# Qiuzhen Lectures on Functional Analysis

Bin Gui

Yau Mathematical Sciences Center, Tsinghua University.

binguimath@gmail.com bingui@tsinghua.edu.cn

July 9, 2025

Current writing

# Contents

**1 Preliminaries** **3**

1.1 Notation . . . 3

1.2 Review of important facts in point-set topology . . . 6

1.3 $*$-algebras and the Stone-Weierstrass theorem . . . 12

1.4 Review of measure theory: general facts . . . 16

1.5 Review of measure theory: Radon measures . . . 21

1.6 Basic facts about increasing functions . . . 26

1.7 The Stieltjes integral . . . 28

1.8 The Riesz representation theorem via the Stieltjes integral . . . 34

**2 Normed vector spaces and their dual spaces** **37**

2.1 The origin of dual spaces in the calculus of variations . . . 37

2.2 Moment problems: a bridge between integral theory and dual spaces 39

2.3 Bounded multilinear maps . . . 49

2.4 Fundamental properties of bounded multilinear maps . . . 53

2.5 The roles of completeness and duality . . . 56

2.6 Dual spaces and the weak-* topology . . . 59

2.7 Weak-* convergence in $L^p$-spaces . . . 62

2.8 Weak-* convergence in $l^p$-spaces . . . 66

2.9 Weak-* convergence of distribution functions . . . 70

2.10 Weak-* approximation of Radon measures by Dirac measures . . . 73

**3 Basics of inner product spaces** **78**
3.1 Sesquilinear forms . . . 78
3.2 Inner product spaces and bounded sesquilinear forms . . . 81
3.3 Orthogonality . . . 83
3.4 Hilbert spaces . . . 89
3.5 Bounded linear maps and bounded sesquilinear forms . . . 92
3.6 Bounded matrices . . . 97

**4 Spectral theorem for bounded self-adjoint operators** **101**
4.1 Prehistory of spectral theory: continued fractions . . . 101
4.2 Prehistory of spectral theory: the polynomial moment problem . . . 101

**Index** **102**

**References** **105**

# 1 Preliminaries

## 1.1 Notation

In this monograph, unless otherwise stated, we understand the field $\mathbb{F}$ as either $\mathbb{R}$ or $\mathbb{C}$.

We use frequently the abbreviations:

iff=if and only if

LHS=left hand side    RHS=right hand side

$\exists$=there exists    $\forall$=for all

i.e.=id est=that is=namely    e.g.=for example

cf.=compare/check/see/you are referred to

resp.=respectively    WLOG=without loss of generality

LCH=locally compact Hausdorff

MCT=monotone convergence theorem

DCT=dominated convergence theorem

When we write $A := B$ or $A \xlongequal{\mathrm{def}} B$, we mean that $A$ is defined by the expression $B$. When we write $A \equiv B$, we mean that $A$ are $B$ are different symbols of the same object.

Unless otherwise stated, an inner product space $V$ denotes a complex inner product space, and its sesquilinear form $\langle\cdot|\cdot\rangle$ is linear on the right argument $|\cdot\rangle$ and antilinear on the left argument $\langle\cdot|$. Note that this convention is different from that of [Gui-A], where the right variable is antilinear.

If $V$ is an $\mathbb{F}$-vector space, then for each $v \in V$ and each linear map $\varphi : V \to \mathbb{F}$, we write

$$\langle v, \varphi\rangle = \langle \varphi, v\rangle := \varphi(v)$$

We assume $a\cdot(+\infty) = (+\infty)\cdot a = +\infty$ if $a \in (0, +\infty]$, and $0\cdot(+\infty) = (+\infty)\cdot 0 = 0$. An increasing function/sequence/net means a non-decreasing one.

- Unless otherwise specified, completeness of a metric space or normed vector space refers to Cauchy completeness.
- $\mathbb{N} = \{0, 1, 2, \dots\}$, $\mathbb{Z}_+ = \{1, 2, \dots\}$.
- $\mathbb{R}_{\geqslant 0} = [0, +\infty)$, $\overline{\mathbb{R}}_{\geqslant 0} = [0, +\infty]$, $\overline{\mathbb{R}} = [-\infty, +\infty]$.
- An **interval** denotes a connected subset of $\overline{\mathbb{R}}$. A **proper interval** denotes an interval with non-zero Lebesgue measure.

- $Y^X$ is the set of functions with domain $X$ and codomain $Y$.
- $2^X$ is the set of subsets of $X$.
- $\mathrm{fin}(2^X)$ is the set of finite subsets of $X$.
- If $f : X \to Y$ is a map, then

  $$\mathrm{Rng}(f) = f(X)$$

  If $X, Y$ are vector spaces and $f$ is linear, then

  $$\mathrm{Ker}(f) = f^{-1}(0)$$

- If $V$ is a vector space and $X$ is a set, then $V^X$ is viewed as a vector space whose linear structure is defined by

  $$(af + bg)(x) = af(x) + bg(x) \qquad \text{for all } f, g \in V^X \text{ and } a, b \in \mathbb{F}$$

- If $X$ is a set and $A \subset X$, the **characteristic function** is

  $$\chi_A : X \to \{0, 1\} \qquad x \mapsto \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in X \backslash A \end{cases}$$

- $\mathrm{Cl}_X(A)$, also denoted by $\mathrm{Cl}(A)$ or $\overline{A}$, is the closure of $A \subset X$ with respect to the topological space $X$.
- If $X$ is a metric space and $p \in X, r \in [0, +\infty]$, we let

  $$B_X(p, r) = \{x \in X : d(x, p) < r\} \qquad \overline{B}_X(p, r) = \{x \in X : d(x, p) \leqslant r\}$$

  For each $E \subset X$, we define the **diameter**

  $$\mathrm{diam}(E) = \sup\{d(x, y) : x, y \in E\}$$

- If $X$ is a topological space, then $\mathcal{T}_X$ denotes the topology of $X$, i.e.,

  $$\mathcal{T}_X = \{\text{open subsets of } X\}$$

  If $x \in X$, a **neighborhood** of $x$ denotes an *open* subset of $X$ containing $x$. We let

  $$\mathrm{Nbh}_X(x) \equiv \mathrm{Nbh}(x) := \{\text{neighborhoods of } x \text{ in } X\}$$

- If $X, Y$ are topological spaces, then

$$C(X,Y) = \{f \in Y^X : f \text{ is continuous}\}$$
$$\mathfrak{B}_X = \text{the Borel } \sigma\text{-algebra of } X$$
$$\mathscr{Bor}(X,Y) = \{f \in Y^X : f \text{ is Borel}\}$$

- $m^n$, as a measure, denotes the Lebesgue measure on $\mathbb{R}^n$, and is abbreviated to $m$ when no confusion arises.

- $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\} \simeq \mathbb{R}/2\pi\mathbb{Z}$. If $f$ is a function on $\mathbb{S}^1$, equivalently, a $2\pi$-periodic function on $\mathbb{R}$, then

$$\widehat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-\mathbf{i}nx}dm(x)$$

is its $n$-th Fourier coefficient (whenever the integral can be defined).

- $(X, \mathfrak{M}, \mu)$, often abbreviated to $(X, \mu)$, denotes a measure space where $\mathfrak{M}$ is the $\sigma$-algebra and $\mu : \mathfrak{M} \to \overline{\mathbb{R}}_{\geqslant 0}$ is the measure.

- Let $V$ be a normed vector space. Let $X$ is either a set or a topological space, depending on the context. Let $1 \leqslant p < +\infty$. For each $f \in V^X$,

$$\mathrm{Supp}_X(f) \equiv \mathrm{Supp}(f) = \mathrm{Cl}_X(\{x \in X : f(x) \neq 0\})$$
$$\|f\|_{l^\infty(X,V)} = \|f\|_{l^\infty} = \sup_{x\in X}\|f(x)\|$$
$$\|f\|_{l^p(X,V)} = \|f\|_{l^p} = \Big(\sum_{x\in X}\|f(x)\|^p\Big)^{\frac{1}{p}}$$
$$|f| \text{ is the function } X \to \mathbb{R}_{\geqslant 0} \text{ such that } |f|(x) = \|f(x)\|$$

We call $|f|$ the **absolute value function** of $f$. For each $E \subset V$, we let

$$C_c(X,E) = \{f \in C(X,E) : \mathrm{Supp}(f) \text{ is compact in } X\}$$
$$l^\infty(X,V) = \{f \in V^X : \|f\|_\infty < +\infty\}$$
$$l^p(X,V) = \{f \in V^X : \|f\|_p < +\infty\}$$

We are particularly interested in the case that $E = V$, $E = [0,1]$, and $E = \mathbb{R}_{\geqslant 0}$.

- Let $V$ be a normed vector space. Let $X$ be a set. We say that a family $(f_\alpha)_{\alpha\in\mathscr{A}}$ in $V^X$ is **uniformly bounded** if $\sup_{\alpha\in\mathscr{A}}\|f_\alpha\|_{l^\infty(X,V)} < +\infty$.

- If $X$ is LCH and $V$ is a normed $\mathbb{F}$-vector space, we understand $C_c(X,V)$ as a normed $\mathbb{F}$-vector space whose linear structure inherits from that of $V^X$, and whose norm is chosen to be the $l^\infty$-norm.

- If $(X, \mathfrak{M})$ and $(Y, \mathfrak{N})$ are measurable spaces, then

$$\mathcal{L}(X, Y) = \{\text{measurable functions } X \to Y\}$$

If $V$ is a normed vector space, for each $f \in \mathcal{L}(X, V)$ and $1 \leqslant p < +\infty$, we let

$$\|f\|_{L^p(X,\mu)} = \|f\|_{L^p} = \Big(\int_X |f|^p d\mu\Big)^{\frac{1}{p}}$$
$$\|f\|_{L^\infty(X,\mu)} = \|f\|_{L^\infty} = \inf\{\lambda \in \overline{\mathbb{R}}_{\geqslant 0} : \mu\{x \in X : \|f(x)\| > a\} = 0\}$$

which are potentially infinite.

- In the notation of function spaces, the codomain is understood to be $\mathbb{C}$ when it is suppressed. For example,

$$C_c(X) = C_c(X, \mathbb{C}) \qquad \mathscr{Bor}(X) = \mathscr{Bor}(X, \mathbb{C}) \qquad L^p(X, \mu) = L^p(X, \mu, \mathbb{C})$$

However, this convention does not apply to $\mathfrak{L}(V)$: If $V$ is a normed vector space, then $\mathfrak{L}(V)$ denotes $\mathfrak{L}(V, V)$, the space of bounded linear operators on $V$.

## 1.2 Review of important facts in point-set topology

Fix a normed vector space $\mathcal{V}$.

### 1.2.1 Miscellaneous definitions and properties

**Definition 1.1.** If $X, Y$ are metric spaces and $f : X \to Y$ is map, we say that $C \in \mathbb{R}_{\geqslant 0}$ is a **Lipschitz constant** of $f$ if

$$d(f(x_1), f(x_2)) \leqslant C d(x_1, x_2) \qquad \text{for all } x_1, x_2 \in X$$

If $f$ has a Lipschitz constant, we say that $f$ is **Lipschitz continuous**.

**Definition 1.2.** If $d$ and $d'$ are two metrics on a set $X$, we say that $d$ and $d'$ are **equivalent** if there exists $\alpha, \beta \in \mathbb{R}_{>0}$ such that

$$d(x, y) \leqslant \alpha d'(x, y) \qquad d'(x, y) \leqslant \beta d(x, y) \qquad \text{for all } x, y \in X$$

**Definition 1.3.** Let $X_1, \ldots, X_N$ be metric spaces. For each $1 \leqslant p < +\infty$, the ***l*$^\infty$-product metric** $d_\infty$ and the ***l*$^p$-product metric** $d_p$ are the metrics on $X_1 \times \cdots \times X_N$ defined by

$$d_\infty((x_1, \ldots, x_N), (y_1, \ldots, y_N)) := \max\{d(x_1, y_1), \ldots, d(x_N, y_N)\}$$
$$d_p((x_1, \ldots, x_N), (y_1, \ldots, y_N)) := \sqrt[p]{d(x_1, y_1)^p + \cdots + d(x_N, y_N)}$$

for all $x_i, y_i \in X_i$. These metrics are equivalent. We equip $X_1 \times \cdots \times X_N$ with any metric equivalent to $l^\infty$ and $l^p$.

**Remark 1.4.** Recall that if $f : X \to Y$ is a map of topological spaces, and $X = \bigcup_{i \in I} U_i$ is an open cover of $X$, then $f$ is continuous iff $f|_{U_i} : U_i \to Y$ is continuous for any $i \in I$.

**Definition 1.5.** Let $f : X \to Y$ be a map where $(Y, \mathcal{T}_Y)$ is a topological space. The **pullback topology** on $X$ is defined to be

$$f^*\mathcal{T}_Y := f^{-1}(\mathcal{T}_Y) = \{f^{-1}(V) : V \in \mathcal{T}_Y\}$$

Then, a net $(x_\alpha)$ in $X$ converges under $f^*\mathcal{T}_Y$ to $x$ iff

$$\lim_\alpha f(x_\alpha) = f(x)$$

### 1.2.2 Product topology and pointwise convergence

Let $(X_\alpha)_{\alpha \in \mathscr{A}}$ be a family of topological spaces. Elements of the product space

$$S = \prod_{\alpha \in \mathscr{A}} X_\alpha$$

are denoted by $x = (x_\alpha)_{\alpha \in \mathscr{A}}$. Let

$$\pi_\alpha : S \to X_\alpha \qquad x \mapsto x(\alpha)$$

It is easy to check that

$$\begin{aligned}\mathcal{B} = \Big\{ \prod_{\alpha \in \mathscr{A}} U_\alpha &: \text{each } U_\alpha \text{ is open in } X_\alpha, \\ &U_\alpha = X_\alpha \text{ for all but finitely many } \alpha \Big\} \\ =& \Big\{ \bigcap_{\alpha \in E} \pi_\alpha^{-1}(U_\alpha) : E \in \operatorname{fin}(2^{\mathscr{A}}),\ U_\alpha \text{ is open in } X_\alpha \text{ for each } \alpha \in E \Big\}\end{aligned}$$

is a base for a topology, namely, for each $W_1, W_2 \in \mathcal{B}$ and $x \in W_1 \cap W_2$, there exists $W_3 \in \mathcal{B}$ such that $W_3 \subset W_1 \cap W_2$. Therefore, $\mathcal{B}$ generates a topology.

**Definition 1.6.** The topology of $S$ generated by $\mathcal{B}$ is called the **product topology** or **pointwise convergence topology** of $S$. Unless otherwise stated, the product of a family of topological spaces is equipped with the product topology.

**Remark 1.7.** If each $X_\alpha$ is Hausdorff, then $S$ is clearly Hausdorff.

**Theorem 1.8.** *Let $(x_\mu)_{\mu \in I}$ be a net in $S$, and let $x \in S$. Then the following conditions are equivalent:*

*(a)* $\lim_{\mu \in I} x_\mu = x$ *under the product topology.*

(b) $(x_\mu)_{\mu\in I}$ *converges pointwise to* $x$*, namely, for each* $\alpha\in\mathscr{A}$ *we have* $\lim\limits_{\mu\in I} x_\mu(\alpha) = x(\alpha)$ *in* $X_\alpha$*.*

*Proof.* (a)⇒(b): Fix $\alpha\in\mathscr{A}$. For each open $U_\alpha\subset X_\alpha$, we have $\pi^{-1}(U_\alpha)\in\mathcal{B}$. Therefore,

$$\pi_\alpha : S\to X_\alpha \qquad \text{is continuous} \tag{1.1}$$

Thus, if $\lim_\mu x_\mu = x$, then $\lim_\mu \pi_\alpha(x_\mu)=\pi_\alpha(x)$. This proves (b).

(b)⇒(a): Assume (b). Choose any $W\in\mathcal{B}$ containing $x$. Then there exists $E\in\mathrm{fin}(2^{\mathscr{A}})$ such that $W=\bigcap_{\alpha\in E}\pi_\alpha^{-1}(U_\alpha)$, where each $U_\alpha\subset X_\alpha$ is open and containing $x_\alpha$. For such $\alpha\in E$, since $\lim_\mu x_\mu(\alpha)=x(\alpha)$, we know that $(x_\mu(\alpha))$ is $\mu$-eventually in $U_\alpha$. Therefore, since $E$ is finite, we conclude that $(x_\mu)$ is eventually in $W$. This proves (a). □

**Corollary 1.9.** *Let* $Z$ *be a topological space. Suppose that for each* $\alpha\in\mathscr{A}$*, a map* $f_\alpha : Z\to X_\alpha$ *is chosen. Then*

$$\bigvee_{\alpha\in\mathscr{A}} f_\alpha : Z\to\prod_{\alpha\in\mathscr{A}} X_\alpha \qquad z\mapsto (f_\alpha(z))_{\alpha\in\mathscr{A}} \tag{1.2}$$

*is continuous iff* $f_\alpha$ *is continuous for each* $\alpha\in\mathscr{A}$*.*

*Proof.* If $F:=\bigvee_{\alpha\in\mathscr{A}} f_\alpha$ is continuous, then since $\pi_\alpha$ is continuous, $f_\alpha=\pi_\alpha\circ f_\alpha$ is also continuous. Conversely, suppose that each $f_\alpha$ is continuous. Let $(z_i)$ be a net in $Z$ converging to $z\in Z$. For each $\alpha$, since $f_\alpha$ is continuous, we see that $\lim_i f_\alpha(z_i)=f_\alpha(z)$. By Thm. 1.8, $F(z_i)$ converges to $F(z)$. This proves that $F$ is continuous. □

**Proposition 1.10.** *Suppose that* $\mathscr{A}$ *is countable. If each* $X_\alpha$ *is second countable, then* $S$ *is second countable. If each* $X_\alpha$ *is metrizable, then* $S$ *is metrizable.*

*Proof.* If $\mathcal{U}_\alpha$ is a base of the topology of $X_\alpha$, then

$$\mathcal{U}:=\Big\{\bigcap_{\alpha\in E}\pi_\alpha^{-1}(U_\alpha) : E\in\mathrm{fin}(2^{\mathscr{A}}), U_\alpha\in\mathcal{U}_\alpha\Big\}$$

is a base of the the product topology, which is countable if each $\mathcal{U}_\alpha$ is countable.

Now assume that each $X_\alpha$ is equipped with a metric $d_\alpha$. Fix any $R\in\mathbb{R}_{>0}$, and let $\widetilde{d}_\alpha$ be metric on $X_\alpha$ inducing the same topology as $d_\alpha$, and satisfies $d_\alpha\leqslant R$. For example,

$$\widetilde{d}_\alpha(x_\alpha,y_\alpha)=\min\{d_\alpha(x_\alpha,y_\alpha),R\} \qquad \text{for each } x_\alpha,y_\alpha\in X_\alpha \tag{1.3a}$$

Let $\nu:\mathscr{A}\to\mathbb{Z}_+$ be an injective map, and define a metric $d$ on $S$ by

$$d(x,y)=\sum_{\alpha\in\mathscr{A}} 2^{-\nu(\alpha)}\widetilde{d}_\alpha(x(\alpha),y(\alpha)) \qquad \text{for each } x,y\in S \tag{1.3b}$$

One shows easily that a net $(x_\mu)$ in $S$ converging to $x\in S$ iff $\lim_\mu \widetilde{d}_\alpha(x_\mu(\alpha),x(\alpha))=0$ for all $\alpha\in\mathscr{A}$. Therefore, by Thm. 1.8, $d$ induces the product topology. □

**Theorem 1.11 (Tychonoff theorem).** *Assume that $X_\alpha$ is compact for each $\alpha \in \mathscr{A}$. Then $S$ is compact.*

$\star$ *Proof.* Assume WLOG that $\mathscr{A}$ is non-empty, that each $X_\alpha$ is non-empty. Let $(x_\mu)_{\mu\in I}$ be a net in $S$. We want to show that $(x_\mu)_{\mu\in I}$ has a cluster point.

For each $\mathscr{E} \subset \mathscr{A}$, let $S_{\mathscr{E}} = \prod_{\alpha\in\mathscr{E}} X_\alpha$. For each $x \in S_{\mathscr{E}}$, we write $\mathrm{Dom}(x) = \mathscr{E}$. For each $\mathscr{E} \subset \mathscr{F} \subset \mathscr{A}$ and $y \in S_{\mathscr{F}}$, let $y|_{\mathscr{E}} = (y(\alpha))_{\alpha\in\mathscr{E}}$. Let

$$P = \bigcup_{\mathscr{E}\subset\mathscr{A}} \{x \in S_{\mathscr{E}} : x \text{ is a cluster point of } (x_\mu|_{\mathscr{E}})_{\mu\in I} \text{ in } S_{\mathscr{E}}\}$$

equipped with the partial order "$\subset$". In other words, if $x, y \in P$, then $x \leqslant y$ means that $\mathrm{Dom}(x) \subset \mathrm{Dom}(y)$ and $x = y|_{\mathscr{E}}$.

Since each $X_\alpha$ is compact, $P$ is clearly non-empty. Let us show that every totally ordered non-empty subset $Q \subset P$ has an upper bound in $P$, so that Zorn's lemma can be applied. Let $x$ be the union of all elements of $Q$. Thus $x \in S_{\mathscr{E}}$ where $\mathscr{E} = \bigcup_{y\in Q} \mathrm{Dom}(y)$, and we have $x|_{\mathrm{Dom}(y)} = y$ for each $y \in Q$.

To show that $x$ is a cluster point of $(x_\mu|_{\mathscr{E}})_{\mu\in I}$ in $S_{\mathscr{E}}$, we pick any neighborhood of $x$ in $S_{\mathscr{E}}$, which, after shrinking if necessary, is of the form $W = \prod_{\alpha\in\mathscr{E}} U_\alpha$ where each $U_\alpha \subset X_\alpha$ is open, and there exists $K \in \mathrm{fin}(2^{\mathscr{E}})$ such that $U_\alpha = X_\alpha$ whenever $\alpha \notin K$. Since $\mathscr{E} = \bigcup_{y\in Q} \mathrm{Dom}(y)$, there exists $y \in Q$ such that $K \subset \mathrm{Dom}(y)$. Namely, $(x_\mu|_{\mathrm{Dom}(y)})_{\mu\in I}$ has cluster point $y$, and $K \subset \mathrm{Dom}(y)$. Therefore $(x_\mu|_K)_{\mu\in I}$ has cluster point $y|_K$ (which equals $x|_K$ because $x|_{\mathrm{Dom}(y)} = y$), and hence is frequently in $\prod_{\alpha\in K} U_\alpha$. Thus $(x_\mu|_{\mathscr{E}})_{\mu\in I}$ is frequently in $W$. This finishes the proof that $x \in P$. Clearly $x$ is an upper bound of $Q$.

Now we can apply Zorn's lemma, which claims that $P$ has a maximal element $x \in P$. The proof of the Tychonoff theorem will be finished by showing that $\mathscr{E} := \mathrm{Dom}(x)$ equals $\mathscr{A}$. Suppose not. Choose $\beta \in \mathscr{A}\backslash\mathscr{E}$. Since $x \in P$, there is a subnet $(x_{\mu_\nu}|_{\mathscr{E}})_{\nu\in J}$ of $(x_\mu|_{\mathscr{E}})_{\mu\in I}$ converging pointwise to $x$. Since $X_\beta$ is compact, $(x_{\mu_\nu}(\beta))_{\nu\in J}$ has a converging subnet $(x_{\mu_{\nu_\upsilon}}(\beta))_{\upsilon\in L}$. Define $\widetilde{x} \in S_{\mathscr{E}\cup\{\beta\}}$ to be $x$ when restricted to $\mathscr{E}$, and $\widetilde{x}(\beta) := \lim_\upsilon x_{\mu_{\nu_\upsilon}}(\beta)$. Then $\widetilde{x} \in P$, and $\widetilde{x}$ is strictly larger than $x$, contradicting the maximality of $x$. $\square$

**Remark 1.12.** If $\mathscr{A}$ is a countable set, and if each $X_\alpha$ is compact and metrizable, the **diagonal method** can be used in place of Zorn's lemma to prove that $S$ (which is metrizable by Prop. 1.10) is compact:

We consider the case that $\mathscr{A} = \mathbb{Z}_+$. (The case that $\mathscr{A}$ is finite is even simpler.) Let $(x_n)_{n\in\mathbb{Z}_+}$ be a sequence in $S$. We construct inductively a double sequence $(x_{m,n})_{m,n\in\mathbb{Z}_+}$ in $S$ as follows. Since $X_1$ is sequentially compact, $(x_n)$ has subseqeunce $(x_{1,n})_{n\in\mathbb{Z}_+}$ whose first component $(x_{1,n}(1))_{n\in\mathbb{Z}_+}$ converges to some $x(1) \in X_1$. Suppose that $(x_{m-1,n})_{n\in\mathbb{Z}_+}$ has been constructed (where $m-1 \geqslant 1$). Since $X_m$ is sequentially compact, $(x_{m-1,n})_{n\in\mathbb{Z}_+}$ has a subsequence $(x_{m,n})_{n\in\mathbb{Z}_+}$ whose $m$-th component $(x_{m,n}(m))_{n\in\mathbb{Z}_+}$ to some $x(m) \in S$. In this way, the double sequence $(x_{m,n})$ in $S$ and the element $x \in S$ are constructed. One checks easily that $(x_{n,n})_{n\in\mathbb{Z}_+}$ is a subsequence of $(x_n)$ converging to $x$. $\square$

### 1.2.3 Precompact sets

Let $X$ be a Hausdorff space.

**Definition 1.13.** Let $A \subset X$. We say that $A$ is **precompact** relative to $X$ and write

$$A \Subset X$$

if $\mathrm{Cl}_X(A)$ is compact, equivalently, if $A$ is contained in a compact subset of $X$.

Recall that a subset of a compact Hausdorff space is closed iff it is compact.

*Proof of equivalence.* "$\Rightarrow$": Obvious. "$\Leftarrow$": Let $B \subset X$ be compact and containing $A$. Then $B$ is closed in $X$. So $\mathrm{Cl}_X(A) \subset B$. Since $\mathrm{Cl}_X(A)$ is closed in $X$ and hence closed in $B$, it is compact. $\square$

**Remark 1.14.** Let $W \subset X$. Then for each $A \subset W$, we have

$$A \Subset W \qquad \Longleftrightarrow \qquad A \Subset X \text{ and } \mathrm{Cl}_X(A) \subset W$$

When either side is true, we have $\mathrm{Cl}_W(A) = \mathrm{Cl}_X(A)$. Thus, both $\mathrm{Cl}_W(A)$ and $\mathrm{Cl}_X(A)$ can be denoted unambiguously by $\overline{A}$.

In practice, we often choose $W$ to be an open subset of $X$.

*Proof.* "$\Leftarrow$": $\mathrm{Cl}_X(A)$ is a compact set inside $W$ and contains $A$. So $A \Subset W$.

"$\Rightarrow$": We have a compact set $B$ such that $A \subset B \subset W$. So $A \Subset X$. Since $B$ is closed in any larger set, we have $\mathrm{Cl}_X(A) \subset B$ and hence $\mathrm{Cl}_X(A) \subset W$.

It is obvious that $\mathrm{Cl}_W(A) \subset \mathrm{Cl}_X(A)$. Assume $A \Subset W$. Then $\mathrm{Cl}_W(A)$ is compact. In the above paragraph, if we choose $B = \mathrm{Cl}_W(A)$. then we have $\mathrm{Cl}_X(A) \subset B = \mathrm{Cl}_X(A)$. This proves $\mathrm{Cl}_W(A) = \mathrm{Cl}_X(A)$. $\square$

**Remark 1.15.** Let $U$ be an open subset of $X$. Let $f \in C_c(U, \mathcal{V})$. Then by zero-extension, $f$ can be viewed as an element of $C_c(X, \mathcal{V})$ supported in $U$. Briefly speaking, we have

$$C_c(U, \mathcal{V}) \subset C_c(X, \mathcal{V})$$

Moreover, for each $f \in C_c(U, \mathcal{V})$, we have

$$\mathrm{Supp}_U(f) = \mathrm{Supp}_X(f)$$

*Proof.* Let $f$ take value $0$ outside $U$. Let $K = \mathrm{Supp}_U(f)$, which is compact by assumption. Since $f|_U$ is continuous and $f|_{K^c} = 0$ are continuous, and since $X = U \cup K^c$ is an open cover on $X$, $f$ is continuous. By the Rem. 1.14, we have $\mathrm{Supp}_U(f) = \mathrm{Supp}_X(f)$. Therefore $f \in C_c(X, \mathcal{V})$. $\square$

Under the setting of Rem. 1.15, it is clear that

$$C_c(U, \mathcal{V}) = \{f \in C_c(X, \mathcal{V}) : \mathrm{Supp}_X(f) \subset U\} \tag{1.4}$$

### 1.2.4 LCH spaces

Let $X$ be LCH.

**Proposition 1.16.** *Any closed or open subset of $X$ is LCH.*

*Proof.* See [Gui-A, Subsec. 8.6.2]. $\square$

**Corollary 1.17.** *Let $W \subset X$ be an open subset. Let $K \subset W$ be compact. Then there exists an open subset $U$ of $X$ such that $K \subset U \Subset W$.*

*Proof.* The case that $K$ is a single point follows from the fact that $W$ is LCH, cf. Prop. 1.16. The general case follows from the compactness of $K$. $\square$

**Corollary 1.18.** *Let $K_1, K_2$ be mutually disjoint compact subsets of $X$. Then there exist open subsets $U_1, U_2$ of $X$ such that $K_1 \subset U_1$ and $K_2 \subset U_2$.*

*Proof.* This corollary in fact holds even without the assumption that $X$ is locally compact, and its proof is a straightforward exercise in point-set topology. However, it also follows directly from the results established above. Indeed, by Prop. 1.16, $X \backslash K_2$ is LCH. Therefore, by Cor. 1.17, there exists an open set $U_1$ such that $K_1 \subset U_1 \Subset X \backslash K_2$. Let $U_2 = X \backslash \overline{U}_1$. $\square$

**Theorem 1.19 (Urysohn's lemma).** *Let $K \subset X$ be compact. Then there exists a (continuous)* ***Urysohn function*** *$f$ with respect to $K$ and $X$, i.e., $f \in C_c(X, [0,1])$ and $f|_K = 1$.*

*Proof.* See [Gui-A, Sec. 15.4]. $\square$

**Remark 1.20.** Urysohn's lemma can be used in the following way. Suppose that $K \subset U \subset X$ where $K$ is compact and $U$ is open in $X$. By Prop. 1.16, $U$ is LCH. Therefore, by Thm. 1.19, there exists $f \in C_c(U, [0,1])$ such that $f|_K = 1$. By Rem. 1.15, $f$ can be viewed as an element of $C_c(X, [0,1])$ satisfying $f|_K = 1$ and $\mathrm{Supp}(f) \subset U$.

**Theorem 1.21.** *Let $K$ be a compact subset of $X$. Let $\mathfrak{U} = (U_1, \dots, U_n)$ be a finite collection of open subsets of $X$ covering $K$ (i.e. $K \subset U_1 \cup \cdots \cup U_n$). Then there exist $h_i \in C_c(U_i, \mathbb{R}_{\geqslant 0})$ (for all $1 \leqslant i \leqslant n$) satisfying the following conditions:*

*(1)* $0 \leqslant \sum_{i=1}^n h_i \leqslant 1$ *on* $X$.

*(2)* $\sum_{i=1}^n h_i\big|_K = 1$.

*Such $h_1, \dots, h_n$ are called a* ***partition of unity of $K$ subordinate to $\mathfrak{U}$***.

In fact, $h_1, \dots, h_n$ should be viewed as a partition of the Urysohn function $h := h_1 + \cdots + h_n$.

*Proof.* See [Gui-A, Sec. 15.4]. Note that condition (1) is not stated in some textbooks on partitions of unity. However, even if (1) is not initially satisfied, one can enforce it by setting $g(x) = \max\{\sum_i h_i(x), 1\}$ and replacing each $h_i$ with $h_i/g$. □

**Theorem 1.22 (Tietze extension theorem).** *Let $K$ be a compact subset of $X$. Let $f \in C(K, \mathbb{F})$. Then there exists $\widetilde{f} \in C_c(X, \mathbb{F})$ such that $\widetilde{f}|_K = f$, and that $\|\widetilde{f}\|_{l^\infty(X)} = \|f\|_{l^\infty(K)}$.*

*Proof.* See [Gui-A, Sec. 15.4]. □

**Definition 1.23.** We let

$$C_0(X, \mathcal{V}) = \begin{cases} \{f \in C(X, \mathcal{V}) : \lim_{x \to \infty} \|f(x)\| = 0\} & \text{if } X \text{ is not compact} \\ C(X, \mathcal{V}) = C_c(X, \mathcal{V}) & \text{if } X \text{ is compact} \end{cases}$$

where $\widehat{X} = X \cup \{\infty\}$ is the one-point compactification of $X$. Equivalently, $C_0(X, \mathcal{V})$ is the set of all $f \in C(X, \mathcal{V})$ such that for any $\varepsilon > 0$ there exists a compact $K \subset X$ such that $\|f\|_{l^\infty(X \backslash K)} < \varepsilon$. See [Gui-A, Subsec. 15.8.1] for more discussions. For each $E \subset \mathcal{V}$, we let

$$C_0(X, E) = C_0(X, V) \cap E^X$$

**Remark 1.24.** $C_0(X, \mathcal{V})$ is the $l^\infty$-closure of $C_c(X, \mathcal{V})$ in $C(X, \mathcal{V})$.

*Proof.* One easily shows that $C_0(X, \mathcal{V})$ is closed in $C(X, \mathcal{V})$. To show that $C_c(X, \mathcal{V})$ is dense in $C_0(X, \mathcal{V})$, we choose any $f \in C_0(X, \mathcal{V})$. Then for each $\varepsilon > 0$ there exists a compact $K \subset X$ such that $\|f\|_{l^\infty(K^c)} < \varepsilon$. By Urysohn's lemma, there exists $h \in C_c(X, [0, 1])$ such that $h|_K = 1$. Then $\|hf\|_{l^\infty(K^c)} < \varepsilon$, and hence $\|f - hf\|_{l^\infty(X)} < 2\varepsilon$. This finishes the proof, since $hf \in C_c(X, \mathcal{V})$. □

**Remark 1.25.** Suppose that $X$ is second countable. Then $X$ is Lindelöf. Therefore, $X$ has a countable open cover $\mathfrak{U} = (U_n)_{n \in \mathbb{Z}_+}$ whose members $U_n$ are precompact open subsets of $X$. In particular, $X$ is $\sigma$-compact, since $X = \bigcup_{n \in \mathbb{Z}_+} \overline{U_n}$ where each $\overline{U_n}$ is compact.

## 1.3 $*$-algebras and the Stone-Weierstrass theorem

Recall that $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. In this section, we let $\mathbb{K}$ be any subfield of $\mathbb{C}$ closed under complex conjugation, such as $\mathbb{R}, \mathbb{C}, \mathbb{Q}, \mathbb{Q} + \mathbf{i}\mathbb{Q}$.

**Definition 1.26.** A $\mathbb{K}$**-algebra** is defined to be a ring $\mathscr{A}$ (not necessarily having $1$) that is also a $\mathbb{K}$-vector space, such that the vector addition agrees with the ring addition, and the scalar multiplication is compatible with the ring multiplication in the following sense: for all $\lambda \in \mathbb{K}$ and $x, y \in \mathscr{A}$, we have

$$\lambda(xy) = (\lambda x)y = x(\lambda y) \tag{1.5}$$

A $\mathbb{K}$-algebra is called **unital** if $\mathscr{A}$, as a ring, has a multiplicative identity $1$. In this case, we write $\lambda \cdot 1$ as $\lambda$ if $\lambda \in \mathbb{K}$.

A $\mathbb{K}$-algebra is called **commutative** or **abelian** if $xy = yx$ for all $x, y \in \mathscr{A}$.

If $\mathscr{A}$ is a $\mathbb{K}$-algebra, then a **(**$\mathbb{K}$**-)subalgebra** is a subset $\mathscr{B}$ which is invariant under the ring addition, ring multiplication, and scalar multiplication. (Namely, $\mathscr{B}$ is a subring and also a subspace of $\mathscr{A}$.) If $\mathscr{A}$ is unital, then a **unital (**$\mathbb{K}$**-)subalgebra** of $\mathscr{A}$ is a $\mathbb{K}$-subalgebra containing the identity of $\mathscr{A}$. □

**Remark 1.27.** A unital $\mathbb{K}$-algebra $\mathscr{A}$ can equivalently be described as a ring with identity, together with a ring homomorphism $\mathbb{C} \to Z(\mathscr{A})$ where $Z(\mathscr{A})$ is the **center** of $\mathscr{A}$, i.e.

$$Z(\mathscr{A}) = \{x \in \mathscr{A} : xy = yx \text{ for every } y \in \mathscr{A}\}$$

We leave the verification of this equivalence to the reader.

**Example 1.28.** If $V$ is a $\mathbb{F}$-vector space, then $\mathrm{End}(V)$, the set of $\mathbb{F}$ linear maps $V \to V$, is naturally an $\mathbb{F}$-algebra. If $V$ is a normed vector space, then $\mathfrak{L}(V)$ is an $\mathbb{F}$-algebra.

**Definition 1.29.** A ***-**$\mathbb{K}$**-algebra** is defined to be a $\mathbb{K}$-algebra together with an **antilinear map** $* : \mathscr{A} \to \mathscr{A}$ sending $x$ to $x^*$ (where "antilinear" means that for every $a, b \in \mathbb{C}$ and $x, y \in \mathscr{A}$ we have $(ax + by)^* = \overline{a}x^* + \overline{b}y^*$) such that for every $x, y \in \mathscr{A}$, we have

$$(x^*)^* = x \qquad (xy)^* = y^*x^*$$

Note that $*$ must be bijective. We call $*$ an **involution**. A ***-**$\mathbb{K}$**-subalgebra** $\mathscr{B}$ is defined to be a subalgebra satisfying $x \in \mathscr{B}$ iff $x^* \in \mathscr{B}$. If $\mathscr{A}$ is a unital algebra with unit $\mathbf{1}$, we say that $\mathscr{A}$ is a **unital *-**$\mathbb{K}$**-algebra** if $\mathscr{A}$ is equipped with an involution $* : \mathscr{A} \to \mathscr{A}$ such that $\mathscr{A}$ is a *-algebra, and that

$$\mathbf{1}^* = \mathbf{1}$$

A unital *-subalgebra is a unital subalegbra and also a *-subalgebra.

**Convention 1.30.** We omit "$\mathbb{K}$-" when $\mathbb{K}$ is $\mathbb{C}$. For example, a **unital *-algebra** means a unital $*$-$\mathbb{C}$-algebra.

**Example 1.31.** The set of complex $n \times n$ matrices $\mathbb{C}^{n\times n}$ is naturally a unital $*$-algebra if for every $A \in \mathbb{C}^{n\times n}$ we define $A^* = \overline{A}^{\mathrm{t}}$, the complex conjugate of the transpose of $A$.

**Example 1.32.** Let $X$ be a set. Then $\mathbb{K}^X$ is naturally a unital $\mathbb{K}$-algebra, and $l^\infty(X,\mathbb{K})$ is its unital $\mathbb{K}$-subalgebra. If $X$ is a topological space, then $C(X,\mathbb{K})$ is a unital $\mathbb{K}$-subalgebra of $\mathbb{K}^X$. If $X$ is compact, then $C(X,\mathbb{K})$ is a unital $\mathbb{K}$-subalgebra of $l^\infty(X,\mathbb{K})$.

**Example 1.33.** Let $X$ be a set. Then $\mathbb{C}^X$ is a unital *-algebra if for every $f \in \mathbb{C}^X$ we define

$$f^* : X \to \mathbb{C} \qquad \boldsymbol{f^*(x)} = \overline{f(x)} \tag{1.6}$$

Then $\mathbb{K}^X$ and $l^\infty(X,\mathbb{K})$ are unital *-$\mathbb{K}$-subalgebras of $\mathbb{C}^X$.

