

Advanced Analysis

Uwe Kaiser

04/29/09

Department of Mathematics
Boise State University
1910 University Drive
Boise, ID 83725-1555, USA

email: kaiser@math.boisestate.edu

Abstract

This is a preliminary version of introductory lecture notes for Advanced Analysis. It does not claim any originality. It is mainly based on Serge Lang's books **Real and Functional Analysis** and **Undergraduate Analysis**. I also have added some more discussion of the linear algebra concepts. There will be no proofs in many cases where they can be found in Lang's Real and Functional Analysis or other standard Analysis textbooks. I would like to thank the students of this course for their helpful comments and remarks.

Chapter 1

Topological Spaces

In this chapter we discuss basics of topological spaces aimed at the application to normed vector spaces and metric spaces. For each set S we let $|S|$ denote the cardinality of S , i. e. the equivalence class of S under bijections of sets.

1.1. Definition. Let X be a set. A *topology* on X is a collection \mathcal{T} of subsets of X , called the *open sets* such that

TOP1: $\emptyset, X \in \mathcal{T}$

TOP2: $U, V \in \mathcal{T} \implies U \cap V \in \mathcal{T}$

TOP3: $U_i \in \mathcal{T}$ for $i \in I$ (I any index set) $\implies \cup_{i \in I} U_i \in \mathcal{T}$.

(X, \mathcal{T}) or just X is called *topological space*. Note that TOP2 implies that all *finite* intersections of open sets are open.

If $x \in U \in \mathcal{T}$ then U is an *open neighborhood* of x . A *neighborhood* of x is any subset of X containing an open neighborhood of x . Note that $U \in \mathcal{T} \iff U$ is neighborhood of each of its points. (If U is open then it is open neighborhood and thus neighborhood of each of its points. If U is neighborhood of each of its points then it contains an open neighborhood in U of each of its points. Thus U is the union of those open neighborhoods contained in U , and thus open by TOP3.)

1.2. Example. On X any set let $\mathcal{T} := \{\emptyset, X\}$.

1.3. Example. On X any set let $\mathcal{T} := \{U \subset X\}$. This is called the *discrete topology*.

We need some preparation for the most important examples of topological spaces we will consider.

1.4. Example. On $X = \mathbb{R}$ define $U \in \mathcal{T} \iff \forall x \in U \exists J$ open interval in \mathbb{R} such that $J \subset U$. This is the *ordinary topology* on \mathbb{R} .

1.5. Definition. A *real vector space* is a set E with two operations: *vector addition* $+E \times E \rightarrow E$, $(v, w) \mapsto v + w$, and *multiplication by scalars* $\mathbb{R} \times E \rightarrow E$, $(a, v) \mapsto av$ such that for all $u, v, w \in E$ and $a, b \in \mathbb{R}$:

$$\text{VS1: } (u + v) + w = u + (v + w)$$

$$\text{VS2: } \exists 0 \in E : 0 + v = v$$

$$\text{VS3: } \forall v \in E \exists w \in E : v + w = 0$$

$$\text{VS4: } v + w = w + v$$

$$\text{VS5: } 1v = v, (ab)v = a(bv), (a + b)v = av + bv, a(v + w) = av + aw$$

1.6. Definition. A *subspace* of a vector space E is a nonempty subset F such that $v, w \in F \implies v + w \in F$ and $v \in F, a \in \mathbb{R} \implies av \in F$.

Note that the zero vector is in a subspace F because $v \in F$ for some $v \in E$. Thus $(-1)v = -v \in F$ and thus $v + (-v) = 0 \in F$.

Complex vector spaces are defined in the same way with the scalars \mathbb{R} replaced by the complex numbers \mathbb{C} .

1.7. Example. Given a set S and a vector space E , the set of all maps $f : S \rightarrow E$ is a vector space with addition and multiplication by scalars defined pointwise, i. e. $(f + g)(x) := f(x) + g(x)$ and $(af)(x) := af(x)$. VS1-VS5 immediately following from the corresponding laws in E . The zero-element is the map, which is constant 0.

1.8. Definition. A *normed vector space* is a vector space E with a map

$$E \ni x \mapsto |x| \in \mathbb{R}$$

such that

$$\text{NVS1: } |x| \geq 0 \text{ and } |x| = 0 \iff x = 0$$

$$\text{NVS2: For } c \in \mathbb{R}, x \in E, \text{ we have } |cx| = |c||x|$$

$$\text{NVS3: For } x, y \in E \text{ we have } |x + y| \leq |x| + |y|$$

Complex normed vector spaces are defined in the same way with $c \in \mathbb{C}$ in NVS2 and $|c|$ denoting the absolute value of a complex number. If E is a normed vector space and $F \subset E$ is a subspace of the vector space E then the norm on E restricts to a norm on F . In this case *subspace* means this normed vector space.

1.9. Example. Given a normed vector space E , $v \in E$ and $r > 0$ a real number let

$$B_r(v) := \{x \in E : |x - v| < r\}$$

be the *ball* of radius r and center v . Define $U \in \mathcal{T} \iff \forall x \in U \exists r > 0 : B_r(v) \subset U$. This collection \mathcal{T} satisfies TOP1-3 above. To see $X \in \mathcal{T}$ you can take any $r > 0$. If $U, V \in \mathcal{T}$ let $x \in U \cap V$. Then $\exists r_1 > 0$ such that $B_{r_1}(x) \subset U$ and $\exists r_2 > 0$ such that $B_{r_2}(x) \subset V$. Then

$$B_{r_1}(x) \cap B_{r_2}(x) = B_{\min(r_1, r_2)}(x) \subset U \cap V.$$

Thus TOP2 holds. If $U_i \in \mathcal{T}$ and $x \in \cup_{i \in I} U_i$ then $x \in U_i$ for some $i \in I$, so $\exists r > 0$ such that $B_r(x) \subset U_i \subset \cup_{i \in I} U_i$. Thus TOP3 holds.

1.10. Remark. Open balls are open. To see this just note that

$$B_{r-|x-y|}(x) \subset B_r(y)$$

for $y \in E$ and $x \in B_r(y)$. An open ball with center x and radius ε is also called an ε -neighborhood of x .

1.11. Definition. (i) A sequence $\{x_n\}$ in E is *Cauchy* if $\forall \varepsilon > 0 \exists N$ such that

$$m, n > N \implies |x_m - x_n| < \varepsilon.$$

(ii) A sequence $\{x_n\}$ converges to $x \in E$ (Notation: $x_n \rightarrow x$) if $\forall \varepsilon > 0 \exists N$ such that

$$n > N \implies |x_n - x| < \varepsilon.$$

\mathbb{R}^n with the Euclidean norm

$$|x| = |(x_1, \dots, x_n)| := \sqrt{\sum_{i=1}^n x_i^2}$$

is a real normed vector space. Similarly \mathbb{C}^n with the norm

$$|z| = |(z_1, \dots, z_n)| = \sqrt{\sum_{i=1}^n |z_i|^2}$$

is a complex normed vector space.

We now give two constructions of new normed vector spaces from old ones. First note that for any vector space E and set S the set of functions $f : S \rightarrow E$ is a vector space by pointwise addition and multiplication by scalars:

$$(f + g)(x) := f(x) + g(x) \in E \quad \text{and} \quad (cf)(x) := cf(x) \in E$$

The axioms VS1-VS5 hold because they hold pointwise. The 0-element is the function, which is constant 0.

1.12. Example. Let S be a set and F a normed vector space. A function $f : S \rightarrow F$ is *bounded* if $\exists C > 0$ such that $|f(x)| \leq C$ for all $x \in S$. Then

$$\|f\|_S = \|f\| := \sup_{x \in S} |f(x)| \in \mathbb{R}$$

defines a norm (the *supremum norm*) on

$$B(S, F) := \{f : S \rightarrow F \text{ bounded}\}.$$

To see this we show that $B(S, F)$ is a subspace of the vector space of all functions $S \rightarrow F$ and $\|\cdot\|_S$ satisfies NVS1-NVS3. Obvious $0 \in B(S, F)$ and $c \in \mathbb{R}$ (or \mathbb{C}) and $f \in B(S, F)$ implies $cf \in B(S, F)$ and $\|cf\| = \sup_{x \in S} |cf(x)| = |c| \sup_{x \in S} |f(x)| = |c| \|f\|$. Moreover for $x \in S$ we have

$$|(f + g)(x)| = |f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\| + \|g\|,$$

and thus $f + g \in B(S, F)$ and by taking the supremum over x in the equation above

$$\|f + g\| \leq \|f\| + \|g\|.$$

A sequence of functions $\{f_n : S \rightarrow F\}$ is *uniformly Cauchy* on S if it is Cauchy with respect to the supremum-norm, thus $\forall \varepsilon > 0 \exists N$ such that

$$m, n > N \implies \|f_n - f_m\| < \varepsilon.$$

Similarly we say $\{f_n\}$ *converges uniformly* to f if $\varepsilon > 0 \exists N$ such that

$$n > N \implies \|f_n - f\| < \varepsilon.$$

1.13. Example. Let $E := \{f : [0, 1] \rightarrow \mathbb{R} \text{ continuous}\}$ and for $f \in E$ let

$$\|f\|_1 := \int_0^1 |f(x)| dx.$$

Then E is a subspace of the vector space of all functions $[0, 1] \rightarrow \mathbb{R}$. It is known from previous analysis classes that E is a subspace. NVS2 and NVS3 follow from standard properties of the integral: $\|cf\|_1 = \int_0^1 |cf(x)| dx = |c| \int_0^1 |f(x)| dx = |c| \|f\|_1$ and $\|f + g\|_1 = \int_0^1 |f(x) + g(x)| dx \leq \int_0^1 (|f(x)| + |g(x)|) dx = \int_0^1 |f(x)| dx + \int_0^1 |g(x)| dx = \|f\|_1 + \|g\|_1$. Of course $\|f\|_1 \geq 0$. Note that also $|f|$ is continuous.

If $f \neq 0$ then $\exists c \in [0, 1]$ such that $|f(c)| \neq 0$. So there exists $\delta > 0$ such that $|f(x)| > \varepsilon$ for $|x - c| < \delta$ and $x \in [0, 1]$, and $\varepsilon > 0$. Thus

$$\int_0^1 |f(x)| dx \geq \frac{1}{2} \delta \varepsilon > 0.$$

This proves NVS1.

We need to recall the definition of metric spaces.

1.14. Definition. Let X be a set. A function $d : X \times X \rightarrow \mathbb{R}$ is a *metric* and (X, d) a *metric space* if the following holds for all $x, y, z \in X$:

DIS1: $d(x, y) \geq 0$, and $d(x, y) = 0 \iff x = y$

DIS2: $d(x, y) = d(y, x)$

DIS3: $d(x, z) \leq d(x, y) + d(y, z)$

Notions of Cauchy sequence and convergence are defined in metric spaces as in normed vector spaces. A metric space is *complete* if every Cauchy sequence in X converges to a point of X . Each normed vector space defines a metric space by

$$d(x, y) := |x - y|.$$

Obviously NVS1 implies DIS1. Also $d(x, y) = |x - y| = |(-1)| |x - y| = |y - x| = d(y, x)$ thus DIS2 holds, and DIS3 follows

$$d(x, z) = |x - z| = |x - y + y - z| \leq |x - y| + |y - z| = d(x, y) + d(y, z).$$

A complete normed vector space is called a *Banach space*.

1.15. Definition. Let $\mathcal{T}, \mathcal{T}'$ be topologies on a set X . If $\mathcal{T} \subset \mathcal{T}'$ then \mathcal{T} is *coarser* than \mathcal{T}' and \mathcal{T}' is *finer* than \mathcal{T} .

For each space X the topology $\{\emptyset, X\}$ is the coarsest and the discrete topology is the finest.

1.16. Proposition. $\mathcal{T} \subset \mathcal{T}' \iff \forall x \in U \in \mathcal{T} \exists U' \in \mathcal{T}'$ such that $x \in U' \subset U$.

Proof. \implies : Choose $U' = U$. \impliedby : Given $U \in \mathcal{T}$, for each $x \in U$ choose $U'_x \in \mathcal{T}'$ such that $x \in U'_x \subset U$. Then

$$U = \cup_{x \in U} U'_x \in \mathcal{T}'$$

by TOP3 for \mathcal{T}' . ■

1.17. Proposition. Two norms $|\cdot|_1$ and $|\cdot|_2$ on a vector space E give rise to the same topology $\iff \exists C_1, C_2 > 0$ such that

$$C_1|x|_1 \leq |x|_2 \leq C_2|x|_1.$$

(The two norms are called equivalent in this case.)

Proof. Let \mathcal{T}_i be the topologies induced by $|\cdot|_i$ and let $B_r^i(x)$ denote the corresponding balls. Let $U \in \mathcal{T}_1$ and $x \in U$. Then there exists $r > 0$ such that $B_r^1(x) \subset U$. Now let $y \in B_{rC_1}^2(x)$, so $|y - x| < rC_1 \implies C_1|y - x|_1 < rC_1 \implies |y - x|_1 < r$. Thus $x \in B_{rC_1}^2(x) \subset B_r^1(x)$, and $\mathcal{T}_1 \subset \mathcal{T}_2$ follows from the previous proposition. Another application of this proposition proves $\mathcal{T}_2 \subset \mathcal{T}_1$. ■

Note that for equivalent norms convergence and Cauchy properties are the same.

Given (X, \mathcal{T}) let $\mathcal{F} := \{X \setminus U : U \in \mathcal{T}\}$. Then the collection \mathcal{F} satisfies the following

CL1: $\emptyset, X \in \mathcal{F}$

CL2: $V_i \in \mathcal{F}$ for $i \in I$ and I any index set then $\cup_{i \in I} V_i \in \mathcal{F}$

CL3 $U, V \in \mathcal{F} \implies U \cup V \in \mathcal{F}$

This follows directly from de'Morgan's laws. The very same argument also shows:

Conversely, given a collection \mathcal{F} of subsets of X satisfying CL1-CL3 then $\mathcal{T} := \{X \setminus V : V \in \mathcal{F}\}$ defines a topology on X .

Let X be a topological space and $S \subset X$. Then $x \in X$ is *adherent to S* if every neighborhood of x contains a point of S . Also $x \in \partial S \iff$ every neighborhood of x contains points of S and of $X \setminus S$, ∂S is the *boundary of S* . Note that S is closed $\iff S \supset \partial S$ (*Proof.* \implies : $X \setminus S$ is open. Suppose $x \in \partial S \setminus S$. Then every neighborhood of x contains a point of S and $x \in X \setminus S$, which contradicts $X \setminus S$ open. \impliedby : Let $x \in X \setminus S$. Then $x \notin \partial S$ because $S \supset \partial S$. Thus there exists a neighborhood of x not containing a point of S , and $X \setminus S$ is open.)

$\bar{S} := S \cup \partial S$ is the *closure* of S and is closed.

$S \subset X$ is *dense* if $\bar{S} = X$. Note that this condition is equivalent to the following: For each $x \in X$, every neighborhood of x contains a point of S (Exercise).

1.18. Definition. Let X be a topological space and $S \subset X$. Define a topology on S by $V \subset S$ open $\iff \exists U \subset X$ open such that $V = U \cap S$. This topology is the *induced topology*. The topological space S with the induced topology is also called a *subspace* of X .

1.19. Example. Let $S := [0, 1) \subset \mathbb{R} =: X$ and \mathbb{R} has the ordinary topology. Let $U = (\frac{1}{2}, 2)$, which is open in \mathbb{R} . Then $U \cap S = (\frac{1}{2}, 1] \subset S$ is open in S but not open in X .

Let $S \subset X$ be open. Then $U \subset S$ is open in the induced topology $\iff U \subset X$ is open. (*Proof.* \implies : Let $V \subset X$ open such that $U = V \cap S$, which is open in X . \impliedby : Let $U \subset S$ be open in X . Then $U = U \cap S$ is open in the induced topology.)

1.20. Definition. Let \mathcal{T} be a topology on X . Then a subset $\mathcal{B} \subset \mathcal{T}$ is a *base for the topology \mathcal{T}* if each $U \in \mathcal{T}$ is a union of elements of \mathcal{B} .

A base satisfies the following properties:

B1: $x \in X \implies \exists B \in \mathcal{B}$ such that $x \in B$.

B2: $B, B' \in \mathcal{B}$ and $x \in B \cap B'$. Then there exists $B'' \in \mathcal{B}$ such that $x \in B'' \subset B \cap B'$.

B1 follows because X is a union of base sets. For B2 let $x \in B \cap B' \in \mathcal{T}$, and write $B \cap B'$ as a union of base sets.

Conversely if \mathcal{B} is a collection of subsets of X satisfying B1 and B2. Then we can define the *topology generated by \mathcal{B}* by $U \in \mathcal{T} \iff U$ is a union of elements of \mathcal{B} . To prove e. g. TOP2 consider $U, V \in \mathcal{T}$ and let $x \in U \cap V$. Then $x \in B \subset U$ and $x \in V \subset V$ for base sets B, B' . Thus $x \in B \cap B' \subset U \cap V$. Then find B''_x according to B2 such that $x \in B''_x \subset B \cap B'$. Then write $U \cap V$ as the union of the B''_x .

1.21. Example. If E is a normed vector space or metric space then $\mathcal{B} := \{B_r(v) : v \in E, r > 0\}$ is a base for the topology of E . Note that $r > 0$ could be replaced by $r \in \mathbb{Q}$ and $r > 0$.

1.22. Definition. Let X, Y be topological spaces. A map $f : X \rightarrow Y$ is *continuous* if $f^{-1}(V) \subset X$ is open for all $V \subset Y$ open.

This can be equivalently stated by closed sets. In metric spaces the definition is equivalent to the usual $\varepsilon\delta$ -definition respectively the condition that f *commutes with limits on sequences*.

A map $f : X \rightarrow Y$ with continuous inverse is a *homeomorphism* or *topological isomorphism*.

Let X be a set and $\{Y_i\}_{i \in I}$ be a collection of topological spaces. Let $\mathcal{W} = \{f_i : X \rightarrow Y_i\}$ be a collection of functions. Then the collection

$$\mathcal{B} := \{U : U \text{ is finite intersection of sets } f_i^{-1}(U_i) \text{ for open } U_i \subset Y_i\}.$$

is the base of a topology on X , called the *weak topology determined by \mathcal{W}* . B2

is satisfied by construction because an intersection of two base sets is again a finite intersection of sets of the form $f_i^{-1}(U_i)$ and thus a base set.

1.23. Example. Let $\{X_i\}_{i \in I}$ be a family of topological spaces. Then $X := \prod_{i \in I} X_i$ (this is the set of maps from I into $\cup_{i \in I} X_i$ such that i maps into X_i) has the *product topology* defined as the weak topology determined by the projections

$$\pi_i : X \rightarrow X_i$$

mapping (x_i) to $x_i \in X_i$. Thus a set $U \subset X$ is open $\iff \forall x \in U \exists i_1, \dots, i_n \in I$ and open sets $U_{i_1}, U_{i_2}, \dots, U_{i_n}$ such that $x \in U_{i_1} \times \dots \times U_{i_n} \times \prod_{i \neq i_k} X_i \subset U$ (the order of the factors got changed to have reasonable notation). Note that sets of the last form $\prod_{i \in U_i} X_i$ where $U_i = X$ for almost all i and U_i is open in X_i , are a base of the product topology. Also note that in the case I finite a subset $U \subset X_1 \times \dots \times X_n$ is open if and only if for each $x = (x_1, \dots, x_n) \in U$ there exist open sets $U_i \subset X_i$ such that $x_i \in U_i$ for $i = 1, \dots, n$ and $U_1 \times \dots \times U_n \subset U$.

1.24. Subexample. (a) Let $X_i = \mathbb{R}$ for $I = \{1, \dots, n\}$. Then the product topology on \mathbb{R}^n is the topology induced from the norm $|x| = \max(x_i)$. In fact the base is given by products of open sets in \mathbb{R} . Note that $(U_1 \times U_2) \cap (V_1 \times V_2) = (U_1 \cap V_1) \times (U_2 \cap V_2)$.

(b) Let E_i be normed vector spaces with norms $|\cdot|_i$ for $i = 1, 2$. So E_1, E_2 have topologies defined from the norms, and $E_1 \times E_2$ has the product topology. But $E_1 \times E_2$ is a vector space with addition of vectors defined by $(v_1, v_2) + (w_1, w_2) := (v_1 + w_1, v_2 + w_2)$ for $v_i \in E_i$ and $i = 1, 2$, and $a(v_1, v_2) = (av_1, av_2)$. Moreover, $|(v_1, v_2)| := \max(|v_1|_1, |v_2|_2)$ defines a norm on $E_1 \times E_2$, which induces the product topology (Homework Problem 6)

1.25. Definition. Let X be a set. A collection of subsets $S_\alpha, \alpha \in A$ any index set, is called a *covering* of X if $\cup_{\alpha \in A} S_\alpha = X$. It is an *open covering* if X is a topological space and all S_α are open. A *subcovering* is a collection $\{S_\beta\}_{\beta \in B}$ with $B \subset A$. A finite subcovering is of the form $\{S_{\alpha_1}, \dots, S_{\alpha_n}\}$.

1.26. Definition. A topological space X is compact $:\iff$ every open covering has a finite subcovering.

1.27. Proposition. $f : X \rightarrow Y$ continuous, X compact $\implies f(X) \subset Y$ compact. ■

1.28. Proposition. A closed subspace of a compact space is compact. ■

The proofs of the above statements can be taken word by word from the corre-

sponding ones for metric spaces as they can be found e. g. Rudin's Principles of Mathematical Analysis.

A topological space X is *Hausdorff* if for all $x, y \in X$ with $x \neq y \exists U, V \in \mathcal{T}$ such that $x \in U, y \in V$ and $U \cap V = \emptyset$.

1.29. Example. Each metric space is Hausdorff because $d(x, y) = r > 0$ for $x \neq y$, and thus $B_{r/2}(x) \cap B_{r/2}(y) = \emptyset$.

1.30. Proposition. *A compact subspace of a Hausdorff space is closed.*

Proof. Let $x \in X \setminus S$. For $y \in S$ find open U_y, V_y such that $U_y \cap V_y = \emptyset$ and $x \in U_y, y \in V_y$. Then the open covering $\{V_y\}_{y \in S}$ has a finite subcovering $\{V_{y_1}, \dots, V_{y_n}\}$. Then $x \in U_{y_1} \cap \dots \cap U_{y_n} \in \mathcal{T}$ and $U_{y_1} \cap \dots \cap U_{y_n} \subset X \setminus S$, which thus is open and S is closed. ■

1.31. Definition. A topological space X is *sequentially compact* if every sequence $\{x_n\}$ in X has an *accumulation point* in X (i. e. $\exists c \in X$ such that for every open neighborhood of c there exist infinitely many n such that $x_n \in U \iff \{x_n\}$ has a subsequence contained in U).

1.32. Remarks. (a) X is sequentially compact \iff every infinite $S \subset X$ has a *point of accumulation* in X (Here $c \in X$ is a point of accumulation of $S \subset X$ if for every open neighborhood U of c , $|U \cap S|$ is infinite). *Proof:* \implies : Choose a sequence in S with distinct terms, which has an accumulation point c in X . Then c is also a point of accumulation of S in X . \impliedby : If a sequence $\{x_n\}$ in X has only finitely many different terms then for some N , $x_N = c$ has $x_n = x_N$ for infinitely many n and c is accumulation point of the sequence. If the sequence has infinitely many terms then let S be the set of its terms and a point of accumulation of S is an accumulation point of the sequence. (*Warning:* points of accumulation in topological spaces in our sense are often called ω -accumulation points in the literature)

1.33. Proposition. *compact \implies sequentially compact.*

Proof. Suppose $S \subset X$ is infinite with no point of accumulation in X but X compact. Then $\forall x \in X \exists$ open $U_x \ni x$ such that $|U_x \cap S|$ is finite. Then $\{U_x\}_{x \in X}$ is an open covering of X with finite subcovering U_{x_1}, \dots, U_{x_n} , and $S = (U_{x_1} \cup \dots \cup U_{x_n}) \cap S$ is finite, which is a contradiction. ■

1.34. Definition. A subset $S \subset X$ of a metric space X is *totally bounded* $:\iff \forall r > 0 \exists$ a finite number of balls of radius r covering S .

1.35. Examples. (a) A totally bounded metric space (X, d) is bounded: Let $r = 1$ and let X be covered by n balls of radius 1 with centers $\{x_1, \dots, x_n\}$. Let $d := \max_{i,j} d(x_i, x_j)$. Then the diameter of X is $\leq d + 2$ by the triangle inequality. ($\text{diam}(X) := \sup_{x,y \in X} d(x, y)$).

(b) A bounded subset of \mathbb{R}^n is totally bounded. In fact the closure of this set is closed and bounded and thus compact by the known theorem (see also below), But then the next result shows that S is totally bounded.

(c) An infinite discrete metric space ($d(x, y) = 1$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$) is bounded but not totally bounded. It is not possible to cover X by finitely many balls of radius $\frac{1}{2}$.

(d) Let $S_1 \subset l^1$ be the unit sphere from homework problem 2, (b), which is complete as a closed subspace of a complete space (see the homework problem). Then S_1 is bounded but not totally bounded. In fact, if S_1 were also totally bounded it would be a compact metric space by the theorem below, which it is not by the solution of the homework problem.

1.36. Theorem. *Let X be a metric space and $S \subset X$. Then the following are equivalent:*

- (i) S is compact.
- (ii) S is sequentially compact.
- (iii) S is complete and totally bounded.

Of course the last theorem applies to subsets of normed vector spaces, which is the usual setting for us.

Proof. (i) \implies (ii) is the previous proposition. (ii) \implies (iii): Let S be sequentially compact and $\{x_n\}$ be Cauchy in S . So the sequence $\{x_n\}$ in S has an accumulation point v in S . Recall that a Cauchy-sequence $\{x_n\}$ with a subsequence $\{x_{n_k}\}$ converging to $v \in S$, converges to $v \in S$: Given $\varepsilon > 0$ let N be such that $|x_n - x_m| < \frac{\varepsilon}{2}$ for $n, m > N$ and let K be such that $|x_{n_k} - v| < \varepsilon$ for $k > K$. Choose k such that $n_k > N$. Then for $n > N$ we have

$$|x_n - x| \leq |x_n - x_{n_k}| + |x_{n_k} - x| < \varepsilon.$$

Thus S is complete. Suppose S is not totally bounded. Let $r > 0$ be such that S cannot be covered by any finite set of balls of radius r . Let $x_1 \in S$ and B_1 the open ball of radius r and center x_1 . Since S is not contained in B_1 (otherwise B_1 alone would cover S) there exists $x_2 \in S, x_2 \notin B_1$. Let B_1, \dots, B_n be balls of radius r and center x_1, \dots, x_n such that $x_{k+1} \notin B_1 \cup \dots \cup B_k$ for $k = 1, \dots, n-1$. Then we can find $x_{n+1} \notin B_1 \cup \dots \cup B_n$ and define B_{n+1} the ball of radius r with center x_{n+1} . Let v be a point of accumulation of

$\{x_n\}$. Then there exist $k > m$ positive integers such that $|x_k - v| < \frac{r}{2}$ and $|x_m - v| < \frac{r}{2}$, which implies $|x_k - x_m| < r$, or $x_k \in B_m$, which contradicts the construction of the sequence of balls. (Remark: We can use notation $|\cdot|$ instead of metric notation because of Homework Problem 3.) (iii) \implies (i): Suppose the open covering $\{U_i\}$ of S has no finite subcovering. We know S is covered by a finite number of closed balls of radius $\frac{1}{2}$. So there exists at least one of those balls C_1 of radius $\frac{1}{2}$ such that $C_1 \cap S$ is not covered by any finite number of the U_i (otherwise collect open sets necessary for all the balls to cover S by finitely many U_i). Let $x_1 \in C_1 \cap S$. Suppose we have constructed a sequence of closed balls C_1, C_2, \dots, C_n with radius of C_n equal to $\frac{1}{2^n}$, $C_n \cap C_{n-1} \neq \emptyset$, and $x_n \in C_n \cap S$ such that $C_n \cap S$ is not covered by a finite number of sets U_i . Note that S can be covered by a finite number of balls of radius $\frac{1}{2^{n+1}}$ and so can $S \cap C_n$. Then there has to be a closed ball C_{n+1} of radius $\frac{1}{2^{n+1}}$ such that $C_{n+1} \cap S$ cannot be covered by finitely many U_i and $C_n \cap C_{n+1} \neq \emptyset$. Let $x_{n+1} \in C_{n+1} \cap S$. Then $\{x_n\}$ is a Cauchy sequence in the complete space S and thus converges to $x \in S$. (Note that $|x_n - x_{n-1}| \leq \frac{1}{2^{n-3}}$ for all $n \geq 4$. Thus for $m > n \geq 4$: $|x_m - x_n| \leq |x_m - x_{m-1}| + \dots + |x_{n+1} - x_n| \leq \frac{1}{2^{m-3}} + \dots + \frac{1}{2^{n-2}} \leq \frac{1}{2^{m-3}} \sum_{k=0}^{\infty} \frac{1}{2^k} = \frac{1}{2^{m-4}}$, which can be made arbitrarily small for m sufficiently large, because we choose N and $m > n > N$ for a given $\varepsilon > 0$.) But $x \in U_{i_0}$ and $U_{i_0} \supset C_n$ for all n sufficiently large. (Use that there is a ball of radius $\delta > 0$ with center x contained in U_{i_0} and that the x_n for large n are all contained in the ball with half radius, and the diameter of C_n will be less than a quarter of this radius for n large.) But those C_n then are covered by just a *single* U_{i_0} , which is a contradiction. ■

1.37. Definition. A subset of a topological space is *relatively compact* if its closure is compact.

1.38. Corollary. Let S be a subset of a complete normed vector space. Then S totally bounded $\implies S$ is relatively compact.

Proof. \bar{S} is complete. (Closed subsets of complete metric spaces are complete because the limit of a Cauchy sequence in the subset exists in the ambient space and by completeness already lies in the subset.) Let $r > 0$ and $S \subset \cup_{i=1}^n B_{r/2}(x_i)$. Then $\bar{S} \subset \overline{\cup_{i=1}^n B_{r/2}(x_i)} = \cup_{i=1}^n \overline{B_{r/2}(x_i)} \subset \cup_{i=1}^n B_r(x_i)$ and thus \bar{S} is totally bounded. Thus \bar{S} is compact by the previous theorem and S is relatively compact. ■

Now it easy to prove that for metric spaces X, Y and compact subsets $S \subset X$ and $T \subset Y$ that $S \times T \subset X \times Y$ is compact. In fact, for a sequence $\{z_n =$

(x_n, y_n) in $S \times T$ one can first find a subsequence $\{x_{n_i}\}$ converging to $a \in S$ and then a subsequence $\{y_{n_{i_k}}\}$ converging to $b \in T$ such that the corresponding subsequence $\{z_{n_{i_k}}\}$ converges to some point $(a, b) \in S \times T$. By induction it follows from this that *finite* products of compact metric spaces are compact. In particular any cube $I^n \subset \mathbb{R}^n$ with $I := [-a, a]$ for some $a > 0$ is compact. (Note that to show compactness of I we can use that every sequence in I has a monotonic subsequence, which thus converges by the completeness axiom for \mathbb{R} .) This leads to the following famous result:

1.39. Theorem. *A subset S of \mathbb{R}^n is compact if and only if it is closed and bounded.*

Proof. \Leftarrow : $S \subset I^n$ for some $I = [-a, a]$, $a > 0$, and thus is closed subset of a compact space and thus compact. \Rightarrow : S is closed because \mathbb{R}^n is Hausdorff (like any other metric space). If it is not bounded construct a sequence $\{x_n\}$ in S such that $\forall M \exists N$ such that $|x_n| > M$ for $n > N$. Then $\{x_n\}$ has no convergent subsequence, and thus the sequence has no accumulation point in S . ■

1.40. Corollary. *All norms on \mathbb{R}^n are equivalent.*

Proof. Let $\|x\| := \max|x_i|$ be the supremum norm and $|\cdot|$ be any norm. Let $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ be the i -th unit vector in \mathbb{R}^n such that $x = \sum_{i=1}^n x_i e_i$. Then

$$|x| \leq |x_1| |e_1| + \dots + |x_n| |e_n| \leq C \|x\|,$$

where $C := n \max(|e_i|)$. It follows that

$$\|x\| - \|y\| \leq |x - y| \leq C \|x - y\|.$$

Thus the function $|\cdot| : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous with respect to $\|\cdot\|$ on \mathbb{R}^n (and the usual absolute value metric on \mathbb{R} . Note that in $\|x\| - \|y\|$ the inner $|\cdot|$ are the norm on \mathbb{R}^n while the outer $|\cdot|$ is the absolute value on \mathbb{R} .) Now consider

$$S_1 := \{x \in \mathbb{R}^n : \|x\| = 1\},$$

which is closed and bounded and thus compact. So the restriction of $|\cdot|$ to S_1 has a minimum taken at some vector $v \in S_1$. Thus for all $x \in \mathbb{R}^n$:

$$\left| \frac{x}{\|x\|} \right| \geq |v|,$$

which implies

$$|v| \|x\| \leq |x|.$$

So we get

$$|v| \|x\| \leq |x| \leq C\|x\|,$$

which proves that the two norms are equivalent. ■

1.41. Definition. The elements v_1, \dots, v_n of some vector space E are *linearly independent* if

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = 0 \implies \lambda_1 = \lambda_2 = \dots = \lambda_n = 0$$

for all $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ (respectively $\in \mathbb{C}$ in the case of a complex vector space).

