.$$

This completes the proof. □

9. Decompositions of Measures

We will be taking much of this material from Rudin, Chapter 6.

In this section $(X,\mathcal{M})$ is a fixed measurable space. We consider the following variations of the notion of a "measure" on $\mathcal{M}$. Recall that a measure is a countably additive set function which takes values in $[0,\infty]$. To distinguish these among the other types to follow, we will use the term "nonnegative measure." If we want to insist on finite nonnegative values, we will say "nonnegative real measure." The following notions still ask for countable additivity, but the values are taken in other sets.

For convenience in discourse, we will use the term "partition of $E\in\mathcal{M}$" to mean a sequence of pairwise disjoint sets $\{E_j\}\subset\mathcal{M}$ such that $E=\cup_{j=1}^{\infty}E_j$.

Definition 9.1. (a) A function $\mu:\mathcal{M}\to\mathbb{R}$ such that

$$\mu(E) = \sum_{j=1}^{\infty}\mu(E_j) \quad \text{for every partition of } E\in\mathcal{M} \tag{9.1}$$

is called a *signed measure* on $\mathcal{M}$.

(b) A function $\mu:\mathcal{M}\to\mathbb{C}$ such that (9.1) holds is called a *complex measure* on $\mathcal{M}$.

A signed measure might have nonnegative values, but it cannot take the value ∞. Signed measures are complex measures, but complex measures can take complex values. If λ,ν are nonnegative real measures on $\mathcal{M}$ (again, the same thing as a plain old "measure" which does not take the value ∞), then

$$\mu(E) := \lambda(E)-\nu(E) \quad \text{defines a signed measure on } \mathcal{M}, \tag{9.2}$$

and if λ,ν are signed measures on $\mathcal{M}$, then

$$\mu(E) := \lambda(E)+i\nu(E) \quad \text{defines a complex measure on } \mathcal{M}. \tag{9.3}$$

Remark 9.2. In fact, (9.3) is general; the real and imaginary parts of a complex measure are clearly signed measures. However, the positive and negative parts of a signed measure are not, in general, measures, as is easy to see (Exercise 9.2). Dealing with signed or complex measures, we lose two things which we are very much used to: $0 \leq \mu(E)$ and $A \subset B \implies \mu(A) \leq \mu(B)$.

In another view, if ν is a plain old measure, and f is in $L^1(\nu)$, then

$$\mu(E) := \int_E f\, d\nu \quad \text{defines a complex measure on } \mathcal{M}. \tag{9.4}$$

If f is real-valued, $f = f^+ - f^-$, then

$$\mu(E) := \int_E f\, d\nu = \int_E f^+\, d\nu - \int_E f^-\, d\nu \tag{9.5}$$

writes μ as the difference of two nonnegative real measures. Moreover, taking $E^+ = \{f \geq 0\}$, $E^- = \{f < 0\}$, (9.5) tells us that

$$\mu(E) = \mu(E \cap E^+) - \mu(E \cap E^-). \tag{9.6}$$

The measures $E \mapsto \mu(E \cap E^+), \mu(E \cap E^-)$ are "concentrated" (defined precisely later) on the disjoint sets E^+, E^-.

We will see below that these examples are pretty general.

9.1. **An Overview of Main Results.** We outline the main results of interest and then prove some of them.

Theorem 9.3. *Let μ be a complex measure on $\mathcal{M}$. Then*

$$|\mu|(E) := \sup\left\{\sum_{j=1}^{\infty} |\mu(E_j)| : \{E_j\} \text{ is a partition of } E\right\} \tag{9.7}$$

defines a measure $|\mu|$ on $\mathcal{M}$ for which $|\mu|(E) \geq |\mu(E)|$ for $E \in \mathcal{M}$. Moreover, if λ is another measure for which $\lambda(E) \geq |\mu(E)|$, for $E \in \mathcal{M}$, then $\lambda \geq |\mu|$.

Proof. For $F \in \mathcal{M}$ let $P(F) = \{\text{partitions of } F\}$. We need to show that

$$\{F_j\} \in P(F) \implies |\mu|(F) = \sum_{j=1}^{\infty} |\mu|(F_j).$$

To this end, notice that $\{E_k\}_{k=1}^{\infty} \in P(F)$ implies $\{E_k \cap F_j\}_{k=1}^{\infty} \in P(F_j)$ for $j = 1, 2, \ldots$, and $\{E_k \cap F_j\}_{j,k=1}^{\infty} \in P(F)$. Moreover, if $\{E_{j,k}\}_{k=1}^{\infty} \in P(F_j)$, $j = 1, 2, \ldots$, then $\{E_{j,k}\}_{j,k=1}^{\infty} \in$

$P(F)$. Thus

$$\begin{aligned}
|\mu|(F) &= \sup\left\{\sum_{k=1}^{\infty}|\mu(E_k)| : \{E_k\}_{k=1}^{\infty} \in P(F)\right\} \\
&= \sup\left\{\sum_{k=1}^{\infty}|\mu(\cup_{j=1}^{\infty}(F_j \cap E_k))| : \{E_k\}_{k=1}^{\infty} \in P(F)\right\} \\
&\leq \sup\left\{\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}|\mu(F_j \cap E_k)| : \{E_k\}_{k=1}^{\infty} \in P(F)\right\} \\
&\leq \sup\left\{\sum_{j,k=1}^{\infty}|\mu(E_{j,k})| : \{E_{j,k}\}_{k=1}^{\infty} \in P(F_j), j = 1, 2, \ldots\right\} \\
&= \sup\left\{\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}|\mu(E_{j,k})| : \{E_{j,k}\}_{k=1}^{\infty} \in P(F_j), j = 1, 2, \ldots\right\} \\
&= \sum_{j=1}^{\infty}|\mu|(F_j).
\end{aligned}$$

The first inequality followed from $\mu(\cup_{j=1}^{\infty}(F_j \cap E_k)) = \sum_{j=1}^{\infty}\mu(F_j \cap E_k)$, the second inequality arises because one sup is over a larger set, and the final equality is due to the fact that the sum over $k's$ in the preceding expression is at most $|\mu|(F_j)$, but can be made arbitrarily close (your basic $\epsilon/2^j$ close) to $|\mu|(F_j)$. Finally, the fourth (equivalently, the fifth) expression on the right is at most $|\mu|(F)$ by definition of $|\mu|$, so all the inequalities are equalities.

Finally, $|\mu|$ is, by its very definition, the smallest measure λ for which $\lambda(E) \geq |\mu(E)|$ for $E \in \mathcal{M}$. Indeed, if λ is any such measure and $\{E_j\}_{j=1}^{\infty} \in P(E)$, then

$$\lambda(E) = \sum_{j=1}^{\infty}\lambda(E_j) \geq \sum_{j=1}^{\infty}|\mu(E_j)|$$

$\square$

Definition 9.4. The measure $|\mu|$ is called the *total variation measure* of μ.

Caution: The notation is a tad dangerous. We would usually use the notation $|f|$ for the function $x \mapsto |f(x)|$. The use of $|\mu|$ above is in conflict with this. $|\mu|$ is *not* the function $E \mapsto |\mu(E)|$.

Remark 9.5. Let μ be a complex measure on $(X, \mathcal{M})$. The natural question arises as to how

$$\int_X f\, d\mu$$

should be defined. It is a consequence of the Radon-Nikodym Theorem below that there is a measurable function h which satisfies $|h| = 1$ everywhere, such that

$$\mu(E) = \int_E h\, d|\mu| \quad (E \in \mathcal{M}). \tag{9.8}$$

Then one can define

$$\int_X f\, d\mu := \int_X fh\, d|\mu| \quad (fh \in L^1(|\mu|)). \tag{9.9}$$

This is what will be done; the notion so defined is "consistent" in that then

$$\int_X \chi_E\, d\mu = \int_X \chi_E h\, d|\mu| = \int_E h\, d|\mu| = \mu(E).$$

Theorem 9.6. *If μ is a complex measure on $(X, \mathcal{M})$, then $|\mu|(X) < \infty$.*

Remark 9.7. If you have a complex measure μ "in your hands," that is, you have a concrete description of how to compute its values, you should be able to quickly see that $|\mu|(X) < \infty$ without using any theorems. This theorem is of the sort that provides general information which is evident in particular cases. This doesn't make it useless, it does help to organize knowledge, but it isn't serious from the point of view of applications. There are lots of such math results.

Proof. Since $\mu = \lambda + i\nu$ where λ, ν are the real and imaginary parts of μ, and $|\mu| \le |\lambda| + |\nu|$ (because $|\mu(E)| \le |\lambda(E)| + |\nu(E)|$), it suffices to assume that μ is a signed measure. Suppose that $|\mu|(X) = \infty$. We claim that then there is a partition A, B of X such that

$$\mu(A) > 1, \ \mu(B) < -1. \tag{9.10}$$

Since then $|\mu|(X) = |\mu|(A) + |\mu|(B)$, at least one of $|\mu|(A), |\mu|(B)$ is infinity. Say $|\mu|(A) = \infty$; applying the claim again, A has a partition A_1, B_1 such that (9.10) holds with A_1, B_1 in place of A, B. Continuing in this manner, we would produce a sequence of pairwise disjoint sets $\{E_j\}$ such that $|\mu(E_j)| > 1$. This is a contradiction, for

$$\mu\left(\cup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \mu\left(E_j\right)$$

requires $|\mu(E_j)| \to 0$.

It remains to produce the splitting A, B of X satisfying (9.10). Since $|\mu|(X) = \infty$, there is a partition $\{F_j\}$ of X such that

$$\begin{aligned} \sum_{j=1}^{\infty} |\mu(F_j)| &= \sum_{\{j:\mu(F_j)\ge 0\}} \mu(F_j) - \sum_{\{j:\mu(F_j)<0\}} \mu(F_j) \\ &= \mu\left(\cup_{\{j:\mu(F_j)\ge 0\}} F_j\right) - \mu\left(\cup_{\{j:\mu(F_j)<0\}} F_j\right) > 2 + |\mu(X)|. \end{aligned} \tag{9.11}$$

Put $A = \cup_{\{j:\mu(F_j)\ge 0\}} F_j$, $B = \cup_{\{j:\mu(F_j)<0\}} F_j$ and $a = \mu(A), b = \mu(B)$. Since

$$\mu(X) = \mu(A) + \mu(B) = a + b$$

and (9.11) tells us $|a-b| = a-b \geq 2 + |\mu(X)|$, the result follows from

$$|a|, |b| \geq \frac{|a-b| - |a+b|}{2},$$

which is valid for $a, b \in \mathbb{R}$. $\square$

Definition 9.8. Let μ is a signed measure on $\mathcal{M}$ and $|\mu|$ be its total variation measure. Then the nonnegative measures

$$\mu^+ := \frac{1}{2}(|\mu| + \mu), \quad \mu^- := \frac{1}{2}(|\mu| - \mu), \tag{9.12}$$

are, respectively, the *positive and negative variations* of μ.

Remark 9.9. Note that $\mu = \mu^+ - \mu^-$, $|\mu| = \mu^+ + \mu^-$. In particular, any signed measure can be written as the difference of nonnegative measures, as in (9.2). This particular "decomposition" $\mu = \mu^+ - \mu^-$ into the difference of two measures is called the *Jordan decomposition* of μ.

Definition 9.10. Let μ be a nonnegative measure on $\mathcal{M}$ (so μ can take the value ∞), and λ be either a nonnegative or a complex measure on $\mathcal{M}$.

(a) If $\lambda(E) = 0$ for every μ-null set E (ie, $\mu(E) = 0$), then λ is *absolutely continuous* with respect to μ, and we write $\lambda \ll \mu$.
(b) If there is a set $A \in \mathcal{M}$ such that $\lambda(E) = \lambda(E \cap A)$ for $E \in \mathcal{M}$, we say that λ *is concentrated on* A.
(c) If λ_1, λ_2 are nonnegative or complex measures on $\mathcal{M}$, and there are disjoint sets $A, B \in \mathcal{M}$ such that λ_1 is concentrated on A and λ_2 is concentrated on B, then λ_1, λ_2 are *mutually singular*, and we write $\lambda_1 \perp \lambda_2$. This description is symmetrical, but one can say it this nonsymmetrical way: λ_1 is concentrated on a λ_2 null set.

Theorem 9.11 (Lebesgue-Radon-Nikodym)**.** *Let μ be a nonnegative σ-finite measure on $\mathcal{M}$ and let λ be a complex measure on $\mathcal{M}$.*

(a) *Then there is a unique pair of complex measures on $\mathcal{M}$ such that*

$$\lambda = \lambda_a + \lambda_s, \ \lambda_a \ll \mu, \ \lambda_s \perp \mu. \tag{9.13}$$

If λ is nonnegative and finite, then so are λ_a and λ_b.