Assume that $X$ is a compact topological space. Then $C(X,\mathbb{F})$ is a unital *-$\mathbb{F}$-subalgebra of $l^\infty(X,\mathbb{F})$. If $f_1,\dots,f_n \in C(X,\mathbb{F})$, then $\mathbb{F}[f_1,\dots,f_n]$, the set of polynomials of $f_1,\dots,f_n$ with coefficients in $\mathbb{F}$, is a unital $\mathbb{F}$-subalgebra of $C(X,\mathbb{F})$. And $\mathbb{F}[f_1,f_1^*,\dots,f_n,f_n^*]$ is a unital *-$\mathbb{F}$-subalgebra of $C(X,\mathbb{F})$. □

More generally, we have:

**Example 1.34.** Let $\mathscr{A}$ be an abelian unital $\mathbb{K}$-algebra. Let $\mathfrak{S} \subset \mathscr{A}$. Then

$$\mathbb{K}\langle\mathfrak{S}\rangle = \mathrm{Span}_{\mathbb{K}}\{x_1^{n_1}\cdots x_k^{n_k} : k \in \mathbb{Z}_+, x_i \in \mathfrak{S}, n_i \in \mathbb{N}\} \tag{1.7}$$

the set of (possibly non-commutative) polynomials of elements in $\mathfrak{S}$, is the smallest unital $\mathbb{K}$-subalgebra containing $\mathfrak{S}$, called the **unital $\mathbb{K}$-subalgebra generated by** $\mathfrak{S}$. (Here, we understand $x^0 = 1$ if $x \in \mathscr{A}$.) Thus, if $\mathscr{A}$ is an abelian unital *-algebra, then $\mathbb{C}\langle\mathfrak{S}\cup\mathfrak{S}^*\rangle$ (where $\mathfrak{S}^* = \{x^* : x \in \mathfrak{S}\}$) is the smallest unital *-algebra containing $\mathfrak{S}$, called the **unital *-$\mathbb{K}$-subalgebra generated by** $\mathfrak{S}$.

**Definition 1.35.** Let $X$ be sets. Let $(f_\alpha)_{\alpha\in\mathfrak{A}}$ be a family of maps where $f_\alpha : X \to Y_\alpha$ and $Y_\alpha$ is a set. We say that $(f_\alpha)_{\alpha\in\mathfrak{A}}$ **separates the points of $X$** if for any distinct $x_1, x_2 \in X$ there exists $\alpha \in \mathfrak{A}$ such that $f_\alpha(x_1) \neq f_\alpha(x_2)$. Equivalently, the map

$$\bigvee_{\alpha\in\mathfrak{A}} f_\alpha : X \to \prod_{\alpha\in\mathfrak{A}} Y_\alpha \qquad x \mapsto (f_\alpha(x))_{\alpha\in\mathfrak{A}} \tag{1.8}$$

is injective.

**Example 1.36.** Let $X$ be an LCH space. Then $C_c(X,[0,1])$ separates the points of $X$.

*Proof.* Choose any distinct points $x, y \in X$. By Urysohn's lemma (Rem. 1.20), there exists $f \in C_c(X,[0,1])$ such that $f(x) = 1$ and $\mathrm{Supp}(f) \subset X\backslash\{y\}$. So $f$ separates $x, y$. □

**Theorem 1.37 (Stone-Weierstrass theorem).** *Let $X$ be a compact Hausdorff space. Let $\mathfrak{S} \subset C(X, \mathbb{F})$. Suppose that $\mathfrak{S}$ separates the points of $X$. Then the $*$-$\mathbb{F}$-subalgebra $\mathbb{F}\langle\mathfrak{S} \cup \mathfrak{S}^*\rangle$ generated by $\mathfrak{S}$ is dense in $C(X, \mathbb{F})$ under the $l^\infty$-norm.*

Note that if $\mathbb{F} = \mathbb{R}$, then $\mathfrak{S}^* = \mathfrak{S}$ by (1.6).

If $\mathbb{F} = \mathbb{C}$, then since $(\mathbb{Q} + \mathrm{i}\mathbb{Q})\langle\mathfrak{S} \cup \mathfrak{S}^*\rangle$ is $l^\infty$-dense in $\mathbb{C}\langle\mathfrak{S} \cup \mathfrak{S}^*\rangle$, it is clear that $(\mathbb{Q} + \mathrm{i}\mathbb{Q})\langle\mathfrak{S} \cup \mathfrak{S}^*\rangle$ is $l^\infty$-dense in $C(X)$. Similarly, if $\mathbb{F} = \mathbb{R}$, then $\mathbb{Q}\langle\mathfrak{S}\rangle$ is $l^\infty$-dense in $C(X, \mathbb{R})$.

*Proof.* See [Gui-A, Ch. 15]. □

The following application of the Stone-Weierstrass theorem will be used in the study of weak-* topology, particularly in the proof of Thm. 2.46. Recall that $C(X, \mathbb{F})$ is equipped with the $l^\infty$-norm.

**Theorem 1.38.** *Let $X$ be a compact Hausdorff space. Then the following are equivalent:*

*(a) $X$ is metrizable.*

*(b) $X$ is second countable.*

*(c) There is a sequence $(f_n)_{n\in\mathbb{Z}_+}$ in $C(X, \mathbb{F})$ separating the points of $X$.*

*(d) $C(X, \mathbb{F})$ is separable.*

*Moreover, if (c) is satisfied, then for each $R \in \mathbb{R}_{>0}$, a compatible metric $d$ on $X$ can be chosen to be*

$$d(x, y) = \sum_{n\in\mathbb{Z}_+} 2^{-n} \min\{|f_n(x) - f_n(y)|, R\} \qquad \text{for each } x, y \in X \tag{1.9}$$

In particular, if (c) is satisfied and $\sup_{n\in\mathbb{Z}_+} \|f_n\|_{l^\infty} < +\infty$, we can choose $R = 2\sup_{n\in\mathbb{Z}_+} \|f_n\|_{l^\infty}$. Then (1.9) becomes

$$d(x, y) = \sum_{n\in\mathbb{Z}_+} 2^{-n}|f_n(x) - f_n(y)| \qquad \text{for each } x, y \in X \tag{1.10}$$

The Stone-Weierstrass theorem will be used in the direction (c)⇒(d). The equivalence of (a,b,c) does not rely on the Stone-Weierstrass theorem.

*Proof.* (a)⇒(b): By .

(b)⇒(c): Since $X$ is second countable, we can choose an infinite countable base $(U_n)_{n\in\mathbb{Z}}$ of the topology. For each $m, n \in \mathbb{Z}_+$, if $U_n \Subset U_m$, we choose $f_{m,n} \in C_c(U_m, [0, 1]) \subset C_c(X, [0, 1])$ such that $f|_{\overline{U}_n} = 1$ (which exists by Urysohn's lemma); otherwise, we let $f_{m,n} = 0$.

Let us prove that $\{f_{m,n} : m, n \in \mathbb{Z}_+\}$ separates the points of $X$: Choose distinct $x, y \in X$. Since $X\backslash\{y\} \in \mathrm{Nbh}_X(x)$, there exists $U_m$ containing $x$ and is contained in

$X\backslash\{y\}$. By Cor. 1.17, there exists $n$ such that $\{x\} \subset U_n \Subset U_m$. Then $f_{m,n}(x) = 1$ and $f_{m,n}(y) = 0$.

(c)⇒(a,b): Since $(f_n)$ separates points, the map

$$\Phi = \bigvee_n f_n : X \to \mathbb{F}^{\mathbb{Z}_+} \qquad x \mapsto (f_n(x))_{n\in\mathbb{Z}_+}$$

is injective. By Cor. 1.9, $\Phi$ is continuous. Since $X$ is compact, the map $\Phi$ restricts to a homeomorphism $\Phi : X \to \Phi(X)$, where $\Phi(X)$ is equipped with the subspace topology of the product topology of $\mathbb{F}^{\mathbb{Z}_+}$. By Prop. 1.10, $\mathbb{F}^{\mathbb{Z}_+}$ is metrizable and second countable, so $\Phi(X)$, and hence $X$, is metrizable and second countable. This proves (a) and (b).

By (1.3), the product topology of $\mathbb{F}^{\mathbb{Z}_+}$ is induced by the metric

$$\delta(u, v) = \sum_{n\in\mathbb{Z}_+} 2^{-n} \min\{|u(n) - v(n)|, R\} \qquad \text{for each } u, v \in \mathbb{F}^{\mathbb{Z}_+}$$

Therefore, the pullback metric $\Phi^*\delta$ on $X$ (defined by $\Phi^*\delta(x, y) = \delta(\Phi(x), \Phi(y))$) induces the topology of $X$. Clearly $\Phi^*\delta(x, y)$ equals (1.9).

(c)⇒(d): Let $\mathbb{K} = \mathbb{F} \cap (\mathbb{Q} + \mathbf{i}\mathbb{Q})$. By Stone-Weierstrass, the countable set $\mathbb{K}[\{f_n : n \in \mathbb{Z}_+\}]$ is dense in $C(X, \mathbb{F})$. Thus $C(X, \mathbb{F})$ is separable.

(d)⇒(c): By Exp. 1.36, $C(X, \mathbb{F})$ separates the points of $X$. Therefore, any dense subset of $C(X, \mathbb{F})$ separates the points of $X$. Since $C(X, \mathbb{F})$ is separable, it has a countable dense subset separating the points of $X$. □

## 1.4 Review of measure theory: general facts

### 1.4.1 Some useful definitions and their basic properties

**Definition 1.39.** Let $X$ be a set. Suppose that $\mathscr{C}$ is an $\mathbb{F}$-linear subspace of $\mathbb{F}^X$. A **positive linear functional** on $\mathscr{C}$ denotes an $\mathbb{F}$-linear map $\Lambda : \mathscr{C} \to \mathbb{F}$ such that $\Lambda(f) \geqslant 0$ for all $f \in \mathscr{C} \cap \mathbb{R}^X_{\geqslant 0}$.

Recall that if $(X, \mathfrak{M})$ is a measurable space, an **$\mathbb{F}$-valued simple function** on $X$ is an $\mathbb{F}$-linear combination of characteristic functions over measurable sets; that is, an element of $\mathrm{Span}_{\mathbb{F}}\{\chi_E : E \in \mathfrak{M}\}$.

**Definition 1.40.** Let $X$ be a set. Let $x \in X$. The **Dirac measure $\boldsymbol{\delta_x}$** of $x$ is defined to be the measure $\delta_x : 2^X \to \overline{\mathbb{R}}_{\geqslant 0}$ satisfying $\delta_x(A) = 1$ if $x \in A$, and $\delta_x(A) = 0$ if $x \notin A$.

**Definition 1.41.** Let $(X, \mathcal{T}_X)$ be a topological space. Let $\mathfrak{M} \subset 2^X$ be a $\sigma$-algebra containing the Borel $\sigma$-algebra $\mathfrak{B}_X$. Let $\mu : \mathfrak{M} \to \overline{\mathbb{R}}_{\geqslant 0}$ be a measure. Assume that one of the following conditions holds:

(1) $X$ is second countable.

(2) $X$ is LCH, and $\mu|_{\mathfrak{B}_X}$ is a Radon measure.

The **support** $\mathbf{Supp}(\boldsymbol{\mu})$ is defined to be

$$\mathrm{Supp}(\mu) = \{x \in X : \mu(U) > 0 \text{ for each } U \in \mathrm{Nbh}_X(x)\}$$

Then $\mathrm{Supp}(\mu)$ is a closed subset of $X$, because we clearly have

$$X \backslash \mathrm{Supp}(\mu) = \bigcup_{U \in \mathcal{T}_X, \mu(U)=0} U$$

Moreover, we have $\mu(X\backslash\mathrm{Supp}(\mu)) = 0$. Thus, $\mathrm{Supp}(\mu)$ is the largest closed subset whose complement is $\mu$-null.

*Proof that $X\backslash\mathrm{Supp}(\mu)$ is null.* It suffices to show that if a family of open subsets $(U_\alpha)_{\alpha\in\mathscr{A}}$ is null, then the union $U := \bigcup_\alpha U_\alpha$ is null.

Assume that condition (1) holds. Since any subset of a second countable space is second countable and hence Lindelöf, the set $U$ is Lindelöf. So $(U_\alpha)$ has a countable subfamily covering $U$. Therefore, by the countable additivity, $U$ is null.

Assume that condition (2) holds. Since Radon measures are inner regular on open sets (cf. Def. 1.53), $\mu(U)$ is the supremum of $\mu(K)$ where $K$ runs through all compact subsets of $U$. Since $K$ is compact, $(U_\alpha)$ has a finite subfamily covering $K$. Therefore $K$ is null, and hence $U$ is null. □

**Lemma 1.42.** *Let $\mu : \mathfrak{M} \to \overline{\mathbb{R}}_{\geqslant 0}$ be as in Def. 1.40, and assume that Condition (1) or (2) of Def. 1.40 holds. The following are equivalent:*

*(a) $\mathrm{Supp}(\mu)$ is a finite set.*

*(b) $\mu$ is a linear combination of Dirac measures (restricted to $\mathfrak{M}$).*

*Proof.* (b)⇒(a): This is obvious.

(a)⇒(b): Write $E = \mathrm{Supp}(\mu)$. Choose any measurable $f : X \to \overline{\mathbb{R}}_{\geqslant 0}$. Then, since $\mu|_{X\backslash E} = 0$, the integral of any measurable function $g : X \to \overline{\mathbb{R}}_{\geqslant 0}$ vanishing ourside $E$ is zero. In particular, we can choose $g$ to be the unique one such that $g + \sum_{x\in E} f(x)\chi_{\{x\}} = f$. Therefore

$$\int_X f d\mu = \int_E \sum_{x\in E} f(x)\chi_{\{x\}} d\mu = \sum_{x\in E} f(x)\cdot\mu(\{x\})$$

This shows that $\mu = \sum_{x\in E} \mu(\{x\})\delta_x$. □

### 1.4.2 Radon-Nikodym derivatives

Fix a measurable space $(X, \mathfrak{M})$.

**Definition 1.43.** Let $\mu, \nu : \mathfrak{M} \to [0, +\infty]$ are measures. We say that $\nu$ is **absolutely continuous** with respect to $\mu$ and write $\boldsymbol{\nu \ll \mu}$ if any $\mu$-null set is $\nu$-null. We say that $h \in \mathcal{L}(X, \overline{\mathbb{R}}_{\geqslant 0})$ is a **Radon-Nikodym derivative** of $\nu$ with respect to $\mu$ if

$$\int_X f d\nu = \int_X fhd\mu \qquad \text{for all } f \in \mathcal{L}(X, \overline{\mathbb{R}}_{\geqslant 0})$$

By MCT, the above condition is equivalent to

$$\nu(E) = \int_E hd\mu \qquad \text{for all } E \in \mathfrak{M}$$

We write $\boldsymbol{d\nu = hd\mu}$.

**Remark 1.44.** If $\mu$ is $\sigma$-finite, and if $h_1, h_2$ are both Radon-Nikodym derivatives of $\nu$ with respect to $\mu$, then $h_1(x) = h_2(x)$ for $\mu$-a.e. $x \in X$.

*Proof.* It suffices to assume that $\mu(X) < +\infty$. For each $k \in \mathbb{N}$, let

$$A_k = \{x \in X : h_1(x) < h_2(x) \text{ and } h_1(x) \leqslant k\}$$

Then $\int_{A_k} h_1 d\mu \leqslant k\mu(X) < +\infty$, and

$$\int_{A_k} h_1 d\mu = \int_{A_k} d\nu = \int_{A_k} h_2 d\mu$$

Taking subtraction, we get $\int_{A_k}(h_2 - h_1)d\mu = 0$. Let $A = \bigcup_k A_k = \{x \in X : h_1(x) < h_2(x)\}$. By MCT, $\int_A (h_2 - h_1)d\mu = 0$. Since $h_2 - h_1 \geqslant 0$ on $A$, we conclude $h_2 - h_1 = 0$ $\mu$-a.e. on $A$, and hence $\mu(A) = 0$. Similarly, $\mu(B) = 0$ where $B = \{x \in X : h_1(x) > h_2(x)\}$. □

**Remark 1.45.** If $\nu$ is $\sigma$-finite, and if $d\nu = hd\mu$, then $h(x) < +\infty$ for $\mu$-a.e. $x \in X$.

*Proof.* Let $A = \{x \in A : h(x) = +\infty\}$. Since $\nu$ is $\sigma$-finite, we can write $A = \bigcup_{k \in \mathbb{N}} A_k$ where $A_k \in \mathfrak{M}$ and $\nu(A_k) < +\infty$. Since $\nu(A_k) = \int_{A_k} hd\mu = +\infty\mu(A_k)$, we have $\mu(A_k) = 0$, and hence $\mu(A) = 0$. □

**Theorem 1.46 (Radon-Nikodym theorem).** *Assume that $\mu, \nu : \mathfrak{M} \to [0, +\infty]$ are $\sigma$-finite measures. Then $\nu \ll \mu$ iff $\nu$ has a Radon-Nikodym derivative with respect to $\mu$.*

*Proof.* "$\Leftarrow$" is obvious. Let us prove "$\Rightarrow$". It is easy to reduce to the case that $\mu(X), \nu(X) < +\infty$. Let $d\psi = d\mu + d\nu$. So $\mu, \nu \leqslant \psi$. Therefore, the linear functional

$$\Lambda : L^2(X, \psi) \to \mathbb{C} \qquad \xi \mapsto \int_X \xi d\mu$$

is bounded. Since $L^2(X,\psi)$ is a Hilbert space (Thm. 1.48), by the Riesz-Fréchet theorem, there exists $f \in L^2(X,\psi)$ such that $\int_X \xi d\nu = \int_X \xi f d\psi$ for all $\xi \in L^2(X,\psi)$. Since $\Lambda$ sends positive functions to $\mathbb{R}_{\geqslant 0}$, after adding an $\psi$-a.e. function to $\xi$, we have $\psi \geqslant 0$ everywhere.

We have found $f \in \mathcal{L}(X,\mathbb{R}_{\geqslant 0})$ such that $d\mu = f d\psi$. Similarly, we have $g \in \mathcal{L}(X,\mathbb{R}_{\geqslant 0})$ such that $d\nu = g d\psi$. Since $\mu \leqslant \psi \ll \mu$, we have $f > 0$ outside a $\psi$-null set $\Delta$. Let $h = g/f$ outside $\Delta$, and $h = 0$ on $\Delta$. Then $d\nu = h d\mu$. □

### 1.4.3 $L^p$-spaces

Let $(X,\mathfrak{M},\mu)$ be a measure space. Let $1 \leqslant p,q \leqslant +\infty$ such that $p^{-1}+q^{-1}=1$.

**Theorem 1.47.** *Let $1 \leqslant p < +\infty$. Then the set of integrable $\mathbb{F}$-valued simple functions is dense in $L^p(X,\mu,\mathbb{F})$. In other words,*

$$\{\chi_E : E \subset \mathfrak{M}, \mu(E) < +\infty\}$$

*spans a dense subspace of $L^p(X,\mu,\mathbb{F})$.*

*Proof.* See [Gui-A, Sec. 27.2]. □

**Theorem 1.48 (Riesz-Fischer theorem**, the modern form). *The normed vector space $L^p(X,\mu,\mathbb{F})$ is (Cauchy) complete. Moreover, any Cauchy sequence in $L^p(X,\mu,\mathbb{F})$ has a subseqence converging $\mu$-a.e..*

*Proof.* See [Gui-A, Sec. 27.3]. □

**Lemma 1.49.** *Assume that $(X,\mu)$ is $\sigma$-finite. Let $\mathcal{S}_+$ be the set of simple functions $X \to \mathbb{R}_{\geqslant 0}$. Then for each $f \in \mathcal{L}(X,\overline{\mathbb{R}}_{\geqslant 0})$ we have*

$$\|f\|_{L^p(X,\mu)} = \sup\Big\{\int_X fg d\mu : g \in \mathcal{S}_+, \|g\|_{L^q(X,\mu)} \leqslant 1\Big\} \tag{1.11}$$

*Consequently, for each $f \in \mathcal{L}(X,\mathbb{C})$ we have*

$$\|f\|_{L^p(X,\mu)} = \sup\Big\{\int_X |fg| : g \in L^q(X,\mu), \|g\|_q \leqslant 1\Big\} \tag{1.12}$$

*Proof.* By Hölder's inequality, we have "$\geqslant$". To prove "$\leqslant$", we note that (1.12) follows immediately from (1.11) by writing $f = u|f|$ where $u \in \mathcal{L}(X,\mathbb{S}^1)$ and applying (1.11) to $|f|$. Thus, in the following, we assume $f \in \mathcal{L}(X,\overline{\mathbb{R}}_{\geqslant 0})$.

Case $1 < p < +\infty$: Choose an increasing sequence $(f_n)$ (i.e. $f_1 \leqslant f_2 \leqslant \cdots$) in $\mathcal{S}_+$ converging pointwise to $f$ such that each $f_n$ vanishes outside a measurable $\mu$-finite set. Let $g_n = (f_n)^{p-1}$. After removing the first several terms, we assume $\|g_n\|_{L^q} > 0$ for all $n$. Then

$$0 < \|g_n\|_q = \|f_n\|_p^{p/q} < +\infty$$

By MCT, we have $\lim_n \|g_n\|_p = \|f\|_p^{p/q}$ and $\lim_n \int_X fg_n = \|f\|_p^p$. Thus, if $\|f\|_p < +\infty$, then

$$\lim_n \|g_n\|_q^{-1} \int_X fg_n = \|f\|_p^{-p/q} \cdot \|f\|_p^p = \|f\|_p$$

This proves (1.11) when $\|f\|_p < +\infty$. If $\|f\|_p = +\infty$, then, by MCT, $\|f_n\|_p < +\infty$ can be sufficiently large. Applying (1.11) to $f_n$, we obtain $g \in \mathcal{S}_+$ such that $\|g\|_q \leqslant 1$ and $\int f_n g$ is sufficiently large, and hence $\int fg$ is sufficiently large. Thus (1.11) holds again.

Case $p = 1$: Let $g = 1$.

Case $p = +\infty$: Write $X = \bigcup_{n\in\mathbb{N}} \Omega_n$ where $\Omega_n \in \mathfrak{M}$ and $\mu(\Omega_n) < +\infty$. Choose any $0 \leqslant \lambda < \|f\|_\infty$. Then $A := \{|f| > \lambda\}$ satisfies $\mu(A) > 0$. Thus, there exists $n$ such that $0 < \mu(A \cap \Omega_n) < +\infty$. Let $g = \chi_{A\cap\Omega_n}/\mu(A \cap \Omega_n)$. Then $g \in \mathcal{S}_+$, $\|g\|_1 = 1$, and $\int fg \geqslant \lambda$. This proves (1.11). $\square$

**Theorem 1.50.** *Assume that $(X, \mu)$ is $\sigma$-finite. Assume $1 < p \leqslant +\infty$. Then we have an isomorphism of normed vector spaces*

$$\Psi : L^p(X, \mu, \mathbb{F}) \to L^q(X, \mu, \mathbb{F})^* \qquad f \mapsto \Big(g \in L^q(X, \mu, \mathbb{F}) \mapsto \int_X fgd\mu\Big) \tag{1.13}$$

When $p < +\infty$, the assumption on $\sigma$-finiteness can be removed. See [Fol-R, Sec. 6.2]. When $p = 2$, this is simply due to the completeness of $L^2(X, \mu, \mathbb{F})$ and the Riesz-Fréchet theorem.

*Proof.* By Hölder's inequality and Lem. 1.49, $\Psi$ is an isometry. Let us show that any $\Lambda \in L^q(X, \mu, \mathbb{F})^*$ belongs to the range of $\Psi$.

Step 1. By considering the real and imaginary parts, we can first assume that $\Lambda$ is real, i.e., $\Lambda(f) \in \mathbb{R}$ for any $f \in L^q(X, \mu, \mathbb{R}_{\geqslant 0})$.

Let us define $\mathbb{R}_{\geqslant 0}$-linear maps $\Lambda^+, \Lambda^- : L^q(X, \mu, \mathbb{R}_{\geqslant 0}) \to \mathbb{R}_{\geqslant 0}$ with operator norms $\leqslant \|\Lambda\|$, i.e.,

$$\|\Lambda^\pm(g)\| \leqslant \|\Lambda\| \cdot \|g\|_q \qquad \text{for all } g \in L^q(X, \mu, \mathbb{R}_{\geqslant 0}) \tag{1.14}$$

and let us check that

$$\Lambda(g) = \Lambda^+(g) - \Lambda^-(g) \qquad \text{for all } g \in L^q(X, \mu, \mathbb{R}_{\geqslant 0}) \tag{1.15}$$

Eq. (1.15) is called the **Jordan decomposition** of $\Lambda$.

Define the $\Lambda^\pm : L^q(X, \mu, \mathbb{R}_{\geqslant 0}) \to \overline{\mathbb{R}}$ by sending each $g \in L^q(X, \mu, \mathbb{R}_{\geqslant 0})$ to

$$\Lambda^\pm(g) = \sup\{\pm\Lambda(h) : h \in L^q(X, \mu, \mathbb{R}_{\geqslant 0}), h \leqslant g\} \tag{1.16}$$

Since $0 \leqslant g$, we clearly have $\Lambda^+(g) \geqslant 0$. Since $\Lambda$ is bounded and $\|h\|_q \leqslant \|g\|_q$, we clearly have $\|\Lambda^+(g)\| \leqslant \|\Lambda\| \cdot \|g\|_q$. In particular, $\Lambda^+$ has range in $\mathbb{R}_{\geqslant 0}$. Since $\Lambda^\pm = (-\Lambda)^\mp$, a similar property holds for $\Lambda^-$. Thus, we have checked (1.14).

Clearly, for each $f, g \in L^1(X, \mu, \mathbb{R}_{\geqslant 0})$, we have $\Lambda^+(f+g) \geqslant \Lambda^+(f) + \Lambda^+(g)$. To prove the other direction, choose any $h \in L^q(X, \mu, \mathbb{R}_{\geqslant 0})$ such that $h \leqslant f + g$. Let $h_1 = fh/(f+g)$ and $h_2 = gh/(f+g)$, understood to be zero where the denominator vanishes. Then $h_1, h_2 \in L^1(X, \mu, \mathbb{R}_{\geqslant 0})$ and $h_1 \leqslant f$ and $h_2 \leqslant g$. This proves $\Lambda^+(f + g) \leqslant \Lambda^+(f) + \Lambda^+(g)$. Thus $\Lambda^+$ (and similarly $\Lambda^-$) is $\mathbb{R}_{\geqslant 0}$-linear.

From (1.16), one easily checks $\Lambda(g) + \Lambda^-(g) \leqslant \Lambda^+(g)$ for each $f \in L^q(X, \mu, \mathbb{R}_{\geqslant 0})$. Replacing $\Lambda$ with $-\Lambda$, we get $-\Lambda(g) + \Lambda^+(g) \leqslant \Lambda^-(g)$. Thus (1.15) holds.

Step 2. Let us prove that $\Lambda^+$ is represented by some $f^+ \in L^p(X, \mu, \mathbb{R}_{\geqslant 0})$, namely,

$$\Lambda^+(g) = \int_X f^+ g d\mu \qquad \text{for all } g \in L^q(X, \mu, \mathbb{R}_{\geqslant 0}) \tag{1.17}$$

Then, similarly, $\Lambda^-$ is represented by some $f^- \in L^p(X, \mu, \mathbb{R}_{\geqslant 0})$. Thus $\Lambda$ is represented by $f^+ - f^-$, finishing the proof.

Write $X = \bigsqcup_n X_n$ where $\mu(X_n) < +\infty$. Suppose that we can find $f_n^+ \in L^p(X_n, \mu)$ representing $\Lambda^+|_{L^q(X_n,\mu)}$, then we can define $f^+ : X \to \mathbb{R}_{\geqslant 0}$ such that $f^+|_{X_n} = f_n$ for all $n$. Clearly $f^+$ represents $\Lambda^+$. In particular, by Lem. 1.49 and (1.14), $\|f^+\|_p \leqslant \|\Lambda\| < +\infty$. Thus $f \in L^p(X, \mu)$.

Therefore, according to the previous paragraph, we may assume at the beginning that $\mu(X) < +\infty$. Define

$$\nu : \mathfrak{M} \to [0, +\infty] \qquad E \mapsto \Lambda(\chi_E)$$

Then one checks easily that $\nu$ is a measure,[1] and that $\nu \ll \mu$. Therefore, by the Radon-Nikodym Thm. 1.46, there exists $f^+ \in \mathcal{L}(X, \mathbb{R}_{\geqslant 0})$ such that $d\nu = f^+ d\mu$. Thus

$$\Lambda^+(g) = \int_X g d\nu = \int_X f^+ g d\mu \quad \text{for each simple function } g \in L^q(X, \mu, \mathbb{R}_{\geqslant 0}) \tag{1.18}$$

Lem. 1.49 and (1.14) then imply $\|f^+\|_p \leqslant \|\Lambda\| < +\infty$, and hence $f \in L^p(X, \mu, \mathbb{R}_{\geqslant 0})$.

Finally, for $g \in L^q(X, \mu, \mathbb{R}_{\geqslant 0})$, find an increasing sequence of simple functions $g_n \in L^q(X, \mu, \mathbb{R}_{\geqslant 0})$ converging pointwise to $g$. By (1.14), $\Lambda^+(g - g_n) \leqslant \|\Lambda\| \cdot \|g - g_n\|_q$ where the RHS converges to zero by DCT. By MCT, $\int_X f^+ g_n d\mu \to \int_X f^+ g d\mu$. Thus, by (1.18), we conclude (1.17). $\square$

## 1.5 Review of measure theory: Radon measures

### 1.5.1 Radon measures and the Riesz-Markov representation theorem

Let $X$ be LCH. The reference for this subsection is [Gui-A, Ch. 25].

[1] To check the countable additivity, we let $E_1 \subset E_2 \subset \cdots$ be measurable and $E = \bigcup_n E_n$. Let $F_n = E \backslash E_n$. By (1.14), $\nu(F_n) \leqslant \|\Lambda\| \mu(F_n)^{\frac{1}{q}} \to 0$. Thus $\nu(E_n) \to \nu(E)$.

**Definition 1.51.** Let $\mathfrak{M} \subset 2^X$ be a $\sigma$ algebra containing $\mathfrak{B}_X$, and let $\mu : \mathfrak{M} \to \overline{\mathbb{R}}_{\geqslant 0}$ be a measure. Let $E \in \mathfrak{M}$. We say that $\mu$ is **outer regular** on $E$ if

$$\mu(E) = \inf\{\mu(U) : U \supset E, U \text{ is open}\}$$

We say that $\mu$ is **inner regular** on $E$ if

$$\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ is compact}\}$$

We say that $\mu$ is **regular** on $E$ if $\mu$ is both outer and inner regular on $E$.

**Lemma 1.52.** *Let $\mu : \mathfrak{B}_X \to \overline{\mathbb{R}}_{\geqslant 0}$ be a Borel measure. Let $U \subset X$ be open. Then*

$$\sup\{\mu(K) : K \subset U, K \text{ is compact}\} = \sup\left\{\int_X f d\mu : f \in C_c(U, [0,1])\right\}$$

*Therefore, $\mu$ is inner regular on $U$ iff*

$$\mu(U) = \sup\left\{\int_X f d\mu : f \in C_c(U, [0,1])\right\}$$

*Proof.* Let $A, B$ denote the LHS and the RHS. If $f \in C_c(U, [0,1])$, then setting $K = \mathrm{Supp}(f)$, we have $\mu(K) = \int_X \chi_K d\mu \geqslant \int_X f d\mu$. This proves $A \geqslant B$.

Conversely, let $K \subset U$. By Urysohn's lemma, there exists $f \in C_c(U, [0,1])$ such that $f|_K = 1$. So $\mu(K) = \int_X \chi_K d\mu \leqslant \int_X f d\mu$. This proves $A \leqslant B$. □

**Definition 1.53.** A Borel measure $\mu : \mathfrak{B}_X \to \overline{\mathbb{R}}_{\geqslant 0}$ is called a **Radon measure** if the following conditions are satisfied:

(a) $\mu$ is outer regular on Borel sets.

(b) $\mu$ is inner regular on open sets. Equivalently, for each open $U \subset X$, we have

$$\mu(U) = \sup\left\{\int_X f d\mu : f \in C_c(U, [0,1])\right\} \tag{1.19}$$

(c) $\mu(K) < +\infty$ if $K$ is a compact subset of $X$. Equivalently, for each $f \in C_c(X, \mathbb{R}_{\geqslant 0})$ we have

$$\int_X f d\mu < +\infty \tag{1.20}$$

*Proof of equivalence.* The equivalence in (b) is due to Lem. 1.52. The equivalence in (c) can be proved in a similar way to Lem. 1.52. □

**Remark 1.54.** There exist canonical bijections among:

- $\mathbb{R}_{\geqslant 0}$-linear maps $C_c(X, \mathbb{R}_{\geqslant 0}) \to \mathbb{R}_{\geqslant 0}$
- Positive linear functionals on $C_c(X, \mathbb{R})$.
- Positive linear functionals on $C_c(X) = C_c(X, \mathbb{C})$.

*Proof.* An $\mathbb{R}_{\geqslant 0}$-linear map $\Lambda : C_c(X, \mathbb{R}_{\geqslant 0}) \to \mathbb{R}_{\geqslant 0}$ can be extended uniquely to a linear map $\Lambda : C_c(X, \mathbb{R}) \to \mathbb{R}$ due to the following Lem. 1.55. The latter can be extended to a linear functional on $C_c(X)$ by setting $\Lambda(f) = \Lambda(\mathrm{Re} f) + \mathbf{i}\Lambda(\mathrm{Im} f)$ for all $C_c(X)$. $\square$

**Lemma 1.55.** *Let $K$ be an $\mathbb{R}_{\geqslant 0}$-linear subspace of an $\mathbb{R}$-vector space $V$. Let $W$ be an $\mathbb{R}$-linear space. Let $\Gamma : K \to W$ be an $\mathbb{R}_{\geqslant 0}$-linear map. Suppose that $V = \mathrm{Span}_{\mathbb{R}} K$. Then $\Gamma$ can be extended uniquely to an $\mathbb{R}$-linear map $\Lambda : V \to W$.*

*Proof.* The uniqueness is obvious. To prove the existence, note that any $v \in V$ can be written as

$$v = v^+ - v^-$$

where $v^+, v^- \in K$. (Proof: Since $V = \mathrm{Span}_{\mathbb{R}} K$, we have $v = a_1 u_1 + \cdots + a_m u_m - b_1 w_1 - \cdots - b_n w_n$ where each $u_i, w_j$ are in $K$, and each $a_i, b_j$ are in $\mathbb{R}_{\geqslant 0}$. One sets $v^+ = \sum_i a_i u_i$ and $v^- = \sum_j b_j w_j$.) We then define $\Lambda(v) = \Gamma(v^+) - \Gamma(v^-)$.

Let us show that this gives a well-defined map $\Lambda : V \to W$. Assume that $v = w^+ - w^-$ where $w^+, w^- \in K$. Then $\Gamma(v^+) - \Gamma(v^-) = \Gamma(w^+) - \Gamma(w^-)$ iff $\Gamma(v^+) + \Gamma(w^-) = \Gamma(v^-) + \Gamma(w^+)$, iff (by the additivity of $\Gamma$) $\Gamma(v^+ + w^-) = \Gamma(v^- + w^+)$. The last statement is true because $v^+ - v^- = w^+ - w^-$ implies $v^+ + w^- = v^- + w^+$.

It is easy to see that $\Lambda$ is additive. If $c \geqslant 0$, then $cv = cv^+ - cv^-$ where $cv^+, cv^- \in K$. So $\Lambda(cv) = \Gamma(cv^+) - \Gamma(cv^-)$, which (by the $\mathbb{R}_{\geqslant 0}$-linearity of $\Gamma$) equals $c\Gamma(v^+) - c\Gamma(v^-) = c\Lambda(v)$. Since $-v = v^- - v^+$, we have $\Lambda(-v) = \Gamma(v^-) - \Gamma(v^+) = -\Lambda(v)$. Hence $\Lambda(-cv) = c\Lambda(-v) = -c\Lambda(v)$. This proves that $\Lambda$ commutes with the $\mathbb{R}$-multiplication. $\square$

**Theorem 1.56 (Riesz-Markov representation theorem).** *For every positive linear $\Lambda : C_c(X, \mathbb{F}) \to \mathbb{F}$ there exists a unique Radon measure $\mu : \mathfrak{B}_X \to \overline{\mathbb{R}}_{\geqslant 0}$ such that*

$$\Lambda(f) = \int_X f d\mu \tag{1.21}$$

*for all $f \in C_c(X, \mathbb{F})$. Moreover, every Radon measure on $X$ arises from some $\Lambda$ in this way.*

*In addition, the operator norm $\|\Lambda\|$ equals $\mu(X)$. Therefore, $\Lambda$ is bounded iff $\mu$ is a finite measure.*

*Proof.* See [Gui-A, Sec. 25.3] for the first paragraph. The second paragraph asserts that

$$\sup_{f \in \overline{B}_{C_c(X)}(0,1)} |\Lambda(f)| = \mu(X)$$

The inequality "$\leqslant$" is obvious. The reverse inequality "$\geqslant$" follows from (1.19). $\square$

### 1.5.2 Basic properties of Radon measures

**Theorem 1.57.** *Let $\mu$ be a Radon measure (or its completion) on $X$. Then $\mu$ is regular on any measurable set $E$ satisfying $\mu(E) < +\infty$.*

*Proof.* See [Gui-A, Sec. 25.4]. A sketch of the proof (different from that in [Gui-A]) is as follows.

Assume WLOG that $E$ is Borel. Since Radon measures are outer regular on Borel sets, it remains to prove that $\mu$ is inner regular on $E$. Pick an open set $\mu(U)$ such that $\mu(U\backslash E)$ is small. Since $\mu$ is inner regular on $U$, there is a compact $K \subset U$ such that $\mu(U\backslash K)$ is small. However, $K$ is not necessarily contained in $E$.

To fix this issue, we note that since $\mu$ is outer regular on $U\backslash E$, we can find an open set $V \subset U$ containing $U\backslash E$ whose measure is close to $\mu(U\backslash E)$. In particular, $\mu(V)$ is small. Then $K\backslash V$ is a compact subset of $E$ whose measure is close to $\mu(E)$. □

**Theorem 1.58.** *Assume that $X$ is second countable. Let $\mu$ be a Borel measure on $X$. Then $\mu$ is Radon iff $\mu(K) < +\infty$ for any compact $K \subset X$.*

In particular, a finite Borel measure on $\mathbb{R}^n$ (where $n \in \mathbb{N}$) is Radon.

*Proof.* See [Gui-A, Sec. 25.5]. □

### 1.5.3 Approximation and density

The main reference for this subsection is [Gui-A, Sec. 27.2].

**Theorem 1.59 (Lusin's theorem).** *Let $X$ be LCH. Let $\mu$ be a Radon measure (or its completion) on $X$ with $\sigma$-algebra $\mathfrak{M}$. Let $f : X \to \mathbb{F}$ be measurable. Let $A \in \mathfrak{M}$ such that $\mu(A) < +\infty$. Then for each $\varepsilon > 0$ there exists a compact $K \subset A$ such that $\mu(A\backslash K) < \varepsilon$ and that $f|_K : K \to \mathbb{F}$ is continuous.*

With the help of the Tietze extension Thm. 1.22, Lusin's theorem implies that for each $\varepsilon > 0$ there exist a compact $K \subset A$ and some $\widetilde{f} \in C_c(X, \mathbb{F})$ such that $\widetilde{f}|_K = f|_K$ and $\mu(A\backslash K) < \varepsilon$.

*Proof.* See [Gui-A, Sec. 25.4]. □

**Theorem 1.60.** *Let $1 \leqslant p < +\infty$. Let $\mu$ be a Radon measure (or its completion) on an LCH space $X$. Then, under the $L^p$-norm, the space $C_c(X, \mathbb{F})$ is dense in $L^p(X, \mu, \mathbb{F})$. More precisely, the map $f \in C_c(X, \mathbb{F}) \mapsto f \in L^p(X, \mu, \mathbb{F})$ has dense range.*

*Proof.* See [Gui-A, Sec. 27.2]. □

**Remark 1.61.** One easily checks that

$$\begin{aligned}&\mathrm{Span}_{\mathbb{F}}\{\chi_I : I \subset \mathbb{R} \text{ is a bounded interval}\}\\ =&\mathrm{Span}_{\mathbb{F}}\{\chi_I : I \subset \mathbb{R} \text{ is a compact interval}\}\\ =&\mathrm{Span}_{\mathbb{F}}\{\chi_I : I \subset \mathbb{R} \text{ is a bounded open interval}\}\end{aligned}$$

An element in these sets is called an $\mathbb{F}$-valued **step function**. Moreover, one checks that

$$\begin{aligned}\{\text{right-continuous } \mathbb{F}\text{-valued step functions}\} &= \mathrm{Span}_{\mathbb{F}}\{\chi_{[a,b)} : a, b \in \mathbb{R}\}\\ \{\text{left-continuous } \mathbb{F}\text{-valued step functions}\} &= \mathrm{Span}_{\mathbb{F}}\{\chi_{(a,b]} : a, b \in \mathbb{R}\}\end{aligned}$$

**Theorem 1.62.** *Let $1 \leqslant p < +\infty$. Let $\mu$ be a Radon measure (or its completion) on $\mathbb{R}$. Then each of the following classes of functions form a dense subset of $L^p(\mathbb{R}, \mu, \mathbb{F})$:*

*(a) Right-continuous $\mathbb{F}$-valued step functions.*

*(b) Left-continuous $\mathbb{F}$-valued step functions.*

*(c) Elements of $\mathrm{Span}_{\mathbb{F}}\{\chi_{(-\infty,b]} : b \in \mathbb{R}\}$.*

*(d) Elements of $\mathrm{Span}_{\mathbb{F}}\{\chi_{(-\infty,b)} : b \in \mathbb{R}\}$.*

*Proof.* With the help of Thm. 1.60, the density of (a) and (b) can be proved by approximating a function $f \in C_c(X, \mathbb{F})$ with left/right-continuous step functions. See [Gui-A, Sec. 27.2] for details.

