Notes for Math 4063 (Undergraduate Functional Analysis)

References

In the notes, we will make references to the following books.

[B] Béla Bollobás, *Linear Analysis*, 2nd ed., Cambridge, 1999.

[Be] Berkeley Mathematics Lecture Notes (by Paul Chernoff and William Arveson), Volume 4, 1993.

[CL] Kung-Ching Chang and Yuan-Qu Lin, *Lectures on Functional Analysis* (in Chinese), vol. 1, Beijing University, 1987.

[RS] Michael Reed and Barry Simon, *Functional Analysis*, Vol. I, Academic Press, 1980.

Also, we will cite some results from the books below:

[BN] George Bachman and Lawrence Narici, *Functional Analysis*, Dover, 2000.

[Co] John Conway, *A Course in Functional Analysis*, 2nd ed., Springer-Verlag, 1990.

[D] Sheldon Davis, *Topology*, McGraw-Hill, 2005.

[F] Gerald Folland, *Real Analysis*, 2nd ed., Wiley, 1999.

[Fr] Avner Friedman, *Foundations of Modern Analysis*, Dover, 1982.

[G] Pierre Grillet, *Algebra*, Wiley, 1999.

[H] Paul Halmos, *A Hilbert Space Problem Book*, 2nd ed., Springer-Verlag, 1982.

[HS] Edwin Hewitt and Karl Stromberg, *Real and Abstract Analysis*, Springer-Verlag, 1965.

[Ho] Kenneth Hoffman, *Banach Spaces of Analytic Functions*, Dover, 1988.

[Hu] Thomas Hungerford, *Algebra*, Springer-Verlag, 1974.

[KR] Richard Kadison and John Ringrose, *Fundamentals of the Theory of Operator Algebras*, Vol. III, AMS, 1991.

[Ku] Robert Kuller, *Topics in Modern Analysis*, Prentice-Hall, 1969.

[L] Ronald Larsen, *Functional Analysis*, Marcel Dekker, 1973.

[M] Robert Megginson, *An Introduction to Banach Space Theory*, Springer-Verlag, 1998.

[Mc] Paul McCarthy, *Algebraic Extension of Fields*, Chelsea, 1976.

[Ru] Walter Rudin, *Functional Analysis*, 2nd ed., McGraw-Hill, 1991.

[SS] Lynn Arthur Steen and J. Arthur Seebach, Jr., *Counterexamples in Topology*, 2nd ed., Springer-Verlag, 1978.

[TL] Angus Taylor and David Lay, *Introduction to Functional Analysis*, 2nd ed., Wiley, 1980.

[W] Albert Wilansky, *Functional Analysis*, Blaisdell, 1964.

[Y] Kôsaku Yosida, *Functional Analysis*, 6th ed., Springer-Verlag, 1980.

Abbreviations and Notations

iff	if and only if
$\square$	end of proof
$\mathbb{K}$	$\mathbb{R}$ or $\mathbb{C}$

Chapter 0. Set and Topological Preliminaries.

§1. **Axiom of Choice and Zorn's Lemma.** We begin by introducing the following axiom from set theory.

Axiom of Choice. *Let A be a nonempty set and for every $\alpha \in A$, let S_α be a nonempty set. Let $\mathcal{S} = \{S_\alpha : \alpha \in A\}$. Then there exists a function $f : A \to \bigcup \mathcal{S} = \bigcup\{S_\alpha : \alpha \in A\}$ such that for all $\alpha \in A$, $f(\alpha) \in S_\alpha$.*

From this we can deduce Zorn's lemma, which is a powerful tool in showing the existence of many important objects. To set it up, we need some terminologies.

Definitions. (1) A *relation* R on a set X is a subset of $X \times X$.

(2) For a relation R on X, we now write $x \preceq y$ (or $y \succeq x$) iff $(x, y) \in R$. Also, $x \prec y$ iff $x \preceq y$ and $x \neq y$. R is a *partial ordering* of X iff it satisfies the *reflexive* property ($x \preceq x$ for all $x \in X$), the *antisymmetric* property ($x \preceq y$ and $y \preceq x$ imply $x = y$) and the *transitive property* ($x \preceq y$ and $y \preceq z$ imply $x \preceq z$). X is a *poset* (or a *partially ordered* set) iff there is a partial ordering R on X.

(3) A poset X is *totally ordered* (or *linearly ordered* or *simply ordered*) iff for all $x, y \in X$, either $x \preceq y$ or $y \preceq x$.

(4) A poset X is *well-ordered* iff every nonempty subset G of X has a *least* element in G, i.e. there is $g_0 \in G$ such that for all $g \in G$, $g_0 \preceq g$. (Taking $G = \{x, y\}$, we see X well-ordered implies X totally ordered.)

(5) A *chain* in a poset X is either the empty set or a totally ordered subset of X.

(6) An element u in a poset X is an *upper bound* for a subset S of X iff $x \in S$ implies $x \preceq u$. An element m of X is *maximal* in X iff $m \preceq x$ implies $x = m$. (Similarly lower bound and minimal element may be defined.)

Examples. (1) For $X = \mathbb{R}$ with the usual ordering (i.e. $x \preceq y$ iff $x \leq y$), $\mathbb{R}$ is totally ordered. $(0, \infty)$ is a chain in $X = \mathbb{R}$ with no upper bound in $\mathbb{R}$. $\mathbb{R}$ has no maximal element.

(2) For every set W, the power set $X = P(W) = \{A : A \subseteq W\}$ has a partial ordering given by inclusion (i.e. $A \preceq B$ iff $A \subseteq B$). X is not totally ordered when W has more than one elements since for distinct elements d, e of W, neither $\{d\} \preceq \{e\}$ nor $\{e\} \preceq \{d\}$. W is the unique maximal element in $X = P(W)$.

(3) Let $X = \{2, 3, 4, \ldots\}$. Define $x \preceq y$ iff x is a multiple of y. For example, $24 \preceq 3$ since $24 = 3 \times 8$. Then this makes X a poset and every prime number is a maximal element of X.

Zorn's Lemma. *For a nonempty poset X, if every chain in X has an upper bound in X, then X has at least one maximal element. (The statement is also true if 'upper' and 'maximal' are replaced by 'lower' and 'minimal' respectively.)*

For a proof, see the appendix at the end of the chapter. Below we will present two examples of Zorn's lemma, namely (1) for any two nonempty sets, there exists an injection from one of them to the other and (2) every nonzero vector space has a basis.

Remark. Generalizing example (2) above, let X be a nonempty collection of subsets of some set W. Very often we consider the set inclusion relation $R = \{(A, B) \mid A, B \in X,\ A \subseteq B\}$ on X (i.e. $A \preceq B$ iff $A \subseteq B$). We can easily check X is partially ordered by this relation:

(a) For every $A \in X$, we have $A = A \Longrightarrow A \subseteq A$.

(b) For every $A, B \in X$, we have $A \subseteq B$ and $B \subseteq A \Longrightarrow A = B$.

(c) For every $A, B, C \in X$, we have $A \subseteq B$ and $B \subseteq C \Longrightarrow A \subseteq C$.

Example 1. For nonempty sets A and B, there exists an injective function either from A to B or from B to A.

Proof. Let $W = A \times B$. For $\emptyset \neq C \subseteq A$, let $g : C \to B$ be a function. Then $\Gamma(g) = \{(c, g(c)) \mid c \in C\} \subseteq W$. Let $X = \{\Gamma(g) \mid g : C \to B$ is injective, where $\emptyset \neq C \subseteq A\}$. Define the set inclusion relation on X, i.e. $\Gamma(g_0) \preceq \Gamma(g_1)$ iff $\Gamma(g_0) \subseteq \Gamma(g_1)$. By the remark above, this is a partial ordering on X.

Next for every chain $C = \{\Gamma(g_\alpha) \mid \alpha \in I,\ g_\alpha : C_\alpha \to B$ is injective, where $\emptyset \subset C_\alpha \subseteq A\}$ in X, we will show $S = \bigcup_{\alpha \in I} \Gamma(g_\alpha)$ is in X. (Observe that a nonempty subset T of $W = A \times B$ is an element of X iff for every pair of distinct points $(a, b), (a', b')$ in T, we have $a \neq a'$ (by the definition of function) and $b \neq b'$ (by injectivity).)

Let (a, b) and (a', b') be distinct points in S. Then there are $\alpha, \alpha' \in I$ such that $(a, b) \in \Gamma(g_\alpha)$ and $(a', b') \in \Gamma(g_{\alpha'})$. Since C is a chain in X, we may suppose $\Gamma(g_{\alpha'}) \subseteq \Gamma(g_\alpha)$. Then (a, b) and (a', b') are distinct points in $\Gamma(g_\alpha)$. Since g_α is injective, $a \neq a'$ and $b \neq b'$. Therefore, S is in X. Finally, since for all $\Gamma(g_\alpha) \in C$, $\Gamma(g_\alpha) \subseteq S$, so S is an upper bound of C.

By Zorn's lemma, X has a maximal element $M = \Gamma(f)$. We claim that either the domain of f is A or the range of f is B. Assume not, then there exist $a \in A$ not in the domain of f and $b \in B$ not in the range of f. It follows $M' = M \cup \{(a, b)\}$ is in X and $M \preceq M'$, a contradiction. So the claim is true.

If the domain of f is A, then $f : A \to B$ is injective. If the range of f is B, then $f^{-1} : B \to A$ is injective.

Example 2. Every nonzero vector space W over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ has a basis.

Proof. For a subset S of W, recall that S is linearly independent iff every finite subset of S is linearly independent. Let $X = \{S \mid S$ is a linearly independent subset of $W\}$. By the remark above, the set inclusion relation on X is a partial ordering on X.

For every chain $C = \{S_\alpha \mid \alpha \in I\}$ in X, let $S_I = \bigcup_{\alpha \in I} S_\alpha$. We will check S_I is in X. For every finite subset $\{x_1, x_2, \ldots, x_n\}$ in S_I, there are $\alpha_1, \alpha_2, \ldots, \alpha_n \in I$ such that $x_1 \in S_{\alpha_1}, x_2 \in S_{\alpha_2}, \ldots, x_n \in S_{\alpha_n}$. Since C is a chain, we may assume $S_{\alpha_2}, \ldots, S_{\alpha_n} \subseteq S_{\alpha_1}$. Then $\{x_1, x_2, \ldots, x_n\} \subseteq S_{\alpha_1}$. Since S_{α_1} is linearly independent, so $\{x_1, x_2, \ldots, x_n\}$ is linearly independent. Therefore, S_I is in X. Clearly, S_I is an upper bound of C.

By Zorn's lemma, X has a maximal element M. We claim that the span of M is W. Assume there exists $x \in W$ not in the span of M. By the maximality of M, $M' = M \cup \{x\}$ cannot be in X, i.e. M' is not linearly independent. So there exists $x_1, x_2, \ldots, x_n \in M$ and $c_1, c_2, \ldots, c_n, c \in \mathbb{K}$ (not all zeros) such that $c_1x_1 + c_2x_2 + \cdots + c_nx_n + cx = 0$. Since M is linearly independent, we must have $c \neq 0$. Then $x = (-1/c)(c_1x_1 + c_2x_2 + \cdots + c_nx_n)$ is in the span of M, a contradiction. So the claim is true.

Finally, since $M \in X$ is linearly independent and M spans W, M is a basis of W.

Exercises. (1) Prove that there exists a collection S of pairwise disjoint open disks on a plane such that every open disk on the plane must intersect at least one open disk in S. (*Hint:* Partial order collections consisted of pairwise disjoint open disks.)

(2) Prove that for every integer $n \geq 3$, there exist a set $S_n \subseteq [0, 1]$ such that S_n contains no n-term arithmetic progression, but for every $x \in [0, 1] \setminus S_n$, $S_n \cup \{x\}$ contains a n-term arithmetic progression.

(3) Prove that a normed space X is nonseparable if and only if there exists uncountably many pairwise disjoint open balls of radius 1 in X.

Remarks. (1) Actually the axiom of choice and Zorn's lemma (as well as a few other principles from set theory) are equivalent, see [HS], pp. 14-17.

(2) Zorn's lemma also holds if antisymmetric property of a partial ordering is omitted. See [M], p. 8, ex. 1.16. If 'chain' is replaced by 'well-ordered subset' everywhere, Zorn's lemma and the proof are still correct.

(3) The axiom of choice is used to prove that every set of positive outer Lebesgue measure in $\mathbb{R}$ has non-measurable subsets. (See [Ku], pp. 287-288.) Important applications of Zorn's lemma include the following:

(a) Every nonzero Hilbert space has an orthonormal basis. (See [RS], pp. 44-45.)

(b) In every nonzero ring with an identity, every ideal is contained in a maximal ideal. (See [Hu], p. 128.)

(c) Every field has an algebraic closure. (See [Mc], pp. 21-22.)

---

§2. **Topology.** In the sequel, the phrase a set S in X will mean $S \subseteq X$. Now we begin by introducing the concept of topology on a set X, which generalizes the concept of all open sets in $\mathbb{R}$.

Definitions. (1) Let X be a set and $\mathcal{T}$ be a collection of subsets of X. $\mathcal{T}$ is a *topology* on X iff

(a) $\emptyset, X \in \mathcal{T}$,

(b) the union of any collection of elements of $\mathcal{T}$ is an element of $\mathcal{T}$,

(c) the intersection of finitely many elements of $\mathcal{T}$ is an element of $\mathcal{T}$.

A set X with a topology is called a *topological space.* In case the topology is clear, we simply say X is a topological space. Below let $\mathcal{T}$ be a topology on X.

(2) Let $S \subseteq X$. S is *open* in X iff $S \in \mathcal{T}$. S is *closed* in X iff $X \setminus S \in \mathcal{T}$. (Using de Morgan's law, we can get topological properties for closed sets, namely (a$'$) $\emptyset, X$ are closed, (b$'$) the *intersection* of any collection of closed sets is closed and (c$'$) the *union* of finitely many closed sets is closed.)

(3) Let $S \subseteq X$. The *interior* S° of S is the union of all open subsets of S. (This is the largest open subset of S.) The *closure* $\overline{S}$ of S is the intersection of all closed sets containing S. (This is the smallest closed set containing S.) S is *dense* iff $\overline{S} = X$ (equivalently every nonempty open set in X contains a point of S).

(4) For every $x \in X$, a subset N of X is a *neighborhood* of x iff there exists $U \in \mathcal{T}$ such that $x \in U \subseteq N$.

(5) A subset $\mathcal{T}_0$ of a topology $\mathcal{T}$ on X is a *base* of $\mathcal{T}$ iff whenever $x \in U \in \mathcal{T}$, there exists $V \in \mathcal{T}_0$ such that $x \in V \subseteq U$ (cf Exercise (4) below).

When we are dealing with more than one topologies $\mathcal{T}_1, \mathcal{T}_2, \ldots$, we shall refer to the elements of $\mathcal{T}_1$ as $\mathcal{T}_1$-open sets, the elements of $\mathcal{T}_2$ as $\mathcal{T}_2$-open sets, etc.

Remarks. (1) If $\mathcal{T}_1$, $\mathcal{T}_2$ are topologies on X and $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then we say $\mathcal{T}_1$ is *weaker* than $\mathcal{T}_2$ (or $\mathcal{T}_2$ is *stronger* than $\mathcal{T}_1$). For every set X, there is a weakest topology on X consisted of $\emptyset$ and X. It is called the *indiscrete* topology on X. Also, there is a strongest topology on X consisted of the collection $P(X)$ of all subsets of X. This is called the *discrete* topology on X.

(2) The set of all open sets in a metric space M is a topology on M. It is called the *metric topology* on M. In the case $M = \mathbb{R}^n$ with the usual metric, it is called the *usual* topology. The set of all open balls is a base of the metric topology on M. Every open set in M is a union of open balls.

Exercises. (4) Prove that a subset $\mathcal{T}_0$ of the topology $\mathcal{T}$ on X is a base if and only if every open set is a union of elements of $\mathcal{T}_0$.

(5) Prove that a collection $\mathcal{B}$ of subsets of X is a base of a topology on X if and only if $\bigcup_{V \in \mathcal{B}} V = X$ and for every $V_0, V_1 \in \mathcal{B}$ and $x \in V_0 \cap V_1$, there exists $V_2 \in \mathcal{B}$ such that $x \in V_2 \subseteq V_0 \cap V_1$. (See [D], pp. 47-48.)

---

§§2.1. **Compactness.** We now introduce an important concept in analysis, namely compactness.

Definitions. Let $\mathcal{T}$ be a topology on X and $S \subseteq X$. A subset J of $\mathcal{T}$ is an *open cover* of S iff $\bigcup_{M \in J} M \supseteq S$. S is *compact* in X iff every open cover J of S has a finite subset J_0 which is also an open cover of S. (Such J_0 is a *finite subcover* of J.) S is *precompact* (or *relatively compact*) iff $\overline{S}$ is compact.

Definitions. Let $\mathcal{T}$ be a topology on X and $W \subseteq X$. Then $\mathcal{T}_W = \{S \cap W : S \in \mathcal{T}\}$ is a topology on W called the *relative topology* on W. A subset V of W is *open* in W iff $V \in \mathcal{T}_W$. If $\mathcal{B}$ is a base of $\mathcal{T}$, then $\mathcal{B}_W = \{S \cap W : S \in \mathcal{B}\}$ is a base of $\mathcal{T}_W$.

Remarks. For $V \subseteq W \subseteq X$, if V is open (or closed) in X, then $V = V \cap W$ is open (or closed) in W, respectively. The converse is false as $(0,1]$ is open and closed in $(0,1]$, but neither open nor closed in $\mathbb{R}$.

Intrinsic Property of Compactness. *Let $\mathcal{T}$ be a topology on X and $W \subseteq X$. W is compact in W with the relative topology $\mathcal{T}_W$ iff W is compact in X with topology $\mathcal{T}$.*

Proof. A collection J of open sets in X covers W in X iff $J_W = \{S \cap W : S \in J\}$ covers W in W. J has a finite subcover iff J_W has a finite subcover. $\square$

Remark. Applying de Morgan's law, S compact in X (equivalently, in S) if and only if every collection $\mathcal{F}$ of closed sets in S having the finite intersection property (i.e. the intersection of finitely many members of $\mathcal{F}$ is always nonempty) must satisfy $\bigcap\{W : W \in \mathcal{F}\} \neq \emptyset$.

---

§§2.2. **Continuity.** Observe that if $a < b$ in $\mathbb{R}$, then $\big(-\infty, (a+b)/2\big)$ and $\big((a+b)/2, +\infty\big)$ are disjoint open sets *separating* a and b. This is a property that makes limit unique if it exists. So we introduce the following.

Definition. A set X with a topology $\mathcal{T}$ is a *Hausdorff* space (or a *T_2-space*) iff for every distinct $a, b \in X$, there exist disjoint $U, V \in \mathcal{T}$ such that $a \in U$ and $b \in V$.

Once we have topologies on sets, we can study "continuous" functions between them.

Definitions. Let $\mathcal{T}_X$ and $\mathcal{T}_Y$ be topologies on X and Y respectively.

(1) $f : X \to Y$ is *continuous at* x iff for every neighborhood N of $f(x)$, $f^{-1}(N)$ is a neighborhood of x. $f : X \to Y$ is *continuous* iff for every $\mathcal{T}_Y$-open set U in Y, $f^{-1}(U)$ is a $\mathcal{T}_X$-open set in X (equivalently, for every $\mathcal{T}_Y$-closed set V in Y, $f^{-1}(V)$ is a $\mathcal{T}_X$-closed set in X).

(2) $f : X \to Y$ is a *homeomorphism* iff f is bijective and both f and f^{-1} are continuous. (In this case, U is open in X iff $f(U)$ is open in Y. We say X and Y are *homeomorphic*.)

Exercises. Prove the following properties of topological spaces S, X, Y, Z (see [Be], pp. 15, 34-35).

(6) If $f : X \to Y$ and $g : Y \to Z$ are continuous, then $g \circ f : X \to Z$ is continuous.

(7) If S is compact and X is a closed subset of S, then X is compact.

(8) If S is Hausdorff and Y is a compact subset of S, then Y is closed.

(9) Let $f : X \to Y$ be continuous. If X is compact, then $f(X)$ is compact.

(10) Let X be compact and Y be Hausdorff. If $f : X \to Y$ is continuous and bijective, then f is a homeomorphism.

---

§§2.3. **Nets and Convergence.** In metric space, we know that the closure of a set is consisted of all limits of sequences in the set. However, this is false in general for topological spaces as shown by the following example!

Example. On $[0,1]$, define open sets to be either empty or sets whose complements in $[0,1]$ are countable. More precisely, let $\mathcal{T} = \{\emptyset\} \cup \{S : S \subseteq [0,1], [0,1] \setminus S \text{ is countable}\}$. We can check $\mathcal{T}$ is a topology on $[0,1]$. It is called the *co-countable* topology on $[0,1]$. Now $\{1\} \notin \mathcal{T}$ so that $[0,1)$ is not closed. Hence the $\mathcal{T}$-closure of $[0,1)$ is $[0,1]$. However, every sequence $\{x_n\}$ in $[0,1)$ cannot converge to 1 in the closure of $[0,1)$ because $[0,1] \setminus \{x_1, x_2, x_3, \ldots\}$ is a $\mathcal{T}$-open neighborhood of 1 that does not contain any term of the sequence $\{x_n\}$.

To remedy the situation, we now introduce a generalization of sequence called net.

Definitions. (a) A *directed set* (or *directed system*) is a poset I such that for every $x, y \in I$, there is $z \in I$ satisfying $x \preceq z$ and $y \preceq z$.

(b) A *net* $\{x_\alpha\}_{\alpha \in I}$ in a set S is a function from a directed set I to S assigning every $\alpha \in I$ to a $x_\alpha \in S$.

(c) A net $\{x_\alpha\}_{\alpha \in I}$ is *eventually* in a set W iff $\exists\, \beta \in I, \forall\, \alpha \succeq \beta$, we have $x_\alpha \in W$. A net $\{x_\alpha\}_{\alpha \in I}$ *converges* to x (and we write $\{x_\alpha\}_{\alpha \in I} \to x$ or $x_\alpha \to x$) iff for every neighborhood N of x, $\{x_\alpha\}_{\alpha \in I}$ is eventually in N.

(d) A net $\{x_\alpha\}_{\alpha \in I}$ is *frequently* in a set W iff $\forall\, \beta \in I, \exists\, \alpha \succeq \beta$ such that $x_\alpha \in W$. We say x is a *cluster point* of $\{x_\alpha\}_{\alpha \in I}$ iff for every neighborhood N of x, $\{x_\alpha\}_{\alpha \in I}$ is frequently in N.

(e) A net $\{y_\beta\}_{\beta \in J}$ is a *subnet* of a net $\{x_\alpha\}_{\alpha \in I}$ iff there is a function $n : J \to I$ such that for every $\beta \in J$, $y_\beta = x_{n(\beta)}$ and for every $\alpha \in I$, there exists $\gamma \in J$ such that $\beta \succeq \gamma$ implies $n(\beta) \succeq \alpha$.

Examples. (1) In the case $I = \mathbb{N}$ is the set of positive integers with the usual order, a net is just a sequence. In the case I is an open interval (a, b) of $\mathbb{R}$ with the usual order, a net in W converges to x is just a function from (a, b) to W with the left-handed limit at b equals x. If we reverse the order on (a, b), this becomes the right-handed limit at a equals x.

(2) Convergent nets need not be bounded! For example, let $I = (-\infty, 0)$ with the usual order and $x_\alpha = \alpha$. Then x_α converges to 0, but $\{x_\alpha : \alpha \in I\} = (-\infty, 0)$ is unbounded!

The following theorem on topological spaces generalize the familiar theorems on uniqueness of limit, closure, continuity, cluster point and compactness for metric spaces.

Exercises. Prove the following statements. Let X and Y be topological spaces.

(11) *X is Hausdorff iff every convergent net in X has a unique limit.*

(12) *For every $S \subseteq X$, $\overline{S} = \{x \in X : \exists \{x_\alpha\}_{\alpha \in I}$ in S such that $x_\alpha \to x\}$.*

(13) *A function $f : X \to Y$ is continuous iff f is continuous at every $x \in X$ iff for every $x \in X$ and $\{x_\alpha\}_{\alpha \in I}$ in X with $x_\alpha \to x$, we have $f(x_\alpha) \to f(x)$. If D is dense in X (i.e. $\overline{D} = X$), Y is Hausdorff and $f, g : X \to Y$ continuous with $f|_D = g|_D$, then $f = g$.*

(14) *x is a cluster point of $\{x_\alpha\}_{\alpha \in I}$ iff $\{x_\alpha\}_{\alpha \in I}$ has a subnet converging to x.*

(15) (*Bolzano-Weierstrass Theorem*) *X is compact iff every $\{x_\alpha\}_{\alpha \in I}$ in X has a subnet converging to some $x \in X$ (equivalently, every net in X has a cluster point).*

For proofs, see [Be], pp. 24-26 and 35-36.

Definition. A topological space X is *sequentially compact* iff every sequence in X has a subsequence converging to some $x \in X$.

Remark. In metric spaces, compactness is the same as sequentially compactness (by the metric compactness theorem). For topological spaces, there exists a compact space that is not sequentially compact. So in such a space there is a sequence having a convergent subnet, but no convergent subsequence! Also, there is a sequentially compact set that is not compact. (See [SS], pp. 69 and 126.)

In analysis, we try to solve problems by approximations. The solutions are often some kind of limits of the approximations. So limits of convergent subsequences or convergent subnets are good candidates for the solutions. Therefore, a large part of analysis studies compactness or sequential compactness conditions.

---

§§2.4. **Product Topology.** We begin by asking the following

<u>Questions</u>: If we take a collection Ω of arbitrary subsets of X, must there exist a topology on X that will contain these arbitrary subsets of X. We know $P(X)$ is one such topology. In fact, it is the largest such topology. Is there a smallest such topology?

To answer this question, we can first check that the intersection of any collection of topologies on X is also a topology on X.

<u>Definition.</u> For every collection Ω of subsets of X, the topology $\mathcal{T}_\Omega$ <u>*generated*</u> by Ω is the intersection of all topologies on X containing Ω. Hence, $\mathcal{T}_\Omega$ is the smallest topology on X containing Ω.

<u>Exercise.</u> (16) Prove that $\mathcal{T}_\Omega$ is the collection of all sets that are $\emptyset$ or X or unions of sets of the form $S_1 \cap S_2 \cap \cdots \cap S_n$, where $S_1, S_2, \ldots, S_n \in \Omega$ (i.e. the set of all finite intersections of $S_i \in \Omega$ is a base of $\mathcal{T}_\Omega$).

If we take an open interval (a, b) in $\mathbb{R}$ and form $(a, b) \times \mathbb{R}$ and $\mathbb{R} \times (a, b)$, then we get "open" strips in $\mathbb{R}^2$. More generally, if S is an open set in $\mathbb{R}$, then $S \times \mathbb{R}$ and $\mathbb{R} \times S$ should be "open" in $\mathbb{R}^2$. For two topological spaces X and Y, we would like to introduce a "product" topology on $X \times Y$ based on these ideas.

<u>Definitions.</u> For X with topology $\mathcal{T}_X$ and Y with topology $\mathcal{T}_Y$, we define the <u>*product*</u> topology on $X \times Y$ to be the topology $\mathcal{T}_{X\times Y}$ generated by $\Omega = \{S_1 \times Y : S_1 \in \mathcal{T}_X\} \cup \{X \times S_2 : S_2 \in \mathcal{T}_Y\}$. The functions $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ defined by $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$ are called the <u>*projection maps*</u> onto X and Y, respectively. Since $\Omega = \{\pi_X^{-1}(S_1) : S_1 \in \mathcal{T}_X\} \cup \{\pi_Y^{-1}(S_2) : S_2 \in \mathcal{T}_Y\} \subseteq \mathcal{T}_{X\times Y}$, π_X and π_Y are continuous. By the exercise above and the identity $\cap_{i=1}^k f^{-1}(A_i) = f^{-1}\Big(\cap_{i=1}^k A_i\Big)$, we see

$$\mathcal{B} = \{\pi_X^{-1}(S_1) \cap \pi_Y^{-1}(S_2) = S_1 \times S_2 : S_1 \in \mathcal{T}_X, S_2 \in \mathcal{T}_Y\}$$

is a base of $\mathcal{T}_{X\times Y}$.

More generally, if X_α is a topological space with topology $\mathcal{T}_\alpha$ for every $\alpha \in A$, then the <u>*product*</u> topology on their Cartesian product $X = \prod_{\alpha \in A} X_\alpha$ is the topology generated by the collection Ω of all sets of the form $\pi_\alpha^{-1}(S_\alpha)$, where $S_\alpha \in \mathcal{T}_\alpha$ and $\pi_\alpha : X \to X_\alpha$ is the projection map $\pi_\alpha(x) = x_\alpha$ with x_α denoting the α-coordinate of $x \in X$. So every π_α is continuous. A typical element in the base of the product topology is

$$\pi_{\alpha_1}^{-1}(S_{\alpha_1}) \cap \cdots \cap \pi_{\alpha_n}^{-1}(S_{\alpha_n}) = \bigcap_{i=1}^{n} \{x \in X : \pi_{\alpha_i}(x) \in S_{\alpha_i}\},$$

where $n \in \mathbb{N}, \alpha_1, \ldots, \alpha_n \in A$ and $S_{\alpha_1} \in \mathcal{T}_{\alpha_1}, \ldots, S_{\alpha_n} \in \mathcal{T}_{\alpha_n}$.

In dealing with nets in product topology, we have

<u>Theorem.</u> *A net $\{x_\gamma\}_{\gamma \in I}$ in $X = \prod_{\alpha\in A} X_\alpha$ converges to x iff for every $\alpha \in A$, $\{\pi_\alpha(x_\gamma)\}_{\gamma\in I} \to \pi_\alpha(x)$.*

<u>Proof.</u> Since sets $\pi_{\alpha_1}^{-1}(S_{\alpha_1}) \cap \cdots \cap \pi_{\alpha_n}^{-1}(S_{\alpha_n})$, where $S_{\alpha_i} \in \mathcal{T}_{X_{\alpha_i}}$, form a base of the product topology,

$$\begin{aligned}
\{x_\gamma\}_{\gamma\in I} \to x \iff & \forall\, n \in \mathbb{N},\ \forall\, \alpha_1, \alpha_2, \ldots, \alpha_n \in A, \forall \text{ neighborhood } \pi_{\alpha_1}^{-1}(S_{\alpha_1}) \cap \cdots \cap \pi_{\alpha_n}^{-1}(S_{\alpha_n}) \text{ of } x,\\
& \qquad \exists\, \beta \in I \text{ such that } \gamma \succeq \beta \text{ implies } x_\gamma \in \pi_{\alpha_1}^{-1}(S_{\alpha_1}) \cap \cdots \cap \pi_{\alpha_n}^{-1}(S_{\alpha_n})\\
\iff & \forall\, \alpha_i \in A,\ \forall\, x \in \pi_{\alpha_i}^{-1}(S_{\alpha_i})\ \ \exists\, \beta_i \in I \text{ such that } \gamma \succeq \beta_i \text{ implies } x_\gamma \in \pi_{\alpha_i}^{-1}(S_{\alpha_i})\\
\iff & \forall\, \alpha_i \in A,\ \forall\, \pi_{\alpha_i}(x) \in S_{\alpha_i}\ \ \exists\, \beta_i \in I \text{ such that } \gamma \succeq \beta_i \text{ implies } \pi_{\alpha_i}(x_\gamma) \in S_{\alpha_i}\\
\iff & \forall\, \alpha \in A,\ \{\pi_\alpha(x_\gamma)\}_{\gamma\in I} \to \pi_\alpha(x),
\end{aligned}$$

where in the second step, we take $n = 1$, $\beta_i = \beta$ in the $\Rightarrow$ direction and take $\beta \succeq \beta_i$ for $i = 1, \ldots, n$ in the $\Leftarrow$ direction (such β exists by the definition of directed set). $\square$

Appendix: Proof of Zorn's Lemma

Let us recall

Zorn's Lemma. *For a nonempty poset X, if every chain in X has an upper bound in X, then X has at least one maximal element. (The statement is also true if 'upper' and 'maximal' are replaced by 'lower' and 'minimal' respectively.)*

Proof. (*Due to H. Lenz, H. Kneser and J. Lewin independently*) Assume X has no maximal element. Since every chain C in X has an upper bound $u \in X$ and u is not maximal in X, the set $S_C = \{x \in X : c \in C \Rightarrow c \prec x\} \neq \emptyset$. (Here, $S_\emptyset = X$.) By the axiom of choice, there is a function f such that $f(C) \in S_C$.

We introduce two terminologies.

(a) For a chain C in X, a set of the form $P(C, c) = \{y \in C : y \prec c\}$ for some $c \in C$ is called an *initial segment* of C.

(b) A subset A of X is *conforming* in X iff (1) A is well-ordered by $\preceq$ and (2) for all $a \in A$, $f(P(A, a)) = a$. For example, $A = \{f(\emptyset)\}$ is conforming because $P(A, f(\emptyset)) = \emptyset$ and so $f(P(A, f(\emptyset))) = f(\emptyset)$.

Claim 1: For conforming subsets A, B of X, if $A \neq B$, then one of them is an initial segment of the other.