(b) *There is a unique $h \in L^1(\mu)$ such that*

$$\lambda_a(E) = \int_E h \, d\mu \quad (E \in \mathcal{M}). \tag{9.14}$$

Definition 9.12. The pair (λ_a, λ_s) of (9.13) is called the *Lebesgue decomposition of λ wrt μ.*

Definition 9.13. Suppose the assumptions of Theorem 9.11 are satisfied and, in addition, $\lambda \ll \mu$. Then (clearly) $\lambda_s = 0$ and (9.14) yields

$$\lambda(E) = \int_E h \, d\mu \quad (E \in \mathcal{M}). \tag{9.15}$$

Then $h \in L^1(\mu)$ is called the *Radon-Nikodym derivative of* λ *wrt* μ. This is denoted by $d\lambda = h d\mu$ or

$$h = \frac{d\,\mu}{d\,\lambda}.$$

This result, which is here built into Theorem 9.11 is called the *Radon-Niodym* theorem.

Remark 9.14. Since it is clear that $\lambda \ll |\lambda|$, then Theorem 9.11 (b) above shows that a general complex measure can be written in the form (9.4) (since $|\lambda|(X) < \infty$ by Theorem 9.6).

Theorem 9.15 (Hahn Decomposition)**.** *Let* μ *be a signed measure on* $\mathcal{M}$. *Then there exist disjoint sets* $A, B \in \mathcal{M}$ *such that* $A \cup B = X$ *and*

$$\mu^+(E) = \mu(E \cap A) \quad \textit{and} \quad \mu^-(E) = -\mu(E \cap B) \quad (E \in \mathcal{M}). \tag{9.16}$$

Thus it is possible to write a general real measure as the difference of nonnegative real measures concentrated on disjoint sets, as in (9.5).

9.2. **Proof of the Lebesgue-Radon-Nikodym Theorem.** This proof is due to von Neumann, and it is therefore so blindingly brilliant and unlike anything in our previous experiece that we will sit in wonder at the end, not quite knowing what happened. However, it is not painful and can easily be learned.

We begin with the lemma:

Lemma 9.16. *Let* μ *be a nonnegative* σ*-finite measure on* $\mathcal{M}$. *Then there is a function* $w \in L^1(\mu)$ *which satisfies* $0 < w(x) < 1$ *for every* $x \in X$.

Proof. If $\mu(X) < \infty$, put $w \equiv 1/2$. Otherwise, let $X = \cup_{n=1}^{\infty} E_n$ where $0 < \mu(E_n) < \infty$. Put $w_n(x) = 0$ if $x \in X \setminus E_n$ and

$$w_n(x) = \frac{2^{-n}}{1 + \mu(E_n)} \quad \text{if} \quad x \in E_n$$

and $w(x) = \sum_{n=1}^{\infty} w_n(x)$. Clearly $0 \le w_n < 2^{-n}$, so $0 < w < 1$. Moreover,

$$\int_X w\, d\mu \le \sum_{n=1}^{\infty} \int_{E_n} w\, d\mu = \sum_{n=1}^{\infty} 2^{-n} \frac{\mu(E_n)}{1 + \mu(E_n)} < 1.$$

□

First assume that λ in Theorem 9.11 is a nonnegative real measure (and hence finite, by Theorem 9.6). Associate w to μ as in Lemma 9.16, and let $d\phi = d\lambda + w d\mu$, which just means that ϕ is the measure

$$\phi(E) = \lambda(E) + \int_E w\, d\mu.$$

In particular,

$$\int_X \chi_E\, d\phi = \int_X \chi_E\, d\lambda + \int_X \chi_E w\, d\mu,$$

and then

$$\int_X f\,d\phi = \int_X f\,d\lambda + \int_X fw\,d\mu$$

holds for simple functions, and then for nonnegative functions and then for $f \in L^1(\phi)$. Now consider the linear functional

$$L^2(\phi) \ni f \mapsto \int_X f\,d\lambda.$$

We have

$$\left|\int_X f\,d\lambda\right| \le \int_X |f|\,d\lambda \le \int_X |f|\,d\phi \le \left(\int_X |f|^2\,d\phi\right)^{1/2} \phi(X)^{1/2}$$

(Exercise 11). We are implicitly using

$$\phi(X) = \lambda(X) + \int_X w\,d\mu < \infty.$$

This shows our linear functional is continuous on the Hilbert space $L^2(\phi)$, so by Corollary 12.8 (Riesz Representation Theorem), there exists $g \in L^2(\phi)$ such that

$$\int_X f\,d\lambda = \int_X fg\,d\phi \quad (f \in L^2(\phi)). \tag{9.17}$$

Put $f = \chi_E$ in (9.17) and use $0 \le \lambda \le \phi$ to find

$$0 \le \lambda(E) = \int_E g\,d\phi \le \phi(E) \implies 0 \le \frac{1}{\phi(E)}\int_E g\,d\phi = \frac{\lambda(E)}{\phi(E)} \le 1. \tag{9.18}$$

We use one more lemma, which we also need again later, to conclude from (9.18) that $0 \le g \le 1$ ae. (Rudin Lemma 1.40).

Lemma 9.17. *Let μ be a nonnegative real measure, $\mu(X) < \infty$, $f \in L^1(\mu)$, and S be a closed convex subset of $\mathbb{C}$. If*

$$\frac{1}{\mu(E)}\int_E f\,d\mu \in S \quad (E \in \mathcal{M},\ \mu(E) > 0) \tag{9.19}$$

then $f(x) \in S$ for almost all $x \in X$.

Proof. $S^c = \mathbb{C} \setminus S$ is the union of countably many closed circular disks. Let $\Delta \subset S^c$ be such a disk; it suffices to prove that $E := f^{-1}(\Delta)$ is a μ null set. Let α be the center of Δ and $r > 0$ be its radius. If $\mu(E) > 0$, then

$$\left|\frac{1}{\mu(E)}\int_E f\,d\mu - \alpha\right| = \left|\frac{1}{\mu(E)}\int_E (f-\alpha)\,d\mu\right| \le \frac{1}{\mu(E)}\int_E |f-\alpha|\,d\mu \le r.$$

However, this contradicts (9.19). □

Hence $0 \le g \le 1$ ae wrt ϕ. Hence we may assume that $0 \le g \le 1$ everywhere. Rewrite (9.17) in the form

$$\int_X (1-g)f\,d\lambda = \int_X fgw\,d\mu. \tag{9.20}$$

Put

$$A = \{x : 0 \le g(x) < 1\}, \quad B = \{x : g(x) = 1\}, \tag{9.21}$$

and define

$$\lambda_a(E) = \lambda(A \cap E),\ \lambda_s(E) = \lambda(B \cap E) \quad (E \in \mathcal{M}). \tag{9.22}$$

Clearly λ_s is concentrated on B. On the other hand, putting $f = \chi_B$ in (9.20), using that $g = 1$ on B and $0 < w$ everywhere, yields

$$\int_B w\, d\mu = 0 \implies \mu(B) = 0.$$

Thus μ is concentrated on $X \setminus B = A$, and so, by definition, $\lambda_s \perp \mu$.

Next we choose

$$f = (1 + g + \cdots + g^n)\chi_E \tag{9.23}$$

in (9.20) to find

$$\int_E (1 - g^{n+1})\, d\lambda = \int_E g(1 + g + \cdots + g^n) w\, d\mu. \tag{9.24}$$

Clearly

$$g(1 + g + \cdots + g^n)w \uparrow h := \begin{cases} \infty & \text{on} \quad \{g = 1\} \\ \dfrac{g}{1-g} w & \text{on} \quad \{g < 1\} \end{cases},$$

while $1 - g^{n+1} \uparrow \chi_A$ as $n \to \infty$. Using the monotone convergence theorem in (9.24) therefore yields

$$\lambda_a(E) = \lambda(E \cap A) = \int_E h\, d\mu \quad (E \in \mathcal{M}).$$

Taking $E = X$, we find $h \in L^1(\mu)$ and then deduce that $\mu(E) = 0$ implies that $\lambda_a(E) = 0$, or $\lambda_a \ll \mu$, as desired. This completes the proof for nonnegative real measures λ, up to the claims of uniqueness. Clearly h satisfying (9.14) is unique a.e. μ, (given λ_a.) The uniqueness of λ_a, λ_s is argued below.

If $\lambda = \lambda_1 + i\lambda_2$ with real λ_j, we use the above result on the positive and negative variations of λ_1, λ_2. To glue the results of this together (first to obtain results about λ_j, then to obtain the desired conclusions about λ), which involves noting:

$$\nu_1 \perp \mu \quad \text{and} \quad \nu_2 \perp \mu \implies \nu_1 + \nu_2 \perp \mu,$$
$$\nu_1 \ll \mu \quad \text{and} \quad \nu_2 \ll \mu \implies \nu_1 + \nu_2 \ll \mu.$$

The second claim is trivial. For the first claim, we note that if ν_1, ν_2 are concentrated on A_1, A_2 respectively, and μ is concentrated on B_1 and B_2, where $A_j \cap B_j = \emptyset$, then $\nu_1 + \nu_2$ is concentrated on $A_1 \cup A_2$ and μ is concentrated on $B_1 \cap B_2$.

As to the uniqueness, let $\lambda_a, \hat{\lambda}_a \ll \mu$, $\lambda_s, \hat{\lambda}_s \perp \mu$, and

$$\lambda_s + \lambda_a = \hat{\lambda}_s + \hat{\lambda}_a = \mu,$$

then, by the above remark,

$$\mu \perp (\lambda_a - \hat{\lambda}_a) = (\hat{\lambda}_s - \lambda_s) \ll \mu.$$

But $\nu \perp \mu$ and $\nu \ll \mu$ clearly implies $\nu = 0$. □

9.3. **Consequences of the Radon-Nikodym Theorem.**

Corollary 9.18. *Let λ, μ be nonnegative σ-finite measures on $\mathcal{M}$. Then $\lambda = \lambda_s + \lambda_a$ for unique measures $\lambda_a \perp \mu$ and $\lambda_s \ll \mu$. Moreover, there exists a unique (a.e. μ) h, $0 \le h$, such that*

$$\lambda_a(E) = \int_E h \, d\mu \quad (E \in \mathcal{M}).$$

Proof. Let $X_1 \subset X_2 \ldots \subset X$, $X = \cup_{n=1}^{\infty} X_n$, and $\lambda(X_n) < \infty$ for $n = 1, 2, \ldots$. Define $E_1 = X_1$, $E_2 = X_2 \setminus X_1$, $E_3 = X_3 \setminus (X_1 \cup X_2)$, etc, constructing a pairwise disjoint partition of X on which λ is finite. Define $\lambda^n(E) := \lambda(E \cap E_n)$ and λ_a^n, λ_s^n, $h_n \in L^1(\mu)$ be the decomposition data for λ^n wrt μ as in the Radon-Nikodym Theorem. Note that we may assume $h_n = 0$ on $X \setminus E_n$. Then it is straightforward to show that

$$\lambda_a := \sum_{n=1}^{\infty} \lambda_a^n, \;\; \lambda_s := \sum_{n=1}^{\infty} \lambda_s^n, \;\; h := \sum_{n=1}^{\infty} h_n \tag{9.25}$$

have the desired properties. Finally, note that h is not necessarily in $L^1(\mu)$, but

$$\int_E h \, d\mu = \lambda_a(E) \le \lambda(E)$$

is finite iff $\lambda(E) < \infty$. □

Remark 9.19. If σ finiteness is dropped, then the Lebesgue - Radon-Nikodym Theorem fails. Indeed, let λ be counting measure on $(0,1)$ and μ be Lebesgue measure and $\mathcal{M}$ consists of the Lebesgue measurable subsets of $(0,1)$. Then λ has no Lebesgue decomposition relative to μ and even though $\mu \ll \lambda$, there is no $h \in L^1(\lambda)$ such that $\mu = h\lambda$. (exercise)

Theorem 9.20. *Let μ be a complex measure on $\mathcal{M}$. Then there is a measurable function h such that $|h(x)| = 1$ for all $x \in X$, and $d\mu = h d|\mu|$.*

Theorem 9.21. *Let μ be a positive measure on $\mathcal{M}$, $g \in L^1(\mu)$, and*

$$\lambda(E) := \int_E g \, d\mu \quad (E \in \mathcal{M}).$$

Then $d|\lambda| = |g| d\mu$.

The Hahn Decomposition (Theorem 9.15).

Corollary 9.22 (to the Hahn Decomposition)**.** *If μ is a signed measure and $\mu = \lambda_1 - \lambda_2$, where λ_1, λ_2 are nonnegative measures, then $\lambda_1 \ge \mu^+, \lambda_2 \ge \mu^-$.*

9.3.1. *The Continuous Linear Functionals on L^p.* This result deserves its own subsection. Throughout this subsection, μ is a σ-finite nonnegative measure on $\mathcal{M}$. We seek to represent the bounded linear functionals $\Phi : L^p(\mu) \to \mathcal{F}$ for $1 \le p < \infty$. We know some of them. If $g \in L^q(\mu)$ where q is the Hölder conjugate of p, then

$$\Phi_g(f) := \int_X fg\, d\mu. \tag{9.26}$$

is well defined by virtue of the Hölder inequality, and

$$|\Phi_g(f)| \le \|f\|_p \|g\|_q, \tag{9.27}$$

so $\|\Phi_g\| \le \|g\|_q$. In fact, all continuous linear functionals on $L^p(\mu)$ may be so represented.