Since (a)$\subset$(c) and (b)$\subset$(d), the density of (c) and (d) follows. □

**Theorem 1.63.** *Let $1 \leqslant p < +\infty$. Let $\mu$ be a Radon measure (or its completion) on a second countable LCH space $X$. Then $L^p(X, \mu, \mathbb{F})$ is separable.*

*Proof.* See [Gui-A, Sec. 27.2]. □

### 1.5.4 Complex Radon measures

**Definition 1.64.** If $X$ is a set and $\mathfrak{M} \subset 2^X$ is a $\sigma$-algebra, a **complex measure** (resp. **signed measure**) is a function $\mathfrak{M} \to \mathbb{C}$ (resp. $\mathfrak{M} \to \mathbb{R}$) that can be written as a $\mathbb{C}$-linear (resp. $\mathbb{R}$-linear) combination of finite measures on $\mathfrak{M}$.

We now assume that $X$ is LCH.

**Definition 1.65.** A complex (resp. signed) measure on $\mathfrak{B}_X$ is called **Radon** if it is a $\mathbb{C}$-linear (resp. $\mathbb{R}$-linear) combination of finite Radon measures.

Suppose that $\mu$ is a complex Radon measure on $X$. Then similar to the proof of Rem. 1.54, for each $f \in C_0(X)$, we can extend the $\mathbb{R}_{\geqslant 0}$-linear functional $f \mapsto \int_X f d\mu$, where $\mu$ are finite Radon measures, to $\mu \mapsto \int_X f d\mu$ for all complex Radon measures $\mu$. This gives a $\mathbb{C}$-bilinear map

$$(f,\mu) \mapsto \int_X f d\mu \qquad \in \mathbb{C}$$

for $f \in C_0(X)$ and complex Radon measures $\mu$.

**Theorem 1.66 (Riesz-Markov representation theorem).** *Let $\mathbb{F} = \mathbb{C}$ (resp. $\mathbb{F} = \mathbb{R}$.) Then the elements of $C_c(X,\mathbb{F})^*$ are precisely linear functionals*

$$\Lambda : C_c(X,\mathbb{F}) \to \mathbb{F} \qquad f \mapsto \int_X f d\mu$$

*where $\mu$ is complex (resp. signed) Radon measure on $X$.*

*Proof.* It suffices to assume that $\Lambda$ is real, i.e., sending $C_c(X,\mathbb{R})$ into $\mathbb{R}$. Similar to the proof of Thm. 1.50, one writes $\Lambda = \Lambda^+ - \Lambda^-$ where $\Lambda^\pm$ are positive. Then apply Thm. 1.56 to $\Lambda^\pm$. See [Gui-A, Subsec. 25.10.2] for details. □

**Remark 1.67.** Since $C_c(X,\mathbb{F})$ is $l^\infty$-dense in $C_0(X,\mathbb{F})$, by Cor. 2.29, the dual spaces $C_c(X,\mathbb{F})^*$ and $C_0(X,\mathbb{F})^*$ are canonically identified. Therefore, Thm. 1.66 holds verbatim if $C_c(X,\mathbb{F})$ is replaced by $C_0(X,\mathbb{F})$.

## 1.6 Basic facts about increasing functions

### 1.6.1 Notation

If $I \subset \mathbb{R}$ is a proper interval, a function $\rho : I \to \mathbb{R}$ is called **increasing** if it is non-decreasing, i.e., $\rho(x) \leqslant \rho(y)$ whenever $x,y \in I$ and $x \leqslant y$. For each $t \in \mathbb{R}$, let

$$I_{\leqslant t} = I \cap (-\infty, t] \quad I_{<t} = I \cap (-\infty, t) \quad I_{\geqslant t} = I \cap [t, +\infty) \quad I_{>t} = I \cap (t, +\infty)$$

Suppose that $a = \inf I$ and $b = \sup I$. Let $\rho : I \to \mathbb{R}$ be increasing. If $x \in (a,b)$, then the left and right limits[2]

$$\rho(x^-) = \lim_{y \to x^-} \rho(y) \qquad \rho(x^+) = \lim_{y \to x^+} \rho(y) \tag{1.22}$$

exist, and

$$\rho(x^-) \leqslant \rho(x) \leqslant \rho(x^+)$$

[2]When taking the limit $\lim_{y \to x^\pm}$, we do not allow $y$ to be equal to $x$.

If $a \in I$, then $\rho(a^+)$ exists, and $\rho(a) \leqslant \rho(a^+)$. If $b \in I$, then $\rho(b^-)$ exists, and $\rho(b^-) \leqslant \rho(b)$. Let

$$\Omega_\rho = \{x \in (a,b) : \rho|_{(a,b)} \text{ is continuous at } x\}$$

Then for each $x \in (a,b)$, we have

$$x \in \Omega_\rho \quad \Leftrightarrow \quad \rho(x^-) = \rho(x^+) \quad \Leftrightarrow \quad \rho(x^-) = \rho(x) = \rho(x^+) \tag{1.23}$$

### 1.6.2 Basic properties of increasing functions

Let $I \subset \mathbb{R}$ be a proper interval with $a = \inf I, b = \sup I$.

**Proposition 1.68.** *If $\rho : I \to \mathbb{R}$ is increasing, then $I \backslash \Omega_\rho$ is countable.*

*Proof.* Replacing $\rho$ with $\arctan \circ \rho$, we may assume that $\rho$ is bounded. Let $C = \mathrm{diam}(\rho(I)) = \sup_{x,y \in I} |\rho(x) - \rho(y)|$. Let $A = (a,b) \backslash \Omega_\rho$. Then for each $B \in \mathrm{fin}(2^A)$, we have

$$\sum_{x \in B} (\rho(x^+) - \rho(x^-)) \leqslant C$$

Applying $\lim_B$, we get $\sum_{x \in A}(\rho(x^+) - \rho(x^-)) \leqslant C < +\infty$. Therefore $A$ is countable. $\square$

**Definition 1.69.** Let $\rho : I \to \mathbb{R}$. The **right-continuous normalization** of $\rho$ is the function $\widetilde{\rho} : I \to \mathbb{R}$ defined by

$$\widetilde{\rho}(x) = \begin{cases} \rho(x^+) & \text{if } x < b \\ \rho(b) & \text{if } x = b \end{cases}$$

The function $\widetilde{\rho}$ is clearly increasing and right-continuous. Moreover, $\widetilde{\rho}$ clearly agrees with $\rho$ on $\Omega_\rho$. Therefore, $\widetilde{\rho}$ and $\rho$ are almost equal, as defined by the following proposition.

**Proposition 1.70.** *Let $\rho_1, \rho_2 : I \to \mathbb{R}$ be increasing. Then the following are equivalent:*

*(a) There exists a dense subset $E \subset I$ such that $\rho_1|_E = \rho_2|_E$.*

*(b) $\Omega_{\rho_1} = \Omega_{\rho_2}$, and $\rho_1|_{\Omega_{\rho_1}} = \rho_2|_{\Omega_{\rho_2}}$.*

*(c) The right-continuous normalizations of $\rho_1$ and $\rho_2$ agree on $I_{<b}$.*

*If any of these statements are true, we say that $\rho_1, \rho_2$ are* ***almost equal****.*

*Proof.* (a)⇒(b): Assume (a). Choose any $x \in I$. If $x > a$ then

$$\rho_1(x^-) = \lim_{E \ni y \to x^-} \rho_1(y) = \lim_{E \ni y \to x^-} \rho_2(y) = \rho_2(x^-) \tag{1.24a}$$

Similarly, if $x < b$ then

$$\rho_1(x^+) = \rho_2(x^+) \tag{1.24b}$$

Thus (b) follows from (1.23).

(b)⇒(a): By Prop. 1.68, $E := (a,b) \cap \Omega_{\rho_1}$ is a dense subset of $(a,b)$.

(b)⇔(c): Let $\widetilde{\rho}_i$ be the right continuous normalization of $\rho_i$. Then by (a)⇒(b), we have $\Omega_{\rho_i} = \Omega_{\widetilde{\rho}_i}$ and $\rho_i|_{\Omega_{\rho_i}} = \widetilde{\rho}_i|_{\Omega_{\widetilde{\rho}_i}}$. Therefore, (b) holds iff

$$\Omega_{\widetilde{\rho}_1} = \Omega_{\widetilde{\rho}_2} \qquad \text{and} \qquad \widetilde{\rho}_1|_{\Omega_{\widetilde{\rho}_1}} = \widetilde{\rho}_2|_{\Omega_{\widetilde{\rho}_2}} \tag{1.25}$$

Clearly (c) implies (1.25). Suppose that (1.25) is true. Then for each $x \in I_{<b}$ we have

$$\widetilde{\rho}_1(x) = \widetilde{\rho}_1(x^+) \xlongequal{(1.24b)} \widetilde{\rho}_2(x^+) = \widetilde{\rho}_2(x)$$

Thus (1.25) implies (c). Therefore (b) and (c) are equivalent. □

## 1.7 The Stieltjes integral

### 1.7.1 Definitions and basic properties

In this subsection, we fix a proper interval $I \subset \mathbb{R}$, and let $\rho : I \to \mathbb{R}_{\geqslant 0}$ be an increasing function.

**Definition 1.71.** Let $J$ be any proper bounded interval. Let $a = \inf J, b = \sup J$. A **partition** of the interval $J$ is defined to be an element of the form

$$\sigma = \{a_0, a_1, \dots, a_n \in [a,b] : a_0 = a < a_1 < a_2 < \cdots < a_n = b, n \in \mathbb{Z}_+\} \tag{1.26}$$

The **mesh** of $\sigma$ is defined to be

$$\max\{a_i - a_{i=1} : i = 1, \dots, n\}$$

If $\sigma, \sigma' \in \mathrm{fin}(2^J)$ are partitions of $J$, we say that $\sigma'$ is a **refinement** of $\sigma$ (or that $\sigma'$ is **finer than** $\sigma$), if $\sigma \subset \sigma'$. In this case, we also write

$$\sigma \prec \sigma'$$

We define $\mathcal{P}(J)$ to be

$$\mathcal{P}(J) = \{\text{partitions of } J\}$$

**Remark 1.72.** If $\sigma, \sigma' \in \mathcal{P}(J)$, then clearly $\sigma \cup \sigma' \in \mathcal{P}(J)$ and $\sigma, \sigma' \prec \sigma \cup \sigma$. Therefore, $\prec$ is a partial order on $\mathcal{P}(J)$. We call $\sigma \cup \sigma'$ the **common refinement** of $\sigma$ and $\sigma'$.

**Definition 1.73.** A **tagged partition** of $I$ is an ordered pair

$$(\sigma, \xi_\bullet) = \big(\{a_0 = a < a_1 < \cdots < a_n = b\}, (\xi_1, \ldots, \xi_n)\big) \tag{1.27}$$

where $\sigma \in \mathcal{P}(J)$ and

$$\xi_i \in (a_{j-1}, a_j]$$

for all $1 \leqslant j \leqslant n$. The set

$$\mathcal{Q}(J) = \{\text{tagged partitions of } J\}$$

equipped with the preorder $\prec$ defined by

$$(\sigma, \xi_\bullet) \prec (\sigma', \xi'_\bullet) \qquad \Longleftrightarrow \qquad \sigma \subset \sigma' \tag{1.28}$$

is a directed set.

**Definition 1.74.** Let $V$ be a Banach space. Assume $[a, b] \subset I$ and $a < b$. Let $f \in C([a, b], V)$. For each $(\sigma, \xi_\bullet) \in \mathcal{Q}(I)$, define the **Stieltjes sum**

$$S_\rho(f, \sigma, \xi_\bullet) = \sum_{j \geqslant 1} f(\xi_j)\big(\rho(a_j) - \rho(a_{j-1})\big)$$

abbreviated to $S(f, \sigma, \xi_\bullet)$ when no confusion arises. The **Stieltjes integral** on $(a, b]$ is defined to be the limit of the net $(S_\rho(f, \sigma, \xi_\bullet))_{(\sigma, \xi_\bullet) \in \mathcal{Q}([a,b])}$:

$$\int_{(a,b]} f d\rho = \lim_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)} S_\rho(f, \sigma, \xi_\bullet) \tag{1.29}$$

The **Stieltjes integral** on $[a, b]$ is defined to be

$$\int_{[a,b]} f d\rho = f(a)\rho(a) + \int_{(a,b]} f d\rho \tag{1.30}$$

Note that when $f(a) \neq 0$, the integral $\int_{(a,b]} f d\rho$ depends not only on $\rho|_{(a,b]}$ but also on the value $\rho(a)$. On the other hand, it is clear that

$$\int_{(a,b]} f d\rho = \int_{(a,b]} f d\rho|_{[a,b]} \qquad \int_{[a,b]} f d\rho = \int_{[a,b]} f d\rho|_{[a,b]} \tag{1.31}$$

*Proof of the convergence of* (1.29). Since $f$ is uniformly continuous, for each $\varepsilon > 0$, there exists $\delta > 0$ such that $\|f(x) - f(y)\| \leqslant \varepsilon$ for all $x, y \in [a, b]$ and $|x - y| \leqslant \delta$. Choose any tagged partition $(\sigma, \xi_\bullet)$ of $[a, b]$ with mesh $\leqslant \delta$. Then one easily sees that for any $(\sigma', \xi'_\bullet) \succ (\sigma, \xi)$ we have

$$\|S(f, \sigma', \xi'_\bullet) - S(f, \sigma, \xi_\bullet)\| \leqslant \varepsilon(\rho(b) - \rho(a))$$

Therefore, the net $(S(f, \sigma, \xi_\bullet))_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)}$ is Cauchy. So it must converge because $V$ is complete. $\square$

**Remark 1.75.** The above proof implies the following useful fact: Let $f \in [a,b]$. Let $\varepsilon, \delta > 0$ such that $\|f(x) - f(y)\| \leqslant \varepsilon$ for all $x, y \in [a,b]$ satisfying $|x-y| \leqslant \delta$. Then for each tagged partition $(\sigma, \xi_\bullet)$ of $[a,b]$ with mesh $\leqslant \delta$, we have

$$\Big\| \int_{(a,b]} f d\rho - S_\rho(f, \sigma, \xi_\bullet) \Big\| \leqslant \varepsilon(\rho(b) - \rho(a)) \tag{1.32}$$

and hence

$$\Big\| \int_{[a,b]} f d\rho - f(a)\rho(a) - S_\rho(f, \sigma, \xi_\bullet) \Big\| \leqslant \varepsilon(\rho(b) - \rho(a)) \tag{1.33}$$

**Example 1.76.** The integrals of the constant function $1$ are

$$\int_{(a,b]} d\rho = \rho(b) - \rho(a) \qquad \int_{[a,b]} d\rho = \rho(b)$$

**Example 1.77.** Suppose that $\rho|_{(a,b]} = 1$. Then

$$\int_{(a,b]} f d\rho = f(a)(1 - \rho(a)) \qquad \int_{[a,b]} f d\rho = f(a)$$

In particular, if $\rho|_{[a,b]} = 1$, then $\displaystyle\int_{(a,b]} f d\rho = 0$ and $\displaystyle\int_{[a,b]} f d\rho = f(a)$.

**Remark 1.78.** It is easy to see that

$$\Lambda : C([a,b], V) \to V \qquad f \mapsto \int_{[a,b]} f d\rho$$

is linear. Moreover, since $\|S(f, \sigma, \xi_\bullet)\| \leqslant (\rho(b) - \rho(a))\|f\|_{l^\infty}$ and hence $\|f(a)\rho(a) + S(f, \sigma, \xi_\bullet)\| \leqslant \rho(b)\|f\|_{l^\infty}$, the operator norm $\|\Lambda\|$ satisfies $\|\Lambda\| \leqslant \rho(b)$, that is

$$\Big\| \int_{[a,b]} f d\rho \Big\| \leqslant \rho(b)\|f\|_{l^\infty} \qquad \text{for all } f \in C([a,b], V)$$

In particular, $\Lambda$ is bounded.

**Remark 1.79.** It is easy to check that $\rho \mapsto \int_{(a,b]} f d\rho$ and $\rho \mapsto \int_{[a,b]} f d\rho$ are $\mathbb{R}_{\geqslant 0}$-linear over increasing functions $\rho : [a,b] \to \mathbb{R}_{\geqslant 0}$. Moreover, if $c \in (a,b)$, one easily shows

$$\int_{(a,b]} f d\rho = \int_{(a,c]} f d\rho + \int_{(c,b]} f d\rho \tag{1.34}$$

by considering tagged partitions finer than $\{a, c, b\}$.

Exp. 1.77 suggests that the value of $\int_{[a,b]} f d\rho$ is independent of $\rho(a)$:

**Lemma 1.80.** *Suppose that $\rho_1, \rho_2 : [a, b] \to \mathbb{R}_{\geqslant 0}$ are increasing and satisfies $\rho_1|_{(a,b]} = \rho_2|_{(a,b]}$. Then for each $f \in C([a, b], V)$ we have $\displaystyle\int_{[a,b]} f d\rho_1 = \int_{[a,b]} f d\rho_2$.*

See Thm. 1.82 for a generalization of this lemma.

*Proof.* Assume WLOG that $\rho_1(a) \leqslant \rho_2(a)$. Let $\lambda = \rho_2(a) - \rho_1(a)$. Then $\rho_2 - \rho_1 = \lambda \cdot \chi_{\{a\}} = \lambda \cdot (1 - \chi_{(a,b]})$, and hence $\rho_1 + \lambda = \rho_2 + \lambda \cdot \chi_{(a,b]}$. By Rem. 1.79,

$$\int_{[a,b]} f d\rho_1 + \lambda \int_{[a,b]} f d1 = \int_{[a,b]} f d\rho_2 + \lambda \int_{[a,b]} f d\chi_{(a,b]}$$

By Exp. 1.79, we obtain $\int_{[a,b]} f d\rho_1 = \int_{[a,b]} f d\rho_2$. □

### 1.7.2 Dependence of the Stieltjes integral on $\rho$

Let $I \subset \mathbb{R}$ be a proper interval, and let $a = \inf I$ and $b = \sup I$. Note that $I$ is not assumed to be bounded.

**Definition 1.81.** For each $f \in C_c(I, V)$ and each increasing $\rho : I \to \mathbb{R}$, we can still define the **Stieltjes integral**

$$\int_I f d\rho := \int_J f d\rho$$

where $J$ is any compact sub-interval of $I$ containing $\mathrm{Supp}_I(f)$. The value of the integral is clearly independent of the choice of such $J$. Moreover, this definition is compatible with the definitions of $\int_{[a,b]} f d\rho$ and $\int_{(a,b]} f d\rho$ in Def. 1.74.

**Theorem 1.82.** *Let $\rho_1, \rho_2 : I \to \mathbb{R}_{\geqslant 0}$ be increasing functions satisfying the following condition:*

- *$\rho_1$ and $\rho_2$ are almost equal, and $\rho_1(b) = \rho_2(b)$ if $b \in I$. (By Prop. 1.70, this is equivalent to that $\rho_1, \rho_2$ have the same right-continuous normalization.)*

*Then for each $f \in C_c(I, V)$, we have*

$$\int_I f d\rho_1 = \int_I f d\rho_2$$

*Proof.* By Lem. 1.80, we may assume that $\rho_1(a) = \rho_2(a)$ if $a \in I$.

Fix $f \in C_c(I, V)$. Choose $\alpha, \beta \in \mathbb{R}$ satisfying $\mathrm{Supp}_I(f) \subset [\alpha, \beta] \subset I$. Due to the assumption on $\rho_1, \rho_2$, we may slightly enlarge the compact interval $J := [\alpha, \beta]$ so that

$$\rho_1(\alpha) = \rho_2(\alpha) \qquad \rho_1(\beta) = \rho_2(\beta)$$

(When $a \in I$ resp. $b \in I$, one can even set $\alpha = a$ resp. $\beta = b$.) Then $\int_I f d\rho_i = \int_J f d\rho_i$.

Let $C = \max\{\rho_i(\beta) - \rho_i(\alpha) : i = 1, 2\}$. Choose any $\varepsilon > 0$. Since $f$ is uniformly continuous, there exists $\delta > 0$ such that $|f(x) - f(y)| \leqslant \varepsilon$ whenever $x, y \in I$ and $|x - y| \leqslant \delta$. Choose a tagged partition

$$(\sigma, \xi_\bullet) = \big(\{a_0 = \alpha < a_1 < \cdots < a_n = \beta\}, (\xi_1, \ldots, \xi_n)\big)$$

of $J$ with mesh $< \delta$. Moreover, due to the assumption on $\rho_1, \rho_2$, by a slight adjustment, we may assume that $\rho_1(a_j) = \rho_2(a_j)$ for each $0 \leqslant j \leqslant n$. This implies

$$S_{\rho_1}(f, \sigma, \xi_\bullet) = S_{\rho_2}(f, \sigma, \xi_\bullet)$$

Therefore, by Rem. 1.75, we obtain

$$\Big\| \int_J f d\rho_1 - \int_J f d\rho_2 \Big\| \leqslant 2\varepsilon \cdot C$$

This completes the proof by choosing arbitrary $\varepsilon$. □

**Theorem 1.83.** *Let $\rho_1, \rho_2 : I \to \mathbb{R}_{\geqslant 0}$ be bounded increasing functions satisfying*

$$\lim_{x \to a^+} \rho_1(x) = \lim_{x \to a^+} \rho_2(x) = 0 \qquad \text{if } a \notin I \tag{1.35}$$

*Then the following are equivalent:*

*(1) $\rho_1$ and $\rho_2$ are almost equal, and $\rho_1(b) = \rho_2(b)$ if $b \in I$. (By Prop. 1.70, this is equivalent to that $\rho_1, \rho_2$ have the same right-continuous normalization.)*

*(2) For each $f \in C_c(I, \mathbb{F})$ we have*

$$\int_I f d\rho_1 = \int_I f d\rho_2$$

*Proof.* By Thm. 1.82, we have "(1)⇒(2)". Assume (2). Let us prove (1). Let $\widetilde{\rho}_i$ be the right-normalization of $\rho_i$. By "(1)⇒(2)", we have $\int_I f d\rho_i = \int_I f d\widetilde{\rho}_i$. Therefore, to prove (1), it suffices to assume that $\rho_1$ and $\rho_2$ are right-continuous on $I$.

We shall prove (1) by choosing an arbitrary bounded increasing right-continuous $\rho : I \to \mathbb{R}_{\geqslant 0}$, and show that for each $x \in I$, the value $\rho(x)$ can be recovered from the integrals $\int_I f d\rho$ where $f \in C_c(I, \mathbb{R})$.

Case 1: Assume $a \notin I$ and $a < x < b$. For each real numbers $v, y$ satisfying

$$a < v < x < y < b$$

choose $\varphi_{v,y} \in C_c(I, [0, 1])$ satisfying

$$\chi_{[v,x]} \leqslant \varphi_{v,y} \leqslant \chi_{(a,y]}$$

Choose $u \in (a,v)$ such that $\varphi_{v,y}$ vanishes outside $[u,y]$. Then by Rem. 1.79,

$$\int_I \varphi_{v,y} d\rho = \int_{[u,y]} \varphi_{v,y} d\rho = \varphi_{v,y}(u) + \int_{(u,v]} \varphi_{v,y} d\rho + \int_{(v,x]} \varphi_{v,y} d\rho + \int_{(x,y]} \varphi_{v,y} d\rho$$
$$= \int_{(u,v]} \varphi_{v,y} d\rho + \rho(x) - \rho(v) + \int_{(x,y]} \varphi_{v,y} d\rho$$

where Exp. 1.76 is used in the last equality. By Rem. 1.75, we have $\int_{(u,v]} \varphi_{v,y} d\rho \leqslant \rho(v) - \rho(u) \leqslant \rho(v)$ and $\int_{(x,y]} \varphi_{v,y} d\rho \leqslant (\rho(y) - \rho(x))$. Since $\rho$ is right-continuous and satisfies (1.35), we have

$$\lim_{v \searrow a^+} \rho(v) = \lim_{y \searrow x^+} (\rho(y) - \rho(x)) = 0$$

Therefore, the above calculation of $\int_I \varphi_{v,y} d\rho$ shows

$$\lim_{\substack{v \searrow a^+ \\ y \searrow x^+}} \int_I \varphi_{v,y} d\rho = \lim_{v \searrow a^+} (\rho(x) - \rho(v)) = \rho(x)$$

Case 2: Assume $a \in I$ and $a \leqslant x < b$. For each $y \in (x,b)$, choose $\varphi_y \in C_c(I,[0,1])$ such that $\chi_{[a,x]} \leqslant \varphi_y \leqslant \chi_{[a,y]}$. Similar to the argument in Case 1, one shows

$$\int_I \varphi_y d\rho = \int_{[a,x]} \varphi_y d\rho + \int_{(x,y]} \varphi_y d\rho = \rho(x) + \int_{(x,y]} \varphi_y d\rho$$

where Exp. 1.76 is used. By Rem. 1.75, $\int_{(x,y]} \varphi_y d\rho \leqslant \rho(y) - \rho(x)$. Therefore, the right-continuity of $\rho$ implies

$$\lim_{y \searrow x^+} \int_I \varphi_y d\rho = \rho(x)$$

Case 3: Assume $I = (a,b]$ and $x = b$. For each $v \in (a,x)$, choose $\varphi_v \in C_c(I,[0,1])$ such that $\chi_{[v,b]} \leqslant \varphi_v \leqslant \chi_I$. Similar to the argument above,

$$\lim_{v \searrow a^+} \int_I \varphi_v d\rho = \rho(b)$$

Case 4: Assume $I = [a,b]$ and $x = b$. Then $\displaystyle\int_I d\rho = \rho(b)$. □

**Remark 1.84.** The assumption (1.35) imposes little restriction. Indeed, suppose $a \notin I$. Then for each $f \in C_c(I)$, since there exists $v \in \mathbb{R}_{>a}$ such that $f$ vanishes on $(a,v]$, for any constant $\varkappa \in \mathbb{R}$ with $\rho + \varkappa \geqslant 0$, we clearly have

$$\int_I f d\rho = \int_I f d(\rho + \varkappa) \tag{1.36}$$

Therefore, when $a \notin I$, given any two increasing functions $\rho_1, \rho_2 : I \to \mathbb{R}_{\geqslant 0}$, we can freely add constants to $\rho_1$ and $\rho_2$ to ensure that (1.35) holds.

## 1.8 The Riesz representation theorem via the Stieltjes integral

In this section, we fix a proper interval $I \subset \mathbb{R}$, and let $a = \inf I$ and $b = \sup I$.

### 1.8.1 The positive case

**Theorem 1.85 (Riesz representation theorem).** *We have a bijection between:*

*(a) A bounded increasing right-continuous function $\rho : I \to \mathbb{R}_{\geqslant 0}$ satisfying $\lim_{x\to a^+} \rho(a) = 0$ if $a \notin I$.*

*(b) A bounded positive linear functional $\Lambda : C_c(I, \mathbb{F}) \to \mathbb{F}$.*

*$\Lambda$ is determined by $\rho$ by*

$$\Lambda : C(I, \mathbb{F}) \to \mathbb{F} \qquad f \mapsto \int_I f d\rho \tag{1.37}$$

*$\rho$ is determined by $\Lambda$ by*

$$\rho(x) = \mu(I_{\leqslant x}) \qquad \text{for all } x \in I \tag{1.38}$$

*where $\mu$ is the finite Borel measure on $I$ associated to $\Lambda$ as in the Riesz-Markov representation Thm. 1.56.*

Note that by Thm. 1.58, finite Borel measures on $I$ and finite Radon measures on $I$ coincide.

*Proof.* Step 1. Thm. 1.56 establishes the equivalence between a bounded positive linear $\Lambda$ and a finite Borel measure $\mu$. We let prove the equivalence between the radon measures $\mu$ and the functions $\rho$ satisfying (a).

More precisely, given a Radon measure $\mu$ on $I$, let $\rho_\mu : I \to \mathbb{R}_{\geqslant 0}$ be defined by (1.38), that is, for each $x \in I$ we have

$$\rho_\mu(x) = \mu(I_{\leqslant x}) \tag{1.39}$$

Then $\rho_\mu$ is clearly bounded and increasing. By DCT, $\rho_\mu$ is right-continuous, and we have $\lim_{x\to a^-} \rho(x) = 0$ when $a \notin I$. Therefore, $\rho_\mu$ satisfies (a).

Conversely, given any $\rho$ satisfying (a), let $\mu_\rho$ be the unique Radon measure corresponding to $\rho$ via (1.37), i.e., for each $f \in C_c(I, \mathbb{F})$ we have

$$\int_I f d\mu_\rho = \int_I f d\rho \tag{1.40}$$

By Rem. 1.78, the linear functional $f \in C_c(I, \mathbb{F}) \mapsto \int_I f d\rho$ is bounded with operator norm $\leqslant \sup_{x\in I} \rho(x)$. Thus, $\mu_\rho$ is a finite measure.

We want to show that $\Phi : \rho \mapsto \mu_\rho$ and $\Psi : \mu \mapsto \rho_\mu$ are inverses of each other. By Thm. 1.83, the map $\Phi$ is injective. Therefore, it suffices to prove that $\Phi \circ \Psi = \mathrm{id}$, i.e., that $\mu_{\rho_\mu} = \mu$. This means proving

$$\int_I f d\mu = \int_I f d\rho_\mu \tag{1.41}$$

for each $f \in C_c(I, \mathbb{F})$.

Step 2. Let us fix $f \in C_c(I, \mathbb{F})$ and prove (1.41). Choose $\alpha, \beta \in \mathbb{R}$ such that $J := [\alpha, \beta]$ is a sub-interval of $I$ containing $\mathrm{Supp}_I(f)$. Choose any $\varepsilon > 0$. Since $f$ is uniformly continuous, there exists $\delta > 0$ such that $|f(x) - f(y)| \leqslant \varepsilon$ whenever $x, y \in I$ and $|x - y| \leqslant \delta$. Choose a tagged partition

$$(\sigma, \xi_\bullet) = \big(\{a_0 = \alpha < a_1 < \cdots < a_n = \beta\}, (\xi_1, \ldots, \xi_n)\big)$$

of $J$ with mesh $\leqslant \delta$. By Rem. 1.75, we have

$$\Big|\int_J f d\rho_\mu - f(\alpha)\rho_\mu(\alpha) - S_{\rho_\mu}(f, \sigma, \xi_\bullet)\Big| \leqslant \varepsilon(\rho_\mu(\beta) - \rho_\mu(\alpha)) = \varepsilon \cdot \mu((\alpha, \beta]) \tag{1.42}$$

Also, we have $\|f - g\|_{l^\infty(I)} \leqslant \varepsilon$ where

$$g = f(\alpha)\chi_{\{a\}} + \sum_{i=1}^n f(\xi_i)\chi_{(a_{i-1}, a_i]}$$

By (1.38), we have

$$\mu(\{\alpha\}) = \rho_\mu(\alpha) - \mu(I_{<\alpha}) \qquad \mu((a_{i-1}, a_i]) = \rho_\mu(a_i) - \rho_\mu(a_{i-1})$$

Note that if $f(\alpha) \neq 0$, then by $\mathrm{Supp}_I(f) \subset [\alpha, \beta]$, we must have $\alpha = a \in I$ and hence $I_{<\alpha} = \emptyset$. Therefore, we must have

$$\int_I g d\mu = f(\alpha)\rho_\mu(\alpha) + S_{\rho_\mu}(f, \sigma, \xi_\bullet)$$

Combining this fact with $\|f - g\|_{l^\infty(I)} \leqslant \varepsilon$, we get

$$\Big|\int_I f d\mu - f(\alpha)\rho_\mu(\alpha) - S_{\rho_\mu}(f, \mu, \xi_\bullet)\Big| \leqslant \varepsilon \cdot \mu(J)$$

This inequality, together with (1.42), implies

$$\Big|\int_I f d\mu - \int_I f d\rho_\mu\Big| \leqslant 2\varepsilon \cdot \mu(J)$$

Since $\varepsilon$ is arbitrary, we conclude (1.41). $\square$

### 1.8.2 The general case

**Definition 1.86.** A real-valued function $I \to \mathbb{F}$ is called of **bounded variation** (or simply **BV**) if it is an $\mathbb{F}$-linear combination of bounded increasing functions $I \to \mathbb{R}_{\geqslant 0}$. The space of BV functions from $I$ to $\mathbb{F}$ is denoted by $BV(I,\mathbb{F})$.

**Remark 1.87.** By Rem. 1.78 and 1.79, we have an $\mathbb{R}_{\geqslant 0}$-bilinear functional

$$(f,\rho) \mapsto \int_I f d\rho \qquad \in \mathbb{R}_{\geqslant 0}$$

for $f \in C_c(I,\mathbb{R}_{\geqslant 0})$ and bounded increasing $\rho : I \to \mathbb{R}_{\geqslant 0}$. Similar to the proof of Rem. 1.54, it can be extended to a positive bililinear functional

$$C_c(I,\mathbb{C}) \times BV(I,\mathbb{C}) \to \mathbb{C} \qquad (f,\rho) \mapsto \int_I f d\rho$$

**Theorem 1.88 (Riesz representation theorem).** *The elements of the dual space $C_c(I,\mathbb{F})^*$ are precisely linear functionals of the form*

$$\Lambda : C(I,\mathbb{F}) \to \mathbb{F} \qquad f \mapsto \int_I f d\rho$$

*where $\rho \in BV(I,\mathbb{F})$. Moreover, the BV function $\rho$ can be chosen such that it is right-continuous on $I$, and that $\lim_{x\to a^+} \rho(x) = 0$ if $a \notin I$.*

*Proof.* This is immediate from Thm. 1.66 and 1.85. $\square$

# 2 Normed vector spaces and their dual spaces

## 2.1 The origin of dual spaces in the calculus of variations

Linear functional analysis treats function spaces as linear spaces with appropriate geometric/topological structures and analytic properties. In the foundational theory of functional analysis, two analytic properties are especially important: (Cauchy) completeness and duality. In this course, our focus is primarily on normed vector spaces $V$. For such spaces, Cauchy completeness is interpreted in the same way as in any metric space. Duality, on the other hand, refers to the natural identification of $V$ as the dual space $V^*$ of another normed vector space $U$.

Many early results in functional analysis were related to duality, while the significance of completeness was not immediately recognized. In fact, the history of functional analysis experienced a paradigm shift from the study of (scalar-valued) functionals to linear maps between vector spaces. Specifically, attention moved from continuous bilinear maps of the form $U \times V \to \mathbb{F}$ to the analysis of continuous linear maps $V \to W$, where $U, V, W$ are normed vector spaces. With this shift, completeness became increasingly central to modern analysis. See Sec. 2.5 for further illustrations.

The early part of this course will also focus more on dual spaces. If $V$ is a normed $\mathbb{F}$-vector space, then the **dual space** $V^* = \mathfrak{L}(V, \mathbb{F})$ is defined to be the space of bounded (i.e. continuous) linear maps $V \to \mathbb{F}$. One of the major themes in early functional analysis was the characterization of dual spaces of various function spaces under appropriate norms. Among the most notable results are F. Riesz's characterization of $C([a,b], \mathbb{R})^*$ (cf. Thm. 1.88) in [Rie09, Rie11], and his proof that $L^q([a,b], m, \mathbb{R})^* \simeq L^p([a,b], m, \mathbb{R})$ (cf. Thm. 1.50) in [Rie10]. These results highlight a profound connection between dual spaces and measure/integration theory. Nevertheless, the study of dual spaces originally arose from a somewhat different field: the calculus of variations in the 19th century.

Consider a nonlinear functional $S : f \mapsto S(f) \in \mathbb{R}$, for example, of the form

$$S(f) = \int_a^b L(f(t), f'(t), \ldots, f^{(r)}(t))dt$$

where $L$ is a "nice" real valued function with $r$-variables, and $f$ is defined on $[a,b]$. If we perturb $f$ slightly by a variation $\eta$, then the corresponding change in $S$ can be approximated by

$$\delta S[f, \eta] := S(f + \eta) - S(f) \approx \int_a^b \beta_f(t) \cdot \eta(t)dt \tag{2.1}$$

where $\beta_f : [a,b] \to \mathbb{R}$ is a function depending on $f$. This function should be interpreted loosely. In some cases, it may involve delta functions or similar objects that are not functions in the classical sense, but rather distributions:

**Example 2.1.** Consider the case where $L$ is smooth and $r = 1$, i.e.

$$S(f) = \int_a^b L(f(t), f'(t))dt$$

(For example, $L(x, y) = T(y) - V(x)$ where $T(y) = \frac{1}{2}my^2$ the kinetic energy for the mass $m \in \mathbb{R}_{>0}$, and $V(x)$ is the potential energy at $x$.) Then

$$\begin{aligned}\delta S[f, \eta] = &\int_a^b L(f + \eta, f' + \eta') \approx \int_a^b (\partial_x L(f, f')\eta + \partial_y L(f, \eta)\eta')\\ =&\partial_y L(f, f')\eta\Big|_a^b + \int_a^b (\partial_x L(f, f') - \partial_t \partial_y L(f, f'))\eta\end{aligned}$$

If we assume that the function $f$ and its variation $\eta$ always vanish at the endpoints $a, b$, then we obtain (2.1) with

$$\beta_f(t) = \partial_x L(f(t), f'(t)) - \partial_t \partial_y L(f(t), f'(t))$$

The equation $\beta_f = 0$ is called the **Euler-Lagrange equation**.

However, if no boundary conditions are imposed on the endpoints, then the term $\partial_y L(f, f')\eta\big|_a^b$ is not necessarily zero. As a result, we have

$$\beta_f = L(f(b), f'(b))\delta_b - L(f(a), f'(a))\delta_a + \partial_x L(f, f') - \partial_t \partial_y (f, f')$$

where, for each $c \in \mathbb{R}$, $\delta_c$ is the "**delta function**" at $c$, namely, the imaginary function $\mathbb{R} \to \mathbb{R}_{\geqslant 0}$ vanishing outside $c$ and satisfying $\int_{\mathbb{R}} \delta_c = 1$. The situation becomes even more singular if we define $S$ by

$$S(f) = \sum_{i=1}^n \lambda_i f(c_i) + \int_a^b L(f(t), f'(t))dt$$

where $\lambda_i \in \mathbb{R}$ and $a < c_i < b$, then

$$\beta_f = \sum_{i=1}^n \lambda_i \delta_{c_i} + L(f(b), f'(b))\delta_b - L(f(a), f'(a))\delta_a + \partial_x L(f, f') - \partial_t \partial_y (f, f')$$

This raises the question: what should the function $\beta_f$, alternatively the integral operator $\eta \mapsto \int_a^b \beta_f \eta$, actually look like in the general case? □

It is in this context that the problem of classifying bounded linear functionals on $C([a, b], \mathbb{R})$, originally posed by Hadamard in 1903, should be understood. Recall that if $V, W$ are normed vector spaces, $\Omega \subset V$ is open, and $S : \Omega \to W$ is a map, one says that $S$ is differentiable at $f \in \Omega$ if

$$S(f + \eta) - S(f) = \Lambda(\eta) + o(\eta)$$

where $\Lambda : V \to W$ is a bounded linear operator (called the **differential** of $S$ at $f$), and $\lim_{\|\eta\| \to 0} o(\eta)/\|\eta\| = 0$. In the calculus of variations, one sets $W = \mathbb{F}$. Then $\Lambda \in V^*$. One can thus understand $\eta \mapsto \delta S[f, \eta]$ as a bounded linear functional on a function space $V$ equipped with a suitable norm.

The problem of expressing $\delta S[f, \eta]$ as an integral involving $\eta$ is therefore transformed to the problem of characterizing the dual space $V^*$. More precisely, the space $V$—and in particular its norm—is not fixed in advance. The situation is not that one starts with a given normed space and is then asked to characterize its dual. Rather, the task is to find an appropriate norm on a suitable function space $V$ such that the bounded linear functionals on $V$, once studied and classified as integrals, are well-suited to capturing the variation of $S$.[1] The two perspectives on $\delta S[f, \eta]$—as a bounded linear functional on $V$, and as an integral involving $\eta$—together offer a deeper and more complete understanding of the variation of $S$.

More discussion of the relationship between dual spaces and the calculus of variations can be found in [Gray84].

## 2.2 Moment problems: a bridge between integral theory and dual spaces

The theory of dual spaces would not have reached its current depth and sophistication if it were developed solely within the framework of the calculus of variations. For instance, Riesz's classification of the duals of $C([a, b])$ and $L^p([a, b], m)$ would have been impossible without the Lebesgue and Stieltjes integrals. In fact, the very form of Riesz's theorems presents a striking connection between integration theory and dual spaces.

But why should such a connection exist in the first place? The way this relationship appears in Riesz's theorems calls for a deeper explanation. My short answer is this: it is the moment problems that form the bridge between integration theory and the theory of dual spaces. (Readers may jump ahead to Subsection 2.2.5 for the detailed final conclusion.)