Note that if v_1, \dots, v_n are *linearly dependent* then the equation above has a nontrivial solution. If e. g. $\lambda_1 = 0$ then

$$v_1 = -\frac{\lambda_2}{\lambda_1} v_2 - \dots - \frac{\lambda_n}{\lambda_1} v_n,$$

and v_1 is a *linear combination* of the elements v_2, \dots, v_n .

1.42. Definition. The vector space E has *dimension* n if there are n linearly independent vectors, but any $n + 1$ vectors are linearly dependent. The vector space E is infinite dimensional if for every positive integer n there exist n linearly independent vectors.

1.43. Remark. Let E have dimension n and let v_1, \dots, v_n be n linearly independent vectors. For each $w \in E$ then we can find scalars $\lambda_1, \dots, \lambda_{n+1}$, not all 0, such that

$$\lambda_1 v_1 + \dots + \lambda_n v_n + \lambda_{n+1} w = 0.$$

Because $\lambda_{n+1} \neq 0$ (otherwise the elements v_1, \dots, v_n would not be linearly independent),

$$w = \sum_{i=1}^n \alpha_i v_i, \quad \alpha_i = -\frac{\lambda_i}{\lambda_{n+1}}$$

Note that this presentation is unique because of the linear independence. We say that the elements v_1, \dots, v_n *span* the vector space E and form an *algebraic basis*. Of course there are many possible bases of a vector space. Also, given a subset S of a vector space E call $\text{lin}(S) := \{v = \sum_{i=1}^r \lambda_i v_i : v_i \in S, \lambda_i \text{ scalars and } r \geq 0 \text{ integer}\}$ the subspace *generated or spanned by* S . This applies in particular to $S = \{v_1, \dots, v_n\}$ a finite set of vectors.

1.44. Definition. A map $f : E \rightarrow F$ between vector spaces E, F is *linear* if

$$f(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 f(v_1) + \lambda_2 f(v_2)$$

for all $v_1, v_2 \in E$ and scalars λ_1, λ_2 . If f is one-to-one onto then f is an *isomorphism* of vector spaces.

Note that for f linear, $f(0) = f(0 \cdot v) = 0 \cdot f(v) = 0$ because $0 \cdot v = 0$ in all vector spaces. ($0 \cdot v = (0+0)v = 0 \cdot v + 0 \cdot v$ by VS5, and now subtract $0 \cdot v$ on both sides.) Also note that compositions of linear maps are linear: $(f \circ g)(\lambda_1 v_1 + \lambda_2 v_2) = f(\lambda_1 g(v_1) + \lambda_2 g(v_2)) = \lambda_1 (f(g(v_1))) + \lambda_2 (f(g(v_2))) = \lambda_1 (f \circ g)(v_1) + \lambda_2 (f \circ g)(v_2)$. Note that linear maps f map subspaces to subspaces, and $f(\text{lin}(S)) = \text{lin}(f(S))$.

1.45. Example. Linear maps $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ are all represented in the form $\mathbb{R}^2 \ni x \mapsto Ax \in \mathbb{R}^2$, where A is a 2×2 -matrix with entries in the scalars. For example multiplication with a diagonal matrix maps the first basis vector $e_1 = (1, 0)$ to ae_1 and the second basis vector $e_2 = (0, 1)$ to be_2 . It maps a circle with center 0 onto an ellipse with center 0.

1.46. Theorem. *Given a real vector space E and a basis v_1, \dots, v_n then the mapping*

$$v = \sum_{i=1}^n \lambda_i v_i \mapsto (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$$

defines an isomorphism of vector spaces $E \rightarrow \mathbb{R}^n$.

Proof. The map is easily seen to be linear, and is well-defined because v_1, \dots, v_n is a basis, and has the inverse map

$$(\lambda_1, \dots, \lambda_n) \mapsto \sum_{i=1}^n \lambda_i v_i, \quad \blacksquare$$

It follows that the previous results on compactness and equivalence of norms apply in any finite dimensional vector space. Use that a norm $|\cdot|$ on a vector space E of dimension n induces a norm $|\cdot|_E$ on \mathbb{R}^n by defining for $x \in \mathbb{R}^n$, $|x|_E := |f(x)|$, where $f : \mathbb{R}^n \rightarrow E$ is a vector space isomorphism. Conversely each norm on \mathbb{R}^n (and isomorphism with E) defines a norm on E .

1.47. Proposition. *Each finite dimensional subspace of a vector space E is closed.*

Proof. Let $F \subset E$ be a finite dimensional subspace. Then all norm functions on F are equivalent, and \mathbb{R}^n is complete thus F is complete in any norm. But a complete subset of a metric space is closed (A sequence $\{x_n\}$ in a complete subset of a metric space X , which converges in X , is a Cauchy sequence in X and thus in S , and so the limit is in S .) \blacksquare

1.48. Definition. A topological space X is *locally compact* \iff every point has a compact neighborhood.

1.49. Examples. (a) \mathbb{R}^n is locally compact because each point has a compact ball neighborhood. Thus each finite dimensional vector space is locally compact.

(b) A normed vector space is locally compact \iff the closed unit ball $\{x \in E : |x| \leq 1\}$ is compact. *Proof.* \implies : Suppose E is locally compact and $x \in E$. Then x has a compact neighborhood, which will contain a closed ball neighborhood. This closed ball is compact as a closed subset of a compact space. Then by translation and multiplication by a positive scalar we can map this ball continuously onto the unit ball. Because continuous images of compact sets are compact, the unit ball is compact. \impliedby : The closed unit ball with center a point x is compact (being a translate of the unit ball), and thus x has a compact neighborhood.

1.50. Corollary (F. Riesz) *A normed vector space is locally compact if and only if it is finite dimensional.*

Proof. \impliedby : This is Examples (a) above. \implies : Let $B \subset E$ be the compact unit ball. By the theorem above we can cover B by a finite number of open balls of radius $\frac{1}{2}$ with centers v_1, \dots, v_n . Let $F := \text{lin}\{v_1, \dots, v_n\}$ be the subspace generated by $\{v_1, \dots, v_n\}$. *Claim.* $F = E$. $F \subset E$ has finite dimension and thus is closed in E . Let $w \in E$ and $w \notin F$ and $d(w, F) = \inf_{v \in F} |v - w|$. Consider $r > 0$ such that for $C := \overline{B_r(w)}$ we have $C \cap F \neq \emptyset$, e. g. $r = |w - v_1|$. Then $d(w, F) = \inf_{v \in F \cap C} |v - w|$. (This is clear because $|v - w| > r$ for all $w \notin C$.) Because $F \cap C$ is a closed subset of a compact space it is compact, and the continuous function $C \cap F \ni v \mapsto |v - w| \in \mathbb{R}$ attains a minimum on F at some $w \neq z \in F$. Because $z - w \neq 0$, we have $\frac{z-w}{|z-w|} \in B$ and can find some i such that

$$\left| \frac{w - z}{|w - z|} - v_i \right| < \frac{1}{2} \implies |w - z - |w - z|v_i| < \frac{|w - z|}{2}$$

But $z + |w - z|v_i \in F$, and so by the choice of z , the left hand side of the last inequality is $\geq |w - z|$, which is a contradiction. ■

Chapter 2

Banach Spaces

For normed vector spaces E, F let

$$L(E, F) := \{\lambda : E \rightarrow F \text{ linear and continuous}\},$$

which is a vector space by pointwise addition, and using that the sums and multiples by scalars of linear maps are linear, and similarly for continuous maps. We want to define a norm on $L(E, F)$.

2.1. Proposition. *A linear map $\lambda : E \rightarrow F$ is continuous $\iff \exists C > 0$ such that*

$$|\lambda(x)| \leq C|x|$$

for all $x \in E$.

Proof. \Leftarrow : If C exists then

$$|\lambda(x) - \lambda(y)| = |\lambda(x - y)| \leq C|x - y|,$$

and so for given $\varepsilon > 0$ we can choose *uniformly on E* , $\delta := \varepsilon/C$. \Rightarrow : λ is in particular continuous at 0, and so there exists $\delta > 0$ such that $|x| \leq \delta \implies |\lambda(x)| < 1$. Hence for each $0 \neq x \in E$ we get

$$\left| \lambda \left(\frac{\delta x}{|x|} \right) \right| < 1$$

because $|\delta \frac{x}{|x|}| \leq \delta$. Thus by NVS2:

$$|\lambda(x)| < \frac{1}{\delta}|x| = C|x|$$

for all $x \neq 0$ with $C := 1/\delta$. But the inequality holds trivially for $x = 0$. ■

The number $C > 0$ is a *bound* for λ . Define $|\lambda| \in \mathbb{R}$ to be the greatest lower bound of all numbers $C > 0$ such that $|\lambda(x)| \leq C|x|$ for all $x \in E$. Note that for a general map $E \rightarrow F$ this infimum might not exist (if no such number C exists). It does exist for all continuous linear maps, which therefore are also called *bounded*. **Careful:** The notion of boundedness of a *linear map* is not the usual notion of boundedness of a map into a metric space.

2.2. Examples. (a) Let E be a finite dimensional vector space with basis e_1, \dots, e_n . Then each linear map $\lambda : E \rightarrow F$ for F a normed vector space is bounded. In fact we can assume that the norm on E is given by $|x| = \max |x_i|$ for $x = \sum_{i=1}^n x_i e_i$. (This is the usual maximum norm on \mathbb{R}^n transported to E using the standard isomorphism between E and \mathbb{R}^n defined by the basis.)

Then

$$|f(x)| \leq \sum_{i=1}^n |x_i| |f(e_i)| \leq C|x|,$$

where $C := n \max_{1 \leq i \leq n} |f(e_i)|$. Thus for E finite dimensional $L(E, F)$ is the space of linear maps.

(b) Let $E = F = \mathbb{R}^n$ with basis e_1, \dots, e_n and $\lambda : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Then λ is represented by multiplication with an $n \times n$ real matrix. Let \mathbb{R}^n be considered as a normed vector space with ℓ^1 -norm $|x| := \sum_{i=1}^n |x_i|$ for $x = (x_1, \dots, x_n) = \sum_{i=1}^n x_i e_i$. Note that $\lambda(e_i)$ is the i -th column of the matrix A . Then $|A| := |\lambda| = \max_i |\lambda(e_i)|$ is a *matrix norm* on the space $L(\mathbb{R}^n, \mathbb{R}^n) = \mathbb{R}^{n^2}$. In fact $|\lambda(x)| = |\lambda(\sum x_i e_i)| = |\sum x_i \lambda(e_i)| \leq \sum_i |x_i| |\lambda(e_i)| \leq |x| \max_i |\lambda(e_i)|$, so $|\lambda| \leq \max_i |\lambda(e_i)|$. But $|e_i| = 1$ and so $|\lambda(e_i)| \leq |\lambda|$ implies $|\lambda| \geq \max_i |\lambda(e_i)|$. Note that all norms on $L(E, E)$ are equivalent for a finite dimensional vector space E .

(c) Let $I = I_E : E \rightarrow E$ be the identity map $I(x) := x$ for all $x \in E$. Then $I(x) = |x|$ for all $x \in E$ and thus $|\lambda| = 1$.

2.3. Proposition.

$$|\lambda| = \sup_{x \in S(E)} |\lambda(x)|,$$

where $S(E) := \{x \in E : |x| = 1\}$ is the unit sphere in E , and

$$L(E, F) \ni \lambda \mapsto |\lambda| \in \mathbb{R}$$

defines a norm on $L(E, F)$ satisfying

$$|\mu \circ \lambda| \leq |\mu| |\lambda|$$

for $\lambda \in L(E, F)$ and $\mu \in L(F, G)$, where E, F, G are all normed vector spaces.

Proof. By definition $|\lambda| = \inf\{C > 0 : \forall x \in E, |\lambda(x)| \leq C|x|\}$. Because for $x \neq 0$,

$$|\lambda(x)| \leq C|x| \iff \lambda\left(\frac{x}{|x|}\right) \leq C,$$

we have that $|\lambda|$ is the greatest lower bound of all upper bounds of $\{|\lambda(x)| : x \in S(E)\}$, and thus the supremum of this set by the definition of the supremum. Now $|\lambda| \geq 0$, and $|\lambda| = 0$ implies that λ vanishes on $S(E)$ thus by linearity at all $x \neq 0$, and because $\lambda(0) = 0$, for all $x \in E$, and thus NVS1 holds. If $\lambda_1, \lambda_2 \in L(E, F)$ and $a \in \mathbb{R}$ then for all $x \in S(E)$

$$|(\lambda_1 + \lambda_2)(x)| = |\lambda_1(x) + \lambda_2(x)| \leq |\lambda_1(x)| + |\lambda_2(x)| \leq |\lambda_1| + |\lambda_2|,$$

and since $|\lambda_1 + \lambda_2|$ is the least upper bound of all $|\lambda_1(x) + \lambda_2(x)|$ for $x \in S(E)$ we get NVS3

$$|\lambda_1 + \lambda_2| \leq |\lambda_1| + |\lambda_2|.$$

Also $|a\lambda| = \sup_{x \in S(E)} |a\lambda(x)| = |a| \sup_{x \in S(E)} |\lambda(x)| = |a| |\lambda|$ implies NVS2. Finally, composition induces a map

$$L(E, F) \times L(F, G) \rightarrow L(E, G),$$

because compositions of continuous maps are continuous, and compositions of linear maps are linear, and

$$|(\mu \circ \lambda)(x)| \leq |\mu(\lambda(x))| \leq |\mu| |\lambda(x)| \leq |\lambda| |\mu| |x|,$$

and so $|\mu \circ \lambda| \leq |\mu| |\lambda|$ because $|\mu \circ \lambda|$ is the smallest number satisfying $|\mu \circ \lambda(x)| \leq C|x|$ for all $x \in E$. ■

2.4. Proposition. *If F is complete then $L(E, F)$ is complete.*

Proof. Let $\{\lambda_n\}$ be a Cauchy sequence in $L(E, F)$. Then for $x \in E$, $\{\lambda_n(x)\}$ is a Cauchy sequence in F , which converges to a vector denoted $\lambda(x) \in F$. This defines a map $\lambda : E \rightarrow F$, which is linear: Fix $x_1, x_2 \in E$ and a_1, a_2 scalars. Then for each $\varepsilon > 0$ there exists a positive integer n such that $|\lambda(a_1x_1 + a_2x_2) - a_1\lambda(x_1) - a_2\lambda(x_2)| = |\lambda(a_1x_1 + a_2x_2) - \lambda_n(a_1x_1 + a_2x_2) + a_1\lambda_n(x_1) + a_2\lambda_n(x_2) - a_1\lambda(x_1) - a_2\lambda(x_2)| \leq |\lambda(a_1x_1 + a_2x_2) - \lambda_n(a_1x_1 + a_2x_2)| + |a_1||\lambda(x_1) - \lambda_n(x_1)| + |a_2||\lambda(x_2) - \lambda_n(x_2)| \leq \varepsilon$. Because this is true for each $\varepsilon > 0$ it follows that λ is linear. (Note that we can assume $a_1, a_2 \neq 0$, and first find N_1, N_2, N_3 such that for $n > N_1 : |\lambda(a_1x_1 + a_2x_2) - \lambda_n(a_1x_1 + a_2x_2)| \leq \frac{\varepsilon}{3}$, for $n > N_2 : |\lambda(x_1) - \lambda_n(x_1)| \leq \frac{\varepsilon}{3|a_1|}$, and for $n > N_3 : |\lambda(x_2) - \lambda_n(x_2)| \leq \frac{\varepsilon}{3|a_2|}$.)

Then we can find $n > \max(N_1, N_2, N_3)$ such that all three inequalities hold.) Note that for m, n , $\lambda_n - \lambda_m \in L(E, F)$ and the inequality

$$||\lambda_n| - |\lambda_m|| \leq |\lambda_n - \lambda_m| < \varepsilon$$

for $m, n > N$ shows that $\{|\lambda_n|\}$ is a Cauchy sequence of real numbers, which has a limit $C := \lim |\lambda_n|$. From $\{\lambda_n\}$ Cauchy it follows that for each $\varepsilon > 0$ there exist N such that for all $m, n > N$ and for all $x \in E$ we have $|\lambda_m(x) - \lambda_n(x)| \leq |\lambda_m - \lambda_n||x| < \varepsilon|x|$. So, if we let $m \rightarrow \infty$ we get $|\lambda_n(x) - \lambda(x)| \leq \varepsilon|x|$ for all x . Then we calculate $|\lambda(x)| \leq |\lambda(x) - \lambda_n(x)| + |\lambda_n(x)| \leq \varepsilon|x| + |\lambda_n| |x| \leq (\varepsilon + |\lambda_n|)|x| \leq (C + 2\varepsilon)|x|$ for n sufficiently large, using $|\lambda_n| \rightarrow C$. Because this holds for each $\varepsilon > 0$ we get that λ is bounded and $|\lambda| \leq C$. Now for $x \in S(E)$ we have $|x| = 1$ and thus from $|\lambda_n(x) - \lambda(x)| \leq \varepsilon$ for all $x \in S(E)$ we get also

$$|\lambda - \lambda_n| = \sup_{x \in S(E)} |\lambda(x) - \lambda_n(x)| \leq \varepsilon,$$

and thus $|\lambda_n - \lambda| \rightarrow 0$. This implies $\lambda_n \rightarrow \lambda$ for $n \rightarrow \infty$ in $L(E, F)$. ■

Recall that a complete normed vector space is called a Banach space. Thus we have proved: F Banach $\implies L(E, F)$ Banach.

2.5. Definition. Let E, F, G be vector spaces. A map $\varphi : E \times F \rightarrow G$ such that for each $x \in E$ the map $F \ni y \mapsto \varphi(x, y) \in G$ is linear, and for each $y \in F$ the map $E \ni x \mapsto \varphi(x, y) \in G$ is linear, is called *bilinear*.

2.6. Example. Let $E = F = \mathbb{R}^n$ and $G = \mathbb{R}$. Given an $n \times n$ -matrix A the map $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $(x, y) \mapsto x^T A y$, where x is seen as a column vector such that x^T is the corresponding row vector, is bilinear. In analysis those matrices A appear as matrices of second order partial derivatives. We will see other examples soon when we discuss Hilbert spaces.

Note that $\varphi(0, y) = \varphi(x, 0) = 0$ for all $x \in E$ and $y \in F$ follows from the linearity as in the case of linear maps.

The set of bilinear maps is a vector space (as a set of maps into a vector space). As with linear maps, a bilinear map φ is continuous \iff there exists $C > 0$ such that $|\varphi(x, y)| \leq C|x| |y|$, where $E \times F$ has the norm $|(x, y)| := \max(|x|, |y|)$. *Proof.* \Leftarrow : Let $(x_0, y_0) \in E \times F$. Then $|\varphi(x, y) - \varphi(x_0, y_0)| = |\varphi(x, y) - \varphi(x_0, y) + \varphi(x_0, y) - \varphi(x_0, y_0)| \leq |\varphi(x, y) - \varphi(x_0, y)| + |\varphi(x_0, y) - \varphi(x_0, y_0)| = |\varphi(x - x_0, y)| + |\varphi(x_0, y - y_0)| \leq C|x - x_0| |y| + C|x_0| |y - y_0|$, we can conclude that for $|(x, x_0) - (y, y_0)| = |(x - x_0, y - y_0)| < \delta$, so $|x - x_0| < \delta$ and $|y - y_0| < \delta$, and thus $|y| \leq |y_0| + \delta$ that $|\varphi(x, y) - \varphi(x_0, y_0)| \leq C(\delta|y_0| + \delta^2 + |x_0|\delta) =$

$\delta C(|y_0| + |x_0| + \delta)$, which can be made arbitrarily small for δ small enough. Thus φ is continuous at (x_0, y_0) . \implies : Use the continuity of φ at 0 to find $\delta > 0$ such that $\max(|x|, |y|) < \delta \implies |\varphi(x, y)| < 1$. Thus for all $0 \neq x \in E$ and $0 \neq y \in F$ we get

$$\left| \varphi \left(\frac{\delta x}{|x|}, \frac{\delta y}{|y|} \right) \right| < 1,$$

which implies $|\varphi(x, y)| < \frac{1}{\delta^2} |x| |y|$ and thus we can take $C := \frac{1}{\delta^2}$. The greatest lower bound of such C defines a norm $\varphi \mapsto |\varphi|$ on the space $L(E, F; G)$ of bilinear continuous maps $E \times F \rightarrow G$, and is equal to

$$|\varphi| = \sup_{(x, y) \in S(E) \times S(F)} |\varphi(x, y)|.$$

This follows from bilinearity using the very same arguments as in the case of linear maps. Similarly, if G is complete it follows that $L(E, F; G)$ is complete: Given a Cauchy sequence $\{\varphi_n\}$ in $L(E, F; G)$ and $x \in E$, $y \in F$ consider the Cauchy sequence $\{\varphi_n(x, y)\}$ in G to construct a function $\varphi : E \times F \rightarrow G$. Bilinearity of this map then is proved just as in the case of linear maps, but now for each argument x and y . The rest of the argument, showing that φ is continuous and $\varphi_n \rightarrow \varphi$ in the norm on $L(E, F; G)$, follows very much the argument that $L(E, F)$ is complete for F complete.

2.7. Example. The *evaluation map* $\chi : L(E, F) \times E \rightarrow F$, $\chi(\lambda, x) := \lambda(x)$ is in $L(L(E, F), E; F)$ and has norm 1 because $|\chi| = \sup_{(\lambda, x) \in S(L(E, F)) \times S(E)} |\lambda(x)| = \sup_{\lambda \in S(L(E, F))} \sup_{x \in S(E)} |\lambda(x)| = \sup_{\lambda \in S(L(E, F))} |\lambda| = 1$ by the definition of the unit sphere in $L(E, F)$.

An element of $L(E, E)$ is an *endomorphism* or *operator*. An element $u \in L(E, F)$ is *invertible* if there exists $v \in L(F, E)$ such that $u \circ v = I_F$ and $v \circ u = I_E$. The set of invertible elements of $L(E, F)$ is denoted $Lis(E, F)$ and its elements are called *toplinear isomorphisms*, $Lis(E, E) =: Laut(E)$ is the set of toplinear *automorphisms*. A norm-preserving isomorphism u between Banach spaces E, F , i. e. $|u(x)| = |x|$ for all $x \in E$, is a *Banach isomorphism* or *isometry*.

2.8. Proposition. *There exists a natural toplinear isomorphism:*

$$L(E, L(F, G)) \rightarrow L(E, F; G)$$

Proof. For $\lambda \in L(E, L(F, G))$ define $\varphi_\lambda \in L(E, F; G)$ by

$$\varphi_\lambda(x, y) := \lambda(x)(y).$$

In fact, φ_λ is bilinear: for example in the first argument, $\varphi_\lambda(a_1x_1 + a_2x_2, y) = \lambda(a_1x_1 + a_2x_2)(y) = (a_1\lambda(x_1) + a_2\lambda(x_2))(y) = a_1\lambda(x_1)(y) + a_2\lambda(x_2)(y) = a_1\varphi_\lambda(x_1, y) + a_2\varphi_\lambda(x_2, y)$. Also

$$|\varphi_\lambda(x, y)| = |\lambda(x)(y)| \leq |\lambda(x)| |y| \leq |\lambda| |x| |y|.$$

Thus φ_λ is bounded and $|\varphi_\lambda| \leq |\lambda|$. This last inequality also shows that the mapping $\lambda \mapsto \varphi_\lambda$ is continuous. Linearity of $\lambda \mapsto \varphi_\lambda$ follows because for a_1, a_2 scalars and $\lambda_1, \lambda_2 \in L(E, L(F, G))$ we have: $\varphi_{a_1\lambda_1 + a_2\lambda_2}(x, y) = (a_1\lambda_1 + a_2\lambda_2)(x)(y) = a_1\lambda_1(x)(y) + a_2\lambda_2(x)(y) = a_1\varphi_{\lambda_1}(x, y) + a_2\varphi_{\lambda_2}(x, y) = (a_1\varphi_{\lambda_1} + a_2\varphi_{\lambda_2})(x, y)$ for all $x \in E$ and $y \in F$, which means $\varphi_{a_1\lambda_1 + a_2\lambda_2} = a_1\varphi_{\lambda_1} + a_2\varphi_{\lambda_2}$. Conversely, for $\varphi \in L(E, F; G)$ define $\lambda_\varphi \in L(E, L(F, G))$ by

$$\lambda_\varphi(x)(y) := \varphi(x, y)$$

The linearity of λ_φ follows because for all $y \in F$, $\lambda_\varphi(a_1x_1 + a_2x_2)(y) = \varphi(a_1x_1 + a_2x_2, y) = a_1\varphi(x_1, y) + a_2\varphi(x_2, y) = a_1\lambda_\varphi(x_1)(y) + a_2\lambda_\varphi(x_2)(y) = (a_1\lambda_\varphi(x_1) + a_2\lambda_\varphi(x_2))(y)$. Note that also for each $x \in E$ the map $\lambda_\varphi(x) : F \rightarrow G$ is linear: $\lambda_\varphi(x)(b_1y_1 + b_2y_2) = \varphi(x, b_1y_1 + b_2y_2) = b_1\varphi(x, y_1) + b_2\varphi(x, y_2) = b_1\lambda_\varphi(x)(y_1) + b_2\lambda_\varphi(x)(y_2)$. Now for the linear map $\lambda_\varphi(x) : F \rightarrow G$ we have:

$$|\lambda_\varphi(x)(y)| = |\varphi(x, y)| \leq (|\varphi| |x|)|y|.$$

This shows that for each $x \in E$ the linear map $\lambda_\varphi(x)$ is bounded. The last inequality, by the very definition of $|\lambda_\varphi(x)|$, also shows that $|\lambda_\varphi(x)| \leq |\varphi| |x|$, and thus the map $\lambda_\varphi : E \rightarrow L(F, G)$ is bounded. Note that until now we only proved that $\varphi \mapsto \lambda_\varphi$ actually is a map $L(E, F; G) \rightarrow L(E, L(F, G))$. Now the last inequality also says $|\lambda_\varphi| \leq |\varphi|$ by definition of $|\lambda_\varphi|$. But the equation $|\lambda_\varphi| \leq |\varphi|$ also shows the continuity of $\varphi \mapsto \lambda_\varphi$. Linearity is a simple calculation: $\lambda_{a_1\varphi_1 + a_2\varphi_2} = a_1\lambda_{\varphi_1} + a_2\lambda_{\varphi_2}$ for all a_1, a_2 scalars and $\varphi_1, \varphi_2 \in L(E, F; G)$. This is easily calculated by application to x and then to y , and follows from the definition of the vector space structure on $L(E, F; G)$.

Finally note that $\lambda \mapsto \varphi_\lambda$ and $\varphi \mapsto \lambda_\varphi$ are inverse to each other, which means $\lambda_{\varphi_\lambda} = \lambda$ and $\varphi_{\lambda_\varphi} = \varphi$. Note that this is from the very definitions, e. g. $\lambda_{\varphi_\lambda}(x)(y) = \varphi_\lambda(x, y) = \lambda(x)(y)$ for all $x \in E$ and $y \in F$. ■

The above result says that the map $\lambda \mapsto \varphi_\lambda$ is an element of $Lis(L(E, L(F, G)), L(E, F; G))$.

2.9. Definition. If E_1, \dots, E_n are vector spaces, a *multilinear map*

$$\varphi : E_1 \times \dots \times E_n \rightarrow F,$$

is a map, which is linear in each variable.

As above, a multilinear map φ for E_1, \dots, E_n, F normed vector spaces is continuous (here $E_1 \times \dots \times E_n$ has the product topology, which is induced from the norm $|(x_1, \dots, x_n)| := \max_i |x_i|$) is continuous $\iff \exists C > 0$ such that

$$|\varphi(x_1, \dots, x_n)| \leq C|x_1| |x_2| \dots |x_n|$$

The following is immediate by induction from the previous propositions:

2.10. Proposition. *There is a natural isomorphism:*

$$L(E_1, L(E_2, \dots, L(E_n, F) \dots)) \rightarrow L(E_1, \dots, E_n; F),$$

and if F is complete then all these spaces are Banach spaces.

2.11. Definition. Let $E' := L(E, \mathbb{R})$ for a real normed vector space, respectively $E' := L(E, \mathbb{C})$ for a complex normed vector space, be the *dual space* of E or the space of *functionals* on E .

Note that these are Banach spaces, even though E might not be a Banach space. In the finite dimensional case functionals are used as *coordinates*. For example the functionals $\lambda_i : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $\lambda_i(x_1, \dots, x_n) = x_i$ for $i = 1, \dots, n$ can be used to *characterize* elements of \mathbb{R}^n : $x = y \iff \lambda_i(x) = \lambda_i(y)$ for $i = 1, \dots, n$. Note that by linearity this is equivalent to $\lambda_i(x) = 0$ for $i = 1, \dots, n \implies x = 0$. In fact, the linear map defined by $e_i \mapsto \lambda_i$ defines an isomorphism $\mathbb{R}^n \rightarrow (\mathbb{R}^n)'$. In particular in the general case, E and E' are isomorphic, with the isomorphism depending on a choice of basis. In infinite dimensions functionals are equally important but it is more difficult to prove a corresponding result.

We will need some set-theoretic preliminaries.

Recall that an *ordering* on a set S is a relation $x \leq y$ among **some** elements of S such that (i) $x \leq x$, (ii) $x \leq y$ and $y \leq x \implies x = y$, and (iii) $x \leq y$ and $y \leq z \implies x \leq z$. If for all $x, y \in S$ it is true that $x \leq y$ or $y \leq x$ (i. e. any two elements are *comparable*) then the ordering is called a *total ordering*. An *upper bound* of $T \subset S$, S an ordered set, is an element $b \in S$ such that $x \leq b$ for all $x \in T$. The set S is *inductively ordered* if every nonempty totally ordered subset has an upper bound. An element $m \in S$ such that $x \in S, x \geq m \implies x = m$ is *maximal* element of S . Also note that each subset of an ordered set is ordered by restriction of the relation.