Theorem 9.23. *Let $1 \le p < \infty$. If $\Phi \in L^p(\mu)^*$, $1 \le p < \infty$, then there is a unique $g \in L^q(\mu)$ such that $\Phi = \Phi_g$. Moreover, $\|\Phi_g\| = \|g\|_q$. That is, the mapping $g \mapsto \Phi_g$ is a linear isometry of $L^q(\mu)$ onto $L^p(\mu)^*$.*

Proof. We first show that $g, g' \in L^q(\mu)$ and $\Phi_g = \Phi_{g'}$ implies $g = g'$ a.e. If $\mu(E) < \infty$, then $\chi_E \in L^p(\mu)$, we have, by assumption,

$$\Phi_g(\chi_E) = \Phi_{g'}(\chi_E) \implies \int_E (g - g')\, d\mu = 0.$$

According to Exercise 34, this guarantees that $g = g'$ ae. Thus $g \mapsto \Phi_g$ is 1-1.

We attempt now to define a measure associated with $\Phi \in L^p(\mu)^*$; then the Radon Nikodym Theorem will be used to obtain the g for which $\Phi = \Phi_g$. We initially assume that $\mu(X) < \infty$. Define λ on $\mathcal{M}$ by

$$\lambda(E) := \Phi(\chi_E). \tag{9.28}$$

If E has a partition $\{E_j\}$, then we observe that

$$\|\chi_E - \sum_{j=1}^n \chi_{E_j}\|_{L^p(\mu)} = \|\chi_{E \setminus \cup_{j=1}^n \chi_{E_j}}\|_{L^p(\mu)} = \mu(E \setminus \cup_{j=1}^n E_j))^{1/p} \to 0 \tag{9.29}$$

as $n \to \infty$. (Here is a place that $p < \infty$ was used). On the other hand, this implies that

$$\lambda(E) - \sum_{j=1}^n \lambda(E_j) = \Phi(\chi_E) - \Phi(\sum_{j=1}^n \chi_{E_j}) \to 0,$$

as Φ is continuous on $L^p(\mu)$. Thus

$$\lambda(E) = \sum_{j=1}^\infty \lambda(E_j), \tag{9.30}$$

and λ is a measure. Clearly $\mu(E) = 0 \implies \lambda(E) = 0$. Thus by the Radon-Nikodym theorem, there is a $g \in L^1(\mu)$ such that

$$\lambda(E) = \Phi(\chi_E) = \int_E g\, d\mu \quad (E \in \mathcal{M}). \tag{9.31}$$

By linearity it follows that

$$\Phi(f) = \int_E fg\,d\mu \tag{9.32}$$

for simple functions f. If $f \in L^\infty(\mu)$, then it is the uniform limit of simple functions f_j, and then $\|f_j - f\|_{L^p(\mu)} \to 0$ (here $\mu(X) < \infty$ is used). Thus, by continuity of Φ, (9.32) holds for $f \in L^\infty(\mu)$.

If $p = 1$, (9.31) shows that

$$\left|\int_E g\,d\mu\right| \le \|\Phi\|\|\chi_E\|_1 = \|\Phi\|\mu(E).$$

By Lemma (9.17), it follows that $|g| \le \|\Phi\|$ ae.

For $1 < p < \infty$, we first define

$$\beta(z) = \begin{cases} \dfrac{\bar{z}}{|z|} & \text{if} \quad z \neq 0 \\ 1 & \text{if} \quad z = 0. \end{cases},$$

so that $\beta(z)z = |z|$. Since β is Borel measurable, $x \mapsto \beta(f(x))$ is measurable and $\alpha(x) = \beta(g(x))$ satisfies $\alpha g = |g|$ and $|\alpha| = 1$ everywhere. Put $E_n = \{|g| \le n\}$ and define $f = \chi_{E_n}|g|^{q-1}\alpha$. Then $|f|^p = |g|^q$ on E_n and $f \in L^\infty(\mu)$. Applying (9.32) to this choice of f, we find

$$\int_{E_n} |g|^q\,d\mu = \int_X fg\,d\mu = \Phi(f) \le \|\Phi\| \left(\int_{E_n} |g|^q\,d\mu\right)^{1/p}.$$

Thus

$$\int_X \chi_{E_n}|g|^q\,d\mu \le \|\Phi\|^q.$$

Using the monotone convergence theorem, we pass to the limit to find $\|g\|_q \le \|\Phi\|$. Thus Φ_g and Φ are both continuous and agree on the dense subspace $L^\infty(\mu)$ of $L^p(\mu)$, and therefore $\Phi = \Phi_g$. Moreover, from the above and (9.27), $\|\Phi_g\| = \|g\|_q$. This completes the proof for the case $\mu(X) < \infty$.

To treat the case $X = \cup_{j=1}^\infty E_j$, where $E_1 \subset E_2, \dots$ and $\mu(E_j) < \infty$, define μ_k to be μ restricted to the sigma algebra of those measurable sets which are subsets of E_k; μ_k is a finite measure. For $f \in L^p(\mu_k)$, define f_k to be the extension of f to X which vanishes off of E_k. The functional $L^p(\mu_k) \ni f \to \Phi(f_k)$, call it Φ^k, is linear and $\|\Phi^k\| \le \|\Phi\|$. By what we already proved, there exists $g_k \in L^q(\mu)$, $\|g^k\|_q \le \|\Phi\|$, such that

$$\Phi^k(f) = \int_{E_k} fg_k\,d\mu_k \quad (f \in L^p(\mu_k)).$$

By the uniqueness of g's, and $E_j \subset E_k$ if $j \le k$, it follows that $g_j = g_k$ a.e. on E_j. Hence defining g on X by

$$g = g_k \quad \text{on} \quad E_k$$

is perfectly sensible. Since for $1 < p < \infty$ (you do the $p = 1$ case)

$$\int_X |g|^q \, d\mu = \lim_{k\to\infty} \int_{E_k} |g|^q \, d\mu = \lim_{k\to\infty} \int_{E_k} |g_k|^q \, d\mu_k \leq \|\Phi\|^q,$$

$g \in L^q(\mu)$ and $\Phi(f) = \Phi_g(f)$ if f vanishes on E_k. By density ($\chi_{E_k} f \to f$ in $L^p(\mu)$), $\Phi = \Phi_g$, and we are done in general. □

10. Differentiation of Measures, Integrals and Functions

Much of this material is taken from Rudin, Chapter 7.

10.1. Notation, Definitions and a Few Preliminaries.

In this section we work in $\mathbb{R}^n$. We will use the notation

$$B\,(x, r) := \{y : |y - x| < r\}$$

to denote the open ball of center x and radius r. We will also use the notation

$$|E| := \mathcal{L}^n(E) \quad \text{for the Lebesgue measure of } E \subset \mathbb{R}^n.$$

The following quantities will appear, involving a (usually complex) Borel measure μ on $\mathbb{R}^n$:

$$(Q_r\mu)(x) := \frac{\mu(B\,(x,r))}{|B\,(x,r)\,|} = \frac{\mu(B\,(x,r))}{\omega_n r^n}, \tag{10.1}$$

where ω_n is the area of the unit sphere in $\mathbb{R}^n$,

$$(D\mu)(x) := \lim_{r\downarrow 0} (Q_r\mu)(x) \quad \text{(for } x \text{ such that the limit exists)}, \tag{10.2}$$

$$(M\mu)(x) = \sup_{0<r<\infty} (Q_r\mu)(x) \quad (\mu \geq 0). \tag{10.3}$$

Definition 10.1. At points x where $D\mu(x)$ is defined, $D\mu(x)$ is called the *symmetric derivative* of μ at x. The function $M\mu$ is called the *maximal function* of μ. It can take the value ∞.

In the case where μ is given by

$$\mu(E) = \int_E |f(x)| \, dx$$

for some $f \in L^1\left(\mathbb{R}^n\right)$, then we write Mf instead of $M\mu$. That is

$$(Mf)(x) := \sup_{0<r<\infty} \frac{1}{\omega_n r^n} \int_{B(x,r)} |f(y)| \, dy. \tag{10.4}$$

In this case, we call Mf the *maximal function of* f.

Definition 10.2. A function $h : \mathbb{R}^n \to [-\infty, \infty]$ is *lower semicontinuous* if any, and hence all, of the following equivalent conditions are satisfied:

$$\begin{aligned} &\text{(i) } \{x : f(x) > \lambda\} \quad \text{is open for} \quad \lambda \in (-\infty, \infty]. \\ &\text{(ii) } \{x : f(x) \leq \lambda\} \quad \text{is closed for} \quad \lambda \in (-\infty, \infty]. \\ &\text{(iii) if } x^k \to x, \text{ then } \liminf_{k\to\infty} f(x^k) \geq f(x). \end{aligned} \tag{10.5}$$

Lemma 10.3. *Let μ be a nonnegative Borel measure on $\mathbb{R}^n$. Then $M\mu$ is lower semicontinuous; thus it is measurable.*

Proof. There are two main points. The first is that Q_r is lower-semicontinuous for fixed $r > 0$. To see this, let $x^k \to x$. Then

$$B(x, r) \subset \cup_{j=1}^{\infty} \cap_{k=j}^{\infty} B(x^k, r), \tag{10.6}$$

which implies that

$$\mu(B(x, r)) = \lim_{j\to\infty} \mu(\cap_{k=j}^{\infty} B(x^k, r)) \leq \liminf_{j\to\infty} \mu(B\left(x^j, r\right)).$$

Since $|B(x, r)|$ is a multiple of r^n, independent of x, Q_r is lower-semicontinuous. The next main point is that the sup of any family $\{f_\alpha, \alpha \in \mathcal{A}\}$ of lower semicontinuous functions is again lower semicontinuous. To see this, we note that

$$f(x) := \sup_{\alpha\in\mathcal{A}} f_\alpha(x)$$

implies

$$\{x : f(x) > \lambda\} = \cup_{\alpha\in\mathcal{A}} \{x : f_\alpha(x) > \lambda\},$$

which is the union of open sets. □

10.2. **Maximal Functions and Lebesgue Points.** The most important result in this section is:

Theorem 10.4. *Let $f \in L^1(\mathbb{R}^n)$. Then for almost all $x \in \mathbb{R}^n$*

$$\lim_{r\downarrow 0} \frac{1}{r^n} \int_{B(x,r)} |f(y) - f(x)|\, dy = 0. \tag{10.7}$$

Remark 10.5. The unmodified expression "almost all" or the form "a.e." of this will always mean relative to Lebesgue measure in this section.

Definition 10.6. If (10.7) holds for x, then x is called a *Lebesgue point* of f.

In the course of proving Theorem 10.4 we will use

Theorem 10.7. *If μ is a nonnegative real Borel measure on $\mathbb{R}^n$, and $\lambda > 0$, then*

$$|\{M\mu > \lambda\}| \leq 3^n \frac{\mu(\mathbb{R}^n)}{\lambda}. \tag{10.8}$$

When $\mu(E) = \int_E |f|\,dx$, (10.7) becomes

$$|\{Mf > \lambda\}| \le 3^n \frac{\|f\|_{L^1(\mathbb{R}^n)}}{\lambda}. \tag{10.9}$$

Note that if $g \in L^1(\mathbb{R}^n)$, then

$$|\{|g| > \lambda\}|\lambda \le \int_{\mathbb{R}^n} |g|\,dx = \|g\|_{L^1(\mathbb{R}^n)}; \tag{10.10}$$

that is, $|g|$ obeys the same sort of inequality as Mf does in (10.9). However, this sort of inequality is not equivalent to being in L^1, and it is therefore referred to as follows: if

$$\lambda|\{|g| > \lambda\}| \quad \text{is bounded for} \quad \lambda > 0,$$

then *"g is in weak L^1."* The function $x \mapsto 1/|x|$ for $\mathbb{R} \ni x \ne 0$ is an example of a weak L^1 function which is not L^1.

The proof of Theorem 10.7 is not more difficult than that of the special case $\mu(E) = \int_E |f|\,dx$. The key ingredient is a *covering lemma*.

Lemma 10.8. *Let $B(x_j, r_j)$, $j = 1, 2, \ldots, N$ be a finite collection of balls in $\mathbb{R}^n$. Let*

$$W := \cup_{j=1}^N B(x_j, r_j).$$

Then there is a set $S \subset \{1, \ldots, N\}$ such that

(a) $B(x_j, r_j) \cap B(x_k, r_k) = \emptyset$ *for $j, k \in S$, $j \ne k$.*
(b) $W \subset \cup_{j \in S} B(x_j, 3r_j)$.