To clarify my point, consider the first major example of a duality theorem: the identification $(L^2)^* \simeq L^2$ proved by Riesz and Fréchet in 1907:

**Theorem 2.2 (Riesz-Fréchet theorem**, the classical form**).** *We have a linear isomorphism*

$$\Lambda : L^2\big([-\pi, \pi], \frac{m}{2\pi}\big) \to L^2\big([-\pi, \pi], \frac{m}{2\pi}\big)^*$$

[1] The same function space $V$, when equipped with different norms, leads to different classifications of bounded linear functionals. For example, let $V = C([a, b])$. If the norm is $l^\infty$, then by Thm. 1.88, the bounded linear functionals are the Stieltjes integrals with respect to BV functions. If the norm is $L^2$, then by Exp. 2.30, the bounded linear functionals are those of the form $f \mapsto \int fg\,dm$ where $g \in L^2([a, b], m)$.

$$\langle \Lambda(f), g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} fg dm$$

In fact, Riesz studied $L^2$ spaces several years before introducing the more general $L^p$ spaces. His interest in $L^2$ spaces was clearly influenced by Hilbert's earlier work on the Hilbert space $l^2(\mathbb{Z})$ and its applications to the theory of integral equations. It was Hilbert's insights that served as the crucial bridge leading to the Riesz-Fréchet Thm. 2.2—the first major result linking Lebesgue integration with dual spaces.

As I will explain in the following, Hilbert's role in this development is best understood through the lens of moment problems.

### 2.2.1 Moment problems and dual spaces

Let me begin by introducing moment problems and explaining how they relate to dual spaces—particularly to the characterization of dual spaces in terms of integral representations.

**Problem 2.3** (**Moment problem**, original version)**.** Let $(\xi_n)$ be a sequence of scalar-valued functions defined on a space, e.g., an interval $I \subset \mathbb{R}$. Choose a sequence of scalars $(c_n)$ satisfying certain conditions. Find a scalar valued function $f$ on $I$ such that for all $n$, we have

$$\int \xi_n f = c_n \qquad \text{resp.} \qquad \int \xi_n df = c_n \tag{2.2}$$

The numbers $c_1, c_2, \dots$ are called the **moments** of $f$ resp. $df$.

There are two typical types of moment problems:

- **Trigonometric moment problem**: Here $I = \mathbb{S}^1 \simeq \mathbb{R}/2\pi\mathbb{Z}$, and $\xi_n(x) = e^{-\mathbf{i}nx}$ for $n \in \mathbb{Z}$. The problem then amounts to finding a function $f$ with prescribed Fourier coefficients $c_1, c_2, \dots$.

- **Polynomial moment problem**: Here $I \subset \mathbb{R}$ is an interval, not necessarily bounded, and $\xi_n(x) = x^n$ for $n \in \mathbb{N}$. One is asked to find an increasing or BV function $f$ such that $df$ has moments $c_1, c_2, \dots$.

Many (but not all) moment problems can be reformulated in the language of bounded linear functionals and dual spaces as follows:

**Problem 2.4** (**Moment problem**, dual space version)**.** Let $(\xi_n)$ be a sequence in a normed vector space $V$, and let $(c_n)$ be a sequence of scalars. Suppose that there exists $M \in \mathbb{R}_{\geqslant 0}$ such that

$$\Big|\sum_n a_n c_n\Big| \leqslant M \Big\|\sum_n a_n \xi_n\Big\| \tag{2.3}$$

for each sequence of scalars $(a_n)$ with finitely many nonzero terms. Find $\varphi \in V^*$ such that

$$\langle \xi_n, \varphi \rangle = c_n \qquad \text{for all } n \tag{2.4}$$

**Remark 2.5.** Note that (2.3) is necessary for the existence of $\varphi$ satisfying (2.4), because

$$\Big|\sum_n a_n c_n\Big| = \Big|\Big\langle \sum_n a_n \xi_n, \varphi \Big\rangle\Big| \leqslant \|\varphi\| \cdot \Big\|\sum_n a_n \xi_n\Big\|$$

where $\|\varphi\|$ is the operator norm. Hence (2.3) holds for any $M$ satisfying $\|\varphi\| \leqslant M$.

Conversely, if we know that $V_0 = \mathrm{Span}\{\xi_n\}$ is dense in $V$, then (2.3) guarantees that the linear functional

$$\varphi : V_0 \to \mathbb{F} \qquad \sum_n a_n \xi_n \mapsto \sum_n a_n c_n$$

is well-defined and bounded, with operator norm $\|\varphi\| \leqslant M$. By boundedness, $\varphi$ extends uniquely to a bounded linear functional on all of $V$. Therefore, Problem 2.4 can always be solved.

The case where $V_0$ is not dense in $V$ is more subtle and will be treated in detail in a later chapter. □

Once Problem 2.4 is resolved—for example, when $\mathrm{Span}\{\xi_n\}$ is dense in $V$—Problem 2.3 can be solved by answering the following:

**Problem 2.6 (Characterization of the dual space).** Characterize the elements of $V^*$ as precisely those linear functionals $\varphi : V \to \mathbb{F}$ of the form

$$\langle \xi, \varphi \rangle = \int \xi f \;\text{ resp. }\; \int \xi df$$

(for all $\xi \in V$), where $f$ is a function satisfying suitable regularity or integrability conditions.

Conversely, Problem 2.6 can be reduced to the moment Problem 2.3 by choosing a densely-spanning $(\xi_n)$ and taking $c_n = \langle \xi_n, \varphi \rangle$. The solution to Problem 2.3 then yields a function $f$ such that $\langle \xi_n, f \rangle = \langle \xi_n, \varphi \rangle$. By the density of $\mathrm{Span}\{\xi_n\}$ in $V$, it follows that $\varphi$ is represented by $f$.

Thus, we conclude that when $(\xi_n)$ spans a dense subspace of $V$, the moment problem (Problem 2.3) and the characterization of dual spaces (Problem 2.6) are equivalent.

### 2.2.2 Moment problems and integral theory/function theory

In the remainder of this section, we focus on the case where the sequence of functions $(\xi_n)$ is "sufficiently rich", for example, when it spans a dense subspace of $V$ in Problem 2.4. Under this assumption, the function $f$ (resp. $df$) in Problem 2.3 or the functional $\varphi$ in Problem 2.4 is uniquely determined by the moments $(c_n)$. Therefore, $(c_n)$ can be understood as the **coordinates** of $f$ (resp. $df$) and $\varphi$ under the **coordinate system** $(\xi_n)$.

We now explain how the moment problems connect to integral theory—in other words, to **function theory**. A central theme in function theory is the approximation of abstract or complicated functions by simpler, more elementary ones. This motivation often arises from practical mathematical problems, particularly those originating in physics, where one seeks to express the solution as a series of elementary functions, such as a power series or a Fourier series. The question of how such series should converge—uniformly, pointwise, or in some other sense—and what kinds of functions they can approximate was a central focus of function theory in the 18th and 19th centuries.

The first step in understanding and solving the approximation problem is to analyze the corresponding moment problem. A typical scenario unfolds as follows. In the setting of Problem 2.3, suppose there exists a sequence of elementary functions $(f_n)$ such that

$$\int \xi_k f_n \quad \text{resp.} \quad \int \xi_k df_n \quad = c_k \qquad \text{when } |k| \leqslant |n| \tag{2.5}$$

This situation arises, for instance, in the study of continued fractions and polynomial moments, where $\xi_k(x) = x^k$. In the case of Fourier series, an even stronger condition holds:

$$\int \xi_k f_n \quad \text{resp.} \quad \int \xi_k df_n \quad = \begin{cases} c_k & \text{if } |k| \leqslant |n| \\ 0 & \text{if } |k| > |n| \end{cases} \tag{2.6}$$

where $\xi_k(x) = e^{-\mathbf{i}kx}$ and $f_n(x) = \sum_{|k|\leqslant n} c_k e^{\mathbf{i}kx}$. The approximation problem asks:

**Problem 2.7.** Does the sequence $(f_n)$ converge to some function $f$? If so, in what sense does it converge?

To approach this problem, observe that if such a function $f$ exists, and if the integral commutes with the convergence of sequence of functions, then

$$\int \xi_k f = \int \lim_{|n|\to\infty} \xi_k f_n = \lim_{|n|\to\infty} \int \xi_k f_n \overset{(2.5)}{=\!=} c_k \tag{2.7a}$$

resp.

$$\int \xi_k df = \int \lim_{|n|\to\infty} \xi_k df_n = \lim_{|n|\to\infty} \int \xi_k df_n \overset{(2.5)}{=\!=} c_k \tag{2.7b}$$

Therefore, the first step in solving Problem 2.7 is to find a function $f$ solving the moment Problem 2.3. Once such an $f$ is found, the next step is to prove that the sequence $(f_n)$ converges to $f$, and investigate the mode of convergence.

Historically, the understanding of convergence, the properties of the limiting function $f$, and the integrals appearing in (2.7) was often insufficient to resolve the approximation problem at the outset. In many cases, addressing the approximation problem required the development of new theories of integration or the extension of the class of integrable functions. Both the Lebesgue and Stieltjes integrals emerged from such needs. For instance, the challenges posed by Fourier series played a central role in motivating the development of the Riemann and later the Lebesgue integral. See [Jah, Ch. 6, 9] and [Haw-L] for a detailed discussion of how Fourier series drove this evolution. The connection between continued fractions and the Stieltjes integral will be explored in Ch. 4.

| Function theory | Moment Problems | Dual spaces |
|---|---|---|
| Lebesgue integral<br>&<br>Fourier series | Fourier coefficients | $L^2([a,b],m)^*$ |
| Stieltjes integral<br>&<br>Continued fractions | Polynomial moments | $C([a,b])^*$ |

Table 2.1: The origin of moment problems in function theory

### 2.2.3 Convergence of functions, moments, and linear functionals

In the previous subsection, we noted that solving moment problems determines the function $f$ that appears in Problem 2.7. But can the moment problem perspective also help us understand the convergence of $f_n$ to $f$? Or conversely, can the convergence behavior of $f_n$ toward $f$ offer deeper insight into the structure of moment problems themselves? Thanks to Hilbert's foundational work on the Hilbert space $l^2(\mathbb{Z})$—especially his groundbreaking 1906 paper [Hil06]—the answer is yes.[2]

A key concept introduced by Hilbert in [Hil06] is **weak convergence**: If $(\psi_n)$ is a sequence in $l^2(\mathbb{Z})$ with uniformly bounded norm, i.e.,

$$\sup_n \|\psi_n\|_2 < +\infty \tag{2.8}$$

we say that $(\psi_n)$ converges weakly to $\psi \in l^2(\mathbb{Z})$ if it converges pointwise $\mathbb{Z}$, i.e.,

$$\lim_n \psi_n(k) = \psi(k) \qquad \text{for all } k \in \mathbb{Z} \tag{2.9}$$

[2]Indeed, Hilbert originally worked with the real Hilbert space $l^2(\mathbb{Z},\mathbb{R})$, rather than the complex one $l^2(\mathbb{Z}) = l^2(\mathbb{Z},\mathbb{C})$. For clarity and simplicity, however, we will work with $l^2(\mathbb{Z})$ in what follows.

Since $l^2(\mathbb{Z})$ is typically interpreted as the space of Fourier series of $L^2$-integrable functions, Hilbert's notion of weak convergence corresponds to the (pointwise) convergence of Fourier coefficients. That is,

$$\lim_n \widehat{f}_n(k) = \widehat{f}(k) \qquad \text{for all } k \in \mathbb{Z}$$

where $f_n$ and $f$ are $L^2$-integrable functions on $[-\pi, \pi]$.[3]

The notion of weak convergence—later extended to weak-* convergence—provided a fundamentally new insight into the study of moment problems and their connection to dual spaces and function theory/integral theory. Since Fourier coefficients are simply trigonometric moments, the weak convergence described by (2.9) can be understood as the (pointwise) **convergence of moments**, which means, in the setting of the moment Problem 2.3, that

$$\lim_n \int \xi_k f_n \quad \text{resp.} \quad \lim_n \int \xi_k df_n \quad = c_n \qquad \text{for all } k \tag{2.10}$$

The translation of (2.10) into the setting of the dual space version of the moment Problem 2.4 is straightforward: One considers a sequence $(\varphi_k)$ in $V^*$ such that $\lim_n \langle \xi_k, \varphi_n \rangle = \langle \xi_k, \varphi \rangle$ holds for all $k$. Since we have assumed at the beginning of Subsec. 2.2.2 that $(\xi_n)$ spans a dense subspace of $V$, it follows from (2.8) that this convergence of moments is equivalent to the **weak-* convergence** of $(\varphi_n)$ to $\varphi$. That is, we say that $(\varphi_n)$ converges weak-* to $\varphi$ if

$$\lim_n \langle \xi, \varphi_n \rangle = \langle \xi, \varphi \rangle \qquad \text{for all } \xi \in V \tag{2.11}$$

Thus, the second and third columns of Table 2.2 are equivalent. See Thm. 2.42 for the formal statement of this equivalence.

On the other hand, (2.10) generalizes the condition (2.5), which, as previously mentioned, arises naturally in the study of Fourier series and continued fractions. As such, its function-theoretic interpretation—highlighted by the following theorems—provides a general framework for understanding the convergence of the sequence $(f_n)$ to $f$ in Problem 2.7.

**Theorem 2.8.** *Let $1 < p \leqslant +\infty$ and $p^{-1} + q^{-1} = 1$. Let $(f_n)$ be a uniformly $L^p$-norm bounded sequence in $L^p([a,b], m)$. Suppose that $(f_n)$ converges pointwise to $f$. Then we have $f \in L^p([a,b], m)$. Moreover, $(f_n)$ converges weak-* to $f$, which means that $\lim_n \int f_n g dm = \int f g dm$ for all $g \in L^q([a,b], m)$.*

*Proof.* See Thm. 2.50. □

[3]Hilbert himself did not initially connect $l^2(\mathbb{Z})$ with the Lebesgue integral. The precise relationship between $l^2(\mathbb{Z})$ and $L^2([-\pi, \pi], \frac{m}{2\pi})$ was later clarified by Riesz and Fischer in 1907.

**Theorem 2.9.** *Let $1 < p \leqslant +\infty$ and $p^{-1} + q^{-1} = 1$. Let $(f_n)$ be a uniformly $L^p$-norm bounded sequence in $L^p([a,b],m)$. Then $(f_n)$ converges weak-* to some element $f \in L^p([a,b],m)$ iff the limit*

$$F(x) := \lim_n \int_a^x f_n dm \tag{2.12}$$

*exists for every $x \in [a,b]$. When $(f_n)$ converges weak-* to $f \in L^p([a,b],m)$, for each $x \in [a,b]$ we have*

$$F(x) = \int_a^x f dm \tag{2.13}$$

*Proof.* If $(f_n)$ converges weak-* to $f$, then $\lim_n \int f_n \chi_{[a,x]} = \int f_n \chi_{[a,x]}$, which implies that $F(x)$ exists and equals $\int_a^x f dm$.

The other direction is more difficult. Indeed, it is almost equivalent to the duality $L^p([a,b],m) \simeq L^q([a,b],m)^*$. See Thm. 2.49. □

**Theorem 2.10.** *Let $(\rho_n)$ be a uniformly $l^\infty$-bounded sequence of increasing functions $[a,b] \to \mathbb{R}_{\geqslant 0}$. The following are true.*

1. *Let $\rho : [a,b] \to \mathbb{R}_{\geqslant 0}$ be bounded and increasing. Then $(d\rho_n)$ converges weak-* to $d\rho$ iff $(\rho_n)$ converges pointwise to $\rho$ at $b$ and at any point where $\rho|_{(a,b)}$ is continuous.*
2. *$(d\rho_n)$ converges weak-* to $d\rho$ for some bounded increasing $\rho : [a,b] \to \mathbb{R}_{\geqslant 0}$ iff $(\rho_n)$ converges pointwise at $b$ and on a dense subset of $I$.*

By saying that $(d\rho_n)$ converges weak-* to $d\rho$, we mean $\lim_n \int g d\rho_n = \int g d\rho$ for all $g \in C([a,b],m)$.

*Proof.* See Thm. 2.65 and Cor. 2.66. □

The above theorems establish an intimate connection between the (pointwise) convergence of moments and the pointwise convergence of the antiderivatives of a sequence of functions.[4] Our understanding of convergence from various perspectives can thus be summarized in Table 2.2.

| Function theory | Moment Problems | Dual spaces |
|---|---|---|
| Pointwise convergence of (antiderivatives of) a sequence of functions | Pointwise convergence of moments | Weak-* convergence |

Table 2.2: Equivalence of convergence notions

[4]We are viewing $\rho_n$ and $\rho$ as the antiderivatives of $d\rho_n$ and $d\rho$.

### 2.2.4 Equivalence of the first and second columns of Table 2.2

Thm. 2.9 and 2.10, which establish the equivalence of the first and second columns of Table 2.2, are not easy to prove. In fact, proving Thm. 2.9 typically requires the duality $L^p([a,b]) \simeq L^q([a,b])^*$, or at least techniques closely related to those used in establishing this duality.

Therefore, the solvability of the moment problems (Problems 2.3 and 2.4)—in other words, the solvability of Problem 2.6 concerning the characterization of dual spaces—is closely related to the equivalence between the first and second columns of Table 2.2. This close connection rests on the following principle:

**Principle 2.11.** Usually, if $V$ is a normed vector space consisting of functions, any element $\varphi$ of $V^*$ can be weak-* approximated by elementary functions with uniformly bounded norms. More precisely, there exists a sequence (or a net) of elementary functions $(f_n)$ such that the operator norms of the linear functionals $\xi \in V \mapsto \int \xi f_n$ are uniformly bounded, and

$$\lim_n \int \xi f_n = \langle \xi, \varphi \rangle \qquad \text{for all } \xi \in V$$

**Remark 2.12.** Here is how, with the help of Principle 2.11, the characterization of $V^*$ can be derived from the equivalence of the first and second columns of Table 2.2:

By this principle, for each $\varphi \in V^*$, we can select a sequence $(f_n)$ approximating weak-* to $\varphi$. Since the second column of Table 2.2 implies the first column, the sequence $(f_n)$ converges to some function $f$ in the sense described in the first column of Table 2.2. Then, by the equivalence of the three modes of convergence in that table, it follows that $(f_n)$ converges weak-* to $f$. Consequently, $\varphi$ is represented by integration against $f$, thereby solving the problem of characterizing the dual space $V^*$. $\square$

The idea outlined in Rem. 2.12 is roughly the approach Riesz employed in 1907 to solve the following trigonometric moment problem.

**Theorem 2.13 (Riesz-Fischer theorem**, Riesz's original version)**.** [5] *For each $(c_k)_{k\in\mathbb{Z}}$ in $l^2(\mathbb{Z})$, there is an (automatically unique) $f \in L^2([-\pi,\pi], \frac{m}{2\pi})$ whose Fourier series is equal to $(c_k)$.*

---

[5]The modern interpretation of the Riesz-Fischer theorem as stating that $L^2(X,\mu)$ (or more generally $L^p(X,\mu)$) is Cauchy-complete for any measure space $(X,\mu)$ has led to a significant misunderstanding. In fact, while Fischer formulated the theorem for $L^2([-\pi,\pi], \frac{m}{2\pi})$ in terms of Cauchy sequences, Riesz understood it quite differently—through the lens of moment problems.

Therefore, once Riesz realized that solving moment problems is equivalent to the characterization of dual spaces, he immediately obtained the Riesz-Fréchet Thm. 2.2. As we have emphasized at the beginning of Sec. 2.1, completeness and duality are fundamentally distinct properties, each serving distinct purposes and arising from different considerations. The fact that they coincide in the case of inner product spaces is purely a coincidence.

***Riesz's idea of the proof.*** [6] Choose $(c_k)_{k\in\mathbb{Z}}$ in $l^2(\mathbb{Z})$. One aims to solve the moment problem that there exists $f \in L^2$ such that $\frac{1}{2\pi}\int fe_{-k} = c_k$ for all $k \in \mathbb{Z}$, where $e_k(x) = e^{\mathbf{i}kx}$. For each $n \in \mathbb{N}$, let

$$f_n = \sum_{-n\leqslant k\leqslant n} c_k e_k$$

Then $(f_n)$ converges weak-* to the bounded linear functional $\varphi \in (L^2)^*$ satisfying $\langle e_{-k}, \varphi\rangle = c_k$ for all $k$. (This is an instance of Principle 2.11.) [7]

On the other hand, the property $\sum_k |c_k|^2 < +\infty$ implies that the antiderivatives of $(f_n)$ converge pointwise to some function $F$ in the sense of (2.12). This establishes the convergence described in the first column of Table 2.2.

Then, applying the fundamental theorem of calculus for the Lebesgue integral, Riesz deduced the convergence in the second column of Table 2.2 for the derivative function $f := F'$ (which exists a.e. and is $L^2$) and for another densely spanning set of functions—the set $\{\chi_{[a,x]} : x \in [a,b]\}$.[8] Namely, he obtained

$$\langle \chi_{[a,x]}, f\rangle = \lim_n \langle \chi_{[a,x]}, f_n\rangle \qquad \text{for all } x \in [a,b]$$

Therefore, since the second column of Table 2.2 is equivalent to the third, $(f_n)$ converges weak-* to $f$. Thus $\varphi$ is represented by $f$, which implies that $f$ solves the desired moment problem—since $\varphi$ does. □

Note that the fundamental theorem of calculus for the Lebesgue integral is crucial to the above proof. Likewise, the Radon-Nikodym Thm. 1.46—a modern form of the fundamental theorem of calculus—also plays a central role in the proof of Theorem 1.50, which establishes the duality 1.50 on $L^p(X,\mu) \simeq L^q(X,\mu)^*$. This reinforces the point that the characterization of dual spaces is deeply connected to the equivalence between the first and second columns of Table 2.2.

See [Gui-A, Sec. 27.3] for further discussion on the relationship between the classical and modern proofs of the duality $L^p \simeq (L^q)^*$, the connection between this duality and the completeness of $L^p$-spaces, and the role of derivatives—both in the classical sense and in the form of Radon-Nikodym derivatives—in this context.

### 2.2.5 Conclusion

We now summarize the discussion so far by addressing the question posed at the beginning of this section: Why are dual spaces related to integral theory? More specifically, from the mathematical-historical perspective, why is it possible to characterize the dual spaces of $L^p(X,\mu)$ and $C(X)$?

[6] See [Haw-L, Ch. 6].

[7] Riesz's original proof does not use the language of linear functionals.

[8] The fact that the fundamental theorem of calculus for the Lebesgue integral—one of the deep-

| Function theory | Moment Problems | Dual spaces |
|---|---|---|
| | Solving moment problems | Characterizing $V^*$ |
| Related by $\Updownarrow$ Principle 2.11 | | |
| Pointwise convergence of (antiderivatives of) a sequence of functions | Pointwise convergence of moments | Weak-* convergence |

Table 2.3: The cells in each row are equivalent

The answer, in my view, is captured in Table 2.3: The power of the Lebesgue and Stieltjes integrals lies in their ability to establish the equivalence between the two gray cells in that table. Once this equivalence is established, with the help of Principle 2.11, the characterization of dual spaces in terms of integrals becomes straightforward.

But why are these two integrals powerful enough to establish the equivalence between the two gray cells in Table 2.3?—Because both the Lebesgue and Stieltjes integrals arise from the study of moment problems, which in turn are rooted in the corresponding approximation problems, as illustrated in Table 2.1. The emphasis of these integral theories on the commutativity of limits and integration anticipates the equivalence of the two gray cells.

In light of the equivalences in Table 2.3, the Lebesgue integral, as the completion of the Riemann integral, can be interpreted as the weak-* completion of trigonometric functions and continuous functions. Similarly, the Stieltjes integral, as the completion of finite sums, can be viewed as the weak-* completion of discrete spectra—a perspective that will be one of the main themes of Ch. 4. See Table 2.4.

| Completion of Integrals | Extension of classes of functions | **Weak-* completion** |
|---|---|---|
| Riemann integral<br>$\cap$<br>Lebesgue integral | Continuous functions<br>$\cap$<br>Measurable functions | of continuous functions |
| Finite sum<br>$\cap$<br>Stieltjes integral | Discrete spectra<br>$\cap$<br>Continuous spectra | of discrete spectra |

Table 2.4

*Side note.* A common viewpoint—motivated by the completeness of $L^1$-spaces—regards

est results in measure theory—is used here highlights how non-trivial the equivalence between the first and second columns of Table 2.2 really is.

the Lebesgue integral and the Lebesgue measurable/integrable functions as the Cauchy completion of Riemann integrals and continuous functions. In my view, this perspective is not only historically inaccurate, but also mathematically misleading.

Historically, the first $L^p$-space considered is $L^2([a,b],m)$, due to its close relation with $l^2(\mathbb{Z})$, the space of trigonometric moments of $L^2$-integrable functions. The space $l^2(\mathbb{Z})$ was introduced by Hilbert in [Hil06], where weak convergence (equivalently, pointwise convergence of moments) plays a central role in his proof of the Hilbert-Schmidt theorem. In [Rie10], Riesz studied the space $L^p([a,b],m)$ for $1<p<+\infty$, and in particular proved the duality $L^p([a,b],m)\simeq L^q([a,b],m)^*$. The completeness of $L^p([a,b],m)$ follows as a corollary. However, $L^1([a,b],m)$ was not considered, likely due to its lack of a satisfactory duality theory. This clearly shows that duality was originally viewed as more fundamental than Cauchy completeness.

Mathematically, to perform a Cauchy completion, one needs a norm, which in this context is defined via an integral. Yet, while integrals are linear functionals, norms only satisfy the subadditivity. As a result, norms and Cauchy completions do not provide the right conceptual framework for understanding the nature of the Lebesgue integral from a functional-analytic perspective.

The more appropriate viewpoint is to regard the Lebesgue integral as arising from weak-* completion, not Cauchy completion.

## 2.3 Bounded multilinear maps

### 2.3.1 Seminorms, norms, and normed vector spaces

**Definition 2.14.** If $V$ is an $\mathbb{F}$-vector space, a function $\|\cdot\|:V\rightarrow\mathbb{R}_{\geqslant 0}$ is called a **seminorm** if

$$\|av\|=|a|\cdot\|v\| \qquad \|u+v\|\leqslant\|u\|+\|v\| \qquad \text{for any } u,v\in V \text{ and } a\in\mathbb{F} \tag{2.14}$$

A seminorm is called a **norm** if any $v\in V$ satisfying $\|v\|=0$ is the zero vector $0$. A vector space $V$, equipped with a norm, is called a **normed vector space**.

If $V$ is a normed vector space, then a **normed vector subspace** of $V$ denotes a linear subspace $U\subset V$ equipped with the norm inherited from $V$, i.e., the restriction of $V$'s norm to $U$.

We say that $V$ is **separable** if it is so under the **norm topology**, namely, the topology induced by the metric $d(u,v)=\|u-v\|$. □

**Remark 2.15.** In Def. 2.14, the condition $\|av\|=|a|\cdot\|v\|$ can be weakened to

$$\|av\|\leqslant|a|\cdot\|v\| \qquad \text{for any } v\in V \text{ and } a\in\mathbb{F} \tag{2.15}$$

Therefore, (2.14) can be weakened to

$$\|au+bv\|\leqslant|a|\cdot\|u\|+|b|\cdot\|v\| \qquad \text{for any } u,v\in V \text{ and } a,b\in\mathbb{F} \tag{2.16}$$

*Proof.* Suppose that (2.15) is true. Then we clearly have $\|av\| = |a| \cdot \|v\|$ when $a = 0$. Suppose that $a \neq 0$. Then $\|v\| = \|a^{-1}av\| \leqslant |a|^{-1}\|av\|$, and hence $\|av\| \geqslant |a| \cdot \|v\|$. Therefore $\|av\| = |a| \cdot \|v\|$. $\square$

**Remark 2.16.** The norm function $\|\cdot\| : V \to \mathbb{R}_{\geqslant 0}$ is continuous. This is because

$$\|u\| - \|v\| \leqslant \|u - v\| \tag{2.17}$$

Therefore, if $(v_\alpha)$ is a net in $V$ converging (in norm) to $v$, then

$$\|v\| = \lim_\alpha \|v_\alpha\|$$

**Proposition 2.17.** *Let $\|\cdot\|_V$ be a seminorm on an $\mathbb{F}$-vector space $V$. Let $V_0 = \{v \in V : \|v\|_V = 0\}$. Then $V_0$ is a linear subspace on $V$, and there is a (clearly unique) norm $\|\cdot\|_{V/V_0}$ on the quotient space $V/V_0$ such that*

$$\|v + V_0\|_{V/V_0} = \|v\|_V \qquad \textit{for all } v \in V \tag{2.18}$$

In the future, unless otherwise stated, we will always equip $V/V_0$ with this norm $\|\cdot\|_{V/V_0}$.

*Proof.* We abbreviate $\|\cdot\|_V$ to $\|\cdot\|$. If $u, v \in V_0$ and $a, b \in \mathbb{F}$, then

$$\|au + bv\| \leqslant |a|\|u\| + |b|\|v\| = 0$$

This shows that $V_0$ is a linear subspace of $V$. On the other hand, if $u, v \in V$ satisfy $u + V_0 = v + V_0$, then $u - v \in V_0$, and hence

$$\|v\| = \|u + v - u\| \leqslant \|u\| + \|v - u\| = \|u\|$$

Similarly, $\|u\| \leqslant \|v\|$. Therefore $\|u\| = \|v\|$. This implies that we have a well-defined function $\|\cdot\|_{V/V_0} : V/V_0 \to \mathbb{R}_{\geqslant 0}$ satisfying (2.18).

If $u, v \in V$ and $a, b \in \mathbb{F}$, then

$$\|a(u + V_0) + b(v + V_0)\|_{V/V_0} = \|au + bv + V_0\|_{V/V_0} = \|au + bv\| \leqslant |a|\|u\| + |b|\|v\|$$

$\square$

### 2.3.2 Bounded multilinear maps

In the rest of this section, $V_1, V_2, \ldots$ and $U, V, W$ all denote normed $\mathbb{F}$-vector spaces.

**Definition 2.18.** Let $N \in \mathbb{Z}_+$. A map $T : V_1 \times \cdots \times V_N \to W$ is called a **multilinear map** if for each $1 \leqslant i \leqslant N$ and each fixed $v_j \in V_j$ (for all $j \neq i$), the map

$$v_i \in V_i \mapsto T(v_1, \ldots, v_N) \in W$$

is $\mathbb{F}$-linear. We let

$$\mathrm{Lin}(V_1 \times \cdots \times V_N, W) = \{\text{multilinear maps } V_1 \times \cdots \times V_N \to W\}$$

For each $T \in \mathrm{Lin}(V_1 \times \cdots \times V_N, W)$, we define the **operator norm**

$$\|T\| := \|T\|_{l^\infty(\overline{B}_{V_1}(0,1)\times\cdots\times\overline{B}_{V_N}(0,1),W)} = \sup_{v_1\in\overline{B}_{V_1}(0,1),\ldots,v_N\in\overline{B}_{V_N}(0,1)} \|T(v_1, \ldots, v_N)\|$$

We say that $T$ is **bounded** if $\|T\| < +\infty$.

**Definition 2.19.** We let

$$\mathfrak{L}(V_1 \times \cdots \times V_N, W) := \{\text{bounded multilinear maps } V_1 \times \cdots \times V_N \to W\} \tag{2.19}$$

viewed as an $\mathbb{F}$-linear subspace of $W^{V_1\times\cdots\times V_N}$. We let

$$\mathfrak{L}(V) := \mathfrak{L}(V, V) \qquad V^* := \mathfrak{L}(V, \mathbb{F})$$

Elements of $\mathfrak{L}(V)$ are called **bounded linear operators on** $V$. An element $T \in \mathfrak{L}(V)$ is called **invertible** if there exists $T^{-1} \in \mathfrak{L}(V)$ such that

$$TT^{-1} = T^{-1}T = \mathrm{id}_V$$

The space $V^*$ is called the **dual space** of $V$.

**Remark 2.20.** In this course, the most frequently encountered cases of (2.19) are $\mathfrak{L}(V)$, $V^*$, and $\mathfrak{L}(U \times V, \mathbb{F})$. In Ch. 4, we also consider spaces such as $\mathfrak{L}(U \times V \times V_*, \mathbb{F})$, where $V_*$ is a normed vector space with dual space $V$. In such cases, Prop. 2.38 gives isomorphisms

$$\mathfrak{L}(U \times V \times V_*, \mathbb{F}) \simeq \mathfrak{L}(U, \mathfrak{L}(V \times V_*, \mathbb{F})) \simeq \mathfrak{L}(U, \mathfrak{L}(V))$$

**Remark 2.21.** $\|T\|$ is the smallest element in $\overline{\mathbb{R}}_{\geqslant 0}$ satisfying

$$\|T(v_1, \ldots, v_N)\| \leqslant \|T\| \cdot \|v_1\| \cdots \|v_N\| \tag{2.20}$$

*Proof.* If one of $v_1, \ldots, v_N$ is zero, then $T(v_1, \ldots, v_N) = 0$ by the multilinearity, and hence (2.20) holds. So we assume that $v_1, \ldots, v_N$ are all non-zero. So their norms are all nonzero. Since $v_i/\|v_i\| \in B_{V_i}(0, 1)$, we have

$$\left\|T\Big(\frac{v_1}{\|v_1\|}, \cdots, \frac{v_N}{\|v_N\|}\Big)\right\| \leqslant \|T\|$$

which implies (2.20) by the multilinearity.

We have proved that $\|T\|$ satisfies (2.20). Now, suppose that $C \in \overline{\mathbb{R}}_{\geqslant 0}$ and

$$\|T(v_1, \ldots, v_N)\| \leqslant C \cdot \|v_1\| \cdots \|v_N\|$$

for all $v_i \in V_i$. Taking $v_i \in \overline{B}_V(0, 1)$, we see that $\|T\| \leqslant C$. $\square$

Recall Def. 1.3.

**Proposition 2.22.** *Let $T : V_1 \times \cdots \times V_N \to W$ be multilinear. The following are equivalent.*

*(a) $T$ is continuous.*

*(b) $T$ is continuous at $0 \times \cdots \times 0$.*

*(c) $T$ is bounded.*

*(d) $T$ is Lipschitz continuous on $\overline{B}_{V_1}(0,R) \times \cdots \times \overline{B}_{V_N}(0,R)$ for every $R \in \mathbb{R}_{>0}$.*

*(e) $T$ is Lipschitz continuous on $\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1)$.*

*Moreover, if $T$ is bounded, and if $V_1 \times \cdots \times V_N$ is equipped with the $l^\infty$-product metric, then the Lipschitz constant in (d) can be chosen to be $NR^{N-1}\|T\|$.*

What matters about the Lipschitz constant above is not its exact formula, but the implication it carries: namely, that any family $(T_\alpha)$ in $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ satisfying $\sup_\alpha \|T_\alpha\| < +\infty$, when restricted to a bounded subset of $V_1 \times \cdots \times V_N$, admits a uniform Lipschitz constant.

*Proof.* Clearly (a)⇒(b).

(b)⇒(c): Assume (b). Then $0 \times \cdots \times 0$ is an interior point of $T^{-1}(B_W(0,1))$, and hence contains $B_{V_1}(0,2\delta_1) \times \cdots \times B_{V_N}(0,2\delta_N)$ for some $\delta_1,\dots,\delta_N > 0$. So $T$ sends $\overline{B}_{V_1}(0,\delta_1) \times \cdots \times \overline{B}_{V_N}(0,\delta_N)$ (which equals $\delta_1\overline{B}_{V_1}(0,1) \times \cdots \times \delta_N\overline{B}_{V_N}(0,1)$) into $B_W(0,1)$. By multilinearity, $T$ sends $\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1)$ into $B_W(0,\delta_1^{-1}\cdots\delta_N^{-1})$. This proves (c).

(c)⇒(d): Assume (c). Choose $v_i \in \overline{B}_{V_i}(0,R_i)$. Then, for each $\xi_i \in \overline{B}_{V_i}(0,R_i)$,

$$
\begin{aligned}
&\|T(\xi_1,\dots,\xi_N) - T(v_1,\dots,v_N)\| \\
\leqslant&\|T(\xi_1 - v_1, \xi_2, \xi_3, \dots, \xi_N)\| + \|T(v_1, \xi_2 - v_2, \xi_3, \dots, \xi_N)\| \\
&+ \|T(v_1, v_2, \xi_3 - v_3, \dots, \xi_N)\| + \cdots + \|T(v_1, v_2, v_3, \dots, \xi_N - v_N)\| \\
\leqslant&NR^{N-1}\|T\| \cdot \max\{\|\xi_1 - v_1\|, \dots, \|\xi_N - v_N\|\}
\end{aligned}
$$

where (2.20) is used in the last inequality. Thus $T$ has Lipschitz constant $NR^{N-1}\|T\|$.

(e)⇔(d): This is clear by scaling the vectors.

(d)⇒(f): This is clear from Rem. 1.4. □

**Corollary 2.23.** *Let $T \in \mathfrak{L}(V,W)$. Then $\mathrm{Ker}(T)$ is a closed linear subspace of $V$.*

*Proof.* By Prop. 2.22, $T$ is continuous. Since the preimage of any closed set under a continuous map is closed, $\mathrm{Ker}(T) = T^{-1}(0)$ is closed. □

**Example 2.24.** A linear map $T : V \to W$ is called a **linear isometry** if it is an isometry of metric spaces, i.e., $\|Tv_1 - Tv_2\| = \|v_1 - v_2\|$ for all $v_1, v_2 \in V$. This is clearly equivalent to

$$\|Tv\| = \|v\| \qquad \text{for all } v \in V$$

A linear isometry is clearly bounded with operator norm $\|T\| = 1$ (unless when $V = \{0\}$). Moreover, a linear isometry is clearly injective. A linear isometry $T : V \to W$ which is also surjective (and hence bijective) is called an **isomorphism of normed vector spaces**. In that case, we say that the normed vector spaces $V, W$ are **isomorphic**.

**Remark 2.25.** Suppose that $\Phi : V \to W$ is a linear map of vector spaces, and $W$ is a normed vector space. Then $V$ has a seminorm defined by

$$\|v\|_V := \|\Phi(v)\|_V$$

Equip $V/\mathrm{Ker}\Phi$ with the norm defined by Prop. 2.17. Then $\Phi$ descends to a linear map $\widetilde{\Phi} : V/\mathrm{Ker}\Phi \to W$, which is clearly a linear isometry.

**Example 2.26.** Let $1 \leqslant p \leqslant +\infty$, let $X$ be an LCH space, let $\mu$ be a Radon measure (or its completion) on $X$. Let $\Phi : C_c(X, \mathbb{F}) \to L^p(X, \mu, \mathbb{F})$ be the obvious map. Then $\Phi$ descends to a linear isometry of normed vector spaces

$$C_c(X, \mathbb{F})/\{f \in C_c(X, \mathbb{F}) : f = 0 \ \mu\text{-a.e.}\} \longrightarrow L^p(X, \mu, \mathbb{F}) \tag{2.21}$$

Now assume $p < +\infty$. Then by Thm. 1.60, the map (2.21) has dense range. This is often expressed by saying that $C_c(X, \mathbb{F})/\{f \in C_c(X, \mathbb{F}) : f = 0 \ \mu\text{-a.e.}\}$ is dense in $L^p(X, \mu, \mathbb{F})$, or simply that $C_c(X, \mathbb{F})$ is dense in $L^p(X, \mu, \mathbb{F})$.

## 2.4 Fundamental properties of bounded multilinear maps

Let $V_1, V_2, \ldots, U, V, W$ be normed vector spaces. In this section, we establish several fundamental properties of bounded multilinear maps that will be used frequently throughout the course. We first note the elementary fact:

**Remark 2.27.** Let $U$ be a linear subspace of $V$. Let $R \in \mathbb{R}_{>0}$. Then $U$ is dense in $V$ iff $\overline{B}_U(0, R)$ is dense in $\overline{B}_V(0, R)$.

*Proof.* The direction "$\Leftarrow$" is obvious. Let us prove "$\Rightarrow$". Let $\xi \in \overline{B}_V(0, R)$, choose a sequence $(\xi_n)$ in $U$ converging to $\xi$. Assume WLOG that $\xi \neq 0$ and $R \in \mathbb{R}_{>0}$; otherwise, the approximation is obvious. Since the norm function is continuous, $\|\xi_n\| \to \|\xi\|$. In particular, $\|\xi_n\|$ is eventually nonzero. Thus $\frac{\|\xi\|}{\|\xi_n\|}\xi_n \to \xi$. $\square$

Recall that two sequences $(x_n), (y_n)$ in a metric space $X$ is called **Cauchy equivalent** if $\lim_n d(x_n, y_n) = 0$.