2.12. Zorn's Lemma. *If $S \neq \emptyset$ is inductively ordered then there exists a maximal element in S . ■*

As a warm-up we first give an algebraic application.

2.13. Definition. Let E be a vector space. A subset $B \subset E$ is called a (*Hamel*) *basis* of E if each $x \in E$ can be written uniquely as a finite linear combination of elements of B , i. e. for each $x \in E$ there exist finitely many elements $v_1, \dots, v_r \in B$ and scalars c_1, \dots, c_r such that $x = \sum_{i=1}^r c_i v_i$ in a unique way.

It follows that the set B is linearly independent in the sense that every finite subset of it is linearly independent. (If a finite subset would be linearly dependent then 0 could be written as a linear combination in two different ways.) More precisely, B is a basis of $E \iff B$ is linearly independent and $\text{lin}(B) = E$.

2.14. Proposition. *Each vector space E has a basis.*

Proof. Let S be the set of linearly independent subsets of E with ordering given by inclusion. Let $T \subset S$ be a totally ordered. Then the set $D := \cup_{A \in T} A$ is also linearly independent. In fact, if D is linearly dependent consider $v_1, \dots, v_n \in D$ and scalars $c_i, i = 1, \dots, n$ such that $\sum c_i v_i = 0$. Now each v_i is contained in one of the sets $A_i \in T$. Since T is totally ordered, $A_i \subset A_j$ or $A_j \subset A_i$. Thus by induction one of the collection of finite sets $\{A_1, \dots, A_r\}$ contains all the others. Let A be this set so that $v_i \in A$ for all i . Since A is linearly independent, all c_i are 0. Thus D is linearly independent hence an element of S , and an upper bound of T by definition. By Zorn's lemma S contains a maximal element B , which is linearly independent by definition. Let $F := \text{lin}(B)$. If F is a proper subspace of E then we can find $v \in E, v \notin F$. Then $B \cup \{v\}$ is linearly independent. (If $B \cup \{v\}$ is linearly dependent then $\{v, v_1, \dots, v_r\}$ is linearly dependent, and we could write v as a linear combination of $v_1, \dots, v_r \in B$ and thus $v \in F$, which is a contradiction.) Thus $E = F$, and B is a basis. ■

2.15. Hahn-Banach theorem. *Let E be a real vector space with subspace F . Let $\lambda \in F' = L(F, \mathbb{R})$. Then λ can be extended to some element $\lambda^* \in E' = L(E, \mathbb{R})$ with the same norm, i. e. $\lambda^*|_F = \lambda$ and $|\lambda^*| = |\lambda|$.*

Proof. If $|\lambda| = 0$ then $\lambda = 0$ is extended to $\lambda^* = 0$ norm-preserving. So we can assume $|\lambda| \neq 0$. If $F = E$ then there is nothing to prove. So we can assume that $F \subset E$ is a proper inclusion. It suffices to extend $\frac{\lambda}{|\lambda|} \in F'$ to $\Lambda^* \in E'$ such that $|\Lambda^*| = 1$. Then $\lambda^* := |\lambda|\Lambda^* \in E'$ extends λ because the restriction

to F is $|\lambda| \frac{\lambda}{|\lambda|} = \lambda$ and $|\lambda^*| = |\lambda| |\Lambda^*| = |\lambda|$. So we can assume that $|\lambda| = 1$. Let $v \in E \setminus F$. Then $F + \mathbb{R}v = \{x + tv : x \in F, t \in \mathbb{R}\} \supset F$ is a proper inclusion and every element in $F + \mathbb{R}v$ has a unique representation in this form. (If $x + tv = x' + t'v$ with $x, x' \in F$ and $t, t' \in \mathbb{R}$ then $x - x' = (t' - t)v$. If $x \neq x'$ then $t \neq t'$ and $(t - t')v \in F$ implies $v \in F$ which is a contradiction to our assumption. Thus $x = x'$ and then $t = t'$.) Now for a real number $a \in \mathbb{R}$ the map $\lambda^* : F + \mathbb{R}v \rightarrow \mathbb{R}$ defined by

$$\lambda^*(x + tv) := \lambda(x) + ta$$

is linear: $\lambda^*(x + tv + x' + t'v) = \lambda(x + x') + (t + t')a = (\lambda(x) + ta) + (\lambda(x') + t'a) = \lambda^*(x + tv) + \lambda^*(x' + t'v)$ and $\lambda^*(c(x + tv)) = \lambda^*(cx + (ct)v) = \lambda(cx) + (ct)a = c(\lambda(x) + ta) = c\lambda^*(x + tv)$.

Claim: a can be chosen such that λ^* is bounded by 1. (This suffices because $|\lambda^*| \geq |\lambda|$ is clear because λ^* extends.) *Proof.* $|\lambda^*| = 1 \iff |\lambda^*(x + tv)| = |\lambda(x) + ta| \leq |x + tv|$ for all $x \in F$ and $t \in \mathbb{R}$. For $t \neq 0$ (otherwise there is nothing to prove) this is equivalent to $|t(\lambda(\frac{1}{t}x) + a)| \leq |t(\frac{1}{t}x + v)|$ and so by division by $|t|$ and noticing that $\frac{1}{t}x$ runs through all elements of F if x runs through all elements of F (F is a vector space) we have to find a such that for all $y \in F$

$$|\lambda(y) + a| \leq |y + v| \in \mathbb{R},$$

which is equivalent to $(a \in B_{|y+v|}(-\lambda(y)) \subset \mathbb{R})$:

$$-\lambda(y) - |y + v| \leq a \leq -\lambda(y) + |y + v|$$

for all $y \in F$.

So it suffices to show that the interval $[-\lambda(y) - |y + v|, -\lambda(y) + |y + v|]$ is nonempty. Now for all $y, z \in F$, $\lambda(y) - \lambda(z) \leq |\lambda(y) - \lambda(z)| = |\lambda(y - z)| \leq |y - z| = |(y + v) - (z + v)| \leq |y + v| + |z + v| \implies -\lambda(z) - |z + v| \leq -\lambda(y) + |y + v|$. Now for $y = z$ we get the inequality we need.

To conclude the proof we need a method to extend this process in an infinite dimensional setting. This is precisely what Zorn's lemma is giving us. Consider the set of pairs (G, λ^*) with $F \subset G \subset E$, G a subspace of E , and $\lambda^* \in G'$ satisfying $\lambda^*|_F = \lambda$ and $|\lambda^*| = |\lambda|$. Order such pairs by $(G_1, \lambda_1) \leq (G_2, \lambda_2)$ if G_1 is a subspace of G_2 and λ_2 is an extension of λ_1 . *Claim.* This defines an inductive ordering. *Proof.* Given a totally ordered set of pairs $\{G_i, \lambda_i\}$ let $G := \cup_i G_i$. Then G is a subspace: If $x_1, x_2 \in G$ then $x_1 \in G_i$ and $x_2 \in G_j$ and because $G_i \subset G_j$ or $G_j \subset G_i$ we can assume that both are in G_i , so all linear combinations $c_1x_1 + c_2x_2 \in G_i \subset G$. Now we define $\lambda^* \in G'$ as follows: If $x \in G$

then $x \in G_i$ for some i , so we can define $\lambda^*(x) = \lambda_i(x)$. This does not depend on i because if also $x \in G_j$ then $G_i \subset G_j$ or $G_j \subset G_i$ and λ_j and λ_i agree on $G_i \cap G_j$. Thus (G, λ^*) is an upper bound of the family $\{G_i, \lambda_i\}$. By Zorn's lemma there exists a maximal element, which we denote also (G, λ^*) , of the set of all pairs. Then $G = E$ because otherwise $G \neq E$ and we can find $v \in E \setminus G$ and apply the procedure of the first part of the theorem to extend to $G + \mathbb{R}v$, which would contradict maximality. ■

2.16. Remark. Note that if $C > 0$ is any bound for λ then because $|\lambda^*| = |\lambda| \leq C$, C is also a bound for λ^* . Thus λ extends such that bounds of λ are bounds of λ^* .

2.17. Corollary. *The Hahn-Banach theorem also holds for complex normed vector spaces.*

Proof. Given $\lambda \in L(E, \mathbb{C})$ for E a complex normed space and $F \subset E$ a complex subspace. Let $\varphi := \operatorname{Re}(\lambda)$ and $\psi := \operatorname{Im}(\lambda)$. Then for all $v \in F$: $\lambda(iv) = i\lambda(v) = i(\varphi(v) + i\psi(v)) = i\varphi(v) - \psi(v) = \varphi(iv) + i\psi(iv)$. Thus by comparing the real parts we get $\psi(v) = -\varphi(iv)$. We use this to extend λ . Let φ' be a real extension of φ to E , and define

$$\lambda'(v) := \varphi'(v) - i\varphi'(iv)$$

for $v \in E$. Then λ' is an extension of λ and is complex linear because for all $v \in E$:

$$\lambda'(iv) = \varphi'(iv) - i\varphi'(-v) = i(\varphi'(v) - i\varphi'(iv)) = i\lambda'(v).$$

Now for each $v \in E$ there exists $\alpha \in \mathbb{R}$ such that $|\lambda'(v)| = e^{i\alpha}\lambda'(v) = \lambda'(e^{i\alpha}v) = \operatorname{Re}(\lambda'(e^{i\alpha}v)) = \varphi'(e^{i\alpha}v) = |\varphi'(e^{i\alpha}v)| \leq |\varphi'| |e^{i\alpha}v| = |\varphi'| |v|$, which shows $|\lambda'| \leq |\varphi'|$, and $|\varphi'| \leq |\lambda'|$ because $|\operatorname{Re}(z)| \leq |z|$ for each complex number. The same argument works for λ and φ , and thus $|\lambda'| = |\varphi'| = |\varphi| = |\lambda|$. ■

2.18. Corollary. *If E is a normed vector space and $0 \neq v \in E$ then there exists $\lambda \in E'$ such that $\lambda(v) \neq 0$.*

Proof. In the real case apply the Hahn-Banach theorem to $F = \mathbb{R}v$ with $\lambda(v) = a$ and $0 \neq a \in \mathbb{R}$, and extend to E . The complex case is similar. ■

For each normed vector space E the dual E' is a Banach space, and so is the double dual $E'' := L(E', \mathbb{R})$ respectively $L(E', \mathbb{C})$. The normed space E and its double dual are naturally related.

2.19. Proposition. *The natural map $E \ni x \rightarrow f_x \in E'' = (E')'$ defined by $f_x(\lambda) := \lambda(x)$ is an injective linear map, which is norm preserving, i. e.*

$|x| = |f_x|$ for all $x \in E$.

Proof. $f_{a_1x_1+a_2x_2}(\lambda) = \lambda(a_1x_1 + a_2x_2) = a_1\lambda(x_1) + a_2\lambda(x_2) = a_1f_{x_1}(\lambda) + a_2f_{x_2}(\lambda)$ shows that $x \mapsto f_x$ is linear. To see $f_x \in (E)'$ we need linearity in λ , which is obvious, and the continuity of f_x . Now continuity of both f_x and $x \mapsto f_x$ follows from $|f_x(\lambda)| = |\lambda(x)| \leq |\lambda| |x| \implies |f_x| \leq |x|$. Let $\lambda \in E'$ be a functional with $\lambda(x) = |x|$ and $|\lambda| = 1$. This exists by the Hahn-Banach theorem by extending the function from $\mathbb{R}x$ to E . Then $|f_x(\lambda)| = |\lambda(x)| = |x| \leq |f_x| |\lambda| = |f_x|$ implies $|f_x| = |x|$. This shows that $x \mapsto f_x$ is norm preserving and injective ($f_x = 0 \implies |f_x| = |x| = 0 \implies x = 0$. Note that this argument always shows: norm-preserving implies injective). ■

2.20. Linear Extension theorem. *Let E be a normed vector space, F a subspace. Then the closure $\overline{F} \subset E$ is a subspace. If G is a **complete** normed vector space and $\lambda : F \rightarrow G$ is a continuous linear map then there exists a **unique** continuous linear extension $\overline{\lambda} : \overline{F} \rightarrow G$, and $|\overline{\lambda}| = |\lambda|$.*

Proof. If $x, y \in \overline{F}$ then there exist sequences $\{x_n\}$ and $\{y_n\}$ in F such that $x_n \rightarrow x$ and $y_n \rightarrow y$. Then $x_n + y_n \rightarrow x + y$, and for a scalar c , $cx_n \rightarrow cx$. Thus \overline{F} is a nonempty ($0 \in F$) subset of E , closed with respect to vector addition and multiplication by scalars, and thus is a subspace of E . Because the continuity of $\overline{\lambda}$ implies $\overline{\lambda}(x) = \lim \overline{\lambda}(x_n) = \lim \lambda(x_n)$, uniqueness is clear if existence is shown. Let $x \in \overline{F}$ with $x = \lim x_n$ and $x_n \in F$. Then

$$|\lambda(x_n) - \lambda(x_m)| = |\lambda(x_n - x_m)| \leq C|x_n - x_m|,$$

and $\{\lambda(x_n)\}$ is a Cauchy sequence in G with limit denoted $\overline{\lambda}(x) \in G$. If also $x = \lim x'_n$ with $x'_n \in F$ then $|x_n - x'_n| \rightarrow 0$ and thus $|\lambda(x_n) - \lambda(x'_n)| = |\lambda(x_n - x'_n)| \leq C|x_n - x'_n|$ shows that $\lim \lambda(x_n) = \lim \lambda(x'_n)$. Now for $y \in \overline{F}$ let $\{y_n\}$ be such that $y_n \rightarrow y$. Then we know that $x_n + y_n \rightarrow x + y$, and for a scalar c , $cx_n \rightarrow cx$. Thus $\overline{\lambda}(x+y) = \lim \lambda(x_n + y_n) = \lim \lambda(x_n) + \lim \lambda(y_n) = \overline{\lambda}(x) + \overline{\lambda}(y)$, and $\overline{\lambda}(cx) = \lim \lambda(cx_n) = \lim(c\lambda(x_n)) = c \lim \lambda(x_n) = c\overline{\lambda}(x)$. Thus $\overline{\lambda}$ is a linear extension of λ . Now $|\cdot| : G \rightarrow \mathbb{R}$ and $|\cdot| : E \rightarrow \mathbb{R}$ are continuous and thus commute with limits. This implies $|\overline{\lambda}(x)| = |\lim \lambda(x_n)| = \lim |\lambda(x_n)| \leq |\lambda| \lim |x_n| = |\lambda| |\lim x_n| = |\lambda| |x|$ for all $x \in \overline{F}$. Thus $|\overline{\lambda}| \leq |\lambda|$, and $|\overline{\lambda}| \geq |\lambda|$ is obvious because $\overline{\lambda}$ extends λ . ■

2.21. Definition. A *completion* of a normed vector space E is a pair (F, ψ) with F a Banach space and $\psi : E \rightarrow F$ a continuous linear (injective) map such that $\psi(E)$ is dense in F and $|\psi(x)| = |x|$ for all $x \in E$.

We will let (\overline{F}, φ) denote a fixed chosen extension, and this is justified because

of the following

2.22. Proposition. *If (F, ψ) is another completion then there exists a unique $\lambda \in \text{Lis}(\overline{E}, F)$ such that $\psi = \lambda \circ \varphi$.*

Proof. The map $\psi \circ \varphi^{-1} : \varphi(E) \rightarrow \psi(E) \subset F$ is continuous and linear, and thus by the linear extension theorem, extends to a unique continuous linear map $\lambda : \overline{E} \rightarrow F$ in a norm-preserving way. (Note that φ norm-preserving implies that φ^{-1} is continuous: $|y| = |\varphi(x)| = |x| = |\varphi^{-1}(y)|$ if $\varphi(x) = y$.) Similarly, the continuous linear map $\varphi \circ \psi^{-1} : \psi(E) \rightarrow \varphi(E) \subset \overline{E}$ extends to $\mu : F \rightarrow \overline{E}$. Then $\mu \circ \lambda = I_{\overline{E}}$ and $\lambda \circ \mu = I_F$ because both maps are extensions of the corresponding identity map and are uniquely determined because of denseness and continuity (compare the uniqueness argument in the proof of the linear extension theorem). Also $|\lambda(x)| = |x|$ for all $x \in \overline{E}$ because λ extends $\psi \circ \varphi^{-1}$, which are both norm preserving. ■

2.23. Completion theorem. *Each normed vector space has a completion.*

First Proof. E'' is complete because the normed space of scalars is complete, and the map $x \mapsto f_x$ as above embeds E into E'' in a norm-preserving way. Thus the closure of the image of this map $\{f_x : x \in E\}$ in E'' is a completion. ■

Second Proof. Let S be the set of Cauchy sequences in E . Then S is a vector space. A *null sequence* in E is a sequence, which converges to 0. The set of null sequences forms a subspace of S . Define two Cauchy sequences $\xi = \{x_n\}$ and $\eta = \{y_n\}$ to be *equivalent* if there exists a null sequence $\alpha = \{a_n\}$ such that $\xi = \eta + \alpha$. This is an equivalence relation on S : (i) ξ is equivalent to ξ by the sequence, which is constant 0, (ii) if α is a null sequence then $-\alpha$ is also a null sequence and $\eta = \xi - \alpha$, thus ξ equivalent to η implies η equivalent to ξ , (iii) if α, β are null sequences then also $\alpha + \beta$ is a null sequence, and thus $\xi = \eta + \alpha$ and $\eta = \rho + \beta$ implies $\xi = \rho + (\beta + \alpha)$, which shows transitivity. Let $\overline{\xi}$ denote the equivalence class of ξ . Then the set of equivalence classes of Cauchy sequences in E will be denoted \overline{E} , and is a vector space by defining $\overline{\xi} + \overline{\eta} := \overline{\xi + \eta}$, and for a scalar c , $c\overline{\xi} := \overline{c\xi}$. (This is the factor space or quotient vector space of the vector space of all Cauchy sequences by the subspace of all null sequences.) That this is well-defined follows easily. If e. g. $\overline{\xi} = \overline{\xi'}$ then $\xi - \xi'$ is a null sequence α , and thus $\xi + \eta = \xi' + \eta + \alpha$ and $\overline{\xi + \eta} = \overline{\xi' + \eta}$. Now if $\xi = \{x_n\}$ and $\eta = \{y_n\}$ are Cauchy sequences in E , which are equivalent then $y_n = x_n + a_n$ for some null sequence $\alpha = \{a_n\}$ and so $|x_n| - |a_n| \leq |y_n| \leq |x_n| + |a_n|$ with $|a_n| \rightarrow 0$, and thus $\lim |x_n| =$

$\lim |y_n|$. So we can define $|\bar{\xi}| := \lim |x_n|$ and this does not depend on the choice of representative in the equivalence class, and defines a norm on \bar{E} . Note that $|\bar{\xi}| = 0$ means that ξ is a null sequence and thus is the 0-vector in \bar{E} . NVS2 and NVS3 are easy to deduce from the corresponding properties in E and the properties of the limit. Now let $\varphi : E \rightarrow \bar{E}$ be defined by $\varphi(x)$ is the equivalence class of the constant sequence $\{x, x, \dots\}$. Then φ is linear and preserves norm by definition. *Claim:* $\varphi(E)$ is a dense subspace of \bar{E} . *Proof:* If $\bar{\xi}$ is represented by the Cauchy sequence $\xi = \{x_n\}$ and $\{\varphi(x_n)\}$ is the corresponding sequence in \bar{E} then $\bar{\xi} = \lim \varphi(x_n) \iff \bar{\xi} - \varphi(x_n) \rightarrow 0$ for $n \rightarrow \infty$ in $\bar{E} \iff |\bar{\xi} - \varphi(x_n)| \rightarrow 0$, where $\bar{\xi}$ is the equivalence class of the Cauchy sequence $\{x_n\}$ and $\varphi(x_n)$ is the equivalence class of the constant sequence $\{x_n, x_n, \dots\}$. Now $|\bar{\xi} - \varphi(x_n)|$ for each n is defined by $\lim_{m \rightarrow \infty} |x_m - \varphi(x_n)_m| = \lim_{m \rightarrow \infty} |x_m - x_n|$, which will have limit 0 for $n \rightarrow \infty$ because $\{x_n\}$ is a Cauchy sequence. This shows that $\varphi(E)$ is dense in \bar{E} , and the *Claim* is proven. Finally let $\{\bar{\xi}_n\}$ be a Cauchy sequence in \bar{E} . For each n there exists $x_n \in E$ such that $|\bar{\xi}_n - \varphi(x_n)| < 1/n$ because of denseness. Then from $|x_n - x_m| = |\varphi(x_n - x_m)| = |\varphi(x_n) - \varphi(x_m)| \leq |\varphi(x_n) - \bar{\xi}_n| + |\bar{\xi}_n - \bar{\xi}_m| + |\bar{\xi}_m - \varphi(x_m)|$ we deduce that $\{x_n\} =: \xi$ is a Cauchy sequence in E . Now let $\varepsilon > 0$ be given. Then $|\bar{\xi}_n - \bar{\xi}| \leq |\bar{\xi}_n - \varphi(x_n)| + |\varphi(x_n) - \bar{\xi}| < 2\varepsilon$ for n sufficiently large (the second term gets small by the argument given above). Thus \bar{E} is complete. ■.

2.24. Remarks. (a) The second proof is more direct and can be easily modified to the case of metric spaces. The elements of the completion are again defined by equivalence classes of Cauchy sequences with the equivalence relation defined by $\{x_n\}$ is equivalent to $\{y_n\}$ if $d(x_n, y_n) \rightarrow 0$ for $n \rightarrow \infty$. The arguments then are very similar, and it follows that each metric space has a unique completion, which is a complete metric space with a distance preserving map of the given metric space into the completion. (Distance preserving implies both continuous and injective.)

(b) It is also possible to define a *completion* of a vector space E with seminorm σ , and construct the space from Cauchy sequences as in the Second Proof above. In this case the vector space will not *embed* into the completion \bar{E} but we will get a map $j : E \rightarrow \bar{E}$ with *kernel* the subspace $\{x \in E : j(x) = 0\} = E_0 = \{x : \sigma(x) = 0\}$. The completion in this space is just the completion of the normed vector space $E/E_0 := \{x + E_0 : x \in E\}$ with addition $(x + E_0) + (y + E_0) := (x + y) + E_0$ and $c(x + E_0) := cx + E_0$ and $|x + E_0| := \sigma(x)$, which is well-defined because $\sigma(x_0) = 0$ for all $x_0 \in E_0$. See also homework problem 12.

(c) Note that it follows that the closure of the image of E in E'' under the natural map is toplinear isomorphic to the space of equivalence classes of Cauchy sequences in E constructed in the Second Proof of the Completion theorem.

Chapter 3

Hilbert Spaces

Throughout we assume that the scalars are in \mathbb{C} , even though the real case typically appears as a special case. A map between L between complex vector spaces is \mathbb{R} -linear if $L(x + y) = L(x) + L(y)$ and $L(\alpha x) = \alpha L(x)$ for all $\alpha \in \mathbb{R}$ and $x \in E$. Following the standard practice we will often write Lx for $L(x)$ in the case of a linear map L .

3.1. Definition. For vector spaces E, F an \mathbb{R} -linear map $L : E \rightarrow F$ is *antilinear* if L is \mathbb{R} -linear and $L(\alpha x) = \bar{\alpha}L(x)$ for $x, y \in E$ and $\alpha \in \mathbb{C}$.

3.2. Example. The trivial example is the the map $L : \mathbb{C} \rightarrow \mathbb{C}$ defined by $L(\alpha) = \bar{\alpha}$.

3.3. Definition. Let E be a vector space. A *sesquilinear form* or *scalar product* on E is a map $E \times E \rightarrow \mathbb{C}$ denoted $(x, y) \mapsto \langle x, y \rangle$, which is linear in the first variable and antilinear in its second variable, so that $\langle x, y_1 + y_2 \rangle = \langle x, y_1 \rangle + \langle x, y_2 \rangle$ and $\langle x, \alpha y \rangle = \bar{\alpha} \langle x, y \rangle$. The form is *hermitian* if $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all $x, y \in E$. The form is *positive* if $\langle x, x \rangle \geq 0$ for all $x \in E$. The form is *positive definite* if it is positive and $\langle x, x \rangle > 0$ for $x \neq 0$.

Throughout we assume that $\langle \cdot, \cdot \rangle$ is **positive hermitian** but not necessarily definite. We will use that a sesquilinear form is in particular \mathbb{R} -bilinear. Note that $\langle v, w \rangle + \langle w, v \rangle = \langle v, w \rangle + \overline{\langle v, w \rangle} = 2\text{Re}\langle v, w \rangle$ while $\langle v, w \rangle - \langle w, v \rangle = 2i\text{Im}\langle v, w \rangle$.

3.4. Example. Let $E = \mathbb{C}^n$ with the hermitian positive definite form

$$\langle x, y \rangle = \sum_{i=1}^n x_i \bar{y}_i,$$

for $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$.

3.5. Definition. $v \in E$ is *perpendicular* or *orthogonal* to $w \in E$ if $\langle v, w \rangle = 0$, Notation: $v \perp w$. For $S \subset E$ the set $S^\perp := \{v \in E : \langle v, w \rangle = 0 \text{ for all } w \in S\}$ is a subspace of E : $0 \in S^\perp$, and $v_1, v_2 \in S^\perp \implies \langle v_1, w \rangle = 0$ and $\langle v_2, w \rangle = 0$ for all $w \in S \implies \langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle = 0$ for all $w \in S$. Thus $v_1 + v_2 \in S^\perp$. If $v \in S^\perp$ and $c \in \mathbb{C}$ then $\langle cv, w \rangle = c\langle v, w \rangle = 0$ and thus $cv \in S^\perp$. We also write $v \perp S$ for $v \in S^\perp$. Also note that $S^\perp = (\text{lin}(S))^\perp$ because $\langle v, w \rangle = 0$ for all $w \in S$ if and only if $\langle v, c_1v_1 + \dots + c_nv_n \rangle = 0$ for all $c_1, \dots, c_n \in \mathbb{C}$ and $v_1, \dots, v_n \in S$. For given sets $S_1, S_2 \subset E$ we write $S_1 \perp S_2$ if $s_i \perp S_j$ for all $s_i \in S_i$, for $i \neq j$, $i, j \in \{1, 2\}$.

Let $E_0 := E^\perp$ be the *null space* of the hermitian product.

3.6. Proposition. *If $w \in E$ such that $\langle w, w \rangle = 0$ then $w \in E_0$.*

Proof. For $t \in \mathbb{R}$ consider $0 \leq \langle v + tw, v + tw \rangle = \langle v, v \rangle + 2t\text{Re}\langle v, w \rangle + t^2\langle w, w \rangle = \langle v, v \rangle + 2t\text{Re}\langle v, w \rangle$. Suppose that $\text{Re}\langle v, w \rangle \neq 0$ then for $|t|$ large with sign opposite to $\text{Re}\langle v, w \rangle$ we get $\langle v, v \rangle + 2t\text{Re}\langle v, w \rangle < 0$, which is a contradiction. Thus $\text{Re}\langle v, w \rangle = 0$ for all $v \in E$. This implies also $\text{Re}\langle iv, w \rangle = 0$ for all $v \in E \implies \text{Im}\langle v, w \rangle = 0$ for all $v \in E$ (The last implication follows from $\text{Re}\langle iv, w \rangle + i\text{Im}\langle iv, w \rangle = \langle iv, w \rangle = i\langle v, w \rangle = i(\text{Re}\langle v, w \rangle + i\text{Im}\langle v, w \rangle) = -\text{Im}\langle v, w \rangle + i\text{Re}\langle v, w \rangle$.) It follows $\langle v, w \rangle = 0$ for all $v \in E$. ■

Let $|v| := \sqrt{\langle v, v \rangle}$ be the *length* or *norm* of v . Because of the last result $|v| = 0 \iff v \in E_0$.

3.7. Schwarz Inequality. *For all $v, w \in E$ we have*

$$|\langle v, w \rangle| \leq |v| |w|$$

Proof. For $\alpha := \langle w, w \rangle = |w|^2 = \bar{\alpha} \geq 0$ and $\beta := -\langle v, w \rangle \in \mathbb{C}$ calculate:
 $0 \leq \langle \alpha v + \beta w, \alpha v + \beta w \rangle = \langle \alpha v, \alpha v \rangle + \langle \beta w, \alpha v \rangle + \langle \alpha v, \beta w \rangle + \langle \beta w, \beta w \rangle$
 $= \alpha \bar{\alpha} \langle v, v \rangle + \beta \bar{\alpha} \langle w, v \rangle + \alpha \bar{\beta} \langle v, w \rangle + \beta \bar{\beta} \langle w, w \rangle$ by substitution we get
 $0 \leq |w|^4 |v|^2 - 2|w|^2 \langle v, w \rangle \overline{\langle v, w \rangle} + |w|^2 \langle v, w \rangle \overline{\langle v, w \rangle}$
 $= |w|^4 |v|^2 - |w|^2 |\langle v, w \rangle|^2$.

If $|w| = 0$ then $w \in E_0$ and $\langle v, w \rangle = 0$ and the inequality is obvious. If $|w| \neq 0$ we can divide by $|w|^2 > 0$ and take square roots. ■

3.8. Remark. Note that the Schwarz inequality for \mathbb{C}^n with the usual hermitian metric is the special case $p = 2$ of the Hölder inequality from homework problem 10.

3.9. Proposition. *The function $E \ni v \mapsto |v| \in \mathbb{R}$ is a seminorm on E , and is*

called the L^2 -norm respectively L^2 -seminorm.

Proof. $|v| \geq 0$ by definition, and $|\alpha v| = \sqrt{\langle \alpha v, \alpha v \rangle} = \sqrt{\alpha \bar{\alpha} \langle v, v \rangle} = |\alpha| \sqrt{\langle v, v \rangle} = |\alpha| |v|$. The triangle inequality follows from Schwarz inequality: $|v + w|^2 = \langle v + w, v + w \rangle = \langle v, v \rangle + \langle w, v \rangle + \langle v, w \rangle + \langle w, w \rangle = |v|^2 + 2\operatorname{Re}\langle v, w \rangle + |w|^2 \leq |v|^2 + 2|\langle v, w \rangle| + |w|^2 \leq |v|^2 + 2|v| |w| + |w|^2 = (|v| + |w|)^2$, where we use that $\operatorname{Re}(a) \leq |a| = \sqrt{\operatorname{Re}(a)^2 + \operatorname{Im}(a)^2}$. Taking square roots gives the triangle inequality. ■

3.10. Remarks. (a) The topology of the vector space with the hermitian form $\langle \cdot, \cdot \rangle$ is defined from the balls defined from the corresponding seminorm (which is a norm if the form is definite) $B_r(v) = \{w \in E : |w - v| < r\}$ in the usual way by defining $U \subset E$ to be open if for each $v \in U$ there exists $r > 0$ such that $B_r(v) \subset U$. This topology is **not** Hausdorff if $|\cdot|$ is not definite. Note that in this case *each* ball $B_r(0)$ contains the subspace E_0 .

(b) The Schwarz inequality shows that the maps $E \ni x \rightarrow \langle x, y \rangle \in \mathbb{C}$ for $y \in E$, and $E \ni y \mapsto \langle x, y \rangle \in \mathbb{C}$ for $x \in E$, are both continuous maps. This has some interesting small consequences. For example, for each set $S \subset E$, the subspace S^\perp is closed: If $\{x_n\}$ is a sequence in S^\perp converging to $x \in E$ then $\langle x, y \rangle = \lim \langle x_n, y \rangle = 0$ for all $y \in S$, and thus $x \in S^\perp$.