Proof of claim 1. Since $A \neq B$, either $A \subseteq B$ or $B \subseteq A$ is false, say the former, then $A \setminus B \neq \emptyset$. Let x be least in $A \setminus B$, then since $a \in A$ and $a \prec x$ imply $a \in B$, we have $P(A, x) \subseteq B$.

We will finish by showing $B = P(A, x)$. Assume $P(A, x) \neq B$. Then there is a least $y \in B \setminus P(A, x)$. Observe that for all $u \in P(B, y)$, since $u \in B$, $u \prec y$ and y least in $B \setminus P(A, x)$, we get $u \in P(A, x)$. Then $u \in A$ and $u \prec x$. $\boxed{\text{(*) For all } v \in A \text{ with } v \prec u, \text{ since } v \prec u \prec x, \text{ we have } v \in P(A, x) \subseteq B.}$ Next, since $\emptyset \neq A \setminus B \subseteq A \setminus P(B, y)$, so $A \setminus P(B, y)$ has a least element z.

We will show $P(A, z) = P(B, y)$. (First, $P(A, z) \subseteq P(B, y)$ because $w \in P(A, z)$ implies $w \in A$ and $w \prec z$, the minimality of z implies $w \in P(B, y)$. For the reverse inclusion, $w \in P(B, y)$ implies $w \in B$ and $w \prec y$. The minimality of y implies $w \in P(A, x)$, particularly $w \in A$. If $z \prec w$, then $z \prec y$ and setting $v = z, u = w$ in (*), we get $z \in B$. Then $z \in P(B, y)$, a contradiction. Since $w, z \in B$, so $w \preceq z$. Now $w \neq z$ as $w \in P(B, y)$ and $z \notin P(B, y)$. Hence $w \prec z$, i.e. $w \in P(A, z)$. This gives us $P(B, y) \subseteq P(A, z)$.)

Next $x \in A \setminus B \subseteq A \setminus P(B, y)$ and z is least in $A \setminus P(B, y)$ imply $z \preceq x$. However, $z = f(P(A, z)) = f(P(B, y)) = y \in B$ and $x \notin B$. So $z \neq x$, hence $z \prec x$. Now $y = z \in P(A, x)$, contradicting the definition of y. Then $B = P(A, x)$. So claim 1 is proved.

Claim 2: Let $U = \bigcup\{S : S \text{ conforming in } X\}$, $y \in U$, A conforming in X, $x \in A$ and $y \prec x$. Then $y \in A$.

Proof of claim 2. Assume $y \notin A$. Now $y \in U$ imply $y \in B$ for some conforming B in X. Then $A \neq B$. By claim 1, $A = P(B, w)$ for some w. Then $y \in B$, $x \in A = P(B, w)$ and $y \prec x \prec w$, so $y \in P(B, w) = A$, a contradiction. So claim 2 is proved.

Claim 3: U is conforming.

Proof of claim 3. Let $x, y \in U$. There are conforming A, B such that $x \in A, y \in B$. As claim 1 implies $A \subseteq B$ or $B \subseteq A$ and A, B are totally ordered, so U is also totally ordered.

To see U is well-ordered, let $x \in G \subseteq U$, then x is in some conforming A. If x is not least in G, then $y \in P(G, x) \subset U$ implies $y \in A$ by claim 2. So $P(G, x) \subseteq A$ and hence $P(G, x)$ has a least element d. For all $g \in G$, either $g \succeq x (\succ d)$ or $x \succ g \Rightarrow g \in P(G, x) \Rightarrow g \succeq d$. So d is least in G.

Next to get $x = f(P(U, x))$, note every $x \in U$ is in some conforming A. We will show $P(U, x) = P(A, x)$. First, $A \subseteq U$ implies $P(A, x) \subseteq P(U, x)$. Also $y \in P(U, x)$ implies $y \in A$ by claim 2. So $P(U, x) \subseteq P(A, x)$. Hence they are equal. Then $f(P(U, x)) = f(P(A, x)) = x$. So claim 3 is proved.

Finally, let $x = f(U) \in S_U$, then for all $u \in U$, $u \prec x$. So $x \notin U$. Note $P(U \cup \{x\}, x) = U$ and for $u \in U$, $P(U \cup \{x\}, u) = P(U, u)$. Hence $U \cup \{x\}$ is conforming. By definition of U, we get $x \in U$, a contradiction. $\square$

Chapter 1. Topological Vector Spaces.

§1. **Vector Topology.** In functional analysis, we deal with (usually infinite dimensional) vector spaces X over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and "continuous" linear transformations between them. So we consider vector spaces with topologies and it is natural to require addition and scalar multiplication be continuous.

Notation. We call $\mathbb{K}$ the *scalar field of X* and $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ for all vector spaces to be considered.

Definitions. A vector space X with a topology is a *topological vector space* (or *linear topological space*) iff the topology on X is a *vector* topology (i.e. addition $f : X \times X \to X$ defined by $f(x, y) = x + y$ and scalar multiplication $g : \mathbb{K} \times X \to X$ defined by $g(c, x) = cx$ are continuous with respect to the topology.) For example, the indiscrete topology on X is a vector topology.

Remarks. Let X be a topological vector space. Note $j_b : X \to \{b\} \times X$ given by $j_b(x) = (b, x)$ is continuous since open sets in $\{b\} \times X$ are of the form $\{b\} \times U$, where U is open in X, then $j_b^{-1}(\{b\} \times U) = U$ is open.

(1) For all $a \in X$, $T_a(x) = a + x = (f \circ j_a)(x)$ is a homeomorphism. U is open in X iff $a + U = T_a(U)$ is open in X. A linear function $h : X \to Y$ is continuous iff it is continuous at 0 (i.e. for every neighborhood V of 0 in Y, $h^{-1}(V)$ is a neighborhood of 0 in X). A *base at 0* (or *local base*) is a set $\mathcal{S}$ of neighborhoods of 0 such that every neighborhood of 0 contains a member of $\mathcal{S}$. So $\mathcal{B} = \{a + N : a \in X, N \in \mathcal{S}\}$ is a base for X.

(2) For $c \neq 0$, $g_c(x) = cx = (g \circ j_c)(x)$ is a homeomorphism. So V is a neighborhood of 0 implies cV is a neighborhood of 0.

Definitions. Let X be a vector space over $\mathbb{K}$, $S \subseteq X$ and $c, r \in \mathbb{K}$.

(1) S is *convex* iff $x, y \in S,\ t \in [0, 1]$ implies $tx + (1 - t)y \in S$.

(2) S is *absorbing* iff for every $x \in X$, there is $r > 0$ such that $|c| \le r$ implies $cx \in S$. (Note $0 \in S$.)

(3) S is *balanced* (or *circled*) iff $x \in S, |c| \le 1 \Longrightarrow cx \in S$. S is *absolutely convex* iff it is convex and balanced.

Theorem *In a topological vector space X, every neighborhood S of 0 is absorbing and contains a balanced neighborhood of 0.*

Proof. Let $x \in X$. Since $g : \mathbb{K} \times X \to X$ is continuous and $g(0, x) = 0 \in S$, so $g^{-1}(S)$ is a neighborhood of $(0, x)$. Then there are $r > 0$ and neighborhood U of x such that $(0, x) \in \pi_1^{-1}(B(0, 2r)) \cap \pi_2^{-1}(U) = B(0, 2r) \times U \subseteq g^{-1}(S)$. For $|c| \le r$, since $(c, x) \in B(0, 2r) \times U$, so $cx = g(c, x) \in S$. Hence, S is absorbing.

Next, since $g(0, 0) = 0$, so there are $r_0 > 0$ and neighborhood V of 0 such that $B(0, r_0) \times V \subseteq g^{-1}(S)$. So $g(\lambda, V) = \lambda V \subseteq S$ for all $|\lambda| < r_0$. Let $W = \bigcup\limits_{|\lambda|<r_0} \lambda V$, then W is a balanced neighborhood of 0 inside S. ∎

Finite Dimension Theorem. *Let Y be a vector subspace of a Hausdorff topological vector space X with $\dim Y = n < \infty$. Then every bijective linear transformation $h : \mathbb{K}^n \to Y$ is a homeomorphism and Y is closed in X. So, two Hausdorff vector topologies on a finite dimensional vector space must be identical.*

Proof. The projection $p_i(z_1, \dots, z_n) = z_i$ on $\mathbb{K}^n$ is continuous. Let $\{e_i\}$ be the standard basis of $\mathbb{K}^n$. Then $h(z) = p_1(z)h(e_1) + \cdots + p_n(z)h(e_n)$ is continuous as addition and scalar multiplication are continuous in X.

Conversely, for $\varepsilon > 0$, $S = \{x \in \mathbb{K}^n : \|x\| = \varepsilon\}$ is compact, so $V = h(S)$ is compact. Since X is Hausdorff, V is closed in X. Since $h(0) = 0$ and h is injective, $0 \notin V$. Hence, there is a balanced neighborhood W of 0 disjoint from V in X. Then $E = h^{-1}(W) = h^{-1}(W \cap Y)$ is a balanced neighborhood of 0 disjoint from S. Now $0 \in E$ and being balanced, E is path connected. So $E \subseteq B(0, \varepsilon)$. Then $(h^{-1})^{-1}(B(0, \varepsilon)) = h(B(0, \varepsilon))$ contains $h(E) = W \cap Y$, which is a neighborhood of 0 in Y. Hence h^{-1} is continuous.

Let $p \in \overline{Y}$, say some net $\{p_\alpha\}$ in Y converges to p. Since W is absorbing, there exists $t > 0$ such that $p \in tW$. Then the net $\{p_\alpha\}$ is eventually in tW. So $p \in \overline{Y \cap tW} = \overline{h(tE)} \subseteq \overline{h(t\overline{B(0, \varepsilon)})} = h(t\overline{B(0, \varepsilon)}) \subseteq Y$, where the last equality follows from $h(t\overline{B(0, \varepsilon)})$ is compact, hence closed in X. So Y is closed in X. ∎

Definitions. Let X, Y be vector spaces. For a linear function $T : X \to Y$, the *kernel* (or *null space*) of T is $\ker T = T^{-1}(\{0\}) = \{x \in X : T(x) = 0\}$ and the *range* of T is $\operatorname{ran} T = T(X) = \{Tx : x \in X\}$. (Another notation for kernel of T is $N(T)$ and for range of T is $R(T)$.)

Closed Kernel Theorem. *For a topological vector space X and a linear function $T : X \to \mathbb{K}$, $\ker T$ is closed if and only if T is continuous. ($\mathbb{K}$ cannot be replaced by X or Y, see [W], p. 113, ex. 3.)*

Proof. The if direction is clear. In the only-if direction, for a $x \in X \setminus \ker T$, there is a balanced neighborhood V of 0 such that $x \in x+V \subseteq X \setminus \ker T$, i.e. $(x+V) \cap \ker T = \emptyset$. Then $0 \notin T(x+V)$. So $T(V)$ cannot contain $-T(x) \in \mathbb{K}$. Since V is balanced, $T(V)$ is balanced in $\mathbb{K}$. So $T(V)$ is a subset of $B(0,r) = \{z \in \mathbb{K} : |z| < r\}$, where $r = |T(x)|$. Then for all $\varepsilon > 0$, $T(\frac{\varepsilon}{r}V) \subseteq B(0,\varepsilon)$. So $T^{-1}(B(0,\varepsilon)) \supseteq \frac{\varepsilon}{r}V$. So T is continuous at 0. ∎

§2. **Normed Spaces.** One common type of topological vector spaces that we will deal with frequently is the family of normed linear spaces.

Definitions. (1) A *semi-norm* on a vector space X is a function that assigns every $x \in X$ a number $\|x\| \in \mathbb{R}$ satisfying (a) $\|x\| \geq 0$ for all $x \in X$, (b) $\|cx\| = |c|\|x\|$ for all $c \in \mathbb{K}, x \in X$ and (c) $\|x+y\| \leq \|x\| + \|y\|$ for all $x, y \in X$. It is a *norm* iff in addition to (a), (b), (c), we also have $\|x\| = 0$ implies $x = 0$.

(2) A *normed space* (or *normed linear space* or *normed vector space*) is a vector space with a norm. A *Banach space* is a complete normed space (where complete means all Cauchy sequences converge). For inner product space V, define $\|x\| = \sqrt{\langle x, x\rangle}$ for all $x \in V$. This makes V a normed space. A *Hilbert space* is a complete inner product space.

(3) For topological vector spaces X and Y, a linear transformation from X to Y is also called a *linear operator*. In case $Y = \mathbb{K}$, it is also called a *linear functional*. Let $L(X,Y)$ denote the set of all continuous linear operators from X to Y. In case $X = Y$, we write $L(X)$ for $L(X,X)$. (Instead of $L(X,Y)$, the notations $B(X,Y)$, $\mathcal{L}(X,Y)$ or $\mathcal{B}(X,Y)$ are also common.)

(4) For a topological vector space X over $\mathbb{K}$, we write X^* for $L(X,\mathbb{K})$ and call it the *dual space* (or *conjugate space*) of X. The elements of X^* are called the *continuous linear functionals on X*.

(5) For a topological vector space X over $\mathbb{K}$, the *twin* of X is X_{twin}, which has the same elements, same addition and same topology as X, but scalar multiplication cx in X_{twin} equals $\overline{c}x$ in X. If $\mathbb{K} = \mathbb{R}$, then $X_{twin} = X$.

Examples. (1) Let X be a normed space. For every $x \in X$ and linear $T : X \to \mathbb{K}$, the function $p_T(x) = |T(x)|$ is easily checked to be a semi-norm on X. It is a norm if and only if $\ker T = \{0\}$.

(2) Let X, Y be normed spaces. For $T \in L(X,Y)$, define $\|T\| = \sup\{\|T(x)\| : \|x\| \leq 1\}$ and we say T is *bounded* as $\|T\| < \infty$. It is easy to check that $L(X,Y)$ is a normed space. If Y is complete, later we will show $L(X,Y)$ (hence X^*) is complete.

(3) $\mathbb{K}^n$ with inner product $\langle (w_1, w_2, \ldots, w_n), (z_1, z_2, \ldots, z_n)\rangle = w_1\overline{z_1} + w_2\overline{z_2} + \cdots, w_n\overline{z_n}$ and norm $\|(z_1, \ldots, z_n)\|$ $= \sqrt{|z_1|^2 + \cdots + |z_n|^2}$ is a Hilbert space. $(\mathbb{K}^n)^* = \mathbb{K}^n_{twin}$. For every Hilbert space H, $H^* = H_{twin}$.

(4) The set $P([0,1])$ of all polynomials on $[0,1]$ with $\|f\| = \sup\{|f(x)| : x \in [0,1]\}$ is a normed space that is not complete. By the Weierstrass approximation theorem, $P([0,1])$ is dense in the set of all continuous functions $C([0,1])$ on $[0,1]$ with the same norm.

In general, for a compact set X, let $C(X)$ be the set of all continuous functions from X to $\mathbb{K}$ with sup-norm $\|f\| = \sup\{|f(x)| : x \in X\}$. Then $C(X)$ is a Banach space. For a description of the dual of $C(X)$, see Rudin's *Real and Complex Analysis*, 3rd. ed, p. 130.

(5) For $1 \leq p < \infty$, $\ell^p = \{(a_1, a_2, a_3, \ldots) : a_i \in \mathbb{K},\ \|(a_1, a_2, a_3, \ldots)\|_p = (|a_1|^p + |a_2|^p + |a_3|^p + \cdots)^{1/p} < \infty\}$ is a Banach space. The dual of ℓ^p is ℓ^q_{twin}, where $\dfrac{1}{p} + \dfrac{1}{q} = 1$ and such q is called the conjugate index of p. (Instead of ℓ^p, the notation ℓ_p is also common.)

(6) $\ell^\infty = \{(a_1, a_2, a_3, \ldots) : a_i \in \mathbb{K},\ \|(a_1, a_2, a_3, \ldots)\|_\infty = \sup\{|a_i| : i \in \mathbb{N}\} = \inf\{M : |a_i| \leq M,\ \forall\, i \in \mathbb{N}\} < \infty\}$ is a Banach space. For its dual, see Alberto Torchinsky's book *Real Variables*, p. 292. The spaces

$$c = \{(a_1, a_2, a_3, \ldots) : a_i \in \mathbb{K},\ \lim_{i\to\infty} a_i \in \mathbb{K}\} \quad \text{and} \quad c_0 = \{(a_1, a_2, a_3, \ldots) : a_i \in \mathbb{K},\ \lim_{i\to\infty} a_i = 0\}$$

are Banach subspaces of ℓ^∞ with the same norm as ℓ^∞. The duals of c and c_0 are ℓ^1.

(7) For $1 \le p < \infty$ and measurable $X \subseteq \mathbb{R}$, the Lebesgue spaces

$$L^p(X) = \{[f] : f \text{ measurable on } X,\ \|f\|_p = \Big(\int_X |f|^p dm\Big)^{1/p} < \infty\},$$

where $[f]$ denotes the set of measurable functions equal to f almost everywhere, is a Banach space. We have $(L^p)^* = L^q_{twin}$, where $\dfrac{1}{p} + \dfrac{1}{q} = 1$, see Rudin's book <u>*Real and Complex Analysis*</u>, 3rd. ed, p. 127.

Also, $L^\infty(X)$ consisted of all $[f]$'s with finite <u>*essential sup-norm*</u> $\|[f]\| = \inf\{M : |f(x)| \le M \text{ a.e.}\}$ is a Banach space. For its dual, see Alberto Torchinsky's book <u>*Real Variables*</u>, p. 292.

(8) Let X, Y be normed spaces. For $1 \le p < \infty$, we may define $X \oplus_p Y = \{(x, y) : x \in X, y \in Y\}$ with $\|(x, y)\|_p = (\|x\|^p + \|y\|^p)^{1/p}$. It is easy to check that $X \oplus_p Y$ is a normed space with $\|\cdot\|_p$ as norm. For $p = \infty$, define $\|(x, y)\|_\infty = \max\{\|x\|, \|y\|\}$ as norm. All these norms are equivalent. We called $X \oplus_2 Y$ the <u>*direct sum*</u> of X and Y. If X, Y are Banach spaces, then $X \oplus_2 Y$ is also a Banach space. For Hilbert spaces X and Y, the direct sum $X \oplus_2 Y$ with the inner product given by $\langle (x_1, y_1), (x_2, y_2)\rangle = \langle x_1, x_2\rangle + \langle y_1, y_2\rangle$ inducing the norm $\|(x, y)\|_2 = (\|x\|^2 + \|y\|^2)^{1/2}$ is a Hilbert space. For $1 < p < \infty, p \ne 2$ and q the conjugate index of p, $L^p \oplus_2 L^q$ is not a Hilbert space, but the dual of $L^p \oplus_2 L^q$ is its twin, like Hilbert spaces.

The <u>*projection map*</u> $P_X : X \oplus_p Y \to X$ defined by $P_X(x, y) = x$ is continuous since $\|x\| \le \|(x, y)\|_p$ and similarly, the projection map $P_Y : X \oplus_p Y \to Y$ defined by $P_Y(x, y) = y$ is continuous.

(9) Let N be a closed vector subspace of a normed space X. For $x \in X$, we define $[x] = x+N = \{x+n : n \in N\}$ and $X/N = \{[x] : x \in X\}$. Note $[x] = [x']$ if and only if $x - x' \in N$. For $c \in \mathbb{K}$ and $x, y \in Y$, defining $[x] + [y] = [x + y]$ and $c[x] = [cx]$ shows X/N is a vector space with $[0] = 0 + N = N$.

Next define $\|[x]\| = \inf\{\|x - n\| : n \in N\}$. We have $\|[x]\| = 0$ implies there is a sequence $\{n_k\}$ in N such that $\|x - n_k\| \to 0$ so that $n_k \to x \in \overline{N} = N$ and $[x] = [0]$. It is easy to see that this makes X/N a normed space. We call X/N the <u>*quotient normed space of X by N*</u> and $\|[\cdot]\|$ the <u>*quotient norm*</u>. The linear surjection $\pi_N : X \to X/N$ defined by $\pi_N(x) = [x]$ is called the <u>*quotient map*</u>. It is continuous since $\|[x]\| = \inf\{\|x - n\| : n \in N\} \le \|x\|$. Also, $\pi_N(B(0, 1)) = B([0], 1)$ implies π_N maps open sets to open sets.

<u>**Theorem.**</u> *If N is a closed vector subspace of a Banach space X, then X/N is also a Banach space.*

<u>**Proof.**</u> Recall that a normed space is complete iff every absolutely convergent series converges in the space. Suppose $\sum_{k=1}^{\infty} \|[x_k]\| < \infty$. By infimum property, for every k, there exists $n_k \in N$ such that $\|x_k - n_k\| \le 2\inf\{\|x_k - n\| : n \in N\} = 2\|[x_k]\|$. Then $\sum_{k=1}^{\infty} \|x_k - n_k\| < \infty$. Since X is complete, this implies $\sum_{k=1}^{\infty}(x_k - n_k)$ converges to some $x \in X$. Using $\|[w]\| \le \|w\|$ for all $w \in X$, we have

$$\Big\|\sum_{k=1}^{m}[x_k] - [x]\Big\| = \Big\|\Big[\sum_{k=1}^{m} x_k - x\Big]\Big\| = \Big\|\Big[\sum_{k=1}^{m} x_k - x - \underbrace{\sum_{k=1}^{m} n_k}_{in N}\Big]\Big\| \le \Big\|\sum_{k=1}^{m}(x_k - n_k) - x\Big\| \to 0 \quad \text{as } m \to \infty.$$

$\square$

<u>**Remarks.**</u> The same reasoning also show that if E is a subspace of a Banach space X such that $E + N$ is closed (hence complete) in X, then $(E + N)/N$ is complete, hence closed in X/N.

<u>**Definition.**</u> For a closed vector subspace N of a Banach space X, define the <u>*codimension*</u> of N in X to be $\operatorname{codim} N = \dim X/N$.

<u>**Remark.**</u> In [RS], pp. 102-103, there is a nice functional analysis proof of the Tietze extension theorem on compact spaces using quotient spaces.

Chapter 2. Basic Principles.

§1. **Consequences of Baire's Category Theorem.** In this and next sections, we will study important principles about linear operators between topological vector spaces. The four pillars of functional analysis are the open mapping theorem, the closed graph theorem, the uniform boundedness principle and the Hahn-Banach theorem. They have many applications in different branches of mathematics. We will cover the first three of these in this section and the last one in the next section.

Definition. For topological spaces X and Y, $T : X \to Y$ is *open* iff U open in X implies $T(U)$ open in Y.

Remarks. (1) In checking $T : X \to Y$ is open, it is enough to check $T(U)$ is open for U's in a base of $\mathcal{T}_X$. Then T open follows from $T(\cup_\alpha U_\alpha) = \cup_\alpha T(U_\alpha)$. For example, every quotient map $\pi : X \to X/N$ of normed spaces is open since $\pi(B(a,r)) = B([a],r)$. Also, a projection $\pi_\beta : \prod_{\alpha \in A} X_\alpha \to X_\beta$ is an open map since for open sets S_{α_i} in X_{α_i}, $\pi_\beta\big(\pi_{\alpha_1}^{-1}(S_{\alpha_1}) \cap \cdots \cap \pi_{\alpha_n}^{-1}(S_{\alpha_n})\big) = S_{\alpha_i}$ or X_β depending if $\beta = \alpha_i$ for some i or not.

(2) An open map may not take closed sets to closed sets. To see this, let $X = P([0,1])$ and $Y = C([0,1])$ be the sets of all polynomials and continuous function on $[0,1]$ with sup-norm, respectively. Then $V = \{(f,f) : f \in P([0,1])\}$ is closed in $X \times Y$ because $(f_n, f_n) \to (f,g)$ in $X \times Y$ implies $f_n \to f$ in $X (\subset Y)$ and $f_n \to g$ in Y, hence, by uniqueness of limit in Y, $f = g$ and so $(f,g) \in V$. The projection map $\pi_Y : X \times Y \to Y$ is open, but $\pi_Y(V) = X$ is not closed in Y since $\overline{X} = Y \neq X$ by the Stone-Weierstrass theorem.

(3) If a vector subspace M contains some $B(a,r)$ in a normed space Y, then $M = \operatorname{span}\{B(a,r) - a\} = \operatorname{span}\{B(0,r)\} = Y$. So if linear $T : X \to Y$ is open (or just $M = T(X)$ contains a ball of Y), then T is surjective. Is there any converse? See the open mapping theorem below.

Lemma 1. *Let X and Y be normed spaces. A linear function $T : X \to Y$ is open if and only if there exist $r, r' > 0$ such that $T(B(0,r)) \supseteq B(0,r')$.*

Proof. If T is open, then $T(B(0,r))$ is open and contains 0. So $T(B(0,r)) \supseteq B(0,r')$ for some $r' > 0$.

If $T(B(0,r)) \supseteq B(0,r')$, then since every open U in X is a union of $B(a,r_a) = a + (r_a/r)B(0,r)$, so $T(U) = T\Big(\bigcup_{a \in U} B(a,r_a)\Big) = \bigcup_{a \in U}\Big(T(a) + \dfrac{r_a}{r}T(B(0,r))\Big) \supseteq \bigcup_{a \in U} B\Big(T(a), \dfrac{r_a r'}{r}\Big) \supseteq \bigcup_{a \in U}\{T(a)\} = T(U)$. Therefore, $T(U)$ is the union of $B(T(a), r_a r'/r)$, hence is open. ∎

Lemma 2. *Let X be a Banach space, Y be a normed space and $T \in L(X,Y)$. If $\overline{T(B(0,r))} \supseteq B(0,r')$, then $T(B(0,r)) \supseteq B(0,r')$.*

Proof. Let $y \in B(0,r')$. Choose c such that $\|y\|/r' < c < 1$. Let $\varepsilon \in (0, 1-c)$. Then $y \in cB(0,r') \subseteq \overline{T(cB(0,r))}$. So y is limit of Tx's with $x \in cB(0,r)$. Then there is $x_1 \in cB(0,r)$ such that $\|y - Tx_1\| < \varepsilon c r'$. So $y - Tx_1 \in \varepsilon c B(0,r') \subseteq \overline{T(\varepsilon c B(0,r))}$. Iterating this, we get by induction a sequence $\{x_n\}$ in X such that $x_n \in \varepsilon^{n-1} c B(0,r)$ and $y - Tx_1 - \cdots - Tx_n \in \varepsilon^n c B(0,r')$. Now $\sum_{n=1}^{\infty} \|x_n\| < \dfrac{cr}{1-\varepsilon} < r$. Since X is complete, $\sum_{n=1}^{\infty} x_n = x$ for some $x \in B(0,r)$. Since T is continuous, $\|y - Tx\| = \lim_{n\to\infty} \|y - Tx_1 - \cdots - Tx_n\| \leq \lim_{n\to\infty} \varepsilon^n c r' = 0$. Then $y = Tx \in T(B(0,r))$. ∎

Open Mapping Theorem. *For Banach spaces X, Y and $T \in L(X,Y)$, if T is surjective, then T is open.*

Proof. Let $U_n = B(0,n)$ in X. Since $T(X) = T\big(\bigcup_{n=1}^{\infty} U_n\big) = \bigcup_{n=1}^{\infty} T(U_n)$ is of the second category in Y by the Baire category theorem, there is n such that $\overline{T(U_n)}$ contains an open ball, say $B(Ta,r) = Ta + B(0,r)$, where $a \in U_n$. Then $B(0,r) = -Ta + B(Ta,r) \subseteq -Ta + \overline{T(U_n)} \subseteq \overline{T(U_n)} + \overline{T(U_n)} \subseteq \overline{T(U_{2n})}$. By the lemmas above, $B(0,r) \subseteq T(U_{2n})$ and T is open. ∎

Remark. Let X be a Banach space, Y be a normed space and $T \in L(X,Y)$. The proof above actually showed if $T(X)$ is of second category in Y, then T is open (and surjective by remark (3) above).

Definitions. Let X, Y be normed spaces. $T \in L(X,Y)$ is *invertible* iff T is bijective and $T^{-1} \in L(Y,X)$. X and Y are *isomorphic* iff there is an invertible $T \in L(X,Y)$. (Such an invertible T is called an *isomorphism* between X and Y. In that case, there exist $c_1, c_2 > 0$ such that for all $x \in X$, $c_1\|x\| \le \|Tx\| \le c_2\|x\|$.)

Inverse Mapping Theorem. *For Banach spaces X and Y, if $T \in L(X,Y)$ is bijective, then $T^{-1} \in L(Y,X)$.*

Proof. For $T \in L(X,Y)$, T bijective is equivalent to T injective and open (by the open mapping theorem and remark (3)). For all open U in X, $(T^{-1})^{-1}(U) = T(U)$ is open in Y. So T^{-1} is continuous. $\square$

Isomorphism Theorem. *For normed spaces X, Y and $T \in L(X,Y)$, the linear function $\widehat{T} : X/\ker T \to Y$ defined by $\widehat{T}([x]) = T(x)$ is bounded and $\|\widehat{T}\| = \|T\|$. (In case X and Y are Banach spaces, if $T \in L(X,Y)$ is surjective, then $\widehat{T} \in L(X/\ker T, Y)$ is an isomorphism and $X/\ker T$ is isomorphic to Y as Banach spaces.)*

Proof. For all $n \in \ker T$, $\|\widehat{T}([x])\| = \|Tx\| = \|T(x-n)\| \le \|T\|\|x-n\|$. Taking infimum over all $n \in \ker T$, we get $\|\widehat{T}([x])\| \le \|T\|\|[x]\|$. So $\widehat{T}$ is bounded and $\|\widehat{T}\| \le \|T\|$. Next, $\|T(x)\| = \|\widehat{T}([x])\| \le \|\widehat{T}\|\|[x]\| \le \|\widehat{T}\|\|x\|$ implies $\|T\| \le \|\widehat{T}\|$. Therefore, $\|\widehat{T}\| = \|T\|$.

In case X and Y are Banach spaces, if $T \in L(X,Y)$ is surjective, then $\widehat{T} \in L(X/\ker T, Y)$ is bijective. By the inverse mapping theorem, $\widehat{T}$ is an isomorphism. $\square$

Remarks. Using the inverse mapping theorem, it can be showed that there exists a complex sequence with limit zero such that it is not the Fourier coefficient sequence of a L^1 function on the unit circle. See applications at the end of the chapter.

Definition. Let X, Y be normed spaces. $T \in L(X,Y)$ is *bounded below* iff there exists $c' > 0$ such that for all $x \in X$, $\|Tx\| \ge c'\|x\|$.

Remarks. (1) Taking $u = x/\|x\|$, the inequality is the same as $\inf\{\|T(u)\| : \|u\| = 1\} > 0$. So T is not bounded below iff there is a sequence $u_n \in X$ such that $\|u_n\| = 1$ and $T(u_n) \to 0$.

(2) If $T \in L(X,Y)$ is bounded below and W is a complete subset of X, then $T(W)$ is also a complete subset in Y (since for $x_n \in W$, $\{Tx_n\}$ Cauchy implies $\{x_n\}$ Cauchy, hence by completeness of W, $x_n \to x$ for some $x \in W$ and by continuity of T, $Tx_n \to Tx \in T(W)$). In case X is a Banach space, T bounded below and W closed subset in X imply $T(W)$ closed in Y.

Lower Bound Theorem. *Let X be a Banach space and Y be a normed space. For $T \in L(X,Y)$, the following are equivalent:*

(a) *T is bounded below,*
(b) *T is injective and $T(X)$ is complete (hence closed in Y),*
(c) *T has a continuous inverse $T^{-1} : T(X) \to X$.*

Proof. (a) $\Rightarrow$ (b) If T is bounded below, then $T(x) = 0$ implies $x = 0$, so T is injective. By remark (2), $T(X)$ is complete (hence closed in Y).

(b) $\Rightarrow$ (c) This follows immediately from the inverse mapping theorem.

(c) $\Rightarrow$ (a) If $T^{-1} \in L(T(X), X)$, then $\|x\| = \|T^{-1}(Tx)\| \le \|T^{-1}\|\|Tx\|$ for all $x \in X$ and we can take $c' = 1/\|T^{-1}\|$ (unless $T^{-1} = 0$, i.e. $X = \{0\}$, then take $c' = 1$). $\square$

Remarks. Let X and Y be Banach spaces. $T \in L(X,Y)$ is invertible if and only if T is injective and $T(X)$ is closed and dense in Y if and only if T is bounded below and $T(X)$ is dense in Y. For injective $T \in L(X,Y)$, $T(X)$ is closed iff T is bounded below.

For the next theorem, we introduce the

Definition. For topological spaces X and Y, $T : X \to Y$ is *closed* iff its graph $\Gamma(T) = \{(x, Tx) : x \in X\}$ is closed in $X \times Y$ (i.e. if $(x_\alpha, Tx_\alpha) \to (x,y) \in \overline{\Gamma(T)}$, then $y = Tx$ so that $(x,y) \in \Gamma(T)$).

Remark. For Hausdorff space Y, if $T : X \to Y$ is continuous, then T is closed (since $(x_\alpha, Tx_\alpha) \to (x, y)$ and $Tx_\alpha \to Tx$ by continuity imply $y = Tx$ by uniqueness of limit). Are there any converse? See the next theorem.