Proof. Relabel the balls so that $r_N \le r_{N-1} \le \ldots \le r_1$. Set $l_1 = 1$. Discard $B(x_j, r_j)$ if it intersects $B(x_1, r_1)$, and let j be the first index $1 < j$ for which $B(x_j, r_j)$ does not meet $B(x_1, r_1)$. Put $l_2 = j$. Repeat this process, discarding all remaining balls which meet $B(x_{l_2}, r_{l_2})$, and let l_3 be the least index from the set of non cast out balls for which $B(x_{l_3}, r_{l_3})$ meets $B(x_{l_2}, r_{l_2})$, etc. The process terminates in some number $m \le N$ steps. We claim $S = \{l_1, \ldots, l_m\}$ has the desired properties. Property (a) is clear. Property (b) is also clear, for any ball which was cast out meets a ball we kept, and which has at least as large a radius. But if two balls meet, then the one with the smallest radius is contained in the ball with thrice the radius and same center as the other ball (in other words, draw a pic). $\square$

Proof of Theorem 10.7. Let K be a compact subset of the open set $\{M\mu > \lambda\}$. Each $x \in K$ is the center of an open ball B for which

$$\mu(B) \ge \lambda|B|.$$

Select a finite subcover of K by such balls, and apply the preceding lemma to extract a disjoint subcollection $\{B_1, \ldots, B_m\}$, for which the balls with the same centers and thrice the radius cover K. Then

$$|K| \le 3^n \sum_{j=1}^m |B_j| \le 3^n \lambda^{-1} \mu(B_j) \le 3^n \frac{\mu(\mathbb{R}^n)}{\lambda}.$$

The result follows upon supping over the compact sets K. $\square$

Proof of Theorem 10.4. For $f \in L^1(\mathbb{R}^n)$ and $r > 0$, define the function $T_r f$ by

$$(T_r f)(x) := \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|\, dy = \frac{1}{\omega_n r^n} \int_{B(x,r)} |f(y) - f(x)|\, dy, \tag{10.11}$$

and then the function Tf by

$$(Tf)(x) := \limsup_{r \downarrow 0} (T_r f)(x). \tag{10.12}$$

We seek to show that $Tf = 0$ a.e. Notice that

$$Tf \le Mf + |f|. \tag{10.13}$$

Next, notice that T_r, and hence T, is subadditive:

$$T_r(f+g) \le T_r f + T_r g \implies T(f+g) \le Tf + Tg. \tag{10.14}$$

so that

$$Tf = T(f - g + g) \le T(f-g) + Tg.$$

Finally, we notice that $Tg = 0$ if g is continuous. Combining this with (10.13), we conclude that if $g \in L^1(\mathbb{R}^n)$ is continuous, then, by the above,

$$Tf \le M(f-g) + |f-g|. \tag{10.15}$$

If $(Tf)(x) > \lambda$, then it must be that at least one of $M(f-g)$ and $|f-g|$ is more than $\lambda/2$; that is

$$\{Tf > \lambda\} \subset \{M(f-g) > \lambda/2\} \cup \{|f-g| > \lambda/2\}.$$

Furthermore, by Theorem 10.7 and (10.10),

$$|\{M(f-g) > \lambda/2\}| \le 3^n 2 \frac{\|f-g\|_{L^1(\mathbb{R}^n)}}{\lambda}, \quad \{|f-g| > \lambda/2\}| \le 2 \frac{\|f-g\|_{L^1(\mathbb{R}^n)}}{\lambda}$$

Hence $\{Tf > \lambda\}$ is contained in a set of measure at most

$$2(3^n + 1) \frac{\|f-g\|_{L^1(\mathbb{R}^n)}}{\lambda},$$

where $g \in C(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ is arbitrary. Since the continuous functions are dense in $L^1(\mathbb{R}^n)$, $\{Tf > \lambda\}$ is contained in a set of measure 0, and, Lebesgue measure being complete, it is then a measurable set of measure 0, and we are done. $\square$

10.2.1. *Some Consequences of Theorem 10.4.* We'll say more in class, but won't write much here ("too simple" :)).

(1) If $f \in L^1(\mathbb{R}^n)$ and x is a Lebesgue point of f and $\{E_j\}$ is a sequence of sets and $\{r_j\}$ is a sequence of positive numbers convergent to 0, and $E_j \subset B(x, r_j)$, then

$$\frac{1}{|E_j|}\int_{E_j}|f(y)-f(x)|\,dy \le \frac{r_j^n}{|E_j|}\frac{1}{r_j^n}\int_{B(x,r_j)}|f(y)-f(x)|\,dy \to 0$$

provided that the ratio $r_j^n/|E_j|$ is bounded above; equivalently, $|E_j| \ge \kappa|B(x,r_j)|$ for some $\kappa > 0$. This is the definition of $\{E_j\}$ *shrinks nicely to* x. Note that we do not even need $x \in E_j$. Of course, this implies

$$\frac{1}{|E_j|}\int_{E_j} f(y)\,dy \to f(x).$$

(2) Let $f \in L^1(0,1)$ and extend it as 0 to the rest of $\mathbb{R}$. Consider

$$\frac{1}{h}\left(\int_0^{x+h} f(y)\,dx - \int_0^x f(y)\,dy\right) = \frac{1}{|h|}\int_{[x,x+h]} f(y)\,dy$$

for $x \in (0,1)$ and $|h| > 0$. As the intervals $[x, x+h]$ shrink nicely to x, we conclude that $F(x) := \int_0^x f(y)\,dy$ is differentiable a.e. on (0,1) (in the calculus sense) and $F'(x) = f(x)$ a.e.

(3) If E is any set, we have

$$\frac{|E \cap B(x,r)|}{|B(x,r)|} = \frac{1}{|B(x,r)|}\int_{B(x,r)} \chi_E\,dy \to \chi_E(x)$$

for almost every x. It follows that the *metric density of* E (namely, the limit of the left-hand side as $r \downarrow 0$) exist almost everywhere, is 1 a.e. on E and is 0 a.e. on the complement of E.

11. The Fundamental Theorem of Calculus, Lebesgue Style

We are out to identify the functions $f : [0,1] \to \mathbb{R}$ for which f' exists ae, belongs to $L^1(0,1)$, and

$$f(x) - f(0) = \int_0^x f'(s)\,ds \quad (0 \le x \le 1), \tag{11.1}$$

where the integral is understood in the Lebesgue sense. To minimize notation, we will work on [0,1] rather than "$[a,b]$", as does Rudin, but if you understand [0,1], you can easily generalize to $[a,b]$. To start, suppose that $g \in L(0,1)$ and

$$f(x) = f(0) + \int_0^x g(s)\,ds \quad (0 \le x \le 1). \tag{11.2}$$

What properties does f have? Well, we know by Section 10.2.1, point (2), that f' exists a.e. and $f' = g$ a.e. This much information about f, that f' exists a.e. and is in $L^1(0,1)$, does not guarantee that (11.1) holds. The key property of those functions f of the form (11.2) is this: suppose that $(a_j, b_j) \subset [0,1]$ for $j = 1, \ldots, m$ is a finite collection of pairwise disjoint intervals. Then

$$\sum_{j=1}^{m} |f(b_j) - f(a_j)| = \sum_{j=1}^{m} \Big| \int_{a_j}^{b_j} g(s)\, ds \Big| \le \sum_{j=1}^{m} \int_{a_j}^{b_j} |g(s)|\, ds. \tag{11.3}$$

We claim that for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$\sum_{j=1}^{m} (b_j - a_j) < \delta \implies \sum_{j=1}^{m} \int_{a_j}^{b_j} |g(s)|\, ds < \epsilon. \tag{11.4}$$

This is a special case of the following general remark.

Lemma 11.1. *Let μ be a nonnegative measure on a σ-algebra $\mathcal{M}$ and $g \in L^1(\mu)$. Then for every $\epsilon > 0$ there is a $\delta > 0$ such that if $\mu(E) < \delta$, then $\int_E |g|\, d\mu < \epsilon$.*

Proof. If the assertion of the theorem fails for some $\epsilon > 0$, then there exist sets E_j such that $\mu(E_j) < 1/2^j$, but

$$\int_{E_j} |g|\, d\mu > \epsilon$$

Let

$$F_k = \cup_{j=k}^{\infty} E_j; \quad \text{then} \quad \mu(F_k) \le \frac{1}{2^{k-1}}.$$

The F_k decrease with k and, by the above, we would have that $F := \cap_{k=1}^{\infty} F_k$ is a μ-null set for which

$$\int_F |g|\, d\mu = \lim_{k\to\infty} \int_{F_k} |g|\, d\mu \ge \limsup_{k\to\infty} \int_{E_k} |g|\, d\mu \ge \epsilon.$$

This is a contradiction. □

Thus (11.4) indeed holds, and by (11.3) we then know that if f is given in the form (11.2), it must be AC (short for absolutely continuous) in the sense of the definition below.

Definition 11.2. A function $f : [0,1] \to \mathbb{C}$ is *absolutely continuous* (AC for short) if for every $\epsilon > 0$ there exists a $\delta > 0$ such that if $(a_j, b_j) \subset [0,1]$, $j = 1, \ldots, m$, is a finite collection of pairwise disjoint intervals and $\sum_{j=1}^{m} (b_j - a_j) < \delta$, then $\sum_{j=1}^{m} |f(b_j) - f(a_j)| \le \epsilon$.

The hard part of the Fundamental Theorem of Calculus, Lebesgue style, is:

Theorem 11.3. *If $f : [0,1] \to \mathbb{C}$ is AC, then f is differentiable at almost all $x \in [0,1]$, $f' \in L^1(0,1)$, and $f(x) = f(0) + \int_0^x f'(s)\, ds$ for $0 \le x \le 1$.*

Remark 11.4. It is not true that if $f : [0,1] \to \mathbb{R}$ is continuous, differentiable a.e. and $f' \in L^1(0,1)$, then f is AC. In Rudin, Section 7.16, or, more congenially, in SS, page 125 (bottom), one finds a continuous monotone f with $f(0) = 0$, $f(1) = 1$, satisfying $f'(x) = 0$ a.e. on [0,1].

The proof proceeds through a couple of stages. The first is to establish (11.3) for nondecreasing functions, and the second reduces the general case to the monotonic one.

This is Rudin, Theorem 7.18.

Theorem 11.5. *Let $f : [0,1] \to I\!R$ be continuous and nondecreasing. Then the following conditions on f are equivalent:*

(a) *f is AC on [0,1].*

(b) *f maps sets of measure 0 to sets of measure 0.*

(c) *f is differentiable a.e. on [0,1] and*

$$f(x) = f(0) + \int_0^x f'(s)\,ds \quad (0 \le x \le 1).$$

Proof. We will show that (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (a). Assuming (a), let E be a Lebesgue null subset of [0,1]. We have to show that $f(E)$ is Lebesgue null. Clearly we may assume that $0, 1 \notin E$. Since Lebesgue measure is complete, it is enough to show that, for $\epsilon > 0$, $f(E)$ is contained in a measurable set W which satisfies $|W| \le \epsilon$. Let $\delta > 0$ go with ϵ per Definition 11.2 and let V be an open subset of (0,1) which contains E and satisfies $|V| < \delta$. V is the disjoint union of open intervals (a_j, b_j) which satisfy $\sum_{j=1}^{\infty}(b_j - a_j) < \delta$. Set $W = \cup_{j=1}^{\infty}[f(a_j), f(b_j)]$; clearly $f(V) \subset W$. By the choice of δ,

$$|W| \le \sum_{j=1}^{\infty}(f(b_j) - f(a_j)) \le \epsilon.$$

Thus (a) $\Longrightarrow$ (b).

To continue, assume that f satisfies (b). We first claim that this guarantees that $f(E)$ is measurable if E is measurable. To see this, let $E \subset [0,1]$ be measurable and K_j be an increasing sequence of compact sets such that $K_j \subset E$ and $|E \backslash \cup_{j=1} K_j| = 0$. Since $f(K_j)$ is compact (by continuity of f), it is measurable and therefore so is $\cup_{j=1}^{\infty} f(K_j) = f(\cup_{j=1}^{\infty} K_j)$. Now

$$f(E) = f(E \setminus \cup_{j=1}^{\infty} K_j) \cup f(\cup_{j=1}^{\infty} K_j)$$

presents $f(E)$ as the union of the measurable set $f(\cup_{j=1}^{\infty} K_j)$ and $f(E \setminus \cup_{j=1}^{\infty} K_j)$, which, due to (b), has measure 0. Thus $f(E)$ is measurable.

If $|E| = 0$, then $E \subset \cap_{j=1}^{\infty} V_j$ for some decreasing sequence of open sets V_j such that $|V_j| \searrow 0$. We claim that

$$|f(V_j)| \searrow |f(\cap_{j=1}^{\infty} V_j)| = 0. \tag{11.5}$$

To see this, we figure out the extent to which $\cap_{j=1}^{\infty} f(V_j)$ and $f(\cap_{j=1}^{\infty} V_j)$ might differ. If they differ by a null set, and we will show the difference is an at most countable set, it follows that

$$|f(V_j)| \searrow |\cap_{j=1}^{\infty} f(V_j)| = |f(\cap_{j=1}^{\infty} V_j)| = 0,$$

whence (11.5).

It is clear that $f(\cap_{j=1}^{\infty} V_j) \subset \cap_{j=1}^{\infty} f(V_j)$.