3.11. Definitions. A vector $w \in E$ is a *unit vector* if $|w| = 1$. If $|v| \neq 0$ then $v/|v|$ is a unit vector. Also if $w \in E$ with $|w| \neq 0$ and $v \in E$ then there exists a *unique* $c \in \mathbb{C}$ such that $v - cw \perp w$ because: $\langle v - cw, w \rangle = 0 \iff \langle v, w \rangle - \langle cw, w \rangle = 0 \iff \langle v, w \rangle = c \langle w, w \rangle \iff c = \frac{\langle v, w \rangle}{\langle w, w \rangle}$. Then c is called the *Fourier coefficient* of v with respect to w . Let v_1, \dots, v_n such that $v_i \notin E_0$ for all i and $v_i \perp v_j$ for all $i \neq j$. Let c_i be the Fourier coefficient with respect to v_i . Then $(v - c_1 v_1 - \dots - c_n v_n) \perp v_k$ for all $k = 1, \dots, n$. This follows from $\langle v - c_1 v_1 - \dots - c_n v_n, v_k \rangle = \langle v - c_k v_k, v_k \rangle = \langle v, v_k \rangle - c_k \langle v_k, v_k \rangle = 0$, for $k = 1, \dots, n$.

3.12. Example. For $a < b$ let $E = C[a, b]$ be the space of continuous complex-valued functions on $[a, b]$ with positive definite hermitian product

$$\langle f, g \rangle := \int_a^b f(t) \overline{g(t)} dt.$$

(Note that the definiteness follows from the continuity of the functions.) Let $a = 0$ and $b = 2\pi$. Then the functions $\chi_k \in E$ defined by $\chi_k(t) := e^{ikt}$ for $k \in \mathbb{Z}$ are *pairwise perpendicular*: For $k \neq \ell$ this follows from $\langle \chi_k, \chi_\ell \rangle = \int_0^{2\pi} e^{i(k-\ell)t} dt = \frac{1}{i(k-\ell)} [e^{i(k-\ell)t}]_0^{2\pi} = 0$, and $|\chi_k|^2 = \langle \chi_k, \chi_k \rangle = \int_0^{2\pi} dt = 2\pi \neq 0$.

Then the k -th Fourier coefficient of f is $c_k(f) = \frac{\langle f, \chi_k \rangle}{\langle \chi_k, \chi_k \rangle} = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{ikt} dt \in \mathbb{C}$. Note that for each number N , $f - \sum_{k=-N}^N c_k \chi_k$ is perpendicular to χ_{-N}, \dots, χ_N .

3.13. Theorem. (a) **Pythagoras theorem.** *If $u \perp w$ then $|u + w|^2 = |u|^2 + |w|^2$.*

(b) **Parallelogram Law.** *For $u, w \in E$: $|u + w|^2 + |u - w|^2 = 2|u|^2 + 2|w|^2$.*

Proof. (a) $|u + w|^2 = \langle u + w, u + w \rangle = \langle u, u \rangle + \langle u, w \rangle + \langle w, u \rangle + \langle w, w \rangle = |u|^2 + |w|^2$, because $u \perp w : \iff \langle u, w \rangle = 0 \iff \langle w, u \rangle = \overline{\langle u, w \rangle} = 0$.

(b) $|u + w|^2 + |u - w|^2 = \langle u + w, u + w \rangle + \langle u - w, u - w \rangle = 2\langle u, u \rangle + 2\langle w, w \rangle = |u|^2 + |w|^2$ because the mixed products cancel by sign. ■

The example above suggests to consider the following generalization:

3.14. Definitions. Let $\{v_i\}_{i \in I}$ be a family of elements of E such that $|v_i| \neq 0$ for all i . For each finite subfamily $\{v_{i_1}, \dots, v_{i_n}\}$ consider the space $\text{lin}\{v_{i_1}, \dots, v_{i_n}\}$ generated by this subfamily, which is the vector space of all *linear combinations* $v = c_1 v_{i_1} + \dots + c_n v_{i_n}$ for all complex coefficients $c_i \in \mathbb{C}$. Then the space $F = \text{lin}\{v_i\}$ is the union of all those subspaces and is the space generated by the family $\{v_i\}$. The family $\{v_i\}$ is *total* in E if the *closure* of F in E is E .

3.15. Definitions. The family $\{v_i\}$ is *orthogonal* if $v_i \perp v_j$ for all $i, j \in I$, $i \neq j$ and $|v_i| \neq 0$ for all i . The family is *orthonormal* if $|v_i| = 1$ for all $i \in I$. Given an orthogonal family, an orthonormal family is constructed by dividing each element by its norm. A total orthonormal family is a *Hilbert basis*, or sometimes also only called orthonormal basis.

An orthogonal family is linearly independent because $\sum_{j=1}^n \alpha_j v_j = 0$ (any finite set of indices from I actually) then $\langle \sum_{j=1}^n \alpha_j \langle v_j, v_k \rangle = \alpha_k \langle v_k, v_k \rangle = 0$ and thus $\alpha_k = 0$ for all $k = 1, \dots, n$. Note that this is not an orthonormal basis in the algebraic sense because not every vector is a *finite* linear combination of elements of the Hilbert basis. (The family $\{\chi_k\}$ in the example above is orthogonal and can be normalized by $e_k := \frac{1}{\sqrt{2\pi}} \chi_k$.)

In the following we will also write i for elements of a given subset of I (in order to avoid double indices).

3.16. Proposition. *Let $\{v_i\}$ be an orthogonal family in E , $x \in E$ and c_i be the Fourier coefficient of x with respect to v_i . Let $\{a_i\}$ be any family of numbers.*

Then

$$\left| x - \sum_{k=1}^n c_k v_k \right| \leq \left| x - \sum_{k=1}^n a_k v_k \right|.$$

Proof. $(x - \sum_{k=1}^n c_k v_k) \perp v_i$ for $i = 1, \dots, n$. Hence from Pythagoras theorem $|x - \sum_{k=1}^n a_k v_k|^2 = |x - \sum_{k=1}^n c_k v_k + \sum_{k=1}^n (c_k - a_k)v_k|^2 = |x - \sum_{k=1}^n c_k v_k|^2 + |\sum_{k=1}^n (c_k - a_k)v_k|^2 \geq |x - \sum_{k=1}^n c_k v_k|^2$. ■

3.17. Remark. The previous result shows that the Fourier coefficients determine *best approximations* of x by linear combinations in all finite dimensional subspaces $\text{lin}\{v_1, \dots, v_n\}$ (here $\{1, \dots, n\}$ can be replaced by any finite subset of I) with respect to the seminorm.

3.18. Definition. A *pre-Hilbert space* is a vector space with a positive definite hermitian form. A *Hilbert space* is a pre-Hilbert space, which is complete with respect to the corresponding norm.

3.19. Examples. (a) $C[a, b]$ with product defined above is a pre-Hilbert space. It is not a Hilbert space: Without restriction assume $a = 0$ and $b = 1$. Consider the sequence of functions $\{f_n\}$ in $C[0, 1]$ defined by $f_n(t) = 0$ for $t \leq \frac{1}{2} - \frac{1}{n}$ and $f(t) = 1$ for $t \geq \frac{1}{2}$, and linearly interpolating in between, i. e. $f_n(t) = \frac{n}{t} + (1 - \frac{n}{2})$. Then for $m < n$ it can be calculated that $|f_n - f_m|^2 \rightarrow 0$ for $\min(n, m) \rightarrow \infty$. (For $m < n$ calculate $\int_{\frac{1}{2} - \frac{1}{m}}^{\frac{1}{2} - \frac{1}{n}} ((n - m)t - \frac{1}{2}(n - m))^2 dt = (n - m)^2 \int_{\frac{1}{2} - \frac{1}{m}}^{\frac{1}{2} - \frac{1}{n}} (t - \frac{1}{2})^2 dt = (n - m)^2 \int_{-\frac{1}{m}}^{-\frac{1}{n}} u^2 du = \frac{(n - m)^2}{3} (\frac{1}{m^3} - \frac{1}{n^3})$, or consider the areas that are computed for n, m large.) It follows that $\{f_n\}$ is Cauchy sequence. But $\{f_n\}$ does not converge in $C[0, 1]$ because a limit function is the step function $f(t) = 0$ for $t < \frac{1}{2}$ and $f(t) = 1$ for $t \geq \frac{1}{2}$.

(b) Let ℓ^2 be the vector space of complex sequences such that $\sum_{n=1}^{\infty} |a_n|^2$ converges. Then a positive definite hermitian product can be defined by

$$\langle a, b \rangle = \sum_{n=1}^{\infty} a_n \bar{b}_n$$

for $a = \{a_n\}$ and $b = \{b_n\}$. This is well-defined because by taking the limit in Hölder's inequality we get

$$\sum_{n=1}^{\infty} |a_n \bar{b}_n| \leq \left(\sum_{n=1}^{\infty} |a_n|^2 \right)^{1/2} \left(\sum_{n=1}^{\infty} |b_n|^2 \right)^{1/2}$$

Then $|a| = |a|_2 = (\sum_{n=1}^{\infty} |a_n|^2)^{1/2}$ is just the norm on ℓ^2 as defined in homework problem 12, where it is also shown that ℓ^2 is complete with respect to this norm. ℓ^2 is the prototype of Hilbert space with countable basis, as we will see shortly.

3.20. Remarks. (a) If a positive hermitian form is given on a vector space E , which is not definite, then a pre-Hilbert space is defined by E/E_0 with the induced form: $\langle x + E_0, y + E_0 \rangle := \langle x, y \rangle$.

(b) If S is a subset of a Hilbert space then $S \cap S^\perp \subset \{0\}$. In fact $v \in S$ and $v \perp S$ imply $\langle v, v \rangle = 0$ and thus $v = 0$ because of definiteness. In a space with positive hermitian form it is always true that $S \cap S^\perp \subset E_0$.

3.21. Lemma. For F a closed subspace of a Hilbert space E and $x \in E$ let $a := \inf_{y \in F} |x - y|$. Then there exists an element $y_0 \in F$ such that $a = |x - y_0|$.

Proof. Let $\{y_n\}$ be a sequence in F such that $|y_n - x| \rightarrow a$ for $n \rightarrow \infty$. Then $\{y_n\}$ is a Cauchy sequence. In fact, by the parallelogram law with $u = y_n - x$ and $v = y_m - x$, and so $u - v = y_n - y_m$ and $u + v = y_n + y_m - 2x = 2(\frac{1}{2}(y_n + y_m) - x)$ and thus (using $\frac{1}{2}(y_n + y_m) \in F$ and the definition of the number a) it follows:

$$|y_n - y_m|^2 = 2|y_n - x|^2 + 2|y_m - x|^2 - 4|\frac{1}{2}(y_n + y_m) - x|^2 \leq 2|y_n - x|^2 + 2|y_m - x|^2 - 4a^2,$$

which $\rightarrow 0$ for $\min(n, m) \rightarrow \infty$. Because E is complete $y_n \rightarrow y_0$ and $y_0 \in F$ because F is closed. ■

3.22. Proposition. Let F be a closed subspace of the Hilbert space E with $F \neq E$. Then there exists $0 \neq z \in E$ such that $z \perp F$.

Proof. Let $x \in E \setminus F$ and $y_0 \in F$ with $|x - y_0| = \inf_{y \in F} |y - x|$ by the previous lemma. Let $z := x - y_0$, which is $\neq 0$ because $y_0 \in F$ and $x \notin F$. For arbitrary $y \in F$ and $\alpha \in \mathbb{C}$ we have $|x - y_0|^2 \leq |x - y_0 + \alpha y|^2$ because $y_0 - \alpha y \in F$ and y_0 is chosen such that $|x - y_0|$ is minimal. Then $|x - y_0 + \alpha y|^2 - |x - y_0|^2 \geq 0$, and by expanding the left hand side

$$\alpha \langle y, z \rangle + \bar{\alpha} \langle z, y \rangle + \alpha \bar{\alpha} \langle y, y \rangle \geq 0.$$

Let $\alpha := t \langle z, y \rangle$ with $t < 0$. Then $\bar{\alpha} = t \langle y, z \rangle$ and the above inequality becomes

$$2t |\langle y, z \rangle|^2 + t^2 |\langle y, z \rangle|^2 \langle y, y \rangle \geq 0,$$

or equivalently, after division by t :

$$(2 + t |y|^2) |\langle y, z \rangle|^2 \leq 0.$$

Now let $t \rightarrow 0$ to get $|\langle y, z \rangle|^2 \leq 0$ and thus $\langle y, z \rangle = 0$ for all $y \in F$. ■

3.23. Corollary. Each Hilbert space $E \neq \{0\}$ has a Hilbert basis.

Proof. Let S be the set of all non-empty orthogonal families in E . For $\mathcal{F}_i \in S$, $i = 1, 2$, define $\mathcal{F}_1 \leq \mathcal{F}_2$ if $\mathcal{F}_1 \subset \mathcal{F}_2$. This defines an inductive ordering on S : If T is a totally ordered subset of S then the union $\cup \mathcal{F}$ of all $\mathcal{F} \in T$ is an upper bound because it is again an element of S . Let \mathcal{B} be a maximal element of S , and let F be the subspace of E generated by \mathcal{B} , i. e. $F := \text{lin}\{v : v \in \mathcal{B}\}$. *Claim:* $\overline{F} = E$. *Proof.* Suppose that $\overline{F} \neq E \implies \exists 0 \neq z \in E : z \perp \overline{F}$ by the proposition. But then the inclusion $\mathcal{B} \cup \{z\} \supset \mathcal{B}$ is proper, and \mathcal{B} is not maximal. ■.

For a subspace F of a vector space E with a positive hermitian form we have by definition $\langle v, w \rangle = 0$ for all $v \in F$ and $w \in F^\perp$, and so $v \in (F^\perp)^\perp = F^{\perp\perp}$. This proves that $F \subset F^{\perp\perp}$ always holds. Also F^\perp is closed because $\langle x_n, y \rangle = 0$ for all n implies $\lim \langle x_n, y \rangle = \langle \lim x_n, y \rangle = 0$. In a pre-Hilbert space also $F \cap F^\perp = \{0\}$ because $v \in F \cap F^\perp \implies \langle v, v \rangle = 0 \implies v = 0$. Now for closed subspaces of Hilbert spaces an even stronger statement can be proved.

3.24. Corollary. *If F is a closed subspace of a Hilbert space E then $E = F + F^\perp$, and $F^{\perp\perp} = F$.*

Proof. If $y_n \in F$ and $z_n \in F^\perp$ then $\{y_n + z_n\}$ is Cauchy $\iff \{y_n\}$ is Cauchy and $\{z_n\}$ is Cauchy. This follows by the Pythagorean theorem:

$$|(y_n + z_n) - (y_m + z_m)|^2 = |(y_n - y_m) + (z_n - z_m)|^2 = |y_n - y_m|^2 + |z_n - z_m|^2.$$

So if $\{y_n + z_n\}$ converges in E then it is Cauchy in both F and F^\perp , and thus converges to a limit in $F + F^\perp$ (recall that F^\perp is closed too.). Hence $F + F^\perp$ is closed. If $F + F^\perp \neq E$ then $\exists 0 \neq w \in E : w \perp (F + F^\perp)$. Then $w \perp F \iff w \in F^\perp$ but also $w \perp F^\perp$ (or $w \in F^{\perp\perp}$), and so $\langle w, w \rangle = 0$, which contradicts the definiteness of $\langle \cdot, \cdot \rangle$. Now recall that $F \cap F^\perp = \{0\}$. So we can write each $x \in E$ *uniquely* as $x = y + z$ with $y \in F$ and $z \in F^\perp$. (If also $x = y' + z'$ then $y - y' = z - z' \in F \cap F^\perp$ and so $y = y'$ and $z = z'$.) Now if $x \in F^{\perp\perp} \subset E$ is written as $y + z$ with $y \in F$ and $z \in F^\perp$ then $0 = \langle x, z \rangle = \langle y + z, z \rangle = \langle y, z \rangle + \langle z, z \rangle = \langle z, z \rangle \implies z = 0$ because the form on E is definite. This shows $F^{\perp\perp} \subset F$, and with the observation above $F^{\perp\perp} = F$. ■

3.25. Definition. For a closed subspace F of a Hilbert space E we let $P = P_F : E \rightarrow E$ or P denote the *orthogonal projection* defined by $P(x) := y$ if $x = y + z$ with $y \in F$ and $z \in F^\perp$. Because $|x|^2 = |y|^2 + |z|^2$ it follows that $|y| = |P(x)| \leq |x|$, and thus P is continuous with norm ≤ 1 . Note that $Px \in F$ and $x - Px \perp F$ for each $x \in E$.

3.26. Theorem. *Let E be a Hilbert space with a sequence $\{F_i\}_{i \in \mathbb{N}}$ of closed*

subspaces such that $F_i \perp F_j$ if $i \neq j$. Let $F := \text{lin}(\cup_i F_i) = \{x_1 + \dots + x_n : x_i \in F_i, n \text{ finite}\}$. Then every element of \overline{F} has a unique expression as a convergent series

$$x = \sum_{i=1}^{\infty} x_i = \lim_{N \rightarrow \infty} \sum_{i=1}^N x_i, \quad x_i \in F_i,$$

with $x_i = P_i(x)$ where $P_i = P_{F_i} : E \rightarrow F_i$ is the orthogonal projection on F_i for all $i \in \mathbb{N}$. Moreover, for any choice of elements $y_i \in F_i$ we have

$$\left| x - \sum_{i=1}^n P_i x \right| \leq \left| x - \sum_{i=1}^n y_i \right|.$$

Proof. Note that $(x - \sum_{i=1}^n P_i x)$ is perpendicular to F_1, \dots, F_n (and thus to $F_1 + \dots + F_n$): For $v_j \in F_j$, $j \in \{1, \dots, n\}$ we have $\langle x - \sum_{i=1}^n P_i x, v_j \rangle = \langle x, v_j \rangle - \sum_{i=1}^n \langle P_i x, v_j \rangle = \langle x, v_j \rangle - \langle P_j x, v_j \rangle$ because $P_i x \perp v_j$ for $i \neq j$, and now $x = (x - P_j x) + P_j x$ with $P_j x \in F_j$ and $x - P_j x \perp F_j$ implies that $\langle x, v_j \rangle = \langle P_j x, v_j \rangle$. Thus because $\sum_{i=1}^n (P_i(x) - y_i) \in F$, the theorem of Pythagoras gives

$$\left| x - \sum_{i=1}^n y_i \right|^2 = \left| x - \sum_{i=1}^n P_i x \right|^2 + \left| \sum_{i=1}^n (P_i x - y_i) \right|^2,$$

and the inequality follows. Now each $x \in F$ can be written as $x_1 + \dots + x_n$ with $x_i \in F_i$. Then $P_i x = \sum_{j=1}^n P_i x_j = P_i x_i = x_i$. Thus we have $x = \sum_{i=1}^{\infty} P_i x$ for each $x \in F$, where the sum is actually finite for each x but the length of the sum will depend on x . Next for each $x \in \overline{F}$ there is a sequence $\{y_i\}$ in F converging to x . Now $x = \lim_i (y_i) = \lim_i \sum_{j=1}^{\infty} P_j(y_i) = \sum_{j=1}^{\infty} \lim_i P_j(y_i) = \sum_{j=1}^{\infty} P_j(\lim_i y_i) = \sum_{j=1}^{\infty} P_j x$. The change of summation and limit is justified because $P_j y_i \rightarrow P_j x$ for $i \rightarrow \infty$ is uniform in j because $|P_j y_i - P_j x| \leq |P_j| |y_i - x| \leq |y_i - x|$. Now if $x = \sum_{i=1}^{\infty} x_i$ is any representation of $x \in \overline{F}$ as a series of elements in F_i then $P_n(x) = P_n(\sum_{i=1}^{\infty} x_i) = \sum_{i=1}^{\infty} P_n x_i = x_n$ because $P_n x_i = 0$ for $n \neq i$ and $P_n x_n = x_n$ by definition of the projections. ■

3.27. Remarks. (a) The theorem shows that the approximations $\sum_{i=1}^n P_i x$ of $x \in \overline{F}$ are *best approximations* of $x \in \overline{F}$ in the subspaces $F_1 + \dots + F_n \subset F$.

(b) The family $\{F_i\}$ is called an *orthogonal decomposition* of \overline{F} (respectively E if $\overline{F} = E$).

(c) Note that the above theorem in particular applies to closed subspaces $F_i := \mathbb{C}v_i$ for a given orthonormal family of vectors $\{v_i\}$ (recall that finite dimensional subspaces are closed). In this case $P_i(x) = \langle x, v_i \rangle v_i$ because $(x - P_i(x)) \perp F_i \iff (x - P_i(x)) \perp v_i$, which holds by the definition of Fourier coefficients.

3.28. Theorem. (a) **Parseval identity** Let a Hilbert space E have a denumerable Hilbert basis $\{v_i\}$. Then every element can be written uniquely as a convergent series

$$x = \sum_{n=1}^{\infty} a_n v_n,$$

where a_n is the Fourier coefficient of x with respect to v_n . Moreover $|x|^2 = \sum_{n=1}^{\infty} |a_n|^2$.

(b) **Bessel inequality** If $\{v_n\}$ is any orthonormal system in a Hilbert space E , and a_n is the Fourier coefficient of x with respect to v_n , then

$$\sum_{n=1}^{\infty} |a_n|^2 \leq |x|^2$$

Proof. (a) follows immediately from the previous theorem with $E = \overline{F}$ and considering previous remarks (c). The equation $|x|^2 = \sum_{n=1}^{\infty} |a_n|^2$ follows from $x = \sum_n a_n v_n$ by using the theorem of Pythagoras on the partial sums and noticing $|v_n| = 1$. For (b) consider $E = \overline{F} + \overline{F}^{\perp}$ and let $x = Px + (x - Px)$ where P is the orthogonal projection on \overline{F} . Then $|x| \geq |Px|$, and now apply (a) to $P(x) \in \overline{F}$. ■

3.29. Theorem. Let E_1, E_2 be two Hilbert spaces with Hilbert bases of the same cardinality (there is a bijection between the two Hilbert bases). Then E_1 and E_2 are Hilbert space isomorphic, i. e. there is an isometry from E_1 to E_2 .

In the midterm problem 2 it will be shown that the hermitian product of a Hilbert space is determined by the norm. It follows from this that an isometry $\varphi : E_1 \rightarrow E_2$ between Hilbert spaces actually preserves the products, i. e. $\langle \varphi(x), \varphi(y) \rangle = \langle x, y \rangle$ for all $x, y \in E_1$.

Proof. If $\{v_n\}$ and $\{w_n\}$ are denumerable Hilbert bases of E_1 and E_2 respectively then the result follows immediately by mapping $\sum_{n=1}^{\infty} a_n v_n$ to $\sum_{n=1}^{\infty} a_n w_n$. In general let $\{v_i\}_{i \in I}$ and $\{w_i\}_{i \in I}$ be the Hilbert bases. Let $F_1 := \text{lin}\{v_i\}$ and $F_2 := \text{lin}\{w_i\}$. The map $v_i \mapsto w_i$ for each $i \in I$ extends *uniquely* to a linear isomorphism $F_1 \rightarrow F_2$ (This follows because the elements of $\{v_i\}$ are linearly independent and thus each $x \in F_1$ can be written uniquely as a *finite* linear combination of v_i . The inverse isomorphism $F_2 \rightarrow F_1$ is constructed similarly from $w_i \mapsto v_i$.) Because of orthonormality, by the theorem of Pythagoras, $|\sum_{i=1}^n a_i v_i|^2 = \sum_{i=1}^n |a_i|^2 = |\sum_{i=1}^n a_i w_i|^2$, and the extension is norm-preserving. Now consider the linear continuous map $F_1 \rightarrow F_2 \subset \overline{F_2} = E_2$. This

extends *uniquely* and norm-preserving to a linear map $E_1 = \overline{F_1} \rightarrow E_2$. Similarly the inverse map is constructed. ■

3.30. Remark. The result above shows that Hilbert spaces behave very similar to finite dimensional vector spaces. In particular, Hilbert spaces with denumerable Hilbert basis are all isomorphic to the Hilbert space ℓ^2 of square summable complex sequences. This type of representation is at the heart of the different representations used in quantum mechanics.

3.31. Example. Consider the subspace of $C[0, 2\pi]$, which is generated by the functions χ_k , $k \in \mathbb{Z}$. Note that this is actually a subspace of the vector space of periodic continuous functions on $[0, 2\pi]$. Then the abstract completion of this normed vector space is a Hilbert space because the hermitian product can easily be extended to the closure using the linear extension theorem. The resulting Hilbert space is isomorphic to ℓ^2 .

Note that in general E is *not* isomorphic to E' for infinite dimensional vector spaces (because e. g. E could be not complete while E' is always a Banach space). If E is a Hilbert space we always have the following, which implies in particular that the isometric embedding $E \rightarrow E''$ is always an isometric isomorphism for Hilbert spaces. (A Banach E space is called *reflexive* if the natural embedding $E \rightarrow E''$ is surjective. Thus the result below implies that Hilbert spaces are reflexive.)

3.32. Theorem. *For every $y \in E$ the map λ_y defined by $\lambda_y(x) := \langle x, y \rangle \in \mathbb{C}$ is a functional, and $E \ni y \mapsto \lambda_y \in E'$ is a norm-preserving antilinear isomorphism.*

Proof. λ_y is obviously a linear map $E \rightarrow \mathbb{C}$. Because $|\lambda_y(x)| \leq |\langle x, y \rangle| \leq |x||y|$ by Schwarz inequality, it follows that λ_y is bounded and $|\lambda_y| \leq |y|$. But for $y \neq 0$, $\lambda_y(\frac{y}{|y|}) = |y|$, so the supremum over $|\lambda_y(x)|$ over $x \in S(E)$ is $|y|$, and $|\lambda_y| = |y|$. Thus $y \mapsto \lambda_y$ is norm-preserving and antilinear because $\langle \cdot, \cdot \rangle$ is antilinear in the second variable, i. e. $\lambda_{\alpha y} = \bar{\alpha}\lambda_y$. Because $y \mapsto \lambda_y$ is norm-preserving it is injective. So it suffices to show that it is surjective. Let $0 \neq \lambda \in E'$ and let $F := \ker(\lambda) = \{x \in E : \lambda(x) = 0\} \subset E$ the kernel subspace. Note that this is a closed subspace because $\ker(\lambda) = \lambda^{-1}(\{0\})$, $\{0\} \subset E$ is a closed subset, and λ is continuous. Because $F \neq E (\iff \lambda \neq 0)$ there exists $0 \neq z \in E$ such that $z \perp F$. *Claim:* There exists $\alpha \in \mathbb{C}$ such that $\lambda(x) = \langle x, \alpha z \rangle$ for all $x \in E$. (This proves the theorem because then $\lambda_{\alpha z} = \lambda$) *Proof.* We find α from $\langle z, \alpha z \rangle = \lambda(z) \iff \bar{\alpha} = \frac{\lambda(z)}{\langle z, z \rangle}$. If now for $x \in E$ we write $x = x - \frac{\lambda(x)}{\lambda(z)}z + \frac{\lambda(x)}{\lambda(z)}z$

with $x - \frac{\lambda(x)}{\lambda(z)} \in F$ (because $\lambda(x - \frac{\lambda(x)}{\lambda(z)}z) = \lambda(x) - \frac{\lambda(x)}{\lambda(z)}\lambda(z) = 0$) we get $\langle x, \alpha z \rangle = \langle (x - \frac{\lambda(x)}{\lambda(z)}z) + \frac{\lambda(x)}{\lambda(z)}z, \frac{\lambda(z)}{\langle z, z \rangle}z \rangle = \lambda(x)$. ■

Chapter 4

Integration

The general idea of Lebesgue integration developed historically from the study of the validity of the formula

$$\int_{-\infty}^{\infty} \lim_{n \rightarrow \infty} f_n(x) dx = \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x) dx$$

for a sequence $\{f_n\}$ of nonuniformly convergent but integrable functions, and the wish to have available a satisfactory theory of integration for unbounded functions. This leads to general ideas of measure theory and modern integration theory. The Lebesgue measure is a generalization of the idea of length. General measures are e. g. important in the study of probability theory.

For $X \subset A$ we often write $\mathfrak{C}_X(A)$ and $X - A$ instead of $X \setminus A$, or just $\mathfrak{C}A$ if X is understood, following usual conventions in measure theory.

Recall that for a mapping of sets $f : X \rightarrow Y$ and subsets $A \subset B \subset X$ respectively $A' \subset B' \subset Y$ the following holds:

- (1) $f(A) \subset f(B) \subset Y$, (2) $f(A \cap B) \subset f(A) \cap f(B)$,
- (3) $f(A \cup B) = f(A) \cup f(B)$, (4) $f^{-1}(A') \subset f^{-1}(B') \subset X$,
- (5) $f^{-1}(A' \cap B') = f^{-1}(A') \cap f^{-1}(B')$, (6) $f^{-1}(A' \cup B') = f^{-1}(A') \cup f^{-1}(B')$,
- (7) $f^{-1}(\mathfrak{C}_Y A') = \mathfrak{C}_X(f^{-1}(A'))$. Note the difference between (2) and (5).

Recall that a set is denumerable if and only if there is a one-to-one correspondence between the set and \mathbb{N} .

4.1. Definition. A σ -algebra on a set $X \neq \emptyset$ is a collection \mathcal{M} of subsets of X satisfying:

σ -Alg1: $\emptyset \in \mathcal{M}$

σ -Alg2: $A \in \mathcal{M} \implies \mathfrak{C}_X A \in \mathcal{M}$, and $\{A_n\}_{n \in \mathbb{N}}$ a sequence of elements $A_n \in \mathcal{M} \implies$

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$$

It follows that $\mathfrak{C}_X \emptyset = X \in \mathcal{M}$ and if $\{A_n\}_{n \in \mathbb{N}}$ is a sequence of sets with $A_n \in \mathcal{M}$ then $\mathfrak{C}A_n \in \mathcal{M}$ for all $n \in \mathbb{N}$ and so:

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \mathfrak{C}\mathfrak{C}A_n = \mathfrak{C} \bigcup_{n=1}^{\infty} \mathfrak{C}A_n \in \mathcal{M}.$$

By choosing $A_n = \emptyset$ for almost all n , the above assertions about unions and intersections of course also apply to finite collections. Also note that if $A, B \in \mathcal{M}$ also $A - B = A \cap \mathfrak{C}B \in \mathcal{M}$.

4.2. Definition. A pair (X, \mathcal{M}) with \mathcal{M} a σ -algebra on X is called a *measurable space*, and the elements of \mathcal{M} are called its *measurable sets*.

4.3. Proposition. A collection \mathcal{M} of subsets of X is a σ -algebra if and only if $\emptyset \in \mathcal{M}$, \mathcal{M} is closed under taking arbitrary complements, **finite** intersections, and such that for each sequence $\{A_n\}$ of **disjoint** elements of \mathcal{M} also $\cup A_n \in \mathcal{M}$.

Proof. Because

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 - A_1) \cup (A_3 - (A_2 \cup A_1)) \cup \dots$$

the assumptions about disjoint unions and complements implies the assumption about general unions. Note that taking complements in X , together with finite intersections, show that the complements in the above union are in \mathcal{M} (see the argument above). ■

4.4. Definition. An *algebra of subsets* of X is a collection \mathcal{A} satisfying:

ALG1: $\emptyset \in \mathcal{A}$

ALG2: $A, B \in \mathcal{A} \implies A \cap B, A \cup B, A - B \in \mathcal{A}$.

4.5. Proposition. For each collection \mathcal{S} of subsets of X there exists a smallest σ -algebra $\mathcal{M} = \mathcal{S}^\sigma$ in X , which contains \mathcal{S} . This is the σ -algebra generated by \mathcal{S} .