Recall that the projection maps $\pi_1 : X \times Y \to X$ defined by $\pi_1(x, y) = x$ and $\pi_2 : X \times Y \to Y$ defined by $\pi_2(x, y) = y$ are continuous.

Closed Graph Theorem. *Let X, Y be Banach spaces and $T : X \to Y$ be linear. If T is closed, then T is continuous.*

Proof. Since X and Y are complete, so $X \times Y$ is complete. Since $\Gamma(T) = \{(x, Tx) : x \in X\}$ is closed in $X \times Y$, $\Gamma(T)$ is complete. Note $\pi_1\big|_{\Gamma(T)} : \Gamma(T) \to X$ is bijective. Also, π_1 continuous implies $\pi_1\big|_{\Gamma(T)} \in L(\Gamma(T), X)$. By the inverse mapping theorem, $\pi_1\big|_{\Gamma(T)}^{-1} \in L(X, \Gamma(T))$. Therefore, $T = \pi_2 \circ \pi_1\big|_{\Gamma(T)}^{-1} \in L(X, Y)$. ◻

Exercises. (1) Let X be a vector space equipped with two complete norms $\|\cdot\|_1$ and $\|\cdot\|_2$. If there exists $c > 0$ such that for all $x \in X$, $\|x\|_1 \le c\|x\|_2$, prove that there exists $c' > 0$ such that for all $x \in X$, $\|x\|_2 \le c'\|x\|_1$. This means the norms are equivalent.

(2) (*Hellinger-Toeplitz Theorem*) Let H be a Hilbert space and $T : H \to H$ be a linear transformation such that for all $x, y \in H$, $\langle x, Ty\rangle = \langle Tx, y\rangle$. Prove that T is bounded. (This theorem has important consequence in mathematical physics. See [RS], p. 84)

Application. See [Fr], pp. 145-149 or [Y], pp. 80-81 for applications of the closed graph theorem to PDE.

Uniform Boundedness Principle (or Resonance Theorem). *Let X, Y be normed spaces, $A \subseteq L(X, Y)$ and S be of the second category in X. If A is pointwise bounded on S (i.e. $\{\|Tx\| : T \in A\}$ is bounded for every $x \in S$), then A is uniformly bounded (i.e. $\{\|T\| : T \in A\}$ is bounded). Thus, if X is a Banach space and A is pointwise bounded on X, then A is uniformly bounded.*

Proof. Note $S_n = \{x \in X : \forall\, T \in A, \|Tx\| \le n\} = \bigcap_{T \in A} \{x \in X : \|Tx\| \le n\}$ is closed. Since $S \subseteq \bigcup_{n=1}^{\infty} S_n$, $\bigcup_{n=1}^{\infty} S_n$ is also of the second category in X. Then there is a S_n containing some ball $B(x, r)$. Hence $S_n \supseteq \overline{B(x, r)} = x + \overline{B(0, r)}$. For every $\|y\| \le 1$, since $x \in S_n$ and $x + ry \in \overline{B(x, r)} \subseteq S_n$, so for all $T \in A$,

$$\|Ty\| = \frac{\|T(ry)\|}{r} \le \frac{\|T(x + ry)\| + \|Tx\|}{r} \le \frac{2n}{r}.$$

Therefore, for every $T \in A$, $\|T\| \le 2n/r$. ◻

Theorem (Banach-Steinhaus). *Let X be a Banach space, Y be a normed space and $T_n \in L(X, Y)$.*

(a) *If for all $x \in X$, $\{T_n x\}$ converges in Y, then $Tx = \lim_{n\to\infty} T_n x \in L(X, Y)$ with $\|T\| \le \liminf_{n\to\infty} \|T_n\| < \infty$.*

(b) *Suppose there is $C > 0$ such that $\|T_n\| \le C$ for $n = 1, 2, 3, \ldots$. For $T_0 \in L(X, Y)$, the vector subspace $M = \{x \in X : \lim_{n\to\infty} T_n x = T_0 x\}$ is closed in X. If M is dense or of the second category in X, then $M = X$ (i.e. T_n converges pointwise on X to T_0).*

Proof. (a) For all $x \in X$, $\{T_n(x)\}$ converges implies it is bounded. By the uniform boundedness principle, $\sup\{\|T_n\| : n = 1, 2, 3, \ldots\} < \infty$. Now there is a subsequence $\{\|T_{n_i}\|\}$ converging to $c = \liminf_{n\to\infty} \|T_n\|$. Then $\|Tx\| = \lim_{i\to\infty} \|T_{n_i} x\| \le \lim_{i\to\infty} \|T_{n_i}\|\|x\| = c\|x\|$, which implies $\|T\| \le c$.

(b) For every $x \in \overline{M}$ and $\varepsilon > 0$, there is $y \in M$ such that $\|x - y\| < \varepsilon/(2C + 2\|T_0\|)$. Since $y \in M$, so $T_n y$ converges to $T_0 y$. Hence, there is N such that $n \ge N$ implies $\|T_n y - T_0 y\| < \varepsilon/2$. Then

$$\|T_n x - T_0 x\| \le \|T_n x - T_n y\| + \|T_n y - T_0 y\| + \|T_0 y - T_0 x\| \le (\|T_n\| + \|T_0\|)\|x - y\| + \varepsilon/2 < \varepsilon.$$

So $\lim_{n\to\infty} T_n x = T_0 x$ and $x \in M$. Then $M = \overline{M}$.

If M is dense in X, then $M = \overline{M} = X$. If M is of the second category (hence not nowhere dense) in X, then M contains some $B(a,r)$ in X. So $M = \operatorname{span}(B(a,r) - a) = X$. $\square$

Remarks. If Y is also a Banach space, then we can replace (b) by

(b') *If there is $C > 0$ such that $\|T_n\| \leq C$ for $n = 1, 2, 3, \ldots$, then the vector subspace*

$$M = \{x \in X : \lim_{n\to\infty} T_n x \text{ exists}\} = \{x \in X : T_n x \text{ is Cauchy}\}$$

is closed in X. If M is dense or of second category in X, then $M = X$ (i.e. T_n converges pointwise on X).

For the proof of (b'), it suffices to show M is closed. For every $x \in \overline{M}$ and $\varepsilon > 0$, there is $y \in M$ such that $\|x - y\| < \varepsilon/(4C)$. Since $y \in M$, there is N such that $n, m \geq N$ implies $\|T_n y - T_m y\| < \varepsilon/2$. Then

$$\|T_n x - T_m x\| \leq \|T_n x - T_n y\| + \|T_n y - T_m y\| + \|T_m y - T_m x\| \leq (\|T_n\| + \|T_m\|)\|x - y\| + \varepsilon/2 < \varepsilon.$$

So $\lim_{n\to\infty} T_n x$ exists and $x \in M$. Then $M = \overline{M}$. The rest is the same.

Remarks. Using the uniform boundedness principle, it can be proved that there exists a 2π-periodic continuous function whose Fourier series does not converge to it everywhere. In fact, it can be used to show that there exists a 2π-periodic continuous function on $\mathbb{R}$ whose Fourier series diverges on an uncountable dense set in $\mathbb{R}$. See applications at the end of the next section.

§2. **Applications of Theorems.** Every function f defined on $(-\pi, \pi]$ corresponds to a 2π-periodic function on $\mathbb{R}$ defined by $f(x + 2n\pi) = f(x)$ for all integers n. Let $e^{i\theta} = \cos\theta + i\sin\theta$ and $\mathbb{T} = \{e^{i\theta} : -\pi < \theta \leq \pi\}$. Every function f defined on $(-\pi, \pi]$ also corresponds to a function f_o on $\mathbb{T}$ defined by $f_o(e^{i\theta}) = f(\theta)$. **In the following we will use these correspondences to identify these three sets of functions.**

Definitions. (1) A function $P : \mathbb{R} \to \mathbb{C}$ is a *trigonometric polynomial* iff it is of the form $P(x) = \sum_{k=-n}^{n} c_k e^{ikx}$, where $c_k \in \mathbb{C}$ and n is a nonnegative integer.

(2) For all $f \in L^1(-\pi, \pi]$ and $n \in \mathbb{Z}$, define the *n-th Fourier coefficient* of f to be $\widehat{f}(n) = \int_{(-\pi,\pi]} f(\theta) e^{-in\theta} \frac{dm}{2\pi}$.

The *Fourier series* of f is $\sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{ikx}$ and its *n-th partial sum* is $s_n(f; x) = \sum_{k=-n}^{n} \widehat{f}(k) e^{ikx}$.

Remarks. (1) Under the identification above, since the trigononmetric polynomials are 2π-periodic on $\mathbb{R}$, they can be considered as functions on $\mathbb{T}$. Below 2π-periodic continuous functions on $\mathbb{R}$ will be considered as functions in $C(\mathbb{T})$. Functions in $L^1(-\pi, \pi]$ can be considered as functions in $L^1(\mathbb{T})$.

(2) The set of all trigonometric polynomials is dense in $C(\mathbb{T})$ with sup-norm by the Stone-Weierstrass theorem since it is a self-adjoint subalgebra of $C(\mathbb{T})$ that separates points of $\mathbb{T}$ and vanishes at no point of $\mathbb{T}$.

(3) The *Dirichlet kernel* is $D_n(x) = \sum_{k=-n}^{n} e^{ikx}$, which is $\dfrac{\sin(n + \frac{1}{2})x}{\sin \frac{1}{2}x}$ if $x \neq 0$ and is $2n + 1$ if $x = 0$. We have

$$s_n(f; x) = \sum_{k=-n}^{n} \widehat{f}(k) e^{ikx} = \sum_{k=-n}^{n} \int_{(-\pi,\pi]} f(\theta) e^{ik(x-\theta)} \frac{dm}{2\pi} = \int_{(-\pi,\pi]} f(\theta) D_n(x - \theta) \frac{dm}{2\pi} = (f * D_n)(x).$$

Riemann-Lebesgue Lemma. *For every $f \in L^1(\mathbb{T})$, $\lim_{n\to\pm\infty} \widehat{f}(n) = 0$. In fact, the function $\mathcal{F} : L^1(\mathbb{T}) \to c_0$ defined by $\mathcal{F}(f) = (\widehat{f}(0), \widehat{f}(1), \widehat{f}(-1), \widehat{f}(2), \widehat{f}(-2), \ldots)$ is continuous and linear.*

Proof. For every $\varepsilon > 0$, from measure theory (see Rudin, *Real and Complex Analysis*, Theorem 3.14), there exists $g \in C(\mathbb{T})$ such that $\|f-g\|_1 < \varepsilon/2$. Next by remark (2) above, there is a trigonometric polynomial $P(x) = \sum_{k=-N}^{N} c_k e^{ikx}$ such that $\|g-P\|_\infty < \varepsilon/2$. For $|n| > N$, we have $\widehat{P}(n) = 0$ and

$$|\widehat{f}(n)| = \left| \int_{(-\pi,\pi]} (f(t) - P(t)) e^{-int} \frac{dm}{2\pi} \right| \leq \|f-P\|_1 \leq \|f-g\|_1 + \|g-P\|_1 \leq \|f-g\|_1 + \|g-P\|_\infty < \varepsilon.$$

So $\widehat{f}(n) \to 0$ as $|n| \to \infty$.

Next, linearity of $\mathcal{F}$ is clear and continuity follows from $\|\mathcal{F}(f)\| = \sup |\widehat{f}(n)| \leq \int_{(-\pi,\pi]} |f| \frac{dm}{2\pi} = \|f\|_1$. $\square$

Questions Is $\mathcal{F}$ injective? Is it surjective?

Theorem. $\mathcal{F} : L^1(\mathbb{T}) \to c_0$ is injective.

Proof. Suppose $f \in \ker \mathcal{F}$, i.e. $\widehat{f}(n) = 0$ for all $n \in \mathbb{Z}$. Then $\int_{(-\pi,\pi]} fP\, dm = 0$ for all trigononmetric polynomials P. There are two ways to finish.

(1) By remark (2) above, we have $\int_{(-\pi,\pi]} fg\, dm = 0$ for all $g \in C(\mathbb{T})$. For those who know the Riesz representation theorem on $C(\mathbb{T})^*$, it follows $f = 0$ almost everywhere.

(2) For every $x \in (-\pi, \pi]$, there are continuous $g_n : [-\pi, \pi] \to [0,1]$ such that $g_n(-\pi) = g_n(\pi) = 0$ and $\lim_{n\to\infty} g_n(t) = \chi_{(-\pi,x)}(t)$ for all $t \in (-\pi, \pi]$. By remark (2) above, there is a trigonometric polynomial P_n such that $\|g_n - P_n\|_\infty < \frac{1}{n}$. For $t \in (-\pi, \pi]$, $|f(t)P_n(t)| \leq |f(t)|\|P_n\|_\infty \leq |f(t)|(\|g_n\|_\infty + \frac{1}{n}) \leq 2|f(t)| \in L^1(-\pi, \pi]$ and $f(t)P_n(t) \to f(t)\chi_{(-\pi,x)}(t)$ for all $t \in (-\pi, \pi]$. By the Lebesgue dominated convergence theorem,

$$\int_{-\pi}^{x} f(t)dt = \int_{(-\pi,\pi]} f\chi_{(-\pi,x)} dm = \lim_{n\to\infty} \int_{(-\pi,\pi]} fP_n dm = 0.$$

Differentiate with respect to x, we get $f = 0$ almost everywhere (see Rudin, *Real and Complex Analysis*, Theorem 7.11). $\square$

Theorem. $\mathcal{F} : L^1(\mathbb{T}) \to c_0$ is not surjective. In fact, the range of $\mathcal{F}$ is not closed.

Proof. Assume $\mathcal{F}$ is surjective. There are two ways to get a contradiction.

(1) By the inverse mapping theorem, $\mathcal{F}$ would be an isomorphism between $L^1(\mathbb{T})$ and c_0. Then $c_0^* = \ell^1$ would be isomorphic to $(L^1(\mathbb{T}))^* = L^\infty(\mathbb{T})$, which is impossible because ℓ^1 is separable, but $L^\infty(\mathbb{T})$ (like ℓ^∞) is not separable as there are uncountably many balls $\{B(\chi_{(-\pi,x)}, \frac{1}{2}) : x \in (-\pi, \pi]\}$ that are pairwise disjoint in $L^\infty(\mathbb{T})$. Therefore, we have a contradiction.

(2) Since $\mathcal{F}$ is injective, if $\mathcal{F}(L^1(\mathbb{T}))$ is c_0 or closed, then by the lower bound theorem, $\mathcal{F}$ would be bounded below, i.e. there exists $c > 0$ such that $\|\mathcal{F}(f)\|_\infty \geq c\|f\|_1$ for all $f \in L^1(\mathbb{T})$. Now $D_n \in C(\mathbb{T}) \subseteq L^1(\mathbb{T})$ and $\|\mathcal{F}(D_n)\|_\infty = \|(1, 1, \ldots, 1, 0, 0, \ldots)\|_\infty = 1$. However, since $|\sin x| \leq |x|$ for all $x \in \mathbb{R}$, we have

$$\|D_n\|_1 > \frac{2}{\pi} \int_0^\pi \left|\sin\big(n + \frac{1}{2}\big)\theta\right| \frac{d\theta}{\theta} = \frac{2}{\pi} \int_0^{(n+1/2)\pi} |\sin\phi| \frac{d\phi}{\phi} > \frac{2}{\pi} \sum_{k=1}^{n} \frac{1}{k\pi} \int_{(k-1)\pi}^{k\pi} |\sin\phi| d\phi = \frac{4}{\pi^2} \sum_{k=1}^{n} \frac{1}{k} \to \infty,$$

which contradicts $\mathcal{F}$ is bounded below. $\square$

Questions Does the Fourier series of $f \in L^1(\mathbb{T})$ converge to f almost everywhere or in L^1-norm?

Theorem (du Bois-Reymond, 1873). *For every $w \in (-\pi, \pi]$, there exists $f \in C(\mathbb{T})$ such that its Fourier series diverges at $x = w$. More precisely, the partial sums of the Fourier series at $x = w$ is unbounded.*

Proof. (*Due to Henri Lebesgue*) First we deal with the case $w = 0$. Define $T_n : C(\mathbb{T}) \to \mathbb{C}$ by $T_n(f) = s_n(f;0) = \sum_{k=-n}^{n} \widehat{f}(k)$. Clearly, T_n is linear. Also, T_n is bounded since

$$|T_n f| = \Big|\int_{(-\pi,\pi]} f(\theta)D_n(-\theta)\frac{dm}{2\pi}\Big| \le \|f\|_\infty \int_{-\pi}^{\pi} |D_n(\theta)|d\theta = \|D_n\|_1\|f\|_\infty.$$

So $\|T_n\| \le \|D_n\|_1$.

In fact, $\|T_n\| = \|D_n\|_1$. To see this, let $g(t) = \operatorname{sgn} D_n(-t)$, which is defined by $g(t) = 1$ if $D_n(-t) \ge 0$ and $g(t) = -1$ if $D_n(-t) < 0$. Then $g(t)D_n(-t) = |D_n(-t)|$. Also, there exists $f_j \in C(\mathbb{T})$ such that $\|f_j\|_\infty = 1$ and $\lim_{j\to\infty} f_j(t) = g(t)$ for every $t \in (-\pi, \pi]$. Since $f_j(\theta)D_n(-\theta) \to g(\theta)D_n(-\theta) = |D_n(-\theta)|$ and $|f_j(\theta)D_n(-\theta)| \le |D_n(-\theta)| \in C(\mathbb{T}) \subset L^1(\mathbb{T})$, by the Lebesgue dominated convergence theorem,

$$\lim_{j\to\infty} T_n f_j = \lim_{j\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} f_j(\theta)D_n(-\theta)d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\theta)D_n(-\theta)d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_n(\theta)|d\theta = \|D_n\|_1.$$

Now $\sup\{\|T_n\| : n = 0,1,2,\ldots\} = \lim_{n\to\infty} \|D_n\|_1 = \infty$. By the uniform boundedness principle, there exists $f \in C(\mathbb{T})$ such that $\sup\{|T_n f| : n = 0,1,2,\ldots\} = \infty$. Therefore, the Fourier series of f diverges when $x = 0$. For $w \ne 0$, $f_w(x) = f(x-w) \in C(\mathbb{T})$ has Fourier coefficients $\widehat{f_w}(k) = \widehat{f}(k)e^{-ikw}$. Hence, its Fourier series is $\sum_{k=-\infty}^{\infty} (\widehat{f}(k)e^{-ikw})e^{ikx}$, which diverges at $x = w$. ◻

Appendix: Divergence of Fourier Series

Principle of Condensation of Singularities. *Let X be a Banach space and Y be a normed space. Let $T_{nj} \in L(X,Y)$ for $n, j = 0,1,2,\ldots$ be such that for all j, $\limsup_{n\to\infty} \|T_{nj}\| = \infty$. Then there is a set U of second category in X such that for all $f \in U$ and all j, $\limsup_{n\to\infty} \|T_{nj}f\| = \infty$.*

Proof. For a fixed j, let $V_j = \{f \in X : \limsup_{n\to\infty} \|T_{nj}f\| < \infty\}$. Then $f \in V_j$ implies $\sup\{\|T_{nj}f\| : n = 0,1,2,\ldots\} < \infty$. If V_j is of the second category in X, then the uniform boundedness principle would imply $\sup\{\|T_{nj}\| : n = 0,1,2,\ldots\} < \infty$, hence $\limsup_{n\to\infty} \|T_{nj}\| < \infty$, a contradiction. So V_j is of first category in X. Then $V = V_0 \cup V_1 \cup V_2 \cup \cdots$ is of first category in X. Since X is complete, $U = X \setminus V$ is of second category in X. For all $f \in U$ and all j, we have $f \notin V_j$, i.e. $\limsup_{n\to\infty} \|T_{nj}f\| = \infty$. ◻

Application Now take a countable dense subset $\{w_j\}$ of $\mathbb{T}$ and define $T_{nj} : C(\mathbb{T}) \to \mathbb{C}$ by $T_{nj}f = s_n(f; w_j)$. As in the proof of the last theorem, $\|T_{nj}\| = \|D_n\|_1$ and so $\limsup_{n\to\infty} \|T_{nj}\| = \infty$ for all j. By the principle of condensation of singularites, there is a set of second category in $C(\mathbb{T})$ such that all these functions f have Fourier series diverging at the dense subset $\{w_j\}$ (with (*) $\sup\{|s_n(f, w_j)| : n = 1,2,3,\ldots\} = \infty$ for all w_j.)

Let f be one such function. We claim that the set of points on $\mathbb{T}$ where the Fourier series of f diverges is actually a set of second category in $\mathbb{T}$, hence uncountable and much more than $\{w_j\}$!

To see this, let $M_{n,k} = \{w \in \mathbb{T} : |s_n(f;w)| \le k\}$, $M_k = \bigcap_{n=1}^{\infty} M_{n,k}$ and $M = \bigcup_{k=1}^{\infty} M_k$.

(1) $M_k = \{w \in \mathbb{T} : \sup\{|s_n(f,w)| : n = 1,2,3,\ldots\} \le k\}$, so by (*), for all j, k, $w_j \notin M_k$.

(2) If the Fourier series of f converges at w, then $\{s_n(f,w) : n = 1,2,3,\ldots\}$ is bounded, hence w is in some M_k, leading to $w \in M$. In particular, the Fourier series of f diverges at all elements of $\mathbb{T} \setminus M$.

(3) $h_{n,f}(w) = s_n(f;w) = \sum_{j=-n}^{n} \widehat{f}(j)e^{ijw}$ is continuous in w. So $M_{n,k} = h_{n,f}^{-1}(\overline{B(0,k)})$ and M_k are closed.

Assume some M_k is of second category in $\mathbb{T}$. Then in particular, it would not be nowhere dense. Since M_k is closed by (3), there is a nonempty open set in M_k. By the density of $\{w_j\}$, one of the w_j would be in M_k, contradicting (1). So all M_k must be of first category in $\mathbb{T}$. Then M will also be of first category in $\mathbb{T}$. By (2), the Fourier series of f diverges on $\mathbb{T} \setminus M$, which is of second category in $\mathbb{T}$, hence uncountable!

Remarks. In 1915, Lusin conjectured that for every $f \in L^2(-\pi,\pi]$, the Fourier series of f converges almost everywhere.

In 1926, Kolmogorov (as an undergraduate student in Moscow State University) proved that there exists a $f \in L^1(-\pi,\pi]$ such that the Fourier series of f diverges everywhere! See Antoni Zygmund, *Trigonometric Series*, second edition, vol. 1, pp. 310-314 for such a function.

In 1927, M. Riesz proved that for every function f in $L^p(-\pi,\pi]$ $(1 < p < \infty)$, the Fourier series of f converges in the L^p-norm to f. From measure theory (see Rudin, *Real and Complex Analysis*, Theorem 3.12), it is known that this implies there is a subsequence of the partial sums of the Fourier series of $f \in L^p(-\pi,\pi]$ converging almost everywhere to f.

In 1966, Lennart Carleson proved the Lusin conjecture. In particular, this implies the Fourier series of a 2π-periodic continuous function converges almost everywhere (to the function itself by Riesz' result). In the same year, Kahane and Katznelson proved that for every set of Lebesgue measure 0 on $(-\pi,\pi]$, there is a 2π-periodic continuous function whose Fourier series diverges there.

In 1968, Richard Hunt proved that for every $f \in L^p(-\pi,\pi]$ with $1 < p \le \infty$, the Fourier series of f converges almost everywhere to f itself.

§3. **Hahn-Banach Theorems.** In the literature, there are a few theorems that are commonly called *the Hahn-Banach theorem.* We will discuss these one at a time.

Definitions. (1) A *Minkowski functional* on a vector space X is a function $p : X \to \mathbb{R}$ such that for all $c \ge 0$ and $x, y \in X$, $p(cx) = cp(x)$ and $p(x+y) \le p(x)+p(y)$. (So semi-norms are Minkowski functionals such that for all $x \in X$ and $|c| = 1$, $p(x) \ge 0$ and $p(cx) = p(x)$.)

(2) A function $F : A \to B$ is an *extension* of another function $f : C \to B$ iff $A \supseteq C$ and $F(x) = f(x)$ for all $x \in C$, equivalently graph of F contains graph of f (in short $f = F|_C$). We say F is a *linear extension* of f when A, B, C are vector spaces and F, f are linear.

Real Hahn-Banach Theorem. *Let Y be a vector subspace of a vector space X over $\mathbb{R}$, p be a Minkowski functional on X and $f : Y \to \mathbb{R}$ be a linear function such that for all $x \in Y$, $f(x) \le p(x)$. Then f has a linear extension $F : X \to \mathbb{R}$ such that for all $x \in X$, $F(x) \le p(x)$.*

Proof. Consider the collection S of all (Z, f_Z), where Z is a vector subspace of X containing Y and there exists a linear extension f_Z of f and $f_Z(x) \le p(x)$ for all $x \in Z$. Since $(Y, f) \in S$, $S \ne \emptyset$. Partial order the elements of S by inclusion (i.e. $(Z_0, f_{Z_0}) \preceq (Z_1, f_{Z_1})$ iff $Z_0 \subseteq Z_1$ and $f_{Z_1}|_{Z_0} = f_{Z_0}$.) If C is a chain in S, then $L = \bigcup_{(Z,f_Z)\in C} Z$ is a vector subspace of X containing Y. Define f_L by taking $\Gamma(f_L) = \bigcup_{(Z,f_Z)\in C} \Gamma(f_Z)$. We see that C has (L, f_L) as an upper bound in S. Hence, by Zorn's lemma, S has a maximal element (M, f_M).

Assume $M \ne X$. Let $x \in X \setminus M$. Consider $Z = \operatorname{span}(M \cup \{x\}) = M + \mathbb{R}x$. For every $a, b \in M$,

$$f_M(a) + f_M(b) = f_M(a+b) \le p(a+b) \le p(a-x) + p(x+b).$$

Then $f_M(a) - p(a-x) \le p(x+b) - f_M(b)$. Taking supremum over $a \in M$, then infimum over $b \in M$, we get $\alpha = \sup\{f_M(a) - p(a-x) : a \in M\} \le \beta = \inf\{p(x+b) - f_M(b) : b \in M\}$. Let $c \in [\alpha, \beta]$ and define

$f_Z(m+rx) = f_M(m) + rc$ for all $m \in M, r \in \mathbb{R}$. It is easy to check f_Z is linear and f_Z extends f_M so that $f_Z(m) = f_M(m) \le p(m)$ for all $m \in M$. If $r > 0$, then taking $b = m/r$ and using $c \le \beta$, we have

$$f_Z(m+rx) = r\Big(f_M\big(\frac{m}{r}\big) + c\Big) \le r\Big(f_M\big(\frac{m}{r}\big) + p\big(x + \frac{m}{r}\big) - f_M\big(\frac{m}{r}\big)\Big) = p(m+rx).$$

If $r < 0$, then $-r > 0$. Taking $a = -m/r$ and using $c \ge \alpha$, we have

$$f_Z(m+rx) = -r\big(f_M\big(-\frac{m}{r}\big) - c\big) \le -r\Big(f_M\big(-\frac{m}{r}\big) - \big(f_M\big(-\frac{m}{r}\big) - p\big(-\frac{m}{r} - x\big)\big)\Big) = p(m+rx).$$

Then $(Z, f_Z) \in S$, which contradicts (M, f_M) maximal in S. So $M = X$. ◻

Complexification Lemma. *Let X be a vector space over $\mathbb{C}$. If $U : X \to \mathbb{R}$ is linear (considering X as a vector space over $\mathbb{R}$), then $F : X \to \mathbb{C}$ defined by $F(x) = U(x) - iU(ix)$ is linear (considering X as a vector space over $\mathbb{C}$).*

Proof. For $c \in \mathbb{R}, x, y \in X$, we have (1) $U(x+y) = U(x) + U(y)$, (2) $iU(i(x+y)) = iU(ix) + iU(iy)$, (3) $U(cx) = cU(x)$ and (4) $iU(icx) = ciU(ix)$. Subtracting (2) from (1), we get $F(x+y) = F(x) + F(y)$. Subtracting (4) from (3), we get $F(cx) = cF(x)$. Also, $F(ix) = U(ix) - iU(-x) = i(U(x) - iU(ix)) = iF(x)$. Therefore, F is linear (considering X as a vector space over $\mathbb{C}$). ◻

Complex Hahn-Banach Theorem. *Let Y be a vector subspace of a vector space X over $\mathbb{C}$, p be a seminorm on X and $f : Y \to \mathbb{C}$ be a linear function such that for all $x \in Y$, $|f(x)| \le p(x)$. Then f has a linear extension $F : X \to \mathbb{C}$ such that for all $x \in X$, $|F(x)| \le p(x)$.*

Proof. Let $u = \operatorname{Re} f$ and $v = \operatorname{Im} f$. Since $f(ix) = if(x)$, we have $u(ix) + iv(ix) = iu(x) - v(x)$ so that $\operatorname{Im} f(x) = v(x) = -u(ix)$. Since for all $x \in Y$, $u(x) \le |f(x)| \le p(x)$, by the real Hahn-Banach theorem, there exists a linear extension $U : X \to \mathbb{R}$ of u (with X as a vector space over $\mathbb{R}$) and $U(x) \le p(x)$ for all $x \in X$.

By the complexification lemma, $F : X \to \mathbb{C}$ defined by $F(x) = U(x) - iU(ix)$ is linear (considering X as a vector space over $\mathbb{C}$). F extends f because for every $x \in Y$,

$$F(x) = U(x) - iU(ix) = u(x) - iu(ix) = \operatorname{Re} f(x) + i \operatorname{Im} f(x) = f(x).$$

If $F(x) = 0$, then $|F(x)| = 0 \le p(x)$. If $F(x) \ne 0$, then let $c = |F(x)|/F(x)$. Since p is a seminorm and $F(cx) = cF(x) = |F(x)| \in \mathbb{R}$, $|F(x)| = F(cx) = (\operatorname{Re} F)(cx) = U(cx) \le p(cx) = |c|p(x) = p(x)$. ◻

Remark. The complexification lemma is useful in reducing problems to the case of vector spaces over $\mathbb{R}$.

Theorem (Hahn-Banach). *Let X be a normed space and Y be a vector subspace of X.*

(a) *For every $f \in Y^*$, there exists an extension $F \in X^*$ of f such that $\|F\| = \|f\|$.*

(b) *Let $x \in X$. We have $x \notin \overline{Y}$ if and only if there exists $F \in X^*$ such that $\|F\| = 1$, $F \equiv 0$ on Y and $F(x) = d(x, Y) = \inf\{\|x - y\| : y \in Y\} \ne 0$. In particular, $\overline{Y} = X$ if and only if $F \in X^*$ with $F \equiv 0$ on Y implies $F \equiv 0$ on X.*

(c) *If $X \ne \{0\}$, then for every $x \in X$, there exists $F \in X^*$ with $\|F\| = 1$ and $F(x) = \|x\|$. Such F is called a support functional at x. Note $x \ne y$ in $X \Rightarrow F(x-y) = \|x-y\| \ne 0$, $F(x) \ne F(y)$ for some $F \in X^*$.*

Proof. (a) For all $x \in X$, $p(x) = \|f\|\|x\|$ defines a seminorm. For case $\mathbb{K} = \mathbb{C}$, since $|f(x)| \le \|f\|\|x\| = p(x)$, by the complex Hahn-Banach theorem, we get a linear $F : X \to \mathbb{C}$ extending f such that $|F(x)| \le p(x) = \|f\|\|x\|$. For case $\mathbb{K} = \mathbb{R}$, similarly we get a linear $F : X \to \mathbb{R}$ extending f such that $F(x) \le p(x)$. Also $-F(x) = F(-x) \le p(-x) = p(x)$. Hence $|F(x)| \le p(x) = \|f\|\|x\|$. These two cases imply F is continuous and $\|F\| \le \|f\|$. Now for all $x \in Y$, $|f(x)| = |F(x)| \le \|F\|\|x\|$, which implies $\|f\| \le \|F\|$. So $\|F\| = \|f\|$.

(b) For the if-direction, by continuity, $F \equiv 0$ on $\overline{Y}$ and so $x \notin \overline{Y}$. For the only-if-direction, let $\delta = d(x, Y) > 0$. Define $f : \mathbb{K}x + Y \to \mathbb{K}$ by $f(cx + y) = c\delta$ for all $c \in \mathbb{K}, y \in Y$. Then $f \equiv 0$ on Y and $f(x) = \delta$. For $c \ne 0$,

$|f(cx+y)| = |c|\delta \leq |c|\|x+\frac{1}{c}y\| = \|cx+y\|$. Then $\|f\| \leq 1$. Taking a sequence $y_n \in Y$ such that $\|x-y_n\| \to \delta$, we may let $v_n = (x-y_n)/\|x-y_n\|$. Then $\|v_n\| = 1$ and $|f(v_n)| = \delta/\|x-y_n\| \to 1$. So $\|f\| = 1$. Applying (a), we get the required F.

(c) For $x \in X \setminus \{0\}$, let $Y = \{0\}$ and apply part (b) to get a $F \in X^*$, $\|F\| = 1$ and $F(x) = \|x\|$. Also using this F, we have $F(0) = 0$.

§4. **Locally Convex Spaces.** We extend the Hahn-Banach theorems to some topological vector spaces.