Suppose that $y \in \cap_{j=1}^{\infty} f(V_j) \setminus f(\cap_{j=1}^{\infty} V_j)$. We claim that then the inverse image of y under f, $f^{-1}(y)$, which is a closed interval I_y by the monotonicity of f, cannot be a singleton, $I_y = \{x\}$. Indeed, if it were, then $y \in \cap_{j=1}^{\infty} f(V_j)$ implies $x \in V_j$ for all j and then $y \in f(\cap_{j=1}^{\infty} V_j)$.

Thus $\left\{I_y : y \in \cap_{j=1}^{\infty} f(V_j) \setminus f(\cap_{j=1}^{\infty} V_j))\right\}$ is a pairwise disjoint collection of subintervals of $[0,1]$ of positive length. There can only be at most countably many such intervals. Hence $\cap_{j=1}^{\infty} f(V_j) \setminus f(\cap_{j=1}^{\infty} V_j)$ is countable, and (11.5) is established.

Now set

$$g(x) := x + f(x).$$

We claim that g also satisfies (b). Indeed, if (a,b) is any component of V_j above, then

$$|g((a,b))| = |(a + f(a), b + f(b))| = b - a + f(b) - f(a) = |(a,b)| + |(f(a), f(b))|$$

and this implies that $|g(V_j)| \le |V_j| + |f(V_j)|$. Invoking (11.5), the claim follows.

The virtue of g is its strict monotonicity. It allows us to define a measure μ on Lebesgue measurable subsets of [0,1] by

$$\mu(E) = |g(E)|.$$

Since the image under g of a measurable set E is measurable by arguments above, and since $\{g(E_j)\}_{j=1}$ is a pairwise disjoint collection of sets whenever $\{E_j\}_{j=1}^{\infty}$ is, μ is a measure. Since g satisfies (b), $\mu(E) = 0$ whenever $|\mu(E)| = 0$. According to Radon-Nikodym, there is then an $h \in L^1(0,1)$ such that

$$\mu([0,x]) = |g([0,x])| = g(x) - g(0) = f(x) - f(0) + x - 0 = \int_0^x h(s)\, ds$$

or

$$f(x) = f(0) + \int_0^x (h(s) - s)\, ds \quad (0 \le x \le 1).$$

By previous results $f'(x) = h'(x) - 1$ a.e, and we have proved that (b) $\Longrightarrow$ (c).

We already established that (c) $\Longrightarrow$ (d). $\square$

To get around the restriction to nondecreasing f in the implications (a) $\Longleftrightarrow$ (b) and prove Theorem 11.3, we use the following notion.

Definition 11.6. Let $f : [0,1] \to \mathbb{C}$. Then

$$V(x) := \sup \left\{ \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})| : 0 = t_0 < t_1 < \ldots < t_n = x, n = 1, 2, \ldots \right\} \quad (0 \le x \le 1)$$

is the *total variation of* f over $[0,x]$ and V is the total variation function of f. Note that $V(x) = \infty$ is possible. If $V(x) < \infty$, f is said to be of *bounded variation* on $[0,x]$. We will call a set $\{0 = t_0 < t_1 < \ldots < t_n = x\}$ a *partition* of $[0,x]$.

Theorem 11.7. *Let $f : [0,1] \to \mathbb{C}$ be AC. Then the total variation function V of f is AC. Moreover, if f is real valued, the functions $V, V \pm f$ are nondecreasing and AC on [0,1].*

Proof. Suppose that $0 \le x < y \le 1$. Then for any partition $\{0 = t_0 < t_1 < \ldots < t_n = x\}$ of $[0, x]$ we have

$$V(y) \ge |f(y) - f(x)| + \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})|$$

which implies

$$V(y) \ge V(x) \quad \text{and} \quad V(y) \ge f(y) - f(x) + V(x) \quad \text{and} \quad V(y) \ge f(x) - f(y) + V(x)$$

so $V, V \pm f$ are nondecreasing.

It remains to show that V is AC; that $V \pm f$ is AC then follows from the fact that sums of AC functions are AC. Let $\epsilon > 0$ and $\delta > 0$ be associated to f as in Definition 11.2. Let $\{(\alpha_j, \beta_j)\}_{j=1}^{n}$ be a collection of disjoint subintervals of $[0, 1]$ such that

$$\sum_{j=1}^{n} (\beta_j - \alpha_j) < \delta. \tag{11.6}$$

Let $\kappa > 0$ and $\alpha_j = t_{j,0} < t_{j,1} < \ldots < t_{j,m_j} = \beta_j$ be a partition such that

$$V(\beta_j) - V(\alpha_j) \le \sum_{k=1}^{m_j} |f(t_{j,k}) - f(t_{j,k-1})| + \kappa.$$

Then

$$\sum_{j=1}^{n} (V(\beta_j) - V(\alpha_j)) \le \sum_{j=1}^{n} \sum_{k=1}^{m_j} |f(t_{j,k}) - f(t_{j,k-1})| + n\kappa.$$

Now the intervals $(t_{j,k-1}, t_{j,k})$, $j = 1, \ldots, n, 0 \le k \le m_j$, are pairwise disjoint and the sum of their lengths is less than δ by (11.6). It follows that

$$\sum_{j=1}^{n} (V(\beta_j) - V(\alpha_j)) \le \epsilon + n\kappa,$$

where $\kappa > 0$ was arbitrary. Thus V is AC. □

Proof of Theorem 11.3 Clearly it suffices to assume that f is real. Let V be the variation function of f and consider the nondecreasing AC functions $V + f$ and V. According to Theorem 11.5, both of these functions are differentiable almost everywhere, the derivatives belong to $L^1(0, 1)$, and they are given by the integrals of their derivatives. Thus the same is true for $V + f - V = f$. □

Remark 11.8. Some of the above discussion goes through if we merely assume that $V(1) < \infty$ (and drop the requirement that f be AC. If $V(1) < \infty$, we say that f is of bounded variation on [0,1]. It is still true that the functions $V \pm f$ are nondecreasing, and then $f = (V + f)/2 - (V - f)/2$ represents f as the difference of nondecreasing functions. According to Exercise 40, it follows that if f is of bounded variation, it is differentiable a.e.

11.1. **Changes of Variables in Integrals.** I can see that we aren't going to get this far. We are all tired. However, it is worth at least explaining Theorem 7.26 of Rudin, pages 153-154. In this statement $V \subset \mathbb{R}^n$ and $T : V \to \mathbb{R}^n$ is assumed to be differentiable at every point of some set $X \subset V$, which means that for $x \in X$, then

$$T(x+h) = T(x) + Ah + \mathrm{o}(h) \quad \text{as} \quad h \to 0, \tag{11.7}$$

for some $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$. The condition (11.7) means

$$\lim_{h \to 0} \frac{|T(x+h) - T(x) - Ah|}{|h|} = 0.$$

If (11.7) holds for some A, then A is unique, T is said to be differentiable at x, and one sets $T'(x) := A$. You can think of $T'(x)$ as an $n \times n$ matrix, called the derivative of T at x; I often call it the Jacobian matrix of T at x. It is a calculus theorem that (11.7) holds with $A_{j,k} = \frac{\partial T_k}{\partial x_j}(x)$ if the components of $T(x)$ have continuous partial derivatives with respect to the components x_j of x on V. There is an irritating glitch in the notation here. One either gets seriously pedantic, or more flexibly thinks of elements of $\mathbb{R}^n$ as either rows or columns depending on the situation. In (11.7), we think of vectors as columns in order to match up right with the formula for the entries of A just recalled. When T is differentiable at x we define the Jacobian determinant of T at x by

$$J_T(x) = \det(T'(x))$$

Theorem 11.9. *Suppose that*

(i) $X \subset V \subset \mathbb{R}^n$, V *is open,* $T : V \to \mathbb{R}^n$ *is continuous;*
(ii) X *is measurable,* T *is 1-1 on* X, T *is differentiable at every point of* X;
(iii) $|T(V \setminus X)| = 0$.

Then if $f : \mathbb{R}^n \to [0, \infty]$,

$$\int_{T(X)} f(y)\, dy = \int_X f(T(x))|J_T(x)|\, dx.$$

As regards the various assumptions, the differentiability of T is usually obvious in concrete cases, and the differentiability holds on the full open set V. Verifying the 1-1 assumption in (ii) requires an ad hoc analysis of T, is cannot be verified "globally" by fairly simple conditions on T' when $n > 1$. However, if $T'(x)$ is continuous on V and $J_T(x) \neq 0$, the proof of the inverse function theorem includes a proof that T is "locally 1-1." The condition (iii) looks a bit scary, but Lemma 7.25 of Rudin shows that it is satisfied if T is differentiable at each point of V and $|V \setminus X| = 0$; it is also clearly satisfied if $X = V$. Subtle points include that the measurability of $x \mapsto f(T(x))|J_T(x)|$ is built into the assertion of the theorem, but it is not necessarily true that $x \mapsto f(T(x))$ is measurable just because f is measurable.

12. Appendix: Inner-product and Hilbert Spaces

We assume that you have seen much of this, in a linear algebra course, so we are terse.

An *inner-product space* $(V, \langle\cdot,\cdot\rangle)$ consists of a vector space V (real or complex) and a mapping

$$V \times V \ni (u,v) \to \langle u,v\rangle \in \mathcal{F}, \tag{12.1}$$

(here $\mathcal{F}$ denotes the scalar field, not the Fourier thingy), which has the following properties

$$\begin{aligned} &\text{(i) } u \to \langle u,v\rangle \quad \text{is linear,}\\ &\text{(ii) } \langle u,v\rangle = \overline{\langle v,u\rangle},\\ &\text{(iii) } \langle u,u\rangle \geq 0 \quad \text{and} \quad \langle u,u\rangle = 0 \quad \text{iff} \quad u = 0. \end{aligned} \tag{12.2}$$

Note that if $\mathcal{F} = \mathbb{R}$, then (ii) (in which the overbar means "complex conjugate") says $\langle u,v\rangle = \langle v,u\rangle$.

Example 12.1. A basic example, for us, is the space $L^2(\mu)$, where $(X, \mathcal{M}, \mu)$ is a measure space, with the inner-product

$$\langle f,g\rangle = \int_X f\overline{g}\,d\mu. \tag{12.3}$$

Hereafter $(V, \langle\cdot,\cdot\rangle)$ is an inner-product space. The associated norm $\|\cdot\|$ is given by

$$\|u\| = \sqrt{\langle u,u\rangle}. \tag{12.4}$$

While we used the word "norm," we have yet to verify that it is a norm. All properties of a norm are immediate, except the triangle inequality, which is proved below. First we notice that

$$\begin{aligned} \|u+v\|^2 = \langle u+v,u+v\rangle &= \langle u,u+v\rangle + \langle v,u+v\rangle\\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle\\ &= \|u\|^2 + 2\Re\langle u,v\rangle + \|v\|^2. \end{aligned} \tag{12.5}$$

Lemma 12.2 (Cauchy-Schwarz inequality)**.** *If $u,v \in V$, then*

$$|\langle u,v\rangle| \leq \|u\|\|v\|. \tag{12.6}$$

In consequence, $\|\cdot\|$, as defined in (12.4), is a norm.

Proof. To prove (12.6) we may assume that $u \neq 0, v \neq 0$. For $t \in \mathbb{R}$, (12.5) yields

$$0 \leq \langle u+tv,u+tv\rangle = \|u\|^2 + 2t\Re\langle u,v\rangle + t^2\|v\|^2.$$

The minimum value of the rhs wrt t is attained at

$$t = -\frac{\Re\langle u,v\rangle}{\|v\|^2}.$$

Plugging this value of t into the above, we find

$$0 \leq \|u\|^2 - 2\frac{\Re\langle u,v\rangle}{\|v\|^2}\Re\langle u,v\rangle + \left(-\frac{\Re\langle u,v\rangle}{\|v\|^2}\right)^2 \|v\|^2 = \|u\|^2 - \frac{\Re\langle u,v\rangle^2}{\|v\|^2}$$

or

$$|\Re\langle u,v\rangle| \leq \|u\|\|v\|. \tag{12.7}$$

We are done in a real inner-product space. In the complex case, chose $\alpha \in \mathbb{R}$ such that

$$|\langle u,v\rangle| = e^{i\alpha}\langle u,v\rangle = \langle e^{i\alpha}u,v\rangle$$

and then, via (12.7),

$$|\langle u,v\rangle| = |\Re\langle e^{i\alpha}u,v\rangle| \leq \|e^{i\alpha}u\|\|v\| = \|u\|\|v\|.$$

Finally, the triangle inequality for the norm follows from

$$\|u+v\|^2 = \|u\|^2 + 2\Re\langle u,v\rangle + \|v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2.$$

□

We will say that two vectors u, v are *orthogonal* if

$$\langle u,v\rangle = 0;$$

this is also the meaning of $u \perp v$. Note here that we allow $u = 0$ or $v = 0$.