Proof. The collection of all subsets of X (the power set of X) is a σ -algebra. Because the intersection of σ -algebras in X is a σ -algebra in X the intersection of all σ -algebras in X containing \mathcal{S} is the σ -algebra, which we are looking for.

(Let $T \neq \emptyset$ be the set of all σ -algebras containing \mathcal{S} . Then $\emptyset \in T$ for all $T \in T$, so $\emptyset \in \cap_{T \in T} T =: \mathcal{M}$. Also if $A \in \mathcal{M}$ then $\mathfrak{C}A \in T$ for all $T \in T$, and thus $\mathfrak{C}A \in \mathcal{M}$. If $\{A_n\}$ is a sequence of subsets of X with $A_n \in \mathcal{M}$ for all n then $A_n \in T$ for all $T \in T$, and thus $\cup_{n=1}^{\infty} A_n \in T$ for all $T \in T$, which implies $\cup_{n=1}^{\infty} A_n \in \mathcal{M}$. Finally $\mathcal{S} \subset T$ for each $T \in T$, and so $\mathcal{M} \supset \mathcal{S}$.) ■

4.6. Examples. (a) If X is a topological space let \mathcal{T} be the collection of all open sets. Then $\sigma(\mathcal{T}) \supset \mathcal{T}$ is the algebra of *Borel sets*. Every denumerable intersection of open sets and every denumerable union of closed sets is a Borel set. Moreover, the Borel algebra on \mathbb{R} is generated by the sets (a, ∞) for $a \in \mathbb{R}$. In fact,

$$[a, \infty) = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, \infty)$$

$$(-\infty, a) = \mathfrak{C}[a, \infty)$$

$$(a, b) = (-\infty, b) \cap (a, \infty)$$

$$[a, b] = [a, \infty) - (b, \infty)$$

Similarly the sets $[a, \infty)$ for $a \in \mathbb{R}$ generate the Borel algebra.

(b) Let $f : X \rightarrow Y$ be a map for (X, \mathcal{M}) a measurable space. $\mathcal{N} := \{S \subset Y : f^{-1}(S) \in \mathcal{M}\}$. Then \mathcal{N} is a σ -algebra on Y : $f^{-1}(\emptyset) = \emptyset$ so $\emptyset \in \mathcal{N}$, and $f^{-1}(\mathfrak{C}_Y S) = \mathfrak{C}_X f^{-1}(S)$, so $S \in \mathcal{N} \iff f^{-1}(S) \in \mathcal{M} \implies \mathfrak{C}_X f^{-1}(S) \in \mathcal{M} \implies \mathfrak{C}_Y(S) \in \mathcal{N}$. Similarly if $A_n \in \mathcal{N}$ for $n \in \mathbb{N}$ then $f^{-1}(A_n) \in \mathcal{M} \implies f^{-1}(\cup A_n) = (\cup f^{-1}(A_n)) \in \mathcal{M} \implies \cup A_n \in \mathcal{N}$. The σ -algebra \mathcal{N} is called the *direct image of \mathcal{M} under f* and is denoted $f_*(\mathcal{M})$.

(c) Let $Y \subset X$ for (X, \mathcal{M}) a measurable space. Then $\mathcal{M}_Y := \{A \cap Y : A \in \mathcal{M}\}$ is a σ -algebra on Y , which is *induced by \mathcal{M} on Y* , and so (Y, \mathcal{M}_Y) is a measurable space. (This follows essentially from the distributive properties of the algebra of sets: $(\cup A_n) \cap Y = \cup(A_n \cap Y)$.)

4.7. Definition. A map $f : X \rightarrow Y$ with (X, \mathcal{M}) and (Y, \mathcal{N}) measurable spaces is *measurable* if $B \in \mathcal{N} \implies f^{-1}(B) \in \mathcal{M}$ ($\iff \mathcal{N} \subset f_*(\mathcal{M})$).

In this language $f_*(\mathcal{M})$ could also be defined to be the *largest* σ -algebra for which f is measurable.

4.8. Proposition. (a) If $f : X \rightarrow Y$ is measurable and $g : Y \rightarrow Z$ is measurable then $g \circ f : X \rightarrow Z$ is measurable.

(b) Let $f : X \rightarrow Y$ be a map from a measurable space into a topological space, with the σ -algebra on Y the Borel algebra. Then f is measurable \iff for every open $V \subset Y$ the set $f^{-1}(V) \subset X$ is measurable.

Proof. (a) If A is measurable subset of Z then $g^{-1}(A) \subset Y$ is measurable and so $f^{-1}(g^{-1}(A)) = (g \circ f)^{-1}(A) \subset X$ is measurable. (b) The σ -algebra $f_*(\mathcal{M}) = \{S \subset Y : f^{-1}(S) \subset X \text{ is measurable}\}$ contains the open sets by assumption. Because the Borel algebra is the smallest σ -algebra containing the open sets, $f_*(\mathcal{M})$ also contains the Borel algebra. Thus f is measurable. ■

Equivalently, a map from a measurable space into a topological space (with the Borel algebra, as will be always assumed from now on) is measurable if and only if inverse images of closed sets are measurable. Note that if $f : X \rightarrow Y$ is a map between measurable spaces both given by Borel algebras, then f continuous implies f measurable. The converse is **not true** in general, see the following Example (b). Also, for a measurable map $f : X \rightarrow Y$ into a topological space, $f^{-1}(\bigcup_{n=1}^{\infty} U_n) = \bigcap_{n=1}^{\infty} U_n$ is measurable for any sequence U_n of open sets, and also the preimage of a countable intersection of closed sets is measurable.

4.9. Examples. (a) Let $J = (a, b] \subset \mathbb{R}$ and $f : X \rightarrow \mathbb{R}$ measurable. Then

$$(a, b] = \bigcup_{n=N}^{\infty} [a + \frac{1}{N}, b]$$

for N sufficiently large, and thus $f^{-1}(J)$ is measurable. Because the Borel algebra is generated by sets $[a, \infty)$ for $a \in \mathbb{R}$ it follows that a mapping f from a measurable space (X, \mathcal{M}) into \mathbb{R} is measurable if and only if $f^{-1}[a, \infty)$ is measurable for all $a \in \mathbb{R}$. Also, if f is measurable then

$$\{x : f(x) = a\} = \{x : f(x) \geq a\} \cap \{x : f(x) \leq a\}$$

is measurable.

(b) Let $f_{x_0} : X \rightarrow \mathbb{R}$, X a metric space, be defined by $f_{x_0}(x_0) = 1$ and $f_{x_0}(x) = 0$ for $x \neq x_0$, which certainly is not continuous. But this map is measurable because $f_{x_0}^{-1}(U) = \emptyset$ if $1 \notin U$ and $0 \notin U$, $f_{x_0}^{-1}(U) = X - \{x_0\}$ if $0 \in U$ but $1 \notin U$ (and this set is measurable because it is countable intersection of balls with center x_0), and $f_{x_0}^{-1}(U) = X$ if both 0 and 1 are in U .

4.10. Proposition. Let $g : X \rightarrow Y$ and $h : X \rightarrow Z$ be maps from a measurable space X into topological spaces Y, Z . Let $f = (g, h) : X \rightarrow Y \times Z$ be defined by $f(x) = (g(x), h(x))$, and $Y \times Z$ be equipped with the product topology. Then:

(a) f measurable $\implies g, h$ measurable.

(b) Assume that every open set in $Y \times Z$ is a countable union of open sets $V \times W$ with $V \subset Y$ open and $W \subset Z$ open. Then g and h measurable $\implies f$ measurable.

Proof. (a) The projections $p_Y : Y \times Z \rightarrow Y$ and $p_Z : Y \times Z \rightarrow Z$ are continuous and thus measurable. So the compositions $g = p_Y \circ f$ and $h = p_Z \circ f$ are measurable. (b) For open sets V, W in Y, Z we have

$$f^{-1}(V \times W) = g^{-1}(V) \cap h^{-1}(W),$$

and thus $f^{-1}(V \times W)$ is measurable. So the result follows because countable unions of measurable sets are measurable. ■

4.11. Example. The topological assumption in (b) is satisfied for topological spaces with a countable base of the topology. Those spaces are called *separable*. For example a metric space with a countable dense subset is separable because we can take for the bases the set of balls with center in the dense subset and rational radii. Note that if \mathcal{B}_X respectively \mathcal{B}_Y are bases for the topology on X respectively Y then $\mathcal{B}_X \times \mathcal{B}_Y$ is a basis for the topology on $X \times Y$. Finite dimensional normed vector spaces, or Hilbert spaces with a denumerable Hilbert basis (also called *separable Hilbert spaces* in the literature) are examples of separable topological spaces. (In a Hilbert space with denumerable Hilbert basis, every element has a unique representation as a series in the Hilbert basis, thus can be approximated by partial sums with coefficients in $\mathbb{Q} + i\mathbb{Q} \subset \mathbb{C}$, and the set of these elements is countable.)

4.12. Corollary. $f : X \rightarrow \mathbb{C}$ is measurable $\iff \operatorname{Re}(f) : X \rightarrow \mathbb{R}$ and $\operatorname{Im}(f) : X \rightarrow \mathbb{R}$ are measurable.

4.13. Proposition. (a) Let f be a measurable map from a measurable space X into a normed vector space E . Then $|f| : X \rightarrow \mathbb{R}$ is measurable.

(b) The set of measurable functions from a measurable space X into a finite dimensional vector space (or more generally into a separable normed vector space) is a vector space E .

(c) The set of measurable functions $X \rightarrow \mathbb{C}$ is a vector space and fg is measurable for measurable functions f and g .

Proof. (a) follows from $|\cdot| : E \rightarrow \mathbb{R}$ continuous. (b) follows from (b) of the previous proposition and continuity of the addition map $E \times E \rightarrow E$ and multiplication by scalars $\mathbb{K} \times E \rightarrow E$ with $\mathbb{K} = \mathbb{R}$ respectively \mathbb{C} for the case of

real respectively complex vector spaces. (c) follows from the continuity of the product $\mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$. We always use that continuous implies measurable. ■

4.14. Theorem. *Let f be a map from a measurable space X into a metric space Y . If there exists a sequence $\{f_n\}$ of measurable mappings $f_n : X \rightarrow Y$ such that $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ (**pointwise limit**) then f is measurable.*

Proof. Let $U \subset Y$ open and $x \in f^{-1}(U)$. Because $\lim_{k \rightarrow \infty} f_k(x) = f(x)$, for k sufficiently large $f_k(x) \in U$ and thus $x \in f_k^{-1}(U)$. Thus for each m ,

$$f^{-1}(U) \subset \bigcup_{k=m}^{\infty} f_k^{-1}(U).$$

(If $x \in f^{-1}(U)$ pick $k \geq m$ large enough such that $x \in f_k^{-1}(U)$.) Thus

$$f^{-1}(U) \subset \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} f_k^{-1}(U).$$

Now suppose that A is closed and for all positive integers m ,

$$x \in \bigcup_{k=m}^{\infty} f_k^{-1}(A).$$

Thus for each m there exists $k \geq m$ such that $f_k(x) \in A$, and thus a subsequence of $f_n(x)$ is completely contained in A , and thus converges to a point in A . But the subsequence and $f_n(x)$ have the same limit and thus $f(x) \in A$. This shows that for A closed

$$\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} f_k^{-1}(A) \subset f^{-1}(A).$$

Let $V \subset Y$ be open. For each positive integer n let

$$A_n := \{y \in Y : d(y, \mathfrak{C}V) \geq \frac{1}{n}\} \subset Y,$$

where d is the metric on Y . Then A_n is closed because it is the preimage of the closed set $[\frac{1}{n}, \infty)$ under the continuous map $Y \ni y \mapsto d(y, \mathfrak{C}V) \in \mathbb{R}$ (see e. g. Lang, II, §5, Exercise 6). Similarly let

$$V_n := \{y \in Y : d(y, \mathfrak{C}V) > \frac{1}{n}\} \subset Y,$$

which is open being preimage of the open set $(\frac{1}{n}, \infty)$ under a continuous map. Obviously

$$V_n \subset A_n$$

and thus

$$V = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} V_n$$

(Note the following: $A_n \subset \bigcup_{n=1}^{\infty} V_n$ follows because $d(y, \mathfrak{C}V) \geq \frac{1}{n} > \frac{1}{2n}$ and thus $y \in A_n \implies y \in V_{2n}$. The first equality holds because V open $\implies \mathfrak{C}V$ closed, and thus for $y \in V$ we have $d(y, \mathfrak{C}V) > 0$, and so for some n , $d(y, \mathfrak{C}V) > \frac{1}{n}$, and $y \in V_n$. Also $V_n \subset V$ for all n , because $y \in \mathfrak{C}V \implies d(y, \mathfrak{C}V) = 0$.) Thus there are the inclusions

$$f^{-1}(V) = \bigcup_{n=1}^{\infty} f^{-1}(A_n) \supset \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} f_k^{-1}(A_n) \supset \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} f_k^{-1}(V_n),$$

because for all k and n : $f_k^{-1}(A_n) \supset f_k^{-1}(V_n)$. On the other hand:

$$f^{-1}(V) = \bigcup_{n=1}^{\infty} f^{-1}(V_n) \subset \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} f_k^{-1}(V_n)$$

Because set equality follows, we see that $f^{-1}(V)$ is a countable union of a countable intersection of a countable union of the measurable sets $f_k^{-1}(V_n)$, and thus $f^{-1}(V)$ is measurable. This shows that f is measurable. ■.

The previous result can be stated briefly as: *Pointwise limits of measurable functions into metric spaces are measurable.*

4.15. Example. Let $f : X \rightarrow \mathbb{R}$ be a pointwise limit of measurable functions. Then $f^{-1}(J) \subset X$ is measurable for $J = (t, \infty)$ or $J = [t, \infty)$, and $t \in \mathbb{R}$.

4.16. Definition. A map f from a measurable space X into a set Z is *simple* if $f(X) \subset Z$ is a finite set, and for each $v \in Z$ the set $f^{-1}(\{v\}) \subset X$ is measurable.

Thus for a simple map $f : X \rightarrow Z$, $X = \bigcup_{i=1}^m X_i$ is a disjoint union with X_i measurable, and f maps X_i to $v_i \in Z$ for $i = 1, \dots, m$. The set of simple maps $X \rightarrow E$ for E a Banach space, is a vector space, because the set of values of $f + g$ and cf is finite if both f and g only take finitely many values, c a scalar. Also, if $v \in Z$ then $(f + g)^{-1}(v)$ is the union, over all $v_1, v_2 \in Z$ such that $v = v_1 + v_2$, of the sets $f^{-1}(v_1) \cap g^{-1}(v_2)$, which thus is measurable, because $f^{-1}(v_1)$ respectively $g^{-1}(v_2)$ can be nonempty for only finitely many values. Note that simple maps into measurable spaces Z are measurable because for each measurable set $B \subset Z$

$$f^{-1}(B) = \bigcup_{v \in B} f^{-1}(\{v\}),$$

and the union is finite because $f^{-1}(\{v\})$ is nonempty for only finitely many $v \in Z$. Note that for a simple real-valued function f , also $|f|$ is simple (because a union of two measurable sets is measurable). It follows that also $\max(f, g) = \frac{1}{2}(f + g + |f - g|)$ is simple for simple f, g .

The proof of the following theorem contains the basic intuition for the *Lebesgue integral* versus the *Riemann integral*. Note that the subdivision or partitioning happens in the codomain of the function, and **not** in the domain like in the theory of the Riemann integral.

4.17. Theorem. *A map f from a measurable space into a finite dimensional vector space is measurable if and only if it is a pointwise limit of simple maps.*

Proof. \Leftarrow follows from the previous theorem. \Rightarrow : Using the result about maps into products it follows that we can assume $E = \mathbb{R}$ and $f : X \rightarrow \mathbb{R}$. For each positive integer $n \geq 1$ cut $[-n, n]$ into disjoint intervals of equal length $\frac{1}{n}$ (so there are $N = 2n^2$ subintervals), denoted J_1, \dots, J_N where each J_k has the form $[a, b)$ for $a < b$ real numbers, $k = 1, \dots, N$. Let $J_{N+1} := \{t \in \mathbb{R} : t \geq n \text{ or } t < -n\}$. Then

$$A_k := f^{-1}(J_k)$$

are disjoint measurable sets with union X , $k = 1, \dots, N+1$. Define $\psi_n : X \rightarrow \mathbb{R}$ by setting $\psi_n(A_k) := \inf_{A_k} f$ for $k = 1, \dots, N$ (i. e. ψ_n is constant on A_k with value $\inf_{A_k} f$). Now $A_{N+1} = B \cup B'$ with $B = \{x \in X : f(x) \geq n\}$ and $B' = \{x \in X : f(x) < -n\}$. Define $\psi_n(B) = n$ and $\psi_n(B') = -n$. Then $\{\psi_n\}$ is a sequence of simple functions, which converges pointwise to f : In fact given $x \in X$ and $\varepsilon > 0$. Note that if $f(x) \in J_k$ for some $k = 1, \dots, N$ and a given n , then $|\inf_{A_k} f - f(x)| \leq \frac{1}{n}$ because $\inf_{A_k} f$ is the infimum over those values $f(y)$ for $y \in A_k = f_k^{-1}(J_k)$ and thus for all these y , $f(y) \in J_k$. Thus for all n sufficiently large, because we can assume that $\psi_n(x) = \inf_{A_k} f$ for $x \in A_k$, we get $|\psi_n(x) - f(x)| < \frac{1}{n} < \varepsilon$. ■

4.18. Remark. Let $f : X \rightarrow \mathbb{R}_{\geq 0}$ be a nonnegative real valued measurable map. Then f is pointwise limit of an increasing sequence of simple maps. *Proof.* The functions ψ_n defined above are all $\leq f$ by construction (here we need $f \geq 0$ because otherwise $\psi_n > f$ on B' possibly nonempty, and the ψ_n will be decreasing on the negative part). Note that the maximum of two simple functions is a simple function. Thus if

$$\varphi_n := \max(\psi_1, \dots, \psi_n)$$

for all n , then $\{\varphi_n\}$ is an increasing sequence of simple maps with limit f .

4.19. Definition. Let $[0, \infty] = [0, \infty) \cup \{\infty\}$ be the ordered set with the usual ordering of real numbers in $[0, \infty)$ and $a < \infty$ for all $a \in [0, \infty)$. Define addition and multiplication in $[0, \infty]$ by extending the usual multiplication and addition of real numbers by $\infty \cdot 0 = 0 \cdot \infty = 0$, $\infty \cdot a = a \cdot \infty = \infty$ for all $0 < a \leq \infty$, and $\infty + a = a + \infty = \infty$ for $0 \leq a \leq \infty$. Then associativity, distributivity and commutativity holds for these operations on $[0, \infty]$. (Note that the cancellation law is not satisfied!). By defining neighborhoods of ∞ to be sets of the form $(a, \infty]$ for $0 < a < \infty$, convergence in $[0, \infty]$ can be defined in the obvious way. Then the sum of a sequence of elements in $[0, \infty]$ can be viewed to converge to a number ≤ 0 or to ∞ . Also continuity of maps into $[0, \infty]$ is defined in the obvious way.

4.20. Definition. Let (X, \mathcal{M}) be a measurable space. A *positive measure* on \mathcal{M} (or on X) is a map

$$\mu : \mathcal{M} \rightarrow [0, \infty]$$

such that $\mu(\emptyset) = 0$ and if $\{A_n\}$ is any sequence of **mutually disjoint** measurable sets then

$$\mu \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n),$$

i.e. μ is *countably additive*. For $A \in \mathcal{M}$, $\mu(A) \in [0, \infty]$ is the *measure* or μ -*measure* of A . The triple (X, \mathcal{M}, μ) is called a *measured space*. Often we only write X if \mathcal{M} and μ are understood.

4.21. Examples. (a) Let $x_0 \in X$. Define $\mu(A) = 1$ if $x_0 \in A$, otherwise define $\mu(A) = 0$. This defines a measure on the σ -algebra of all subsets of X , which is called the *Dirac measure with respect to x_0* .

(b) Define $\mu(A) = |A|$ = number of elements in A if A is finite, and define $\mu(A) = \infty$ if the set is infinite. This is again a measure on the σ -algebra of all subsets of X , and is called the *counting measure*.

4.22. Proposition. (a) $A \subset B \implies \mu(A) \leq \mu(B)$

(b) If $\{A_n\}$ is an increasing sequence of measurable sets and $A = \bigcup_{n=1}^{\infty} A_n$ then

$$\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n).$$

(c) For an arbitrary sequence of measurable sets $\{A_n\}$

$$\mu \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Proof. (a) Additivity holds with respect to finite unions because we can take almost all A_n to be empty. Thus with $B = A \cup (B - A)$ we get $\mu(B) = \mu(A) + \mu(B - A) \geq \mu(A)$ because of positivity. For (b) let

$$A = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \dots \cup (A_{n+1} - A_n) \cup \dots$$

be the corresponding disjoint union then with $A_0 := \emptyset$ we get

$$\mu(A) = \lim_{n \rightarrow \infty} \sum_{n=0}^N \mu(A_{n+1} - A_n) = \lim_{N \rightarrow \infty} \mu(A_N),$$

which shows the result. (Note that we use finite additivity under the limit.) For (c) note that by (a)

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2 - A_1) \leq \mu(A_1) + \mu(A_2)$$

and by induction

$$\mu(A_1 \cup \dots \cup A_n) \leq \mu(A_1) + \dots + \mu(A_n).$$

Since $A_1 \cup \dots \cup A_n$ is an increasing sequence by taking limit and (b) we get the result. ■

4.23. Proposition. $\mu : \mathcal{M} \rightarrow [0, \infty]$ is a measure if and only if $\mu(\emptyset) = 0$, μ is finitely additive, and for each increasing sequence $\{A_n\}$ of measurable sets with union A , we have $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$.

Proof. \implies : follows immediately from the previous theorem. \impliedby : Countable additivity for a family $\{A_n\}$ of disjoint measurable sets follows by applying the assumption to the increasing sequence of sets $B_n = A_1 \cup \dots \cup A_n$ using finite additivity under the limit. ■

4.24. Proposition. If $\{A_n\}$ is a decreasing sequence of measurable sets $A_{n+1} \subset A_n$, and if some A_n has finite measure, then

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$$

Proof. Let $A := \bigcap_{n=1}^{\infty} A_n$. Without restriction we can assume that A_1 has finite measure. (Note that $\mu(B) \geq \mu(A)$ if $B \supset A$. So the sequence $\mu(A_n)$ might start with terms ∞ . But by assumption it will become a sequence of real numbers at some point. Just relabel the sets correspondingly.) Consider the ascending sequence $A_1 - A_n$ with union $A_1 - A$. Then for each n , $\mu(A_1 - A_n) =$

$\mu(A_1) - \mu(A_n)$. Thus we get $\mu(A_1) = \lim_{n \rightarrow \infty} \mu(A_1 - A_n) + \lim_{n \rightarrow \infty} \mu(A_n)$ because we know that the first limit exists. Then by the previous results we get

$$\mu(A_1) = \mu(A_1 - A) + \lim_{n \rightarrow \infty} \mu(A_n) = \mu(A_1) - \mu(A) + \lim_{n \rightarrow \infty} \mu(A_n). \quad \blacksquare$$

4.25. Remark. The result can be wrong if we do not assume $\mu(A_n)$ is finite for some n . In fact let $A_n = [n, \infty)$. Then $\mu(A_n) = \infty$ for all n but $\bigcap_{n=1}^{\infty} A_n = \emptyset$ and thus $\mu(\bigcap_{n=1}^{\infty} A_n) = \mu(\emptyset) = 0$.

4.26. Definition. A property of elements of X is said to hold *almost everywhere* or *for almost all* $x \in X$ if there exists a set S of measure zero, i. e. $\mu(S) = 0$, such that the property holds for all $x \in X - S$. This is often abbreviated a.e. or μ -a.e. to emphasize the dependence on the measure.

4.27. Example. Let $f : X \rightarrow \mathbb{R}$. Then $f \geq 0$ a. e. if $f(x) \geq 0$ for all $x \in X - S$ and $\mu(S) = 0$.

4.28. Definition. (a) A *partition* of a set A with $\mu(A) < \infty$ is a finite sequence $\{A_i\}_{i=1, \dots, r}$ of measurable disjoint sets with $A = \bigcup_{i=1}^r A_i$.

(b) A map f from a measured space (X, \mathcal{M}, μ) into a Banach space E is a *step map* if there exists a set $A \subset X$ with $\mu(A) < \infty$ and a partition of A such that $f(x) = 0$ for all $x \in X - A$ and $f(A_i)$ has just one element for $i = 1, \dots, r$. Then f is also called *step with respect to the partition of A* .

The set of all step maps from X to E is denoted $\text{St}(\mu, E)$ or $\text{St}(\mu)$ if E is understood.

4.29. Remark. For each subset $A \subset X$ let $\chi_A : X \rightarrow \mathbb{R}$ be the *characteristic function* on A , i. e. $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$. Then a function, which is step with respect to a partition $\{A_i\}_{i=1, \dots, r}$ of disjoint sets is of the form

$$\sum_{i=1}^r c_i \chi_{A_i},$$

where $\mu(A_i) < \infty$ and $c_i \in E$ for $i = 1, \dots, r$.

4.30. Proposition. $\text{St}(\mu, E)$ is a vector space, and $f \in \text{St}(\mu, E) \implies |f| \in \text{St}(\mu, E)$. Also, $f, g \in \text{St}(\mu, \mathbb{C}) \implies fg \in \text{St}(\mu, \mathbb{C})$.

Proof. If $\{A_i\}$ and $\{B_j\}$ are two partitions of a set A then $\{A_i \cap B_j\}$ is also a partition. Also, if $f(x) = 0$ for $x \in X - A$ and $g(x) = 0$ for $x \in X - B$ then $A \cup B$ has finite measure, and there is a partition of $A \cup B$ such that f, g are step with respect to this partition of $A \cup B$. This shows that $\text{St}(\mu, E)$ is a vector

space. The last claim follows using the same idea. If f is step with respect to some partition then $|f|$ is step with respect to the same partition. ■

4.31. Remark. The vector space $\text{St}(\mu, E)$ can also be described as the vector space of all maps $\sum_{i=1}^r c_i \chi_{A_i}$ with A_i sets of finite measure (but not necessarily disjoint), and $c_i \in E$. This is obvious from the previous proposition.

4.32. Definition. A mapping f from a measured space X into a Banach space E is μ -measurable if it is the pointwise limit of a sequence of step maps almost everywhere, i. e. there exists a set $Z \subset X$, $\mu(Z) = 0$ and a sequence $\{\varphi_n\}$ with $\varphi_n \in \text{St}(\mu, E)$ such that $\varphi_n(x) \rightarrow f(x)$ for all $x \in X - Z$. The sequence $\{\varphi_n\}$ is called an approximating sequence for f .

If $f : X \rightarrow E$ is μ -measurable, $F \subset E$ is a subspace such that $f(A) \subset F$ then the induced map $f|_A : A \rightarrow F$ is μ -measurable: If $\{\varphi_n\}$ is a sequence in $\text{St}(\mu, E)$, which converges a.e. to f then $\{\varphi_n|_A\}$ is a sequence of step mappings defined on A with values in E . Replace the constant values c_i of φ_n on the set $A \cap A_i$ of the partition given for φ_n by some element $d_i \in F$ such that $|\inf_{y \in F} d(y, c_i) - d_i| < \frac{1}{n}$. The resulting sequence $\{\psi_n\}$ will be in $\text{St}(\mu, F)$ and approximate $f|_A : A \rightarrow F$. Also, if $g : E \rightarrow F$ is a continuous map between Banach spaces then $g \circ f$ is μ -measurable (use $\{g \circ \varphi_n\}$ for the approximating sequence).

4.33. Proposition. (a) *The set of μ -measurable maps from X to E is a vector space.*

(b) *If $f : X \rightarrow E$ and $g : X \rightarrow F$ are μ -measurable maps into Banach spaces E, F , and if $\mathbf{m} : E \times F \rightarrow G$ is a continuous bilinear map of Banach spaces then their \mathbf{m} -product $\mathbf{m} \circ (f, g) : X \rightarrow G$ is μ -measurable.*

(c) *If $f : X \rightarrow E$ is μ -measurable then $|f| : X \rightarrow \mathbb{R}$ is μ -measurable.*

(d) *If $f, g : X \rightarrow \mathbb{C}$ are μ -measurable then $fg : X \rightarrow \mathbb{C}$ is μ -measurable.*

(e) *If f is μ -measurable and $f(x) \neq 0$ for all x then $\frac{1}{f}$ is μ -measurable.*

Proof. (a) Suppose that Z_f and Z_g are μ -zero sets and $\{\varphi_n\}$ respectively $\{\psi_n\}$ are sequences in $\text{St}(\mu, E)$ with limits f respectively g . Then $Z_f \cup Z_g$ is a μ -zero set, and $\{\varphi_n + \psi_n\}$ is a sequence in $\text{St}(\mu, E)$, which converges to $f + g$ on $X - (Z_f \cup Z_g)$. Obviously cf is μ -measurable if f is μ -measurable and c is a scalar. (b) Using a common refinement we see that if $\varphi \in \text{St}(\mu, E)$ and $\psi \in \text{St}(\mu, F)$ then $\mathbf{m}(\varphi, \psi) \in \text{St}(\mu, G)$. Thus, with approximating sequences and notation similar to (a), $\mathbf{m}(\varphi_n, \psi_n)$ is a sequence in $\text{St}(\mu, G)$ converging to $\mathbf{m}(f, g)$ on $X - (Z_f \cup Z_g)$. (c) and (d) are proved in the same way using

the corresponding statement for step maps. (e) If $\{\varphi_n\}$ is a sequence of step maps converging pointwise to f on a set then $\{\psi_n\}$ defined by $\psi_n(x) := \frac{1}{\varphi_n(x)}$ if $\varphi_n(x) \neq 0$ and $\psi_n(x) = 0$ otherwise, is a sequence of step functions converging pointwise to $\frac{1}{f}$ on this set (Note that if $\{\varphi_n\}$ is step with respect to a partition $\{A_i\}$ of a set A of finite measure then $\{x : \varphi_n(x) = 0\}$ is a union of $X - A$ and sets A_i . Thus ψ_n is step with respect to the same partition.) ■

4.34. Definition. A measurable subset Y of X is called σ -finite if it is countable union of sets of finite measure. We also say that μ is σ -finite on Y . We say μ is σ -finite if it is σ -finite on X .

4.35. Proposition. Suppose that $f : X \rightarrow E$ is μ -measurable. Then

- (a) $f(x) = 0$ for x outside a countable union of sets of finite measure.
- (b) There exists a set Z with $\mu(Z) = 0$ such that $f(X - Z) \subset \overline{D}$, where $D \subset E$ is countable (i. e. the image of $X - Z$ is contained in separable subspace of E).

Proof. (a) For an approximating sequence of step maps $\{\varphi_n\}$ for f , let for each n , $\{A_n\}$ be the set with $\mu(A_n) < \infty$ and $\varphi_n = 0$ outside A_n . Let $A := \cup_{n=1}^{\infty} A_n$, which is a countable union of sets of finite measure. If $x \notin A$ then $x \notin A_n$ for all n and thus $\varphi_n(x) = 0$ for all n . Thus $f(x) = 0$. For the proof of (b) define $D \subset E$ to be the set of all values in an approximating sequence of f . This is a countable set. Let Z be the set with $\mu(Z) = 0$ for which the approximating sequence is pointwise converging to f . Then $f(X - Z) \subset \overline{D}$ because each $f(x)$ for $x \in X - Z$ is contained in a ball of radius ε with center c for some $c \in D$ (Just find n sufficiently large such that $|f(x) - \varphi_n(x)| < \varepsilon$, and let $c := \varphi_n(x)$.) ■

The following result tries to clarify the relation between *measurable* and μ -*measurable*. Recall that we have already proved that a map f from a measurable set X into a finite dimensional vector space is measurable if and only if it is pointwise limit of a sequence of *simple maps*. The **subtle** difference between simple maps and step maps comes from the finite measure of the partition sets, on which the function is not 0. In fact a simple map is a step map if $f^{-1}(v)$ is a set of finite measure for each $v \neq 0$. On the other hand a step map is always a simple map. Thus because a μ -measurable map $f : X \rightarrow E$ is pointwise limit a.e. of a sequence of simple maps it's restriction to $X - Z$ for a set of measure 0 will be measurable. Note that on a set of finite measure simple maps are step maps, and so measurable implies μ -measurable. This distinction is visible in the quite difficult proof of the following theorem.