Definition. X is a *locally convex* space iff X is a topological vector space such that every neighborhood of 0 contains a convex neighborhood of 0.

Remark. In a locally convex space, it is even true that every neighborhood of 0 contains an *open* convex neighborhood of 0. This follows from the fact that if A is convex, then A° is also convex because for $0 < t < 1$, $tA^\circ + (1-t)A^\circ = \bigcup_{a \in A^\circ} (ta + (1-t)A^\circ)$ is open in A, hence $tA^\circ + (1-t)A^\circ \subseteq A^\circ$.

Theorem. *In a vector space X, for every convex absorbing set U, let $p_U(x) = \inf\{t > 0 : x \in tU\}$. Then*

(a) *$p_U(x)$ is a Minkowski functional. (It is called the Minkowski functional of U.)*

(b) $\{x : p_U(x) < 1\} \subseteq U \subseteq \{x : p_U(x) \leq 1\}$.

(c) *If X is a topological vector space and U is also open, then $U = \{x : p_U(x) < 1\}$.*

Proof. (a) Since U is absorbing, $0 \in U$ and so $p_U(0) = 0$. For $c > 0$, since $x \in tU$ iff $cx \in ctU$, so $p_U(cx) = cp_U(x)$. Next observe that for $s, t > 0$, if $x \in sU$ and $y \in tU$, then since U is convex, $x + y \in sU + tU = (s+t)\big(\frac{s}{s+t}U + \frac{t}{s+t}U\big) = (s+t)U$. Taking infima of such s and t, we get $p_U(x+y) \leq p_U(x) + p_U(y)$.

(b) If $p_U(x) < 1$, then there is $t \in [p_U(x), 1)$ such that $x \in tU$. If $t = 0$, then $x = 0 \in U$. If $t > 0$, $(1/t)x \in U$ and $x = t(1/t)x + (1-t)0 \in U$ by convexity. Next, if $x \in U$, then $1 \in \{t > 0 : x \in tU\}$ and so $p_U(x) \leq 1$.

(c) Let $x \in U$. Since the scalar multiplication map g is continuous, U is open and $g(1,x) = x \in U$, there is a neighborhood $B(1,r) \times V$ of $(1,x)$ such that $B(1,r) \times V \subseteq g^{-1}(U)$. Let $t = 1 + (r/2)$. Then $tx = g(t,x) \in U$. So $x \in (1/t)U$, which implies $p_U(x) \leq 1/t < 1$. Combining with (b), we get $U = \{x : p_U(x) < 1\}$. ∎

Remarks. For convex absorbing U, if U is balanced, then $p_U(x)$ is a semi-norm. (The reason is as follow. Clearly $p_U(x) \geq 0$. For $c \in \mathbb{K} \setminus \{0\}$, let $c = |c|a$. Since $p_U(cx) = |c|p_U(ax)$, it suffices to show $p_U(ax) = p_U(x)$. Since $|a| = 1$ and U is balanced, we have $aU \subseteq U$ and $(1/a)U \subseteq U$. They imply $(1/a)U = U$. So $ax \in tU$ iff $x \in (t/a)U = tU$. So $p_U(ax) = p_U(x)$.)

The converse is false. For example, $U = B(0,1) \cup \{1\}$ is not balanced in $\mathbb{C}$. Yet, for $x \in (0, +\infty)$, $x \in tU, t > 0$ iff $t \in [x, +\infty)$. For $z \in \mathbb{C} \setminus [0, +\infty)$, we have $z \in tU, t > 0$ iff $t \in (|z|, +\infty)$. So $p_U(z) = |z|$.

Lemma. *Let X be a topological vector space over $\mathbb{R}$ and A be a nonempty open convex subset of X. If $f \in X^*$ and $f \not\equiv 0$,, then $f(A)$ is an open interval.*

Proof. Now A convex implies it is path connected. Since f is continuous, $f(A)$ is path connected in $\mathbb{R}$. Hence $f(A)$ is an interval. For every $a \in A$, $U = -a + A$ is an open neighborhood of 0. Since $f \not\equiv 0$, there is $x_0 \in X$ such that $f(x_0) = 1$. Let g be the scalar multiplication map $g(t,x) = tx$. Since $g(0, x_0) = 0 \in U$, $g^{-1}(U)$ contains a neighborhood $(-\varepsilon, \varepsilon) \times N_{x_0}$ of $(0, x_0)$. This implies $tx_0 \in U$ for $t \in (-\varepsilon, \varepsilon)$. Now $(f(a) - \varepsilon, f(a) + \varepsilon) = \{f(a) + t = f(a + tx_0) : t \in (-\varepsilon, \varepsilon)\} \subseteq f(a + U) = f(A)$. So $f(A)$ is open. ∎

Separation Theorem. *Let A, B be disjoint, nonempty convex subsets of a topological vector space X.*

(a) *If A is open, then there is $f \in X^*$ such that for all $x \in A$, $\operatorname{Re} f(x) < \inf \operatorname{Re} f(B)$.*

(b) *(Strong Separation Theorem) If A is compact, B is closed and X is locally convex, then there is $f \in X^*$ such that $\max \operatorname{Re} f(A) < \inf \operatorname{Re} f(B)$. This was proved by V. L. Klee in 1951.*

Proof. It suffices to prove the case $\mathbb{K} = \mathbb{R}$. (Then for the case $\mathbb{K} = \mathbb{C}$, we may regard X as a vector space over $\mathbb{R}$ and keep the *same topology* so that it is a topological vector space over $\mathbb{R}$. Then apply the case $\mathbb{K} = \mathbb{R}$ and use the complexification lemma to get the desired complex linear functional. This complex linear functional is continuous because its real and imaginary parts are continuous.)

(a) Fix $a_0 \in A$ and $b_0 \in B$. Let $x_0 = b_0 - a_0$, then $C = \underbrace{A - B + x_0}_{convex} = \bigcup_{b\in B} \underbrace{(A - b + x_0)}_{open}$ is an open convex neighborhood of $a_0 - b_0 + x_0 = 0$. Let $p(x)$ be the Minkowski funcional of C, then by (c) of the above theorem, $C = \{x : p(x) < 1\}$. Next $A \cap B = \emptyset$ implies $x_0 \notin C$. So $p(x_0) \geq 1$.

We claim there exists $f \in X^*$ such that $f(x_0) = 1$ and for all $x \in C$, $f(x) < 1$. To see this, let M be the linear span of $\{x_0\}$. Define $f : M \to \mathbb{R}$ by $f(tx_0) = t$. Then $f(x_0) = 1 \leq p(x_0)$ implies $f(x) \leq p(x)$ on M. So f can be extended linearly to X with $f(x) \leq p(x)$ on X. Since $f(x) \leq p(x) < 1$ for all $x \in C$, so $f(-x) = -f(x) > -1$ for all $-x \in -C$. Then $|f| < 1$ on $U = C \cap (-C)$, a neighborhood of 0. Thus, for all $\varepsilon > 0$, εU is a neighborhood of 0 and $x \in \varepsilon U$ implies $|f(x)| < \varepsilon$. So f is continuous at 0, hence continuous on X.

For all $a \in A$ and $b \in B$, $a - b + x_0 \in C$ implies $f(a) - f(b) + 1 = f(a - b + x_0) < 1$. So $f(a) < f(b)$. Taking supremum over $a \in A$, then infimum over $b \in B$, we get $\sup f(A) \leq \inf f(B)$. Since A is nonempty open convex, $f \in X^*$ and $f(x_0) = 1$, by the lemma, $f(A) = (\alpha, \beta)$ say. Then for all $x \in A$, we have $f(x) < \beta = \sup f(A) \leq \inf f(B)$.

(b) Since $A \cap B = \emptyset$, B is closed and X is locally convex, $X \setminus B$ is a neighborhood of every $a \in A$. So there is an open convex neighborhood V_a of 0 such that $a + V_a \subseteq X \setminus B$. Now $\{a + \frac{1}{2}V_a : a \in A\}$ covers A. From a subcover $\{a_i + \frac{1}{2}V_{a_i} : i = 1, 2, \ldots, n\}$, we intersect the $\frac{1}{2}V_{a_i}$'s to get an open convex neighborhood V of 0. Note

$$A + V \subseteq \bigcup_{i=1}^{n}(a_i + \frac{1}{2}V_{a_i} + V) \subseteq \bigcup_{i=1}^{n}(a_i + \frac{1}{2}V_{a_i} + \frac{1}{2}V_{a_i}) \subseteq \bigcup_{i=1}^{n}(a_i + V_{a_i}) \subseteq X \setminus B.$$

Then $A + V = \bigcup_{a\in A}(a + V) \subseteq X \setminus B$ is an open convex set disjoint from B. By (a), there is a continuous linear functional $f : X \to \mathbb{R}$ such that $\sup f(A + V) \leq \inf\{f(y) : y \in B\}$. Since $f(A)$ is compact in $f(A + V) = (\alpha, \beta)$ by lemma, we have $\max f(A) < \beta = \sup f(A + V) \leq \inf f(B)$. □

Corollary (Consequences of Separation Theorem). *Let X be a locally convex space.*

(a) *If X is Hausdorff, then X^* separates points of X in the sense that if $x \neq y$ in X, then there exists $f \in X^*$ such that $f(x) \neq f(y)$. In particular, if $f(x) = 0$ for all $f \in X^*$, then $x = 0$.*

(b) *Let Y be a vector subspace of X and $x \in X$. We have $x \notin \overline{Y}$ if and only if there exists $f \in X^*$ such that $f(x) \neq 0$ and $f \equiv 0$ on Y. Also, $\overline{Y} = X$ if and only if $f \in X^*$ with $f \equiv 0$ on Y implies $f \equiv 0$ on X.*

Proof. (a) For distinct $x, y \in X$, let $A = \{x\}$ and $B = \{y\}$ and apply (b) of the separation theorem.

(b) For the if direction, by continuity, $f \equiv 0$ on $\overline{Y}$ and so $x \notin \overline{Y}$. For the only-if direction, let $A = \{x\}$ and $B = \overline{Y}$ and apply (b) of the separation theorem to get $f \in X^*$ to separate A and B. Since $f(Y)$ is a vector subspace of $\mathbb{K}$, we must have $f(Y) = \{0\}$ and $f(x) \neq 0$.

Using the separation theorem, we can obtain an important theorem of M. Krein and D. Milman.

Definitions. Let V be a vector space over $\mathbb{K}$ and $V \supseteq S \supseteq M \neq \emptyset$.

(a) M is an *extreme set* in S iff M has the property that "if there exist $s_1, s_2 \in S$ and there exists $t \in (0, 1)$ such that $ts_1 + (1 - t)s_2$ is in M, then both s_1 and s_2 are in M." An extremal set consisted of a single point is called an *extreme point*.

(b) The *__convex hull__* of S is the smallest (or intersection of every) convex set in V containing S. (It is easy to see that the convex hull of S is $\Big\{\sum_{i=1}^{n} t_i s_i : n = 1,2,3,\ldots,\ s_i \in S, t_i \in [0,1], \sum_{i=1}^{n} t_i = 1\Big\}$.) For S in a topological vector space, the *__closed convex hull__* of S is the closure of the convex hull of S.

__Examples.__ Every side of a triangular region on a plane is an extreme set of the region and every vertex is an extreme point. Every point of a circle is an extreme point of the closed disk having the circle as boundary.

__Remarks.__ (1) If for every $\alpha \in A$, E_α is an extreme set in S and $E = \bigcap_{\alpha\in A} E_\alpha \neq \emptyset$, then E is an extreme set in S. This is because $s_1, s_2 \in S, t \in (0,1)$ and $ts_1 + (1-t)s_2 \in E$ imply $ts_1 + (1-t)s_2 \in E_\alpha$ for every $\alpha \in A$, which implies $s_1, s_2 \in E_\alpha$ for every α, hence $s_1, s_2 \in E$.

(2)If P is an extreme set in M and M is an extreme set in S, then P is an extreme set in S. This is because $ts_1 + (1-t)s_2 \in P$ for some $s_1, s_2 \in S, 0 < t < 1$ implies $ts_1 + (1-t)s_2 \in M$ so that $s_1, s_2 \in M$ (by the extremity of M in S), then $s_1, s_2 \in P$ (by the extremity of P in M).

__Theorem (Krein-Milman).__ *Let X be a Hausdorff locally convex space and $\emptyset \neq S \subseteq X$. If S is compact and convex, then S has at least one extreme point and S is the closed convex hull of its extreme points.*

__Proof.__ We first show S has an extreme point. Note S is an extreme subset of itself. Let $\mathcal{C} = \{W : W \text{ is a nonempty compact extreme subsets of } S\}$. Order $\mathcal{C}$ by reverse inclusion, i.e. for $E_1, E_2 \in \mathcal{C}$, define $E_1 \preceq E_2$ iff $E_1 \supseteq E_2$. Since X is Hausdorff, every $W \in \mathcal{C}$ is closed. For every nonempty chain $\mathcal{E}$ in $\mathcal{C}$, let $L = \bigcap_{W\in\mathcal{E}} W$, then L is closed and compact. Assume $L = \emptyset$. Then $\bigcup_{W\in\mathcal{E}} (S \setminus W) = S$. Since S is compact and $S \setminus W$ is open in S, there are $W_1, W_2, \ldots, W_n \in \mathcal{E}$ such that $\emptyset \neq W_1 \subseteq W_2 \subseteq \cdots \subseteq W_n \subseteq S$ and $S = \bigcup_{i=1}^{n} (S \setminus W_i) = S \setminus W_1$. Then $W_1 = \emptyset$, contradiction. Hence $L \neq \emptyset$. By remark (1), L is an extreme subset of S. So L is an upper bound of the chain $\mathcal{E}$ in $\mathcal{C}$. By Zorn's lemma, $\mathcal{C}$ has a maximal element E.

Assume E has distinct elements x, y. By (b) of the separation theorem, there exists $f \in X^*$ such that $\operatorname{Re} f(x) < \operatorname{Re} f(y)$. This implies $y \notin E_0 = \{s \in E : \operatorname{Re} f(s) = \inf \operatorname{Re} f(E)\} \subset E$. Now E_0 is __nonempty__ due to continuity of $\operatorname{Re} f$ on the compact set E. Since $E_0 = (\operatorname{Re} f)^{-1}(\{\inf \operatorname{Re} f(E)\})$, it is closed (hence __compact__). Finally, E_0 is an __extreme__ subset of S because $s = ts_1 + (1-t)s_2 \in E_0 \subset E$ implies $s_1, s_2 \in E$ (as $E \in \mathcal{C}$ is extreme) and

$$\inf \operatorname{Re} f(E) \le \min\{\operatorname{Re} f(s_1), \operatorname{Re} f(s_2)\} \le t \operatorname{Re} f(s_1) + (1-t) \operatorname{Re} f(s_2) = \operatorname{Re} f(s) = \inf \operatorname{Re} f(E)$$

implies $\operatorname{Re} f(s_1) = \inf \operatorname{Re} f(E) = \operatorname{Re} f(s_2)$, i.e. $s_1, s_2 \in E_0$. So $E_0 \in \mathcal{C}$. Since $E_0 \succ E$, this contradicts the maximality of E in $\mathcal{C}$. Therefore, E can only contain an extreme point of S.

Now we show S equals the closed convex hull H of all of its extreme points. Since S is closed and convex, $H \subseteq S$. Assume there is $s \in S \setminus H$. By (b) of the separation theorem, there is $f \in X^*$ such that $\operatorname{Re} f(s) < \inf \operatorname{Re} f(H)$. Then $H_1 = \{x \in S : \operatorname{Re} f(x) = \inf \operatorname{Re} f(S)\}$ is convex and disjoint from H. Similar to E_0 above, H_1 is a nonempty closed (hence compact) extreme subset of S. By the first part, H_1 has at least one extreme point p. By remark (2), p is an extreme point of S, which contradicts $H_1 \cap H = \emptyset$. So $S = H$.$\square$

__Remarks.__ (1) Compactness is needed in the Krein-Milman theorem as the set $\{(x,y) \in \mathbb{R}^2 : x \geq 0\}$ is closed and convex in $\mathbb{R}^2$, but it has no extreme point.

(2) The Krein-Milman theorem can be used to prove the Stone-Weierstrass theorem. Combining with the Banach-Alaoglu theorem in the next chapter, it can be used to show that there exist Banach spaces that are not the dual spaces of Banach spaces. For details of these two applications, see [Be], p. 110.

Chapter 3. Weak Topologies and Reflexivity.

§1. **Canonical Embedding.** For a normed space X over $\mathbb{K}$, $x \in X$ and $y \in X^*$, let $\langle x, y \rangle = y(x)$. This notation is to illustrate that many similar properties exist between X and X^*. For example, $\langle x, y \rangle$ is linear in x and y. For $y \in X^*$, $\|y\| = \sup\{|y(x)| : x \in X, \|x\| \le 1\} = \sup\{|\langle x, y\rangle| : x \in X, \|x\| \le 1\}$. In remark (1) below, we will show that $\|x\| = \sup\{|y(x)| : y \in X^*, \|y\| \le 1\} = \sup\{|\langle x, y\rangle| : y \in X^*, \|y\| \le 1\}$.

Theorem. *Let X, Y be normed spaces. If Y is complete, then $L(X, Y)$ is a Banach space. (In particular, $X^* = L(X, \mathbb{K})$ is a Banach space.)*

Proof. Clearly $L(X, Y)$ is a normed vector space. For completeness, suppose $\{T_n\}$ is a Cauchy sequence in $L(X, Y)$. Then $\{T_n\}$ is bounded so that there is $K \ge 0$ such that for all $n \ge 1$, $\|T_n\| \le K$. Then for all $x \in X, n \ge 1$, we have $\|T_n(x)\| \le K\|x\|$. Since $\|T_n(x) - T_m(x)\| \le \|T_n - T_m\|\|x\|$, the sequence $\{T_n(x)\}$ is a Cauchy sequence in Y. Since Y is complete, $\lim_{n\to\infty} T_n(x)$ exists and we may define $T(x) = \lim_{n\to\infty} T_n(x)$. Clearly, T is linear. Also, T is bounded as $\|T(x)\| = \lim_{n\to\infty} \|T_n(x)\| \le K\|x\|$. So $T \in L(X, Y)$. For every $\varepsilon > 0$, since $\{T_n\}$ is Cauchy, there is N such that $m, n \ge N$ implies $\|T_n - T_m\| < \varepsilon$. Then $\|T_n(x) - T_m(x)\| \le \varepsilon\|x\|$ for all $x \in X$. So $\|T_n(x) - T(x)\| = \lim_{m\to\infty} \|T_n(x) - T_m(x)\| \le \varepsilon\|x\|$. Hence, if $n \ge N$, then $\|T_n - T\| \le \varepsilon$. Therefore, $\{T_n\}$ converges to T in $L(X, Y)$. ∎

Exercise. For $X \ne \{0\}$, if $L(X, Y)$ is a Banach space, then prove that Y is complete.

Canonical Embedding Theorem. *For a normed space X, the "canonical embedding" $i : X \to X^{**} = (X^*)^*$ defined by $i(x) = i_x$, where $i_x(y) = y(x)$, is a linear isometry. If $X \ne \{0\}$, then for all $x \in X$, $\|x\| = \sup\{|y(x)| : y \in X^*, \|y\| = 1\}$. In fact,* sup *can be replaced by* max.

Proof. It is easy to see that i_x is linear from X^* to $\mathbb{K}$ and i is linear from X to X^{**}. To show i is an isometry, it is enough to deal with the case $X \ne \{0\}$. Note $|i_x(y)| = |y(x)| \le \|y\|\|x\|$ for all $y \in X^*$ so that $\|i_x\| \le \|x\|$. By part (c) of the Hahn-Banach theorem, for every $x \in X$, there is $y \in X^*$ such that $\|y\| = 1$ and $y(x) = \|x\|$. Then $\|x\| = y(x) = i_x(y) \le \|i_x\|\|y\| = \|i_x\|$. Therefore, $\|x\| = \|i_x\|$. ∎

Remarks. (1) In the case $X = \{0\}$, we have $X^* = \{0\}$. So to cover all normed spaces, the second statement should be changed to $\|x\| = \sup\{|y(x)| : y \in X^*, \|y\| \le 1\} = \sup\{|\langle x, y\rangle| : y \in X^*, \|y\| \le 1\}$.

(2) To simplify notations, we will often identify $x \in X$ with $i_x \in X^{**}$ and X with $i(X)$ below.

Definitions. The closure $\widehat{X}$ of X in X^{**} is a Banach space containing X as a dense subset and it is called a *completion* of X. Banach spaces X satisfying $i(X) = X^{**}$ are called *reflexive*. (For example, Hilbert spaces, $L^p([0,1])$ and ℓ^p with $1 < p < \infty$ are reflexive.)

§2. **Locally Convex Spaces Generated by Seminorms.** Occasionally, we will come across vector spaces X that have many important semi-norms like those of the form $|T(x)|$, where $T : X \to \mathbb{K}$ is linear. Then we may want vector topologies on the vector spaces so that all these semi-norms are continuous. Below is a theorem for that purpose. First, let p be a semi-norm and let $V(p) = \{x : p(x) < 1\}$. Observe that if $r > 0$, then $rV(p) = \{x : p(x) < r\}$ because $x \in rV(p)$ iff $x/r \in V(p)$ iff $p(x) = rp(x/r) < r$.

Theorem. (a) *Let $p_1(x), p_2(x), \ldots, p_n(x)$ be semi-norms on a vector space X and $r_1, r_2, \ldots, r_n > 0$. Then $S = r_1V(p_1) \cap \cdots \cap r_nV(p_n)$ is convex, balanced (i.e. absolutely convex) and absorbing.*

(b) *On a topological vector space X, a semi-norm $p(x)$ is continuous iff $p(x)$ is continuous at 0 iff for all $r > 0$, $rV(p)$ is open.*

(c) *Let $\mathcal{P}$ be a family of semi-norms on a vector space X. The collection*

$$\mathcal{U} = \{r_1V(p_1) \cap \cdots \cap r_nV(p_n) : n \in \mathbb{N}, r_1, \cdots, r_n > 0,\ p_1, \cdots, p_n \in \mathcal{P}\}$$

is a base at 0 of a topology that makes X into a locally convex space. Furthermore, it is the weakest vector topology on X for which all semi-norms in $\mathcal{P}$ are continuous.

Proof. (a) For $i = 1, 2, \ldots, n$, let $S_i = r_i V(p_i)$. If $x, y \in S$ and $t \in [0,1]$, then $x, y \in S_i$ and $p_i(tx+(1-t)y) \le tp_i(x) + (1-t)p_i(y) < r_i$, i.e. $tx + (1-t)y \in S_i$ for all i. Hence, $tx + (1-t)y \in S$, i.e. S is convex. Next, if $|c| \le 1$, then $p_i(cx) = |c|p_i(x) < r$, i.e. $cx \in S_i$ for all i. Hence $cx \in S$, i.e. S is balanced. Finally, if $z \in X$, then $p_i(z) = 0$ implies $cz \in S_i$ for $0 < |c| \le r_i = 1$ and $p_i(z) > 0$ implies $cz \in S_i$ for $0 < |c| \le r_i = 1/p_i(z)$. So for $r = \min\{r_1, r_2, \ldots, r_n\}$, we have $cz \in S$ for $0 < |c| \le r$. Hence S is absorbing.

(b) This follows from $|p(x) - p(x_0)| \le p(x - x_0)$ and $p^{-1}(-\varepsilon, \varepsilon) = \{x \in X : p(x) < \varepsilon\} = \varepsilon V(p)$.

(c) By (b), all $p \in \mathcal{P}$ are continuous iff all elements of $\mathcal{U}$ are open. The weakest vector topology on X for which all semi-norms in $\mathcal{P}$ are continuous is the one generated by $\Omega = \{x + U : x \in X, U \in \mathcal{U}\}$. Let $\mathcal{T}$ be consisted of all subsets S of X satisfying the condition that for every $x \in S$, there exists $U \in \mathcal{U}$ such that $x + U \subseteq S$. We can check $\mathcal{T}$ is a topology on X. Let $U = r_1 V(p_1) \cap \cdots \cap r_n V(p_n)$ be an arbitrary element in $\mathcal{U}$. Every $x + U$ is in $\mathcal{T}$ because $a \in x + U$ implies $a + U_a \in x + U$, where $U_a = c_1 V(p_1) \cap \cdots \cap c_n V(p_n) \in \mathcal{U}$ and $c_i = r_i - p_i(a - x)$. The definition of $\mathcal{T}$ makes Ω a base for $\mathcal{T}$ and $\mathcal{U}$ a base at 0.

Next we check the addition map f and scalar multiplication map g are continuous. Let $a, b \in X$. To see f is continuous at (a, b), for $U \in \mathcal{U}$, let $V = \frac{1}{2}U$ and observe that $f((a + V) \times (b + V)) = a + b + U$. So $f^{-1}(a + b + U)$ contains $(a + V) \times (b + V)$, which is a neighborhood of (a, b). Hence f is continuous.

Suppose $c \in \mathbb{K}$ and $x \in X$. Since V is absorbing, there is $s > 0$ such that $x \in sV$. Let $t = s/(1+|c|s) > 0$. For every (c', x') in the neighborhood $B(c, 1/s) \times (x + tV)$ of (c, x), we have $|c' - c| < 1/s$, $|c'|t \le (|c| + \frac{1}{s})t = 1$ and $c'x' - cx = c'(x' - x) + (c' - c)x \in |c'|tV + |c' - c|sV \subseteq V + V = U$. This implies $g^{-1}(cx + U)$ contains the neighborhood $B(c, 1/s) \times (x + tV)$ of (c, x). So g is continuous. Therefore, $\mathcal{T}$ is the desired topology. $\square$

Remarks. (1) The topology given in (c) is called the *topology generated by the family $\mathcal{P}$ of semi-norms*.

(2) The converse of (c) is true, i.e. a topological vector space X is a locally convex space iff there exists a family of semi-norms that generates the topology on X (see [TL], p. 113).

(3) In the case $\mathcal{P}$ is consisted of exactly one norm, then we get the usual normed topology. So all theorems on locally convex spaces apply to normed spaces!

Theorem. *Let X be a locally convex space whose topology is generated by a family $\mathcal{P}$ of semi-norms. X is Hausdorff iff $\mathcal{P}$ is separating (i.e. for each nonzero $x \in X$, there is $p \in \mathcal{P}$ such that $p(x) \neq 0$).*

Proof. For $\mathcal{P}$ separating, let $a, b \in X$ with $x = b - a \neq 0$. So there is $p \in \mathcal{P}$ such that $p(x) > 0$. Then $A = (-\infty, p(x)/2)$ and $B = (p(x)/2, +\infty)$ are disjoint open in $\mathbb{R}$. So $p^{-1}(A)$ and $p^{-1}(B)$ are disjoint neighborhoods of 0 and x respectively. So $a + p^{-1}(A)$ and $a + p^{-1}(B)$ are disjoint neighborhoods of a and b respectively.

For $\mathcal{P}$ not separating, there is $x \neq 0$ such that for all $p \in \mathcal{P}$, $p(x) = 0$. Then for all $r > 0$ and $p \in \mathcal{P}$, $x \in rV(p)$. Hence every neighborhood of 0 contains x. So X is not Hausdorff. $\square$

Definition. A set S in a topological vector space X is *bounded* iff for every neighborhood N of 0, there is $r > 0$ such that $S \subseteq rN$.

Theorem. *Let X be a locally convex space whose topology is generated by a family $\mathcal{P}$ of seminorms.*

(a) *A set W is bounded in X iff for every $p \in \mathcal{P}$, $p(W)$ is bounded in $\mathbb{K}$.*

(b) *A net $\{x_\alpha\}_{\alpha \in I} \to x$ in X iff for every $p \in \mathcal{P}$, $\{p(x_\alpha - x)\}_{\alpha \in I} \to 0$. (Then $|p(x_\alpha) - p(x)| \le p(x_\alpha - x) \to 0$.)*

Proof. (a)

$$\begin{aligned} W \text{ is bounded } &\iff \forall\, p_1, \ldots, p_n \in \mathcal{P},\ r_1, \ldots, r_n > 0,\ \exists\, r > 0 \text{ such that } W \subseteq r \bigcap_{i=1}^{n} \underbrace{\{x : p_i(x) < r_i\}}_{= r_i V(p_i)} \\ &\iff \forall\, p_i \in \mathcal{P},\ \exists\, R_i > 0 \text{ such that } \forall x \in W,\ p_i(x) < R_i \\ &\iff \forall\, p \in \mathcal{P},\ p(W) \text{ is bounded in } \mathbb{K}, \end{aligned}$$

where in the second step, take $n = 1$, $R_1 = rr_1$ in the $\Rightarrow$ direction and take $r > R_i/r_i$ for $i = 1, \ldots, n$ in the $\Leftarrow$ direction.

(b) $$\begin{aligned} \{x_\alpha\}_{\alpha\in I} \to x &\iff \{x_\alpha - x\}_{\alpha\in I} \to 0 \\ &\iff \forall\, p_1, \ldots, p_n \in \mathcal{P},\ r_1, \ldots, r_n > 0,\ \ \exists\, \beta \in I \text{ such that} \\ &\qquad\qquad \alpha \succeq \beta \text{ implies } x_\alpha - x \in \bigcap_{i=1}^{n} \{y : p_i(y) < r_i\} \\ &\iff \forall\, p_i \in \mathcal{P},\ r_i > 0,\ \ \exists\, \beta_i \in I \text{ such that } \alpha \succeq \beta_i \text{ implies } x_\alpha - x \in \{y : p_i(y) < r_i\} \\ &\iff \forall\, p \in \mathcal{P},\ \{p(x_\alpha - x)\}_{\alpha\in I} \to 0, \end{aligned}$$

where in the third step, take $n = 1$ in the $\Rightarrow$ direction and take $\beta \succeq \beta_i$ for $i = 1, \ldots, n$ in the $\Leftarrow$ direction.□

§3. **Weak and Weak-star Topologies.** We now ask the

Questions: Why are we interested in locally convex spaces? Why are normed spaces not good enough?

(1) Some important classes of functions in analysis, such as the collection of distributions or generalized functions is not a normed space. They can be topologized by semi-norms.

(2) In analysis, we solve many problems by taking limit. Very often we consider bounded sequences and try to extract convergent subsequences or subnets to get a limit point. For an infinite dimensional normed space X, an application of the Riesz lemma showed the closed unit ball is not compact. So bounded sequences on normed spaces may not have convergent subsequences or subnets in the norm topology!

However, Banach and Alaoglu proved that the closed unit ball of X^* is compact in another topology $\mathcal{T}$ generated by some semi-norms. So bounded sequences on dual spaces have $\mathcal{T}$-cluster points. This is very useful for solving many analysis problems.

For a normed space X, there is a weakest vector topology w on X that makes all elements of X^* continuous. We simply take $\mathcal{P} = \{|f| : f \in X^*\}$ and apply the theorems on locally convex spaces. This topology w on X is called the *weak topology* on X. Then X with this topology is a locally convex space. Using the description of a base of 0 in a locally convex space, we see sets of the form $U = \bigcap_{i=1}^{n} \{x \in X : |f_i(x)| < r_i\}$, where $r_i > 0$ and $f_i \in X^*$, form a base at 0 for the weak topology.

So on a normed space X, there are two topologies, namely the original norm-topology and the w-topology. When we mean X with the w-topology, we shall write (X, w).

Properties of Weak Topologies.

(1) By definition of weak topology, we have the w-topology is a subset of the norm-topology. So w-open sets are open in X, w-closed sets are closed in X, but compact sets in X are w-compact.

(2) By part (c) of the Hahn-Banach theorem, $\mathcal{P} = \{|f| : f \in X^*\}$ is separating, which implies the weak topology is Hausdorff. So w-compact sets are w-closed. In case $\dim X < \infty$, by the finite dimension theorem, the norm and weak topologies are equal.

(3) For every net $\{x_\alpha\}_{\alpha\in I}$ in X, by a theorem in the section on locally convex spaces, we have $\{x_\alpha\}_{\alpha\in I}$ w-converges to x in X (write as $x_\alpha \xrightarrow{w} x$) iff for every $f \in X^*$, $|f(x_\alpha - x)| \to 0$, i.e. $f(x_\alpha) \to f(x)$.

(4) For a normed space X, a sequence $x_n \xrightarrow{w} x$ in X iff there is $C > 0$ such that $\|x_n\| < C$ for $n = 1, 2, 3, \ldots$ and $M = \{f \in X^* : \lim_{n\to\infty} f(x_n) = f(x)\}$ is dense in X^*. This follows from the uniform boundedness principle, part (b) of the Banach-Steinhaus theorem and the canonical embedding theorem that $\|i_{x_n}\| = \|x_n\|$.

(5) For a convex subset C of a normed space X, we have $\overline{C} = \overline{C}^w$. Hence C is closed iff it is w-closed. Also, C is dense iff it is w-dense.

Proof. For the first statement, since the weak topology is a subset of the norm topology, $\overline{C} \subseteq \overline{C}^w$.