**Theorem 2.28.** *Suppose that $W$ is complete. For each $i$, let $U_i$ be a dense linear subspace of $V_i$. Then we have an isomorphism of normed vector spaces*

$$\begin{gathered}\mathfrak{L}(V_1\times\cdots\times V_N,W)\xrightarrow{\simeq}\mathfrak{L}(U_1\times\cdots\times U_N,W)\\ T\mapsto T\big|_{U_1\times\cdots\times U_N}\end{gathered}\tag{2.22}$$

*Proof.* Denote the map (2.22) by $\Phi$ which is clearly linear. By Rem. 2.27, $\overline{B}_{U_1}(0,1)\times\cdots\times\overline{B}_{U_N}(0,1)$ is dense in $\overline{B}_{U_1}(0,1)\times\cdots\times\overline{B}_{U_N}(0,1)$. This shows that $\Psi$ is a linear isometry, i.e., $T$ and $T\big|_{U_1\times\cdots\times U_N}$ have the same operator norm.

We now show that $\Phi$ is surjective. Here, the completeness of $W$ is need. Let $T\in\mathfrak{L}(U_1\times\cdots\times U_N,W)$. We want to extend $T$ to a bounded multilinear map $V_1\times\cdots\times V_N\to W$. We only need to extend $T$ on the first component, i.e., extend $T$ to a bounded multilinear $V_1\times U_2\times U_3\times\cdots\times U_N\to W$. Then, a similar argument applies to the second component extend $T$ to a bounded multilinear $V_1\times V_2\times U_3\times\cdots\times U_N\to W$. By repeating this procedure, we obtain bounded multilinear $V_1\times\cdots\times V_N\to W$ extending $T$.

Let $\xi\in V_1,u_2\in U_2,\dots,u_N\in U_N$. Let $(\xi_n)$ be a sequence in $U_1$ converging to $\xi$. In particular, $(x_n)$ is a Cauchy sequence. By Rem. 2.21, $T(\xi_n,v_2,\dots,v_N)$ is a Cauchy sequence in $W$. Therefore, by the completeness of $W$, $T(\xi_n,v_2,\dots,v_N)$ converges to some element, which we denote by $T(\xi,v_2,\dots,v_N)$.

Let us show that the definition of $T(\xi,v_2,\dots,v_N)$ is independent of the choice of sequence converging to $\xi$. Suppose that $(\xi_n')$ is another sequence converging to $\xi$. Then $(\xi_n)$ and $(\xi_n')$ are Cauchy equivalent. By Rem. 2.21, $T(\xi_n,v_2,\dots,v_N)$ and $T(\xi_n',v_2,\dots,v_N)$ are Cauchy equivalent. So they converge to the same element.

Thus, we have defined a map $T:V_1\times U_2\times\cdots\times U_N\to W$. We leave it to the reader to check that $T$ is bounded multi-linear map. □

**Corollary 2.29.** *Let $U$ be a dense linear subspace of $V$. Then we have an isomorphism of normed vector spaces*

$$V^*\xrightarrow{\simeq}U^*\qquad\varphi\mapsto\varphi|_U\tag{2.23}$$

*Proof.* This follows immediate from Thm. 2.28. □

**Example 2.30.** Let $1\leqslant q<+\infty$ and $p^{-1}+q^{-1}=1$. Let $X$ be an LCH space. Let $\mu$ be a Radon measure (or its completion) on $X$. By Exp. 2.26, the $L^q$-seminorm on $C_c(X,\mathbb{F})$ descends to the $L^q$-norm on $V=C_c(X,\mathbb{F})/\{f\in C_c(X,\mathbb{F}):f=0\ \mu\text{-a.e.}\}$, and $V$ is dense in $L^p(X,\mu)$. Therefore, by Thm. 1.50 and Cor. 2.29, the map (1.13) gives an isomorphism of normed vector spaces $V^*\simeq L^p(X,\mu)$.

The following Prop. 2.31 and Thm. 1.83 will imply Thm. 2.42, which establishes the equivalence of the second and third columns of Table 2.2.

**Proposition 2.31.** *For each $i$, let $E_i$ be a densely spanning subset of $V_i$. Let $(T_\alpha)$ be a net in $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ with **uniformly bounded operator norms**, i.e., $\sup_\alpha \|T_\alpha\| < +\infty$. Choose $T \in \mathfrak{L}(V_1 \times \cdots \times V_N, W)$, and assume that $(T_\alpha)$ converges pointwise on $E_1 \times \cdots \times E_N$ to $T$. Then $(T_\alpha)$ converges pointwise on $V_1 \times \cdots \times V_N$ to $T$.*

*Proof.* Let $U_i = \mathrm{Span}(E_i)$, which is dense in $V_i$. Then $(T_\alpha)$ converges pointwise on $U_1 \times \cdots \times U_N$ to $T$.

Choose any $\xi_i \in V_i$. Choose $R \in \mathbb{R}_{>0}$ such that $\|\xi_i\| \leqslant R$ for each $i$. Since $\sup_\alpha \|T_\alpha\| < +\infty$, by Prop. 2.22, $\{T_\alpha, T : \alpha \in I\}$ has a uniform Lipschitz constant $C \in \mathbb{R}_{\geqslant 0}$ (with respect to the $l^\infty$-product metric) when restricted to $\overline{B}_{V_1}(0,R) \times \cdots \times \overline{B}_{V_N}(0,R)$. By Rem. 2.27, for each $\varepsilon > 0$, there exists $v_i \in \overline{B}_{U_i}(0,R)$ such that $\|\xi_i - v_i\| \leqslant \varepsilon$. Then

$$
\begin{aligned}
&\limsup_\alpha \|T(\xi_1,\dots,\xi_N) - T_\alpha(\xi_1,\dots,\xi_N)\| \\
\leqslant &\limsup_\alpha \|T(v_1,\dots,v_N) - T_\alpha(v_1,\dots,v_N)\| + 2C\varepsilon = 2C\varepsilon
\end{aligned}
$$

Since $\varepsilon$ is arbitrary, we conclude that $T_\alpha(\xi_1,\dots,\xi_N) \to T(\xi_1,\dots,\xi_N)$. □

**Theorem 2.32.** *Suppose that $W$ is complete. For each $i$, let $E_i$ be a densely spanning subset of $V_i$. Let $(T_\alpha)$ be a net in $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ satisfying $\sup_\alpha \|T_\alpha\| < +\infty$. Suppose that $(T_\alpha)$ converges pointwise on $E_1 \times \cdots \times E_N$. Then $(T_\alpha)$ converges pointwise on $V_1 \times \cdots \times V_N$ to some $T \in \mathfrak{L}(V_1 \times \cdots \times V_N, W)$, and*

$$
\|T\| \leqslant \liminf_\alpha \|T_\alpha\| \tag{2.24}
$$

Inequality (2.24) is sometimes referred to as **Fatou's lemma**.

*Proof.* Let $U_i = \mathrm{Span}(E_i)$, which is dense in $V_i$. Let $T : U_1 \times \cdots \times U_N \to W$ be the pointwise limit of $(T_\alpha)_{\alpha \in I}$ restricted to $U_1 \times \cdots \times U_N$, which is clearly linear. Moreover, for each $v_i \in \overline{B}_{U_i}(0,1)$ we have

$$
\|T(v_1,\dots,v_N)\| = \liminf_\alpha \|T_\alpha(v_1,\dots,v_N)\| \leqslant \liminf_\alpha \|T_\alpha\|
$$

Taking sup over all $v_i \in \overline{B}_{U_i}(0,1)$, we see that $\|T\| \leqslant \sup_\alpha \|T_\alpha\| < +\infty$. In particular, $T \in \mathfrak{L}(U_1 \times \cdots \times U_N, W)$. By Thm. 2.28, $T$ can be extended to a bounded multilinear map $T : V_1 \times \cdots \times V_N \to W$ with $\|T\|$ unchanged. By Prop. 2.31, this extended $T$ is the pointwise limit of $(T_\alpha)$ on the whole domain $V_1 \times \cdots \times V_N$. □

**Remark 2.33.** Recall that if $X$ is a set, then $l^\infty(X, W)$, equipped with the $l^\infty$-norm, is a normed vector space.

By the definition of operator norms, we have a linear isometry of normed vector spaces

$$
\begin{gathered}
\mathfrak{L}(V_1 \times \cdots \times V_N, W) \to l^\infty(\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1), W) \\
T \mapsto T|_{\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1)}
\end{gathered}
\tag{2.25}
$$

Therefore, by identifying $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ with its image under (2.25), we view $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ as a normed vector subspace of $l^\infty(\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}, W)$.

Consequently, if $(T_\alpha)$ is a net in $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$, and if $T \in \mathfrak{L}(V_1 \times \cdots \times V_N, W)$, then $\lim_\alpha \|T - T_\alpha\| = 0$ is equivalent to that $(T_\alpha)$ converges uniformly to $T$ on $\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1)$. □

**Theorem 2.34.** *$\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ is a closed linear subspace of $l^\infty(\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1), W)$.*

*Proof.* Let $T \in l^\infty(\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1), W)$ be the limit of a sequence $(T_n)$ in $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$. Then $(T_n)$ converges uniformly on $\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1)$ to $T$. By scaling the vectors, we see that $(T_n)$ converges uniformly on $\overline{B}_{V_1}(0,R) \times \cdots \times \overline{B}_{V_N}(0,R)$ for any $R > 0$. Let $T : V_1 \times \cdots \times V_N \to W$ be the pointwise limit of $(T_n)$, which automatically extends the original $T$ defined on $\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}(0,1)$.

Since each $T_n$ is multilinear, clearly $T$ is multilinear. Thus $T \in \mathfrak{L}(V_1 \times \cdots \times V_N, W)$. This proves that $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ is a closed. □

**Corollary 2.35.** *Suppose that $W$ is complete. Then $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ is complete.*

*Proof.* Since $W$ is complete, by the following Prop. 2.36, $l^\infty(\overline{B}_{V_1}(0,1) \times \cdots \times \overline{B}_{V_N}, W)$ is complete. Since any closed subset of a complete space is complete, by Thm. 2.34, $\mathfrak{L}(V_1 \times \cdots \times V_N, W)$ is complete. □

**Proposition 2.36.** *Suppose that $W$ is complete. Then for each $1 \leqslant p \leqslant +\infty$, the normed vector space $l^p(X, W)$ is complete.*

*Proof.* Let $(f_n)$ be a Cauchy sequence in $l^p(X, W)$. Then for each $x \in X$, $(f_n(x))$ is a Cauchy sequence in $W$, and hence converges to some $f(x) \in W$. This defines $f : X \to W$.

Case $p = +\infty$: For each $\varepsilon > 0$, choose $N \in \mathbb{Z}_+$ such that for all $m, n \geqslant N$ we have $\|f_n - f_m\|_{l^\infty} \leqslant \varepsilon$, i.e., $\|f_n(x) - f_m(x)\| \leqslant \varepsilon$ for every $x \in X$. Applying $\lim_{m\to\infty}$, we get $\|f_n(x) - f(x)\| \leqslant \varepsilon$ for all $x \in X$ and $n \geqslant N$. Thus, for all $n \geqslant N$ we have $\|f_n - f\|_{l^\infty} \leqslant \varepsilon$; in particular, we have $f \in l^\infty(X, W)$. Thus $\|f_n - f\|_{l^\infty} \to 0$.

Case $p < +\infty$: For each $\varepsilon > 0$, choose $N \in \mathbb{Z}_+$ such that for all $m, n \geqslant N$ we have $\|f_n - f_m\|_{l^p(X)} \leqslant \varepsilon$, equivalently, $\|f_n - f_m\|_{l^p(A)} \leqslant \varepsilon$ for each $A \in \mathrm{fin}(2^X)$. Applying $\lim_{m\to\infty}$, we get $\|f_n - f\|_{l^p(A)} \leqslant \varepsilon$ for all $n \geqslant N$ and $A \in \mathrm{fin}(2^X)$. Thus $\|f_n - f\|_{l^p(X)} \leqslant \varepsilon$ for all $n \geqslant N$; in particular, we have $f \in l^p(X, W)$. This proves $\|f_n - f\|_p \to 0$. □

**Corollary 2.37.** *The dual space $V^*$, equipped with the operator norm, is complete.*

*Proof.* This follows immediately from Cor. 2.35. □

## 2.5 The roles of completeness and duality

Let $V_1, \ldots, V_N$ and $V, W$ be normed vector spaces.

### 2.5.1 The role of Cauchy completeness

In functional analysis, Cauchy completeness plays two primary roles:

1. Completeness as a domain property, where it is often used in conjunction with the Baire category theorem.

2. Completeness as a codomain property, which ensures that linear operators can be restricted from the whole space to a dense subspace without loss. Thm. 2.28 and 2.32 are typical examples illustrating this usage.

Among these two, completeness as a codomain is the more widely encountered in practice. This suggests that the recognition and widespread appreciation of **Cauchy completeness** in function spaces developed alongside the study of linear operators—that is, linear maps from $V$ to $W$—rather than with linear, bilinear, or multilinear functionals, such as $V \times W \rightarrow \mathbb{F}$. In the early days of functional analysis, particularly in Hilbert's foundational work [Hil06], the dominant perspective was centered not on linear operators, but on bilinear forms and linear functionals. Within this (bi)linear framework, completeness is not required—indeed, in Thm 2.28, 2.32, and Corollary 2.35, when $W = \mathbb{F}$, none of the remaining vector spaces involved (namely $V_1, \ldots, V_N$) are assumed to be complete.

Historically, the focus on bilinear forms gradually gave way to the linear operator viewpoint. As this shift took place, Cauchy completeness came to occupy a central role in functional analysis. The fact that the bilinear form or multilinear functional viewpoint can be reformulated in terms of linear operators is a consequence of the following elementary observation:

**Proposition 2.38.** *Let $U_1, \ldots, U_M$ be normed vector spaces. Then we have an isomorphism of normed vector spaces*

$$\begin{gathered}\mathfrak{L}(U_1 \times \cdots \times U_M \times V_1 \times \cdots \times V_N, W) \xrightarrow{\simeq} \mathfrak{L}(U_1 \times \cdots \times U_M, \mathfrak{L}(V_1 \times \cdots \times V_N, W)) \\ T \mapsto \Big((u_1, \ldots, u_M) \mapsto T(u_1, \ldots, u_M, -, \ldots, -)\Big)\end{gathered} \tag{2.26}$$

*where $T(u_1, \ldots, u_M, -, \ldots, -)$ denotes the multilinear map $V_1 \times \cdots \times V_N \rightarrow W$ sending $(v_1, \ldots, v_N)$ to $T(u_1, \ldots, u_M, v_1, \ldots, v_N)$.*

*Proof.* It is easy to verify that the second line of (2.26) defines a linear isomorphism

$$\begin{aligned}\Psi :&\mathrm{Lin}(U_1 \times \cdots \times U_M \times V_1 \times \cdots \times V_N, W) \\ &\xrightarrow{\simeq} \mathrm{Lin}(U_1 \times \cdots \times U_M, \mathrm{Lin}(V_1 \times \cdots \times V_N, W))\end{aligned}$$

To explain the idea of comparing the operator norms, we assume for simplicity that $M = N = 1$, and write $U_1 = U$ and $V_1 = V$.

Choose any $T \in \mathrm{Lin}(U \times V, W)$. Then $\Psi(T) : U \to \mathrm{Lin}(V, W)$ sends each $u \in V$ to the linear map

$$\Psi(T)(u) : v \in \mathrm{Lin}(V, W) \mapsto T(u, v)$$

Thus, for each $u \in U$ and $v \in V$, we have

$$\|T(u, v)\| = \|\Psi(T)(u)(v)\| \leqslant \|\Psi(T)(u)\| \cdot \|v\| \leqslant \|\Psi(T)\| \cdot \|u\| \cdot \|v\|$$

This proves $\|T\| \leqslant \|\Psi(T)\|$. Conversely, for each $u \in U$,

$$\begin{aligned} \|\Psi(T)(u)\| &= \sup_{v \in \overline{B}_V(0,1)} \|\Psi(T)(u)(v)\| = \sup_{v \in \overline{B}_V(0,1)} \|T(u, v)\| \\ &\leqslant \sup_{v \in \overline{B}_V(0,1)} \|T\| \cdot \|u\| \cdot \|v\| = \|T\| \cdot \|u\| \end{aligned}$$

This proves $\|\Psi(T)\| \leqslant \|T\|$.

We have proved that $\|\Psi(T)\| = \|T\|$. In particular, if $T$ is bounded, then $\Psi(T)(u)$ is bounded for each $u \in U$, and $\Psi(T)$ is bounded. Conversely, if $\Psi(T)(u)$ is bounded for each $u$, and if $\Psi(T)$ is bounded, then $T$ is bounded. This proves that $\Psi$ restricts to the linear isomorphism (2.26), which is an isometry because $\|\Psi(T)\| = \|T\|$. □

### 2.5.2 The role of duality

The following two corollaries follow immediate from Prop. 2.38.

**Corollary 2.39.** *We have an isomorphism of normed vector spaces*

$$\mathfrak{L}(U \times V, \mathbb{F}) \xrightarrow{\simeq} \mathfrak{L}(U, V^*) \qquad T \mapsto \big(u \mapsto T(u, -)\big) \tag{2.27}$$

**Corollary 2.40.** *Suppose that $V$ is the dual space of another normed vector space $V_*$. Then we have an isomorphism of normed vector spaces*

$$\mathfrak{L}(V \times V_*, \mathbb{F}) \xrightarrow{\simeq} \mathfrak{L}(V) \qquad T \mapsto \big(v \mapsto T(v, -)\big) \tag{2.28}$$

In Sec. 2.1 and 2.2, we explored the motivation for introducing dual spaces from the perspectives of the calculus of variations and moment problems. Cor. 2.40 now offers yet another compelling reason for the study of duality: when a space $V$ possesses a **dual structure**—specifically, when $V$ is the dual of some normed space $V_*$—it allows us to approach problems from both the bilinear form and linear operator perspectives.

What are the respective advantages of these two viewpoints? To address this, I would like to revisit the arguments presented in [Gui-A], particularly in the Introduction and in Ch. 21 and 25 of [Gui-A]:

1. The bilinear form framework allows us to draw upon the full strength of measure theory. In fact, measure theory can be understood as a method of **monotone convergence extension**—a procedure for extending linear functionals in such a way that the monotone convergence theorem (or its variants) holds. This type of extension aligns naturally with the structure of bilinear forms.

2. The space $\mathfrak{L}(V)$ of bounded linear operators on $V$ is not just a vector space but also an algebra, with multiplication given by composition. This algebraic structure enables the use of **symbolic calculus**, a technique developed in the mid-19th century in the study of linear algebras, and it connects directly to the representation-theoretic perspectives that flourished in the 20th century.

As discussed in [Gui-A, Sec. 25.8, 25.9], and as we will also explore in Ch. 4, Riesz's spectral theorem provides a striking example of how these two advantages can be fruitfully combined.

## 2.6 Dual spaces and the weak-* topology

Let $V_1, V_2, \dots, U, V, W$ be normed $\mathbb{F}$-vector spaces.

**Definition 2.41.** By viewing $V^*$ as a subset of $\mathbb{C}^V$, the subspace topology on $V^*$ inherited from the product topology of $\mathbb{C}^V$ is called the **weak-* topology** on $V^*$. By Thm. 1.8, this is the unique topology such that for any net $(\varphi_\alpha)$ in $V^*$ and any $\varphi \in V$, the net $(\varphi_\alpha)$ **converges weak-*** to $\varphi$—that is, converges to $\varphi$ in the weak-* topology—iff

$$\lim_\alpha \langle \varphi_\alpha, v\rangle = \langle \varphi, v\rangle \qquad \text{for any } v \in V \tag{2.29}$$

Since $\mathbb{C}^V$ is Hausdorff, the weak-* topology is also Hausdorff.

Weak-* topology is mainly considered for closed balls of $V^*$, rather than the whole dual space $V^*$, because for such subsets, pointwise convergence of moments is equivalent to weak-* convergence—that is, the second and third columns of Table 2.2 are equivalent. This equivalence is formally stated in the following theorem.

**Theorem 2.42.** *Suppose that $E$ is a densely spanning subset of $V$. Let $(\varphi_\alpha)$ be a net in $V^*$ satisfying $\sup_\alpha \|\varphi_\alpha\| < +\infty$. Then $(\varphi_\alpha)$ converges weak-* in $V^*$ iff the limit $\lim_\alpha \langle \varphi_\alpha, v\rangle$ exists for any $v \in E$.*

*Moreover, if $\varphi \in V^*$ satisfies that*

$$\lim_\alpha \langle \varphi_\alpha, v\rangle = \langle \varphi, v\rangle \qquad \text{for any } v \in E$$

*then $(\varphi_\alpha)$ converges weak-* to $\varphi$.*

*Proof.* This is clear from Prop. 2.31 and Thm. 2.32. $\square$

**Remark 2.43.** Let $U$ be a dense linear subspace of $V$. (For example, take $V = C_0(X,\mathbb{F})$ and $U = C_c(X,\mathbb{F})$.) Recall the canonical isomorphism $V^* \simeq U^*$ given in Cor. 2.29. Then by Prop. 2.42, for each $R \in \mathbb{R}_{\geqslant 0}$, the weak-* topology on $\overline{B}_{V^*}(0,R)$ agrees with the weak-* topology on $\overline{B}_{U^*}(0,R)$. However, the weak-* topology on $V^*$ is in general not equal to the weak-* topology on $U^*$.

In Prop. 2.42, one might further ask whether a net $(\varphi_\alpha)$ in $\overline{B}_{V^*}(0,R)$ that converges weak-* has its limit also in $\overline{B}_{V^*}(0,R)$. The answer is yes:

**Proposition 2.44 (Fatou's lemma for weak-* convergence).** *Let $(\varphi_\alpha)$ be a net in $V^*$ converging weak-* to some $\varphi \in V^*$. Then*

$$\|\varphi\| \leqslant \liminf_\alpha \|\varphi_\alpha\| \tag{2.30}$$

*In other words, the norm function $\|\cdot\| : V^* \to \mathbb{R}_{\geqslant 0}$ is lower semicontinuous with respect to the weak-* topology on $V^*$.*

In contrast, if $(\varphi_\alpha)$ converges in the operator norm to $\varphi$, then $\|\varphi\| = \lim_\alpha \|\varphi_\alpha\|$. Cf. Rem. 2.16.

*Proof.* For each $v \in \overline{B}_V(0,1)$, we have

$$|\langle\varphi,v\rangle| = \lim_\alpha |\langle\varphi_\alpha,v\rangle| = \liminf_\alpha |\langle\varphi_\alpha,v\rangle| \leqslant \liminf_\alpha \|\varphi_\alpha\|\cdot\|v\| = \|\varphi_\alpha\|$$

Applying $\sup_{v\in\overline{B}_V(0,1)}$ to the LHS above yields (2.30). (See also Thm. 2.32.) $\square$

**Theorem 2.45 (Banach-Alaoglu theorem).** *$\overline{B}_{V^*}(0,1)$ is **weak-* compact**—that is, it is compact in the weak-* topology.*

Thus, $\overline{B}_{V^*}(0,1)$ is a compact Hausdorff space.

***First proof.*** Let $(\varphi_\alpha)$ be a net $\overline{B}_{V^*}(0,1)$. Since $|\langle\varphi_\alpha,v\rangle| \leqslant \|v\|$ for each $v \in V$, we can view $(\varphi_\alpha)$ as a net in

$$S = \prod_{v\in V} \overline{B}_{\mathbb{F}}(0,\|v\|)$$

By Tychonoff's Thm. 1.11, $S$ is compact. Therefore, $(\varphi_\alpha)$ has a subnet $(\varphi_{\alpha_\mu})$ converging pointwise on $V$ to some function $\varphi : V \to \mathbb{F}$. The function $\varphi$ is clearly linear and satisfies $\|\varphi\| \leqslant \sup_\mu \|\varphi_{\alpha_\mu}\| \leqslant 1$, cf. Thm. 2.32. Thus $(\varphi_{\alpha_\mu})$ converges weak-* to $\varphi \in \overline{B}_{V^*}(0,1)$. This finishes the proof that $\overline{B}_{V^*}(0,1)$ is compact. $\square$

The above proof relies on Tychonoff's theorem, which in turn relies on Zorn's lemma. When $V$ is separable, one can prove the Banach-Alaoglu theorem without using Zorn's lemma:

*Second proof assuming that $V$ is separable.* Let $E$ be a countable dense subset of $V$. Then

$$\Phi : \overline{B}_{V^*}(0,1) \to \mathbb{F}^E \qquad \varphi \mapsto \varphi|_E$$

is injective. Moreover, if $(\varphi_\alpha)$ is a net in $\overline{B}_{V^*}(0,1)$ and $\varphi \in \overline{B}_{V^*}(0,1)$, then Prop. 2.31 indicates that $(\varphi_\alpha)$ converges weak-* to $\varphi$ iff $(\varphi_\alpha)$ converges pointwise on $E$ to $\varphi$. Therefore, $\Phi$ restricts to a homeomorphism from $\overline{B}_{V^*(0,1)}$ to its image. Thus, since $\mathbb{F}^E$ is metrizable (cf. Prop. 1.10), so is any subset—in particular, $\overline{B}_{V^*}(0,1)$.

Therefore, showing that $\overline{B}_{V^*}(0,1)$ is compact is equivalent to showing that it is sequentially compact. Let $(\varphi_n)$ be a sequence in $\overline{B}_{V^*}(0,1)$. By the diagonal method (cf. Rem. 1.12), $(\varphi_n)$ has a subsequence $(\varphi_{n_k})$ converging pointwise on $E$. Thm. 2.32 now implies that $(\varphi_{n_k})$ converges weak-* to some $\varphi \in \overline{B}_{V^*}(0,1)$. □

The above proof shows that if $V$ is separable, then $\overline{B}_{V^*}(0,1)$ is metrizable and therefore sequentially compact under the weak-* topology. The converse is also true:

**Theorem 2.46.** *The following statements are equivalent.*

*(a) The normed vector space $V$ is separable.*

*(b) When equipped with the weak-* topology, the compact Hausdorff space $\overline{B}_{V^*}(0,1)$ is metrizable.*

*Proof.* (a)⇒(b) has been proved above. Here, we give a more direct argument of the equivalence (a)⇒(b). By the following Lem. 2.47, $V$ can be viewed as a subset of $C(X,\mathbb{F})$ where $X = \overline{B}_{V^*}(0,1)$ is compact by Banach-Alaoglu. Clearly $V$ separates the points of $X$. Therefore, if $V$ is separable, then $X$ is metrizable by (c)⇒(a) of Thm. 1.38. Conversely, if $X$ is metrizable, then $C(X,\mathbb{F})$ is separable the (a)⇒(d) of Thm. 1.38. Therefore, the subset $V$ of $C(X,\mathbb{F})$ is also separable. □

**Lemma 2.47.** *For each $\varphi \in V$, the function*

$$\overline{B}_{V^*}(0,1) \to \mathbb{F} \qquad \varphi \mapsto \langle \varphi, v\rangle$$

*is continuous with respect to the weak-* topology.*

*Proof.* This is clear by (2.29). □

**Remark 2.48.** When $V$ is separable, a metric $d$ generating the weak-* topology of $\overline{B}_{V^*}(0,1)$ can be explicitly given: Let $(v_n)_{n\in\mathbb{Z}_+}$ be a dense sequence in $V$. Replacing $v_n$ with $v_n/\|v_n\|$ if $v_n \neq 0$, we assume that $\|v_n\| \leqslant 1$. Then, by (1.10), the metric $d$ can be chosen to be

$$d(\varphi_1,\varphi_2) = \sum_{n\in\mathbb{Z}_+} 2^{-n}|\varphi_1(v_n) - \varphi_2(v_n)| \qquad \text{for each } \varphi_1,\varphi_2 \in \overline{B}_{V^*}(0,1) \tag{2.31}$$

## 2.7 Weak-* convergence in $L^p$-spaces

Let $(X, \mathfrak{M}, \mu)$ be a $\sigma$-finite measure space.[9] Let $I \subset \mathbb{R}$ be a closed proper interval. Let $1 < p \leqslant +\infty$ and $p^{-1} + q^{-1} = 1$.

We identify $L^p(X, \mu, \mathbb{F})$ with the dual space $L^q(X, \mu, \mathbb{F})^*$ via the isomorphism described in Thm. 1.50. This defines the **weak-* topology on $L^p(X, \mu, \mathbb{F})$** . In particular, a net $(f_\alpha)$ in $L^p(X, \mu, \mathbb{F})$ converges weak-* to $f \in L^p(X, \mu, \mathbb{F})$ iff

$$\lim_\alpha \int_X f_\alpha g d\mu = \int_X f g d\mu \qquad \text{for all } g \in L^q(X, \mu, \mathbb{F})$$

### 2.7.1 Pointwise convergence and weak-* convergence

Let us prove Thm. 1.54 in a slightly more general setting. Note that a finite Borel measure $\mu$ on an interval $I \subset \mathbb{R}$ can be extended by zero to a finite Borel measure on $\mathbb{R}$, which is Radon by Thm. 1.58. Therefore, to generalize Thm. 1.54, it suffices to consider finite Borel (equivalently, finite Radon) measures on $\mathbb{R}$.

**Theorem 2.49.** *Let $\mu$ be a finite Borel measure on $\mathbb{R}$. Let $(f_\alpha)$ be a net in $L^p(\mathbb{R}, \mu, \mathbb{F})$ satisfying $\sup_\alpha \|f_\alpha\|_{L^p} < +\infty$. Then $(f_\alpha)$ converges weak-* to some element $f \in L^p(\mathbb{R}, \mu, \mathbb{F})$ iff the following limit exists for every $x \in \mathbb{R}$:*

$$F(x) := \lim_\alpha \int_{(-\infty, x]} f_\alpha d\mu \tag{2.32}$$

*When $(f_\alpha)$ converges weak-* to $f \in L^p(\mathbb{R}, \mu, \mathbb{F})$, for each $x \in \mathbb{R}$ we have*

$$F(x) = \int_{(-\infty, x]} f d\mu \tag{2.33}$$

Note that since $\mu$ is finite, the constant function $1$ belongs to $L^q$. Therefore, by Hölder's inequality, any function in $L^p(\mathbb{R}, \mu, \mathbb{F})$ is integrable.

*Proof.* First, assume that $(f_\alpha)$ converges weak-* to $f$ in $L^p(\mathbb{R}, \mu, \mathbb{F})$. Then for each $x \in \mathbb{R}$, we have $\lim_\alpha \int f_\alpha \chi_{(-\infty, x]} d\mu = \int f \chi_{(-\infty, x]} d\mu$. This proves that (2.32) exists and (2.33) holds.

Next, we assume that (2.32) exists for every $x$. In the following, we give two proofs of the weak-* convergence of $(f_\alpha)$.

First proof. Let $\varphi_\alpha \in L^q(\mathbb{R}, \mu, \mathbb{F})^*$ be the linear functional associated to $f_\alpha$, i.e., $\langle \varphi_\alpha, g \rangle = \int f_\alpha g d\mu$ for each $g \in L^q$. By assumption, $\varphi_\alpha$ converges when evaluated with any member of

$$\mathcal{E} = \mathrm{Span}_{\mathbb{F}}\{\chi_{(-\infty, x]} : x \in \mathbb{R}\}$$

[9] The condition on $\sigma$-finiteness can be removed at least when $p = 2$. See the paragraph after Thm. 1.50.

By Thm. 1.62, $\mathcal{E}$ is dense in $L^q$. Therefore, since

$$\sup_\alpha \|\varphi_\alpha\| = \sup_\alpha \|f_\alpha\|_p < +\infty$$

by Thm. 2.42, $(\varphi_\alpha)$ converges weak-* to some $\varphi \in (L^q)^*$. By Thm. 1.50, $\varphi$ is represented by some $f \in L^p(\mathbb{R}, \mu, \mathbb{F})$. Thus $(f_\alpha)$ converges weak-* to $f$.

Second proof. In this proof, we use the fact that any bounded closed ball of $L^p(\mathbb{R}, \mu, \mathbb{F})$ is weak-* compact, which is due to Thm. 1.50 and the Banach-Alaoglu theorem.

Since $\sup_\alpha \|f_\alpha\|_p < +\infty$, the net $(f_\alpha)$ has a subnet $(f_{\alpha_\nu})$ converging weak-* to some $f \in L^p$. By the first paragraph, for each $x \in \mathbb{R}$ we have

$$\lim_\nu \int_{(-\infty,x]} f_{\alpha_\nu} d\mu = \int_{(-\infty,x]} f d\mu$$

Since (2.32) converges, we conclude

$$\lim_\alpha \int_{(-\infty,x]} f_\alpha d\mu = \int_{(-\infty,x]} f d\mu$$

That is, if we let $\varphi_\alpha \in (L^q)^*$ represent $f_\alpha$ and let $\varphi \in (L^q)^*$ represent $f$, then $(\varphi_\alpha)$ converges to $\varphi$ when evaluated on $\mathcal{E}$. By Thm. 1.62, $\mathcal{E}$ is dense in $L^q$. Therefore, by Thm. 2.42, $(\varphi_\alpha)$ converges weak-* to $\varphi$. That is, $(f_\alpha)$ converges weak-* to $f$. $\square$

We now present another connection between pointwise convergence and weak-* convergence.

**Theorem 2.50.** *Let $(f_n)$ be a sequence in $L^p(X, \mu, \mathbb{F})$ satisfying $\sup_n \|f_n\|_p < +\infty$. Suppose that $(f_n)$ converges pointwise to $f$. Then $f \in L^p(X, \mu, \mathbb{F})$, and $(f_n)$ converges weak-* to $f$.*

*Proof.* By Fatou's lemma, we have $f \in L^p$, since

$$\int |f|^p \leqslant \liminf_n \int |f_n|^p < +\infty$$

Thm. 2.49 suggests that when $X = \mathbb{R}$ and $\mu$ is a finite Borel measure, to prove that $(f_n)$ converges weak-* to $f$, it suffices to verify that $\lim_n \int_{(-\infty,x]} f_n = \int_{(-\infty,x]} f$ for each $x \in \mathbb{R}$. Motivated by this, we claim that in the general case, it suffices to prove

$$\lim_n \int_E f_n d\mu = \int_E f d\mu \tag{2.34}$$

for each $E \in \mathfrak{M}$ satisfying $\mu(E) < +\infty$. (Note that any $L^p$ function is integrable in $E$ by Hölder's inequality.) Indeed, suppose (2.34) is true. Then, by the density of integrable simple functions in $L^p$ (Thm. 1.47), and by Thm. 2.42, the sequence $(f_n)$ converges weak-* to $f$.

Let us prove (2.34). For each $\lambda \geqslant 0$, let $\alpha_\lambda : \mathbb{R}_{\geqslant 0} \to [0,1]$ be the (continuous) piecewise linear increasing function such that $\alpha_\lambda|_{[0,\lambda]} = 0$ and $\alpha_\lambda|_{[\lambda+1,+\infty)} = 1$. Let $\beta_\lambda = 1 - \alpha_\lambda$. Since $0 \leqslant \alpha_\lambda \leqslant \chi_{[\lambda,+\infty)}$, we have

$$\lambda^{p-1}\alpha_\lambda \leqslant \lambda^{p-1}\chi_{[\lambda,+\infty)} \leqslant x^{p-1}$$

where $x$ denotes the identity function $\mathrm{id} : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$. Hence $\lambda^{p-1}x\alpha_\lambda \leqslant x^p$, which implies $\lambda^{p-1}|f_n|(\alpha_\lambda \circ |f_n|) \leqslant |f_n|^p$. Let $C := \sup_n \int_E |f_n|^p$. Then

$$\lambda^{p-1} \sup_n \int_E |f_n| \cdot (\alpha_\lambda \circ |f_n|) \leqslant \sup_n \int_E |f_n|^p = C < +\infty$$

Therefore, $(f_n)$ is **uniformly integrable** on $E$, which means that for each $\varepsilon > 0$ we have

$$\sup_n \int_E |f_n| \cdot (\alpha_\lambda \circ |f_n|) \leqslant \varepsilon \qquad \text{for sufficiently large } \lambda$$

Since $|f| \cdot (\alpha_\lambda \circ |f|)$ decreases to $0$ as $\lambda \to +\infty$, and since $\int_E |f| < +\infty$ (due to Hölder's inequality), by DCT or MCT,

$$\int_E |f| \cdot (\alpha_\lambda \circ |f|) \leqslant \varepsilon \qquad \text{for sufficiently large } \lambda$$

On the other hand, since $0 \leqslant x\beta_\lambda \leqslant \lambda + 1$, and since $\lim_n \beta_\lambda \circ |f_n|$ converges pointwise to $\beta_\lambda \circ |f|$ (due to the continuity of $\beta_\lambda$), by DCT we have

$$\lim_n \int_E f_n \cdot (\beta_\lambda \circ |f_n|) = \int_E f \cdot (\beta_\lambda \circ |f|)$$

Therefore, since $\alpha_\lambda + \beta_\lambda = 1$,

$$\begin{aligned}\limsup_n \Big| \int_E f_n - \int_E f \Big| \leqslant & \limsup_n \Big| \int_E f_n \cdot (\beta_\lambda \circ |f_n|) - \int_E f \cdot (\beta_\lambda \circ |f|) \Big| \\ & + \limsup_n \int_E |f_n| \cdot (\alpha_\lambda \circ |f_n|) + \int_E |f| \cdot (\alpha_\lambda \circ |f|)\end{aligned}$$

where the RHS is $\leqslant 2\varepsilon$ for sufficiently large $\lambda$. □

### 2.7.2 Weak-* approximation by elementary functions

Let $X$ be an LCH space, and let $\mu$ be a Radon measure (or its completion) on $X$ with $\sigma$-algebra $\mathfrak{M}$. We assume that $\mu$ is $\sigma$-finite. This condition holds, for example, when $X$ is $\sigma$-compact (in particular, when $X$ is second countable; cf. Rem. 1.25.)

In this subsection, we examine Principle 2.11 in the context of $L^p$-spaces. We begin with the following observation:

**Remark 2.51.** Let $V$ be a normed vector space, and let $U$ be a linear subspace of $V^*$. Let $R \in \mathbb{R}_{>0}$. By Rem. 2.27, $U$ is norm-dense in $V^*$ iff $\overline{B}_U(0,R)$ is norm-dense in $\overline{B}_{V^*}(0,R)$.

It is clear from linearity that if $\overline{B}_U(0,R)$ is weak-* dense in $\overline{B}_{V^*}(0,R)$, then $U$ is weak-* dense in $V^*$. However, the weak-* density of $U$ in $V^*$ does not imply the weak-* density of $\overline{B}_U(0,R)$ in $\overline{B}_{V^*}(0,R)$. Therefore, when studying weak-* approximation in $V^*$, we aim—when possible—to approximate any $\varphi \in V^*$ by a net $(\varphi_\alpha)$ in $U$ such that $\|\varphi_\alpha\| \leqslant \|\varphi\|$. This ensures not only convergence but also control of norms. □

**Theorem 2.52.** *The closed unit ball of $C_c(X,\mathbb{F})$ is weak-* dense in the closed unit ball of $L^p(X,\mu,\mathbb{F})$. More precisely, the obvious map $C_c(X,\mathbb{F}) \to L^p(X,\mu,\mathbb{F})$ sends $\overline{B}_{C_c(X,\mathbb{F})}(0,1)$ to a weak-* dense subset of $\overline{B}_{L^p(X,\mu,\mathbb{F})}(0,1)$.*

*Proof.* By Thm. 1.60, if $p < +\infty$, then $\overline{B}_{C_c(X,\mathbb{F})}(0,1)$ is norm-dense in $\overline{B}_{L^p(X,\mu,\mathbb{F})}(0,1)$, and hence also weak-* dense.

Now, we assume $p = +\infty$. let $\mathscr{I}$ be the directed set

$$\begin{gathered}\mathscr{I} = \{(\mathcal{G},\varepsilon) : \mathcal{G} \in \mathrm{fin}(2^{C_c(X,\mathbb{F})}), \varepsilon \in \mathbb{R}_{\geqslant 0}\} \\ (\mathcal{G}_1,\varepsilon_1) \leqslant (\mathcal{G}_2,\varepsilon_2) \qquad \text{means} \qquad \mathcal{G}_1 \subset \mathcal{G}_2, \varepsilon_1 \geqslant \varepsilon_2\end{gathered}$$

Fix any $f \in \overline{B}_{L^\infty(X,\mu,\mathbb{F})}(0,1)$. By adding a $\mu$-a.e. zero function to $f$, we assume that $\|f\|_{l^\infty(X)} = \|f\|_{L^\infty(X,\mu,\mathbb{F})} \leqslant 1$. We claim that for any $(\mathcal{G},\varepsilon) \in \mathscr{I}$, there exists $f_{\mathcal{G},\varepsilon} \in \overline{B}_{C_c(X,\mathbb{F})}(0,1)$ such that

$$\Big|\int_X (f - f_{\mathcal{G},\varepsilon}) g d\mu\Big| \leqslant \varepsilon \qquad \text{for all } g \in \mathcal{G}$$

If this is true, then $(f_{\mathcal{G},\varepsilon})_{(\mathcal{G},\varepsilon)\in\mathscr{I}}$ converges to $f$ when integrated against any element of $C_c(X,\mathbb{F})$. Since $C_c(X,\mathbb{F})$ is dense in $L^1(X,\mu,\mathbb{F})$ (Thm. 1.60), it follows from Thm. 2.42 that $(f_{\mathcal{G},\varepsilon})_{(\mathcal{G},\varepsilon)\in\mathscr{I}}$ converges weak-* to $f$, finishing the proof.