We will call an indexed family of vectors in V, say $\{v_\alpha : \alpha \in \mathcal{A}\}$, *pairwise orthogonal* if NO $v_\alpha = 0$ and

$$v_\alpha \perp v_\beta \quad \text{for} \quad \alpha, \beta \in \mathcal{A}, \alpha \neq \beta. \tag{12.8}$$

If $\{v_\alpha : \alpha \in \mathcal{A}\}$ is pairwise orthogonal, we say, more elegantly, that $\{v_\alpha : \alpha \in \mathcal{A}\}$ is an *orthogonal sysem.* If, in addition, $\|v_\alpha\| = 1$ for $\alpha \in \mathcal{A}$, we say that $\{v_\alpha : \alpha \in \mathcal{A}\}$ is an *orthonormal system.* It follows from (12.5) that if $\{v_1, \cdots, v_n\}$ is an orthogonal system, then

$$\|\sum_{j=1}^{n} a_j v_j\|^2 = \sum_{j=1}^{n} |a_j|^2 \|v_j\|^2. \tag{12.9}$$

Lemma 12.3. *Let $\{v_1, \ldots, v_n\}$ be a pairwise orthogonal system in V. Let $u \in V$. Then*

$$\left(u - \sum_{j=1}^{n} a_j v_j\right) \perp v_k \quad \textit{for} \quad k = 1, 2, \cdots, v_n, \tag{12.10}$$

iff

$$a_k = \frac{\langle u, v_k\rangle}{\|v_k\|^2} \quad \textit{for} \quad k = 1, 2, \cdots, n. \tag{12.11}$$

Proof. We have

$$\langle u - \sum_{j=1}^{n} a_j v_j, v_k\rangle = \langle u, v_k\rangle - \langle\sum_{j=1}^{n} a_j v_j, v_k\rangle = \langle u, v_k\rangle - a_k\langle v_k, v_k\rangle = \langle u, v_k\rangle - a_k\|v_k\|^2,$$

and the equivalence of (12.10) and (12.11) follows. □

Let $u \in V$ and $\{v_1, \ldots, v_n\}$ be an orthogonal system. Let the a_k be given by (12.11) and $b_1, \ldots, b_n \in \mathcal{F}$. Then

$$\begin{aligned} \|u - \sum_{j=1}^{n} b_j v_j\|^2 &= \| \left(u - \sum_{j=1}^{n} a_j v_j\right) + \left(\sum_{j=1}^{n}(a_j - b_j)v_j\right) \|^2 \\ &= \|u - \sum_{j=1}^{n} a_j v_j\|^2 + \|\sum_{j=1}^{n}(a_j - b_j)v_j\|^2 \\ &= \|u - \sum_{j=1}^{n} a_j v_j\|^2 + \sum_{j=1}^{n} |a_j - b_j|^2\|v_j\|^2 \end{aligned} \tag{12.12}$$

because the two terms displayed in parentheses on the right of the first line are orthogonal by remarks just made and the v_j are an orthogonal system. It follows, in this view, that the the choice $b_j = a_j$, as given by (12.11), makes the left hand side above as small as possible among all choices of the b_j. Taking the $b_j = 0$, we also find that

$$\|u\|^2 = \|u - \sum_{j=1}^{n} a_j v_j\|^2 + \sum_{j=1}^{n} |a_j|^2\|v_j\|^2 \geq \sum_{j=1}^{n} |a_j|^2\|v_j\|^2. \tag{12.13}$$

The inequality of the extremes is called *Bessel's inequality.* We conclude more; rather than use a general index set, we formulate the next two results for the indexing associated with Fourier series. The first result is immediate from (12.13).

Proposition 12.4. *Let $\{v_j\}_{j=-\infty}^{\infty}$ be an orthogonal system in V, $u \in V$ and a_j be given by (12.11), $j = 1, 2, \ldots$. Then*

$$u = \sum_{j=-\infty}^{\infty} a_j v_j \quad \textit{(convergence in } (V, \|\cdot\|)\textit{)}$$

iff

$$\|u\|^2 = \sum_{j=-\infty}^{\infty} |a_j|^2\|v_j\|^2. \tag{12.14}$$

If an inner-product space is complete, it is called a *Hilbert space.*

Proposition 12.5. *Let $(V, \langle\cdot,\cdot\rangle)$ be a Hilbert space and $\{v_j\}_{j=-\infty}^{\infty}$ be an orthogonal system. Let $a_j \in \mathcal{F}, j = 1, \dots$. Then $\lim_{N\to\infty} \sum_{j=-N}^{N} a_j v_j$ exists, that is $\sum_{j=-\infty}^{\infty} a_j v_j$ converges, iff*

$$\sum_{j=-\infty}^{\infty} |a_j|^2 \|v_j\|^2 < \infty. \tag{12.15}$$

Proof. We have, for $M \le N \in \mathbb{N}$,

$$\|\sum_{j=-N}^{N} a_j v_j - \sum_{j=-M}^{M} a_j v_j\|^2 = \sum_{M+1\le|j|\le N} |a_j|^2 \|v_j\|^2,$$

so the sequence of partial sums is Cauchy iff (12.15) holds. □

The following theorm has to do with the "geometry" of Hilbert spaces. The key ingredient of the proof is the fact that if $\mathcal{H}$ is a Hilbert space, $x_1, x_2 \in \mathcal{H}$ and x_1, x_2 and the midpoint of the line segment joining them, $(x_1 + x_2)/2$, all have about the same norm (which is bounded), then $\|x_1 - x_2\|$ has to be small. We single this out in a preliminary lemma.

Lemma 12.6. *Let $\mathcal{H}$ be a Hilbert space, $\lambda > 0, M > 0$, $x_1, x_2 \in \mathcal{H}$ and*

$$\|x_1\|, \|x_2\| \le \lambda M \quad \textit{and} \quad M \le \frac{1}{2}\|x_1 + x_2\|. \tag{12.16}$$

Then

$$\|x_1 - x_2\|^2 \le 4^2 M^2 (\lambda^2 - 1). \tag{12.17}$$

Proof. We may assume that $\mathcal{H}$ is a real Hilbert space (use the real part of the inner-product if $\mathcal{H}$ is originally complex). The first estimate we need is

$$4M^2 \le \|x_1 + x_2\|^2 = \|x_1\|^2 + \|x_2\|^2 + 2\langle x_1, x_2\rangle \le 2\lambda^2 M^2 + 2\langle x_1, x_2\rangle,$$

which implies that

$$2M^2(2 - \lambda^2) \le 2\langle x_1, x_2\rangle.$$

In consequence, using the above,

$$\begin{aligned}\|x_1 - x_2\|^2 &\le \|x_1\|^2 + \|x_2\|^2 - 2\langle x_1, x_2\rangle \\ &\le 2\lambda^2 M^2 - 2M^2(2 - \lambda^2) = 4M^2(\lambda^2 - 1).\end{aligned}$$

□

Theorem 12.7. *Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$. Then there is a unique point $x \in C$ of minimal norm; that is,*

$$\|x\| \le \|y\| \quad \textit{for} \quad y \in C. \tag{12.18}$$

Moreover, $x \in C$ satisfies (12.18) if and only if

$$\|x\|^2 \le \Re\langle y, x\rangle \quad \textit{for} \quad y \in C. \tag{12.19}$$

Proof. Let $M = \inf\{\|y\| : y \in C\}$. Let $\{x_j\}$ be a norm minimizing sequence in C, that is, $\|x_j\| \to M$. If $M = 0$, then $x_j \to 0 \in C$, and we are done, 0 is the unique element of minimal norm. If $M > 0$, define $\lambda_j \geq 1$ by

$$\|x_j\| = \lambda_j M.$$

By assumption, $\lambda_j \to 1$. Since $(x_j + x_k)/2 \in C$, we have

$$M \leq \frac{1}{2}\|x_j + x_k\|.$$

Applying Lemma 12.6 with $\lambda = \max(\lambda_j, \lambda_k)$ and x_j, x_k in place of x_1, x_2, we conclude that

$$\|x_j - x_k\|^2 \leq 4M^2(\max(\lambda_j, \lambda_k)^2 - 1) \to 0 \quad \text{as} \quad j, k \to \infty.$$

Thus $\{x_j\}$ is Cauchy and converges to some $x \in C$ satisfying $\|x\| = \lim_{j\to\infty} \|x_j\| = M$. Thus x is an element of minimal norm. If $\hat{x} \in C$ is any element of minimal norm, we apply (12.6) with $\lambda = 1$ and $x, \hat{x}$ in place of x_1, x_2 to find $x = \hat{x}$.

The characterization (12.19) is established as follows. If (12.19) holds and $y \in C$, then

$$\|x\|^2 \leq \Re\langle x, y\rangle \leq \|y\|\|x\|,$$

so $\|x\| \leq \|y\|$. In the other direction, if $y \in C$ and $0 \leq t \leq 1$, then $x + t(y - x) \in C$ and so

$$\|x\|^2 \leq \|x + t(y-x)\|^2 = \|x\|^2 + 2t\Re\langle x, y - x\rangle + t^2\|y - x\|^2$$

for $0 \leq t \leq 1$. Clearly, then, (12.19) holds. □

In consequence, we can find all the continuous linear functionals on a Hilbert space; it turn out that they are just the elements of the space, as explained below.

Corollary 12.8 (Riesz Representation Theorem)**.** *Let $f \in \mathcal{H}^*$, that is, $f : \mathcal{H} \to \mathcal{F}$ is linear and continuous. Then there is a unique $y \in \mathcal{H}$ such that $f(x) = \langle x, y\rangle$ for $x \in \mathcal{H}$.*

Proof. First, if $f = 0$, then clearly $y = 0$ is the unique vector in $\mathcal{H}$ for which $f(x) = \langle x, y\rangle$ for $y \in \mathcal{H}$.

If $f \neq 0$, let $N_f = \{x \in \mathcal{H} : f(x) = 0\}$. N_f is a closed subspace of $\mathcal{H}$. Suppose $g \in \mathcal{H}^*$ and $N_f \subset N_g$. Choose $z \in \mathcal{H}$ such that $f(z) \neq 0$. We claim that then

$$g(x) = \frac{g(z)}{f(z)} f(x) \ \forall x \in \mathcal{H}. \tag{12.20}$$

Indeed, for any $x \in \mathcal{H}$,

$$f\left(x - \frac{f(x)}{f(z)}z\right) = f(x) - f(x) = 0 \implies x - \frac{f(x)}{f(z)}z \in N_f \subset N_g. \tag{12.21}$$

Hence

$$g\left(x - \frac{f(x)}{f(z)}z\right) = g(x) - \frac{f(x)}{f(z)}g(z) = 0,$$

as claimed in (12.20).

If we can find $0 \neq y \perp N$, ie, $\langle n, y\rangle = 0$ for $n \in N_f$, as we will show that we can do, then $g(x) = \langle x, y\rangle$ satisfies $N_f \subset N_g$ and so, replacing z by y in (12.20),

$$\langle x, y\rangle = \frac{g(y)}{f(y)} f(x) = \frac{\|y\|^2}{f(y)} f(x) \implies f(x) = \langle x, \frac{\overline{f(y)}}{\|y\|^2} y\rangle$$

and renaming

$$\frac{\overline{f(y)}}{\|y\|^2} y$$

to y, we have established the existence of a y such that $f(x) = \langle x, y\rangle$. If $\hat{y}$ also has this property, then $f(x) - f(x) = \langle x, y - \hat{y}\rangle = 0$ for every x, in particular $x = y - \hat{y}$, so $y = \hat{y}$.

It remains to find $0 \neq y \perp N$. Consider the closed convex set $C = z - N$, with z as above, and let $z - n$, $n \in N$, be the element of minimal norm of C. By (12.19),

$$\|z - n\|^2 \leq \Re\langle z - n, z - w\rangle = \Re\langle z - n, z - n + n - w\rangle = \|z - n\|^2 + \Re\langle z - n, n - w\rangle$$

for $w \in N$. Since $n \in N$ while $w \in N$ is arbitrary, we conclude that

$$z - n \perp N \quad \text{ie, } z - n \text{ is perpendicular to every element of } N.$$

Thus $y = z - n$ has the desired properties. $\square$

Exercises

Exercise 1. In Proposition (3.1), we assumed that X was a countable union of sets in $\mathcal{S}$. The effect of this is that the set on the right of (3.2) never empty. However, if we define $\inf \emptyset = \infty$, then (3.2) makes sense without assuming that X is a countable union of sets in $\mathcal{S}$. Show that ν^* is still an outer measure in this situation. (This is almost nothing, given that you may refer to the notes and don't need to reproduce what is there. The point is just for you to learn the definitions.)

Exercise 2. Take $X = S_0 = \left\{(x, y) \in \mathbb{R}^2 : |x|, |y| < 3/2\right\}$; that is, S_0 is the open square centered at the origin of side length 3. Let $\mathcal{S}$ to be the set of open subsquares in S_0 with sides parallel to the axes and $\nu(S) = \operatorname{diam}(S)$. Let S_1, S_2, S_3, S_4 be the open squares of side length 1 in the four corners of S_0 (draw a picture). Verify that $\nu^*(S) = \nu(S)$ for $S \in \mathcal{S}$ and deduce from this that

$$\nu^*(S_0) < \sum_{i=1}^{4} \nu^*(S_i).$$

Finally, use this to conclude that S_i is not ν^*-measurable for some $i \in \{1, 2, 3, 4\}$, and therefore for all $i = 1, 2, 3, 4$. A point of this is that the "$\delta \downarrow 0$" aspect of the definition of Hausdorff measures is necessary to have Borel sets be measurable.