Conversely, assume there is $x_0 \in \overline{C}^w \setminus \overline{C}$. By the separation theorem, there is $f \in X^*$ such that $\operatorname{Re} f(x_0) < s = \inf\{\operatorname{Re} f(x) : x \in \overline{C}\}$. Since f is w-continuous, $U = \{x \in X : \operatorname{Re} f(x) < s\} = f^{-1}(\{z \in \mathbb{K} : \operatorname{Re} z < s\})$ is a w-open neighborhood of x_0 and disjoint from C, hence also from $\overline{C}^w$. So $x_0 \notin \overline{C}^w$, a contradiction. Therefore $\overline{C} = \overline{C}^w$. The second and third statements follow easily from the first statement. ∎

Similarly, on a dual space $X^* = L(X, \mathbb{K})$ (which is a normed space), for each $x \in X$, consider i_x as in the canonical embedding. We can take $\mathcal{P} = \{|i_x| : x \in X\}$ to generate a topology w^* on X^* so that all i_x are continuous. This topology w^* on X^* is called the *weak-star topology on X^**. Then X^* with this topology is a locally convex space. Using the description of a base of 0 in a locally convex space, we see sets of the form $U^* = \bigcap_{i=1}^{n} \{f \in X^* : |f(x_i)| < r_i\}$, where $r_i > 0$ and $x_i \in X$, form a base at 0 for the weak-star topology.

Thus, on a dual space X^*, there are more than one topologies we will be using, namely the original norm-topology and the w^*-topology. When we mean X^* with w^*-topology, we shall write (X^*, w^*).

Properties of Weak-star Topologies.

(1) By definition of weak-star topology, we have the w^*-topology is a subset of the norm-topology. So w^*-open sets are open in X^*, w^*-closed sets are closed in X^*, but compact sets in X are w^*-compact.

(2) For nonzero $f \in X^*$, there is $x \in X$ such that $f(x) \neq 0$. Then $|i_x(f)| \neq 0$. So $\mathcal{P} = \{|i_x| : x \in X\}$ is separating. This implies the w^* topology is Hausdorff and w^*-compact sets are w^*-closed. In case $\dim X^* < \infty$, by the finite dimension theorem, the norm, weak and weak-star topologies are equal.

(3) For a net $\{f_\beta\}_{\beta \in J}$ in X^*, we have $\{f_\beta\}_{\beta \in J}$ w^*-converges to f in X^* (write as $f_\beta \xrightarrow[\mathrm{w*}]{} f$) iff for every $x \in X$, $f_\beta(x) \to f(x)$.

(4) Let X be a *Banach* space. A sequence $f_n \xrightarrow[\mathrm{w*}]{} f$ in X^* iff there is $C > 0$ such that $\|f_n\| < C$ for $n = 1, 2, 3, \ldots$ and $M = \{x \in X : \lim_{n\to\infty} f_n(x) = f(x)\}$ is dense in X. This follows from the uniform boundedness principle and part (b) of the Banach-Steinhaus theorem.

Next, we will show that for a convex subset C of a dual space X^*, $\overline{C} = \overline{C}^{w^*}$ may not hold.

Lemma. *Let $g, g_1, \ldots, g_n$ be linear functionals on a vector space X. The following are equivalent.*

(a) *There are $c_1, \ldots, c_n \in \mathbb{K}$ such that $g = c_1 g_1 + \cdots + c_n g_n$.*

(b) *There exists $c > 0$ such that for all $z \in X$, $|g(z)| \leq c \max\{|g_j(z)| : j = 1, 2, \ldots, n\}$.*

(c) $\bigcap_{j=1}^{n} \ker g_j \subseteq \ker g$.

Proof. For (a) $\Rightarrow$ (b), take $c = |c_1| + |c_2| + \cdots + |c_n|$. Next, (b) $\Rightarrow$ (c) is obvious. For (c) $\Rightarrow$ (a), define $T : X \to \mathbb{K}^n$ by $T(x) = \big(g_1(x), \ldots, g_n(x)\big)$. Then $\ker T = \ker g_1 \cap \cdots \cap \ker g_n$. If $T(x) = T(x')$, then $x - x' \in \ker T \subseteq \ker g$ and so $g(x) = g(x')$. Choose a basis for $\operatorname{ran} T$ and extend it to a basis for $\mathbb{K}^n$. Define a linear transformation $G : \mathbb{K}^n \to \mathbb{K}$ such that $G(T(x)) = g(x)$ for $x \in X$ and $G(v) = 0$ for v in the extended part of the basis. Then $g = G \circ T$. For the standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{K}^n$, let $c_i = G(e_i)$, then $G(x_1, \ldots, x_n) = G(x_1 e_1 + \cdots + x_n e_n) = c_1 x_1 + \cdots + c_n x_n$. Therefore, $g = G \circ T = c_1 g_1 + \cdots + c_n g_n$. ∎

Weak-star Functional Theorem. *Let X be a normed space. If $g : X^* \to \mathbb{K}$ is linear and continuous with the weak-star topology on X^*, then $g = i_x$ for some $x \in X$.*

Proof 1. As $g^{-1}(B(0,1))$ is a w^*-open neighborhood of 0, we get $0 \in \bigcap_{j=1}^{n} \{z \in X^* : |z(x_j)| < r_j\} \subseteq g^{-1}(B(0,1))$ for some $x_1, \ldots, x_n \in X$ and $r_1, \ldots, r_n > 0$. Let $0 < s < \min\{r_1, r_2, \ldots, r_n\}$ and $c = 1/s$. For every $z \in X$, we claim $|g(z)| \leq cr$, where $r = \max\{|i_{x_j}(z)| : j = 1, 2, \ldots, n\}$. If $r = 0$, then for all $t > 0$, $|tz(x_j)| = 0 < r_j$, hence $tz \in g^{-1}(B(0,1))$, which implies $|g(z)| < 1/t \to 0$ as $t \to \infty$. If $r > 0$, then $|i_{x_j}(sz/r)| = |sz(x_j)/r| \leq s < r_j$ for $j = 1, 2, \ldots, n$. So, $|g(sz/r)| < 1$, i.e. $|g(z)| < cr$. By lemma, this implies $g = c_1 i_{x_1} + \cdots + c_n i_{x_n} = i_x$, where $x = c_1 x_1 + \cdots + c_n x_n$. ∎

Proof 2. For a fixed $r > 0$ and a linear transformation $\phi : X^* \to \mathbb{K}$, we have

$$w \in \ker\phi \quad \Leftrightarrow \quad \phi(w) = 0 \quad \Leftrightarrow \quad \forall\, t > 0,\ |\phi(w)| < tr \quad \Leftrightarrow \quad \forall\, t > 0,\ w = tz,\ \text{where } |\phi(z)| < r,$$

i.e. $\ker\phi = \bigcap_{t>0} t\{z \in X^* : |\phi(z)| < r\}$. For $g^{-1}(B(0,1))$, we have $0 \in \bigcap_{j=1}^{n}\{z \in X^* : |z(x_j)| < r_j\} \subseteq g^{-1}(B(0,1))$ for some $x_1, \ldots, x_n \in X$ and $r_1, \ldots, r_n > 0$. Then

$$\begin{aligned}\bigcap_{j=1}^{n} \ker i_{x_j} &= \bigcap_{j=1}^{n}\bigcap_{t>0} t\{z \in X^* : |i_{x_j}(z)| = |z(x_j)| < r_j\} = \bigcap_{t>0} t \bigcap_{j=1}^{n}\{z \in X^* : |z(x_j)| < r_j\}\\ &\subseteq \bigcap_{t>0} t g^{-1}(B(0,1)) = \bigcap_{t>0} t\{z \in X^* : |g(z)| < 1\} = \ker g.\end{aligned}$$

By the last lemma, this implies $g = c_1 i_{x_1} + \cdots + c_n i_{x_n} = i_x$, where $x = c_1x_1 + \cdots + c_nx_n$. ∎

Remark. Now we show for a convex subset C of a dual space X^*, $\overline{C} = \overline{C}^{w^*}$ may not hold. Let X be a nonreflexive normed space. Take a $g \in X^{**} \setminus i(X)$. Then $C = \ker g$ is convex and norm-closed in X^*. If $C = \ker g$ is w^*-closed, then by the closed kernel theorem, g would be a w^*-continuous linear functional, hence in $i(X)$ by the weak-star functional theorem, a contradiction.

Theorem (Tychonoff). *The Cartesian product S of a family of compact spaces $\{S_\alpha : \alpha \in A\}$ is compact.*

Proof. (*Due to Paul Chernoff*) Below an element of S will be viewed as a function $s : A \to \cup\{S_\alpha : \alpha \in A\}$ with $s(\alpha) = s_\alpha \in S_\alpha$ for all $\alpha \in A$. Let $\{x_i\}_{i\in I}$ be a net in S.

For $B \subseteq A$, let S_B be the Cartesian product of $\{S_\alpha : \alpha \in B\}$. We say $p \in S_B$ is a *partial cluster point* of $\{x_i\}_{i\in I}$ iff $B \subseteq A$ and $\{x_i|_B\}_{i\in I}$ in S_B has p as a cluster point, i.e. for every neighborhood U of p in S_B and every $i \in I$, there exists $j \succeq i$ such that $x_j|_B \in U$. (Suffice to check U in the base of product topology.)

Order the set X of all partial cluster points of $\{x_i\}_{i\in I}$ by inclusion (i.e. $p_0 \preceq p_1$ iff $\operatorname{dom} p_0 \subseteq \operatorname{dom} p_1 \subseteq A$ and $p_1|_{\operatorname{dom} p_0} = p_0$). For a chain C, define p by $\Gamma(p) = \cup\{\Gamma(q) : q \in C\}$. Let $E = \operatorname{dom} p$. Let $p \in U = \cap_{i=1}^{n} \pi_{\alpha_i}^{-1}(N_{\alpha_i})$ in S_E, where $\alpha_i \in E, p(\alpha_i) \in N_{\alpha_i} \subset S_{\alpha_i}$. C totally ordered implies $\operatorname{dom} q \supseteq \{\alpha_1, \ldots, \alpha_n\}$ for a $q \in C$. Now $q \in X$ implies $p \in X$ due to $p(\alpha) \in S_\alpha$ for $\alpha \in E \setminus \{\alpha_1, \ldots, \alpha_n\}$. So p is an upper bound of C.

By Zorn's lemma, X has a maximal element $P \in S_D$. We claim $D = A$ (which will lead to P as a cluster point of $\{x_i\}_{i\in I}$ and so S is compact). Assume there is $\delta \in A \setminus D$. Since P is a cluster point of $\{x_i|_D\}_{i\in I}$, by exercise 14 on page 6, some subnet $\{y_j\}_{j\in J}$ of $\{x_i\}_{i\in I}$ satisfies $\{y_j|_D\}_{j\in J} \to P$. Since S_δ is compact, some subnet $\{z_k\}_{k\in K}$ of $\{y_j\}_{j\in J}$ satisfies $\{z_k(\delta)\}_{k\in K}$ converging to some $p \in S_\delta$. Define $Q \in S_{D\cup\{\delta\}}$ by $Q|_D = P$ and $Q(\delta) = p$. Then $Q \in X$ and $Q \succeq P$, contradicting the maximality of P. ∎

Remark. In 1950, John Kelley proved that Tychonoff's theorem was equivalent to the axiom of choice.

Theorem (Banach-Alaoglu). *Let X be a normed space. The closed unit ball B^* of X^* is w^*-compact, i.e. B^* is compact in the weak-star topology.*

Proof. For each $x \in X$, let D_x be the closed disk with center 0 and radius $\|x\|$ in $\mathbb{K}$. By Tychonoff's theorem, $D = \prod_{x\in X} D_x$ is compact. For $x \in X$ and $d \in D$, let d_x denote the x-coordinate of d, i.e. $d_x = \pi_x(d)$. For every $y \in B^*$ and $x \in X$, since $\|y\| \le 1$, $|y(x)| \le \|y\|\|x\| \le \|x\|$. So we may define $f : B^* \to D$ by letting $f(y) \in D$ to satisfy $f(y)_x = \pi_x(f(y)) = y(x)$ for all $x \in X$. Now f is injective because $f(y_1) = f(y_2)$ implies for all $x \in X$, $y_1(x) = f(y_1)_x = f(y_2)_x = y_2(x)$, i.e. $y_1 = y_2$. Also, f is a homeomorphism from B^* (with the relative w^*-topology) onto $f(B^*)$ (with the relative product topology) because

$$\begin{aligned}\{z_\alpha\}_{\alpha\in I} \xrightarrow{w*} z \text{ in } B^* &\iff \forall\, x \in X,\ \{z_\alpha(x)\}_{\alpha\in I} \to z(x) \text{ in } \mathbb{K} \qquad (\text{by property 3 of } w^*\text{-topology})\\ &\iff \forall\, x \in X,\ \{\pi_x(f(z_\alpha))\}_{\alpha\in I} \to \pi_x(f(z)) \text{ in } \mathbb{K} \qquad (\text{by definition of } f(z_\alpha))\\ &\iff \{f(z_\alpha)\}_{\alpha\in I} \to f(z) \text{ in } f(B^*) \qquad (\text{by theorem on net convergence on page 7}).\end{aligned}$$

To see B^* is w^*-compact, we show $f(B^*)$ is closed (hence compact) in D. Suppose $\{f(y_\beta)\}_{\beta\in J} \to w \in D$. Then for all $x \in X$, $f(y_\beta)_x \to w_x \in D_x$. Since $y_\beta \in B^* \subseteq X^*$, for every $a, b, x \in X$ and $c \in \mathbb{K}$, $f(y_\beta)_{a+b} = y_\beta(a+b) = y_\beta(a) + y_\beta(b) = f(y_\beta)_a + f(y_\beta)_b$. Taking limit, we get $w_{a+b} = w_a + w_b$. Similarly, $w_{cx} = cw_x$. Define $W : X \to \mathbb{K}$ by $W(x) = w_x$. Then W is linear and $w_x \in D_x$ implies $|W(x)| = |w_x| \le \|x\|$. So $W \in B^*$. Therefore, $w = f(W) \in f(B^*)$. ∎

Remarks. Using the Krein-Milman theorem and the Banach-Alaoglu theorem, it follows that the Banach spaces $C([0,1],\mathbb{R}), L^1([0,1]), c_0$ are not dual spaces of Banach spaces since their closed unit balls have too few extreme points and hence, the closed unit balls cannot be the closed convex hulls of the extreme points, see [Be], p. 110.

Theorem (Helly). *Let X be a Banach space. If X is separable, then the closed unit ball B^* of X^* is w^*-sequentially compact (and hence all bounded sequences in X^* have w^*-convergent subsequences.)*

Proof. Let S be a countable dense subset of X. Let $g_n \in B^*$. By a diagonalization argument (as in the proof of the Arzela-Ascoli theorem), there is a subsequence g_{n_k} such that $\lim_{k\to\infty} g_{n_k}(s)$ exists for all $s \in S$. Next, for every $x \in X$, we will show $\{g_{n_k}(x)\}$ is a Cauchy sequence, hence it converges. This is because for every $\varepsilon > 0$, there are $s \in S$ such that $\|x - s\| < \varepsilon/3$ and $N \in \mathbb{N}$ such that $j, k \ge N$ implies $|g_{n_k}(s) - g_{n_j}(s)| < \varepsilon/3$. So $j, k \ge N$ implies

$$\begin{aligned}|g_{n_k}(x) - g_{n_j}(x)| &\le |g_{n_k}(x) - g_{n_k}(s)| + |g_{n_k}(s) - g_{n_j}(s)| + |g_{n_j}(s) - g_{n_j}(x)|\\ &\le \|g_{n_k}\|\|x - s\| + |g_{n_k}(s) - g_{n_j}(s)| + \|g_{n_j}\|\|s - x\|\\ &< 1(\varepsilon/3) + (\varepsilon/3) + 1(\varepsilon/3) = \varepsilon.\end{aligned}$$

By part (a) of the Banach-Steinhaus theorem, $g(x) = \lim_{k\to\infty} g_{n_k}(x) \in X^*$ and $\|g\| \le \liminf_{k\to\infty} \|g_{n_k}\| \le 1$. By property (3) of weak-star topologies, we have $g_{n_k} \xrightarrow[\text{w*}]{} g \in B^*$ in X^*. ∎

Remarks. The converse of the theorem is false. If X is a nonseparable reflexive space, then by the Eberlein-Smulian theorem in the next section, the closed unit ball of X^* is still w^*-sequentially compact.

§4. **Reflexivity.** Next we may inquire when the closed unit ball B of a normed space X is w-compact. To answer this, let B^{**} be the closed unit balls of X^{**}. We have the following theorem.

Theorem (Goldstine). *Let X be a normed space. Then $B^{**} = \overline{i(B)}^{w^*}$, where i is the canonical embedding. (Hence, $i(X)$ is w^*-dense in X^{**} because $X^{**} = \bigcup_{n=1}^{\infty} nB^{**} = \bigcup_{n=1}^{\infty} \overline{i(nB)}^{w^*} \subseteq \overline{i(X)}^{w^*} \subseteq X^{**}$.)*

Proof. By the Banach-Alaoglu theorem, B^{**} is w^*-compact, hence w^*-closed. Also, since i an isometry, $B^{**} \supseteq i(B)$. Hence $B^{**} \supseteq \overline{i(B)}^{w^*}$. Assume there is $y \in B^{**} \setminus \overline{i(B)}^{w^*}$. Since $\overline{i(B)}^{w^*}$ is convex and w^*-closed, by the separation theorem, there is a w^*-continuous linear functional g on X^{**} such that $\operatorname{Re} g(y) < \inf\{\operatorname{Re} g(u) : u \in \overline{i(B)}^{w^*}\}$. By the weak-star functional theorem, $-g = i_z$ for some $z \in X^*$. For all $u \in X^{**}$, let $f(u) = -g(u) = i_z(u) = u(z)$. Observe that there is $c \in \mathbb{K}$ with $|c| = 1$ such that $|z(x)| = z(cx) = \operatorname{Re} z(cx)$. Using $cB = B$ in the second equality below, we have

$$\begin{aligned}\|f\|\|y\| \ge |f(y)| \ge \operatorname{Re} f(y) &> \sup\{\operatorname{Re} f(u) : u \in \overline{i(B)}^{w^*}\}\\ &\ge \sup\{\operatorname{Re} u(z) : u = i_x \in i(B)\} = \sup\{\operatorname{Re} z(x) : x \in B\}\\ &= \sup\{|z(x)| : x \in B\} = \|z\| = \|i_z\| = \|f\|.\end{aligned}$$

Then $\|y\| > 1$, i.e. $y \notin B^{**}$, a contradiction. Therefore, $B^{**} = \overline{i(B)}^{w^*}$. ∎

Remarks. (1) We have $i(B) = B^{**}$ if and only if $i(X) = X^{**}$. This is because $i(B) = B^{**}$ implies $i(X) = \operatorname{span} i(B) = \operatorname{span} B^{**} = X^{**}$ and conversely, if $i(X) = X^{**}$, then for all $f \in B^{**} \subseteq X^{**} = i(X)$, we have $f = i_x$ for some $x \in X$ (with $\|x\| = \|f\| \le 1$ due to i is an isometry) so that $f \in i(B)$.

(2) The canonical embedding $i : X \to i(X)$ is a homeomorphism when we take the w-topology on X and the w^* topology on X^{**}. This is because it is bijective and

$$x_\alpha \xrightarrow[\text{w}]{} x \iff \forall f \in X^*,\ f(x_\alpha) \to f(x) \iff \forall f \in X^*,\ i_{x_\alpha}(f) \to i_x(f) \iff i_{x_\alpha} \xrightarrow[\text{w*}]{} i_x.$$

Theorem (Banach-Smulian). *A normed space is reflexive iff its closed unit ball B is w-compact.*

Proof. By the remarks and Goldstine's theorem, B is w-compact in X iff $i(B)$ is w^*-compact (hence w^*-closed) in X^{**} and B^{**} iff $i(B) = \overline{i(B)}^{w*} = B^{**}$ iff $i(X) = X^{**}$. $\square$

Question. For a reflexive space X, is the closed unit ball of X^* w-sequentially compact? Yes.

First we need to know more facts. Now reflexive spaces are dual spaces, hence they are complete. Which Banach spaces are reflexive? Also, observe that in addition to the w^*-topology on X^*, there is also the weak topology on X^*. Since $\{|f| : f \in X^{**}\} \supseteq \{|i_x| : x \in X\}$, so the weak-star topology on X^* is a subset of the weak topology (which is a subset of the norm topology) on X^*. Hence, on X^*, w^*-open sets are w-open, w^*-closed sets are w-closed, but w-compact sets are w^*-compact. When are the w-topology and w^*-topology equal in X^*? The following theorem will answer both questions.

Theorem. *Let X be a Banach space. The following are equivalent.*

(a) *X is reflexive (i.e. $X = X^{**}$.*

(b) *On X^*, the weak topology is the same as the weak-star topology.*

(c) *X^* is reflexive (i.e. $X^* = X^{***}$).*

Proof. (a) $\Rightarrow$ (b) By (a), $\{|f| : f \in X^{**}\} = \{|i_x| : x \in X\}$. So both topologies are generated by the same seminorms.

(b) $\Rightarrow$ (c) By the Banach-Alaoglu theorem, the closed unit ball B^* of X^* is w^*-compact, hence w-compact by (b). By the Banach-Smulian theorem, X^* is reflexive.

(c) $\Rightarrow$ (a) Since the canonical embedding is an isometry and the closed unit ball B of X is closed, hence complete, in X, so $i(B)$ is complete, hence closed, in X^{**}. As $i(B)$ is convex, by property (5) of weak topology, it is w-closed in X^{**}. Since X^* is reflexive, applying (a) $\Rightarrow$ (b) to X^*, we see $i(B)$ is also w^*-closed in X^{**}. By Goldstine's theorem, $i(B) = \overline{i(B)}^{w^*} = B^{**}$. By remark (1) above, $i(X) = X^{**}$. $\square$

Theorem (Pettis). *If X is reflexive and M is a closed vector subspace of X, then M is reflexive.*

Proof. Let $z \in M^{**}$. We have to show $z = i_w$ for some $w \in M$. Define $T : X^* \to M^*$ by $Tf = f|_M$. Since $\|f|_M\| \le \|f\|$, we get $T \in L(X^*, M^*)$. Then $z \circ T \in X^{**} = i(X)$. So there is $w \in X$ such that $z \circ T = i_w$, i.e. $z(Tf) = f(w)$ for all $f \in X^*$.

Assume $w \in X \setminus M$. By the Hahn-Banach theorem, there is $g \in X^*$ such that $g|_M = 0$ and $g(w) \neq 0$. Then $Tg = g|_M = 0$. Then $0 = z(Tg) = g(w) \neq 0$, a contradiction. Hence $w \in M$. Now for every $h \in M^*$, by the Hahn-Banach theorem, there exists $H \in X^*$ extending h (i.e. $TH = H|_M = h$). Then $z(h) = z(TH) = H(w) = h(w) = i_w(h)$ for all $h \in M^*$. Therefore, $z = i_w$. $\square$

Exercise. Prove that X is reflexive iff for any closed vector subspace M of X, M and X/M are reflexive. See [KR], pp. 8-9.

Theorem (Banach). *For a normed space X, if X^* is separable, then X is separable.*

Proof. $X = \{0\}$ is a trivial case. For $X \neq \{0\}$, let D be a countable dense subset of X^*. For every $f \in D$, by the definition of $\|f\|$ and the supremum property, there is $x_f \in X$ such that $\|x_f\| = 1$ and $|f(x_f)| \ge \|f\|/2$. Let S be the set of all finite linear combinations of the x_f's with $\mathbb{K} \cap (\mathbb{Q} + i\mathbb{Q})$ coefficients. Then S is countable.

Next we will show S is dense in X. By part (b) of the Hahn-Banach theorem, it suffices to show $F \in X^*$ satisfying $F \equiv 0$ on $\overline{S}$ must be the zero functional. Since D is dense in X^*, there exists a sequence $\{f_n\}$ in D converging to F. We have $\|f_n - F\| \geq |(f_n - F)(x_{f_n})| = |f_n(x_{f_n})| \geq \|f_n\|/2$, which implies $\|f_n\| \to 0$. Then $F = 0$. $\square$

Remarks. The converse is false in general. For example, ℓ^1 is separable, but $(\ell^1)^* = \ell^\infty$ is not separable. However, if X is a reflexive and separable Banach space, then since i is an isometry, $X^{**} = i(X)$ is separable and hence X^* is separable by Banach's theorem.

Theorem (Eberlein-Smulian). *If X is reflexive, then the closed unit ball B of X is w-sequentially compact (and hence all bounded sequences in X have w-convergent subsequences).*

Proof. Let $\{x_n\}$ be a sequence in B. Let M be the closed linear span of $\{x_n\}$. By Pettis' theorem, M is reflexive. Also M is separable as the set of all finite linear combinations of $\{x_n\}$ with $\mathbb{K} \cap (\mathbb{Q} + i\mathbb{Q})$ coefficients is dense. By the remark above, M^* is separable. By Helly's theorem, $\{i_{x_n}\}$ in the closed unit ball B^{**} of M^{**} has a w^*-convergent subsequence $\{i_{x_{n_k}}\}$. By remark (2) before the Banach-Smulian theorem, $\{x_{n_k}\}$ is a w-convergent subsequence of $\{x_n\}$ in M, say $x_{n_k} \xrightarrow[\text{w}]{} x \in M$. For all $f \in X^*$, we have $f|_M \in M^*$. By property 3 of weak topology, $f(x_{n_k}) = f|_M(x_{n_k}) \to f|_M(x) = f(x)$, i.e. $\{x_{n_k}\}$ w-converges to x in X.

Remarks. In fact, Eberlein-Smulian proved a much deeper theorem, namely on any normed space (not necessarily reflexive), a subset is w-compact iff it is w-sequentially compact. See [M], pp. 248-250.

Clearly every finite dimensional normed space is reflexive. If X is an infinite dimensional normed space, must X have some reflexive closed linear subspaces, other than the finite dimensional subspaces? The answer turns out to be negative. Below, we will show the only reflexive subspaces of ℓ^1 are the finite dimensional subspaces. Let M be a reflexive closed linear subspace of $X = \ell^1$. By the Eberlein-Smulian theorem, the closed unit ball of M is w-sequentially compact. Schur's lemma below asserts that every w-convergent sequence in ℓ^1 is convergent in the norm topology of ℓ^1. Hence, the closed unit ball of M would be compact. By Riesz' lemma, M would be finite dimensional. This implies ℓ^1 is not reflexive and its only reflexive closed linear subspaces are the finite dimensional subspaces.

Theorem (Schur's Lemma). *If $\{x^{(n)}\}$ is w-convergent in ℓ^1, where $x^{(n)} = (x_1^{(n)}, x_2^{(n)}, x_3^{(n)}, \ldots)$ for $n = 1, 2, 3, \ldots$, then $\{x^{(n)}\}$ is convergent in the norm topology of ℓ^1.*

Proof. (Sliding Hump Argument) Assume $x^{(n)} \xrightarrow[\text{w}]{} x$ in ℓ^1, but $x^{(n)} \to x$ is false. Replacing $x^{(n)}$ by $x^{(n)} - x$ if necessary, we may assume $x = 0$. Since $\|x^{(n)}\|_1 \to 0$ is false, passing to a subsequence, we may assume there is an $\varepsilon > 0$ such that (i) $\|x^{(n)}\|_1 = \displaystyle\sum_{j=1}^{\infty} |x_j^{(n)}| > \varepsilon$ for $n = 1, 2, 3, \ldots$.

Since $x^{(n)} \xrightarrow[\text{w}]{} 0$ in ℓ^1, by property 3 of weak topology, $\langle x^{(n)}, z \rangle = \displaystyle\sum_{j=1}^{\infty} z_j x_j^{(n)} \to 0$ as $n \to \infty$ for every $z = (z_1, z_2, z_3, \ldots) \in \ell^\infty = (\ell^1)^*$. Our goal is to construct a special z with all $|z_j| = 1$ to get a contradiction of the last sentence.

First, by taking $z = (0, \ldots, 0, 1, 0, \ldots)$, where 1 is in the j-th coordinate, we have for all $j = 1, 2, 3, \ldots$, $x_j^{(n)} \to 0$ as $n \to \infty$.

Next, define sequences $\{m_k\}, \{n_k\}$ as follows. Set $m_0 = 1, n_0 = 0$. Inductively, for $k \geq 1$, suppose m_{k-1} and n_{k-1} are determined. By the last paragraph, $\displaystyle\lim_{n\to\infty} \sum_{j=1}^{m_{k-1}} |x_j^{(n)}| = \sum_{j=1}^{m_{k-1}} \lim_{n\to\infty} |x_j^{(n)}| = 0$. Then there exists an integer $n = n_k > n_{k-1}$ such that (ii) $\displaystyle\sum_{j=1}^{m_{k-1}} |x_j^{(n)}| < \frac{\varepsilon}{5}$. Since $\displaystyle\sum_{j=1}^{\infty} |x_j^{(n_k)}| = \|x^{(n_k)}\|_1 < \infty$, there exists an integer $m = m_k > m_{k-1}$ such that (iii) $\displaystyle\sum_{j=m+1}^{\infty} |x_j^{(n_k)}| < \frac{\varepsilon}{5}$. By (i), (ii), (iii), we have (iv) $\displaystyle\sum_{j=m_{k-1}+1}^{m_k} |x_j^{(n_k)}| > \frac{3\varepsilon}{5}$.

Now observe that $1 = m_0 < m_1 < m_2 < \cdots$. Recall that sgn α is the *signum* function defined to be $|\alpha|/\alpha$ if $\alpha \neq 0$ and 1 if $\alpha = 0$. Let $z = (z_1, z_2, \ldots) \in \ell^\infty$ be defined by $z_1 = 1$ and for $k = 1, 2, 3, \ldots$ and $m_{k-1} < j \leq m_k$, $z_j = \operatorname{sgn} x_j^{(n_k)}$. By the conditions on n_k and m_k, we have (v) $z_j x_j^{(n_k)} = |x_j^{(n_k)}|$ for $m_{k-1} < j \leq m_k$. For $k = 1, 2, 3, \ldots$, by (ii), (iii), (iv) and (v),

$$\begin{aligned}\Big|\sum_{j=1}^{\infty} z_j x_j^{(n_k)}\Big| &\geq \Big|\sum_{j=m_{k-1}+1}^{m_k} z_j x_j^{(n_k)}\Big| - \Big|\sum_{j=1}^{m_{k-1}} z_j x_j^{(n_k)}\Big| - \Big|\sum_{j=m_k+1}^{\infty} z_j x_j^{(n_k)}\Big| \\ &\geq \underbrace{\sum_{j=m_{k-1}+1}^{m_k} |x_j^{(n_k)}|}_{hump} - \underbrace{\sum_{j=1}^{m_{k-1}} |x_j^{(n_k)}|}_{front} - \underbrace{\sum_{j=m_k+1}^{\infty} |x_j^{(n_k)}|}_{tail} \\ &> \frac{3\varepsilon}{5} - \frac{\varepsilon}{5} - \frac{\varepsilon}{5} = \frac{\varepsilon}{5},\end{aligned}$$

which is a contradiction to $\displaystyle\sum_{j=1}^{\infty} z_j x_j^{(n)} \to 0$ as $n \to \infty$. ∎

Here is the reason why the proof is called a sliding hump argument. For each $x^{(n_k)} \in \ell^1$, if we plot the graph of $f_{n_k}(x) = \displaystyle\sum_{j=1}^{\infty} |x_j^{(n_k)}| \mathcal{X}_{(j-1,j]}(x)$ on the coordinate plane, then the area under the curve is greater than ε and the areas under the curve on $(0, m_{k-1}]$ and (m_k, ∞) are both less than $\varepsilon/5$ so that the area under the curve on $(m_{k-1}, m_k]$ is greater than $3\varepsilon/5$. Thus, we can say there is a hump in the middle portion over $(m_{k-1}, m_k]$. As k takes on the values $1, 2, 3, \ldots$, since $1 = m_0 < m_1 < m_2 < \ldots$, the humps of $f_{n_k}(x)$ start to slide along the intervals $(m_0, m_1], (m_1, m_2], (m_2, m_3], \ldots$. Since the union of these intervals is $(0, \infty)$, we can patch up the $\operatorname{sgn} x_j^{(n_k)}$ on the intervals to get a $z \in \ell^\infty$ to get a contradiction.

Appendix : Examples of Sliding Hump Technique

Below are some historical examples of the sliding hump arguments.