4.36. Technical Theorem. Let $f : X \rightarrow E$ be a mapping from a measured space into a Banach space. Then the following two conditions are equivalent:

- (i) $f(x) = 0$ outside a σ -finite subset of X , and there exists a set $Z \subset X$ with $\mu(Z) = 0$, and a countable set $D \subset E$ such that $f|(X - Z)$ is measurable, and $f(X - Z) \subset \overline{D}$.
- (ii) f is μ -measurable, i.e. is a pointwise limit a.e. of a sequence of step maps.

Proof. (ii) \implies (i) has been shown above. Conversely, assume (i). Replace X by a σ -finite subset outside of which $f(x) = 0$. So we can assume that $X = \cup_{k=1}^{\infty} X_k$ with X_k disjoint (inductively make them disjoint if they are not to begin with) and $\mu(X_k) < \infty$ for $k = 1, 2, \dots$. If we can show that $f|X_k$ is μ -measurable for all k , then for each k can find a sequence of step maps $\{\varphi_j^{(k)}\}_{j=1,2,\dots}$ on X_k , which converge a.e. to $f|X_k$. Then we can define $\varphi_n := \varphi_n^{(k)}$ on X_k for $k = 1, \dots, n$, and $\varphi_n(x) = 0$ for $x \notin X_1 \cup \dots \cup X_k$. Each φ_n is a step map, and the sequence $\{\varphi_n\}$ converges a.e. to f . Thus we only have to show that $f : X \rightarrow E$ is μ -measurable for a set X of **finite** measure. Note that for E a finite dimensional vector space we know this by the considerations above. It remains to show that the assumption of finite dimensionality can be replaced by the assumption that the image is contained in a separable subspace of E . This turns out to be quite difficult (Don't worry if you get lost...!) Let $D = \{v_k\}$ be a countable subset of E such that $\overline{D} \supset f(X)$. For each positive integer n let $B_{1/n}(v_k)$ be the ball of radius $1/n$ centered at v_k . Then the union of these balls for $k = 1, 2, \dots$ covers $f(X)$. In fact, if $z \in f(X)$ then $z = \lim z_n$ where $\{z_n\}$ is a sequence in D . Thus given $\varepsilon > 0$ we know that $|z - z_i| < \varepsilon$ for all $i > N$. If we choose $\varepsilon < \frac{1}{n}$ we will find some $i > N$ such that $|z - z_i| < \frac{1}{n}$. But z_i is some $v_k \in D$, so $z \in B_{1/n}(v_k)$ for this k . So, for each n , we have $f(X) \subset \cup_{k=1}^{\infty} B_{1/n}(v_k)$, and thus $X = \cup_{j=1}^{\infty} f^{-1}(B_{1/n}(v_j))$. Now $\{\cup_{j=1}^k B_{1/n}(v_j)\}_{k=1,2,\dots}$ is an increasing sequence of finite measure sets (here is where we need our main assumption of measurability of the function f) and thus $\mu(\cup_{j=1}^k f^{-1}(B_{1/n}(v_j))) \rightarrow \mu(X)$ for $k \rightarrow \infty$, in particular for k large $\mu(X) - \mu(\cup_{j=1}^k f^{-1}(B_{1/n}(v_j))) < \frac{1}{2^n}$. This implies

$$\bigcup_{j=1}^k f^{-1}(B_{1/n}(v_j)) = X - Y_n$$

for some set $Y_n := X - \cup_{j=1}^k f^{-1}(B_{1/n}(v_j))$ with $\mu(Y_n) < \frac{1}{2^n}$ because by definition $X = Y_n \cup (\cup_{j=1}^k f^{-1}(B_{1/n}(v_j)))$ is a disjoint union of the two sets, giving the relation

$$\mu(X) = \mu\left(\bigcup_{j=1}^k f^{-1}(B_{1/n}(v_j))\right) + \mu(Y_n).$$

Then let $Z_n := Y_n \cup Y_{n+1} \cup \dots$ so that $\mu(Z_n) \leq \frac{1}{2^{n-1}}$, and $Z_n \supset Z_{n+1} \supset \dots$ is a decreasing sequence. On $X - Y_n = \cup_{j=1}^k f^{-1}(B_{1/n}(v_j))$ find a step function φ_n such that $|f(x) - \varphi_n(x)| < \frac{1}{n}$. Define inductively $\varphi_n(x) = v_1$ for all $x \in f^{-1}(B_{1/n}(v_1))$, $\varphi_n(x) = v_2$ for all $x \in f^{-1}(B_{1/n}(v_2) - B_{1/n}(v_1))$, and so on (recall that complements of measurable sets in measurable sets are measurable for each σ -algebra). Let $\psi_n(x) := \varphi_n(x)$ for $x \in X - Z_n \subset X - Y_n$, and $\psi_n(x) = 0$ for $x \in Z_n$. Then ψ_n is a step function, and $\psi_n(x)$ converges to $f(x)$ except possibly on the set $\cap_{n=1}^{\infty} Z_n$, which has measure 0. ■.

4.37. Remarks. (a) The sequence of step maps constructed in the previous proof converges *uniformly* outside a set of arbitrarily small measure. (Outside of Z_n we can approximate f by ψ_m for $m \geq n$ and $|\psi_m(x) - f(x)| < \frac{1}{m}$ for all $x \notin Z_n$, and so the approximation is uniform on $X - Z_n$.)

(b) If μ is σ -finite and f is complex-valued then f is μ -measurable if and only if there exists a subset Z of μ -measure 0 outside of which f is measurable.

(c) The sets Z_n have to be introduced at the end of the proof to achieve convergence outside of a *fixed* set of measure 0. The φ_n are all all $\frac{1}{n}$ -close to f outside of Y_n but we need a procedure to find *one fixed set of measure zero* for all n simultaneously. This can't be $\cup_{n=1}^{\infty} Y_n$ because all we can say is that the measure of this set is ≤ 1 . I hope this explains a little the technical construction in the proof.

(d) If we are in a reasonable setting then the condition on σ -finiteness and countable denseness hold automatically. Think about mappings from a σ -finite set into \mathbb{C} or a finite dimensional vector space or a separable Hilbert space. Then the statement to have in mind is: *μ -measurability is equivalent to measurability outside of a set of μ -measure 0.*

4.38. Proposition. *Let $\{f_n\}$ be a sequence of μ -measurable functions converging a.e. to a map f . Then f is μ -measurable.*

Proof. This follows from the previous theorem. In fact if applied to each function f_n we know that (i) is satisfied for each f_n . From this we want to deduce that (i) holds for f , and thus f is μ -measurable by (ii). Because each f_n vanishes outside of a σ -finite set, also f vanishes outside of a union of a countable union of σ -finite sets and a set of measure 0, and so f vanishes outside of a σ -finite set. If $\{f_n\}$ converges pointwise to f outside of Z_0 with $\mu(Z_0) = 0$, and each f_n is measurable outside of a set Z_n with $\mu(Z_n) = 0$ then f_n is pointwise limit of the sequence of measurable functions outside of $Z = \cup_{n=0}^{\infty} Z_n$, which is a set of measure 0. Finally the condition $f(X - Z) \subset \overline{D}$ for a countable set D

follows, because a denumerable union of sets W_k with $\overline{D_k} \supset W_k$ for each k and D_k countable is also contained in the closure of a countable set. In fact, if $\{D_k\}$ is a sequence of denumerable sets in a metric space then $\overline{\cup_{k=1}^{\infty} D_k} \supset \overline{D_n}$ for all n , and thus $\overline{\cup_{k=1}^{\infty} D_k} \supset \cup_{n=1}^{\infty} \overline{D_n}$, and since the left hand side is closed it contains also $\cup_{n=1}^{\infty} \overline{D_n}$. Thus $\overline{\cup_{k=1}^{\infty} D_k} = \overline{\cup_{n=1}^{\infty} \overline{D_n}}$ because the other inclusion is obvious by $D_n \subset \overline{D_n}$ for all n . This applies in particular to D_k countable with $\overline{D_k} \supset f_k(X - Z)$, and $f(X - Z) \subset \overline{\cup_k f_k(X - Z)}$. ■

We are now ready to define the integral for step maps. Given a map $f : X \rightarrow E$, which is step with respect to a partition $\{A_i\}_{i=1, \dots, r}$ of a finite measure set $A \subset X$ let

$$\int_X f d\mu := \sum_{i=1}^r \mu(A_i) f(A_i) \in E$$

be the *integral* of the step map (note that $\mu(A_i) \in [0, \infty)$ and $f(A_i) \in E$ and the product is multiplication by scalars in a vector space). If $B \subset X$ is measurable and f is a step map on X let $f_B := f\chi_B$. If f is step with respect to $\{A_i\}$ then f_B is step with respect to the partition $\{B \cap A_i\}$ of $B \cap A$ (which has finite measure), and define

$$\int_B f d\mu := \int_X f_B d\mu.$$

If μ is fixed then $d\mu$ will often be omitted from the notation.

4.39. Theorem. *The map*

$$\int : \text{St}(\mu, E) \rightarrow E,$$

is well-defined and linear, and satisfies the following properties:

- (i) *If A, B are disjoint then $\int_{A \cup B} f = \int_A f + \int_B f$*
- (ii) *If $E = \mathbb{R}$ and $f \leq g$ then $\int f \leq \int g$, and if $A \subset B$ and $f \geq 0$ then $\int_A f \leq \int_B f$.*
- (iii) *$|\int_A f d\mu| \leq \int_A |f| d\mu \leq \|f\| \mu(A)$, where $\|f\|$ is the supremum norm, which will be infinite in the case that f is not bounded.*

Proof. If $\{B_j\}_{j=1, \dots, s}$ is another partition of A then f is step with respect to the partition $\{A_i \cap B_j\}$, and $\sum_{j=1}^s \mu(A_i \cap B_j) f(A_i) = \mu(A_i) f(A_i)$ and similarly $\sum_{i=1}^r \mu(A_i \cap B_j) f(B_j) = \mu(B_j) f(B_j)$. Thus $\sum_{i=1}^r \sum_{j=1}^s \mu(A_i \cap B_j) f(A_i) = \sum_{j=1}^s \sum_{i=1}^r \mu(A_i \cap B_j) f(B_j)$ shows that each partition gives rise to the same integral. (Note that if $A_i \cap B_j \neq \emptyset$ then $f(A_i) = f(B_j)$.) Linearity follows easily from this. In fact, we can assume that two given step maps f and g are step with respect to a common partition, and so $f = \sum_{i=1}^r f(A_i) \chi_{A_i}$ and

$g = \sum_{i=1}^r g(A_i)\chi_{A_i}$. Thus $\int f + g = \sum_{i=1}^r \mu(A_i)(f(A_i) + g(A_i)) = \int f + \int g$. Also for c a scalar, $(cf)(A_i) = cf(A_i)$ and the result is obvious. But now (i) follows because $f_{A \cup B} = f_A + f_B$. (ii) follows from $\int f \geq 0$ if $f \geq 0$, which is clear by the very definition using positivity of the measure. To see (iii) just assume f is step with respect to $\{A_i\}$, so that $|\int_A f d\mu| = |\sum_{i=1}^r \mu(A_i)f(A_i)| \leq \sum_{i=1}^r \mu(A_i)|f(A_i)| = \int_A |f| d\mu$ and then $\sum_{i=1}^r \mu(A_i)|f(A_i)| \leq \max_i |f(A_i)| \sum_{i=1}^r \mu(A_i) = \|f\| \mu(A)$. ■

4.40. Proposition. *Let E be a Banach space with norm $|\cdot|$. Then the function*

$$\text{St}(\mu, E) \ni f \mapsto \int_X |f| = \int |f| =: \|f\|_1 \in \mathbb{R}$$

is a seminorm, called the L^1 -seminorm.

Proof. $\|f\|_1 \geq 0$ is immediate from the definition or (ii) above with $|f| \geq 0$. The triangle inequality follows on a common partition of f and g from the triangle inequality for the norm on E : $\int |f + g| = \sum_{i=1}^r \mu(A_i)|f(A_i) + g(A_i)| \leq \sum_{i=1}^r \mu(A_i)(|f(A_i)| + |g(A_i)|) = \sum_{i=1}^r \mu(A_i)|f(A_i)| + \sum_{i=1}^r \mu(A_i)|g(A_i)| = \int |f| + \int |g|$. $\int |cf| = |c| \int |f|$ is a special case of (i) above. ■

The estimate (iii) in the theorem above says that $\int : \text{St}(\mu, E) \rightarrow E$ is *continuous* with respect to $|\cdot|$ on E and the seminorm $\|\cdot\|_1$ on $\text{St}(\mu, E)$. Our approach to integration will extend the integral to a large class of functions using the ideas of *completion* and *linear extension* from Chapter 2. Recall that the completion of a vector space with seminorm is the Banach space of equivalence classes of Cauchy sequences in the vector space by the following equivalence relation: two Cauchy sequences are equivalent if they differ by a sequence converging to 0 (in the seminorm). Let $L^1(\mu)$ be the completion resulting from $\text{St}(\mu)$ with respect to the L^1 -seminorm. Then the natural map $\text{St}(\mu) \rightarrow L^1(\mu)$ maps all step maps f with $\|f\|_1 = 0$ to 0 (because those are equivalent to the step function, which is identically zero). (See also the remark following the completion theorem in Chapter 2). Because the completion is a normed vector space we cannot see functions with trivial L^1 -norm. The next step is to construct a space of **functions**, which can be considered more directly as a *seminorm completion* of $\text{St}(\mu)$. If every L^1 -Cauchy sequence of step maps would converge pointwise the limit would be a measurable function, and the space of these functions would naturally provide the candidate for the completion. (Recall that a sequence $\{f_n\}$ of step maps is L^1 -Cauchy if $\|f_n - f_m\|_1 \rightarrow 0$ for $\min(n, m) \rightarrow \infty$.) Unfortunately this is not the case and we need the following constructions.

4.41. Definition. Let $\mathcal{L}^1(\mu) = \mathcal{L}^1$ be the set of mappings $f : X \rightarrow E$ such that

there exists an L^1 -Cauchy-sequence of step mappings, which converges pointwise a.e. to f . Such a sequence will be said to *approximate* $f \in \mathfrak{L}^1$.

4.42. Proposition. $\mathfrak{L}^1(\mu)$ is a vector space, which contains $\text{St}(\mu)$.

Proof. Constant sequences are L^1 -Cauchy and so the inclusion is clear. If $\{f_n\}$ respectively $\{g_n\}$ are L^1 -Cauchy sequences converging a.e. pointwise to f respectively g then $\{f_n + g_n\}$ respectively cf_n are L^1 -Cauchy sequences converging pointwise a.e. to $f + g$ respectively cf , c a scalar. In fact, if Z_f respectively Z_g are sets with $\mu(Z_f) = \mu(Z_g) = 0$ and $f_n(x) \rightarrow f(x)$ respectively $g_n(x) \rightarrow g(x)$ for all $x \in X - Z_f$ respectively $x \in X - Z_g$ then $f_n(x) + g_n(x) \rightarrow f(x) + g(x)$ on $X - (Z_f \cup Z_g)$, and $\mu(Z_f \cup Z_g) \leq \mu(Z_f) + \mu(Z_g) = 0$. Also $cf_n(x) \rightarrow cf(x)$ for $x \in X - Z_f$. Moreover $\|(f_n + g_n) - (f_m + g_m)\|_1 \leq \|f_n - f_m\|_1 + \|g_n - g_m\|_1 \rightarrow 0$ and $\|cf_n - cf_m\|_1 = |c| \|f_n - f_m\|_1 \rightarrow 0$ for $\min(n, m) \rightarrow \infty$. ■

In the following, please pay special attention to the difference between a.e. and outside of a set of arbitrarily small measure. If a property holds a.e. it holds outside of a set of arbitrarily small measure but the converse is not necessarily true. Our goal is to extend \int from $\text{St}(\mu)$ to $\mathfrak{L}^1(\mu)$ by choosing a

4.43. Fundamental lemma of integration. *Each L^1 -Cauchy sequence $\{f_n\}$ of step maps has a subsequence which converges pointwise a.e. and satisfies the following: for each $\varepsilon > 0$ there exists a set Z with $\mu(Z) < \varepsilon$ such that this subsequence converges uniformly on $X - Z$.*

Proof. For each positive integer k there exists a positive integer N_k such that $n, m \geq N_k \implies \|f_m - f_n\|_1 < \frac{1}{2^{2k}}$. Let $g_k := f_{N_k}$, choosing the N_k inductively increasing. Then for all $m \geq n$ we have $\|g_m - g_n\|_1 < \frac{1}{2^{2n}}$. *Claim:* The series $g_1(x) + \sum_{k=1}^{\infty} (g_{k+1}(x) - g_k(x)) = \lim_{N \rightarrow \infty} g_{N+1}(x)$ converges absolutely for all x to some element in E , and the convergence is uniform except on a set of arbitrarily small measure. *Proof of Claim.* Define $Y_n := \{x \in X : |g_{n+1}(x) - g_n(x)| \geq \frac{1}{2^n}\}$. Because $g_{n+1} - g_n \in \text{St}(\mu)$ we know that $g_{n+1} - g_n$ vanishes outside of a set of finite measure and takes only finitely many values whose inverse images are measurable. It follows that the sets Y_n are measurable. Because $\frac{1}{2^n} \leq |g_{n+1} - g_n|$ we get on Y_n :

$$\frac{1}{2^n} \mu(Y_n) \leq \int_{Y_n} \frac{1}{2^n} \leq \int_X |g_{n+1} - g_n| \leq \frac{1}{2^{2n}}$$

and $\mu(Y_n) \leq \frac{1}{2^n}$. Let $Z_n := Y_n \cup Y_{n+1} \cup \dots$. Then $\mu(Z_n) \leq \sum_{i=n}^{\infty} \frac{1}{2^i} = \frac{1}{2^n} \sum_{i=0}^{\infty} \frac{1}{2^i} = \frac{1}{2^n} \cdot 2 = \frac{1}{2^{n-1}}$. If $x \notin Z_n$ or $x \in X - Z_n = \bigcap_{k=n}^{\infty} (X - Y_k)$ then $k \geq n \implies |g_{k+1}(x) - g_k(x)| < \frac{1}{2^k}$ and thus $\sum_{k=n}^{\infty} (g_{k+1}(x) - g_k(x))$ is absolutely and

uniformly convergent for $x \notin Z_n$. Note that we talk about absolute convergence of a series here, which of course implies convergence. Thus for $\varepsilon > 0$ we can find n such that $\frac{1}{2^{n-1}} < \varepsilon$, then $\mu(Z_n) < \varepsilon$, and $\{g_n\}$ converges absolutely and uniformly on $X - Z_n$. Finally let $Z_0 := \bigcap_{n=1}^{\infty} Z_n$ so that $\mu(Z_0) = 0$. If $x \notin Z_0$ or $x \in X - Z_0 = \bigcup (X - Z_n)$ then $x \notin Z_n$ for some n . For those x the series converges. This proves that the subsequence converges pointwise a.e. ■

4.44. Remark. Convergence of a *sequence* always implies *absolute convergence* by the triangle inequality if we define absolute convergence by $|a_n| \rightarrow |a|$ for $a_n \rightarrow a$. So don't get confused by Lang's statement about absolute convergence of the subsequence. Of course if the convergence of a sequence of functions $\{f_n\}$ is uniform then the convergence of $\{|f_n|\}$ is also uniform, again by the triangle inequality. For series this is different: Convergence of a series of course does not imply absolute convergence of the series. But absolute convergence of a series is *very different* from absolute convergence of its sequence of partial sums.

4.45. Main lemma. *Suppose that $\{g_n\}$ and $\{h_n\}$ are L^1 -Cauchy sequences of step mappings from X to E converging pointwise a.e. to the same map. Then $\lim \int_X g_n$ and $\lim \int_X h_n$ exist and are equal. Moreover, the sequence $\{g_n - h_n\}$ is an L^1 -null sequence, i. e. $\{g_n\}$ and $\{h_n\}$ are equivalent.*

Proof. Because $|\int g_n - \int g_m| \leq \int |g_n - g_m| = \|g_n - g_m\|_1$ it follows that $\{\int g_n\}$ is a Cauchy sequence in E , which converges since E is complete, and similarly for $\{\int h_n\}$. So we know that the limits of the integrals exist. Let $f_n := g_n - h_n$. Because $\|(g_n - h_n) - (g_m - h_m)\|_1 \leq \|g_n - g_m\|_1 + \|h_n - h_m\|_1$, $\{f_n\}$ is L^1 -Cauchy, and converges a.e. to 0. (If $g_n \rightarrow g$ pointwise on $X - Z_g$ and $h_n \rightarrow g$ pointwise on $X - Z_h$ with $\mu(Z_g) = \mu(Z_h) = 0$, then for $x \in X - (Z_g \cup Z_h)$ with $\mu(Z_g \cup Z_h) = 0$ we have $|f_n(x)| = |g_n(x) - h_n(x)| = |(g_n(x) - g(x)) - (h_n(x) - g(x))| \leq |g_n(x) - g(x)| + |h_n(x) - g(x)| \rightarrow 0$.) Let $\varepsilon > 0$. Then $\exists N$ such that $m, n \geq N \implies \|f_n - f_m\|_1 < \varepsilon$. Let A be a set with $\mu(A)$ finite and $f_N(x) = 0$ for $x \in X - A$. Then $n \geq N \implies \int_{\mathcal{C}A} |f_n| = \int_{\mathcal{C}A} |f_n - f_N| \leq \int_X |f_n - f_N| = \|f_n - f_N\|_1 < \varepsilon$. By the fundamental lemma there exists $Z \subset A$ such that $\mu(Z) < \frac{\varepsilon}{1 + \|f_N\|_1}$ and a subsequence of $\{n\}$ such that the corresponding subsequence of $\{f_n\}$ converges to 0 uniformly on $A - Z$. Then for n in the subsequence we conclude that $\int_{A-Z} |f_n| < \varepsilon$ (using uniformity, the properties of the integral and $\mu(A - Z) \leq \mu(A) < \infty$). Also for large n in the subsequence we have

$$\int_Z |f_n| \leq \int_Z |f_n - f_N| + \int_Z |f_N| \leq \|f_n - f_N\|_1 + \mu(Z) \|f_N\|_1 < 2\varepsilon$$

Thus

$$\|f_n\|_1 = \int_X |f_n| = \int_Z |f_n| + \int_{A-Z} |f_n| + \int_{\mathfrak{c}A} |f_n| < 2\varepsilon + \varepsilon + \varepsilon = 4\varepsilon.$$

This proves the lemma because we have $\|f_n\|_1 \rightarrow 0$ on the subsequence implies the same for the whole sequence because it is L^1 -Cauchy ($\|f_m\|_1 \leq \|f_n\|_1 + \|f_m - f_n\|_1$), and thus also for the original sequences:

$$\left| \int g_n - \int h_n \right| = \left| \int f_n \right| \leq \int |f_n| = \|f_n\|_1 \rightarrow 0. \quad \blacksquare$$

From the Main lemma we get immediately the following extension of \int from $\text{St}(\mu)$ to $\mathfrak{L}^1(\mu)$:

4.46. Corollary. *The map $f \in \mathfrak{L}^1(\mu) \mapsto \int_X f d\mu = \int_X f := \lim \int_X f_n d\mu$, where $\{f_n\}$ is **any** approximating sequence of step maps with pointwise a.e. limit f , is well-defined. Elements of \mathfrak{L}^1 are thus called **integrable maps**.*

The map $\mathfrak{L}^1 \ni f \mapsto \int_X f$ is linear. In fact, if $\{f_n\}$ respectively $\{g_n\}$ are approximating sequences for f respectively g then $\{f_n + g_n\}$ is an approximating sequence for $f + g \in \mathfrak{L}^1$ and $\{cf_n\}$ is an approximating sequence for $cf \in \mathfrak{L}^1$. Thus linearity follows from the properties of the limit.

4.47. Lemma. *If $f \in \mathfrak{L}^1$ with approximating sequence $\{f_n\}$ then $|f| \in \mathfrak{L}^1$ with approximating sequence $\{|f_n|\}$, and thus*

$$\int_X |f| = \lim \int_X |f_n| = \lim \|f_n\|_1.$$

Proof. The pointwise a.e. convergence $|f_n| \rightarrow |f|$ follows from the triangle inequality for the norm in E : $\||f_n| - |f|\| \leq |f_n - f|$ holds for $x \in X - Z$ and $\mu(Z) = 0$. The triangle inequality in E and positivity of the integral for step maps also implies:

$$\||f_n| - |f_m|\|_1 = \int_X \||f_n| - |f_m|\| = \int_X |f_n - f_m| = \|f_n - f_m\|_1,$$

thus $\{|f_n|\}$ is an L^1 -Cauchy sequence.. \blacksquare

Note that if $\{f_n\}$ is an approximating sequence of step maps for f then $\{f_n\}$ also L^1 -converges to f . To see this use that $\{|f_n - f_m|\}_{m=1,2,\dots}$ is an approximating sequence for $|f_n - f|$. Given $\varepsilon > 0$ there exists N such that $n, m > N$ implies $\|f_n - f_m\|_1 = \int |f_n - f_m| < \varepsilon$. Then for $n > N$:

$$\|f_n - f\|_1 = \int |f_n - f| = \lim_{m \rightarrow \infty} \int |f_n - f_m| \leq \varepsilon$$

which proves $\|f_n - f\|_1 \rightarrow 0$ for $n \rightarrow \infty$.

4.48. Theorem. *The map*

$$\mathfrak{L}^1 \ni f \mapsto \|f\|_1 := \int_X |f| dx = \lim \|f_n\|_1 \in \mathbb{R}$$

is a well-defined and defines a seminorm such that \mathfrak{L}^1 is complete (i. e. L^1 -Cauchy sequences converge).

Proof. Well-definedness follows from the previous lemma, and the seminorm properties follow from the corresponding ones for the seminorm on $\text{St}(\mu)$ and the properties of limits using approximating sequences of step mappings. We show L^1 -completeness: Let $\{f_n\}$ be a Cauchy sequence in \mathfrak{L}^1 . For each n find $g_n \in \text{St}(\mu)$ such that $\|f_n - g_n\|_1 < \frac{1}{n}$. Then $\|g_n - g_m\|_1 \leq \|g_n - f_n\|_1 + \|f_n - f_m\|_1 + \|f_m - g_m\|_1$ shows that $\{g_n\}$ is L^1 -Cauchy. By the fundamental lemma of integration, for a subsequence of $\{n\}$ the corresponding subsequence of $\{g_n\}$ converges pointwise a.e. to $f \in \mathfrak{L}^1$, and for n in the subsequence $\|f_n - f\|_1 \leq \|f_n - g_n\|_1 + \|g_n - f\|_1 < 2\varepsilon$ for n sufficiently large in the subsequence. Hence the subsequence is L^1 -convergent to f . Thus the sequence $\{f_n\}$ also is L^1 -convergent to f . (Recall from Chapter 1 the fact that if a subsequence of a Cauchy sequence in a metric space converges then the sequence itself converges and to the same limit.) ■

4.49. Remark. (a) Note that the seminorm $\|\cdot\|_1$ on \mathfrak{L}^1 is **not** a norm in general because there can be functions with $\|f\|_1 = 0$, which are not 0 (just consider a constant sequence of step maps, which is nonzero only on a set of measure 0). Thus we have the linear map of complete vector spaces:

$$\gamma : \mathfrak{L}^1(\mu) \rightarrow L^1(\mu),$$

which to $f \in \mathfrak{L}^1$ associates the equivalence class of an approximating L^1 -Cauchy sequence with $\|\gamma(f)\|_1 = \|f\|_1$. This map usually has the nontrivial kernel of maps $f \in \mathfrak{L}^1$ with $\|f\|_1 = 0$ (note that $\|\cdot\|_1$ is a norm on $L^1(\mu)$ so $\|f\|_1 = \|\gamma(f)\|_1 = 0 \iff \gamma(f) = 0$). The integral $\int : \text{St}(\mu) \rightarrow E$ extends to a continuous map $L^1(\mu)$ by the linear extension theorem from Chapter 2 (continuity is just the statement $|\int_X f| \leq \|f\|_1$). The above result show that the integral can also be *lifted* to a continuous map $\mathfrak{L}^1(\mu) \rightarrow E$.

(b) Special attention has to be paid throughout to the type of convergence. If $\{f_n\}$ is L^1 -convergent to f then $\| \|f_n\|_1 - \|f\|_1 \| \leq \|f_n - f\|_1 \rightarrow 0$ so $\int |f_n| \rightarrow \int |f|$ holds, and also $\int f_n \rightarrow \int f$ because

$|\int f_n - \int f| = |\int(f_n - f)| \leq \int |f_n - f| = \|f_n - f\|_1$. This means integral and limit can be interchanged. In the following there will be a subtle and detailed study of this if we do not have given convergence in L^1 -seminorm but various pointwise conditions are given as assumption. But note that we apply 4.50 (ix) because f is not necessarily a step function.

- 4.50. Proposition.** (i) If $f = g$ a.e. and $f \in \mathfrak{L}^1$ then $g \in \mathfrak{L}^1$, and $\int_X f = \int_X g$ and $\|f\|_1 = \|g\|_1$.
(ii) If $f \in \mathfrak{L}^1$ then $\exists g \in \mathfrak{L}^1$ measurable such that $f = g$ a.e.
(iii) $f \in \mathfrak{L}^1(\mu, E) \implies |f| \in \mathfrak{L}^1(\mu, \mathbb{R})$
(iv) $f, g \in \mathfrak{L}^1(\mu, \mathbb{R}) \implies \sup(f, g), \inf(f, g) \in \mathfrak{L}^1(\mu, \mathbb{R})$, in particular $0 \leq f^+ := \sup(f, 0) \in \mathfrak{L}^1$ and $0 \geq f^- := -\inf(f, 0) \in \mathfrak{L}^1$, where $f = f^+ - f^-$
(v) If $A \subset X$ is measurable then $f \in \mathfrak{L}^1(\mu) \implies f_A = f\chi_A \in \mathfrak{L}^1(\mu)$
(vi) If $A, B \subset X$ are measurable and $A \cap B = \emptyset$ then $\int_{A \cup B} f = \int_A f + \int_B f$.
(vii) If $E = \mathbb{R}$ and $f \leq g$ a.e. then $\int f \leq \int g$.
(viii) If $f \geq 0$ and $A \subset B$ then $\int_A f \leq \int_B f$.
(ix) $|\int_A f d\mu| \leq \int_A |f| d\mu \leq \|f\| \mu(A)$, where $\|f\| := \sup_{x \in A} |f(x)| \in [0, \infty]$.