Exercise 3. Let ν^* be an outer measure on X where (X, d) is a metric space. Show that if the Borel sets are ν^*-measurable, then ν^* is a metric outer measure. Hint: Suppose that $A, B \subset X$ and $\operatorname{dist}(A, B) > 0$. Show that for $\delta > 0$, the set $A_\delta = \{x \in X : \operatorname{dist}(x, A) < \delta\}$ is open, and for δ small $A_\delta \cap B = \emptyset$. Now use the assumed measurability of A_δ.

Exercise 4. Show the the definition (3.19) is unchanged if the restriction $s : X \to [0, \infty)$ is added on the right.

Exercise 5. Show that with the definition (3.19) one has

$$\int_E f\, d\mu = \int_X \chi_E f\, d\mu = \sup \left\{ \int_E s\, d\mu : s \quad \text{is simple and} \quad 0 \le \chi_E s \le \chi_E f \right\}.$$

(Rudin top of pg 20)

Exercise 6. Let $f : X \to [0, \infty]$. If $\mu(E) = 0$, then $\int_E f\, d\mu = 0$ even if $f(x) = \infty$ for all $x \in E$. If $f(x) = 0$ for all $x \in E$, then $\int_E f\, d\mu = 0$ even if $\mu(E) = \infty$. (Rudin top of pg 20)

Exercise 7. Verify the claims about $d_p(f, g) = 0$ and $\sim, X/\sim, \tilde{d}$ around (4.8).

Exercise 8. Show that the conditions

$$\forall \epsilon > 0\ \exists\ J \ni \mu\left(\{x : |f_j(x) - f(x)| > \epsilon\}\right) < \epsilon \quad \text{for} \quad j > J \tag{12.22}$$

and

$$\forall \epsilon > 0\ \exists\ N \ni \mu\left(\{x : |f_j(x) - f_k(x)| > \epsilon\}\right) < \epsilon \quad \text{for} \quad j, k > N. \tag{12.23}$$

are equivalent to (4.16), (4.17), respectively.

Exercise 9. (a) Show that if $f_j \to f$ in measure, then $\{f_j\}$ is Cauchy in measure.
(b) Show that if $f_j \to f$ and $f_j \to g$ in measure, then $f = g$ a.e.
(c) Is it necessarily true that if f is the function constructed in the proof of Lemma 4.3, then $f_j \to f$ in measure? (See also Rudin, Pg 74, #18.)

Exercise 10. Show that $L^\infty(\mu)$ is complete.

Exercise 11. Let $1 \le r, s \le \infty$. If $(X, \mathcal{M}, \mu)$ is a finite measure space, show that $L^r(\mu) \subset L^s(\mu)$ if $r \ge s$. Hint: Hölder inequality.

Exercise 12. In the case of Lebesgue measure on a subset E of $\mathbb{R}^n$, we simply write $L^p(E)$ rather than display Lebesgue measure restricted to E. Let $1 \le r, s \le \infty$.
(a) Let B be the unit ball in $\mathbb{R}^n$. Show that $L^r(B) \subset L^s(B)$ only if $r \ge s$.
(a) Show that $L^r(\mathbb{R}^n) \subset L^s(\mathbb{R}^n)$ only if $r = s$.

Hint: Consider functions of the form $|x|^{-\alpha}$ near infinity and near 0. You may assume that changes of variables to spherical coords are ok and Riemann integrals of nonnegative functions are Lebesgue integrals.

Exercise 13. (a) The identity map I of $\mathbb{R}^3$, that is, $Ix = x$ for $x \in \mathbb{R}^3$, is a bounded linear mapping from any of l_3^2, l_3^1, l_3^∞ into any of l_3^2, l_3^1, l_3^∞ (see Definition 4.4). Compute the norm of I as a mapping between any two of these spaces.
(b) Determine the set of $1 \le r, s \le \infty$ for which $l^r \subset l^s$ for the sequence spaces of Definition 4.4.

Exercise 14. Let Z be a vector space. If $z, w \in Z$, the line segment joining z and w is

$$[z, w] := \{tz + (1-t)w : 0 \le t \le 1\}.$$

A subset C of Z is *convex* if it contains the line segment joining any two of its points. For general $C \subset Z$, the *convex hull* $\operatorname{conv} C$ of C is the smallest convex set containing C. Show that $\operatorname{conv} C$ is the set of all vectors of the form

$$t_1 z_1 + \cdots + t_m z_m$$

where $z_j \in C$, $t_j \ge 0$ for $j = 1, \ldots, m$,

$$t_1 + \cdots + t_m = 1,$$

and $m = 1, 2, \ldots$.

Exercise 15. Let $(Z, \|\ \|$ be an nls. If $C \subset Z$, then $\overline{C}$ is its closure, C° denotes its interior, $\operatorname{conv} C$ is its convex hull, and $\overline{\operatorname{conv}}\, C$ is the closure of the convex hull of C. If $a \in \mathcal{F}$, then $aC = \{az : z \in C\}$. If $C, D \subset Z$, then $C + D = \{z + w : z \in C, w \in D\}$ and $C - D = \{z - w : z \in C, w \in D\}$.

Below, $a \in \mathcal{F}$, and $C, D \subset Z$.

(a) Show that $\overline{aC} = a\overline{C}$.
(b) Show that $(aC)^\circ = aC^\circ$.
(c) If C is convex, show that $\overline{C}$ and C° are convex.
(d) If C is open, then so is $C + D$.

Exercise 16. Show that $\int_{-\pi}^{\pi} |D_n(t)|\, dt \to \infty$ as $n \to \infty$.

Exercise 17. Consider the mapping $f \mapsto s_N(f) := \sum_{-N}^{N} \hat{f}(n) e^{int}$, which takes f into the N^{th} partial sum of its Fourier series. Show that it is not true that for all f in $L^1(-\pi, \pi)$ we have $\|f - s_N(f)\|_{L^1(-\pi,\pi)} \to 0$ as $N \to \infty$. Hint: In outline, but not in detail, the proof mimics the proof that Fourier series of a continuous function do not converge pointwise to the function. You need to show that $\|S_N\|$ is unbounded in N, and this is related to $\|D_N\|_{L^\infty(-\pi,\pi)}$.

Exercise 18. Let X be a normed linear space, Y be a Banach space, $M \in \mathbb{R}$, and $\{T_j\}_{j=1}^\infty \subset \mathcal{L}(X, Y)$ satisfy

$$\|T_j\| \le M \quad (j = 1, 2, \ldots).$$

Show that $C := \{x \in X : \lim_{j\to\infty} T_j x \text{ exists}\}$ is a closed linear subspace of X. Show that if C is dense, then $Tx := \lim_{j\to\infty} T_j x$ is a bounded linear operator and $\|T\| \le \limsup_{j\to\infty} \|T_j\|$.

Exercise 19. Let X, Y be a Banach spaces, and $\{T_j\}_{j=1}^\infty \subset \mathcal{L}(X, Y)$. Let $T : X \to Y$ be linear. One says that $T_j \to T$

(a) in the uniform operator topology if $\|T_j - T\| \to 0$ (which gets abbreviated to $T_j \to T$ in norm).

(b) in the strong operator topology if $T_j x \to Tx$ for all $x \in X$ (which gets abbreviated to $T_j \to T$ strongly).
(c) in the weak operator topology if $f(T_j x) \to f(Tx)$ for all $x \in X$ and all $f \in Y^*$, the dual space of Y (which gets abbreviated to $T_j \to T$ weakly).

Here are questions about these notions of convergence of operators.

(i) Show that $T_j \to T$ in norm implies $T_j \to T$ strongly implies $T_j \to T$ weakly.
(ii) According to Theorem 6.4 the operators s_N satisfy $s_N \to I$ (the identity) strongly. Show that $s_N \to I$ does not hold in norm.
(iii) Show that the translations $\tau_h f(x) = f(x-h)$ satisfy $\tau_h \to I$ strongly in $L^p(\mathbb{R}^n)$, $1 \leq p < \infty$, as $h \to 0$, but $\tau_h \to I$ does not hold in norm.
(iv) Do the same for the mappings $f \to \rho_\epsilon * f$ of Theorem 7.2 (as $\epsilon \downarrow 0$.)
(v) Consider the shift operator $S(\{x_1, x_2, \ldots\}) = \{0, x_1, x_2, \ldots\}$, which is an isometry of l^2 into l^2. Show that $S^j \to 0$ weakly, but not strongly. You may assume Corollary 12.8, which tells you what the continuous linear functionals on l^2 are.

Exercise 20. (a) Provide details about the claim (7.3).
(b) In the context of (7.3), provide a formula for $\int_E f(T^{-1}x)\,dx$ in terms of the integral of f over some set.

Exercise 21. Show that if $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue measurable, then there is a Borel measurable function f_0 such that $f = f_0$ a.e. You may quote - do so in detail by giving page numbers or statements - results from KF or your lecture notes from last quarter, but not other sources. Close any gaps to accommodate "infinite measures."

Exercise 22. Let $1 \leq p, q \leq \infty$ (not necessarily conjugate) be such that

$$\frac{1}{p} + \frac{1}{q} \geq 1 \quad \text{and define } r \geq 1 \text{ by} \quad \frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1.$$

Show that $f * g$ exists a.e. and

$$\|f * g\|_{L^r(\mathbb{R}^n)} \leq \|f\|_{L^p(\mathbb{R}^n)} \|g\|_{L^q(\mathbb{R}^n)}.$$

Exercise 23. Let $f \in L^1(\mathbb{R}^n)$ and consider the mapping $Tg = f * g$ of $L^p(\mathbb{R}^n)$ into itself where $1 \leq p < \infty$. Show that the norm of T is $\|f\|_{L^1(\mathbb{R}^n)}$.

Exercise 24. Show that if $f \in L^1(\mathbb{R}^n)$, then $u(t, x)$ as given by (7.12) is continuous on $\mathbb{R}^n \times (0, \infty)$ and has continuous first partial derivatives there, which may be computed by differentiating "under the integral sign."

STARRED EXERCISES BELOW NEED NOT BE TURNED IN - WHICH ONES ARE STARRED MAY CHANGE A BIT (until March 8). IF YOU SEE WRONG THINGS, LET ME KNOW - TIA. Right now there are 14 unstarred problems below, but I judge about 7 of them to be short and "easy."

Exercise 25. There is something amiss with our sketch of proof in Section 7.2. It does not appear to use that $f(\pi) = f(-\pi)$, which certainly must be the case if f is the uniform limit of trig polys. Clarify this point.

Exercise 26. Show that

$$P(x) = \limsup_{j\to\infty} x_j$$

is a sublinear function on (real) l^∞. Show that if $L : l^\infty \to \mathbb{R}$ is linear and $L \le P$, then L is continuous and satisfies

$$L(x) = \lim_{j\to\infty} x_j$$

whenever the limit exists. Continuous linear functionals with this property are called "Banach limits."

Exercise 27. Use Zorn's Lemma, as formulated during the proof of Lemma 8.4, to prove that if " $\le$ " is a partial order on a set S, and every chain in S has an upper-bound, then S contains maximal elements (with the obvious definition).

Exercise 28. Let H be a Hilbert space. For any $S \subset H$, define

$$S^\perp = \{x \in H : \langle x, s\rangle = 0 \text{ for } s \in S\}.$$

(a) Show that $S^\perp$ is a closed linear subspace of H.
(b) Show that $\left(S^\perp\right)^\perp$ is the closure of the linear span of S.

Exercise 29.* Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Let $D(A)$ be a subspace of H, not necessarily closed, and $A : D(A) \to H$ be linear. We will abbreviate this to "A is a linear operator in H." Define the graph $G(A)$ of A by

$$G(A) = \{(x, Ax) : x \in D(A)\} \subset H \times H.$$

Note that $G(A)$ is a linear subspace of $H \times H$ and $H \times H$ is a Hilbert space when equipped with the inner-product

$$\langle (x,y),(u,v)\rangle_{H\times H} := \langle x,u\rangle + \langle y,v\rangle.$$

(a) Let V be a linear subspace of $H \times H$. Show that it is the graph of a linear operator iff $(0, v) \in V$ implies $v = 0$.
(b) Show that

$$\left\{(x,y) : (-y,x) \in G(A)^\perp\right\}$$

is the graph of a linear operator iff $D(A)$ is dense. If $D(A)$ is dense, the operator A^* whose graph is $\left\{(x,y) : (-y,x) \in G(A)^\perp\right\}$ is called *'the adjoint of A."*
(c) If $D(A)$ is dense, show that $D(A^*)$ is dense if and only if the closure of $G(A)$ is a graph. This property of A, that the closure of $G(A)$ is a graph, is referred to by saying "A is *closeable*," while the operator $\bar{A}$ whose graph is the closure of the graph of A is called the *closure of* A. If $G(A)$ itself is closed, then A is said to be a *closed linear operator.*
(d) Show that if A is a densely defined closeable linear operator, then $(A^*)^* = \bar{A}$.