Uniform Boundedness Principle. *Let X be a Banach space, Y a normed space and $A \subseteq L(X, Y)$. If $c_x = \sup\{\|Tx\| : T \in A\} < \infty$ for every $x \in X$, then $\sup\{\|T\| : T \in A\} < \infty$.*

Proof. If $\{\|T\| : T \in A\}$ is unbounded, then there are $T_1 \in A$ (with $\|T_1\| \geq 12$) and $\|x\| \leq 1$ such that $\|T_1\| \geq \|T_1 x\| \geq \frac{3}{4}\|T_1\|$. Let $x_1 = \frac{1}{3}x$, then $\|x_1\| \leq \frac{1}{3}$ and $\|T_1 x_1\| \geq \frac{1}{4}\|T_1\|$. Inductively we can find $T_2, T_3, \ldots \in A$ and $x_2, x_3, \ldots \in X$ such that for all $n \geq 2$, $\|T_n\| \geq 4 \cdot 3^n \Big(\displaystyle\sum_{k=1}^{n-1} c_{x_k} + n\Big)$, $\|x_n\| \leq \dfrac{1}{3^n}$ and $\|T_n x_n\| \geq \dfrac{3}{4 \cdot 3^n}\|T_n\|$. Since $\displaystyle\sum_{k=1}^{\infty} \|x_k\| < \infty$ and X is a Banach space, so $\displaystyle\sum_{k=1}^{\infty} x_k$ converges to some $x \in X$. Observe that

$$T_n x = \underbrace{\overbrace{T_n x_1 + \cdots + T_n x_{n-1}}^{small}}_{call\ this\ I} + \overbrace{T_n x_n}^{hump} + \underbrace{\overbrace{T_n x_{n+1} + T_n x_{n+2} + \cdots}^{small}}_{call\ this\ J}.$$

We have

$$\|I\| = \Big\|\sum_{k=1}^{n-1} T_n x_k\Big\| \leq \sum_{k=1}^{n-1} \|T_n x_k\| \leq \sum_{k=1}^{n-1} c_{x_k} \leq \frac{1}{4 \cdot 3^n}\|T_n\| \leq \frac{1}{3}\|T_n x_n\|,$$

$$\|J\| = \Big\|\sum_{k=n+1}^{\infty} T_n x_k\Big\| \leq \sum_{k=n+1}^{\infty} \|T_n x_k\| \leq \sum_{k=n+1}^{\infty} \frac{1}{3^k}\|T_n\| \leq \frac{1}{2\cdot 3^n}\|T_n\| \leq \frac{2}{3}\|T_n x_n\|.$$

These lead to the contradiction that for every $n \in \mathbb{N}$,

$$c_x \geq \|T_n x\| \geq \underbrace{\|T_n x_n\|}_{hump} - \underbrace{\|I\|}_{front} - \underbrace{\|J\|}_{tail} \geq \frac{3}{4\cdot 3^n}\|T_n\| - \sum_{k=1}^{n-1} c_{x_k} - \frac{1}{2\cdot 3^n}\|T_n\| \geq n.$$

Remark. As n increases, the hump slides to the right!

For all $f \in L^1(-\pi,\pi]$ and $n \in \mathbb{Z}$, define the *n-th Fourier coefficient* of f to be $\widehat{f}(n) = \int_{(-\pi,\pi]} f(\theta)e^{-in\theta}\frac{dm}{2\pi}$. The *Fourier series* of f is $\sum_{k=-\infty}^{\infty} \widehat{f}(k)e^{ikx}$ and its *n-th partial sum* is $s_n(f;x) = \sum_{k=-n}^{n} \widehat{f}(k)e^{ikx}$. The function $D_n(x) = \sum_{k=-n}^{n} e^{ikx} = \frac{\sin(n+\frac{1}{2})x}{\sin(x/2)}$ is called the *Dirichlet kernel*. Using it, we have

$$s_n(f;x) = \sum_{k=-n}^{n} \widehat{f}(k)e^{ikx} = \sum_{k=-n}^{n} \int_{(-\pi,\pi]} f(\theta)e^{ik(x-\theta)}\frac{dm}{2\pi} = \int_{(-\pi,\pi]} f(\theta)D_n(x-\theta)\frac{dm}{2\pi}.$$

Observe that

$$\|D_n\|_1 > \frac{2}{\pi}\int_0^{\pi} \Big|\sin\big(n+\frac{1}{2}\big)\theta\Big|\frac{d\theta}{\theta} = \frac{2}{\pi}\int_0^{(n+1/2)\pi} |\sin\phi|\frac{d\phi}{\phi} > \frac{2}{\pi}\sum_{k=1}^{n}\frac{1}{k\pi}\int_{(k-1)\pi}^{k\pi}|\sin\phi|d\phi = \frac{4}{\pi^2}\sum_{k=1}^{n}\frac{1}{k} \to \infty.$$

Using these facts, Lebesgue constructed a 2π-periodic continuous function on $\mathbb{R}$ with the Fourier series diverging at $x=0$. Here is the version of his sliding hump argument as appeared in Hardy and Rogosinski's book *Fourier Series*, pp. 51-52.

Let $g(\theta) = \operatorname{sgn} D_n(-\theta)$. Then $gD_n = |D_n|$ and $s_n(g;0) = \int_{(-\pi,\pi]} |D_n(-\theta)|\frac{dm}{2\pi} = \|D_n\|_1$. Since g is piecewise constant, we can get a 2π-periodic continuous function f_n on $\mathbb{R}$ such that for every $\theta \in \mathbb{R}$, $|f_n(\theta)| \leq 1$, $\lim_{n\to\infty} f_n(\theta) = g(\theta)$ and $\int_{(-\pi,\pi]} \big|\big(g(\theta)-f_n(\theta)\big)D_n(-\theta)\big|\frac{dm}{2\pi} \leq \frac{1}{2}\|D_n\|_1$. Then $s_n(f_n;0) \geq s_n(g;0) - \frac{1}{2}\|D_n\|_1 = \frac{1}{2}\|D_n\|_1$. If the Fourier series of any of the f_n diverges at $x=0$, then we have a desired function. Otherwise, let the Fourier series of f_n at $x=0$ converge to γ_n.

Observe that the sequence $\alpha_k = 7^{-k}$ has the properties that $\sum_{k=1}^{\infty}\alpha_k < \infty$ and $\sum_{j=k+1}^{\infty}\alpha_j \leq \frac{1}{6}\alpha_k$. For any strictly increasing sequence $\{n_k\}$, by the Weierstrass M-test, $\sum_{k=1}^{\infty}\alpha_k f_{n_k}(t)$ converge uniformly on $\mathbb{R}$ to a 2π-periodic continuous function $f(t)$. By the definition of γ_n, we have $\lim_{n\to\infty} s_n\Big(\sum_{j=1}^{k-1}\alpha_j f_{n_j};0\Big) = \sum_{j=1}^{k-1}\alpha_j\gamma_{n_j}$.

Now we choose the sequence $\{n_k\}$ so that $\alpha_k\|D_{n_k}\|_1 \to \infty$, $\sum_{j=1}^{k-1}\alpha_j|\gamma_{n_j}| \leq \frac{1}{12}\alpha_k\|D_{n_k}\|_1$ and

$$\Big|s_{n_k}\Big(\sum_{j=1}^{k-1}\alpha_j f_{n_j};0\Big)\Big| \leq 2\sum_{j=1}^{k-1}\alpha_j|\gamma_{n_j}| \leq \frac{1}{6}\alpha_k\|D_{n_k}\|_1.$$

Also, $\left|s_{n_k}\Big(\sum_{j=k+1}^{\infty}\alpha_j f_{n_j};0\Big)\right| \le \sum_{j=k+1}^{\infty}\alpha_j\int_{(-\pi,\pi]}|f_{n_j}(\theta)D_{n_k}(-\theta)|\ \dfrac{dm}{2\pi} \le \sum_{j=k+1}^{\infty}\alpha_j\|D_{n_k}\|_1 \le \dfrac{1}{6}\alpha_k\|D_{n_k}\|_1$. Also, $s_{n_k}(\alpha_k f_{n_k};0) = \alpha_k s_{n_k}(f_{n_k};0) \ge \dfrac{1}{2}\alpha_k\|D_n\|_1$. By the last three inequalities,

$$\begin{aligned} s_{n_k}(f;0) = \sum_{j=1}^{\infty}s_{n_k}(\alpha_j f_{n_j};0) &= \underbrace{s_{n_k}(\alpha_k f_{n_k};0)}_{hump} + \underbrace{s_{n_k}\Big(\sum_{j=1}^{k-1}\alpha_j f_{n_j};0\Big)}_{front} + \underbrace{s_{n_k}\Big(\sum_{j=k+1}^{\infty}\alpha_j f_{n_j};0\Big)}_{tail} \\ &\ge \frac{1}{2}\alpha_k\|D_{n_k}\|_1 - \frac{1}{6}\alpha_k\|D_{n_k}\|_1 - \frac{1}{6}\alpha_k\|D_{n_k}\|_1 \\ &= \frac{1}{6}\alpha_k\|D_{n_k}\|_1 \to \infty. \end{aligned}$$

Therefore, the Fourier series of f diverges at $x = 0$.

Remark. In hindsight, we can see

$$s_{n_k}(f;0) = \overbrace{s_{n_k}(\alpha_1 f_{n_1};0) + \cdots + s_{n_k}(\alpha_{k-1}f_{n_{k-1}};0)}^{small} + \overbrace{s_{n_k}(\alpha_k f_{n_k};0)}^{hump} + \overbrace{s_{n_k}(\alpha_{k+1}f_{n_{k+1}};0) + \cdots}^{small},$$

where $|small| \le \frac{1}{6}\alpha_k\|D_{n_k}\|_1 \le \frac{1}{2}\alpha_k\|D_{n_k}\|_1 \le hump$. As k increases, the hump moves to the right!

This argument led to the birth of the uniform boundedness principle (see Dieudonne's book, *History of Functional Analysis*, Chapter VI, §4).

Chapter 4. Duality and Adjoints.

§1. **Duality.** For a closed vector subspace M of a normed space X, we would like to identify M^* and $(X/M)^*$. For this, we first introduce the concept of annihilator of a set.

Definitions. For a nonempty subset M of a normed space X, the *annihilator* of M is

$$M^\perp = \{y \in X^* : \underbrace{\langle x, y\rangle}_{=i_x(y)} = 0 \text{ for all } x \in M\} = \bigcap_{x\in M} \ker i_x,$$

which is w^*-closed and norm-closed. For a nonempty subset N of X^*, the *(pre)annihilator* of N is

$$^\perp N = \{x \in X : \underbrace{\langle x, y\rangle}_{=y(x)} = 0 \text{ for all } y \in N\} = \bigcap_{y\in N} \ker y,$$

which is w-closed and norm-closed.

Remarks. (1) $M^\perp = \overline{M}^\perp$ since for all $y \in X^*$, $y|_M = 0$ iff $y|_{\overline{M}} = 0$. Similarly, $^\perp N = {}^\perp\overline{N}$.

(2) By the definitions above, $\{0\}^\perp = X^*$, $^\perp\{0\} = X$, $X^\perp = \{0\}$. Also, $^\perp(X^*) = \{0\}$, where the left-to-right inclusion is due to part (c) of the Hahn-Banach theorem.

Notations. For a subset M of X, we write $M^{\perp\perp}$ to mean $^\perp(M^\perp)$. For a subset N of X^*, we write $N^{\perp\perp}$ to mean $(^\perp N)^\perp$. From these, we have $M \subseteq M^{\perp\perp} \subseteq X$ and $N \subseteq N^{\perp\perp} \subseteq X^*$.

Although in the definitions of annihilator and preannihilator, M and N may be any nonempty subset of the normed space, in the following, we will only consider the cases M and N are vector subspaces.

Double-Perp Theorem. *Let X be a normed space.*

(a) *If M is a vector subspace of X, then $M^{\perp\perp} = \overline{M}^w = \overline{M}$, the weak-closure or norm-closure of M.*

(b) *If N is a vector subspace of X^*, then $N^{\perp\perp} = \overline{N}^{w^*}$, the weak-star closure of N.*

Proof. (a) Since $M \subseteq M^{\perp\perp}$, so $\overline{M} \subseteq M^{\perp\perp}$. Assume there is $x \in M^{\perp\perp} \setminus \overline{M}$. By part (b) of the Hahn-Banach theorem, there is $y \in X^*$ such that $y|_M = 0$ and $y(x) \neq 0$. Then $y \in M^\perp$ and $x \notin M^{\perp\perp}$, a contradiction. Therefore, $\overline{M} = M^{\perp\perp}$.

(b) Since $N \subseteq N^{\perp\perp}$, so $\overline{N}^{w^*} \subseteq N^{\perp\perp}$. Assume there is $y \in N^{\perp\perp} \setminus \overline{N}^{w^*}$. Applying part (b) of the corollary to the separation theorem to X^* with the w^*-topology and the weak-star functional theorem, there is w^*-continuous linear functional $g = i_x$ on X^* such that $g = i_x \equiv 0$ on N and $y(x) = g(y) \neq 0$. Then $x \in {}^\perp N$ and $y \notin N^{\perp\perp}$, a contradiction. Therefore, $\overline{N}^{w^*} = N^{\perp\perp}$. ∎

Remarks. We have $M^\perp = \{0\}$ iff $\overline{M} = X$, which can be checked by taking (pre)annihilators of both sides. Similarly, $M^\perp = X^*$ iff $M = \{0\}$; $^\perp N = \{0\}$ iff $\overline{N}^{w^*} = X^*$; $^\perp N = X$ iff $N = \{0\}$.

Duality Theorem. *Let M be a closed vector subspace of a normed space X. We have the following isometric isomorphisms and equations.*

(a) $M^* \cong X^*/M^\perp$. *For every $F \in X^*$,* $\sup\{|\langle x, F\rangle| : x \in M, \|x\| \le 1\} = \min\{\|F - G\| : G \in M^\perp\}$.

(b) $(X/M)^* \cong M^\perp$. *For every $x \in X$,* $\inf\{\|x - m\| : m \in M\} = \max\{|\langle x, G\rangle| : G \in M^\perp, \|G\| \le 1\}$.

Proof. (a) Define $\phi : M^* \to X^*/M^\perp$ by $\phi(f) = F + M^\perp$, where $F \in X^*$ is any linear extension of $f \in M^*$. (If F and F' are linear extensions of f, then $F - F' \equiv 0$ on M. So $F - F' \in M^\perp$ and $F + M^\perp = F' + M^\perp$. Hence ϕ is well defined.) Clearly ϕ is linear.

By the Hahn-Banach theorem, we have $M^* = \{F|_M : F \in X^*\}$. For every $F \in X^*$, we have $\phi(F|_M) = F + M^\perp$. So ϕ is surjective. Next we show ϕ is isometric. (This will show ϕ is an isometric isomorphism.) For every $F \in X^*$, let $f = F|_M$. By part (a) of the Hahn-Banach theorem, there is a linear extension $F_f \in X^*$ of $f \in M^*$ such that $\|F_f\| = \|f\|$. Note $\|F|_M\| = \|f\| = \|F_f\|$ and $g = F - F_f \in M^\perp$. For every $G \in M^\perp$, we have $\|F|_M\| = \|(F-G)|_M\| \le \|F-G\|$. Since $F_f \in F+M^\perp$, $\|F|_M\| \le \inf\{\|F-G\| : G \in M^\perp\} = \|F+M^\perp\| \le \|F-g\| = \|F_f\| = \|F|_M\|$. Thus, we have equality throughout. So $\|\phi(F|_M)\| = \|F+M^\perp\| = \|F|_M\|$ showing ϕ is an isometry and the infimum is attained by $g = F - F_f \in M^\perp$.

(b) Recall the quotient map $\pi : X \to X/M$ is defined by $\pi(x) = x + M$. Define $\tau : (X/M)^* \to M^\perp$ by $\tau(F) = F \circ \pi$. (<u>Check $F \circ \pi \in M^\perp$</u>: $\pi \in L(X, X/M)$ and $F \in (X/M)^*$ imply $F \circ \pi \in X^*$. If $x \in M$, then $(F \circ \pi)(x) = F(x + M) = F(M) = F([0]) = 0$ and so $F \circ \pi \in M^\perp$.) Clearly, τ is linear.

Next we will show τ is surjective and isometric. For every $f \in M^\perp$, since $M \subseteq \ker f$, the function $F : X/M \to \mathbb{K}$ given by $F(x + M) = f(x)$ is well-defined, linear and $F \circ \pi = f$. We <u>claim</u> $\|F\| = \|f\|$ (then $F \in (X/M)^*$, $\tau(F) = f$ and τ is an isometric isomorphism).

For all $m \in M$, $|F(x + M)| = |f(x)| = |f(x - m)| \le \|f\|\|x - m\|$. Taking infimum over all $m \in M$, $|F(x + M)| \le \|f\|\|x + M\|$. Then $\|F\| \le \|f\|$ (and so F is continuous). Also, $|f(x)| = |F(x + M)| \le \|F\|\|x + M\| \le \|F\|\|x\|$. So $\|F\| = \|f\| = \|\tau(F)\|$.

For the equation in the second part of (b), let $x \in X$. By part (c) of the Hahn-Banach theorem, there is $F_x \in (X/M)^*$ such that $\|F_x\| = 1$ and $F_x(x + M) = \|x + M\|$. Let $f_x = \tau(F_x) \in M^\perp$, then $f_x(x) = \tau(F_x)(x) = F_x(x + M) = \|x + M\|$. Also, τ isometric implies $\|f_x\| = \|F_x\| = 1$.

For all $G \in M^\perp$, $\|G\| \le 1$ and $m \in M$, we have $|\langle x, G\rangle| = |G(x)| = |G(x - m)| \le \|x - m\|$. Since f_x is such a G, we have

$$f_x(x) \le \sup\{|\langle x, G\rangle| : G \in M^\perp, \|G\| \le 1\} \le \inf\{\|x - m\| : m \in M\} = \|x + M\| = F_x(x + M) = f_x(x).$$

(Thus, there is equality throughout and the supremum is attained by $G = f_x$.)

<u>**Remarks.**</u> If M is a finite dimensional subspace of X, then $\dim M = \dim M^* = \dim(X^*/M^\perp) = \operatorname{codim} M^\perp$ by (a). If M is a closed subspace of finite codimension in X, then $\operatorname{codim} M = \dim(X/M) = \dim(X/M)^* = \dim M^\perp$.

<u>**Example.**</u> Let $W = \Big\{g \in L^2([0,1]) : \displaystyle\int_{[0,1]} g\, dm = 0\Big\}$, where dm is Lebesgue measure. Let $f : [0,1] \to \mathbb{R}$ be defined by $f(t) = t^2$. Find $d(f, W) = \inf\{\|f - g\|_2 : g \in W\}$ in $L^2([0,1])$.

<u>**Solution.**</u> Recall $L^2([0,1])^* = L^2([0,1])$. For $h \in L^2([0,1])$ and $g \in L^2([0,1])^*$, we have $\langle h, g\rangle = \displaystyle\int_{[0,1]} hg\, dm$. Observe that W is a vector subspace of $L^2([0,1])^*$ and for all $g \in W$, we have $\langle 1, g\rangle = 0$ (and also $\langle c, g\rangle = 0$ for all $c \in \operatorname{span} 1 = \mathbb{K}$). So $W = (\operatorname{span} 1)^\perp = \mathbb{K}^\perp$. By part (a) of the duality theorem,

$$d(f, W) = \min\{\|f - g\|_2 : g \in W = \mathbb{K}^\perp\} = \sup\{|\langle c, f\rangle| : c \in \mathbb{K}, |c| \le 1\}.$$

Now $|\langle c, f\rangle| = \Big|\displaystyle\int_{[0,1]} ct^2\, dm\Big| = \dfrac{|c|}{3}$. Therefore, $d(f, W) = \dfrac{1}{3}$.

---

§2. **Adjoints.** Next we introduce adjoint operators. Also, we study how certain properties, such as surjectivity, density of ranges or closure of ranges of operators can be expressed equivalently in terms of adjoint operators.

<u>**Definition.**</u> Let X, Y be normed spaces over $\mathbb{K}$. For $T \in L(X, Y)$ and $y \in Y^*$, define $T^* : Y^* \to X^*$ by $T^*(y) = y \circ T \in X^*$. Thus, for all $x \in X$, $\langle x, T^*(y)\rangle = y(T(x)) = \langle T(x), y\rangle$. T^* is called the <u>*adjoint*</u> of T.

<u>**Notations.**</u> For convenience, we will write $T(x)$ as Tx and $S \circ T$ as ST when no confusion arises.

Theorem (Properties of Adjoint Operators). *If X, Y, Z are normed spaces over $\mathbb{K}$, $c_1, c_2 \in \mathbb{K}$, $S \in L(Y, Z)$ and $T, T_1, T_2 \in L(X, Y)$, then*

(a) $\|T^*\| = \|T\|$ *and hence* $T^* \in L(Y^*, X^*)$

(b) $(c_1T_1 + c_2T_2)^* = c_1T_1^* + c_2T_2^*$

(c) $(ST)^* = T^*S^*$ *and for the identity operator* $I \in L(X)$, $I^* = I$

(d) $T^{**} \in L(X^{**}, Y^{**})$ *and identifying X with $i(X) \subseteq X^{**}$, we have* $T^{**}|_X = T$

(e) *if T is invertible, then T^* is also invertible and* $(T^*)^{-1} = (T^{-1})^* \in L(X^*, Y^*)$

(f) *if T^* is invertible, then T is bounded below, hence injective. In case X is a Banach space, T^* invertible implies T invertible and Y complete.*

Proof. (a)
$$\begin{aligned}\|T\| &= \sup\{\|T(x)\| : \|x\| \le 1\} = \sup\{|\langle T(x), y\rangle| : \|x\| \le 1, \|y\| \le 1\}\\ &= \sup\{|\langle x, T^*(y)\rangle| : \|x\| \le 1, \|y\| \le 1\} = \sup\{\|T^*(y)\| : \|y\| \le 1\} = \|T^*\|,\end{aligned}$$
where the second equality is due to $|\langle T(x), y\rangle| \le \|T(x)\| = \lim\limits_{n\to\infty} |\langle T(x), y_n\rangle|$ and similarly for the fourth.

(b) $(c_1T_1 + c_2T_2)^*(y) = y \circ (c_1T_1 + c_2T_2) = c_1 y \circ T_1 + c_2 y \circ T_2 = (c_1T_1^* + c_2T_2^*)(y)$.

(c) $(S \circ T)^*(y) = y \circ (S \circ T) = T^*(y \circ S) = T^* \circ S^*(y)$. $I^*(y)(x) = y(I(x)) = y(x)$ for all $x \in X$. So $I^*(y) = y$.

(d) For $x \in X$, $T^{**}(x) = T^{**}(i_x) = i_x \circ T^* = i_{T(x)} = T(x)$.

(e) Applying (c) to $T \circ T^{-1} = I$ and $T^{-1} \circ T = I$, we get $(T^{-1})^* \circ T^* = I^* = I$ and $T^* \circ (T^{-1})^* = I^* = I$. So $(T^*)^{-1} = (T^{-1})^*$ and it is in $L(X^*, Y^*)$ by the inverse mapping theorem.

(f) By (e), T^* invertible implies T^{**} invertible. Hence T^{**} is bounded below. By (d), $T^{**}|_X = T$ is bounded below, so T is injective.

In case X is a Banach space, by the lower bound theorem, $T(X)$ is complete and hence closed. Suppose $F \in Y^*$ and $F \equiv 0$ on $T(X)$. Then for all $x \in X$, $0 = F(T(x)) = T^*(F(x))$, i.e. $T^*(F) = 0$. Since T^* is invertible (in particular, injective), we get $F = 0$. By part (b) of the Hahn-Banach theorem, we have $Y = \overline{T(X)} = T(X)$. Then Y is complete and T is bijective. By the inverse mapping theorem, $T^{-1} \in L(Y, X)$ and T is invertible. ∎

Theorem (Kernel-Range Relations). *Let X, Y be normed spaces and $T \in L(X, Y)$. Then*
$$\ker T = {}^{\perp}(\operatorname{ran} T^*), \quad \ker T^* = (\operatorname{ran} T)^{\perp}, \quad (\ker T)^{\perp} = \overline{\operatorname{ran} T^*}^{w^*} \quad \textit{and} \quad {}^{\perp}(\ker T^*) = \overline{\operatorname{ran} T}.$$

Proof. By the remarks after the double-perp theorem, we have $M = \{0\} \subseteq Y$ iff $M^{\perp} = Y^*$ and $N = \{0\} \subseteq X^*$ iff ${}^{\perp}N = X$. For the first and second equations, using part (a) of the corollary to the separation theorem and the definition of adjoint,
$$x \in \ker T \;\Leftrightarrow\; 0 = Tx \;\Leftrightarrow\; \forall y \in Y^*,\; 0 = y(Tx) = (T^*y)x \;\Leftrightarrow\; x \in {}^{\perp}T^*(Y^*) = {}^{\perp}(\operatorname{ran} T^*);$$
$$y \in \ker T^* \;\Leftrightarrow\; 0 = T^*y \;\Leftrightarrow\; \forall x \in X,\; 0 = (T^*y)x = y(Tx) \;\Leftrightarrow\; y \in T(X)^{\perp} = (\operatorname{ran} T)^{\perp}.$$

For the third equation, by the first equation, $\ker T = {}^{\perp}(\operatorname{ran} T^*)$. By the double-perp theorem, $(\ker T)^{\perp} = (\operatorname{ran} T^*)^{\perp\perp} = \overline{\operatorname{ran} T^*}^{w^*}$.

For the fourth equation, by the second equation, $\ker T^* = (\operatorname{ran} T)^{\perp}$. By the double-perp theorem, ${}^{\perp}(\ker T^*) = (\operatorname{ran} T)^{\perp\perp} = \overline{\operatorname{ran} T}$. ∎

Corollary 1. *Let X, Y be normed spaces and $T \in L(X, Y)$. Then*

(a) $\ker T = (\ker T)^{\perp\perp}$ *and* $\ker T^* = (\ker T^*)^{\perp\perp}$,

(b) $\operatorname{ran} T$ *is w-dense (or dense) in Y iff T^* is injective,*

(c) $\operatorname{ran} T^*$ *is w^*-dense in X^* iff T is injective.*

Proof. (a) $\ker T$ is norm-closed in X. So $\ker T = \overline{\ker T} = (\ker T)^{\perp\perp}$. Also, $\ker T^* = (\operatorname{ran} T)^\perp$ is w^*-closed in Y^*. So $\ker T^* = \overline{\ker T^*}^{w^*} = (\ker T^*)^{\perp\perp}$.

(b) If T^* is injective, then by the last theorem, $\overline{\operatorname{ran} T} = {}^\perp(\ker T^*) = {}^\perp\{0\} = Y$. Conversely, if $\operatorname{ran} T$ is dense (equivalently, w-dense) in Y, then by (a), $\ker T^* = (\ker T^*)^{\perp\perp} = (\overline{\operatorname{ran} T})^\perp = Y^\perp = \{0\}$.

(c) If T is injective, then by the last theorem, $\overline{\operatorname{ran} T^*}^{w^*} = (\ker T)^\perp = \{0\}^\perp = X^*$. Conversely, if $\operatorname{ran} T^*$ is w^*-dense in X^*, then by (a), $\ker T = (\ker T)^{\perp\perp} = {}^\perp(\overline{\operatorname{ran} T^*}^{w^*}) = {}^\perp(X^*) = \{0\}$. $\square$

Corollary 2. *Let X be a Banach space, Y a normed space and $T \in L(X,Y)$. The following are equivalent:*

(a) *T is invertible,*

(b) *T^* is invertible,*

(c) *$\operatorname{ran} T$ is dense in Y (or T^* is injective) and T is bounded below,*

(d) *T and T^* are both bounded below.*

Proof. (a) $\iff$ (b) is due to properties (e) and (f) of the adjoint operators. (a) $\iff$ (c) and also (a), (b) $\implies$ (d) are due to the lower bound theorem and its remarks. (d) $\implies$ (c) is due to (b) of corollary 1.

Closed Range Theorem. *Let X, Y be Banach spaces and $T \in L(X,Y)$. The following are equivalent.*

(a) *$\operatorname{ran} T$ is norm-closed (or w-closed),*

(b) *$\operatorname{ran} T^*$ is w^*-closed,*

(c) *$\operatorname{ran} T^*$ is norm-closed.*

Proof. Let $X_0 = X/\ker T$ and $Y_0 = \overline{\operatorname{ran} T} \subseteq Y$. By the kernel-range relations, $Y_0^\perp = \ker T^*$. By Hahn-Banach theorem, $Y_0^* = \{F|_{Y_0} : F \in Y^*\}$. The map $T_0 : X_0 \to Y_0$ given by $T_0(x + \ker T) = T(x)$ is well-defined and linear. Also, T_0 is injective and $\operatorname{ran} T_0 = \operatorname{ran} T$. Then, $T_0^* : Y_0^* \to X_0^*$ is given by $T_0^*(F|_{Y_0}) = F|_{Y_0} \circ T_0$ for all $F \in Y^*$. By the duality theorem, we have $Y_0^* \leftrightarrow Y^*/Y_0^\perp = Y^*/\ker T^*$ (with $F|_{Y_0} \leftrightarrow F + \ker T^*$) and $X_0^* \leftrightarrow (\ker T)^\perp$ (with $G \leftrightarrow G \circ \pi$, where $\pi : X \to X/\ker T$ is the quotient map). Under duality correspondence, we can view $T_0^* : Y_0^* = Y^*/\ker T^* \to X_0^* = (\ker T)^\perp$ as given by $T_0^*(F + \ker T^*) = F|_{Y_0} \circ T_0 \circ \pi = F|_{Y_0} \circ T = F \circ T = T^*(F)$, which is well-defined and linear. More importantly, T_0^* is injective and $\operatorname{ran} T_0^* = \operatorname{ran} T^*$.

(a) $\Rightarrow$ (b) Since $\operatorname{ran} T$ is norm-closed, $\operatorname{ran} T_0 = \operatorname{ran} T = \overline{\operatorname{ran} T} = Y_0$. Then T_0 is surjective (hence bijective). By the inverse mapping theorem, T_0 is invertible. So T_0^* is invertible, hence surjective. So $\operatorname{ran} T^* = \operatorname{ran} T_0^* = (\ker T)^\perp$ is w^*-closed in X^*.

(b) $\Rightarrow$ (c) The weak-star topology is a subset of the norm topology in X^*.

(c) $\Rightarrow$ (a) Since T_0^* is injective and $\operatorname{ran} T_0^* = \operatorname{ran} T^*$ is norm-closed in X^*, by the lower bound theorem, T_0^* is bounded below. Hence there is $\delta > 0$ such that $\|T_0^*(u)\| \geq \delta \|u\|$ for all $u \in Y_0^*$.

To show $\operatorname{ran} T$ is norm-closed, it suffices to show T_0 is open (as it would implies T_0 is surjective and hence $\operatorname{ran} T = \operatorname{ran} T_0 = Y_0 = \overline{\operatorname{ran} T}$). Now to show T_0 is open, let U be the open unit ball in X_0. It is enough to show $T_0(U)$ is a neighborhood of 0 in Y_0. Using the lemmas prior to the open mapping theorem, it is further enough to show $\overline{T_0(U)}$ contains $B(0,\delta)$.

Let $v \in Y_0 \setminus \overline{T_0(U)}$. By the separation theorem, there is a $g \in Y_0^*$ such that $\operatorname{Re} g(v) < \inf\{\operatorname{Re} g(T_0(u)) : u \in U\}$. Let $f = -g/\|T_0^* g\|$, then $\|T_0^* f\| = 1$ and $|T_0^* f(u)| = e^{i\theta} T_0^* f(u) = T_0^* f(e^{i\theta} u) = \operatorname{Re} T_0^* f(e^{i\theta} u)$. So $\operatorname{Re} f(v) > \sup\{\operatorname{Re} f(T_0(u)) : u \in U\} = \sup\{\operatorname{Re}(T_0^* f)(u) : u \in U\} = \sup\{|(T_0^* f)(u)| : u \in U\} = \|T_0^* f\| = 1$. As T_0^* is bounded below, we get $1 = \|T_0^* f\| \geq \delta \|f\|$. So $\frac{1}{\delta} \geq \|f\|$. Then $\frac{1}{\delta}\|v\| \geq \|f\|\|v\| \geq |f(v)| \geq \operatorname{Re} f(v) > 1$. We get $\|v\| > \delta$. Hence, $v \in Y_0 \setminus B(0,\delta)$. Therefore, $\overline{T_0(U)} \supseteq B(0,\delta)$, which gives $\operatorname{ran} T$ is norm-closed. $\square$

Corollary. *Let X, Y be Banach spaces and $T \in L(X,Y)$. Then T is surjective iff T^* is bounded below. Similarly, T^* is surjective iff T is bounded below.*

Proof. T is surjective iff $\operatorname{ran} T$ is dense and norm-closed in Y. By corollary 1 and the closed range theorem, this is iff T^* is injective and $\operatorname{ran} T^*$ is norm-closed. By the lower bound theorem, this is iff T^* is bounded below. The second statement can be proved similarly.

Chapter 5. Basic Operator Facts on Banach Spaces.

§1. **Spectrum.** We will study operators in Banach spaces over $\mathbb{C}$ in this chapter. So all vector spaces refered to below when not specified will mean Banach spaces over $\mathbb{C}$. We begin with the observation that for a Banach space X, $L(X) = L(X,X)$ is not only a Banach space, but it also has a continuous multiplication structure.

Definition. A *Banach algebra* is a Banach space with a multiplication such that $\|xy\| \leq \|x\| \cdot \|y\|$ for all x and y in the space. (Note $x_n \to x$ and $y_n \to y$ implies $\|x_n\|$, $\|y_n\|$ bounded and

$$\|x_n y_n - xy\| = \|x_n(y_n - y) + (x_n - x)y\| \leq \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\| \to 0.$$

So multiplication is continuous.) By math induction, we also have $\|x^n\| \leq \|x\|^n$.