Proof. (i): An L^1 -Cauchy sequence approximating f pointwise a.e. will also approximate g pointwise a.e. and thus $g \in \mathfrak{L}^1$. Because the integrals and L^1 -norms are defined by integrals over X of approximating \mathfrak{L}^1 -Cauchy sequences the rest of (i) is immediate. (ii): An approximating \mathfrak{L}^1 -Cauchy converges to f outside of a set Z of measure 0. If we redefine the approximating sequence on Z by the value 0 on Z then the pointwise limit of the redefined sequence of step functions is a measurable function and equal to f outside of Z . (iii) has been shown in 4.47. (iv) follows from (iii) and the formulas $\sup(f, g) = \frac{1}{2}(f + g + |f - g|)$ respectively $\inf(f, g) = \frac{1}{2}(f + g - |f - g|)$ (v): Let $\{\varphi_n\}$ be an approximating \mathfrak{L}^1 -Cauchy sequence for f then $\{\varphi_{n,A}\}$ converges a.e. to f_A , and is \mathfrak{L}^1 -Cauchy because

$$\int_X |\varphi_{n,A} - \varphi_{m,A}| \leq \int_X |\varphi_n - \varphi_m| = \|\varphi_n - \varphi_m\|_1,$$

and thus $\{\varphi_{n,A}\}$ is an approximating sequence for f_A . Then (vi) follows from $f_{A \cup B} = f_A \cup f_B$. (vii) follows by linearity from the corresponding statement for $g = 0$ a.e. To see this replace an approximating sequence of step functions $\{\varphi_n\}$ by $\{\sup(\varphi_n, 0)\}$. (viii) follows from (vii) with $f_A \leq f_B$. Finally the first inequality in (ix) follows by continuity from the corresponding inequality for approximating step mappings, and the second inequality follows by (vii) and $|f| \leq \|f\|$ and $\int_A d\mu = \int_X \chi_A d\mu = \mu(A)$. Actually this argument does not apply if $\|f\| = \infty$ or $\mu(A) = \infty$. But $\|f\| \mu(A) = \infty$ if $\mu(A) > 0$ and the

inequality holds because $\int_A f \in \mathbb{R}$. In the case $\|f\| = \infty$ and $\mu(A) = 0$ we have $\|f\|\mu(A) = 0$. But in this case we can choose an approximating sequence for $f\chi_A$, which is 0 outside of A . Thus the sequence of their integrals is constant 0 and so their limit, which is $\int_A f$. ■

Note that if A is measurable and $\mu(A) < \infty$ then χ_A is measurable because the constant sequence χ_A is an L^1 -Cauchy sequence of step maps converging to χ_A . See also 4.54 below.

Note the *abuse of notation* in (b) below where (g, h) appears as two very different mathematical objects.

4.51. Proposition. (a) Let $\lambda \in L(E, F)$ for Banach spaces E, F . Then the map $f \mapsto \lambda \circ f$ is in $L(\mathfrak{L}^1(\mu, E), \mathfrak{L}^1(\mu, F))$, and

$$\lambda\left(\int_X f d\mu\right) = \int_X \lambda \circ f d\mu.$$

(b) Given $g \in \mathfrak{L}^1(\mu, E)$ and $h \in \mathfrak{L}^1(\mu, F)$ then $f := (g, h) \in \mathfrak{L}^1(\mu, E \times F)$ (so $f(x) = (g(x), h(x))$ for all $x \in X$), and the map

$$\mathfrak{L}^1(\mu, E) \times \mathfrak{L}^1(\mu, F) \ni (g, h) \mapsto f \in \mathfrak{L}^1(\mu, E \times F)$$

is a bilinear toplinear mapping of Banach spaces, i. e. an element in

$$Lis(\mathfrak{L}^1(E, \mu) \times \mathfrak{L}^1(\mu, F), \mathfrak{L}^1(\mu, E \times F)).$$

Moreover $\int f = (\int g, \int h)$.

Proof. (a) follows from the corresponding statement for step mappings by approximation and continuity. For (b), in order to see that $(g, h) \mapsto f$ is a linear mapping with the claimed properties approximate each by \mathfrak{L}^1 -Cauchy sequences of step mappings and consider the corresponding sequence $\{(\varphi_n, \psi_n)\}$ of step mappings in $\text{St}(\mu, E)$ which will be an approximating sequence for f . Linearity and the property about the integral follows by approximation and continuity. The inverse map is defined by using the linear projections $p_E : E \times F \rightarrow E$ and $p_F : E \times F \rightarrow F$ inducing linear mappings of \mathfrak{L}^1 -spaces by (a) such that $p_E(\int_X f d\mu) = \int_X p_E \circ f d\mu$ with $p_E \circ f = g$, correspondingly for F . Details of the proof are cumbersome and will be deferred to an exercise. ■

4.52. Examples. The previous proposition applies in particular to $\mathbf{C} = \mathbf{R} \times \mathbf{R}$, in which case we have $\int (g + ih) = \int g + i \int h$ for real functions g, h . Note that a sequence of complex step functions approximates $g + ih$ if and only if the real

part of the sequence approximates g and the imaginary part approximates h . Similarly the result applies to \mathbb{R}^n by induction.

4.53. Lemma. *Let $f \in \mathfrak{L}^1(\mu)$ and $c > 0$. Then the set*

$$S_c = S_c(f) := \{x \in X : |f(x)| \geq c\}$$

has finite measure, and f vanishes outside a σ -finite set.

Proof. It follows from the fundamental lemma of integration that there exists a set Z with $\mu(Z) < \varepsilon$ and a subsequence of an approximating \mathfrak{L}^1 -Cauchy sequence, denoted $\{\varphi_n\}$, which converges uniformly to f on $X - Z$. Thus $|\varphi_n(x)| \geq \frac{c}{2}$ for $x \in S_c - Z$. Thus $S_c - Z \subset |\varphi_n|^{-1}[\frac{c}{2}, \infty)$ and the set on the right has finite measure. Thus $S_c - Z$ and so also S_c has finite measure. By applying this to $c = \frac{1}{k}$, $k = 1, 2, \dots$ we see that that $|f(x)| > 0$ on a countable union of sets of finite measure, and thus $f(x) = 0$ outside a σ -countable set. ■

4.54. Example. If $A \subset X$ and $\chi_A \in \mathfrak{L}^1$ then $\mu(A) = \mu(S_{\frac{1}{2}}(\chi_A))$ has finite measure.

The following result extends the fundamental lemma of integration from *step maps* to integrable functions.

4.55. Theorem. *Suppose $\{f_n\}$ is a Cauchy sequence in the semi-normed space \mathfrak{L}^1 and $\{f_n\}$ is L^1 -convergent to $f \in \mathfrak{L}^1$. Then there exists a subsequence of $\{f_n\}$, which converges to f a.e., such that for each $\varepsilon > 0$ there exists $Z \subset X$ with $\mu(Z) < \varepsilon$ such that the convergence is uniform on $X - Z$.*

Proof. $\{f_n - f\}$ is L^1 -Cauchy sequence of functions in \mathfrak{L}^1 , which is L^1 -convergent to 0. So assume without restriction that $\{f_n\}$ is an L^1 -Cauchy sequence, which is L^1 -convergent to 0. Thus $\|f_n\|_1 \rightarrow 0$ for $n \rightarrow \infty$. So by selecting a subsequence we can assume that $\|f_n\|_1 < \frac{1}{2^{2n}}$.

Let $Y_n := \{x \in X : |f_n(x)| \geq \frac{1}{2^n}\} \subset X$. Then

$$\frac{1}{2^n} \mu(Y_n) \leq \int_{Y_n} |f_n| \leq \int_X |f_n| \leq \frac{1}{2^{2n}},$$

which implies $\mu(Y_n) \leq \frac{1}{2^n}$. Let $Z_n := Y_n \cup Y_{n+1} \cup \dots$. Then $\mu(Z_n) \leq \frac{1}{2^{n-1}}$. If $x \notin Z_n$ or $x \in X - (Y_n \cup Y_{n+1} \cup \dots) = (X - Y_n) \cap (X - Y_{n+1}) \cap \dots$ then for $k \geq n$, $x \notin Y_n$ and thus $|f_k(x)| < \frac{1}{2^k}$. Thus $\{f_k\}$ converges to 0 uniformly on $X - Z_n$ and the measure of Z_n can be made arbitrarily small. If $Z = \bigcap_{n=1}^{\infty} Z_n$ then $\mu(Z) = 0$ and $f_n \rightarrow 0$ pointwise on $X - Z$. ■

Note that the function f in the assumptions of the last theorem always exists because we know that \mathfrak{L}^1 is complete with respect to the L^1 -seminorm, see 4.48.

4.56. Corollary. For $f \in \mathfrak{L}^1$: $\|f\|_1 = 0 \iff f = 0$ a.e.

Proof. \implies : Suppose $\|f\|_1 = 0$. Then the sequence $\{0, 0, \dots\}$ is L^1 -convergent to f , and by 4.55 a subsequence converges to f pointwise a.e., so $f = 0$ a.e.
 \impliedby : Let $f = 0$ for $x \notin Z$ with $\mu(Z) = 0$. Then

$$\|f\|_1 = \int_Z |f| + \int_{X-Z} |f| \leq \mu(Z)\|f\| + \int_{X-Z} 0 = 0,$$

because even if $\|f\| = \infty$ we have $\mu(Z)\|f\| = 0$. ■

4.57. Corollary. The kernel of $\gamma : \mathfrak{L}^1(\mu) \rightarrow L^1(\mu)$ is the space of maps $f \in \mathfrak{L}^1(\mu)$ such that $f = 0$ a.e.

Proof. Let $f \in \ker(\gamma)$ then $\gamma(f) = 0$ in $L^1(\mu)$ thus $\|\gamma(f)\|_1 = \|f\|_1 = 0$ thus $f = 0$ a.e. Conversely, if $f = 0$ a.e. then $\|f\|_1 = \|\gamma(f)\|_1 = 0$ and thus $\gamma(f) = 0$ in $L^1(\mu)$ because $\|\cdot\|$ is a norm on $L^1(\mu)$. Thus $\|f\|_1 = 0$ and so $\gamma(f) = 0$ in $L^1(\mu)$. ■

Define two maps $f, g : X \rightarrow E$ equivalent if $f = g$ a.e. Thus $L^1(\mu)$, the norm completion of \mathfrak{L}^1 under the L^1 -norm, is the quotient of $\mathfrak{L}^1(\mu)$ by the kernel of γ .

4.58. Corollary. Let $\{f_n\}$ be a Cauchy sequence in \mathfrak{L}^1 converging to f a.e. Then $f \in \mathfrak{L}^1$ and the convergence is also in the L^1 -seminorm.

Proof. We already know that $\{f_n\}$ is L^1 -convergent to some $g \in \mathfrak{L}^1$ because \mathfrak{L}^1 is complete in the L^1 -seminorm. By the previous theorem then a subsequence converges a.e. to g . Because this subsequence also converges a.e. to f it follows that $f = g$ a.e. Thus $f \in \mathfrak{L}^1$ and is also L^1 -limit of $\{f_n\}$ by the previous theorem. ■

4.59. Monotone convergence theorem. Let $\{f_n\}$ be an increasing respectively decreasing sequence of real valued functions in \mathfrak{L}^1 such that the sequence of integrals $\{\int_X f_n d\mu\}$ is a sequence of real numbers, which is bounded from above respectively below. Then $\{f_n\}$ is L^1 -Cauchy, and is both L^1 - and a.e. convergent to some function $f \in \mathfrak{L}^1$.

Proof. Suppose that $\{f_n\}$ is increasing, i. e. $f_n \leq f_{n+1}$ for all $n = 1, 2, \dots$. Let

$\alpha := \sup_k \int_X f_k$. Then $n \geq m$ implies

$$\|f_n - f_m\|_1 = \int (f_n - f_m) = \int f_n - \int f_m \leq \alpha - \int f_m,$$

which implies that $\{f_n\}$ is L^1 -Cauchy because $\alpha - \int f_m \rightarrow 0$ for $m \rightarrow \infty$. By 4.55. a subsequence converges a.e. (L^1 -convergence holds because L^1 -Cauchy implies L^1 -convergence by completeness), and because $\{f_n\}$ is increasing $\{f_n\}$ itself converges a.e. (Recall the **very** important property of monotonic real sequences, which is applied here pointwise a.e.: Suppose $\{a_n\}$ is a real sequence and $a_{n_k} \rightarrow a$ for $k \rightarrow \infty$. Given $\varepsilon > 0$ there exists K such that $k \geq K$ implies $a - a_{n_k} < \varepsilon$. But then for all $n \geq n_k$ it follows $a - a_{n_k} \leq a - a_n < \varepsilon$, thus $a_n \rightarrow a$.) Then by 4.58 the convergence is also in L^1 -seminorm. For the decreasing case replace f_n by $-f_n$. ■

Recall from 4.49 (b) that L^1 -convergence implies that limit and integral can be interchanged. **The slogan of the monotonic convergence theorem is:** *If the sequence of integrals of an increasing sequence of functions is bounded from above then the sequence is L^1 -Cauchy and thus converges to an integrable function a.e. and limit and integral can be interchanged.*

4.60. Corollary. *Let $\{f_n\}$ be a sequence of real valued functions in \mathfrak{L}^1 . Suppose $\exists 0 \leq g \in \mathfrak{L}^1$ such that $|f_n| \leq g$ for all n . Then $\sup(f_n), \inf(f_n) \in \mathfrak{L}^1$ and*

$$\sup\left(\int f_n\right) \leq \int \sup(f_n) \quad \text{and} \quad \int \inf(f_n) \leq \inf\left(\int f_n\right)$$

Proof. The sequence $g_n := \sup(f_1, \dots, f_n) \in \mathfrak{L}^1$ is pointwise bounded by g because $f_n \leq |f_n| \leq g$. Because $\int g$ exists and the integral is monotonic also $\{\int g_n\}$ is an increasing bounded sequence bounded from above by $\int g$. By 4.59. $\{g_n\}$ converges a.e. to $\lim g_n = \sup f_n \in \mathfrak{L}^1$. To get the estimate use $\int \sup(f_n) = \int \lim(g_n) = \lim \int(g_n) \geq \int g_n \geq \int f_n$ using that $\{g_n\}$ is increasing and $g_n \geq f_n$ for all n . Thus $\sup \int f_n \leq \int \sup(f_n)$. The proof for \inf is similar, using $h_n = \inf(f_1, \dots, f_n)$. ■

4.61. Definition. For a real valued sequence of functions $\{f_n\}$ define

$$\liminf(f_n) := \lim_{k \rightarrow \infty} \inf_{n \geq k} f_n,$$

which is a real-valued function if the limits exist. If $\{f_n\}$ converges pointwise then $\liminf(f_n) = \lim(f_n)$. If the limit or \liminf exist a.e. the resulting functions will still be denoted $\lim(f_n)$ or $\liminf(f_n)$ (define the functions 0 where limit or \liminf do not exist).

4.62. Fatou's lemma. Let $\{f_n\}$ be a sequence of nonnegative functions ($f_n \geq 0$) in \mathfrak{L}^1 . Suppose that $0 \leq \liminf(\|f_n\|_1) \in \mathbb{R}$ exists. Then $\liminf(f_n(x))$ exists a.e., $\liminf(f_n) \in \mathfrak{L}^1$ and

$$\int_X \liminf(f_n) d\mu \leq \liminf \int_X f_n d\mu = \liminf(\|f_n\|_1).$$

Proof. Apply the monotone convergence theorem 4.59. first for fixed k to the decreasing sequence $g_m := \inf(f_k, f_{k+1}, \dots, f_{k+m})$ Now $\{g_m\}$ converges to $\inf_{n \geq k} f_n$, and for $j = 1, 2, \dots, m$ we have by monotonicity of the integral:

$$\int \inf(f_k, f_{k+1}, \dots, f_{k+m}) \leq \int f_{k+j}.$$

Also, $0 \leq \int g_m$, so the assumption of the monotone convergence theorem is satisfied. Thus $\lim_{m \rightarrow \infty} g_m = \inf_{n \geq k} f_n \in \mathfrak{L}^1$ for all n , and using the previous inequality,

$$\begin{aligned} \int \inf_{n \geq k} f_n &= \int \lim_{m \rightarrow \infty} g_m = \lim_{m \rightarrow \infty} \int g_m = \lim_{m \rightarrow \infty} \int \inf(f_k, f_{k+1}, \dots, f_{k+m}) \leq \\ &\leq \lim_{m \rightarrow \infty} \inf_{0 \leq j \leq m} \left(\int f_{k+j} \right) = \inf_{n \geq k} \int f_n. \end{aligned}$$

Now $h_k := \inf_{n \geq k} f_n \leq f_k$ is an increasing sequence of integrable functions and $\int h_k \leq \int f_k$. For a subsequence of $\{k\}$ such that $\int f_k = \int |f_k|$ converges to $\liminf \|f_k\|_1 \in \mathbb{R}$ the assumption of the monotone convergence theorem is satisfied. Because the corresponding subsequence of $\{h_k\}$ also converges to $\liminf f_n$, we get $\liminf f_n \in \mathfrak{L}^1$ and thus $\liminf(f_n(x))$ exists in \mathbb{R} for almost all x . Finally by letting k in this subsequence go to ∞ we get from the monotone convergence theorem and the previous lemma:

$$\int \liminf f_n = \int \lim_{k \rightarrow \infty} h_k = \lim_{k \rightarrow \infty} \int h_k = \lim_{k \rightarrow \infty} \int \inf_{n \geq k} f_n \leq \liminf \int f_n$$

This proves Fatou's lemma. ■

Fatou's lemma applies in particular in the case when $\{f_n\}$ actually is pointwise convergent a.e. and $\{\|f_n\|_1\}$ is bounded proving that in this case $\lim f_n \in \mathfrak{L}^1$.

4.63. Dominated convergence theorem. Let $\{f_n\}$ with $f_n \in \mathfrak{L}^1$. Suppose that there exists $g \in \mathfrak{L}^1$ such that $|f_n| \leq g$ for all n . If $f_n \rightarrow f$ a.e. then $f \in \mathfrak{L}^1$ and $\{f_n\}$ is also L^1 -convergent to f so that in particular

$$\lim \int f_n = \int \lim f_n$$

Proof. Consider the sequence $\{g_k\}$ of nonnegative functions defined by $g_k := \sup_{m,n \geq k} |f_n - f_m|$. This is a decreasing sequence of real valued functions. Because $|f_n - f_m| \leq 2g$ two applications of 4.60 show that $g_k \in \mathcal{L}^1$. Because $g_k \geq 0$ and thus $\int g_k \geq 0$ the sequence $\{g_k\}$ converges a.e. and $\lim \int g_k = \int \lim g_k$. But $g_k \rightarrow 0$ a.e. because $f_n \rightarrow f$ a.e. and $|f_n - f_m| \leq |f_n - f| + |f_m - f| \rightarrow 0$ a.e. for $\min(n, m) \rightarrow \infty$ (in other words convergence at a point implies Cauchy at that point). Thus $\int g_k \rightarrow 0$ for $k \rightarrow \infty$, and for $n, m \geq k$ we have

$$\|f_n - f_m\|_1 = \int |f_n - f_m| \leq \int \sup_{n,m \geq k} |f_n - f_m| = \int g_k$$

and the right hand side converges to 0 for $k \rightarrow \infty$. Thus $\{f_n\}$ is L^1 -Cauchy and by 4.58, $f \in \mathcal{L}^1$ and $\{f_n\}$ is L^1 -convergent to f . ■

The slogan of the dominated convergence theorem is: *If the norms of a sequence of integrable functions are pointwise bounded by some real-valued integrable function and if an a.e. limit exists then the sequence is L^1 -Cauchy, so f is integrable and limit and integral can be changed.*

4.64. Corollary. *Suppose that f is μ -measurable. Then the following holds:*

- (i) $f \in \mathcal{L}^1(\mu) \iff |f| \in \mathcal{L}^1(\mu, \mathbb{R})$.
- (ii) *If $\exists g \in \mathcal{L}^1(\mu, \mathbb{R})$ such that $|f| \leq g$ then $f \in \mathcal{L}^1(\mu)$.*

Proof. (i) \implies has been shown in 4.47. (i) \longleftarrow follows from (ii) with $g = |f|$. Let $\{\varphi_n\}$ be a sequence of step maps converging a.e. to f . After possibly changing f, g and the φ_n on a set of measure 0 we can assume that φ_n converges to f and g is measurable. Define maps h_n by $h_n(x) = \varphi_n(x)$ if $|\varphi_n(x)| \leq 2g(x)$ and $h_n(x) = 0$ if $|\varphi_n(x)| > 2g(x)$. The set $S_n := \{x \in X : 2g(x) - |\varphi_n(x)| \geq 0\}$ is measurable and so $h_n \in \mathcal{L}^1(\mu)$ for each n because h_n is a step map. (We multiply a step map by a characteristic function of a measurable set, which we saw is integrable in 4.50 (v).) Because $|f| \leq g$ also $\{h_n\}$ converges pointwise to f and $|h_n| \leq 2g$. Thus $f \in \mathcal{L}^1(\mu)$ by the dominated convergence theorem. ■

4.65. Corollary. *Let $f_n \in \mathcal{L}^1(\mu)$ and $f_n \rightarrow f$ a.e. If $\exists C \in \mathbb{R}$ such that $\|f_n\|_1 \leq C$ for all n then $f \in \mathcal{L}^1(\mu)$ and $\|f\|_1 \leq C$.*

Proof. Because f_n is μ -measurable for each n it follows from 4.38 that f is μ -measurable. Also $|f| = \lim |f_n|$ a.e. by the triangle inequality and the special case of Fatou's lemma proves $|f| \in \mathcal{L}^1$. By the previous corollary it follows that $f \in \mathcal{L}^1(\mu)$. ■

4.66. Example. Consider a sequence of functions $f_n : \mathbb{R} \rightarrow \mathbb{R}$ such that $f_n(x) = 1$ for $x \in [n, n+1]$ and $f_n(x) = 0$ otherwise. Then $f_n \rightarrow 0$ pointwise

everywhere because for each $x \in \mathbb{R}$ we have $f_n(x) = 0$ for $n > x$ but using a usual measure we can have $\|f_n\|_1 = 1$ for all n thus f_n is not L^1 -convergent to 0. Note that $\{f_n\}$ is also not L^1 -Cauchy. So notice that no conclusion about L^1 -convergence is made in 4.65.

4.67. Corollary. (i) If $f \in \mathfrak{L}^1(\mu, E)$ and g is a bounded and measurable real or complex-valued function then $fg \in \mathfrak{L}^1$.

(ii) More generally, if $\varphi : E \times F \rightarrow G$ is a continuous bilinear map of Banach spaces, and $f \in \mathfrak{L}^1(\mu)$ and g is a bounded μ -measurable function of X into F then $\varphi \circ (f, g) : X \rightarrow G$ is in $\mathfrak{L}^1(\mu)$

Proof. (i) Choose an approximating sequence of step maps $\{\varphi_n\}$ for f . Using 4.17 find a sequence of simple maps $\{\psi_n\}$ converging pointwise to g . Then $\{\varphi_n\psi_n\}$ is a sequence of step maps converging pointwise a.e. to fg (This is because $(\varphi_n\psi_n)(x) = 0$ if $\varphi_n(x) = 0$ so outside of a set of finite measure. Also for each n , $\varphi_n\psi_n$ takes only finitely many values and $\{x : \varphi_n(x)\psi_n(x) = b\}$ is the union of all sets $\varphi_n^{-1}(b_1) \cap \psi_n^{-1}(b_2)$ over all b_1, b_2 such that $b = b_1b_2$, but there are only finitely many possibilities. So the sets $(\varphi_n\psi_n)^{-1}(\{b\})$ are measurable). By changing f and g on sets of measure 0 we can assume that the convergence is pointwise everywhere (give f and the approximating sequence the values 0 on that set of measure 0, which thus is measurable, where we do not know convergence of the approximating sequences.) If $|g(x)| \leq C$ for all x then $|fg| \leq C|f|$. Because $|f| \in \mathfrak{L}^1$ it follows from 4.64 (ii) that $fg \in \mathfrak{L}^1$. (Alternatively we can assume that $|\varphi_n| \leq 2|f|$ and then apply the dominated convergence theorem.) The above proof is easily modified to prove (ii). The details will be an exercise. ■

4.68. Corollary. If $\{f_n\}$ is a sequence of μ -integrable maps such that $\sum_{n=1}^{\infty} \int_X |f_n| d\mu$ converges then $\sum_{n=1}^{\infty} f_n(x)$ converges pointwise a.e. to the μ -integrable function f , and

$$\int_X f d\mu = \sum_{n=1}^{\infty} \int_X f_n d\mu$$

Proof. Apply the dominated convergence theorem to the sequence of partial sums and use $g(x) := \lim_{n \rightarrow \infty} \sum_{k=1}^n |f_k(x)|$ for the assumption. The fact that integral and limit can be changed follows from the L^1 -convergence in the dominated convergence theorem. Note that g is integrable by the monotone convergence theorem: In fact $g_n := \sum_{k=1}^n |f_k|$ is an increasing sequence of integrable functions, and the sequence of integrals $\int \sum_{k=1}^n |f_k| = \sum_{k=1}^n \int |f_k|$ is bounded by $\sum_{k=1}^{\infty} \int |f_k|$. ■

4.69. Example. A standard example of the situation above is given by $\{A_n\}$ a sequence of disjoint measurable sets with union X . If $f_n : X \rightarrow E$ is a sequence of functions, integrable over A_n and 0 outside of A_n , then $f = \sum f_n$ exists a.e. and is integrable, under the assumption that the series $\sum_{n=1}^{\infty} \int_X |f_n| = \sum_{n=1}^{\infty} \int_{A_n} |f_n|$ converges. Alternatively given $f : X \rightarrow E$ we can define $f_n := f\chi_{A_n}$ and often prove integrability of f by showing convergence of the sequence of integrals.

4.70. Corollary. For each integrable f and $\varepsilon > 0$ there exists a set $A \subset X$ of finite measure such that

$$\left| \int_X f d\mu - \int_A f d\mu \right| < \varepsilon$$

Proof. Because f is a.e. limit of a sequence of step maps we can change f on a set of measure 0 such that f vanishes outside a countable union of sets A_n of finite measure. Consider the increasing sequence $B_n := A_1 \cup \dots \cup A_n$. For the estimate of the integral we can assume that $X = \cup_{n=1}^{\infty} A_n$ because $f = 0$ outside of this union. Now we can calculate:

$$\left| \int_X f d\mu - \int_{B_n} f d\mu \right| = \left| \int_{X-B_n} f d\mu \right| \leq \int_{X-B_n} |f| d\mu \leq \int_X |f|(1 - \chi_{B_n}) d\mu.$$

Now $|f|(1 - \chi_{B_n})$ is an decreasing sequence of integrable functions, bounded below by 0 and converging pointwise to the function 0 for $n \rightarrow \infty$. Thus by the monotone convergence theorem the sequence of integrals converges to 0 for $n \rightarrow \infty$. So for n large enough and $A = B_n$ we can assume the right side of the inequality to be $< \varepsilon$. ■

4.71. Theorem. Assume f is integrable and S is a closed subset of E . Suppose that for all measurable sets $A \subset X$ of finite measure $\neq 0$ we have

$$\frac{1}{\mu(A)} \int_A f d\mu \in S.$$

If $0 \in S$ or X is σ -finite then $f(x) \in S$ for almost all x .

Proof. If X is not σ -finite by integrability we can assume that $f = 0$ outside of a σ -finite set. But in this case by assumption $0 \in S$ and we only have to consider f on the σ -finite set where f is possibly non-zero. If we can prove the result for sets with $\mu(X) < \infty$ then it will follow because countable unions of sets of measure 0 have measure 0. Let $D \subset E$ be a countable set with $f(X) \subset \overline{D}$, see 4.36. (Note that f is μ -measurable because it is a.e. pointwise limit of a sequence of step maps, ignore a set of measure 0 in X where this does not hold.) Consider

$v \in D \cap (E - S)$ and r a positive rational number such that $B_r(v) \subset E - S$. Let $A = A(v, r)$ be the set of $x \in X$ such that $f(x) \in B_r(v)$. Then A has measure 0. Indeed, if $\mu(A) > 0$ then $0 < \mu(A) < \infty$ and

$$\left| \frac{1}{\mu(A)} \int_A f d\mu - v \right| = \left| \frac{1}{\mu(A)} \int f d\mu - \frac{1}{\mu(A)} \int v d\mu \right| \leq \frac{1}{\mu(A)} \int_A |f - v| d\mu < r,$$

which contradicts $\frac{1}{\mu(A)} \int f d\mu \in S$. Hence $\mu(A) = 0$. But the set of all x such that $f(x) \notin S$ is in the union of the sets $A(v, r)$ over all $v \in D$ and r a positive rational number. This follows from $\overline{D} - S \supset f(X) - S$. ■

4.72. Corollaries. (i) *If $f \in \mathfrak{L}^1(\mu)$ and $\int_A f d\mu = 0$ for every measurable set A of finite measure then $f = 0$ a.e.*

(ii) *If $f \in \mathfrak{L}^1(\mu, E)$ and $g \in \text{St}(\mu, \mathbb{R})$ then $fg \in \mathfrak{L}^1(\mu, E)$ and if $\int_X fgd\mu = 0$ for all $g \in \text{St}(\mu)$ then $f(x) = 0$ a.e.*

(iii) *If $f \in \mathfrak{L}^1(\mu)$ and $b \geq 0$, and if $|\int_A f d\mu| \leq b\mu(A)$ for all sets A of finite measure then $|f(x)| \leq b$ a.e.*

(iv) *If E is a Hilbert space and $f \in \mathfrak{L}^1(\mu, E)$ such that $\int_X \langle f, g \rangle d\mu = 0$ for all $g \in \text{St}(\mu, E)$ then $f = 0$ a.e.*

(v) *If E is a Hilbert space and $f \in \mathfrak{L}^1(\mu, E)$. For $f \in E$ and e a unit vector let let $f_e = \langle f, e \rangle$ be the Fourier coefficient. Let $b \geq 0$. Suppose that for each unit vector e and each set of finite measure A we have $|\int_A f_e d\mu| \leq b\mu(A)$. Then $|f(x)| \leq b$ a.e.*

Proof. (i): Take $S = \{0\}$ in 4.71. (ii): We can apply (i) because for all characteristic functions $\chi_A \in \text{St}(\mu, \mathbb{R})$. (iii): Let $S := \{v \in E : |v| \leq b\} = \overline{B}_b(0) \subset E$ the closed ball in E of radius b . and apply 4.71., by assumption $|\frac{1}{\mu(A)} \int_A f d\mu| \leq b$ and thus $\frac{1}{\mu(A)} \int_A f d\mu \in S$ for all A of finite measure. (iv) Because we can assume $f(X) \subset \overline{D}$ for $D \subset E$ countable we can replace E by the subspace of E generated by \overline{D} . This is a separable Hilbert space. Let e be a unit vector in a Hilbert basis and A a measurable set of finite measure. The step map $e\chi_A$ is bounded and measurable. If $f_e = \langle f, e \rangle$ is the Fourier coefficient of f with respect to e then $0 = \int_X \langle f, e\chi_A \rangle d\mu = \int_A f_e d\mu$. Thus $f_e = 0$ a.e. for all e , and thus because $f = \sum_{n=1}^{\infty} f_n e_n$ for a Hilbert basis $\{e_n\}$ and Fourier coefficients f_n , which implies $f = 0$ a.e. (take the union of the sets for which $f_n \neq 0$, which is a set of measure 0). (v): As in 4.71 assume that E is separable and $\mu(X) < \infty$. Let $v \in E$ with $|v| > b$ and $r > 0$ such that $B_r(v) \cap \overline{B}_b(0) = \emptyset$ (we can choose $r := \frac{|v|-b}{2}$). Let $A := \{x \in X : f(x) \in B_r(v)\}$ and $e := \frac{v}{|v|}$. Let $x \in A$. Then $f_e(x) - |v| = \langle f(x), \frac{v}{|v|} \rangle - |v| = \langle f(x), \frac{v}{|v|} \rangle - \langle v, \frac{v}{|v|} \rangle = \langle f(x) - v, \frac{v}{|v|} \rangle$ thus the

Schwarz inequality implies $|f_e(x) - |v|| \leq |f(x) - v| < r$ thus $f_e(x) \in B_r(|v|) \subset \mathbb{R}$. Using our assumption we know from the proof of 4.71 in the case (iii) that $\mu(A) = 0$. Because the set where $|f(x)| > b$ is countable union of sets of the form A the result follows. ■

Recall that a collection \mathcal{A} of subsets of X is an *algebra of sets* if $\emptyset \in \mathcal{A}$ and $A, B \in \mathcal{A} \implies A \cup B, A \cap B, A - B \in \mathcal{A}$.