Let us prove the claim. We write $\mathcal{G} = \{g_1,\dots,g_n\}$. Let $A_i = \mathrm{Supp}(g_i)$ and $A = A_1 \cup \cdots \cup A_n$. Since $A$ is compact, we have $\mu(A) < +\infty$. Let $M = \|g_1\|_\infty + \cdots + \|g_n\|_\infty$. By Lusin's Thm. 1.59 and the Tietze extension Thm. 1.22, there exist a compact set $K \subset A$ and a function $f_{\mathcal{G},\varepsilon} \in C_c(X,\mathbb{F})$ satisfying

$$f_{\mathcal{G},\varepsilon}|_K = f|_K \qquad \|f_{\mathcal{G},\varepsilon}\|_{l^\infty} = \|f\|_{l^\infty} \qquad \mu(A\backslash K) \leqslant \varepsilon/2M$$

Recall that $\|f\|_{l^\infty} \leqslant 1$. Thus, for each $1 \leqslant i \leqslant n$, we have

$$\begin{aligned}&\Big|\int_X (f - f_{\mathcal{G},\varepsilon}) g_i\Big| = \Big|\int_{A\backslash K} (f - f_{\mathcal{G},\varepsilon}) g_i\Big| \leqslant M \int_{A\backslash K} (|f| + |f_{\mathcal{G},\varepsilon}|) \\ \leqslant& 2M \cdot \mu(A\backslash K) \leqslant \varepsilon\end{aligned}$$

□

**Corollary 2.53.** *Let $\mu$ be a finite Borel measure on $\mathbb{S}^1$. Let $U = \mathrm{Span}_{\mathbb{C}}\{e_n : n \in \mathbb{Z}\}$ where $e_n : z \in \mathbb{S}^1 \mapsto z^n \in \mathbb{C}$. Then for each $f \in L^p(\mathbb{S}^1, \mu)$, there exists a sequence $(f_n)$ in $U$ converging weak-* to $f$ and satisfying $\sup_n \|f\|_{L^p} \leqslant \|f\|_{L^p}$.*

*Proof.* By Thm. 1.63, the normed vector space $V = L^q(\mathbb{S}^1, \mu)$ is separable. Therefore, by Thm. 2.46, the weak-* topology of $\overline{B}_{L^q(\mathbb{S}^1,\mu)}(0,1)$ is metrizable. Therefore, to prove the corollary, it suffices to show that $\overline{B}_U(0,1)$ is weak-* dense in $\overline{B}_{L^q(\mathbb{S}^1,\mu)}(0,1)$.

By Thm. 2.52, $\overline{B}_{C(\mathbb{S}^1)}(0,1)$ is weak-* dense in $\overline{B}_{L^q(\mathbb{S}^1,\mu)}(0,1)$. By the Stone-Weierstrass Thm. 1.37, $U$ is $l^\infty$-dense (and hence $L^p$-dense) in $C(\mathbb{S}^1)$. Thus, $\overline{B}_U(0,1)$ is $L^p$-norm-dense (and hence weak-* dense) in $\overline{B}_{C(\mathbb{S}^1)}(0,1)$. This finishes the proof. $\square$

## 2.8 Weak-* convergence in $l^p$-spaces

Let $X$ be a set, and let $1 \leqslant p \leqslant +\infty$ and $p^{-1} + q^{-1} = 1$. In this section, we prove the equivalence of the first two columns of Table 2.2 for $V = L^q(X, \mathbb{F})$, cf. Thm. 2.58. The most important case is when $X$ is countable and $p = q = 2$. For example, $l^2(\mathbb{Z}^n)$ corresponds to the space of Fourier coefficients of $L^2$-functions on $\mathbb{T}^n := (\mathbb{S}^1)^n$.

### 2.8.1 The linear isometry $l^p(X,\mathbb{F}) \to l^q(X,\mathbb{F})^*$

**Proposition 2.54.** *Assume that $1 \leqslant p < +\infty$. Then $C_c(X,\mathbb{F})$ is dense in $l^p(X,\mathbb{F})$, where*

$$C_c(X,\mathbb{F}) := \{f \in \mathbb{F}^X : \mathrm{Supp}(f) \text{ is a finite set}\} \tag{2.35}$$

The notation of $C_c(X,\mathbb{F})$ in (2.35) is compatible with our usual notation for LCH spaces if $X$ is equipped with the discrete topology $\mathcal{T}_X = 2^X$.

*Proof.* Choose $f \in l^p(X,\mathbb{F})$. Then, since

$$\lim_{A \in \mathrm{fin}(2^X)} \sum_A |f|^p = \sum_X |f|^p$$

we have

$$\lim_{A \in \mathrm{fin}(2^X)} \|f - f\chi_A\|_{l^p}^p = \lim_{A \in \mathrm{fin}(2^X)} \sum_{X \backslash A} |f| = \sum_X |f|^p - \lim_{A \in \mathrm{fin}(2^X)} \sum_A |f|^p = 0$$

Thus, $(f\chi_A)_{A \in \mathrm{fin}(2^X)}$ is a net in $C_c(X,\mathbb{F})$ converging to $f$. $\square$

**Remark 2.55.** We have a linear map

$$\begin{aligned} \Psi : l^p(X,\mathbb{F}) &\to l^q(X,\mathbb{F})^* \\ f &\mapsto \Big(g \in l^q(X,\mathbb{F}) \mapsto \sum_{x \in X} f(x)g(x)\Big) \end{aligned} \tag{2.36}$$

Indeed, by Hölder's inequality, for each $A \in \mathrm{fin}(2^X)$,

$$\Big|\sum_A fg\Big| \leqslant \sum_A |fg| \leqslant \|f\|_{l^p(A)} \cdot \|g\|_{l^q(X)} \leqslant \|f\|_{l^p(X)} \cdot \|g\|_{l^q(X)}$$

Applying $\lim_A$, we see that $\sum_X fg$ is absolutely convergence (i.e. $\sum_X |fg| < +\infty$), and

$$\Big|\sum_X fg\Big| \leqslant \sum_X |fg| \leqslant \|f\|_{l^p(X)} \cdot \|g\|_{l^q(X)}$$

This justifies the claim that $\Psi$ has range in $l^q(X,\mathbb{F})^*$ (rather than just in $\mathrm{Lin}(l^q(X,\mathbb{F}),\mathbb{F})$), and that $\|\Psi\| \leqslant 1$.

**Proposition 2.56.** *The map $\Psi$ in (2.36) is a linear isometry.*

*Proof.* We already know $\|\Psi\| \leqslant 1$, and we want to show $\|\Psi\| = 1$.

Case $p < +\infty$: By Prop. 2.54 and Thm. 2.28, we have $\|\Psi\| = \|\Psi|_{C_c(X,\mathbb{F})}\|$. Therefore, it suffices to show that $\|\Psi(f)\| = \|f\|$ for each $f \in C_c(X,\mathbb{F})$. We assume WLOG that $f \neq 0$. Then

$$\langle \Psi(f), g\rangle = \|f\|_{l^p} \cdot \|g\|_{l^q}$$

if we write $f = u|f|$ (where $u : X \to \mathbb{S}^1$) and let $g = \overline{u} \cdot |f|^{p-1}$. Since $\|\Psi(f)\| \cdot \|g\|_{l^q} \geqslant |\langle \Psi(f), g\rangle|$ and $\|g\|_{l^q} > 0$, we conclude that $\|\Psi(f)\| \geqslant \|f\|_{l^p}$, and hence $\|\Psi(f)\| = \|f\|_{l^p}$.

Case $p = +\infty$: For each $0 \leqslant \lambda < 1$, let $x \in X$ such that $|f(x)| \geqslant \lambda\|f\|_{l^\infty}$. Take $g = \chi_{\{x\}}$. Then

$$\langle \Psi(f), g\rangle = \lambda\|f\|_{l^p} \cdot \|g\|_{l^q}$$

and hence $\|\Psi(f)\| \geqslant \lambda\|f\|_{l^p}$. Since $\lambda$ is arbitrary, we conclude $\|\Psi(f)\| = \|f\|_{l^p}$. □

### 2.8.2 Weak-* convergence in $l^p(X,\mathbb{F})$

**Definition 2.57.** Assume that $1 < p \leqslant +\infty$. The **weak-* topology on $l^p(X,\mathbb{F})$** is defined to be the pullback topology via the (injective) map $\Phi : l^p(X,\mathbb{F}) \to l^q(X,\mathbb{F})^*$ of the weak-* topology of $l^q(X,\mathbb{F})^*$. In other words, a net $(f_\alpha)$ in $l^p(X,\mathbb{F})$ converges weak-* to $f \in l^p(X,\mathbb{F})$ iff for each $g \in l^q(X,\mathbb{F})$ we have

$$\lim_\alpha \sum_X f_\alpha g = \sum_X fg \tag{2.37}$$

**Theorem 2.58.** *Assume $1 < p \leqslant +\infty$. Let $(f_\alpha)$ be a net in $L^p(X,\mathbb{F})$ satisfying $\sup_\alpha \|f_\alpha\|_{l^p} < +\infty$. Then $(f_\alpha)$ converges weak-* to some $f \in l^p(X,\mathbb{F})$ iff $\lim_\alpha f_\alpha(x)$ converges for each $x \in X$.*

*Moreover, if $(f_\alpha)$ converges weak-* to $f$, then $f(x) = \lim_\alpha f_\alpha(x)$ for each $x \in X$.*

Consequently, if $p > 1$ and $(f_\alpha)$ is a uniformly $l^p$-bounded net in $L^p(X,\mathbb{F})$ converging pointwise to $f : X \to \mathbb{F}$, then $f \in l^p(X,\mathbb{F})$. (Indeed, by Thm. 2.58, $(f_\alpha)$ converges weak-* to some $\widetilde{f} \in l^p(X,\mathbb{F})$, and $\widetilde{f}$ is the pointwise limit of $(f_\alpha)$. Therefore $f = \widetilde{f}$ belongs to $l^p(X,\mathbb{F})$.)

However, as we will see below, this conclusion must in fact be established first in order to complete the proof of Thm. 2.58

*Proof.* First, assume that $(f_\alpha)$ converges weak-* to $f \in l^p(X,\mathbb{F})$. Applying (2.37) to $g = \chi_{\{x\}}$ (for each $x \in X$), we see that $(f_\alpha)$ converges pointwise to $f$.

Conversely, assume that $(f_\alpha)$ converges pointwise on $X$. Let $f \in \mathbb{F}^X$ be the pointwise limit of $(f_\alpha)$. Recall that $C = \sup_\alpha \|f_\alpha\|_{l^p}$ is finite. We claim that $f \in l^p(X,\mathbb{F})$. Indeed, if $p = +\infty$, then for each $x \in X$, we have

$$|f(x)| = \lim_\alpha |f_\alpha(x)| \leqslant \sup_\alpha \|f_\alpha\|_{l^\infty} < +\infty$$

If $p < +\infty$, then for each $A \in \mathrm{fin}(2^X)$,

$$\sum_A |f|^p = \lim_\alpha \sum_A |f_\alpha|^p \leqslant \sup_\alpha \|f_\alpha\|_{l^p}^p \leqslant C^p$$

Applying $\lim_A$, we see that $\sum_X |f|^p \leqslant C^p$, and hence $f \in l^p(X,\mathbb{F})$.

Let $\Psi$ be as in (2.36). By Prop. 2.54, $C_c(X,\mathbb{F})$ is dense in $L^q(X,\mathbb{F})$. Therefore, to show that $(f_\alpha)$ converges weak-* to $f$, by Thm. 2.42 and the observation that

$$\sup_\alpha \|\Psi(f_\alpha)\| = \sup_\alpha \|f_\alpha\|_{l^p} < +\infty$$

it suffices to show that $\langle \Psi(f_\alpha), g\rangle$ converges to $\langle \Psi(f), g\rangle$ (that is, $\sum f_\alpha g$ converges to $\sum fg$) for each $g \in C_c(X,\mathbb{F})$. But this follows from the fact that $(f_\alpha)$ converges pointwise to $f$. $\square$

As an application of Thm. 2.58, we prove a variant of Prop. 2.54.

**Proposition 2.59.** *Let $1 < p \leqslant +\infty$. Then $\overline{B}_{C_c(X,\mathbb{F})}(0,1)$ is weak-* dense in $\overline{B}_{l^\infty(X,\mathbb{F})}$.*

*Proof.* Let $f \in \overline{B}_{l^\infty(X,\mathbb{F})}$. Then $(f\chi_A)_{A\in\mathrm{fin}(2^X)}$ is a net in $\overline{B}_{C_c(X,\mathbb{F})}(0,1)$ converging pointwise to $f$. By Thm. 2.58, this net converges weak-* to $f$. $\square$

### 2.8.3 The isomorphism $l^p(X,\mathbb{F}) \simeq l^q(X,\mathbb{F})^*$

Now that the equivalence of the first two columns of Table 2.2 for $V = L^q(X,\mathbb{F})$ has been established in Thm. 2.58 for $p > 1$, we can prove the isomorphism $l^p(X,\mathbb{F}) \simeq l^q(X,\mathbb{F})^*$ by following the strategy outlined in Rem. 2.12.

Of course, at least when $X$ is countable, this isomorphism is a special case of the duality $L^p(X,\mu,\mathbb{F}) \simeq L^q(X,\mu,\mathbb{F})^*$ from Thm. 1.50, by taking $\mu : 2^X \to [0,+\infty]$

to be the counting measure. However, there are good reasons to study the proof of $l^q(X,\mathbb{F})^* \simeq l^p(X,\mathbb{F})$ independently.

First, the proof of Thm. 1.50 is significantly more involved than the direct proof in the $l^p$ setting. Whenever a result admits a simpler proof in a special case, it is worthwhile to examine that proof directly. Second, Thm. 1.50 depends crucially on the Radon–Nikodym Thm. 1.46, which in turn can be derived from the Riesz-Fréchet Theorem. The latter can be proved with the help of the isomorphism $l^2(X,\mathbb{F}) \simeq l^2(X,\mathbb{F})^*$. Third, since the proof below follows the idea in Rem. 2.12, it also serves as another concrete illustration of Table 2.3.

**Theorem 2.60.** *Assume that $1 < p \leqslant +\infty$. Then the map $\Psi : l^p(X,\mathbb{F}) \to l^q(X,\mathbb{F})^*$ is an isomorphism of normed vector spaces.*

*Proof.* By Prop. 2.56, it remains to show that $\Psi$ is surjective. Choose $\varphi \in l^q(X,\mathbb{F})^*$. We want to find $f \in l^p(X,\mathbb{F})$ such that $\Psi(f) = \varphi$.

Step 1. In this step, we verify Principle 2.11, which says in the current setting that $\varphi$ can be weak-* approximated by a uniformly bounded net in $C_c(X,\mathbb{F})$.

For each $A \in \mathrm{fin}(2^X)$, let $\Psi_A : l^p(A,\mathbb{F}) \to l^q(A,\mathbb{F})^*$ be defined as in (2.36), which is a linear isometry by Prop. 2.56. Moreover, since $l^p(A,\mathbb{F})$ and $l^q(A,\mathbb{F})^*$ both have dimension $|A|$, $\Psi_A$ is an isomorphism. Therefore, there exists $f_A \in C_c(X,\mathbb{F})$, supported in $A$, such that

$$\Psi_A(f_A) = \varphi|_{l^q(A,\mathbb{F})}$$

This relation clearly shows that

$$\lim_{A\in\mathrm{fin}(2^X)} \langle \Psi(f_A), g\rangle = \langle \varphi, g\rangle$$

for each $g$ of the form $\chi_{\{x\}}$ where $x \in X$, and hence for each $g \in C_c(X,\mathbb{F})$. Moreover, the net $(\Psi(f_A))_{A\in\mathrm{fin}(2^X)}$ is uniformly bounded, since

$$\|\Psi(f_A)\|_{l^p} = \|\varphi|_{l^q(A,\mathbb{F})}\| \leqslant \|\varphi\|$$

Therefore, since $C_c(X,\mathbb{F})$ is dense in $l^q(X,\mathbb{F})$ (cf. Prop. 2.54), by Thm. 2.42, the net $(\Psi(f_A))_{A\in\mathrm{fin}(2^X)}$ converges weak-* to $\varphi$. In other words, $(f_A)_{A\in\mathrm{fin}(2^X)}$ is a uniformly $l^p$-bounded net in $C_c(X,\mathbb{F})$ converging weak-* to $\varphi$.

Step 2. For each $x \in X$, the limit

$$\lim_{A\in\mathrm{fin}(2^X)} f_A(x) = \lim_{A\in\mathrm{fin}(2^X)} \sum_X f_A\chi_{\{x\}}$$

converges by the weak-* convergence of $(f_A)_{A\in\mathrm{fin}(2^X)}$. Therefore, since $(f_A)_{A\in\mathrm{fin}(2^X)}$ is a uniformly bounded, by Thm. 2.58, the net $(f_A)_{A\in\mathrm{fin}(2^X)}$ converge weak-* to some $f \in l^p(X,\mathbb{F})$. Thus $\varphi = \Psi(f)$. □

## 2.9 Weak-* convergence of distribution functions

In this section, we fix a proper interval $I \subset \mathbb{R}$, and let $a = \inf I, b = \sup I$. We use freely the notation in Subsec. 1.6.1. In particular, for each function $\rho$ on $I$, we let

$$\Omega_\rho = \{x \in (a,b) : \rho|_{(a,b)} \text{ is continuous at } x\}$$

A family of functions $(\rho_\alpha)$ from $I$ to $\mathbb{R}$ is called **uniformly bounded** if $\sup_\alpha \|\rho_\alpha\|_{l^\infty(I,\mathbb{R})} < +\infty$.

The goal of this section is to prove Thm. 2.10, which characterizes the relationship between pointwise convergence and weak-* convergence for increasing functions. To this end, we begin with several preparatory results concerning the pointwise convergence of such functions.

### 2.9.1 Almost convergence of increasing functions

**Lemma 2.61.** *Let $(\rho_\alpha)$ be a uniformly bounded net of increasing functions $I \to \mathbb{R}_{\geqslant 0}$. Suppose that $(\rho_\alpha)$ converges pointwise on a dense subset $E \subset I$. Then there exists a bounded increasing function $\rho : I \to \mathbb{R}_{\geqslant 0}$ such that $(\rho_\alpha)$ converges pointwise on $E$ to $\rho$.*

*Proof.* Let $\rho : E \to \mathbb{R}_{\geqslant 0}$ be the pointwise limit of $(\rho_\alpha)$, which is clearly bounded and increasing. Extend $\rho$ to a function $\rho : I_{<b} \cup (E \cap \{b\}) \to \mathbb{R}_{\geqslant 0}$ by setting

$$\rho(x) = \lim_{E \ni y \to x^+} \rho(y)$$

if $x \in I \backslash E$. Extend $\rho$ further to $\rho : I \to \mathbb{R}_{\geqslant 0}$ by setting $\rho(b) = \lim_{x \to b^-} \rho(x)$ if $b \in I \backslash E$. Then $\rho$ is bounded and increasing, and $(\rho_\alpha)$ converges pointwise to $\rho$ on $E$. □

**Proposition 2.62.** *Let $(\rho_\alpha)$ be a uniformly bounded net of increasing functions $I \to \mathbb{R}_{\geqslant 0}$. Let $\rho : I \to \mathbb{R}_{\geqslant 0}$ be increasing. Then the following are equivalent:*

*(a) There exists a dense subset $E \subset I$ such that $(\rho_\alpha)$ converges pointwise on $E$ to $\rho$.*

*(b) The net $(\rho_\alpha)$ converges pointwise on $\Omega_\rho$ to $\rho$.*

*If either of these two statements are true, we say that $(\rho_\alpha)$ **almost converges** to $\rho$.*

*Proof.* Since $\Omega_\rho$ is dense (Prop. 1.68), clearly (b) implies (a).

Now assume (a). Choose any $x \in \Omega_\rho$. We will show that every convergent subnet $(\rho_{\alpha_\nu}(x))$ of $(\rho_\alpha(x))$ converges to $\rho(x)$. This will immediately imply (b).

By Lem. 2.61, there exists an increasing function $\widetilde{\rho} : I \to \mathbb{R}_{\geqslant 0}$ such that $(\rho_{\alpha_\nu})$ converges on $E \cup \{x\}$ to $\widetilde{\rho}$. Since $(\rho_{\alpha_\nu})$ converges pointwise on $E$ to $\rho$, the functions $\rho$ and $\widetilde{\rho}$ agree on $E$. Namely, $\rho$ and $\widetilde{\rho}$ are almost equal. Therefore, by Prop. 1.70, $\rho$ and $\widetilde{\rho}$ agree on $\Omega_\rho$, and in particular at $x$. This proves $\lim_\nu \rho_{\alpha_\nu}(x) = \rho(x)$. □

The following theorem can be viewed as a concrete manifestation of the Banach-Alaoglu Thm. 2.45 in the setting of $C_c(I)^*$. It will be used in the proof of Thm. 2.65.

**Theorem 2.63 (Helly selection theorem).** *Let $(\rho_\alpha)$ be a uniformly bounded net (resp. sequence) of increasing functions $I \to \mathbb{R}_{\geqslant 0}$. Then $(\rho_\alpha)$ admits a pointwise convergent subnet (resp. subsequence).*

*Proof.* The existence of a pointwise convergent subnet follows directly from the Tychonoff Thm. 1.11. Therefore, let us assume that $(\rho_\alpha)$ is a sequence $(\rho_n)$. Let $E = I \cap \mathbb{Q}$. Then, by the diagonal method (cf. Rem. 1.12), $(\rho_n)$ has a subsequence $(\rho_{n_k})$ converging pointwise on $E$. By Lem. 2.61, there exists a bounded increasing $\rho : I \to \mathbb{R}_{\geqslant 0}$ such that $(\rho_{n_k})$ converges pointwise on $E$ to $\rho$. Therefore, by Prop. 2.62, $(\rho_{n_k})$ converges pointwise on $\Omega_\rho$ to $\rho$. Since $I\backslash\Omega_\rho$ is countable, by the diagonal method again, $(\rho_{n_k})$ has a subsequence converging pointwise on $I\backslash\Omega_\rho$, and hence on $I$. $\square$

### 2.9.2 Almost convergence and weak-* convergence

**Definition 2.64.** Let $(\rho_\alpha)$ be a net in $BV(I,\mathbb{F})$. Let $\rho \in BV(I,\mathbb{F})$. Let $\Lambda_\alpha$ and $\Lambda$ be the elements of $C_c(I,\mathbb{F})^*$ corresponding to $\rho_\alpha$ and $\rho$, respectively, via the Riesz representation Thm. 1.88. We say that the net $(d\rho_\alpha)$ **converges weak-*** to $d\rho$ if $(\Lambda_\alpha)$ converges weak-* to $\Lambda$. Namely, for each $f \in C_c(I,\mathbb{F})$, we have

$$\lim_\alpha \int_I f d\rho_\alpha = \int_I f d\rho \tag{2.38}$$

The following Thm. 2.65 is parallel to Thm. 2.49. However, unlike Thm. 2.49 whose proof relies on the isomorphism $L^p \simeq (L^q)^*$, Thm. 2.65 does not rely on the Riesz representation theorem.

**Theorem 2.65.** *Let $(\rho_\alpha)_{\alpha\in\mathscr{A}}$ be a uniformly bounded net of bounded increasing functions $I \to \mathbb{R}_{\geqslant 0}$. Let $\rho : I \to \mathbb{R}_{\geqslant 0}$ be bounded and increasing. Then the following are equivalent:*

*(a) There exists a bounded family $(\varkappa_\alpha)_{\alpha\in\mathscr{A}}$ in $\mathbb{R}$ (assumed to be zero if $a \in I$) satisfying the following conditions:*

- *$(\rho_\alpha + \varkappa_\alpha)$ almost converges to $\rho$.*
- *$\lim_\alpha(\rho_\alpha(b) + \varkappa_\alpha) = \rho(b)$ if $b \in I$.*

*(b) The net $(d\rho_\alpha)$ converges weak-* to $d\rho$.*

The boundedness of $(\varkappa_\alpha)_{\alpha\in\mathscr{A}}$ means that $\sup_\alpha |\varkappa_\alpha| < +\infty$.

*Proof.* (a)⇒(b): Assume (a). We verify (2.38) for each $f \in C_c(I, \mathbb{F})$, which established (b). Recall from Rem. 1.84 that if $a \notin I$, adding constants to $\rho_\alpha$ and $\rho$ does not affect the values of $\int_I f d\rho_\alpha$ and $\int_I f d\rho$.

Since $(\rho_\alpha)$ is uniformly bounded $(\varkappa_\alpha)$ is bounded, there exists $c \geqslant 0$ such that $\rho_\alpha + \varkappa_\alpha + c \geqslant 0$ for all $\alpha$. Therefore, replacing $\rho_\alpha$ with $\rho_\alpha + \varkappa_\alpha + c$ and $\rho$ with $\rho + c$, we assume that there exists a dense subset $E \subset I$ such that $(\rho_\alpha)$ converges pointwise on $E$ to $\rho$, and that $b \in E$ if $b \in I$.

Choose any $f \in C_c(I, \mathbb{F})$. Choose $u, v \in \mathbb{R}$ satisfying $\mathrm{Supp}_I(f) \subset [u, v] \subset I$, and let $J = [u, v]$. By increasing $v$ if possible, we may assume that $v \in E$. (When $b \in I$, one simply choose $v = b$.)

In the case where $a \in I$, by Lem. 1.80, the values of $\int_J f d\rho_\alpha$ and $\int_J f d\rho$ remain unchanged if we change the values of $\rho_\alpha(a)$ and $\rho(a)$ to $0$. Therefore, we may assume that $\rho_\alpha(a) = \rho(a) = 0$ (so that $a$ can be included to $E$), and we may also choose $u = a$. In the case where $a \notin I$, by the density of $E$, we can slightly decrease $u$ so that $u \in E$. To summarize, whether $a$ or $b$ belongs to $I$ or not, we can assume

$$u, v \in E$$

Since $f$ is uniformly continuous, for each $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| \leqslant \varepsilon$ for each $x, y \in I$ satisfying $|x - y| \leqslant \delta$. Choose a tagged partition

$$(\sigma, \xi_\bullet) = \big(\{a_0 = u < a_1 < \cdots < a_n = v\}, (\xi_1, \ldots, \xi_n)\big)$$

of $J$ with mesh $< \delta$. Since $E$ is dense, by a slight adjustment, we may assume that $a_0, a_1, \ldots, a_n \in E$. This implies

$$\lim_\alpha f(u)\rho_\alpha(u) = f(u)\rho(u) \qquad \lim_\alpha S_{\rho_\alpha}(f, \sigma, \xi_\bullet) = S_\rho(f, \sigma, \xi_\bullet)$$

Therefore, if we let $C = \sup\{\rho_\alpha(v) - \rho_\alpha(u), \rho(v) - \rho(u) : \alpha \in \mathscr{A}\}$, then Rem. 1.75 implies

$$\limsup_\alpha \Big| \int_J f d\rho_\alpha - \int_J f d\rho \Big| \leqslant 2\varepsilon \cdot C$$

This finishes the proof of (2.38).

(b)⇒(a): Assume (b). We first consider the case where $a \notin I$. Fix $t \in \Omega_\rho$, and let

$$\varkappa_\alpha = \rho(t) - \rho_\alpha(t)$$

Then $(\varkappa_\alpha)$ is bounded. Therefore, $(\rho_\alpha + \varkappa_\alpha)$ is uniformly bounded, and hence there exists $c \geqslant 0$ such that $\rho_\alpha + \varkappa_\alpha + c \geqslant 0$ for all $\alpha$. Replacing $\rho_\alpha$ with $\rho_\alpha + c$ and $\rho$ with $\rho + c$, we assume that $\rho_\alpha + \varkappa_\alpha \geqslant 0$ for all $\alpha$. (Of course, we still have $\rho \geqslant 0$.)

Choose any $x \in \Omega_\rho$. To show that $(\rho_\alpha(x) + \varkappa_\alpha)_\alpha$ converges to $\rho(x)$, it suffices to show that every convergent subnet $(\rho_\beta(x) + \varkappa_\beta)_\beta$ converges to $\rho(x)$.

By the Helly selection Thm. 2.63, the net of functions $(\rho_\beta+\varkappa_\beta)_\beta$ has a pointwise convergent subnet $(\rho_\gamma+\varkappa_\gamma)_\gamma$. Let $\widetilde{\rho}: I\to\overline{\mathbb{R}}_{\geqslant 0}$ be the pointwise limit of this subnet, which is clear bounded and increasing. By (a)⇒(b), the net $(d(\rho_\gamma+\varkappa_\gamma))_\gamma$ converges weak-* to $d\widetilde{\rho}$. By assumption, it also converges weak-* to $d\rho$. Therefore, we have $\int_I fd\widetilde{\rho}=\int_I fd\rho$ for each $f\in C_c(I)$.

By Thm. 1.83 (and noting Rem. 1.84), we have

$$\widetilde{\rho}-\lim_{y\to a^+}\widetilde{\rho}(y)=\rho-\lim_{y\to a^+}\rho(y)\qquad\text{on }\Omega_\rho$$

In other words, there exists a constant $c\in\mathbb{R}$ such that

$$\widetilde{\rho}+c=\rho\qquad\text{on }\Omega_\rho \tag{2.39}$$

Since $\rho_\alpha(t)+\varkappa_\alpha=\rho(t)$ is constant over $\alpha$, and since its subnet $(\rho_\gamma(t)+\varkappa_\gamma)_\gamma$ converges to $\widetilde{\rho}(t)$, we conclude $\widetilde{\rho}(t)=\rho(t)$. Therefore, since $t\in\Omega_\rho$, by (2.39), we have $c=0$. Since $x\in\Omega_\rho$, by (2.39), we obtain $\widetilde{\rho}(x)=\rho(x)$. This proves that $(\rho_\gamma(x)+\varkappa_\gamma)_\gamma$ converges to $\rho(x)$, and hence $(\rho_\beta(x)+\varkappa_\beta)_\beta$ converges to $\rho(x)$.

Now consider the case where $a\in I$. We set $\varkappa_\alpha=0$. Similar to the above argument, we choose any $x\in\Omega_\rho$, choose a subnet $\rho_\beta$ converging at $x$, and further choose a subnet $\rho_\gamma$ converging pointwise on $I$ to $\widetilde{\rho}: I\to\overline{\mathbb{R}}_{\geqslant 0}$. By (a)⇒(b), we have $\int_I fd\widetilde{\rho}=\int_I fd\rho$ for each $f\in C_c(I)$. Consequently, Thm. 1.83 implies that $\widetilde{\rho}=\rho$ on $\Omega_\rho$. Since $x\in\Omega_\rho$, we obtain again $\lim_\beta\rho_\beta(x)=\lim_\gamma\rho_\gamma(x)=\widetilde{\rho}(x)=\rho(x)$. Therefore $(\rho_\alpha(x))_\alpha$ converges to $\rho(x)$ for each $x\in\Omega_\rho$. □

**Corollary 2.66.** *Let $(\rho_\alpha)_{\alpha\in\mathscr{A}}$ be a uniformly bounded net of increasing functions $I\to\mathbb{R}_{\geqslant 0}$. Then the following are equivalent:*

*(1) There exists a bounded family $(\varkappa_\alpha)_{\alpha\in\mathscr{A}}$ in $\mathbb{R}$ (assumed to be zero if $a\in I$) such that $(\rho_\alpha+\varkappa_\alpha)$ converges pointwise on a dense subset $E\subset I$, and also at $b$ if $b\in I$.*

*(2) There exists a bounded increasing $\rho: I\to\mathbb{R}_{\geqslant 0}$ such that $(d\rho_\alpha)_{\alpha\in\mathscr{A}}$ converges weak-* to $d\rho$.*

*Proof.* "(2)⇒(1)" follows immediately from Thm. 2.65. Conversely, assume (1). By Lem. 2.61, there exists a bounded increasing $\rho: I\to\mathbb{R}_{\geqslant 0}$ such that $(\rho_\alpha+\varkappa_\alpha)$ converges pointwise on $E\cup\{I\cap\{b\}\}$ to $\rho$. Then Thm. 2.65 implies (2). □

## 2.10 Weak-* approximation of Radon measures by Dirac measures

Fix an LCH space $X$. Recall that we have assumed throughout the notes that $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$. Let

$$\begin{aligned}
\mathcal{RM}(X,\overline{\mathbb{R}}_{\geqslant 0})&=\{\text{Radon measures on }X\}\\
\mathcal{RM}(X,\mathbb{R}_{\geqslant 0})&=\{\text{finite Radon measures on }X\}\\
\mathcal{RM}(X,\mathbb{R})&=\{\text{signed Radon measures on }X\}\\
\mathcal{RM}(X,\mathbb{C})&=\{\text{complex Radon measures on }X\}
\end{aligned}\tag{2.40}$$

which are vectors spaces over $\overline{\mathbb{R}}_{\geqslant 0}, \mathbb{R}_{\geqslant 0}, \mathbb{R}, \mathbb{C}$ respectively. Note the inclusion relation

$$\mathcal{RM}(X, \mathbb{R}_{\geqslant 0}) \subset \mathcal{RM}(X, \overline{\mathbb{R}}_{\geqslant 0}) \qquad \mathcal{RM}(X, \mathbb{R}_{\geqslant 0}) \subset \mathcal{RM}(X, \mathbb{R}) \subset \mathcal{RM}(X, \mathbb{C})$$

Recall that for each $x \in X$, the Dirac measure at $x$ is denoted by $\delta_x$.

The goal of this section is to prove Principle 2.11 for $V = C_c(X, \mathbb{F})$. In this context, elementary functions are understood as linear combinations of Dirac measures. When $X$ is an interval $I \subset \mathbb{R}$, these elementary functions correspond to bounded increasing functions $I \to \mathbb{R}_{\geqslant 0}$ whose ranges are finite sets.

### 2.10.1 Definitions and basic properties

**Definition 2.67.** Recall the $\mathbb{F}$-linear isomorphism

$$\mathcal{RM}(X, \mathbb{F}) \simeq C_c(X, \mathbb{F})^*$$

defined by the Riesz-Markov representation Thm. 1.66. The pullback of the operator norm on $C_c(X, \mathbb{F})^*$ to $\mu \in \mathcal{RM}(X, \mathbb{F})$ is called the **total variation** of $\mu$, and is denoted by $\|\boldsymbol{\mu}\|$. In other words,

$$\|\mu\| = \sup\Big\{\Big|\int f d\mu\Big| : f \in C_c(X, \mathbb{F}), |f| \leqslant 1\Big\}$$

A family of complex Radon measures $(\mu_\alpha)_{\alpha \in \mathscr{A}}$ is called **uniformly bounded** if

$$\sup_{\alpha \in \mathscr{A}} \|\mu_\alpha\| < +\infty$$

The weak-* topology on $C_c(X, \mathbb{F})^*$ defines the **weak-* topology on $\mathcal{RM}(X, \mathbb{F})$**. Thus, if $(\mu_\alpha)$ is a uniformly bounded net in $\mathcal{RM}(X, \mathbb{F})$, and if $\mu \in \mathcal{RM}(X, \mathbb{F})$, then $(\mu_\alpha)$ converges weak-* to $\mu$ [10] iff for each $f \in C_c(X, \mathbb{F})$ we have[11]

$$\lim_\alpha \int_X f d\mu_\alpha = \int_X f d\mu \tag{2.41}$$

**Example 2.68.** By Thm. 1.56, if $\mu \in \mathcal{RM}(X, \mathbb{R}_{\geqslant 0})$, then

$$\|\mu\| = \mu(X)$$

**Example 2.69.** Let $E \subset X$ be a finite set, and let $c : E \to \mathbb{F}$ be a function. Then

$$\Big\|\sum_{x \in E} c(x)\delta_x\Big\| = \sum_{x \in E} |c(x)| \tag{2.42}$$

[10] We also say that $(d\mu_\alpha)$ converges weak-* to $d\mu$.

[11] By Rem. 2.43, this is equivalent to that (2.41) holds for each $f \in C_0(X, \mathbb{F})$.

*Proof.* Let $\mu = \sum_{x \in E} c(x)\delta_x$. By Exp. 2.68, we have $\|\delta_x\| = 1$. Since norms satisfy the sub-additivity, we have

$$\|\mu\| \leqslant \sum_{x \in E} |c(x)| \cdot \|\delta_x\| = \sum_{x \in E} |c(x)|$$

By Urysohn's lemma, there exists $f \in C_c(X, \mathbb{F})$ such that $\|f\|_{l^\infty} \leqslant 1$, and that for each $x \in E$, we have $|f(x)| = 1$ and $f(x)c(x) = |c(x)|$. Then $\int_X f d\mu = \sum_{x \in E} |c(x)|$. This proves $\|\mu\| \geqslant \sum_{x \in E} |c(x)|$. □

**Lemma 2.70.** *Let $\mu \in \mathcal{RM}(X, \mathbb{F})$. Let $A_1, \ldots, A_k$ be mutually disjoint Borel subsets of $X$. Then*

$$\|\mu\| \geqslant \sum_{j=1}^{k} |\mu(A_j)|$$

*Proof.* Since $\mu$ is a linear combination of finite Radon measures, there exists $\widehat{\mu} \in \mathcal{RM}(X, \mathbb{R}_{\geqslant 0})$ such that $|\mu(A)| \leqslant \widehat{\mu}(A)$ for each Borel $A \subset X$. Since Radon measures are regular on Borel sets with finite measures (Thm. 1.57), for each $\varepsilon > 0$ there exists compact $K_j \subset A_j$ such that $\widehat{\mu}(A_j \backslash K_j) \leqslant \varepsilon$.

By Cor. 1.18, there exist mutually disjoint open subsets $U_1, \ldots, U_n \subset X$ such that $U_j \supset K_j$. Since $\widehat{\mu}$ is regular on $K_j$, we may assume that $\widehat{\mu}(U_j \backslash K_j) < \varepsilon$. By Urysohn's lemma, there exists $f_j \in C_c(U_j, \mathbb{F})$ such that $|f_j| \leqslant 1$, that $f_j|_{K_j}$ equals a constant $c_j \in \mathbb{F}$, and that $c_j \mu(K_j) = |\mu(K_j)|$. Let $f = f_1 + \cdots f_k$, which is an element of $C_c(X, \mathbb{F})$ satisfying $|f| \leqslant 1$. Then

$$\int_{\bigcup_j K_j} f d\mu = \sum_j |\mu(K_j)| \qquad \Big| \int_{X \backslash \bigcup_j K_j} f d\mu \Big| \leqslant k\varepsilon$$

Since $|\mu(A_j) - \mu(K_j)| = |\mu(A_j \backslash K_j)| \leqslant \widehat{\mu}(A_j \backslash K_j) \leqslant \varepsilon$, we obtain $|\mu(K_j)| \geqslant |\mu(A_j)| - \varepsilon$, and hence

$$\|\mu\| \geqslant \Big| \int_X f d\mu \Big| \geqslant \Big| \int_{\bigcup_j K_j} f d\mu \Big| - \Big| \int_{X \backslash \bigcup_j K_j} f d\mu \Big| \geqslant \sum_j |\mu(A_j)| - 2k\varepsilon$$

Since $\varepsilon$ is arbitrary, we obtain the desired inequality. □

### 2.10.2 Approximation of Radon measures by Dirac measures

In this section, we let $\mathbb{K} \in \{\mathbb{R}_{\geqslant 0}, \mathbb{R}, \mathbb{C}\}$.

**Theorem 2.71.** *Define*

$$\mathcal{D}(X, \mathbb{K}) = \mathrm{Span}_{\mathbb{K}}\{\delta_x : x \in X\}$$

*Then the closed unit ball of $\mathcal{D}(X, \mathbb{K})$ is weak-* dense in the closed unit ball of $\mathcal{RM}(X, \mathbb{K})$. In other words, $\overline{B}_{\mathcal{D}(X,\mathbb{K})}(0, 1)$ is weak-* dense in $\overline{B}_{\mathcal{RM}(X,\mathbb{K})}(0, 1)$.*

*Proof.* Fix $\mu \in \mathcal{RM}(X, \mathbb{K})$ satisfying $\|\mu\| \leqslant 1$. Similar to the proof of Thm. 2.52, we let $\mathscr{I}$ be the directed set

$$\begin{gathered}\mathscr{I} = \{(\mathcal{G}, \varepsilon) : \mathcal{G} \in \mathrm{fin}(2^{C_c(X,\mathbb{K})}), \varepsilon \in \mathbb{R}_{\geqslant 0}\}\\ (\mathcal{G}_1, \varepsilon_1) \leqslant (\mathcal{G}_2, \varepsilon_2) \qquad \text{means} \qquad \mathcal{G}_1 \subset \mathcal{G}_2, \varepsilon_1 \geqslant \varepsilon_2\end{gathered}$$

We claim that for any $(\mathcal{G}, \varepsilon) \in \mathscr{I}$, there exists $\mu_{\mathcal{G},\varepsilon} \in \overline{B}_{\mathcal{D}(X,\mathbb{K})}(0, 1)$ such that

$$\Big|\int_X f d\mu - \int_X f d\mu_{\mathcal{G},\varepsilon}\Big| \leqslant \varepsilon \qquad \text{for all } f \in \mathcal{G}$$

If this is true, then $(\mu_{\mathcal{G},\varepsilon})_{(\mathcal{G},\varepsilon)\in\mathscr{I}}$ is a net in $\overline{B}_{\mathcal{D}(X,\mathbb{K})}(0, 1)$ converging weak-* to $\mu$. This will finish the proof.