Example. Let X, Y, Z be normed spaces, $T \in L(X,Y)$ and $S \in L(Y,Z)$, then $S \circ T \in L(X,Z)$. For every $x \in X$, $\|(S \circ T)(x)\| = \|S(T(x))\| \leq \|S\|\|T(x)\| \leq \|S\|\|T\|\|x\|$ Thus, $\|S \circ T\| \leq \|S\|\|T\|$. In particular, if X is a Banach space, then $L(X)$ is a Banach algebra with composition as multiplication.

As in linear algebra, for an operator $T \in L(X)$, the related operator $T - cI$ is important.

Definitions. Let X be a Banach space over $\mathbb{C}$ and $T \in L(X)$.

(1) The *resolvent set* of T is $\rho(T) = \{c \in \mathbb{C} : T - cI \text{ is invertible}\}$. For $c \in \rho(T)$, the operator $R_c(T) = (cI - T)^{-1}$ is called the *resolvent* of T.

(2) The *spectrum* of T is $\sigma(T) = \{c \in \mathbb{C} : T - cI \text{ is non-invertible}\}$. A common alternative notation is $sp(T)$.

(3) The *point spectrum* of T is the set $\sigma_p(T) = \{c \in \mathbb{C} : \ker(T - cI) \neq \{0\}\}$ of *eigenvalues* of T.

(4) The *approximate point spectrum* of T is $\sigma_{ap}(T) = \{c \in \mathbb{C} : T - cI \text{ is not bounded below}\} = \{c \in \mathbb{C} : \exists\, x_1, x_2, x_3, \ldots \in X, \|x_i\| = 1, (T - cI)(x_i) \to 0\}$ of all *approximate eigenvalues* of T.

(5) The *compression spectrum* of T is $\sigma_{com}(T) = \{c \in \mathbb{C} : \overline{\operatorname{ran}(T - cI)} \neq X\}$.

(6) The *residual spectrum* of T is $\sigma_r(T) = \sigma_{com}(T) \setminus \sigma_p(T) = \{c \in \mathbb{C} : \ker(T - cI) = \{0\}, \overline{\operatorname{ran}(T - cI)} \neq X\}$.

(7) The *continuous spectrum* of T is $\sigma_c(T) = \sigma(T) \setminus (\sigma_p(T) \cup \sigma_{com}(T)) = \{c \in \mathbb{C} : \ker(T - cI) = \{0\},\ \operatorname{ran}(T - cI) \subset \overline{\operatorname{ran}(T - cI)} = X\}$.

Remarks. Since an operator is invertible iff it is injective and surjective (i.e. its range is closed and dense) iff it is bounded below and its range is dense, so $\sigma(T) = \sigma_{ap}(T) \cup \sigma_{com}(T)$. Clearly, $\sigma_p(T) \subseteq \sigma_{ap}(T)$, but $\sigma_p(T) \cap \sigma_{com}(T)$ may not be empty (eg. T has rank 1) so that $\sigma_{ap}(T), \sigma_{com}(T)$ may not be disjoint. To get disjoint decomposition of $\sigma(T)$, we can write $\sigma(T)$ as the union of the pairwise disjoint sets $\sigma_p(T), \sigma_r(T), \sigma_c(T)$.

Theorem. *For every operator $T \in L(X)$, $\sigma(T) = \sigma(T^*)$.*

Proof. This follows easily from the properties (e) and (f) of adjoint operators that $T - cI$ is invertible if and only if $T^* - cI = (T - cI)^*$ is invertible on a Banach space X. ∎

Concerning the spectrum of an operator, we have the following important facts.

Gelfand's Theorem. *For every $T \in L(X)$, $\sigma(T)$ is a nonempty compact set in $\mathbb{C}$.*

Gelfand-Mazur Theorem. *Let $r(T) = \max\{|z| : z \in \sigma(T)\}$. Then*

$$r(T) = \inf\{\|T^m\|^{1/m} : m = 1, 2, 3, \ldots\} = \lim_{m \to \infty} \|T^m\|^{1/m} \leq \|T\|.$$

($r(T)$ is the furthest distance of any point in $\sigma(T)$ from the origin and is called the spectral radius of T.)

Using these theorems, we will look at some examples first. In each example, to find the spectrum, we try to find the norm of the operator first. We use the norm and spectral radius to bound the spectrum. Then we try to find eigenvalues of the operator.

Examples. (1) Define the *backward shift operator* $T : \ell^1 \to \ell^1$ by $T(x_1, x_2, x_3, \ldots) = (x_2, x_3, x_4, \ldots)$. It is easy to see T is linear. Also,

$$\|T(x_1, x_2, x_3, \ldots)\|_1 = \|(x_2, x_3, x_4, \ldots)\|_1 = \sum_{i=2}^{\infty} |x_i| \le \sum_{i=1}^{\infty} |x_i| = \|(x_1, x_2, x_3, \ldots)\|_1.$$

So $\|T\| \le 1$, hence T is continuous. If $x_1 = 0$, then the above inequality becomes an equality. So $\|T\| = 1$. Since $r(T) = \lim_{n\to\infty} \|T^n\|^{1/n} \le \|T\| = 1$, so $\sigma(T)$ is a nonempty compact subset of $\overline{B(0,1)} = \{z \in \mathbb{C} : |z| \le 1\}$. If $|z| < 1$, then $T(1, z, z^2, \ldots) = (z, z^2, z^3, \ldots) = z(1, z, z^2, \ldots)$. So $T - zI$ is not invertible as $(1, z, z^2, \ldots) \in \ker(T - zI)$. Then $B(0,1) = \{z \in \mathbb{C} : |z| < 1\}$ is a subset of $\sigma(T)$. As $\sigma(T)$ is closed, $\sigma(T) = \overline{B(0,1)}$.

Define *forward (or unilateral) shift operator* $S : \ell^\infty \to \ell^\infty$ by $S(y_1, y_2, y_3, \ldots) = (0, y_1, y_2, \ldots)$. $S = T^*$ because from $(\ell^1)^* = \ell^\infty$ under the pairing $\langle (a_1, a_2, a_3, \ldots), (b_1, b_2, b_3, \ldots)\rangle = a_1b_1 + a_2b_2 + a_3b_3 + \cdots$, we have for all $(x_1, x_2, x_3, \ldots) \in \ell^1$,

$$\begin{aligned}
\langle (x_1, x_2, x_3, \ldots), T^*(y_1, y_2, y_3, \ldots)\rangle &= \langle T(x_1, x_2, x_3, \ldots), (y_1, y_2, y_3, \ldots)\rangle \\
&= x_2y_1 + x_3y_2 + x_4y_3 + \cdots \\
&= \langle (x_1, x_2, x_3, \ldots), (0, y_1, y_2, \cdots)\rangle \\
&= \langle (x_1, x_2, x_3, \ldots), S(y_1, y_2, y_3, \cdots)\rangle.
\end{aligned}$$

Now $\|S\| = \|T^*\| = \|T\| = 1$ and $\sigma(S) = \sigma(T^*) = \sigma(T) = \overline{B(0,1)}$.

---

(2) Define the *Volterra operator* $V : C[0,1] \to C[0,1]$ by $(Vf)(x) = \int_0^x f(t)\,dt$. V is linear and

$$\|Vf\|_\infty = \sup_{x\in[0,1]} \Big|\int_0^x f(t)\,dt\Big| \le \int_0^1 |f(t)|\,dt \le \|f\|_\infty.$$

So $\|V\| \le 1$. For $f \equiv 1$, $(Vf)(x) = x$, $\|Vf\|_\infty = 1 = \|f\|_\infty$ and so $\|V\| = 1$. We claim $|(V^nf)(x)| \le \|f\|_\infty \dfrac{x^n}{n!}$ for all $x \in [0,1]$. For $n = 1$, $|(Vf)(x)| = \Big|\int_0^x f(t)\,dt\Big| \le \|f\|_\infty x$. Assuming case n, we have

$$|(V^{n+1}f)(x)| = \Big|\int_0^x (V^nf)(t)\,dt\Big| \le \int_0^x |(V^nf)(t)|\,dt \le \int_0^x \|f\|_\infty \frac{t^n}{n!}\,dt = \|f\|_\infty \frac{x^{n+1}}{(n+1)!}.$$

This implies $\|V^nf\|_\infty \le \|f\|_\infty \dfrac{1}{n!}$. For $f \equiv 1$, we get equality. Hence $\|V^n\| = \dfrac{1}{n!}$. Since $\lim_{n\to\infty} \dfrac{1/(n+1)!}{1/n!} = \lim_{n\to\infty} \dfrac{1}{n+1} = 0$, we get $\lim_{n\to\infty} (\dfrac{1}{n!})^{1/n} = 0$. So $r(V) = 0$ and $\sigma(V) = \{0\}$, but $\ker V = \{0\}$ implies $\sigma_p(V) = \emptyset$.

Remarks. If $\sigma(T) = \{0\}$, then T is called a *quasinilpotent* operators. We can also define $V : L^2[0,1] \to L^2[0,1]$ by $(Vf)(x) = \int_{[0,x]} f\,dm$. Then $\|V\| = \dfrac{2}{\pi}$ and $\sigma(V) = \{0\}$. See [H], problems 186 to 188.

---

(3) For $f \in L^\infty[0,1]$, define the *multiplication operator* $M_f : L^1[0,1] \to L^1[0,1]$ by $M_f(g) = fg$. We will show $\|M_f\| = \|f\|_\infty$. The case $f = 0$ is clear. So we consider $f \ne 0$ in $L^\infty[0,1]$.

Clearly, $\|M_f(g)\|_1 = \displaystyle\int_{[0,1]} |fg|\,dm \le \|f\|_\infty \|g\|_1$. So $\|M_f\| \le \|f\|_\infty$. Conversely, we may think of f as a bounded measurable function on $[0,1]$ (by taking a representative in the equivalence class of $f \in L^\infty[0,1]$). Let $A_n = \{x \in [0,1] : |f(x)| > \|f\|_\infty - \frac{1}{n}\}$ and $g_n = \dfrac{\chi_{A_n}}{m(A_n)}$. Then $\|g_n\|_1 = 1$ and

$$\|f\|_\infty - \frac{1}{n} \le \frac{1}{m(A_n)} \int_{A_n} |f|\,dm = \int_{[0,1]} |fg_n|\,dm \le \|f\|_\infty \|g_n\|_1 = \|f\|_\infty.$$

So $\|M_f(g_n)\|_1 = \displaystyle\int_{[0,1]} |fg_n|\,dm \to \|f\|_\infty$ as $n \to \infty$. Therefore, $\|M_f\| = \|f\|_\infty$.

For $\sigma(M_f)$, consider the *essential range* of f, which is $S = \{z \in \mathbb{C} : m\big(f^{-1}(B(z,r)\big) > 0 \text{ for all } r > 0\}$. If $z \in S$, then let $D_n = f^{-1}(B(z, \frac{1}{n}))$ and $h_n = \dfrac{\chi_{D_n}}{m(D_n)}$. Then $\|h_n\|_1 = 1$ and

$$\|(M_f - zI)h_n\|_1 = \int_{[0,1]} |f-z||h_n|\,dm = \frac{1}{m(D_n)} \int_{D_n} |f-z|\,dm \le \frac{1}{n}.$$

Assume $M_f - zI$ has an inverse L, then $1 = \|h_n\|_1 = \|L(M - zI)h_n\|_1 \le \|L\|\|(M-zI)h_n\|_1 \le \|L\|\frac{1}{n}$, which implies $n \le \|L\|$ for all n, a contradiction. So $S \subseteq \sigma(M_f)$.

Conversely, if $z \notin S$, then there is $r > 0$ such that $m(f^{-1}(B(z,r)) = 0$. On $[0,1] \setminus f^{-1}(B(z,r))$, define $g(x) = \dfrac{1}{f(x) - z}$ and on $f^{-1}(B(z,r))$, define $g(x) = 0$. Then g is measurable on $[0,1]$ and $\|g\|_\infty \le \frac{1}{r}$. So $M_g(M_f - zI)(h) = h = (M_f - zI)M_g(h)$ almost everywhere. Then $M_f - zI$ is invertible. Hence $\sigma(M_f) = S$.

M_f may also be defined on $L^p[0,1]$, $1 \le p < \infty$, by $M_f(g) = fg$. The norm and spectrum are the same as in the $L^1[0,1]$ case. Finally, $M_f^* : L^q[0,1] \to L^q[0,1]$ is the same as M_f because

$$\langle g, M_f^*(h)\rangle = \langle M_f(g), h\rangle = \int_{[0,1]} (fg)h\,dm = \int_{[0,1]} g(fh)\,dm = \langle g, fh\rangle = \langle g, M_f(h)\rangle.$$

Now we present the proofs of the Gelfand and Gelfand-Mazur theorems. First we need some facts.

Lemma on Inverses. (1) *If $T \in L(X)$ is invertible and $S \in L(X)$ such that $\|S\| < \|T^{-1}\|^{-1}$, then $T - S$ is invertible. So the set of invertible operators in $L(X)$ is an open set.*

(2) *The map $T \mapsto T^{-1}$ on the set of invertible operators is continuous.*

Proof. (1) Let $R = T^{-1}S$, then $\|R\| \le \|T^{-1}\|\|S\| < 1$ and $\displaystyle\sum_{i=0}^\infty R^i$ converges absolutely in $L(X)$. The sum is easily checked to be $(I - R)^{-1}$. Then $T - S = T(I - R)$ is invertible.

(2) For T invertible and $\|S\| < \|T^{-1}\|^{-1}$, let $R = T^{-1}S$. As $\|S\| \to 0$, $\|R\| \le \|T^{-1}\|\|S\| \to 0$, which implies $\|(T-S)^{-1} - T^{-1}\| = \Big\|\big((I-R)^{-1} - I\big)T^{-1}\Big\| \le \Big\|\displaystyle\sum_{i=1}^\infty R^i\Big\|\|T^{-1}\| \le \dfrac{\|R\|}{1 - \|R\|}\|T^{-1}\| \to 0.$ ∎

Resolvent Identity. $R_a(T) - R_b(T) = (b-a)R_a(T)R_b(T)$. As $a \to b$, $R_a(T) \to R_b(T)$ in norm topology.

Proof. Let $A = aI - T = R_a(T)^{-1}$ and $B = bI - T = R_b(T)^{-1}$, then $B - A = (b-a)I$ and $A^{-1} - B^{-1} = A^{-1}BB^{-1} - A^{-1}AB^{-1} = A^{-1}(B-A)B^{-1} = (b-a)A^{-1}B^{-1}$, which is the identity. As $a \to b$, $A \to B$ (since $\|A - B\| = |a - b|$), hence $A^{-1} = R_a(T) \to B^{-1} = R_b(T)$ by part (2) of the lemma on inverses. ∎

Remarks. Two operators T_0 and T_1 are said to *commute* iff $T_0T_1 = T_1T_0$. The resolvent identity implies $R_a(T)$ and $R_b(T)$ commute since $R_a(T)R_b(T) = \dfrac{R_a(T) - R_b(T)}{b - a} = R_b(T)R_a(T)$. Also, $\displaystyle\lim_{a\to b} \frac{R_a(T) - R_b(T)}{a - b} = -R_b(T)^2$, the limit being taken in the norm of $L(X)$.

Lemma 1. *For every $T \in L(X)$, $\sigma(T) \subseteq \{z : |z| \le r\}$, where $r = \limsup_{n\to\infty} \|T^n\|^{1/n} < \infty$, i.e. $r(T) \le r$. Also, for $|z| > r$, $(T - zI)^{-1} = -\sum_{n=0}^{\infty} z^{-n-1}T^n$. For $r = 0$, we interpret $\overline{B(0,r)}$ as $\{0\}$.*

Proof. Note $\|T^n\| \le \|T\|^n$ implies that $r = \limsup_{n\to\infty} \|T^n\|^{1/n} \le \|T\| < \infty$. For $|z| > r$, there is $\varepsilon > 0$ such that $|z| > r + \varepsilon$. By properties of limsup, we see that $\|T^n\|^{1/n} \le r + \varepsilon$ for all except finitely many n. Since $\|z^{-n-1}T^n\| = \underbrace{\frac{\|T^n\|}{(r+\varepsilon)^{n+1}}}_{bounded} \cdot \underbrace{\left(\frac{r+\varepsilon}{|z|}\right)^{n+1}}_{geometric}$ and $\frac{r+\varepsilon}{|z|} < 1$, so $S = -\sum_{n=0}^{\infty} z^{-n-1}T^n$ converges absolutely in $L(X)$.

For $|z| > r$, both $S(T - zI)$ and $(T - zI)S$ equal $-\sum_{n=0}^{\infty} z^{-n-1}T^{n+1} + \sum_{n=0}^{\infty} z^{-n}T^n = I$. So $T - zI$ is invertible, i.e. $z \in \rho(T) = \mathbb{C} \setminus \sigma(T)$. Hence, $\sigma(T) \subseteq \{z : |z| \le r\}$.

Lemma 2. *Let Ω be a nonempty open subset of $\mathbb{C}$ contained in $\rho(T)$. For $f \in L(X)^*$, the function $g : \Omega \to \mathbb{C}$ defined by $g(z) = f((T - zI)^{-1}) = -f(R_z(T))$ is holomorphic with derivative $g'(z) = f(R_z(T)^2)$.*

Proof. This follows from the continuity of f and the remark below the resolvent identity.

Proof of Gelfand's Theorem. By lemma 1, $\sigma(T)$ is bounded in $\mathbb{C}$.

Next we show $\sigma(T)$ is closed by showing $\rho(T) = \mathbb{C}\setminus\sigma(T)$ is open. Let $z \in \rho(T)$. Then $T - zI$ is invertible. By the lemma on inverses, we get $T - wI = (T - zI) - (w - z)I$ is also invertible if $|w - z| < \|(T - zI)^{-1}\|^{-1}$. Then $B(z, \|(T - zI)^{-1}\|^{-1}) \subseteq \rho(T)$. So $\rho(T)$ is open and $\sigma(T) = \mathbb{C} \setminus \rho(T)$ is closed.

Finally, we show $\sigma(T) \neq \emptyset$. Assume $\sigma(T) = \emptyset$. Let $\Omega = \rho(T) = \mathbb{C}$. Then the g function in lemma 2 is entire. By lemmas 1 and 2, for $|z| > \|T\| \ge r$,

$$|g(z)| \le \|f\|\|(T - zI)^{-1}\| \le \|f\| \sum_{n=0}^{\infty} |z|^{-n-1}\|T\|^n = \frac{\|f\|}{|z| - \|T\|} \to 0 \quad \text{as } z \to \infty.$$

Hence, $g(z)$ is bounded. By Liouville's theorem, $f((T-zI)^{-1}) = g(z) = 0$. Then $\|(T-zI)^{-1}\| = \sup\{|f((T-zI)^{-1})| : f \in L(X)^*,\ \|f\| \le 1\} = 0$, which is absurd. So $\sigma(T) \neq \emptyset$. ∎

Proof of the Gelfand-Mazur Theorem. By lemma 1, $r(T) \le r = \limsup_{n\to\infty} \|T^n\|^{1/n}$. Now, to reverse this inequality, it suffices to show there is a $z \in \sigma(T)$ with $|z| = r$, which implies $r \le r(T)$. If $r = 0$, then $\emptyset \neq \sigma(T) \subseteq \{z : |z| \le r\}$ implies $\sigma(T) = \{0\}$.

Next we consider $r > 0$. Assume $\sigma(T) \cap \{z : |z| = r\} = \emptyset$. Then there exists R such that $r(T) = \max\{|z| : z \in \sigma(T)\} < R < r$. So $\sigma(T) \subseteq \{z : |z| \le r(T)\}$. For all $f \in L(X)^*$, by lemma 2, $g(z) = f((T - zI)^{-1})$ is holomorphic on $\rho(T) \supseteq \{z : |z| > r(T)\}$. By lemma 1, $g(z) = f((T - zI)^{-1}) = -\sum_{n=0}^{\infty} f(T^n)z^{-n-1}$ on $\{z : |z| > r\}$, hence also on $\{z : |z| > r(T)\}$ by the uniqueness of Laurent series on annulus. Then it converges absolutely on $|z| = R$. So $\sup\{|f(T^n)/R^{n+1}| : n = 0, 1, 2, \ldots\} < \infty$. By the uniform boundedness principle, we get $c = \sup\{\|T^n/R^{n+1}\| : n = 0, 1, 2, \ldots\} < \infty$. Hence, $\|T^n\| \le cR^{n+1}$. Then $\|T^n\|^{1/n} \le c^{1/n}R^{1+1/n}$. Taking limsup, we get $r \le R$, a contradiction.

Next we show $r = \inf\{\|T^m\|^{1/m} : m = 1, 2, 3, \ldots\}$. For positive integers m, n, we have $n = qm + k$ with $k = 0, 1, \ldots, m - 1$. Then $\|T^n\| \le \|T^m\|^q\|T\|^k$. So $\|T^n\|^{1/n} \le \|T^m\|^{q/n}\|T\|^{k/n}$. Fix m and let $n \to \infty$, since $1 = m(q/n) + (k/n)$, we get $k/n \to 0$ and $q/n \to 1/m$. So $r = \limsup_{n\to\infty} \|T^n\|^{1/n} \le \|T^m\|^{1/m}$. Taking infimum over m, we get the result $r \le \inf\{\|T^m\|^{1/m} : m = 1, 2, 3, \ldots\} \le \liminf_{m\to\infty} \|T^m\|^{1/m} \le \limsup_{m\to\infty} \|T^m\|^{1/m} = r$. ∎

§2. **Projections and Complemented Subspaces.** In the literature, vector subspaces are sometimes called linear manifolds. For convenience, below the term "subspaces" will mean *closed* vector subspaces of Banach spaces.

Definition. A subspace E of a Banach space X is *complemented* iff there is a subspace F of X such that $E \cap F = \{0\}$ and $E + F = X$. Such F is called a *complementary subspace for E*. (In algebra, we write $X = E \oplus F$ and call it an *internal direct sum*.)

Remarks. (1) In the definition, if $x = y + z = y' + z'$ for $y, y' \in E$ and $z, z' \in F$, then $y - y' = z' - z \in E \cap F = \{0\}$ implies $y = y'$ and $z = z'$. So every x has a unique representation as $y + z$ with $y \in E, z \in F$.

(2) We have $\dim F = \operatorname{codim} E$ (i.e. $\dim X/E$) since if B is a basis of F, then $\pi(B)$ is a basis of X/E, where $\pi : X \to X/E$ is the quotient map.

Examples. (1) If $\dim E = n < \infty$, then E is complemented. (To see this, let $\{x_1, \ldots, x_n\}$ be a basis of E. By the Hahn-Banach theorem, for $i = 1, \ldots, n$, there is $f_i \in X^*$ such that $f_i(x_i) = 1$ and $f_i(x_j) = 0$ for $i \neq j$. Let $F = \bigcap_{i=1}^{n} \ker f_i$. If $e = c_1x_1 + \cdots + c_nx_n \in E$ is in F, then $c_i = f_i(e) = 0$ for $i = 1, \ldots, n$, i.e. $E \cap F = \{0\}$. For $x \in X$, we have $y = f_1(x)x_1 + \cdots + f_n(x)x_n \in E$ and $z = x - y \in F$ because $f_i(z) = f_i(x) - f_i(y) = f_i(x) - f_i(x) = 0$ for $i = 1, \ldots, n$, i.e. $x = y + z \in E + F$. So F is a complementary subspace of E.)

(2) If $\operatorname{codim} E < \infty$, then E is complemented. (To see this, suppose $\dim(X/E) = n < \infty$. Let $\{x_1 + E, \ldots, x_n + E\}$ be a basis of X/E. Then $b_1x_1 + \cdots + b_nx_n = 0$ implies $b_1(x_1 + E) + \cdots + b_n(x_n + E) = 0 + E$. It follows $\{x_1, \ldots, x_n\}$ is linearly independent. Let F be the linear span of $\{x_1, \ldots, x_n\}$. Then $\dim F = n < \infty$. So F is complete, hence closed. If $c_1x_1 + \cdots + c_nx_n \in F$ is in E, then $c_1(x_1 + E) + \cdots + c_n(x_n + E) = (c_1x_1 + \cdots + c_nx_n) + E = E$, which implies $c_i = 0$ for $i = 1, \ldots, n$. So $E \cap F = \{0\}$. For $x \in X$, $x + E$ can be written as $a_1(x_1 + E) + \cdots + a_n(x_n + E) = z + E$ in X/E, where $z = a_1x_1 + \cdots + a_nx_n \in F$. Then $x + E = z + E$ implies $y = x - z \in E$. So $x = y + z \in E + F$. Hence F is a complementary subspace of E.)

(3) Every subspace M in a Hilbert space H is complemented by its orthogonal complement $M^\perp$, i.e. we have $H = M \oplus M^\perp$. (In 1971, Lindenstrauss and Tzafriri proved the converse, namely if every subspace of a Banach space is complemented, then the Banach space is isomorphic to a Hilbert space.)

(4) c_0 is uncomplemented in ℓ^∞. See [M], pp. 301-302.

(5) In $L^p = L^p(-\pi, \pi]$, let H^p be the closed linear span of $e^{in\theta}$ $(n \geq 0)$. M. Riesz proved that for $1 < p < \infty$, H^p is complemented in L^p by the closed linear span of $e^{in\theta}$ $(n < 0)$. D. J. Newman proved that H^1 is uncomplemented in L^1. R. Arens and P. C. Curtis proved that H^∞ is uncomplemented in L^∞.

Definition. An operator $P \in L(X)$ is a *projection* iff $P^2 = P$, i.e. $P|_{\operatorname{ran} P} = I|_{\operatorname{ran} P}$.

Remarks. If P is a projection, then $Q = I - P$ is a projection since $(I - P)^2 = I - 2P + P^2 = I - P$. Also, $\ker P = \operatorname{ran}(I - P)$ since $Px = 0$ imples $x = x - Px = (I - P)x$ and conversely, $P((I - P)x) = Px - P^2x = 0$. Similarly, $\operatorname{ran} P = \operatorname{ran}(I - Q) = \ker Q = \ker(I - P)$. So $\operatorname{ran} P$ is always closed.

Theorem. *If P is a projection, then $\operatorname{ran} P$ and $\ker P$ complement each other, i.e. $X = \operatorname{ran} P \oplus \ker P$.*

Proof. Since $\ker P = \operatorname{ran}(I - P)$, $x = Px + (I - P)x$ and $x \in (\operatorname{ran} P) \cap (\ker P)$ implies $x = Px = 0$, we get $X = \operatorname{ran} P \oplus \ker P$. ∎

Theorem. *A subspace E of X is complemented iff $E = \operatorname{ran} P$ for some projection $P \in L(X)$.*

Proof. The if direction follows from the last theorem. For the only-if direction, let F be a complementary subspace of E. Then each $x \in X$ can be written as $x = y + z$ for some unique $y \in E$ and $z \in F$. Define $Px = y$. Then P is linear by uniqueness of representation. Now $\operatorname{ran} P = E$ since for every $y \in E$, $y = y + 0$ in X implies $Py = y$. Also, $P^2x = Py = y = Px$, i.e. $P^2 = P$.

For continuity, consider the graph of P. If $(x_n, Px_n) \to (x, y)$, then write $x_n = y_n + z_n$, where $y_n \in E$ and $z_n \in F$. As E is closed, $y_n \in E$, $y_n = Px_n \to y$, so $y \in E$. As F is closed, $z_n \in F$, $z_n = x_n - y_n \to x - y$, so $x - y \in F$ As $x = y + (x - y)$, we get $y = Px$. By the closed graph theorem, P is continuous. $\square$

Corollary. *If E is complemented in X, then $E^\perp$ is complemented in X^*.*

Proof. Let $P \in L(X)$ be a projection with $\operatorname{ran} P = E$, then $(P^*)^2 = P^*P^* = (PP)^* = P^*$, i.e. $P^* \in L(X^*)$ is a projection and $E^\perp = (\operatorname{ran} P)^\perp = \ker P^*$ is closed and complemented by $\operatorname{ran} P^*$ in X^*. $\square$

Left Inverse Theorem. *$T \in L(X, Y)$ is left invertible (i.e. there is $S \in L(Y, X)$ such that $ST = I$) iff T is injective and $\operatorname{ran} T$ is closed and complemented in Y (iff T is bounded below and $\operatorname{ran} T$ is complemented in Y by the lower bound theorem).*

Proof. For the if direction, T is injective. Let $P \in L(Y)$ be the projection onto $\operatorname{ran} T$, then $T_0 = P \circ T : X \to \operatorname{ran} T$ is bijective (since $\operatorname{ran} T = \operatorname{ran} P$ makes $T_0(x) = T(x)$). Let $S = T_0^{-1} \circ P$, then $ST = I$.

For the only-if direction, if $S \in L(Y, X)$ is such that $ST = I$, then T is injective and $(TS)^2 = TSTS = TS$ is the projection with $\operatorname{ran} TS = \operatorname{ran} T$ (since $\operatorname{ran} TS \subseteq \operatorname{ran} T = \operatorname{ran} TST \subseteq \operatorname{ran} TS$) so that $\operatorname{ran} T$ (being the range of a projection) is closed and complemented. $\square$

Exercise. Prove that $T \in L(X, Y)$ is right-invertible (i.e. there is $S \in L(Y, X)$ such that $TS = I$) iff T is surjective and $\ker T$ is complemented. (*Hint:* (if) let $\operatorname{ran} Q$ be complement to $\ker T$, then $S = \widehat{Q} \circ \widehat{T}^{-1}$, where $X/\ker T$ is considered; (only-if) check ST is a projection and $\operatorname{ran} ST = \operatorname{ran} S$ is complement to $\ker T$.)

---

§3. **Compact Operators.** Finite rank operators (i.e. operators whose ranges are finite dimensional) are easy to understand by using linear algebra. In this section, we will study a class of operators related to the finite rank operators. First we recall the following facts:

(1) For any normed vector space V, if the closed unit ball of V is compact, then $\dim V < \infty$.

(2) (*Metric Compactness Theorem*) In a metric space M, a set S in M is compact iff S is sequentially compact (i.e. every sequence in S has a convergent subsequence with limit in S) iff S is complete and totally bounded (i.e. for every $\varepsilon > 0$, there are $x_1, \ldots, x_n \in S$ such that $B(x_1, \varepsilon) \cup \cdots \cup B(x_n, \varepsilon) \supseteq S$ and we say S has an ε-dense set $\{x_1, x_2, \ldots, x_n\}$). S is totally bounded implies S is separable (by taking $\varepsilon = 1/k$ and union of all centers over all $k \in \mathbb{N}$ gives a countable dense set). It is easy to check that if either S or $\overline{S}$ has an $\varepsilon/2$-dense set, then the other has an ε-dense set. Hence, S is totally bounded if and only if $\overline{S}$ is totally bounded.

(3) (*Arzela-Ascoli Theorem*) For a compact set M, a set S in $C(M, \mathbb{K})$ is (sequentially) compact iff S is closed, bounded and equicontinuous in $C(M, \mathbb{K})$, where equicontinuity means for every $\varepsilon > 0$, there is a $\delta > 0$ such that for all $f \in S$ and for all $x, y \in M$, $d(x, y) < \delta$ implies $|f(x) - f(y)| < \varepsilon$.

Definition. Let X, Y be Banach spaces and B be the open unit ball of X. A linear function $K : X \to Y$ is ***compact*** iff $K(B)$ is precompact, i.e. $\overline{K(B)}$ is compact, in Y (hence $K(B)$ bounded, K bounded). (By the metric compactness theorem, this is equivalent to the condition that for every bounded sequence $\{x_n\}$ in X, the sequence $\{K(x_n)\}$ has a convergent subsequence in Y or to $K(B)$ is totally bounded).

Remark. Since $\overline{K(B)}$ is compact, it cannot contain any closed ball (which is never compact) in infinite dimensional spaces. So compact operators are considered "small" operators.

Theorem (Properties of Compact Operators). *Let X, Y, Z be Banach spaces.*

(a) *Finite rank operators $F \in L(X, Y)$ (i.e. $\dim \operatorname{ran} F < \infty$) are compact. If $K \in L(X, Y)$ is compact, then $\operatorname{ran} K$ contains no infinite dimensional closed subspaces of Y. In particular, if $\operatorname{ran} K$ is closed in Y, then K has finite rank.*

(b) *If K_1, K_2 are compact and $c \in \mathbb{C}$, then $K_1 + cK_2$ is compact.*

(c) *If $K \in L(X,Y)$ is compact and $T \in L(Y,Z)$, then TK is compact.*

(d) *If $K \in L(Y,Z)$ is compact and $T \in L(X,Y)$, then KT is compact.*

(e) *If $K \in L(X,Y)$ is compact and invertible, then* $\dim X = \dim Y < \infty$.

(f) *The restriction $K|_V$ of a compact operator $K \in L(X,Y)$ to a closed subspace V of X is compact.*

(g) *If $K \in L(X,Y)$ is compact, then* $\operatorname{ran} K$ *is separable.*

(h) *If for $n = 1,2,3,\ldots,$ $K_n \in L(X,Y)$ is compact and K_n converges to K, then K is compact.*

(i) *$K \in L(X,Y)$ is compact iff $K^* \in L(Y^*, X^*)$ is compact.*

Remarks. (1) In the case $X = Y = Z$, parts (b), (c), (d), (h) imply the set of all compact operators is a closed two-sided ideal of $L(X)$.