(e) If $A \in \mathcal{L}(H,H)$, show that $A^* \in \mathcal{L}(H,H)$ and $\|A^*\| = \|A\|$.

(f) If A is a closed linear operator, and $R(A) := \{Ax : x \in D(A)\}$, show that $R(A)^\perp = N(A^*) := \{x \in D(A^*) : A^*x = 0\}$. If $R(A)$ is closed, we conclude that $Ax = y$ has a solution $x \in D(A)$ iff $y \perp N(A^*)$. This is the *Fredholm Alternative.*

(g) Let $D(A)$ be the set of twice continuously differentiable functions u on $[0,1]$ satisfying $u(0) = u(1) = 0$ (you may assume that $D(A)$ is dense in $L^2(0,1)$) and

$$(Au)(x) = -(1+x^2)u''(x) + \int_0^x \sin(s)e^x u(s)\,ds \quad (u \in D(A), 0 \le x \le 1).$$

Show that A is closeable by showing that $D(A^*)$ is dense.

Exercise 30.* Let H be a Hilbert space. Define

$$S = \big\{\{v_\alpha\}_{\alpha\in\mathcal{A}} : \{v_\alpha\}_{\alpha\in\mathcal{A}} \text{ is an orthonormal system in } H\big\} \tag{12.24}$$

I guess we have to say what this notation means. "$\{v_\alpha\}_{\alpha\in\mathcal{A}}$" denotes an "indexed family of vectors," where v_α has the index α from the "index set" $\mathcal{A}$. In other words, we are talking about a function $f : \mathcal{A} \to H$, where v_α is another name for $f(\alpha)$. We are about to partially order S by

$$\{v_\alpha\}_{\alpha\in\mathcal{A}} \le \{v'_\alpha\}_{\alpha'\in\mathcal{A}'}$$

where "$\le$" means inclusion of the sets of vectors, independent of the particular indexing. That is, if $f : \mathcal{A} \to H$, $f' : \mathcal{A}' \to H$ are the functions corresponding to the "systems," then $f \le f'$ means $f(\mathcal{A}) \subset f'(\mathcal{A}')$.

(a) Show that S has a maximal element with this ordering.

(b) Show that $\{v_\alpha\}_{\alpha\in\mathcal{A}}$ is maximal iff H is the closure of the linear span of $\{v_\alpha\}_{\alpha\in\mathcal{A}}$. Hint: If H is not the closure of the linear span of $\{v_\alpha\}_{\alpha\in\mathcal{A}}$, rediscover where we proved that then there is a unit vector v which satisfies $\langle v, v_\alpha\rangle = 0$ for $\alpha \in \mathcal{A}$.

(c) If $\{v_\alpha\}_{\alpha\in\mathcal{A}}$ is any orthonormal system and $v \in H$, use Bessel's inequality (12.13) to show that $\langle v, v_\alpha\rangle = 0$ except for at most countably many α. (Hint: how many α's can there be for which $|\langle v, v_\alpha\rangle| \ge \delta$ where $\delta > 0$?

(d) If $\{v_\alpha\}_{\alpha\in\mathcal{A}}$ is maximal and $v \in H$, show that

$$v = \sum_{\alpha\in\mathcal{A}, \langle v, v_\alpha\rangle \ne 0} \langle v, v_\alpha\rangle v_\alpha.$$

By (c), the sum above is over an at most countable set. Not indicating the order of summation means that the statement is true no matter how the summands are ordered.

(e) Conclude that if $\{v_\alpha\}_{\alpha\in\mathcal{A}}$ is maximal, then the mapping $v \mapsto \{\langle v, v_\alpha\rangle\}_{\alpha\in\mathcal{A}}$ is an isometry of H onto the Hilbert space of functions $g : \mathcal{A} \to \mathbb{C}$ such that $\sum_{\alpha\in\mathcal{A}} |g(\alpha)|^2 < \infty$, which is equipped with the inner-product

$$\langle h, g\rangle = \sum_{\alpha\in\mathcal{A}} h(a)\overline{g(a)}.$$

This space is just L^2(counting measure on $\mathcal{P}(\mathcal{A})$).

Exercise 31. Let μ be a complex measure on $(X, \mathcal{M})$. Show that if $\{A_j\}_{j=1}^{\infty}$ is an increasing sequence of measurable sets, then

$$\mu(\cup_{j=1}^{\infty} A_j) = \lim_{j\to\infty} \mu(A_j).$$

Exercise 32. Give an example to show that if μ be a signed measure on $(X, \mathcal{M})$, then $A \mapsto \mu(A)^+$ is not necessarily a nonnegative real measure.

Exercise 33. Let $(X, \mathcal{M})$ be a measurable space. Consider the space

$$M = \{\mu : \mu \text{ is a complex measure on } (X, \mathcal{M})\}$$

M is a vector space in the obvious way. Show that

$$\|\mu\| = |\mu|(X)$$

is a norm on M and $(M, \|\cdot\|)$ is a Banach space. Hint: suppose $\{\mu_j\}_{j=1}^{\infty}$ is a Cauchy sequence in M and $\{E_k\}_{k=1}^{\infty}$ is a sequence of pairwise disjoint measurable sets. Along the way, show that for $\epsilon > 0$ there exists N_ϵ such that

$$|\mu_j|(\cup_{k=K}^{\infty} E_k) \leq \epsilon \quad \text{for} \quad K \geq N_\epsilon, j = 1, 2, \ldots.$$

Exercise 34. Let $(X, \mathcal{M}, \mu)$ be an arbitrary measure space and $f \in L^p(\mu)$ where $1 \leq p \leq \infty$ and f is allowed to take complex values.

(a) Show that if $1 \leq p < \infty$ and $\int_E f\, d\mu = 0$ for every set of finite measure E, then $f = 0$ a.e.

(b) Show, however, that the assertion of (a) is in general false for $f \in L^\infty(\mu)$ by considering the weird space preceding (3.15).

(c) Show that the assertion of (a) does hold for $p = \infty$ if μ is σ-finite.

Exercise 35. Let $f, g \in L^1(\mathbb{R}^n)$, $0 \leq f$. Let

$$\lambda(E) = \int_E g\, dx, \;\; \mu(E) = \int_E f\, dx.$$

Find the Lebesgue decomposition of λ with respect to μ.

Exercise 36. (a) Part of the proof of the Arzéla-Ascoli Theorem (AAT) is this: Let (X, d), (Y, ρ) metric spaces and $\{f_j\}_{j=1}^{\infty}$ be a sequence of functions $f_j : X \to Y$ such that $\{f_j(x)\}_{j=1}^{n}$ lies in a compact subset of Y for each $x \in X$. Let $\{x_k\}_{k=1}^{\infty}$ be a countable subset of X. Then there is a subsequence f_{j_l} of the f_j's such that $\lim_{l\to\infty} f_{j_l}(x_k)$ exists for each k. The proof goes by selecting a subsequence of the f_j's along which $f_j(x_1)$ converges, then a subsequence of the one just selected along which $f_j(x_2)$ converges, and so on, and then "diagonalize." Do not write this up, but do review/understand it.

(b) Let $(X, \|\cdot\|)$ be a separable normed space; that is, there is a countable dense subset of X, call it $\{x_k\}_{k=1}^{\infty}$. Let X^* be the space of continuous linear functionals on X and $\left\{x_j^*\right\}_{j=1}^{\infty} \subset X^*$ be bounded. then $\left\{\langle x_j^*, x\rangle\right\}$ is bounded for each $x \in X$

(we are denoting the value of $x^* \in X^*$ at $x \in X$ by $\langle x^*, x\rangle$ in this exercise). Using Part (a) of this exercise and Exercise 18, conclude that there is a subsequence $x^*_{j_l}$ of the x^*_j and $x^* \in X^*$ such that $\langle x^*_{j_l}, x\rangle \to \langle x^*, x\rangle$ for $x \in X$. In the nomenclature of Exercise 18, $x^*_{j_l} \to x^*$ *strongly.*

Exercise 37. Explain how Corollary 8.8 provides a nonzero $\Phi \in L^\infty(0,1)^*$ which is 0 on $C([0,1])$. Prove that there is no $g \in L^1(0,1)$ for which $\Phi_g = \Phi$.

Exercise 38. Let $(X, \|\cdot\|)$ be a Banach space, X^* be the space of continuous linear functionals on X and "$\langle x^*, x\rangle$" have the same meaning as in Exercise 36. A sequence $\{x_k\}_{k=1}^\infty$ is said to *converge weakly to* $x \in X$ if $\langle x^*, x_k\rangle \to \langle x^*, x\rangle$ for $x^* \in X^*$, and a sequence $\{x^*_j\}_{j=1}^\infty$ is said to *converge weakly* to* $x^* \in X^*$ if $\langle x^*_j, x\rangle \to \langle x^*, x\rangle$ for $x \in X$. Note that weak* convergence of a sequence in X^* is the same as strong convergence as a sequence of operators and express your irritation at the nomenclature (to yourself). This nomenclature horror is compounded by the common usage of $x_k \to x$ "strongly" to mean $\|x_k - x\| \to 0$.

(a) Show that if $\{x_k\}_{k=1}^\infty \subset X$ and $x_k \to x$ weakly, then the sequence is bounded. Hint: Each $x \in X$ defines a continuous linear functional on X^* by $x^* \to \langle x^*, x\rangle$. What is the norm of this linear functional?

(b) Show that $L^p(\mathbb{R}^n)$ is separable for $1 \le p < \infty$ and $L^\infty(\mathbb{R}^n)$ is not separable. I'll be satisfied with $n = 1$, for simplicity in writing. Hints: (i) $C_0(\mathbb{R})$ is dense. (ii) Approximate compactly supported continuous functions uniformly by step functions with rational values on some well chosen countable collection of intervals.

(c) Conclude that if $\{f_j\}_{j=1}^\infty$ is a bounded sequence in $L^p(\mathbb{R}^n)$ and $1 < p < \infty$, then it has a subsequence $\{f_{j_l}\}$ for which there is an $f \in L^p(\mathbb{R}^n)$ such that $f_{j_l} \to f$ weakly; explain why this amounts to

$$\int_{\mathbb{R}^n} f_{j_l} g\, dx \to \int_{\mathbb{R}^n} f g\, dx \quad (g \in L^{p/(p-1)}(\mathbb{R}^n)).$$

(d) Show that the assertion of (c) is false if $p = 1$. Hint: we have already discussed a bounded sequence in $L^1(\mathbb{R})$ that provides a counterexample when discussing approximate identities.

Exercise 39. Let $f \in L^p(\mathbb{R}^n)$ where $1 \le p < \infty$.

(a) Use Theorem 10.4 to show that almost every $x \in \mathbb{R}^n$ is a Lebesgue point of f.

(b) Let ρ_ϵ be as in Theorem 7.2; in addition, for simplicity, assume that ρ is bounded and compactly supported. Show that $\lim_{\epsilon\downarrow 0}(\rho_\epsilon * f)(x) = f(x)$ holds at every Lebesgue point of f.

(c) Show that $|\rho_\epsilon * f|^p \le \rho_\epsilon * |f|^p$.

(d) Use dominated convergence to show that $\rho_\epsilon * f \to f$ in $L^p(\mathbb{R}^n)$ (without using approximation by continuous functions - this part of the proof is now buried in (a).).

Exercise 40.* Assume Theorem 7.14 of Rudin:

Theorem 12.9. *Let μ be a complex Borel measure on $I\!R^n$. Let $d\mu = f d\mathcal{L}^n + d\mu_s$ be its Lebesgue decompostion. For each $x \in I\!R^n$ let $\{E_j(x)\}$ be a sequence of sets which shrinks nicely to x. Then*

$$\lim_{j\to\infty} \frac{\mu(E_j(x))}{|E_j(x)|} = f(x) \quad \textit{for almost all } x \in I\!R^n.$$

(a) Let $f : [0,1] \to \mathrm{I\!R}$ be nondecreasing and continuous from the left; putting $f(x) = f(0)$ for $x \leq 0$ and $f(x) = f(1)$ for $1 \leq x$, we can assume that f is nondecreasing and left-continuous on $\mathrm{I\!R}$. Let μ be the nonnegative real Borel measure on $\mathrm{I\!R}$ for which $\mu([a,b)) = f(b) - f(a)$ (see KF pg 13 - here clearly μ is concentrated on [0,1]). Apply Theorem 12.9 to conclude that f is differentiable ae on $[0,1]$.

(b) Show that if $f : [0,1] \to \mathrm{I\!R}$ is nondecreasing, then f is continuous except at an at most countable set of discontinuities. Hint: f is discontinuous at $x \in (0,1)$ only if f "jumps" at x, that is, the jump at x,

$$f(x+) - f(x-) := \lim_{y\downarrow x} f(y) - \lim_{y\uparrow x} f(y) > 0.$$

Now bound the sum of the jumps.

(c) Noting that if $f : [0,1] \to \mathrm{I\!R}$ is nondecreasing, then $x \mapsto f(x+)$ is continuous from the left (understand this, but do not write the proof), deduce that f is differentiable a.e.

Remark: A more elementary - and probably more friendly - presentation of these results is found in SS, Chapter 3.