4.73. Definition. A *measure* on \mathcal{A} is a function $\mu : \mathcal{A} \rightarrow [0, \infty]$ with $\mu(\emptyset) = 0$, and if $\{A_k\}$ is a sequence of disjoint sets in \mathcal{A} such that $A = \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$ then $\mu(A) = \sum_{k=1}^{\infty} \mu(A_k)$ (μ is countably additive. Note that all our measures are positive.)

Let \mathcal{I} be the collection of finite disjoint unions of bounded intervals of the real numbers. Then \mathcal{I} is an algebra of sets. We want to prove that the usual length function l defines a measure on \mathcal{I} . Certainly l is additive with respect to finite unions of sets.

4.74. Theorem. Let $\{f_n\}$ be a sequence of non-negative functions on a closed bounded interval I , monotonically decreasing to 0. Assume that each f_n is a step function with respect to some interval. Then the sequence of usual integrals of step maps

$$\int_I f_n(x) dx$$

decreases to 0.

Proof. Note that the claim is trivial in the case that the convergence is uniform. Our proof will precisely use the fact that the set of points where the convergence is not uniform can be covered by a set of arbitrarily small length. Consider for each n the endpoints of the intervals on which f_n is constant. This is a finite set, and thus the union of these sets over all n is countable. Cover the set by a sequence of open intervals $\{J_k\}$ such that $\sum_{k=1}^{\infty} l(J_k) < \varepsilon$. (Choose n such that $\frac{1}{2^{n-1}} < \varepsilon$ and intervals J_k of lengths $\frac{1}{2^{n+k}}$ then $\frac{1}{2^n} \sum_{k=0}^{\infty} \frac{1}{2^k} = \frac{1}{2^{n-1}}$.) Let $U := \bigcup J_k$. For each $x \notin U$ all functions f_n are constant in a neighborhood of x and so there exists $n_x \in \mathbb{N}$ and some open neighborhood V_x of x such that $f_m(t) < \varepsilon$ for all $t \in V_x$ and $m \geq n_x$. The collection $\{J_k, V_x\}_{k=1,2,\dots; x \in I-U}$ is an open covering of the compact set I , so we find a finite subcovering $\{J_{k_1}, \dots, J_{k_r}, V_{x_1}, \dots, V_{x_s}\}$. Let $N := \max(n_{x_1}, n_{x_2}, \dots, n_{x_s})$. If $n \geq N$ then $f_n(t) < \varepsilon$ for all $t \in V_{x_1} \cup \dots \cup V_{x_s}$. Thus if C is a constant with $f_1 \leq C$ and thus $f_n \leq C$ for all n (recall that

step functions are bounded) then

$$\int_I f_n(x) dx \leq \int_{J_{k_1} \cup \dots \cup J_{k_r}} f_n(x) dx + \int_{V_{x_1} \cup \dots \cup V_{x_s}} f_n(x) dx \leq C\varepsilon + l(I)\varepsilon. \blacksquare$$

4.75. Corollary. *The length function $l : \mathcal{I} \rightarrow [0, \infty) \subset [0, \infty]$ is a measure.*

Proof. Let $\{A_n\}$ be a collection of disjoint sets in \mathcal{I} such that $A := \cup_{n=1}^{\infty} A_n \in \mathcal{I}$. Let $B_n := A_1 \cup \dots \cup A_n$. Then $\{\chi_A - \chi_{B_n}\} = \{\chi_{A-B_n}\}$ is a decreasing sequence of step maps defined on some closed bounded interval $I \supset A$, converging pointwise to 0. Thus the sequence of integrals decreases to 0: $\int (\chi_A - \chi_{B_n}) \rightarrow 0$ or $\sum_{k=1}^n \mu(A_k) = \mu(B_n) = \int \chi_{B_n} \rightarrow \int \chi_A = \mu(A)$. \blacksquare

The existence of the Lebesgue integral on \mathbb{R} is now a consequence of the following *abstract* theorem and our previous general constructions. Recall that the σ -algebra generated by a collection of subsets of X is the smallest σ -algebra containing this collection.

4.76. Theorem (Hahn). *Let μ be a measure on an algebra \mathcal{A} of subsets of X , and assume that X is denumerable union of sets of \mathcal{A} . Then μ can be extended to a measure on the σ -algebra \mathcal{M} generated by \mathcal{A} such that for all $Y \in \mathcal{M}$:*

$$\mu(Y) = \inf \sum_{n=1}^{\infty} \mu(A_n),$$

*with the infimum taken over all sequences $\{A_n\}$ in \mathcal{A} such that $\cup A_n \supset Y$. If X is countable union of sets in \mathcal{A} of **finite** measure then the extension is unique.*

We will not prove Hahn's theorem. See Lang Theorem VI, 7.1. for the proof.

Let \mathcal{A} be an algebra of subsets of X with a measure μ such that all sets in \mathcal{A} have finite measure. Then X is called *σ -finite with respect to \mathcal{A}* if X is countable union of elements of \mathcal{A} . Define a function $\mu^* : \mathcal{N} \rightarrow [0, \infty]$ on the σ -algebra of all subsets of X by $\mu^*(Y) := \inf \sum_{n=1}^{\infty} \mu(A_n)$, with the infimum taken over all sequences $\{A_n\}$ of elements of \mathcal{A} whose union contains Y (μ^* is what is called an outer measure. It satisfies the usual properties of a measure except σ -additivity is replaced by $\mu^*(\cup B_n) \leq \sum_{n=1}^{\infty} \mu^*(B_n)$ for any sequence of elements in \mathcal{N} , see Lang Lemma VI, 7.2). We will need the following consequence of Hahn's theorem:

4.77. Corollary. *Let \mathcal{A} be a measured algebra of sets of finite measure such that X is σ -finite with respect to \mathcal{A} . Then a subset $Z \subset X$ has μ^* -measure 0 if and only if for each $\varepsilon > 0$ there exists a sequence $\{A_n\}$ in \mathcal{A} with $\cup A_n \supset Z$*

such that $\sum_{n=1}^{\infty} \mu(A_n) < \varepsilon$. If (X, \mathcal{M}, μ) is measured and \mathcal{A} as above generates \mathcal{M} , then the same is true for the μ -measure and measurable sets Z .

Proof. The assertion for the outer measure is clear by the very definition of the outer measure. In the case that (X, \mathcal{M}, μ) is measured the assertion follows for measurable Z by Hahn's theorem because the measure on \mathcal{A} has a unique extension to \mathcal{M} to the measure by the outer measure defined from the measure on \mathcal{A} . ■

For the existence of the Lebesgue measure on \mathbb{R}^p we need to develop a theory of product measures and integration on product spaces.

Let X, Y be sets with sets of algebras \mathcal{A}, \mathcal{B} . Let $\mathcal{A} \times \mathcal{B}$ be the collection of **finite disjoint unions** of rectangles with respect to \mathcal{A} and \mathcal{B} , i. e. sets $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

4.78. Lemma. $\mathcal{A} \times \mathcal{B}$ is an algebra of sets in $X \times Y$.

Proof. $\emptyset \in \mathcal{A} \times \mathcal{B}$ is clear. From the set equalities (draw Venn diagrams)

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$$

and

$$(A_1 \times B_1) - (A_2 \times B_2) = ((A_1 - A_2) \times B_1) \cup ((A_1 \cap A_2) \times (B_1 - B_2))$$

it follows that $P, Q \in \mathcal{A} \times \mathcal{B}$ then $P \cap Q$ and $P - Q$ are in $\mathcal{A} \times \mathcal{B}$. Then also $P \cup Q = (P - Q) \cup Q \in \mathcal{A} \times \mathcal{B}$ because $(P - Q) \cap Q = \emptyset$. Because disjoint unions of disjoint unions of rectangles are disjoint unions of rectangles the result follows. ■

For each algebra of sets \mathcal{A} on a set X let let \mathcal{A}^σ denote the σ -algebra generated by \mathcal{A} .

4.79. Lemma. For \mathcal{A}, \mathcal{B} algebras of sets in X, Y we have

$$(\mathcal{A}^\sigma \times \mathcal{B}^\sigma)^\sigma = (\mathcal{A} \times \mathcal{B})^\sigma =: \mathcal{A} \otimes \mathcal{B},$$

so equivalently

$$\mathcal{A}^\sigma \otimes \mathcal{B}^\sigma = (\mathcal{A} \times \mathcal{B})^\sigma.$$

Proof. $\mathcal{A} \times \mathcal{B} \subset \mathcal{A}^\sigma \times \mathcal{B}^\sigma \subset \mathcal{A}^\sigma \otimes \mathcal{B}^\sigma$ is clear. For each $B \in \mathcal{B}$ consider the σ -algebra in $X \times B$ generated by all sets $A \times B$ for $A \in \mathcal{A}$. Using $X \times B \subset X \times Y$ this collection is contained in $(\mathcal{A} \times \mathcal{B})^\sigma$, or $(\mathcal{A} \times \mathcal{B})^\sigma \supset \mathcal{A}^\sigma \times \{B\}$ for all $B \in \mathcal{B}$.

But then for each $A \in \mathcal{A}^\sigma$ it follows that $\{A\} \times \mathcal{B}^\sigma \subset (\mathcal{A} \times \mathcal{B})^\sigma$ and thus $\mathcal{A}^\sigma \times \mathcal{B}^\sigma \subset (\mathcal{A} \times \mathcal{B})^\sigma$. This implies $\mathcal{A}^\sigma \otimes \mathcal{B}^\sigma \subset (\mathcal{A} \times \mathcal{B})^\sigma$. ■

4.80. Lemma. *Let \mathcal{M}, \mathcal{N} be σ -algebras on X, Y .*

(i) *Let $Q \in \mathcal{M} \otimes \mathcal{N}$ and for $x \in X$ let $Q_x := \{y \in Y : (x, y) \in Q\}$. Then $Q_x \in \mathcal{N}$.*

(ii) *Let $f : X \times Y \rightarrow Z$ be a $\mathcal{M} \otimes \mathcal{N}$ -measurable map into a topological space Z . Then for each $x \in X$ the map $f_x : Y \rightarrow Z$ defined by $f_x(y) = f(x, y)$ is \mathcal{N} -measurable.*

Proof. (i): Let $\mathcal{S} := \{Q \in \mathcal{M} \otimes \mathcal{N} : Q_x \in \mathcal{N} \text{ for all } x \in X\} \supset \{A \times B : A \in \mathcal{M}, B \in \mathcal{N}\}$ because $(A \times B)_x = B$ if $x \in A$ and $(A \times B)_x = \emptyset$ if $x \notin A$. It suffices to show that \mathcal{S} is a σ -algebra then $\mathcal{S} = \mathcal{M} \otimes \mathcal{N}$ by minimality proves (i). First, $X \times Y \in \mathcal{S}$ because $(X \times Y)_x = Y \in \mathcal{N}$ for each $x \in X$. If $Q \in \mathcal{S}$ then $\mathfrak{C}Q \in \mathcal{S}$ because $(\mathfrak{C}Q)_x = \mathfrak{C}(Q_x)$ (because: $y \in (\mathfrak{C}Q)_x \iff (x, y) \in \mathfrak{C}Q \iff (x, y) \notin Q \iff y \notin Q_x \iff y \in \mathfrak{C}(Q_x)$). Similarly, if $Q, P \in \mathcal{S}$ then $(P \cap Q)_x = P_x \cap Q_x \in \mathcal{N} \implies P \cap Q \in \mathcal{S}$. If $\{Q_n\}$ is a sequence in \mathcal{S} then $(\cup Q_n)_x = \cup(Q_n)_x \in \mathcal{N}$ and thus $\cup Q_n \in \mathcal{S}$. This shows that \mathcal{S} is a σ -algebra. (ii): If $V \subset Z$ is open then $(f^{-1}(V))_x = \{y \in Y : (x, y) \in f^{-1}(V)\} = f_x^{-1}(V) \in \mathcal{N}$ because $f_x^{-1}(V) \in \mathcal{M} \otimes \mathcal{N}$. ■

For a subalgebra of sets \mathcal{A} of a σ -algebra \mathcal{M} with all sets in \mathcal{A} of finite measure we say that f is a *step map with respect to \mathcal{A}* if $f = 0$ outside some $A \in \mathcal{A}$ and there is a partition $\{A_1, \dots, A_r\}$ of A consisting of elements of \mathcal{A} such that f is step with respect to this partition.

In the following assume that (X, \mathcal{M}, μ) and (Y, \mathcal{N}, ν) are σ -finite, and \mathcal{A} respectively \mathcal{B} are algebras of sets of finite measure contained in \mathcal{M} respectively \mathcal{N} . We want to define *iterated* integrals.

If f is step with respect to $\mathcal{A} \times \mathcal{B}$ then for each $x \in X$ the map f_x is step on Y with respect to \mathcal{B} : If $f = v\chi_{A \times B}$ for $v \in E$ and $A \in \mathcal{A}, B \in \mathcal{B}$ then $f_x = v(\chi_{A \times B})_x = v\chi_A(x)\chi_B$ and $f_x(y) = v\chi_A(x)\chi_B(y)$, so the result follows by linearity. Thus for $x \in X$ we can first consider $\int_Y f_x d\nu$, which for $f = v\chi_{A \times B}$ becomes $v\chi_A(x)\mu(B)$. If f is step with respect to $\mathcal{A} \times \mathcal{B}$ then $x \mapsto \int_Y f_x d\nu$ is step with respect to \mathcal{A} and can be integrated with the notation being given by one of the following:

$$\int_X \left[\int_Y f_x d\nu \right] d\mu(x), \int_X d\mu(x) \int_Y f_x d\nu, \int_X \int_Y f(x, y) d\nu(y) d\mu(x), \int_X \int_Y f d\nu d\mu$$

It is clear that on step maps the order of integration can be reversed because from

$$\int_X \int_Y \chi_{A \times B} d\nu d\mu = \mu(A)\nu(B) = \int_Y \int_X \chi_{A \times B} d\mu d\nu$$

we can deduce this by linearity (each single integral is linear and composition of linear maps is linear).

4.81. Theorem. *Given (X, \mathcal{M}, μ) and (Y, \mathcal{N}, ν) , both σ -finite measured spaces. Then there exists a unique measure $\mu \otimes \nu$ on $\mathcal{M} \otimes \mathcal{N}$ such that for all sets A, B of finite measure in \mathcal{M} and \mathcal{N} we have:*

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B).$$

Proof. Hahn's theorem applies if we can show that $\mu \times \nu : \mathcal{A} \times \mathcal{B} \rightarrow [0, \infty)$ defined by $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$ is a measure, where \mathcal{A} respectively \mathcal{B} are the subalgebras of sets of finite measure of \mathcal{M} respectively \mathcal{N} . (σ -finiteness then implies that $\mathcal{M} \otimes \mathcal{N} = (\mathcal{A} \times \mathcal{B})^\sigma$.) The interesting point is countable additivity. Let $\{Q_n\}$ be increasing in $\mathcal{A} \times \mathcal{B}$ with union $Q \in \mathcal{A} \times \mathcal{B}$. Let $f_n := \chi_{Q_n}$. Then $\{f_n\}$ is increasing to $f = \chi_Q$. Also for each $x \in X$, $\{(f_n)_x\}$ is increasing to f_x . By monotone convergence with respect to ν we have for each x that the sequence of real numbers $\int_Y (f_n)_x d\nu$ is increasing to $\int_Y f_x d\nu$ (you changed integral and limit!). Then by applying monotone convergence with respect to μ we have that the sequence of real numbers $\int_X \int_Y f_n d\nu d\mu$ converges to $\int_X \int_Y f d\nu d\mu$. Reinterpreting this in terms of measures proves the claim. ■

The existence of the Lebesgue-measure and integral on \mathbb{R}^p follows from the previous theorem because \mathbb{R} is σ -finite with respect to the algebra of disjoint unions of bounded intervals. We also know that the measure of rectangles is the product of the lengths of its sides. The σ -algebra on the product space is generated by rectangles and thus is the Borel algebra of \mathbb{R}^p (because each open set is a countable union of *iterated* rectangles).

In the following we want to show that Lebesgue integrals on \mathbb{R}^p can be calculated from iterated integrals as known in Riemann integration. This is Fubini's theorem. We assume throughout that X respectively Y are σ -finite with respect to the subalgebras \mathcal{A} respectively \mathcal{B} of finite measure subsets of X respectively Y .

4.82. Lemma. *If $Z \subset X \times Y$ has $(\mu \otimes \nu)(Z) = 0$ then $\nu(Z_x) = 0$ for μ -almost all x .*

Proof. For $n \in \mathbb{N}$ let $S_n := \{x \in X : \nu(Z_x) \geq \frac{1}{n}\}$, and $S = \cup S_n$ (these are the bad guys). We will show that S is contained in a set of μ -measure 0. Let $\varepsilon > 0$. By the Corollary of Hahn's theorem 4.77 we can find a sequence of rectangles $\{R_k\}$ with $\cup R_k \supset Z$ such that $\sum_{k=1}^{\infty} (\mu \times \nu)(R_k) < \frac{\varepsilon}{n2^n}$. Then $Z_x \subset \cup_{k=1}^{\infty} R_{k,x}$. Consider $T_n := \{x \in X : \frac{1}{n} \leq \sum_{k=1}^{\infty} \nu(R_{k,x})\}$. Then $S_n \subset T_n$ and T_n is measurable. [We can write T_n as a countable union of the sets of x such that $\sum_{k=1}^N \nu(R_{k,x}) \geq \frac{1}{n} - \frac{1}{N}$. But then $\nu(R_{k,x}) = \int_Y \chi_{R_{k,x}}$ and we have seen before that $x \mapsto \int_Y f_x d\nu$ is step with respect to \mathcal{A} if f is step with respect to $\mathcal{A} \times \mathcal{B}$. This proves the measurability of T_n .] By monotone convergence we also know that $x \mapsto \sum_{k=1}^{\infty} \nu(R_{k,x})$ is integrable and

$$\begin{aligned} \frac{1}{n} \mu(T_n) &= \int_{T_n} \frac{1}{n} d\mu \leq \int_{T_n} \sum_{k=1}^{\infty} \nu(R_{k,x}) d\mu = \sum_{k=1}^{\infty} \int_{T_n} \nu(R_{k,x}) d\mu \\ &\leq \sum_{k=1}^{\infty} \int_X \nu(R_{k,x}) d\mu = \sum_{k=1}^{\infty} (\mu \times \nu)(R_k) < \frac{\varepsilon}{n2^n}. \end{aligned}$$

Thus $\mu(T_n) < \frac{\varepsilon}{2^n}$. Thus S is contained in the set $\cup T_n$ of measure $< \varepsilon$. Because this holds for each $\varepsilon > 0$ we get $\mu(S) = 0$. But if $\nu(Z_x) > 0$ then $x \in S_n$ for some n and so $x \in S$. Thus $\nu(Z_x) = 0$ outside of a set of μ -measure 0. ■

4.83. Example. Suppose $f \in \mathcal{L}^1(\mu \otimes \nu, E)$ and $g = f$ outside of a set Z with $(\mu \otimes \nu)(Z) = 0$. Then for each $x \in X$ we have $f_x = g_x$ except for $y \in Y$ such that $(x, y) \in Z \iff y \in Z_x$. By the lemma there exists $S \subset X$ with $\mu(S) = 0$ such that $\nu(Z_x) = 0$ if $x \notin S$. Thus for each $x \notin S$ we conclude that $f_x = g_x$ holds ν -a.e. Thus $f_x \in \mathcal{L}^1(\nu) \iff g_x \in \mathcal{L}^1(\nu)$, and so the integrals with respect to ν are equal.

For the proof of Fubini's theorem we need one more technical lemma. For the proofs we refer to Lang, Theorem 6.3. and Corollary 6.4.

4.84. Lemma. *Let \mathcal{A} be a generating subalgebra of a σ -algebra \mathcal{M} of a set X such that X is σ -finite with respect to \mathcal{A} (in particular the elements of \mathcal{A} have finite measure).*

(i) *Then the space $\text{St}(\mathcal{A}) = \text{St}(\mathcal{A}, E)$ of step mappings with respect to \mathcal{A} with values in E (i. e. the step map vanishes outside some element of \mathcal{A} and the partition is with elements from \mathcal{A}) is dense in $\mathcal{L}^1(\mu, E)$.*

(ii) *If $f \in \mathcal{L}^1(\mu)$ and $\int_A f d\mu = 0$ for all $A \in \mathcal{A}$ then $f = 0$ a.e.*

4.85. Fubini's Theorem, Part I *Let $f \in \mathcal{L}^1(\mu \otimes \nu, E)$. Then for μ -almost all x we have $f_x \in \mathcal{L}^1(\nu, E)$. Moreover, the map $X \ni x \mapsto \int_Y f_x d\nu \in E$ for*

μ -almost all x and arbitrarily otherwise, is in $\mathfrak{L}^1(\mu, E)$, and

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_X \int_Y f_x d\nu d\mu(x).$$

In fact for each Banach space E there is a norm preserving isomorphism of Banach spaces:

$$L^1(\mu \otimes \nu, E) \rightarrow L^1(\mu, L^1(\nu, E)),$$

defined by $L^1(\mu \otimes \nu, E) \ni f \mapsto (X \ni x \mapsto f_x \in L^1(\nu, E))$, where we have identified functions with their images in the Banach space completions.

Proof. By the lemma we can find a sequence $\{\varphi_n\}$ of step mappings with respect to $\mathcal{A} \times \mathcal{B}$, which converges to f in the L^1 -seminorm and a.e. on $X \times Y$. Let $Z \subset X \times Y$ be such that $\{\varphi_n\}$ converges pointwise to f outside of Z and $(\mu \otimes \nu)(Z) = 0$. We want to show first that f_x is integrable for fixed x a.e. For this we construct an explicit L^1 -Cauchy sequence converging to f_x a.e. The idea is to use 4.82 to find $S \subset X$ with $\mu(S) = 0$ such that $\nu(Z_x) = 0$ for $x \notin S$. Now if $x \notin S$ then $\{\varphi_{n,x}\}$ converges pointwise to f_x on $Y - Z_x$. For each n the map $\Phi_n : x \mapsto \varphi_{n,x}$ is a step map from X to $\text{St}(\mathcal{B})$. Indeed, $\varphi_{n,x}$ is a step map with respect to \mathcal{B} , and for $v \in E$ from $(v\chi_{\mathcal{A} \times \mathcal{B}})_x = v\chi_{\mathcal{A}}(x)\chi_{\mathcal{B}} = (v\chi_{\mathcal{B}})\chi_{\mathcal{A}}(x)$ we see by using linearity that Φ_n is a step map with respect to \mathcal{A} . Thus we see that $\Phi_n \in \text{St}(\mathcal{A}, \text{St}(\mathcal{B}))$. Note that the vector space $\text{St}(\mathcal{B})$ has the L^1 -seminorm, denoted $|\cdot|_1$. *Claim:* $\{\Phi_n\}$ is an L^1 -Cauchy sequence. This is immediate from

$$\begin{aligned} \|\Phi_n - \Phi_m\|_1 &= \int_X |\Phi_n - \Phi_m|_1 d\mu = \int_X \int_Y |\varphi_n(x, y) - \varphi_m(x, y)| d\nu(y) d\mu(x) = \\ &= \|\varphi_n - \varphi_m\|_1 \rightarrow 0 \text{ for } \min(n, m) \rightarrow \infty. \end{aligned}$$

Here $\|\cdot\|_1$ on the left hand side of the equation is the L^1 -norm on $\mathfrak{L}^1(\mu, \text{St}(\mathcal{B}))$, $|\cdot|_1$ is the restriction of the $\|\cdot\|_1$ -norm from $\mathfrak{L}^1(\nu, E)$ to $\text{St}(\mathcal{B})$, and $\|\cdot\|_1$ on the right hand side of the equations is the L^1 -norm on $\mathfrak{L}(\mu \otimes \nu, E)$.

It follows from the fundamental lemma of integration that we can assume, possibly going over to a subsequence, that for fixed x , outside of a set $T \subset X$ with $\mu(T) = 0$, we have that $\{\Phi_n(x) = \varphi_{n,x}\}$ is a pointwise Cauchy sequence in $\text{St}(\mathcal{B}) \subset \mathfrak{L}^1(\nu)$. [Note that the fundamental lemma actually has been stated only for Banach space valued functions. It is easy to check that the proof also works in the seminormed case. Alternatively one can compose with the natural map $\mathfrak{L}^1(\nu) \rightarrow L^1(\nu)$ and apply the fundamental lemma for the Banach space $L^1(\nu)$.] Because the norm on $\text{St}(\mathcal{B})$ is also the L^1 -norm it follows that for each fixed $x \notin T$ the sequence $\{\Phi_n(x) = \varphi_{n,x}\}$ is L^1 -Cauchy with respect to ν . If $x \notin S \cup T$ then $\{\varphi_{n,x}(y) \in E\}$ converges to $f_x(y)$ for ν -almost all $y \in Y$. So at

this point we have found for fixed $x \notin S \cup T$ the $L^1(\nu)$ -Cauchy sequence of step maps $\{\varphi_{n,x}\}$, which converges pointwise ν -a.e. to f_x . By the definition of $\mathfrak{L}^1(\nu)$ it follows $f_x \in \mathfrak{L}^1(\nu)$, $\{\varphi_{n,x}\}$ is L^1 -convergent to f_x , and from the very definition of the integral, $\{\int_Y \varphi_{n,x} d\nu\}$ converges to $\int_Y f_x d\nu$, all for fixed $x \notin (S \cup T)$. **Here is the main point again:** By construction, *pointwise convergence of the sequence $\{\Phi_n\}$ at x is $L^1(\nu)$ -convergence of $\{\varphi_{n,x}\}$.*

Next consider the map $\Psi_n : X \rightarrow E, x \mapsto \int_Y \varphi_{n,x} d\nu$. This is a step map with respect to \mathcal{A} being the composition of Φ_n and the linear integration map $\int_Y d\nu : \mathfrak{L}^1(\nu) \rightarrow E$. A similar argument as above applies to see that $\{\Psi_n\}$ is L^1 -Cauchy in $\mathfrak{L}^1(\mu)$:

$$\begin{aligned} \|\Psi_n - \Psi_m\|_1 &= \int_X |\Psi_n(x) - \Psi_m(x)| d\mu(x) = \int_X \left| \int_Y \varphi_{n,x} d\nu - \int_Y \varphi_{m,x} d\nu \right| d\mu(x) \\ &= \int_X \left| \int_Y (\varphi_{n,x} - \varphi_{m,x}) d\nu \right| d\mu(x) \leq \int_X \int_Y |\varphi_{n,x} - \varphi_{m,x}| d\nu d\mu(x) = \\ &= \int_{X \times Y} |\varphi_n - \varphi_m| d(\mu \otimes \nu). \end{aligned}$$

For all $x \notin S \cup T$ we know that $\{\Psi_n(x)\}$ is a sequence in E converging to $\Psi(x) = \int_Y f_x d\nu$. Because $\{\Psi_n\}$ is also $L^1(\mu)$ -Cauchy, by letting $n \rightarrow \infty$, we get that $\int_X \int_Y \varphi_{n,x} d\nu d\mu(x)$ converges to $\int_X \int_Y f_x d\nu d\mu(x)$. But φ_n is a step-map with respect to $\mathcal{A} \times \mathcal{B}$ and thus $\int_X \Psi_n d\mu = \int_X \int_Y \varphi_{n,x} d\nu d\mu(x) = \int_{X \times Y} \varphi_n d(\mu \otimes \nu)$. This proves the claimed equality of the integrals by letting $n \rightarrow \infty$. Then the left hand integral converges to $\int_X \int_Y f_x d\nu d\mu(x)$ while the right hand side converges to $\int_{X \times Y} f d(\mu \otimes \nu)$.

The statement about the norm-preserving isomorphism follows by applying the above to $|f|$ and checking with the definitions of the norms. ■

If we apply Fubini's theorem to the characteristic function of a measurable subset of $X \times Y$ we get

4.86. Corollary. *Let $Q \subset X \times Y$ be measurable with finite measure. Then*

$$\int_{X \times Y} \chi_Q d(\mu \otimes \nu) = \int_X \int_Y (\chi_Q)_x d\nu d\mu(x).$$

We have shown that the map $x \mapsto \int_Y (\chi_Q)_x d\nu$ is μ -measurable. Note that this implies measurable after changing on a set of measure zero. It is in fact true that this map is measurable *on the nose* but it requires work to show this, see Lang, VI, Exercise 11. Recall that μ -measurability is weaker than integrability. The assumption of Fubini's Theorem, Part I, is integrability of the function

on the product, and we conclude integrability of the *partial* functions. In the following the assumption of integrability in Fubini's theorem will be replaced by $(\mu \otimes \nu)$ -measurability and existence of the partial integrals.

4.87. Lemma. *If $f : X \times Y \rightarrow E$ is $\mu \otimes \nu$ -measurable then f_x is μ -measurable for almost all $x \in X$.*

Proof. Let $Z \subset X \times Y$ such that $f|(X - Z)$ is measurable and $f(X - Z) \subset \overline{D}$ for $D \subset E$ countable. By 4.82 for almost all $x \in X$ we have $\mu(Z_x) = 0$ and by 4.80 (ii) $f_x|(Y - Z_x)$ is measurable and so ν -measurable by the Technical Theorem 4.36. ■

4.88. Fubini's Theorem, Part II *Suppose that $f : X \times Y \rightarrow E$ is μ -measurable. Assume that for almost all $x \in X$ the map f_x is in $\mathfrak{L}^1(\nu)$ and the map defined by $x \mapsto \int_Y |f_x| d\nu$ for almost all x and arbitrarily otherwise, is in $\mathfrak{L}^1(\mu)$. Then $f \in \mathfrak{L}^1(\mu \otimes \nu, E)$ and the conclusions of Fubini's Theorem, Part I, apply.*

Proof. Recall that by the Corollary 4.64 of the dominated convergence theorem it suffices to prove that $|f| \in \mathfrak{L}^1(\mu \otimes \nu, \mathbb{R})$. So assume without restriction that f is a real-valued function with values in $[0, \infty)$ and satisfies the assumptions. After possibly changing f on a set of measure 0 we know that there exists an increasing sequence of simple nonnegative functions converging to f pointwise. Using the σ -finiteness of $X \times Y$, just like in the proof of the Technical theorem, we can assume that each φ_n vanishes outside of a set of finite measure and thus is a step map. Now for each x , $\{\varphi_{n,x}\}$ is increasing to f_x . For those x with $f_x \in \mathfrak{L}^1$ and $\varphi_{n,x}$ μ -measurable it follows for $n \rightarrow \infty$ that $\int_Y \varphi_{n,x} d\nu$ is increasing and convergent to $\int_Y f_x d\nu$, just by the very definition of $\int_Y f_x d\nu$ and monotone convergence. By 4.86 and monotone convergence again we conclude that the sequence $\{\int_{X \times Y} \varphi_n d(\mu \otimes \nu) = \int_X \int_Y \varphi_{n,x} d\nu d\mu(x)\}$ is increasing and convergent to $\int_X \int_Y f_x d\nu d\mu(x)$. We can apply monotone convergence once more to see that $f \in \mathfrak{L}^1$. ■