Let us prove the claim. Since $\mu$ is a linear combination of finite Radon measures, there exists $\widehat{\mu} \in \mathcal{RM}(X, \mathbb{R}_{\geqslant 0})$ such that

$$\Big|\int_X g d\mu\Big| \leqslant \int_X |g| d\widehat{\mu}$$

for each bounded Borel function $g : X \to \mathbb{C}$.

Let $K \subset X$ be compact and containing $\mathrm{Supp}(f)$ for all $f \in \mathcal{G}$. By the compactness of $K$, there exist open sets $U_1, \dots, U_k \subset X$ whose union contains $K$, such that $\mathrm{diam}(f(U_j)) \leqslant \varepsilon/\widehat{\mu}(K)$ for each $j$ and $f \in \mathcal{G}$. Choose a Borel set $A_j \subset U_j$ such that $K = A_1 \sqcup \cdots \sqcup A_k$.[12] Choose any $x_j \in A_j$, and let

$$\mu_{\mathcal{G},\varepsilon} = \sum_{j=1}^{k} \mu_j(A_j)\delta_{x_j} \tag{2.43}$$

Then, for each $f \in \mathcal{G}$,

$$\begin{aligned}&\Big|\int_X f d(\mu - \mu_{\mathcal{G},\varepsilon})\Big| \leqslant \sum_{j=1}^{k}\Big|\int_{A_i} f d(\mu - \mu_{\mathcal{G},\varepsilon})\Big| = \sum_{j=1}^{k}\Big|\int_{A_i} f d\mu - \mu_j(A_j) f(x_i)\Big|\\ =&\sum_{j=1}^{k}\Big|\int_{A_i} (f - f(x_j)) d\mu\Big| \leqslant \sum_{j=1}^{k}\int_{A_j} |f - f(x_j)| d\widehat{\mu} \leqslant \frac{\varepsilon}{\widehat{\mu}(K)}\sum_{j=1}^{k}\widehat{\mu}(A_j) = \varepsilon\end{aligned}$$

This proves the desired inequality. Moreover, by Exp. 2.69 and Lem. 2.70,

$$\|\mu_{\mathcal{G},\varepsilon}\| = \sum_{j=1}^{k} |\mu_j(A_j)| \leqslant \|\mu\| \leqslant 1$$

This proves that $\mu_{\mathcal{G},\varepsilon} \in \overline{B}_{\mathcal{D}(X,\mathbb{K})}(0, 1)$. □

[12] For example, take $A_1 = K \cap U_1$ and $A_j = K \cap U_j \backslash (U_1 \cup \cdots \cup U_{j-1})$ if $j > 1$.

The proof of Thm. 2.71 immediately implies:

**Theorem 2.72.** *For each $\mu \in \mathcal{RM}(X,\mathbb{C})$, we have*

$$\|\mu\| = \sup\Big\{\sum_{j=1}^k |\mu(A_j)| : k \in \mathbb{Z}_+, \text{ and } A_1,\dots,A_k \in \mathfrak{B}_X \text{ are mutually disjoint}\Big\} \tag{2.44}$$

*Proof.* Lem. 2.70 implies "$\geqslant$". Let us prove "$\leqslant$". Let $(\mu_{\mathcal{G},\varepsilon})_{(\mathcal{G},\varepsilon)\in\mathfrak{I}}$ be the net in $\mathcal{D}(X,\mathbb{C})$ converging weak-* to $\mu$ and satisfying $\|\mu_{\mathcal{G},\varepsilon}\| \leqslant \|\mu\|$. Each $\mu_{\mathcal{G},\varepsilon}$ is of the form (2.43), by Lem. 2.69, the RHS of (2.44) is $\geqslant \|\mu_{\mathcal{G},\varepsilon}\|$. By Fatou's lemma for weak-* convergence (Prop. 2.44), the RHS of (2.44) is $\geqslant \|\mu\|$. $\square$

# 3 Basics of inner product spaces

## 3.1 Sesquilinear forms

Let $V$ be $\mathbb{C}$-vector spaces.

### 3.1.1 Sesquilinear forms

**Definition 3.1.** A map of $\mathbb{C}$-vector spaces $T : V \to W$ is called **antilinear** or **conjugate linear** if for every $a, b \in \mathbb{F}$ and $u, v \in V$ we have

$$T(au + bv) = \overline{a}u + \overline{b}v$$

where $\overline{a}, \overline{b}$ are the complex conjugates of $a, b$.

**Definition 3.2.** A function $\langle \cdot | \cdot \rangle : V \times V \to \mathbb{C}$ (sending $u \times v \in V^2$ to $\langle u|v \rangle$) is called a **sesquilinear form** if it is antilinear on the first variable, and linear on the second one.[1] Namely, for each $a, b \in \mathbb{C}$ and $u, v, w \in V$ we have

$$\langle au + bv|w \rangle = \overline{a}\langle u|w \rangle + \overline{b}\langle v|w \rangle \qquad \langle w|au + bv \rangle = a\langle w|u \rangle + b\langle w|v \rangle$$

More generally, if $V, W$ are complex vector spaces, a map $V \times W \to \mathbb{C}$ is also called **sesquilinear** if it is antilinear on the $V$-component and linear on the $W$-component. The function

$$V \to \mathbb{C} \qquad v \mapsto \langle v|v \rangle$$

is called the **quadratic form** associated to the sesquilinear form $\langle \cdot | \cdot \rangle$.

Notice the difference between the notations $\langle u|v \rangle$ and $\langle u, v \rangle$: the latter always means a bilinear form, i.e., a function which is linear on both variables.

**Remark 3.3.** For each sesquilinear form $\langle \cdot | \cdot \rangle$ on $V$, we have the **polarization identity**

$$\begin{aligned} \langle u|v \rangle = &\frac{1}{4} \sum_{t=0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}} \langle u + e^{\mathbf{i}t}v|u + e^{\mathbf{i}t}v \rangle e^{\mathbf{i}t} \\ = &\frac{1}{4}\Big(\langle u + v|u + v \rangle - \langle u - v|u - v \rangle + \mathbf{i}\langle u + \mathbf{i}v|u + \mathbf{i}v \rangle - \mathbf{i}\langle u - \mathbf{i}v|u - \mathbf{i}v \rangle\Big) \end{aligned} \tag{3.1}$$

Therefore, sesquilinear forms are determined by their associated quadratic forms.

**Definition 3.4.** Let $\omega(\cdot|\cdot) : V \times W \to \mathbb{C}$ be a sesquilinear form. The **adjoint sesequilinear form $\omega^*$** is defined to be

$$\omega^* : W \times V \to \mathbb{C} \qquad \omega^*(w|v) = \overline{\omega(v|w)}$$

[1] This is different from [Gui-A], where the second variable is assumed to be antilinear

**Definition 3.5.** A sesquilinear form $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{C}$ is called a **Hermitian form** if is equal to it adjoint, namely,

$$\langle v|u\rangle = \overline{\langle u|v\rangle} \qquad \text{for each } u, v \in V$$

**Proposition 3.6.** *Let $\langle\cdot|\cdot\rangle$ be a sesquilinear form on $V$. The following are equivalent:*

*(1) $\langle\cdot|\cdot\rangle$ is a Hermitian form.*

*(2) The quadratic form associated to $\langle\cdot|\cdot\rangle$ is real-valued, that is, for each $v \in V$ we have $\langle v|v\rangle \in \mathbb{R}$.*

*Proof.* Let $\omega = \langle\cdot|\cdot\rangle$. By the polarization identity, we have $\omega^* = \omega$ iff $\omega^*(v|v) = \omega(v|v)$ (i.e. $\overline{\omega(v|v)} = \omega(v|v)$) for each $v \in V$. □

### 3.1.2 Positive sesquilinear forms

**Definition 3.7.** A sesquilinear form $\langle\cdot|\cdot\rangle$ on $V$ is called **positive semi-definite** (or simply **positive**) and written as $\langle\cdot|\cdot\rangle \geqslant 0$, if $\langle v|v\rangle \geqslant 0$ for all $v \in V$. If a positive sesquilinear form $\langle\cdot|\cdot\rangle$ on $V$ is fixed, we define

$$\|v\| = \sqrt{\langle v|v\rangle} \qquad \text{for all } v \in V \tag{3.2}$$

Then it is clear that $\|\lambda v\| = |\lambda| \cdot \|v\|$ for each $v \in V$ and $\lambda \in \mathbb{C}$. A vector $v \in V$ satisfying $\|v\| = 1$ is called a **unit vector**.

By Prop. 3.6, a positive sesquilinear form is Hermitian. More generally, we have the following definition:

**Definition 3.8.** Let $\omega_1, \omega_2$ be Hermitian forms on $V$. We write

$$\boldsymbol{\omega_1 \leqslant \omega_2}$$

(equivalently, $\omega_2 \geqslant \omega_1$) if the (real-valued) quadratic forms associated to $\omega_1$ and $\omega_2$ satisfy the corresponding inequality, that is,

$$\omega_1(\xi|\xi) \leqslant \omega_2(\xi|\xi) \qquad \text{for each } \xi \in V$$

Thus, "$\leqslant$" defines a partial order on the set of sesquilinear forms on $V$. Moreover, the meaning of $0 \leqslant \omega$ agrees that in Def. 3.7

**Theorem 3.9 (Cauchy-Schwarz inequality).** *Let $\langle\cdot|\cdot\rangle$ be a positive sesquilinear form on $V$. Then for each $u, v \in V$ we have*

$$|\langle u|v\rangle| \leqslant \|u\| \cdot \|v\|$$

*Proof.* By linear algebra, if $f : \mathbb{R}^2 \to \mathbb{R}$ is a quadratic form

$$f(x,y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2bxy + cy^2$$

where $a, b, c \in \mathbb{R}$, then $f \geqslant 0$ iff $a \geqslant 0, b \geqslant 0$ and

$$ac - b^2 \equiv \det \begin{pmatrix} a & b \\ b & c \end{pmatrix} \geqslant 0$$

In fact, we only need the fact that if $f \geqslant 0$ then $ac - b^2 \geqslant 0$. To see this, note that if $f$ is not always $0$, then one of $a, c$ must be nonzero; otherwise, $f(x,y) = 2bxy$ cannot be always $\geqslant 0$. Thus, assume WLOG that $a \neq 0$. Then $f(x,1) = ax^2 + 2bx + c = a(x + b/a)^2 + c - b^2/a$, which implies $a > 0$ and $c - b^2/a \geqslant 0$, and hence $ac - b^2 \geqslant 0$.

Now, we let $f : \mathbb{R}^2 \to \mathbb{R}_{\geqslant 0}$ be the quadratic form defined by pulling back the form $\xi \in V \mapsto \langle \xi | \xi \rangle$ via the map $(x,y) \in \mathbb{R}^2 \mapsto xu + yv \in V$, that is,

$$f(x,y) = \langle xu + yv | xu + yv \rangle = \|u\|^2 \cdot x^2 + 2\mathrm{Re}\langle u|v \rangle \cdot xy + \|v\|^2 \cdot y^2$$

Then, the above paragraph shows that $\|u\|^2 \cdot \|v\|^2 - (\mathrm{Re}\langle u|v\rangle)^2 \geqslant 0$, equivalently,

$$|\mathrm{Re}\langle u|v\rangle| \leqslant \|u\| \cdot \|v\|$$

Choose $\lambda \in \mathbb{S}^1$ such that $\lambda \langle u|v\rangle \in \mathbb{R}$. Since the above inequality holds when $v$ is replaced by $\lambda v$, we get

$$|\langle u|v\rangle| = |\mathrm{Re}\langle u|\lambda v\rangle| \leqslant \|u\| \cdot \|\lambda v\| = \|u\| \cdot \|v\|$$

$\square$

**Corollary 3.10.** *Let $\langle \cdot | \cdot \rangle$ be a positive sesquilinear form on $V$. Then we have*

$$\{v \in V : \|v\| = 0\} = \{v \in V : \langle v|\xi\rangle = 0 \text{ for all } \xi \in V\}$$

*where the RHS is clearly a linear subspace of $V$. We call this space the **null space** of $\langle \cdot | \cdot \rangle$.*

*Proof.* Let $v \in V$. If $\langle v|V\rangle = 0$, then $\|v\|^2 = \langle v|v\rangle = 0$. Conversely, if $\|v\| = 0$, then by the Cauchy-Schwarz inequality, for each $\xi \in V$ we have $|\langle u|\xi\rangle| \leqslant \|u\| \cdot \|\xi\| = 0$. $\square$

**Corollary 3.11.** *Let $\langle \cdot | \cdot \rangle$ be a positive sesquilinear form on $V$. Then $v \in V \mapsto \|v\| \in \mathbb{R}_{\geqslant 0}$ is a seminorm on $V$.*

*Proof.* It remains to check the subadditivity: for each $u, v \in V$, the Cauchy-Schwarz inequality imlies

$$\begin{aligned} \|u+v\|^2 &= \langle u+v|u+v\rangle = \|u\|^2 + 2\mathrm{Re}\langle u|v\rangle + \|u\|^2 \\ \leqslant &\|u\|^2 + 2\|u\| \cdot \|v\| + \|v\|^2 = (\|u\| + \|v\|)^2 \end{aligned}$$

$\square$

## 3.2 Inner product spaces and bounded sesquilinear forms

### 3.2.1 Inner product spaces

**Definition 3.12.** Let $\langle\cdot|\cdot\rangle$ be a positive sesquilinear form on a $\mathbb{C}$-vector space $V$. We call $\langle\cdot|\cdot\rangle$ an **inner product** if it is **non-degenerate**, i.e., the null space is $0$. We call the pair $(V,\langle\cdot|\cdot\rangle)$ (or simply call $V$) an **inner product space** or a **pre-Hilbert space** .

**Exercise 3.13.** Let $\langle\cdot|\cdot\rangle$ be a positive sesquilinear form on $V$ with null space $\mathscr{N}$. Prove that there is a (necessarily unique) inner product $\langle\cdot|\cdot\rangle_{V/\mathscr{N}}$ on the quotient space $V/\mathscr{N}$ such that for any $u,v\in V$, the cosets $u+\mathscr{N}$ and $v+\mathscr{N}$ satisfy

$$\langle u+\mathscr{N}|v+\mathscr{N}\rangle_{V/\mathscr{N}}=\langle u|v\rangle$$

**Example 3.14.** Let $X$ be a set. Then $l^2(X)=l^2(X,\mathbb{C})$ is an inner product space, where

$$\langle f|g\rangle=\sum_{x\in X}\overline{f(x)}g(x)\qquad\text{for any } f,g\in l^2(X)$$

**Example 3.15.** Let $(X,\mu)$ be a measure space. Then $L^2(X,\mu)$ is an inner product space, where

$$\langle f|g\rangle=\int_X\overline{f}gd\mu\qquad\text{for any } f,g\in L^2(X,\mu)$$

**Remark 3.16.** By Rem. 3.11, an inner product space $V$ is equipped with the norm defined by $\|v\|=\sqrt{\langle v|v\rangle}$. In particular, $V$ is a metric space with metric $d(u,v)=\|u-v\|$. The topology on $V$ induced by this metric is called the **norm topology** of $V$.

**Remark 3.17.** Let $V,W$ be inner product spaces. If $T:V\to V$ is a linear map, then $T$ is an isometry of metric spaces iff $T$ is an isometry of normed vector spaces, i.e.,

$$\langle Tv|Tv\rangle=\langle v|v\rangle\qquad\text{for all } v\in V$$

By the polarization identity, this is equivalent to

$$\langle Tu|Tv\rangle=\langle u|v\rangle\qquad\text{for all } u,v\in V$$

A surjective linear isometry $T:V\to W$ is called a **unitary map**. If $T:V\to W$ is unitary, we say that $V,W$ are **isomorphic inner product spaces** (or that $V,W$ are **unitarily equivalent**).

Similarly, if $T:V\to V$ is antilinear map between inner product spaces, then $T$ is an isometry of metric spaces iff

$$\langle Tv|Tv\rangle=\langle v|v\rangle\qquad\text{for all } v\in V$$

By the polarization identity, this is equivalent to

$$\langle Tu|Tv\rangle = \langle v|u\rangle \qquad \text{for all } u, v \in V$$

A surjective antilinear isometry $T : V \to W$ is called an **antiunitary map**. If $T : V \to W$ is antiunitary, we say that $V$ and $W$ are **antiunitarily equivalent**. □

### 3.2.2 Bounded sesquilinear forms

Let $V, W$ be inner product spaces.

**Definition 3.18.** The **(complex) conjugate** of $V$ is the inner product space $V^{\complement}$ defined as follows. The elements of $V^{\complement}$ correspond bijectively to those of $V$ by the map

$$\complement : V \to V^{\complement} \qquad v \mapsto v^{\complement} \equiv \overline{v}$$

where $v^{\complement} \equiv \overline{v}$ is an abstract element, called the **conjugate** of $v$. Moreover, the structure of an inner product space on $V^{\complement}$ is defined in such a way that $\complement$ is antiunitary. In other words, for each $u, v \in V$ and $a, b \in \mathbb{C}$, we have

$$\overline{a} \cdot \overline{u} + \overline{b} \cdot \overline{v} := \overline{au + bv}$$
$$\langle \overline{u}|\overline{v}\rangle_{V^{\complement}} := \overline{\langle u|v\rangle_V} = \langle v|u\rangle_V$$

The conjugate of $V^{\complement}$ is defined to be $V$, that is,

$$(V^{\complement})^{\complement} = V$$

Moreover, the conjugate map $\complement : V^{\complement} \to V$ is defined by

$$\complement : V^{\complement} \to V \qquad \overline{v} \mapsto v$$

Thus $\overline{\overline{v}} = v$ for each $v \in V$. □

**Remark 3.19.** An antilinear map $T : V \to W$ is equivalent to the linear map

$$V \to W^{\complement} \qquad v \mapsto \overline{Tv} \tag{3.3a}$$

and is also equivalent to the linear map

$$V^{\complement} \to W \qquad \overline{v} \mapsto Tv \tag{3.3b}$$

It is clear that $T$ is an antilinear isometry (resp. antiunitary) iff (3.3a) is a linear isometry (resp. unitary) iff (3.3b) is a linear isometry (resp. unitary).

**Remark 3.20.** A sesquilinear form $\omega : V \times W \to \mathbb{C}$ is equivalent to a bilinear form

$$\widetilde{\omega} : V^{\complement} \times W \to \mathbb{C} \qquad (\overline{v}, w) \mapsto \langle v|w\rangle$$

Unless otherwise stated, we always view $\omega$ and $\widetilde{\omega}$ as the same.

**Definition 3.21.** Let $\omega : V \times W \to \mathbb{C}$ be a sesquilinear form. The **norm** $\|\omega\|$ is defined to be the norm of the associated bilinear form $V^{\complement} \times W \to \mathbb{C}$. Therefore,

$$\|\omega\| = \sup_{v \in \overline{B}_V(0,1), w \in \overline{B}_W(0,1)} |\omega(u|v)|$$

Recalling the notation (2.19), we let

$$\mathfrak{Ses}(V|W) := \mathfrak{L}(V^{\complement} \times W, \mathbb{C})$$

which is the space of bounded sesquilinear forms $V \times W \to \mathbb{C}$. We write

$$\mathfrak{Ses}(V) := \mathfrak{Ses}(V|V)$$

The elements of $\mathfrak{Ses}(V|W)$ (resp. $\mathfrak{Ses}(V)$) are called **bounded sesquilinear forms** on $V \times W$ (resp. on $V$).

**Example 3.22.** The inner product

$$\langle \cdot|\cdot \rangle : V \times V \to \mathbb{C} \qquad (u, v) \mapsto \langle u|v\rangle$$

has norm $1$, and hence belongs to $\mathfrak{Ses}(V)$. Therefore, by Prop. 2.22, this map is continuous.

## 3.3 Orthogonality

Let $V$ be an inner product spaces.

### 3.3.1 Orthogonal and orthonormal vectors

**Definition 3.23.** A set $\mathfrak{S}$ of vectors of $V$ are called **orthogonal** if $\langle u|v\rangle = 0$ for any distinct $u, v \in V$. An orthogonal set $\mathfrak{S}$ is called **orthonormal** if $\|v\| = 1$ for all $v \in V$.

**Remark 3.24.** We will also talk about an **orthogonal** resp. **orthonormal family of vectors** $(e_i)_{i\in I}$. This means that $\langle e_i|e_j\rangle = 0$ for any distinct $i, j \in I$ (resp. $\langle e_i|e_j\rangle = \delta_{i,j}$ for any $i, j \in I$).

In particular, two vectors $u, v \in V$ are called orthogonal and written as

$$u \perp v$$

when $\langle u|v\rangle = 0$. A fundamental fact about orthogonal vectors is

**Proposition 3.25 (Pythagorean identity).** *Suppose that $u, v \in V$ are orthogonal. Then*

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 \tag{3.4}$$

*In particular,*

$$\|v\| \leqslant \|u+v\| \tag{3.5}$$

*Proof.* $\|u+v\|^2 = \langle u+v|u+v\rangle = \langle u|u\rangle + \langle v|v\rangle + 2\mathrm{Re}\langle u|v\rangle = \langle u|u\rangle + \langle v|v\rangle$. □

Note that by applying (3.4) repeatedly, we see that if $v_1, \dots, v_n \in V$ are orthogonal, then

$$\|v_1 + \cdots + v_n\|^2 = \|v_1\|^2 + \cdots + \|v_n\|^2 \tag{3.6}$$

**Remark 3.26.** Suppose that $\mathfrak{S}$ is an orthonormal set of vectors of $V$. Then $\mathfrak{S}$ is clearly linearly independent. (If $e_1, \dots, e_n \in \mathfrak{S}$ and $\sum_i a_i e_i = 0$, then $a_j = \sum_i \langle e_j | a_i e_i\rangle = \langle e_j|0\rangle = 0$.) Thus, by linear algebra, if $\mathfrak{S} = \{e_1, \dots, e_n\}$ is finite, then one can find uniquely $a_1, \dots, a_n \in \mathbb{C}$ and $u \in V$ such that $v = a_1 e_1 + \cdots + a_n e_n + u$ and that $u$ is orthogonal to $e_1, \dots, e_n$. The expressions of $a_1, \dots, a_n, u$ can be expressed explicitly:

**Proposition 3.27 (Gram-Schmidt).** *Let $e_1, \dots, e_n$ be orthonormal vectors in $V$. Let $v \in V$. Then*

$$v - \sum_{i=1}^n e_i \cdot \langle e_i|v\rangle \tag{3.7}$$

*is orthogonal to $e_1, \dots, e_n$.*

*Proof.* This is a direct calculation and is left to the readers. □

**Remark 3.28.** "Gram-Schmidt" usually refers to the following process. Let $v_1, \dots, v_n$ be a set of linearly independent vectors of $V$. Then there is an algorithm of finding an orthonormal basis of $U = \mathrm{Span}\{v_1, \dots, v_n\}$: Let $e_1 = v_1/\|v_1\|$. Suppose that a set of orthonormal vectors $e_1, \dots, e_k$ in $U$ have been found. Then $e_{k+1}$ is defined by $\widetilde{v}_{k+1}/\|\widetilde{v}_{k+1}\|$ where $\widetilde{v}_{k+1} = v_{k+1} - \sum_{i=1}^k e_i \cdot \langle e_i|v_{k+1}\rangle$.

Combining Pythagorean with Gram-Schmidt, we have:

**Corollary 3.29 (Bessel's inequality).** *Let $(e_i)_{i\in I}$ be a family of orthonormal vectors of $V$. Then for each $v \in V$ we have*

$$\sum_{i\in I} |\langle e_i|v\rangle|^2 \leqslant \|v\|^2 \tag{3.8}$$

*In particular, the set $\{i \in I : \langle e_i|v\rangle \neq 0\}$ is countable.*

*Proof.* The LHS of (3.8) is $\lim_{J\in\mathrm{fin}(2^I)}\sum_{j\in J}|\langle e_j|v\rangle|^2$. Thus, it suffices to show that for each $J\in\mathrm{fin}(2^I)$ we have $\sum_{j\in J}|\langle e_j|v\rangle|^2\leqslant\|v\|^2$. Let

$$u_1=\sum_{j\in J}e_j\cdot\langle e_j|v\rangle\qquad u_2=v-u_1$$

(Namely, $v=u_1+u_2$ is the orthogonal decomposition of $v$ with respect to $\mathrm{Span}\{e_j:j\in J\}$.) By Gram-Schmidt, we have $\langle u_1|u_2\rangle=0$. By Pythagorean, we have $\|u_1\|^2\leqslant\|v\|^2$. But Pythagorean (3.6) also implies

$$\|u_1\|^2=\sum_{j\in J}|\langle e_j|v\rangle|^2$$

The last statement about countability follows from . $\square$

### 3.3.2 Orthogonal decomposition

**Definition 3.30.** Let $U$ be a linear subspace of $V$. Let $v\in V$. An **orthogonal decomposition** of $v$ with respect to $U$ is an expression of the form

$$v=u+w\qquad\text{where } u\in U \text{ and } w\perp U$$

Orthogonal decompositions of $v$ are unique if exist. We call $u$ the **orthogonal projection** of $v$ onto $U$.

*Proof of uniqueness.* Suppose that $v=u'+w'$ is another orthogonal decomposition. Then $u-u'$ equals $w'-w$. Let $\xi=u-u'$. Then $\xi\in U$ and $\xi\perp U$. So $\langle\xi|\xi\rangle=0$, and hence $\xi=0$. So $u=u'$ and $w=w'$. $\square$

**Definition 3.31.** Let $U$ be a linear subspace of $V$. We say that $V$ **has a projection onto** $U$ if every vector has an orthogonal decomposition with respect to $U$. In that case, we define the map

$$P:V\to V$$

determined by the fact that each $v\in V$ has orthogonal decomposition $v=Pv+(v-Pv)$ where $Pv\in U$ and $v-Pv\perp U$. Clearly $P$ is linear. By the Pythagorean identity, we have $\|Pv\|\leqslant\|v\|$, and hence

$$\|P\|\leqslant 1$$

Thus $P\in\mathfrak{L}(V)$. We say that $P$ is the **projection (operator) associated to** $U$.

**Example 3.32.** Let $e_1,\dots,e_n$ be orthonormal vectors of $V$. Let $U=\mathrm{Span}\{e_1,\dots,e_n\}$. Choose any $v\in V$. Then by Gram-Schmidt,

$$v=u+w\qquad\text{where } u=\sum_{i=1}^n e_i\cdot\langle e_i|v\rangle \text{ and } w=v-u\tag{3.9}$$

is the orthogonal decomposition of $v$ with respect to $U$. Therefore, the projection operator associated to $U$ is

$$V \to V \qquad v \mapsto \sum_{i=1}^{n} e_i \cdot \langle e_i | v \rangle$$

**Proposition 3.33.** *Let $U$ be a linear subspace of $V$. Suppose that $v \in V$ has orthogonal decomposition $v = u + w$ with respect to $U$. Then*

$$\|v - u\| = \inf_{\xi \in U} \|v - \xi\| \tag{3.10}$$

*Proof.* Clearly "$\geqslant$" holds. Choose any $\xi \in U$. Then $v - \xi = v - u + u - \xi = w + (u - \xi)$. Since $u - \xi \in U$, we have $w \perp u - \xi$. Thus, by Pythagorean, we have $\|w\| \leqslant \|v - \xi\|$. □

### 3.3.3 Direct sums and orthogonal decomposition

Next, we give a more explicit description of orthogonal decomposition in terms of direct sum.

**Definition 3.34.** Let $V_1, \dots, V_n$ be inner product spaces. Their **direct sum** $\boldsymbol{V_1 \oplus \cdots \oplus V_n}$ is an inner product space defined as follows. As a set, $V_1 \oplus \cdots \oplus V_n$ equals $V_1 \times \cdots \times V_n$. So it consists of elements of the form $(v_1, \dots, v_n)$ where $v_i \in V_i$. We write $(v_1, \dots, v_n)$ as $v_1 \oplus \cdots \oplus v_n$. The linear structure is defined by

$$\begin{aligned}(v_1 \oplus \cdots \oplus v_n) + (v_1' \oplus \cdots \oplus v_n') &= (v_1 + v_1') \oplus \cdots \oplus (v_n + v_n') \\ a(v_1 \oplus \cdots \oplus v_n) &= av_1 \oplus \cdots \oplus av_n\end{aligned}$$

where $v_i, v_i' \in V_i$ and $a \in \mathbb{C}$. The inner product is defined by

$$\langle v_1 \oplus \cdots \oplus v_n | v_1' \oplus \cdots \oplus v_n' \rangle = \langle v_1 | v_1' \rangle + \cdots + \langle v_n | v_n' \rangle$$

We view $V_i$ as an inner product subspace of $V_1 \oplus \cdots \oplus V_n$ by identifying $v_i \in V_i$ with $0 \oplus \cdots \oplus v_i \oplus \cdots \oplus 0 \in V_1 \oplus \cdots \oplus V_n$. Then, it is clear that $V_i \perp V_j$ if $i \neq j$.

**Remark 3.35.** Suppose that $U_1, \dots, U_n$ are mutually orthogonal linear subspaces of $V$. Then we clearly have a linear isometry

$$U_1 \oplus \cdots \oplus U_n \longrightarrow V \qquad u_1 \oplus \cdots \oplus u_n \mapsto u_1 + \cdots + u_n \tag{3.11}$$

Therefore, if $V$ is spanned by $U_1, \dots, U_n$, then (3.11) is surjective, and hence is an isomorphism of normed vector spaces. In that case, we say that (3.11) is the **canonical isomorphism** from $U_1 \oplus \cdots \oplus U_n$ to $V$. With abuse of notation, we also say that $V$ "is" the direct sum $U_1 \oplus \cdots \oplus U_n$.

**Example 3.36.** Let $U_1, U_2$ be inner product spaces and $V = U_1 \oplus U_2$. Then $V$ has a projection onto $U_1$. The projection operator associated to $U_1$ is defined by sending each $u_1 \oplus u_2$ to $u_1$.

We now show that any projection is unitarily equivalent to the one given in Exp. 3.36.

**Definition 3.37.** If $U$ is a linear subspace of $V$, we define the **orthogonal complement** of $U$ (in $V$) to be

$$U^\perp = \{\xi \in V : \langle \xi | u \rangle = 0 \text{ for all } u \in U\}$$

**Remark 3.38.** Let $U$ be a linear subspace of $V$. Then $U^\perp$ is closed in $V$, since it is the kernel of the bounded linear map $\xi \in V \mapsto \langle \xi | u \rangle$. Moreover, by the continuity of $\langle \cdot | \cdot \rangle : V \times V \to \mathbb{C}$, a vector of $V$ is orthogonal to $U$ iff it is orthogonal to $\overline{U} = \mathrm{Cl}_V(U)$, that is,

$$U^\perp = \overline{U}^\perp$$

**Example 3.39.** If $U_1, U_2$ are inner product spaces, then $U_1$ and $U_2$ are the orthogonal complements of each other in $U_1 \oplus U_2$.

**Proposition 3.40.** *Let $U$ be a linear subspace of $V$. Suppose that $V$ has a projection onto $U$, and let $P$ the projection operator onto $U$. Then $V$ is canonically isomorphic to $U \oplus U^\perp$. Moreover, identifying $U \oplus U^\perp$ with $V$ (by identifying $u \oplus v$ with $u + v$ if $u \in U, v \in U^\perp$), then*

$$P : U \oplus U^\perp \to U \oplus U^\perp \qquad u \oplus v \mapsto u = u \oplus 0$$

*Consequently, $1 - P$ is the projection of $V$ onto $U^\perp$, and we have*

$$\mathrm{Rng}(U) = \mathrm{Ker}(1 - P) = U \qquad \mathrm{Ker}(P) = \mathrm{Rng}(1 - P) = U^\perp$$

It follows from $V = U \oplus U^\perp$ that $U$ is the orthogonal complement of $U^\perp$, i.e., $U = U^{\perp\perp}$.

*Proof.* The surjectivity of the linear isometry

$$U \oplus U^\perp \to V \qquad u \oplus v \mapsto u + v$$

follows from the fact that $V$ has a projection onto $U$. Clearly $P$ sends $u + v$ to $u$. The rest of this proposition is obvious. $\square$

**Corollary 3.41.** *Suppose that $U$ is a finite-dimensional linear subspace of $V$. Then $U$ is closed in $V$.*

*Proof.* By Exp. 3.32, there is a projection operator $P$ of $V$ onto $U$. By Prop. 3.40, $U$ is the orthogonal complement of $U^\perp$, and hence is closed. $\square$

### 3.3.4 Orthonormal basis

**Definition 3.42.** A set $\mathfrak{S}$ (or a family $(e_i)_{i\in I}$) of orthonormal vectors of $V$ is called an **orthonormal basis** of $V$ if it spans a dense subspace of $V$.

**Example 3.43.** If $X$ is a set, by Prop. 2.54, $l^2(X)$ has an orthonormal basis $(\chi_{\{x\}})_{x\in X}$.

**Example 3.44.** If $V$ is separable, then $V$ has a countable orthonormal basis.

*Proof.* Let $\{v_1, v_2, \dots\}$ be a dense subset of $V$ where $v_1 \neq 0$. Then by Gram-Schmidt (Rem. 3.28), we can find $e_1, e_2, \dots \in V$ such that the set $\{e_1, e_2, \dots\}$ is orthnormal (after removing the duplicated terms), and that $\mathrm{Span}\{v_1, \dots, v_n\} = \mathrm{Span}\{e_1, \dots, e_n\}$ for each $n$. Then $\{e_1, e_2, \dots\}$ clearly spans a dense subspace of $V$. $\square$

We remark that there are non-separable and non-complete inner product spaces that do not have orthonormal bases. See [Gud74].

**Theorem 3.45.** *Suppose that $(e_i)_{i\in I}$ is an orthonormal basis of $V$. Then for each $v \in V$, the RHS of the following converges (under the norm of $V$) to the LHS:*

$$v = \sum_{i\in I} e_i \cdot \langle e_i|v\rangle \tag{3.12}$$

*Proof.* Note that for $J \in \mathrm{fin}(2^I)$, the expression

$$\Big\| v - \sum_{j\in J} e_j \cdot \langle e_j|v\rangle \Big\|^2 = \|v\|^2 - \sum_{j\in J} |\langle e_j|v\rangle|^2$$

decreases when $J$ increases. Thus, it suffices to prove that the $\inf_{J\in\mathrm{fin}(2^I)}$ of this expression is 0.

By assumption, we can find $J \in \mathrm{fin}(2^I)$ and $(\lambda_j)_{j\in J}$ in $\mathbb{C}$ such that $\|v - \sum_{j\in J} \lambda_j e_j\|$ is small enough. On the other hand, applying Prop. 3.33 to the orthogonal projection $v = u + w$ where $w = \sum_{j\in J} e_j \cdot \langle e_j|v\rangle$ (cf. Exp. 3.32), we have

$$\Big\| v - \sum_{j\in J} e_j \cdot \langle e_j|v\rangle \Big\| \leqslant \Big\| v - \sum_{j\in J} \lambda_j e_j \Big\| \tag{3.13}$$

Thus, the infimum of the LHS over $J \in \mathrm{fin}(2^I)$ is zero. $\square$

**Corollary 3.46 (Parseval's identity).** *Suppose that $(e_i)_{i\in I}$ is an orthonormal basis of $V$. Then for each $u, v \in V$ we have*

$$\langle u|v\rangle = \sum_{i\in I} \langle u|e_i\rangle \cdot \langle e_i|v\rangle \tag{3.14}$$

*In particular,*

$$\|v\|^2 = \sum_{i\in I} |\langle e_i|v\rangle|^2 \tag{3.15}$$

*Proof.* By Thm. 3.45, $v = \lim_{J\in\mathrm{fin}(2^I)} v_J$ where $v_J = \sum_{j\in J} e_j \cdot \langle v|e_j\rangle$. By the continuity of $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{C}$ (Exp. 3.22), we have

$$\langle u|v\rangle = \lim_{J\in\mathrm{fin}(2^I)} \langle u|v_J\rangle = \lim_{J\in\mathrm{fin}(2^I)} \sum_{j\in J}\langle u|e_j\rangle \cdot \langle e_j|v\rangle = \sum_{i\in I}\langle u|e_i\rangle \cdot \langle e_i|v\rangle$$

□

**Corollary 3.47.** *Suppose that $(e_x)_{x\in X}$ is an orthonormal basis of $V$. Then there is a linear isometry*

$$\Phi : V \to l^2(X) \qquad v \mapsto (\langle e_x|v\rangle)_{x\in X} \tag{3.16}$$

*whose range is dense in $l^2(X)$.*

*Proof.* Parseval's identity shows that $(\langle e_x|v\rangle)_{x\in X}$ has finite $l^2$-norm $\|v\|$. So the map $\Phi$ defined by (3.16) is clearly a linear isometry. The density of the range of $\Phi$ follows from the fact that $l^2(X)$ contains all $\chi_{\{x\}} = \Phi(e_x)$, and that $\mathrm{Span}\{\chi_{\{x\}} : x \in X\}$ is dense in $l^2(X)$ (cf. Prop. 2.54). □

## 3.4 Hilbert spaces

**Theorem 3.48.** *Let $\mathcal{H}$ be an inner product space. Then the following three conditions are equivalent:*

*(a) $\mathcal{H}$ is (Cauchy) complete.*

*(b) For each orthonormal family $(e_i)_{i\in I}$ in $\mathcal{H}$, and for each family $(a_i)_{i\in I}$ in $\mathbb{C}$ satisfying $\sum_{i\in I} |a_i|^2 < +\infty$, the unordered sum $\sum_{i\in I} a_i e_i$ converges (under the norm of $\mathcal{H}$).*

*(c) $\mathcal{H}$ is unitarily equivalent to $l^2(X)$ for some set $X$.*

*If $\mathcal{H}$ satisfies any of these conditions, we say that $\mathcal{H}$ is a **Hilbert space**.*

*Proof.* (c)⇒(a): By Thm. 2.60, $l^2(X)$ is the dual space of $l^2(X)$. Since any dual space is complete (Cor. 2.37), $l^2(X)$ is complete.

(a)⇒(b): Since $\sum_i |a_i|^2 < +\infty$, for each $\varepsilon > 0$ there exists $J \in \mathrm{fin}(2^I)$ such that for all finite $K \subset I\backslash J$ we have $\sum_{k\in K} |a_k|^2 < \varepsilon$, and hence, by the Pythagorean identity,

$$\Big\|\sum_{k\in K} a_k e_k\Big\|^2 = \sum_{k\in K} |a_k e_k|^2 < \varepsilon$$

Thus $(\sum_{j\in J} a_j e_j)_{J\in\mathrm{fin}(2^I)}$ is a Cauchy net. By the completeness of $\mathcal{H}$, we see that $\sum_{i\in I} a_i e_i$ converges.

(b)⇒(c): Assume (b). We first show that $\mathcal{H}$ has an orthonormal basis. By Zorn's lemma, we can find a maximal (with respect to the partial order $\subset$) set of orthonormal vectors, written as a family $(e_i)_{i\in I}$. The maximality implies that every nonzero vector $\xi \in \mathcal{H}$ is not orthogonal to some $e_i$. (Otherwise, $\{e_i : i \in I\}$ can be extended to $\{e_i : i \in I\} \cup \{\xi/\|\xi\|\}$.)

Let us prove that $(e_i)_{i\in I}$ is an orthonormal basis. Suppose not. Then $U = \mathrm{Span}\{e_i : i \in I\}$ is not dense in $\mathcal{H}$. Let $\xi \in \mathcal{H}\backslash\overline{U}$. By Bessel's inequality, we have

$$\sum_{i\in I} |\langle e_i|\xi\rangle|^2 < +\infty$$

Therefore, by (b),

$$\sum_{i\in I} e_i \cdot \langle e_i|\xi\rangle \tag{3.17}$$

converges to some vector $\eta \in \mathcal{H}$. By the continuity of $\langle\cdot|\cdot\rangle$ (Exp. 3.22), we see that $\langle e_i|\eta\rangle = \langle e_i|\xi\rangle$ for all $i$, and hence

$$\langle e_i|\xi - \eta\rangle = 0 \qquad \text{for all } i \in I \tag{3.18}$$

Since $\eta \in \overline{U}$ and $\xi \notin \overline{U}$, we conclude that $\xi - \eta$ is a nonzero vector orthogonal to all $e_i$. This contradicts the maximality of $(e_i)_{i\in I}$.

Now we have an orthonormal basis $(e_i)_{i\in I}$. By Cor. 3.47, we have a linear isometry

$$\Phi : \mathcal{H} \to l^2(I) \qquad \xi \mapsto \big(\langle e_i|\xi\rangle\big)_{i\in I}$$

with dense range. If $(a_i)_{i\in I}$ belongs to $l^2(I)$, by (b), the unordered sum $\sum_{i\in I} a_i e_i$ converges to some $\xi \in \mathcal{H}$. Clearly $\Phi(\xi) = (a_i)_{i\in I}$. This proves that $\Phi$ is surjective, and hence is a unitary map. So $\mathcal{H} \simeq l^2(I)$. □

In the proof of Thm. 3.48, we use Zorn's lemma to show that every Hilbert space $\mathcal{H}$ admits an orthonormal basis. The same argument yields a stronger result:

**Proposition 3.49.** *Let $\mathcal{H}$ be a Hilbert space. Then any orthonormal family of vectors in $\mathcal{H}$ can be extended to an orthonormal basis.*

When $\mathcal{H}$ is separable, this proposition can be proved without invoking Zorn's lemma, by applying mathematical induction together with the Gram-Schmidt process (Rem. 3.28). We leave the details of the proof of Prop. 3.49 to the reader.

**Example 3.50.** By Thm. 3.48, if $X$ is a set, then $l^2(X)$ is a Hilbert space.

**Example 3.51.** Let $(X, \mu)$ be a measure space. By the Riesz-Fischer Thm. 1.48, the inner product space $L^2(X, \mu)$ is a Hilbert space.

**Example 3.52.** If $V$ is a closed linear subspace of $\mathcal{H}$ whose inner product is inherited from that of $\mathcal{H}$, then $V$ is a Hilbert space. This is either due to Thm. 3.48-(b), or due to the fact that a closed subset of a complete metric space is complete.

**Corollary 3.53.** *Every Hilbert space $\mathcal{H}$ has an orthonormal basis. Moreover, $\mathcal{H}$ is separable iff the orthonormal basis can be chosen to be countable.*

*Proof.* That $\mathcal{H}$ has an orthonormal basis follows from the proof of Thm. 3.48 or from the fact that $l^2(X)$ has an orthonormal basis $(\chi_{\{x\}})_{x\in X}$. If $X$ is countable, then $l^2(X)$ has dense subset $\mathrm{Span}_{\mathbb{Q}+\mathbf{i}\mathbb{Q}}\{\chi_{\{x\}} : x \in X\}$ and hence is separable. Conversely, we have proved in Exp. 3.44 that every separable inner product space has a countable orthonormal basis. $\square$

**Theorem 3.54.** *Let $(e_x)_{x\in X}$ be an orthonormal basis of a Hilbert space $\mathcal{H}$. Then we have a unitary map*

$$\mathcal{H} \xrightarrow{\simeq} l^2(X) \qquad \xi \mapsto \big(\langle e_x|\xi\rangle\big)_{x\in X} \tag{3.19}$$

*Proof.* This is clear from the proof of Thm. 3.48. $\square$

**Theorem 3.55.** *Let $V$ be a closed linear subspace of $\mathcal{H}$. Then $\mathcal{H}$ has a projection onto $V$. Consequently, by Prop. 3.40, $V \oplus V^\perp$ is canonically isomorphic to $\mathcal{H}$.*

*Proof.* By Exp. 3.52, $V$ is a Hilbert space, and hence admits an orthonormal basis $(e_i)_{i\in I}$. For each $\xi \in \mathcal{H}$, since Bessel's inequality implies $\sum_i |\langle e_i|\xi\rangle|^2 < \|\xi\|^2 < +\infty$, by Thm. 3.48-(b), the following sum converges:

$$P\xi = \sum_{i\in I} e_i \cdot \langle e_i|\xi\rangle$$

and is clearly in $V$. Similar to the argument around (3.18), $\xi - P\xi$ is orthogonal to every $e_i$. Hence $V_0 := \mathrm{Span}\{e_i : i \in I\}$ is orthogonal to $\xi - P\xi$, i.e., $\xi - P\xi \in V_0^\perp$. Since $V$ is the closure of $V_0$, by Rem. 3.38, we have $\xi - P\xi \in V^\perp$. Therefore, $\xi = P\xi + (\xi - P\xi)$ is the orthogonal decomposition of $\xi$ with respect to $V$. $\square$

**Corollary 3.56.** *Let $V$ be a linear subspace of $\mathcal{H}$. Then $(V^\perp)^\perp = \mathrm{Cl}_{\mathcal{H}}(V)$.*

Note that since $V^\perp$ is closed, Cor. 3.56 implies $V^{\perp\perp\perp} = V^\perp$.

*Proof.* By Rem. 3.38, we have $V^\perp = \overline{V}^\perp$. By Thm. 3.55, $\mathcal{H}$ has a projection onto $\overline{V}$. Therefore, by Prop. 3.40, we have $\mathcal{H} = \overline{V} \oplus \overline{V}^\perp = \overline{V} \oplus V^\perp$. Therefore, $\overline{V}$ is the orthogonal complement of $V^\perp$. $\square$

**Corollary 3.57.** *Let $V$ be a linear subspace of $\mathcal{H}$. Then $V$ is dense in $\mathcal{H}$ iff $V^\perp = \{0\}$.*

*Proof.* If $V$ is dense, then $V^\perp = \overline{V}^\perp = \mathcal{H}^\perp = 0$. Conversely, if $V^\perp = \{0\}$, then $V^{\perp\perp} = 0^\perp = \mathcal{H}$. By Cor. 3.56, we have $\overline{V} = V^{\perp\perp} = \mathcal{H}$. Hence $V$ is dense. $\square$

## 3.5 Bounded linear maps and bounded sesquilinear forms

In this section, we let $U, V, W$ be inner product spaces.

In Subsec. 2.5.2, we discussed the close relationship between bounded linear maps and bounded bilinear forms in the general setting of normed vector spaces. This connection allows us to combine the strengths of both perspectives. One key advantage of the perspective of linear operators is that the space $\mathfrak{L}(V)$ is particularly well-suited for symbolic calculus.

In this section, we explore this relationship in the context of inner product spaces and Hilbert spaces. We will see that the passage from $\mathfrak{L}(V)$ to bounded sesquilinear forms fundamentally relies on the Riesz–Fréchet theorem, a pivotal result that enables this correspondence.

### 3.5.1 The Riesz-Fréchet representation theorem

**Definition 3.58.** If $T \in \mathrm{Lin}(V, W)$, we let $\omega_T$ be the sesquilinear form

$$\omega_T : W \times V \to \mathbb{C} \qquad (w, v) \mapsto \langle w|Tv\rangle$$

**Proposition 3.59.** *For each $T \in \mathrm{Lin}(V, W)$, we have*

$$\|T\| = \|\omega_T\|$$

*Consequently, $T$ is bounded iff $\omega_T$ is so, and the map $T \in \mathrm{Lin}(V, W) \mapsto \omega_T$ is injective.*

*Proof.* For each $v \in V, w \in W$, we have

$$|\omega_T(w|v)| = |\langle w|Tv\rangle| \leqslant \|Tv\| \cdot \|w\| \leqslant \|T\| \cdot \|v\| \cdot \|w\|$$

Applying $\sup$ over all $v, w$ in the closed unit balls, we get $\|\omega_T\| \leqslant \|T\|$. Moreover,

$$\|Tv\|^2 = \omega_T(Tv|v) \leqslant \|\omega_T\| \cdot \|Tv\| \cdot \|v\|$$

and hence $\|Tv\| \leqslant \|\omega_T\| \cdot \|v\|$. Applying $\sup$ over all $v$ in the closed unit ball, we get $\|T\| \leqslant \|\omega_T\|$. $\square$

By Prop. 3.59, the map $T \in \mathrm{Lin}(V, W)$ restricts to a linear isometry of normed vector spaces

$$\mathfrak{L}(V, W) \to \mathfrak{Ses}(W|V) \qquad T \mapsto \omega_T \tag{3.20}$$

On the other hand, Cor. 2.39 implies

$$\mathfrak{Ses}(W|V) = \mathfrak{L}(W^{\complement} \times V, \mathbb{C}) \simeq \mathfrak{L}(V, (W^{\complement})^*)$$

and hence a linear isometry

$$\mathfrak{L}(V, W) \to \mathfrak{L}(V, (W^{\complement})^*) \tag{3.21}$$

**Exercise 3.60.** Show that the map (3.21) sends each $T \in \mathfrak{L}(V, W)$ to $\Phi \circ T$, where $\Phi : W \to (W^{\complement})^*$ is defined below.

**Theorem 3.61 (Riesz-Fréchet representation theorem).** *The following map is a linear isometry:*

$$\Phi : W \to (W^{\complement})^* \qquad \xi \mapsto \langle \overline{\xi} | - \rangle \tag{3.22a}$$

*where $\langle \overline{\xi} | - \rangle$ denotes the bounded linear functional*

$$\langle \overline{\xi} | - \rangle : W^{\complement} \to \mathbb{C} \qquad \overline{w} \mapsto \langle \overline{\xi} | \overline{w} \rangle_{W^{\complement}} = \langle w | \xi \rangle_W \tag{3.22b}$$

*Moreover, $W$ is a Hilbert space iff $\Phi$ is surjective (and hence an isomorphism of normed vector spaces).*

In other words, $\Phi$ is determined by the fact that for each $w, \xi \in W$,

$$\langle \overline{w}, \Phi\xi \rangle = \langle w | \xi \rangle \tag{3.23}$$

*Proof.* First, note that for each $\xi \in W$,

$$\|\xi\| = \sup_{w \in \overline{B}_W(0,1)} |\langle w | \xi \rangle| \tag{3.24}$$

Indeed, the Cauchy-Schwarz inequality implies "$\geqslant$ 0". The equality can be achieved by choosing $w = \xi / \|\xi\|$ if $\xi \neq 0$. Therefore,

$$\|\Phi(\xi)\| = \sup_{\overline{w} \in \overline{B}_{W^{\complement}}(0,1)} |\langle \overline{w}, \Phi(\xi) \rangle| = \sup_{w \in \overline{B}_W(0,1)} |\langle w | \xi \rangle| = \|\xi\|$$

This proves that $\Phi$ is a linear isometry.

If $\Phi$ is surjective, then the normed vector space $W$ is isomorphic to the dual space $(W^{\complement})^*$ where the latter is complete by Cor. 2.37. Therefore, $W$ is a Hilbert space.

Conversely, assume that $W$ is a Hilbert space. Then we can assume that $W = l^2(X)$ for some set $X$. The surjectivity of $\Phi$ then follows from the surjectivity of the map

$$l^2(X) \to l^2(X)^* \qquad \xi \mapsto \langle -, \xi \rangle$$

due to Thm. 2.60. $\square$

**Definition 3.62.** The map $\Phi$ in Thm. 3.61 is called the **Riesz isometry** of $W$. If $W$ is a Hilbert space, then $\Phi$ is called the **Riesz isomorphism** of $W$. An equivalent description of $\Phi$ is as follows: In view of the isomorphism

$$\mathfrak{Ses}(W) = \mathfrak{L}(W^{\complement} \times W, \mathbb{C}) \simeq \mathfrak{L}(W, (W^{\complement})^*)$$

due to Cor. 2.39, the Riesz isometry $\Phi$ is the element of $\mathfrak{L}(W, (W^{\complement})^*)$ corresponding the the inner product $\langle \cdot | \cdot \rangle_W$ as an elemet of $\mathfrak{Ses}(W)$.

### 3.5.2 Equivalence between bounded linear maps and bounded sesquilinear forms

With the help of the Riesz-Fréchet theorem, we can establish the equivalence between bounded linear maps and bounded sesquilinear forms.

**Theorem 3.63.** *Let $\mathcal{H},\mathcal{K}$ be Hilbert spaces. Then we have an isomorphism of normed vector spaces*

$$\mathfrak{L}(\mathcal{H},\mathcal{K}) \xrightarrow{\simeq} \mathfrak{Ses}(\mathcal{K}|\mathcal{H}) \qquad T \mapsto \omega_T \tag{3.25}$$

*In particular, when $\mathcal{H}=\mathcal{K}$, the above isomorphism becomes*

$$\mathfrak{L}(\mathcal{H}) \xrightarrow{\simeq} \mathfrak{Ses}(\mathcal{H}) \qquad T \mapsto \omega_T \tag{3.26}$$

*Proof.* By Cor. 2.39, we have

$$\mathfrak{L}(\mathcal{H},(\mathcal{K}^{\complement})^*) \simeq \mathfrak{L}(\mathcal{K}^{\complement}\times\mathcal{H},\mathbb{C}) = \mathfrak{Ses}(\mathcal{K}|\mathcal{H})$$

where each $S\in\mathfrak{L}(\mathcal{H},(\mathcal{K}^{\complement})^*)$ corresponds to the bounded bilinear form

$$\mathcal{K}^{\complement}\times\mathcal{H}\to\mathbb{C} \qquad (\overline{\eta},\xi)\mapsto\langle\overline{\eta},S\xi\rangle$$

equivalently, the bounded sesquilinear form

$$\mathcal{K}\times\mathcal{H}\to\mathbb{C} \qquad (\eta,\xi)\mapsto\langle\overline{\eta},S\xi\rangle$$

Now, suppose that $S=\Phi\circ T$ where $T\in\mathfrak{L}(\mathcal{H},\mathcal{K})$, and $\Phi:\mathcal{K}\xrightarrow{\simeq}(\mathcal{K}^{\complement})^*$ is the Riesz-isomorphism of $\mathcal{K}$ defined in Thm. 3.61. Then $\langle\overline{\eta}|\Phi\mu\rangle=\langle\eta|\mu\rangle$ for each $\mu,\eta\in\mathcal{K}$, and hence

$$\langle\overline{\eta},S\xi\rangle=\langle\overline{\eta},\Phi\circ T\xi\rangle=\langle\eta|T\xi\rangle=\omega_T(\eta|\xi)$$

Therefore, the isomorphism

$$\mathfrak{L}(\mathcal{H},\mathcal{K}) \xrightarrow[\simeq]{T\mapsto\Phi\circ T} \mathfrak{L}(\mathcal{H},(\mathcal{K}^{\complement})^*) \simeq \mathfrak{Ses}(\mathcal{K}|\mathcal{H})$$

sends $T$ to $\omega_T$. □

### 3.5.3 Adjoint operators, self-adjoint operators and positive operators

Let $\mathcal{H},\mathcal{K}$ be Hilbert spaces. With the help of Thm. 3.63, we can define adjoint operators:

**Definition 3.64.** Recall that for each $\omega \in \mathfrak{Ses}(\mathcal{K}|\mathcal{H})$, the **adjoint sesquilinear form** $\omega^* \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$ is defined by $\omega^*(\xi|\eta) = \overline{\omega(\eta|\xi)}$ for each $\xi \in \mathcal{H}, \eta \in \mathcal{K}$. It is clear that

$$\|\omega^*\| = \|\omega\|$$

Now, for each $T \in \mathfrak{L}(\mathcal{H}, \mathcal{K})$, define the **adjoint operator** $T^* \in \mathfrak{L}(\mathcal{K}, \mathcal{H})$ such that

$$\omega_{T^*} = (\omega_T)^*$$

More explicitly, $T^*$ is determined by the fact that for each $\xi \in \mathcal{H}, \eta \in \mathcal{K}$,

$$\langle \eta | T\xi \rangle = \langle T^*\eta | \xi \rangle$$

Then, we clearly also have $\|T\| = \|T^*\|$.

**Exercise 3.65.** Show that

$$* : \mathfrak{L}(\mathcal{H}, \mathcal{K}) \to \mathfrak{L}(\mathcal{K}, \mathcal{H}) \qquad T \mapsto T^*$$

is a bijective antilinear map, and that $(T^*)^* = T$. Prove that if $\mathcal{M}$ is a Hilbert space and $T \in \mathfrak{L}(\mathcal{H}, \mathcal{K}), S \in \mathfrak{L}(\mathcal{K}, \mathcal{M})$, then

$$(ST)^* = T^*S^*$$

**Definition 3.66.** A bounded linear operator $T \in \mathfrak{L}(\mathcal{H})$ is called **self-adjoint** if $T = T^*$, equivalently, if $\omega_T$ is self-adjoint.

**Definition 3.67.** Let $A, B \in \mathfrak{L}(\mathcal{H})$. We write

$$\boldsymbol{A \leqslant B}$$

if $\omega_A \leqslant \omega_B$ in the sense of Def. 3.8, that is, $\langle \xi | A\xi \rangle \leqslant \langle \xi | B\xi \rangle$ for all $\xi \in \mathcal{H}$. We say that $A \in \mathfrak{L}(\mathcal{H})$ is **positive** if $A \geqslant 0$, equivalently, if $\omega_A$ is positive.

**Example 3.68.** Let $A \in \mathfrak{L}(\mathcal{H}, \mathcal{K})$. Then $A^*A \in \mathfrak{L}(\mathcal{H})$ is positive, because

$$\langle \xi | A^*A\xi \rangle = \|A\xi\|^2 \geqslant 0$$

### 3.5.4 Composition of bounded linear operators and bounded sesquilinear forms

Let $\mathcal{H}, \mathcal{K}, \mathcal{M}$ be Hilbert spaces.

One of the major advantages of working with bounded linear operators rather than bounded sesquilinear forms is the ease with which one can handle problems involving operator composition. This does not mean, however, that a notion of composition cannot be defined on the side of sesquilinear forms. In fact, the following lemma illustrates how such a composition can be defined.

**Lemma 3.69.** *Let $T \in \mathfrak{L}(\mathcal{K}, \mathcal{H})$ and $S \in \mathfrak{L}(\mathcal{M}, \mathcal{K})$. Let $(e_i)_{i \in I}$ be an orthonormal basis of $\mathcal{K}$. Then for each $\xi \in \mathcal{M}$, we have*

$$T \circ S\xi = \sum_{i \in I} Te_i \cdot \langle e_i | S\xi \rangle \tag{3.27}$$

*where the unordered sum on the RHS converges in norm to the LHS.*

*Proof.* By Thm. 3.45, we have $S\xi = \sum_i e_i \cdot \langle e_i | S\xi \rangle$. Therefore, by the linearity and the continuity of $T$, we get (3.27). □

**Definition 3.70.** Let $\omega \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$ and $\sigma \in \mathfrak{Ses}(\mathcal{K}|\mathcal{M})$. Then the **composition** $\boldsymbol{\omega \circ \sigma}$ is the element of $\mathfrak{Ses}(\mathcal{M}|\mathcal{H})$ defined by

$$(\omega \circ \sigma)(\psi|\xi) = \sum_{i \in I} \omega(\psi|e_i) \cdot \sigma(e_i|\xi) \qquad \text{for all } \psi \in \mathcal{H}, \xi \in \mathcal{M}$$

where $(e_i)_{i \in I}$ is a basis of $\mathcal{K}$. This definition is independent of the choice of basis. Moreover, by Lem. 3.69, for each $T \in \mathfrak{L}(\mathcal{K}, \mathcal{H})$ and $S \in \mathfrak{L}(\mathcal{M}, \mathcal{K})$, we have

$$\omega_{T \circ S} = \omega_T \circ \omega_S$$

However, many properties about composition that are straightforward from the perspective of bounded linear operators become far less transparent when viewed in terms of sesquilinear forms. For instance, consider the following basic result:

**Proposition 3.71.** *Let $\mathcal{U}, \mathcal{V}, \mathcal{W}$ be normed vector spaces. Let $T \in \mathfrak{L}(\mathcal{V}, \mathcal{W})$ and $S \in \mathfrak{L}(\mathcal{U}, \mathcal{V})$. Then*

$$\|TS\| \leqslant \|T\| \cdot \|S\|$$

*Proof.* Apply sup over all $\xi \in \overline{B}_{\mathcal{U}}(0, 1)$ to

$$\|TS\xi\| \leqslant \|T\| \cdot \|S\xi\| \leqslant \|T\| \cdot \|S\| \cdot \|\xi\|$$

□

**Corollary 3.72.** *Let $T \in \mathfrak{L}(\mathcal{H})$. Let $\Omega = \{z \in \mathbb{C} : |z| > \|T\|\}$. Then for each $z \in \Omega$, the operator $z - T$ is invertible (cf. Def. 2.19). Moreover, for each $\xi, \eta \in \mathcal{H}$, the function*

$$z \in \Omega \mapsto \langle \eta | (z - T)^{-1} \xi \rangle = \omega_{(z-T)^{-1}}(\eta|\xi)$$

*is holomorphic.*

The expression $(z - T)^{-1}$ is called the **resolvent** of $T$.

*Proof.* By Prop. 3.71, we have $\|T^k\| \leqslant \|T\|^k$. Therefore, if $z \in \Omega$, then

$$\sum_{k=0}^{\infty} \|z^{-k-1}T^k\| \leqslant \sum_{k=0}^{\infty} |z|^{-k-1}\|T\|^k < +\infty$$

Therefore, if we define

$$S_n(z) = \sum_{k=0}^{n} z^{-k-1}T^k \tag{3.28}$$

Then $(S_n(z))_{n\in\mathbb{N}}$ is a Cauchy sequence in the normed vector space $\mathfrak{L}(\mathcal{H})$. By Cor. 2.35, $\mathfrak{L}(\mathcal{H}) \simeq \mathfrak{Ses}(\mathcal{H})$ is complete. Therefore, $(S_n(z))$ converges under the operator norm to some $S(z) \in \mathfrak{L}(\mathcal{H})$. Since

$$(z-T)S_n(z) = S_n(z)\cdot(z-T) = 1 - z^{-n-1}T^{n+1}$$

and since $\|z^{-n-1}T^{n+1}\| \leqslant |z|^{-n-1}\|T\|^{n+1} \to 0$, we have $(z-T)S(z) = S(z)(z-T) = 1$. This proves that $z-T$ is invertible.

For each $\xi,\eta \in \mathcal{H}$, and for each compact $K \subset \Omega$, we have

$$\sup_{z\in K}\sum_{k=0}^{\infty} |z^{-k-1}\langle\eta|T^k\xi\rangle| \leqslant \sup_{z\in K}\sum_{k=0}^{\infty} |z|^{-k-1}\|T\|^k \cdot \|\eta\| \cdot \|\xi\| < +\infty$$

Therefore, the series of functions

$$z \in \Omega \mapsto \sum_{k=0}^{\infty}\langle\eta|z^{-k-1}T^k\xi\rangle$$

converges absolutely and uniformly on compact subsets of $\Omega$. Since the limit of this series of functions is $z \in \Omega \mapsto \omega_{S_n(z)}(\eta|\xi)$, the latter is holomorphic. $\square$

Before we explore further examples, let us examine another foundational perspective that played a central role in the early development of functional analysis: the viewpoint of bounded matrices.

## 3.6 Bounded matrices

Let $\mathcal{H},\mathcal{K},\mathcal{M}$ be Hilbert spaces.

As mentioned in Subsec. 2.5.1, early developments in functional analysis focused primarily on bounded sesquilinear forms rather than bounded linear operators. Closely tied to this approach was the study of infinite matrices, which provided a concrete representation of these abstract objects.

The notion of boundedness was first defined in this matrix context. Hilbert introduced this concept in [Hil06], where he also introduced the space $l^2(\mathbb{Z})$. As established in Prop. 2.22, boundedness in the context of linear maps or sesquilinear forms is equivalent to (Lipschitz) continuity. However, as we will see below, in the setting of infinite matrices, boundedness takes on a stronger meaning—it implies equicontinuity. More precisely, it ensures that a family of linear maps or sesquilinear forms shares a uniform Lipschitz constant. See Step 2 of the proof of Thm. 3.75.

This distinction highlights a deeper philosophical insight under the perspective of infinite matrices: a bounded linear operator or sesquilinear form is regarded as the limit of a sequence (or net) of finite-rank operators or forms. This philosophy is central to our treatment of spectral theory in Ch. 4, and it resonates not only with the historical approaches of Hilbert and F. Riesz, but also with the viewpoint that the Stieltjes integral arises as the weak-* completion of finite sums (see Table 2.4). For this reason, it is worthwhile to study bounded matrices and their relationship to bounded linear operators and sesquilinear forms.

**Definition 3.73.** Let $X, Y$ be sets. The elements of $\mathbb{C}^{I\times J}$, which are of the form

$$A = (A(x,y))_{x\in X, y\in Y} \qquad \text{where } A_{x,y} \in \mathbb{C}$$

are called ***I* × *J* (complex) matrices**. For each $A \in \mathbb{C}^{X\times Y}$, the **norm $\|A\|$** is defined to be

$$\|A\| = \sup_{I\in\mathrm{fin}(2^X), J\in\mathrm{fin}(2^Y)} \|A_{I,J}\| \tag{3.29a}$$

where each $\|A_{I,J}\|$ is defined by

$$\|A_{I,J}\| = \sup_{\substack{\xi\in\overline{B}_{l^2(I)}(0,1)\\ \eta\in\overline{B}_{l^2(J)}(0,1)}} \Big|\sum_{x\in I, y\in J} \overline{\xi(x)} A(x,y)\eta(y)\Big| \tag{3.29b}$$

We say that $A$ is **bounded** if $\|A\| < +\infty$.

**Definition 3.74.** Suppose that $(e_x)_{x\in X}$ and $(e_y)_{y\in Y}$ are orthonormal basis of $\mathcal{H}$ and $\mathcal{K}$, respectively. The **matrix representation** of each $\omega \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$ is the element $[\omega] \in \mathbb{C}^{X\times Y}$ defined by

$$[\omega](x,y) = \omega(e_x|e_y) \qquad \text{for each } x\in X, y\in Y$$

If $\omega = \omega_T$ where $T \in \mathfrak{L}(\mathcal{K},\mathcal{H})$, we also say that $[\omega]$ is the **matrix representation** of $T$ and write it as $[T]$. In other words, we say that $[T] \in \mathbb{C}^{X\times Y}$ is the matrix representation of $T$ if

$$[T](x,y) = \langle e_x|Te_y\rangle \qquad \text{for each } x\in X, y\in Y$$

**Theorem 3.75.** *Suppose that $(e_x)_{x\in X}$ and $(e_y)_{y\in Y}$ are orthonormal basis of $\mathcal{H}$ and $\mathcal{K}$, respectively. Then the linear map*

$$\mathfrak{Ses}(\mathcal{H}|\mathcal{K}) \to \mathbb{C}^{X\times Y} \qquad \omega \mapsto [\omega] \tag{3.30}$$

*is injective, and its range is the set of all bounded matrices. Moreover, for each $\omega \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$, we have*

$$\|\omega\| = \|[\omega]\|$$

*Proof.* Step 1. We assume WLOG that $\mathcal{H} = l^2(X), \mathcal{K} = l^2(Y)$ and $e_x = \{\chi_{\{x\}}\}_{x\in X}$ and $e_y = \{\chi_{\{y\}}\}_{y\in Y}$. Recall that $C_c(X)$ is dense in $l^2(X)$ and $C_c(Y)$ is dense in $l^2(Y)$.

Let $\omega \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$. From (3.29), it is clear that $\|[\omega]\|$ is the norm of the restriction of $\omega$ to $C_c(X) \times C_c(Y)$. Therefore, by the continuity of $\omega$, we have $\|[\omega]\| = \|\omega\| < +\infty$. In particular, we have proved that the matrices in the range of (3.30) are bounded. Moreover, if $[\omega] = 0$, then $\|\omega\| = \|\omega\| = 0$, and hence $\omega = 0$. This proves that (3.30) is injective.

Step 2. Choose any bounded $A \in \mathbb{C}^{I\times J}$. We want to find $\omega \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$ such that $[\omega] = A$.

For each $I \in \mathrm{fin}(2^X)$ and $J \in \mathrm{fin}(2^Y)$, let $A_{I,J} \in \mathbb{C}^{X\times Y}$ be the **truncation of $A$ by $I, J$**. In other words, for each $x \in X, y \in Y$,

$$A_{I,J}(x,y) = \begin{cases} A(x,y) & \text{if } x \in I, y \in J \\ 0 & \text{otherwise} \end{cases}$$

Then there exists $\omega_{I,J} \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$ whose matrix representation is $A_{I,J}$, namely,

$$\omega_{I,J}(\xi|\eta) = \sum_{x\in I, y\in J} \overline{\xi(x)} A(x,y)\eta(y)$$

We clearly have $\|\omega_{I,J}\| = \|A_{I,J}\|$.

Consider the net $(\omega_{I,J})_{I\times J\in \mathrm{fin}(2^X)\times \mathrm{fin}(2^Y)}$ of sesquilinear forms. The assumption $\|A\| < +\infty$ implies that $\sup_{I,J} \|\omega_{I,J}\| < +\infty$. Moreover, since

$$\lim_{I\in \mathrm{fin}(2^X), J\in \mathrm{fin}(2^Y)} \omega_{I,J}(\chi_{\{x\}}|\chi_{\{y\}}) = A(x,y) \qquad \text{for all } x \in X, y \in Y \tag{3.31}$$

the net $(\omega_{I,J})$ converges pointwise on $C_c(X) \times C_c(Y)$. Therefore, by Thm. 2.32, the net $(\omega_{I,J})$ converges pointwise on $\mathcal{H} \times \mathcal{K}$ to some $\omega \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K})$. By (3.31), we have $[\omega] = A$. □

**Definition 3.76.** Let $X, Y, Z$ be sets. Let $A \in \mathbb{C}^{X\times Y}$ and $B \in \mathbb{C}^{Y\times Z}$ be bounded matrices. Then the **matrix multiplication** $AB \in \mathbb{C}^{X\times Z}$ is defined to be

$$(AB)(x,z) = \sum_{y\in Y} A(x,y)B(y,z)$$

where the RHS is convergent for each $x \in X, z \in Z$. This definition is clearly compatible with Def. 3.70, that is, if $\mathcal{H}, \mathcal{K}, \mathcal{M}$ have orthonormal basis $(e_x)_{x\in X}, (e_y)_{y\in Y}, (e_z)_{z\in Z}$ respectively, and if $\omega \in \mathfrak{Ses}(\mathcal{H}|\mathcal{K}), \sigma \in \mathfrak{Ses}(\mathcal{K}|\mathcal{M})$, then the corresponding matrix representations satisfy

$$[\omega \circ \sigma] = [\omega] \cdot [\sigma]$$

We now return to the topic discussed at the end of Sec. 3.5.4: the subtlety of defining and understanding composition on the side of bounded sesquilinear forms—a subtlety that also arises in the context of bounded matrices. For simplicity, we restrict attention to a fixed Hilbert space $\mathcal{H}$ with orthonormal basis $(e_x)_{x\in X}$.

For $R, S, T \in \mathfrak{L}(\mathcal{H})$, associativity of composition,

$$(RS)T = R(ST)$$

is almost tautological. However, when working with bounded sesquilinear forms or bounded matrices, associativity is far less transparent. To see this, consider $\sigma, \omega, \tau \in \mathfrak{Ses}(\mathcal{H})$. Then the associativity $(\sigma\omega)\tau = \sigma(\omega\tau)$ amounts to the commutativity of the two unordered sums: for all $\xi, \eta \in \mathcal{H}$,

$$\sum_{y\in X}\sum_{x\in X} \sigma(\xi|e_x)\omega(e_x|e_y)\tau(e_y|\eta) = \sum_{x\in X}\sum_{y\in X} \sigma(\xi|e_x)\omega(e_x|e_y)\tau(e_y|\eta)$$

Similarly, if $A, B, C \in \mathbb{C}^{X\times X}$ are bounded matrices, associativity of matrix multiplications means that for each $i, j \in X$,

$$\sum_{y\in X}\sum_{x\in X} A(i,x)B(x,y)C(y,j) = \sum_{x\in X}\sum_{y\in X} A(i,x)B(x,y)C(y,j)$$

The commutativity of $\sum_{x\in X}$ and $\sum_{y\in Y}$ is not at all obvious.

The issue of commutativity of unordered sums—which appears in the frameworks of sesquilinear forms and matrices—disappears in the perspective of linear maps. Where is this Fubini-type property hidden in the linear map viewpoint? And how can one understand the commutativity of such unordered sums in a more general context? We will answer this question in the next section.

# 4 Spectral theorem for bounded self-adjoint operators

## 4.1 Prehistory of spectral theory: continued fractions

## 4.2 Prehistory of spectral theory: the polynomial moment problem

# Index


$l^\infty$-product metric, 6
$l^p$-product metric, 6
*-$\mathbb{K}$-algebra, 13
*-$\mathbb{K}$-subalgebra, 13
Abelian algebra, 13
Absolute continuity of measures, 18
Adjoint operator, 95
Adjoint sesquilinear form, 78
Algebra, 13
Almost convergence of a net of increasing functions, 70
Almost equal increasing functions, 27
Antilinear map, 13, 78
Antiunitary map, 82
Banach-Alaoglu theorem, 60
Bounded matrix, 98
Bounded multilinear map, 51
Bounded sesquilinear forms, 83
Cauchy equivalence of two sequences, 53
Cauchy-Schwarz inequality, 79
Common refinement, 29
Commutative algebra, 13
Complex measure, 25
Composition of sesquilinear forms, 96
Conjugate inner product space $V^{\complement}$, 82
DCT=dominated convergence theorem, 3
Delta function $\delta_c$, 38
Diagonal method, 9
Diameter $\mathrm{diam}(E)$, 4
Diract measure, 16
Direct sum inner product spaces, 86
Equivalent metrics, 6
Fatou's lemma for weak-* convergence, 60
Finer partition, 28
Gram-Schmidt, 84
Helly selection theorem, 71
Hermitian form, 79
Inner product, 81
Inner product space, also called pre-Hilbert space, 81
Interval, 3
Involution, 13
Isomorphic normed vector spaces, 53
Isomorphism of normed vector spaces, 53
LCH=locally compact Hausdorff, 3
Lipschitz continuous, 6
Lusin's theorem, 24
Matrix multiplication, 99
Matrix representation, 98
MCT=monotone convergence theorem, 3
Mesh, 28
Moment, 40
Moment problem, 40
Monotone convergence extension, 59
Multilinear map, 51
Neighborhood, 4
Norm of matrices, 98
Norm of sesquilinear forms, 83
Norm topology, 49, 81
Normed vector subspace, 49
Null space of a positive sesquilinear form, 80
Operator norm $\|T\|$, 51
Orthogonal, 83
Orthogonal complement, 87
Orthogonal decomposition and orthonal projection, 85
Orthonormal, 83
Orthonormal basis, 88
Outer and inner regularity of measures, 22
Parseval's identity, 88
Partition of an interval, 28
Partition of unity in LCH spaces, 11
Pointwise convergence topology, 7

Polarization identity, 78
Polynomial moment problem, 40
Positive bounded linear operators, 95
Positive linear functional, 16
Positive sesquilinear form, 79
Product topology, 7
Projection (operator) associated to a linear subspace, 85
Projection onto a linear subspace, 85
Proper interval, 3
Pullback topology, 7
Pythagorean identity, 84
Quadratic form associated to a sesquilinear form, 78
Radon measure, 22
Radon-Nikodym theorem, 18
Refinement of a partition, 28
Resolvent, 96
Riesz isometry/isomorphism, 93
Riesz-Fischer theorem, Riesz's original version, 46
Riesz-Fischer theorem, the modern form, 19
Riesz-Fréchet representation theorem, 93
Riesz-Fréchet theorem, the classical form, 39
Riesz-Markov representation theorem, 23, 26
Right-continuous normalization, 27
Self-adjoint bounded linear operators, 95
Seminorm, 49
Separable normed vector spaces, 49
Seperate points, 14
Sesquilinear form, 78
Signed measure, 25
Simple function, 16
Step functions, 25
Stieltjes integral, 29, 31
Stieltjes sum, 29
Stone-Weierstrass, 15
Subalgebra, 13
Subalgebra generated by..., 14
Support of a measure, 17
Tagged partition, 29
Tietze extension theorem, 12
Total variation of a complex measure, 74
Trigonometric moment problem, 40
Truncation of matrices, 99
Tychonoff theorem, 9
Uniformly bounded families of functions, 5
Uniformly bounded family of complex measures, 74
Uniformly bounded operator norms, 55
Unit vector, 79
Unital algebra, 13
Unitarily equivalent, 81
Unitary maps, 81
Urysohn function, 11
Urysohn's lemma, 11
Weak-* compact, 60
Weak-* convergence, 59
Weak-* convergence of $(d\rho_\alpha)$ where $(\rho_\alpha)$ is a net of BV functions, 71
Weak-* topology, 59
Weak-* topology on $l^p(X, \mathbb{F})$, 67
Weak-* topology on $L^p(X, \mu, \mathbb{F})$, 62

$B_X(p, r), \overline{B}_X(p, r)$, 4
$\mathscr{Bor}(X, Y)$, the Borel maps $X \to Y$, 5
$BV(I, \mathbb{F})$, 36
BV=bounded variation, 36

$C_0(X, \mathcal{V}), C_0(X, E)$, 12

$\mathrm{diam}(E)$, 4

$f^*(x) = \overline{f(x)}$, 14
$\mathbb{K}\langle \mathfrak{S} \rangle$, 14

$\mathrm{Lin}(V_1 \times \cdots \times V_N, W)$, 51
Lipschitz constant, 6
$\mathfrak{L}(V)$, 51

$\mathfrak{L}(V_1 \times \cdots \times V_N, W)$, 51

$\mathrm{Nbh}_X(x) = \mathrm{Nbh}(x)$, 4

$\mathcal{P}(I)$, the set of partitions of the bounded interval $I$, 28

$\mathcal{Q}(I)$, the directed set of tagged partitions, 29

$\mathcal{RM}(X, \mathbb{K})$ where $\mathbb{K}$ is $\overline{\mathbb{R}}_{\geqslant 0}, \mathbb{R}_{\geqslant 0}, \mathbb{R}, \mathbb{C}$, 73
$\mathrm{Rng}(f)$, the range of $f$, 4

$\mathfrak{Ses}(V|W)$ and $\mathfrak{Ses}(V)$, 83
$S_\rho(f, \sigma, \xi_\bullet)$, 29
$\mathrm{Supp}(\mu)$, 17

$\|T\|$, the operator norm of $T$, 51
$\mathcal{T}_X$, the topology of $X$, 4

$u \perp v$, 83

$V^* = \mathfrak{L}(V, \mathbb{F})$, 51
$v^{\complement} = \overline{v}$, 82
$V^{\complement}$, 82

$\Subset$, 10
$\|\mu\|$, the total variation of $\mu$, 74
$\bigvee_{\alpha \in \mathscr{A}} f_\alpha$, 8
$\chi_A$, 4
$\delta_x$, the Dirac measure at $x$, 16
$\nu \ll \mu$, 18
$\omega^*$, the adjoint sesquilinear form of $\omega$, 78
$\omega_1 \leqslant \omega_2$ where $\omega_1, \omega_2$ are Hermitian forms, 79
$\omega_T$, 92
$\sigma \prec \sigma'$, 28

# References


[Fol-R] Folland, G. B. (1999). Real analysis: modern techniques and their applications. 2nd ed.

[Gray84] Gray, J. D. (1984). The shaping of the Riesz representation theorem: A chapter in the history of analysis. Archive for History of Exact Sciences, 31, 127-187.

[Gud74] Gudder, S. (1974). Inner product spaces. The American Mathematical Monthly, 81(1), 29-36.

[Gui-A] Bin Gui, Qiuzhen Lectures on Analysis. See https://binguimath.github.io/Pages/2023_Analysis.html

[Haw-L] Hawkins, T. (1979) Lebesgue's theory of integration: Its origins and development. Corrected reprint of the 2nd edition.

[Hil06] Hilbert, D. (1906) Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. The fourth part (Vierter Abschnitt, Theorie der quadratischen Formen mit unendlich vielen Variablen).

[Jah] Jahnke, H. N. (2003). A history of analysis (No. 24). American Mathematical Soc..

[Rie09] Riesz, F. (1909). Sur les opérations functionnelles linéaires. Gauthier-Vllars.

[Rie10] Riesz, F. (1910). Untersuchungen über systeme integrierbarer funktionen. Mathematische Annalen, 69(4), 449-497.

[Rie11] Riesz, F. (1911). Sur certains systémes singuliers d'équations intégrales. In Annales scientifiques de l'École Normale Supérieure (Vol. 28, pp. 33-62).

[Rie13] Riesz, F. (1913). Les systèmes d'équations linéaires à une infinite d'inconnues.

[Rie14] Riesz, F. (1914). Démonstration nouvelle d'un théorème concernant les opérations fonctionelles linéaires. Ann. Sci. Éc. Norm. Supér. (3).

[Rie18] Riesz, F. (1918). Über lineare funktionalgleichungen. Acta math, 41(1), 71-98.

[Rud-R] Rudin, W. (1987). Real and complex analysis. 3rd ed.