(2) Part (i) of the theorem is called Schauder's theorem in some literatures.

Examples. (1) Let $X = Y = \ell^p$ $(1 \le p \le \infty)$. For $a = (a_1, a_2, a_3, \ldots) \in c_0$, define $K(x_1, x_2, x_3, \ldots)$= $(a_1x_1, a_2x_2, a_3x_3, \ldots)$ and $K_n(x_1, x_2, x_3, \ldots)$=$(a_1x_1, a_2x_2, \ldots, a_nx_n, 0, 0, \ldots)$. Then $\|K\|, \|K_n\| \le \|a\|_\infty$. Now K_n is finite rank, hence compact. Then $\|K - K_n\| \le \sup\{|a_j| : j > n\} \to 0$ as $\limsup\limits_{n\to\infty} |a_n| = \lim\limits_{n\to\infty} |a_n| = 0$. By property (h), K is compact.

(2) Let $X = Y = C([0,1])$ and $G \in C([0,1]^2)$. Define $(Kf)(x) = \displaystyle\int_0^1 G(x,y)f(y)\,dy$. This is called the *Fredholm integral operator.* Note that $K \in L(X)$ and $\|K\| \le \|G\|_\infty$. If $G(x,y) = F(x)H(y)$ for some $F, H \in C([0,1])$, then K has at most rank 1. Similarly, if $G(x,y) = \displaystyle\sum_{j=1}^n F_j(x)H_j(y)$, then K has finite rank. By the Stone-Weierstrass theorem, we can approximate $G \in C([0,1]^2)$ uniformly by functions of the form $\displaystyle\sum_{j=1}^n F_j(x)H_j(y)$. So we can approximate K by finite rank operators. Therefore, by (a) and (h), K is compact.

(3) Let $X = Y = L^2([0,1])$ and $G \in L^2([0,1]^2)$. Define K as in (2). Then $K \in L(X)$ and $\|K\| \le \|G\|_2$ since

$$\sqrt{\int_{[0,1]} \Big|\int_{[0,1]} G(x,y)f(y)\,dy\Big|^2\,dx} \le \sqrt{\int_{[0,1]} \Big(\int_{[0,1]} |G(x,y)|^2\,dy\Big)\Big(\int_{[0,1]} |f(y)|^2\,dy\Big)\,dx} = \|G\|_2\|f\|_2.$$

By the reasoning above, K is compact (as continuous functions are dense in L^2) by (h).

(4) Let $X = C^1([0,1])$ be the set of functions with continuous derivatives on $[0,1]$. For $f \in C^1([0,1])$, let $\|f\|_{C^1([0,1])} = \|f\|_\infty + \|f'\|_\infty$. This is a complete norm by properties of uniform convergence. So $C^1([0,1])$ is a Banach space. Let $Y = C([0,1])$ and $K : X \to Y$ be the inclusion map $K(f) = f$. Then K is compact by the Arzela-Ascoli theorem because for all $f \in \overline{B}$, $\|f\|_{C^1([0,1])} \le 1$ implies $\|f\|_\infty \le 1$ (hence $\overline{K(B)}$ bounded in $C([0,1])$) and $\|f'\|_\infty \le 1$ (hence $\overline{K(B)}$ is equicontinuous in $C([0,1])$ by the mean-value theorem).

Proof of Properties of Compact Operators. Let B and B' denote the open unit balls of X and Y respectively.

(a) For the first statement, $\dim \operatorname{ran} F < \infty$ implies $\operatorname{ran} F$ closed. Also, $\overline{F(B)} \subseteq \overline{B(0, \|F\|)}$ implies $\overline{F(B)}$ is bounded. Hence $\overline{F(B)}$ is compact. For the second statement, let Z be a closed subspace of Y in $\operatorname{ran} K$, then $W = K^{-1}(Z)$ is closed in X. So $K|_W : W \to Z$ is surjective. By the open mapping theorem, $K|_W$ sends the open unit ball B_W of W to an open neighborhood $K(B_W)$ of 0 in Z. Then $\overline{K(B_W)}$, being a closed subset of $\overline{K(B)}$, is a compact neighborhood of 0 in Z. Hence, $\overline{K(B_W)}$ contains some compact $\overline{B(0,r)}$ of Z, which implies Z is finite dimensional.

(b) $K_1 + cK_2$ compact follows from $(K_1 + cK_2)(B) \subseteq \overline{K_1(B)} + c\overline{K_2(B)}$, which is compact as it is the image of $\overline{K_1(B)} \times \overline{K_2(B)}$ under the continuous function $g(x,y) = x + cy$.

(c) TK compact follows from $TK(B) \subseteq T(\overline{K(B)})$, which is compact.

(d) KT compact follows from $KT(B) \subseteq K(\|T\|B') \subseteq \|T\|\overline{K(B')}$, which is compact.

(e) By (c) and (d), $K^{-1}K = I$ and $KK^{-1} = I$ are compact and hence the closed unit balls of X and Y are compact. Then X, Y are finite dimensional. K invertible implies the dimensions are the same.

(f) $K|_V$ compact follows from $K|_V(B \cap V) \subseteq \overline{K(B)}$, which is compact.

(g) This follows from $K(B)$ totally bounded, hence separable, and $\operatorname{ran} K = \bigcup_{n=1}^{\infty} nK(B)$.

(h) To show $\overline{K(B)}$ compact, it is enough to show $K(B)$ is totally bounded. For $\varepsilon > 0$, take n with $||K_n - K\| < \varepsilon/3$. Since $\overline{K_n(B)}$ is compact, it is totally bounded. So there is a finite set $\{x_1, \ldots, x_m\} \subseteq B$ such that $\{K_n(x_1), \ldots, K_n(x_m)\}$ is $(\varepsilon/3)$-dense in $K_n(B)$. Hence, for every $y \in B$, there is j with $\|K_n(y) - K_n(x_j)\| < \varepsilon/3$, so

$$\|K(y) - K(x_j)\| \leq \|K(y) - K_n(y)\| + \|K_n(y) - K_n(x_j)\| + \|K_n(x_j) - K(x_j)\| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Hence, $\{K(x_1), \ldots, K(x_m)\}$ is ε-dense in $K(B)$. Therefore, $K(B)$ is totally bounded.

(i) Let K be compact and U be the closed unit ball of X, then $\overline{K(U)} = \overline{K(B)}$ is compact in Y. Let $\{y_n\}$ be a sequence in Y^* with $\|y_n\| \leq 1$. Since for every $x, z \in \overline{K(U)}$, $|y_n(x) - y_n(z)| \leq \|y_n\|\|x - z\| \leq \|x - z\|$, the functions y_n are equicontinuous in $C(\overline{K(U)}, \mathbb{K})$. By the Arzela-Ascoli theorem, there is a subsequence $\{y_{n_i}\}$ convergent in $C(\overline{K(U)}, \mathbb{K})$. Since $K^* y_{n_i} = y_{n_i} \circ K$, the sequence $\{K^* y_{n_i}\}$ converges uniformly on U. Since norm of T in X^* is sup-norm of T on U, $K^* y_{n_i}$ converges in X^*. Hence K^* is compact.

Conversely, K^* compact implies K^{**} is compact, which implies $K = K^{**}|_X$ is compact. ∎

From property (h), we know the limit of finite rank operators is compact. This raised the question of whether compact operators are always limit of finite rank operators or not. In the case $Y = X$ is a Hilbert space, it is true and will be shown in the next chapter. Below we will prove it for a separable Hilbert space with the help of the following theorem.

Theorem. *If $K \in L(X, Y)$ is compact, then K is completely continuous, i.e. for every $\{x_n\}$ w-converges to x in X, Kx_n norm-converges to Kx in Y. For reflexive X, the converse is true.*

Proof. For the first statement, assume Kx_n does not converge to Kx. Then there are $\varepsilon > 0$ and subsequence $\{x_{n_k}\}$ such that $\|Kx_{n_k} - Kx\| \geq \varepsilon$. Since $\{x_{n_k}\}$ w-converges to x, by the uniform boundedness principle, $\{x_{n_k}\}$ is bounded. By compactness of K, there is a subsequence $x_{n_{k_j}}$ such that $Kx_{n_{k_j}}$ norm-converges (hence also w-converges) to some z. Since $\|z - Kx\| = \lim_{j\to\infty} \|Kx_{n_{k_j}} - Kx\| \geq \varepsilon$, $z \neq Kx$. Since $x_n \xrightarrow{w} x$, for every $f \in Y^*$, we have $K^*(f) \in X^*$ and $f(Kx_{n_{k_j}} - Kx) = K^*(f)(x_{n_{k_j}} - x) \to 0$, i.e. $Kx_{n_{k_j}}$ w-converges to Kx. This leads to $Kx = z$, a contradiction.

For the second statement, since X is reflexive, if K is completely continuous, then for every bounded sequence $\{x_n\}$ in X, by the Eberlein-Smulian theorem, there is a subsequence $\{x_{n_k}\}$ w-converges to some y. Then $\{Kx_{n_k}\}$ converges to Ky by complete continuity of K. Therefore, K is compact. ∎

Theorem. *Let H be a separable Hilbert space and $K \in L(H)$ be a compact operator. Then K is the limit of a sequence of finite rank operators in $L(H)$ under the norm topology.*

Proof. For K with finite rank, take every term to be K. For compact K, not finite rank, by property (g) of compact operators, $\operatorname{ran} K$ is separable. Let $\{y_1, y_2, y_3, \ldots\}$ be an orthonormal basis of $\overline{\operatorname{ran} K}$ and $P_n x = \sum_{j=1}^{n} (x, y_j) y_j$ be the projection onto $\overline{\operatorname{span}\{y_1, \ldots, y_n\}}$. Then $\|P_n\| = 1 = \|I - P_n\|$. For $1 \leq m \leq n$, $\operatorname{ran} P_n \supseteq \operatorname{ran} P_m$ implies $P_n P_m = P_m$ and so $(I - P_n)(I - P_m) = I - P_n - P_m + P_n P_m = I - P_n$. Then

$$\|K - P_n K\| = \|(I - P_n)K\| = \|(I - P_n)(I - P_m)K\| \leq \|(I - P_m)K\| = \|K - P_m K\|.$$

Hence $\|K - P_nK\| \to \eta \in [0, +\infty)$. Assume its limit is $\eta > 0$. Then for every n, there is $x_n \in H$ such that $\|x_n\| = 1$ and $\|(I - P_n)Kx_n\| > \eta/2$. By the Eberlein-Smulian theorem, since Hilbert spaces are reflexive, $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ converges weakly to some x. By the last theorem, Kx_{n_k} converges in norm to $u = Kx$. Now P_nu converges to u in norm. Then

$$\eta/2 < \|(I - P_{n_k})Kx_{n_k}\| \le \|(I - P_{n_k})(Kx_{n_k} - u)\| + \|(I - P_{n_k})u\| \le \|Kx_{n_k} - u\| + \|u - P_{n_k}u\| \to 0,$$

a contradiction. Therefore, $\|K - P_nK\| \to \eta = 0$ and P_nK is finite rank. □

Definition. (1) A Banach space Y has the *approximation property* iff for every Banach space X, every compact operator in $L(X, Y)$ is the limit of a sequence of finite rank operators in $L(X, Y)$.

(2) A sequence $\{x_n\}$ in a Banach space Y is a *Schauder basis* of Y iff for every $y \in Y$, there is a unique sequence $\{c_n\}$ of scalars such that $y = \sum_{n=1}^{\infty} c_n x_n$. (Such spaces are clearly separable.)

Remarks. It is known that if Y has a Schauder basis, then Y has the approximation property (see [M], p. 364) and in particular, every compact operator in $L(Y)$ is the limit of a sequence of finite rank operators in $L(Y)$. See [CL], pp. 212-213. In 1932, Banach conjectured that every Banach space Y has the approximation property and further conjectured that every separable Banach space has a Schauder basis. On November 6, 1936, Mazur offered a goose as a prize for a solution of these problems in (problem 153 of) the famous "Scottish book" of open problems kept at the Scottish Coffee House in Lwów, Poland by Banach, Mazur, Ulam and other mathematicians.

In 1955, A. Grothendieck proved that Y has the approximation property iff for every compact subset W of Y and every $\varepsilon > 0$, there is a finite rank operator $T \in L(Y)$ such that for all $y \in W$, $\|Ty - y\| < \varepsilon$. Thus to check the approximation property, there is no need to involve other Banach spaces X. Separable Hilbert spaces, c_0 and ℓ^p $(1 \le p < \infty)$ have the approximation property.

Finally, in 1971, Swedish mathematician and pianist Per Enflo showed that there is a separable reflexive Banach space Y and a compact operator in $L(Y)$ that is not the limit of any sequence of finite rank operators in $L(Y)$. This refuted both conjectures. About a year after solving the problem, Enflo traveled to Warsaw to give a lecture on his solution, after which he was awarded the goose. Enflo's solution was published in Acta Mathematica, vol. 130 (1973), pp. 309-317.

Next we will look at theorems about compact operators, which are useful for differential equations.

Lemma (Riesz-Fredholm). *If $K \in L(X)$ is compact and $c \ne 0$, then $N = \ker(K - cI)$ is finite dimensional and $M = \operatorname{ran}(K - cI)$ is closed and finite codimensional (i.e. $\operatorname{codim} M = \dim(X/M) < \infty$).*

Proof. For N, by property (f), $K|_N$ is compact. Also, $K|_N = cI$ is invertible. By property (e), N is finite dimensional. Next, $M^\perp = \ker(K^* - cI)$ is finite dimensional by property (i) and last sentence. If we can show M is closed, then $(X/M)^* = M^\perp$ is finite dimensional and hence $\infty > \dim(X/M)^* = \dim(X/M)^{**} \ge \dim(X/M) = \operatorname{codim} M$. Let Z be a complementary subspace of $N = \ker(K - cI)$. Since $Z \cap N = \{0\}$, $S = (K - cI)|_Z : Z \to X$ is injective. To show M is closed, since $M = \operatorname{ran}(K - cI) = \operatorname{ran} S$, by the lower bound theorem, it suffices to show S is bounded below.

Assume S is not bounded below. Then there is $z_n \in Z$, $\|z_n\| = 1$ and $S(z_n) \to 0$. Since K is compact, passing to a subsequence, we may assume $K(z_n) \to w$. Then $z_n = (K - S)(z_n)/c \to w/c$, which is in Z as Z is closed. As $\|z_n\| = 1$, so $\|w\| = |c| \ne 0$. Also, $K(z_n) \to K(w/c)$. By the uniqueness of limit, $w = K(w/c)$. Then $w \in \ker(K - cI) \cap Z = \{0\}$, contradicting $\|w\| \ne 0$. □

Theorem (Riesz-Fredholm). *Let $K \in L(X)$ be compact, $c \ne 0$, $N_i = \ker(K - cI)^i$ and $M_i = \operatorname{ran}(K - cI)^i$.*

(a) *$K(N_i) \subseteq N_i$ and $\dim N_i < \infty$. $N_1 \subseteq N_2 \subseteq N_3 \subseteq \cdots$ and there is a least j such that $N_j = N_{j+1} = N_{j+2} = \cdots$.*

(b) $K(M_i) \subseteq M_i$, *M_i is closed and* $\operatorname{codim} M_i < \infty$. *$M_1 \supseteq M_2 \supseteq M_3 \supseteq \cdots$ and there is a least k such that $M_k = M_{k+1} = M_{k+2} = \cdots$.*

(c) *$j = k$ and $X = M_j \oplus N_j$. Also, $(K - cI)|_{M_j} \in L(M_j)$ is invertible and $(K - cI)|_{N_j} \in L(N_j)$ is nilpotent of index j (i.e. $(K - cI)|_{N_j}^{j-1} \neq 0$, but $(K - cI)|_{N_j}^{j} \equiv 0$.)*

(d) $\dim\ker(K - cI) = \operatorname{codim}\operatorname{ran}(K - cI) = \dim\ker(K^* - cI) = \operatorname{codim}\operatorname{ran}(K^* - cI) < \infty$. *In particular, $K - cI$ is injective iff $K - cI$ is surjective iff $K^* - cI$ is injective iff $K^* - cI$ is surjective.*

Proof. (a) Observe that $z \in N_i$ implies $(K - cI)^i(Kz) = K(K - cI)^i(z) = 0$ (i.e. $Kz \in N_i$). So $K(N_i) \subseteq N_i$.

Next, K compact implies $T = (K - cI)^i - (-c)^i I$ compact. So $N_i = \ker(K - cI)^i = \ker(T + (-c)^i I)$ is finite dimensional by the lemma.

For all $i > 1$, $N_{i-1} \subseteq N_i$ because $(K - cI)^{i-1}(x) = 0$ implies $(K - cI)^i(x) = 0$. Assume $N_{i-1} \subset N_i$ for all $i > 1$. Then $N_i/N_{i-1} \neq \{0\}$. we may pick $x_i \in N_i$ with $\|x_i\| \leq 2$ and $\|x_i + N_{i-1}\| = 1$. (This is possible by taking $x + N_{i-1} \in N_i/N_{i-1}$ with $\|x + N_{i-1}\| = 1$, then there is $y \in N_{i-1}$ such that $\|x + y\| \leq 2$ and we can let $x_i = x + y$, then $x_i + N_{i-1} = x + N_{i-1}$.) If $i < j$, then $x_i \in N_i$ implies $Kx_i \in N_i$ and

$$Kx_j - Kx_i = cx_j + (Kx_j - cx_j) - Kx_i \in cx_j + N_{j-1} + N_i = cx_j + N_{j-1} = c(x_j + N_{j-1}).$$

So $\|Kx_j - Kx_i\| \geq \|c(x_j + N_{j-1})\| = |c| > 0$. Then $\{Kx_i\}$ has no convergent subsequence, contradicting K is compact. Therefore, there is a least j such that $N_j = N_{j+1}$. Since $x \in N_{j+2}$ implies $(K - cI)x \in N_{j+1} = N_j$, which implies $x \in N_{j+1}$, so $N_{j+1} = N_{j+2}$ and so on.

(b) Observe that K compact implies $T = (K - cI)^i - (-c)^i I$ compact and $M_i = \operatorname{ran}(K - cI)^i = \operatorname{ran}(T + (-c)^i I)$ is closed by the lemma. The rest is similar to (a).

(c) To show $j = k$, suppose $a \in N_{k+1}$, i.e. $(K - cI)^{k+1}(a) = 0$. Take $m > 0$ such that $m + k \geq j$. Since $(K - cI)^k(a) \in M_k = M_{m+k}$, we have $(K - cI)^k(a) = (K - cI)^{m+k}(b)$ for some $b \in X$. Since $N_j = \cdots = N_{m+k} = N_{m+k+1}$, so $0 = (K - cI)^{k+1}(a) = (K - cI)^{m+k+1}(b) = (K - cI)^{m+k}(b) = (K - cI)^k(a)$. So $N_{k+1} = N_k$. By minimality of j, we get $j \leq k$ (or more precisely, $j_K \leq k_K$ for every compact K).

For the converse, as $N_i^\perp = (\ker(K - cI)^i)^\perp = \overline{\operatorname{ran}(K^* - cI)^i}^{w*} = \overline{\operatorname{ran}(K^* - cI)^i} = \operatorname{ran}(K^* - cI)^i$ by the closed range theorem, so $N_i \neq N_{i+1}$ for $i < k_{K^*}$. This implies $k_{K^*} \leq j_K$. Similarly, ${}^\perp(\ker(K^* - cI)^i) = \overline{\operatorname{ran}(K - cI)^i} = \operatorname{ran}(K - cI)^i$ implies $k_K \leq j_{K^*}$. So $k_K \leq j_{K^*} \leq k_{K^*} \leq j_K$. Therefore, the j, k for K and K^* are all equal.

Next, we show $X = M_j \oplus N_j$. Let $x \in X$. Since $(K - cI)^j(x) \in M_j = M_{2j}$, $(K - cI)^j(x) = (K - cI)^{2j}(y)$ for some $y \in X$. Write $x = (K - cI)^j(y) + z$. Then $(K - cI)^j(z) = (K - cI)^j(x) - (K - cI)^{2j}(y) = 0$, i.e. $z \in N_j$. So $X = M_j + N_j$. Next for $r \in M_j \cap N_j$, there is $s \in X$ such that $r = (K - cI)^j(s)$ and $0 = (K - cI)^j(r) = (K - cI)^{2j}(s)$. Then $s \in N_{2j} = N_j$. So $r = (K - cI)^j(s) = 0$. Therefore, $X = M_j \oplus N_j$.

Next we show $(K - cI)|_{M_j} : M_j \to M_j$ is injective and surjective. For $x \in \ker(K - cI)|_{M_j}$, there is y such that $x = (K - cI)^j(y) \in M_j$ and $(K - cI)x = 0$. Then $y \in N_{j+1} = N_j$ so that $x = (K - cI)^j(y) = 0$. Hence, $(K - cI)|_{M_j}$ is injective. Also, for $z \in M_j = M_{j+1}$, we have $z = (K - cI)^{j+1}(w) = (K - cI)(K - cI)^j(w)$ for some w and so $z \in \operatorname{ran}(K - cI)|_{M_j}$. Hence, $(K - cI)|_{M_j}$ is surjective. Therefore, $(K - cI)|_{M_j}$ is invertible.

Finally, since $N_{j-1} \subset N_j$, there is $x \in N_j \setminus N_{j-1}$. So $(K - cI)|_{N_j}^{j-1}(x) \neq 0$. By definition of N_j, $(K - cI)|_{N_j}^{j} \equiv 0$. So $(K - cI)|_{N_j}$ is nilpotent of index j.

(d) By (c), $X = M_j \oplus N_j$. For the left equality, we have

$$\infty > \dim\ker(K - cI) = \dim\ker(K - cI)|_{N_j} = \operatorname{codim}\operatorname{ran}(K - cI)|_{N_j} = \operatorname{codim}\operatorname{ran}(K - cI), \qquad (1)$$

where the invertibility of $(K - cI)|_{M_j}$ is used in the first and third equalities and $\dim N_j < \infty$ is used in the second equality. Similarly, $\dim\ker(K^* - cI) = \operatorname{codim}\operatorname{ran}(K^* - cI) < \infty$. For the middle equality, by the kernel-range relations and the duality theorem,

$$\ker(K^* - cI) = \ker(K - cI)^* = (\operatorname{ran}(K - cI))^\perp = (X/\operatorname{ran}(K - cI))^*. \qquad (2)$$

By (1), $\infty > \operatorname{codim} \operatorname{ran}(K - cI) = \dim(X/\operatorname{ran}(K - cI)) = \dim(X/\operatorname{ran}(K - cI))^* = \dim\ker(K^* - cI)$, where the last equality is by (2). $\square$

The following theorem of Riesz and Schauder on the spectrums of compact operators together with the Riesz-Fredholm theorem provided our understanding to the Sturm-Liouville boundary value problems.

Theorem (Riesz-Schauder). *Let $K \in L(X)$ be compact.*

(a) *If $\dim X = \infty$, then $0 \in \sigma(K)$. If $c \in \sigma(K)$ and $c \neq 0$, then c is an eigenvalue of K and K^* of finite multiplicities (i.e. the dimensions of the spaces of eigenvectors are finite).*

(b) *$\sigma(K)$ is a countable compact set and 0 is the only possible limit point of $\sigma(K)$.*

Proof. (a) If $0 \notin \sigma(K)$ (i.e. K is invertible), then by property (e), $\dim X < \infty$. The contrapositive asserts that if $\dim X = \infty$, then $0 \in \sigma(K)$.

Next, if $c \in \sigma(K) \setminus \{0\}$, then $K - cI$ is either not injective or not subjective. By part (d) of the Riesz-Fredholm theorem, $0 < \dim\ker(K - cI) = \dim\ker(K^* - cI) < \infty$. Therefore, c is an eigenvalue of K and K^* of finite multiplicities.

(b) For $c \in \sigma(K) \setminus \{0\}$, by part (c) of the Riesz-Fredholm theorem, $A = (K - cI)|_{M_j}$ is invertible. By the lemma on inverses, for $|z - c| < \|A^{-1}\|^{-1}$, we know $(K - zI)|_{M_j} = A - (z - c)I$ is invertible.

Also, by part (c) of the Riesz-Fredholm theorem, $T = (K - cI)|_{N_j}$ is nilpotent of index j, i.e. $T^j \equiv 0$. Observe that for $\alpha \neq 0$, $(T - \alpha I)^{-1} = -\alpha^{-j}(T^{j-1} + \alpha T^{j-2} + \cdots + \alpha^{j-1} I)$. Hence $\sigma(T) = \{0\}$. Then, for $z \neq c$, $(K - zI)|_{N_j} = T - (z - c)I$ is invertible. So for $0 < |z - c| < \|A^{-1}\|^{-1}$, $K - zI$ is invertible on $X = M_j \oplus N_j$, i.e. $z \notin \sigma(K)$. Hence c is an isolated point in $\sigma(K)$. For $n = 1, 2, 3, \ldots$, the set $S_n = \sigma(K) \cap \{z : |z| \geq 1/n\}$ is finite (otherwise, by the Bolzano-Weierstrass theorem, S_n has a limit point c, which cannot be isolated). Therefore, $\sigma(K) \setminus \{0\} = S_1 \cup S_2 \cup S_3 \cup \cdots$ is countable and 0 is the only possible limit point of $\sigma(K)$. $\square$

In the beginning of the twentieth century, Fredholm inspired many mathematicians to investigate integral equations. These works led to the solutions of the Neumann and Dirichlet problems by single and double layer potential methods (see Folland's <u>*Introduction to Partial Differential Equations*</u>, Chapter 3). The integral equations were mostly of the form $\int_a^b G(s,t)x(t)\,dt - cx(s) = y(s)$. In case G and x are continuous, the first term on the left is a compact operator. The studies on these equations led to the theory of compact operators. The following were the results obtained for these equations.

Corollary (Fredholm Alternatives). *Let X be a Banach space, $K \in L(X)$ be compact and $c \neq 0$. Either* (a) *$K - cI$ is invertible or* (b) $0 < \dim\ker(K - cI) < \infty$.

If (a) *holds, then $K^* - cI$ is invertible. If* (b) *holds, then* $0 < \dim\ker(K - cI) = \dim\ker(K^* - cI) < \infty$.

Furthermore, there exists $x \in X$ such that $(K - cI)x = y$ if and only if $y \in {}^{\perp}(\ker(K^* - cI))$. Also, there exists $x^* \in X^*$ such that $(K^* - cI)x^* = y^*$ if and only if $y^* \in (\ker(K - cI))^{\perp}$.

Proof. By part (d) of the Riesz-Fredholm theorem, $0 \leq \dim\ker(K - cI) = \operatorname{codim}\operatorname{ran}(K - cI) < \infty$. Alternative (a) is the case $0 = \dim\ker(K - cI) = \operatorname{codim}\operatorname{ran}(K - cI)$. Alternative (b) is the case $0 < \dim\ker(K - cI) < \infty$.

If (a) holds, then $0 = \dim\ker(K^* - cI) = \operatorname{codim}\operatorname{ran}(K^* - cI)$. If (b) holds, then $0 < \dim\ker(K - cI) = \dim\ker(K^* - cI) < \infty$.

The furthermore statement follows as $\operatorname{ran}(K - cI) = \overline{\operatorname{ran}(K - cI)} = {}^{\perp}(\ker(K^* - cI))$ and $\operatorname{ran}(K^* - cI) = \overline{\operatorname{ran}(K^* - cI)} = \overline{\operatorname{ran}(K^* - cI)}^{w^*} = (\ker(K - cI))^{\perp}$ by using the closed range theorem and the kernel-range relations. $\square$

In ordinary differential equation, the Sturm-Liouville boundary value problems (see Boyce and DiPrima's *Elementary Differential Equations and Boundary Value Problems,* Chapter 11) are important. It is well-known that the corresponding Sturm-Liouville operators have real eigenvalue sequence tending to infinity. Being unbounded operators, when they are injective, it is known (see Gohberg, Goldberg and Kaashoek's *Basic Classes of Linear Operator*, Chapter 6) to have inverses, which are compact integral operators.

One of the most important problems in operator theory is to determine if every operator $T \in L(X)$ has a nontrivial closed invariant subspace M (i.e. $\{0\} \subset M \subset X$ and $T(M) \subseteq M$). For $X = \ell^1$, Enflo proved that there exists operators without nontrivial closed invariant subspaces. The case X is a Hilbert space is still open. For compact operators, not only do they have nontrivial closed invariant subspaces, but we also have the following stronger results.

Lomonosov's Theorem. *Let X be an infinite dimensional Banach space over $\mathbb{C}$ and K be a nonzero compact operator. Then there exists a closed subspace M of X such that $\{0\} \subset M \subset X$ and for every $T \in L(X)$ commuting with K (i.e. satisfying $TK = KT$), we have $T(M) \subseteq M$. Such a closed subspace M is called a nontrivial hyperinvariant subspace of K.*

Proof. (*Due to H. M. Hilden*) Let $\Gamma = \{S \in L(X) : SK = KS\}$, which is called the *commutant* of K. For every $y \in X$, $\Gamma_y = \overline{\{Sy : S \in \Gamma\}}$ is a closed subspace of X which contains $I(y) = y$. If $y \neq 0$, then $\{0\} \subset \Gamma_y$. Also, for all $T, S \in \Gamma$, since $TSK = TKS = KTS$ implies $TS \in \Gamma$, we get $T(\Gamma_y) \subseteq \Gamma_y$.

If there is a $y \neq 0$ such that $\Gamma_y \subset X$, then $M = \Gamma_y$ is a nontrivial hyperinvariant subspace of K.

Otherwise, $\Gamma_y = X$ for all $y \neq 0$. Since $K \neq 0$, there exists $x_0 \in X \setminus \{0\}$ such that $Kx_0 \neq 0$. Since K is bounded, $K^{-1}(B(Kx_0, \|Kx_0\|/2))$ is an open set containing x_0. Then for some $r > 0$, $B = B(x_0, r)$ is inside $B(x_0, \|x_0\|/2) \cap K^{-1}(B(Kx_0, \|Kx_0\|/2))$. So for all $x \in B$, $\|x - x_0\| < r \leq \|x_0\|/2$, which implies $\|x\| \geq \|x_0\| - \|x - x_0\| \geq \|x_0\|/2 > 0$. Then $0 \notin B$. Similarly, for all $x \in B$, $\|Kx - Kx_0\| < \|Kx_0\|/2$ implies $\|Kx\| \geq \|Kx_0\| - \|Kx - Kx_0\| \geq \|Kx_0\|/2 > 0$. Then $0 \notin \overline{K(B)}$.

For every $y \in \overline{K(B)}$, since $\Gamma_y = X$, hence $\{Sy : y \in \Gamma\}$ is dense in X, there is some $S_y \in \Gamma$ such that $S_y(y) \in B$. Then $W_y = S_y^{-1}(B)$ is open and contains y. Since $\{W_y : y \in \overline{K(B)}\}$ is an open cover of $\overline{K(B)}$, there are $W_{y_1}, \ldots, W_{y_n}$ such that $\overline{K(B)} \subseteq W_{y_1} \cup \cdots \cup W_{y_n}$. For simplicity, write W_i for W_{y_i} and S_i for S_{y_i}. Since $S_i(W_i) \subseteq B$ and $0 \notin B$, $S_i \neq 0$. So $d = \max\{\|S_1\|, \ldots, \|S_n\|\} > 0$.

Recall x_0 is the center of B. "Now $Kx_0 \in \overline{K(B)}$. So there are S_{i_1} and W_{i_1} such that $Kx_0 \in W_{i_1}$, $S_{i_1}Kx_0 \in S_{i_1}(W_{i_1}) \subseteq B$ and $KS_{i_1}Kx_0 \in \overline{K(B)}$. So there are S_{i_2} and W_{i_2} such that $KS_{i_1}Kx_0 \in W_{i_2}$ so that $S_{i_2}KS_{i_1}Kx_0 \in B$." Inductively, for every positive integer j, there is $x_j = S_{i_j}K \cdots S_{i_1}Kx_0 = S_{i_j} \cdots S_{i_1}K^j x_0 \in B$. Hence, $d^j\|K^j\|\|x_0\| \geq \|x_j\| \geq \|x_0\| - \|x_j - x_0\| \geq \|x_0\|/2$. By the Gelfand-Mazur theorem, $r(K) = \lim\limits_{j\to\infty} \|K^j\|^{1/j} \geq 1/d > 0$. Then $\sigma(K)$ contains some $c \neq 0$.

By the Riesz-Fredholm Lemma, c is an eigenvalue of K. Then $M = \ker(K - cI) = \{v \in X : Kv = cv\}$ is finite dimensional. Hence, M is a closed subspace satisfying $\{0\} \subset M \subset X$. For every $T \in \Gamma$ and $v \in M$, we have $KTv = TKv = T(cv) = cTv$, which implies $T(M) \subseteq M$. So, M is hyperinvariant. ∎

Remark. In fact, Lomonosov proved a even stronger result, namely if $A \neq 0$ commutes with $B \neq 0$, which commutes with a nonzero compact operator, then A has a nontrivial closed invariant subspace.

Appendix : Fredholm Operators

In this appendix, we study a special class of operators, for which we can associate an index that has deep connections with elliptic differential operators on manifolds. In the 1960s, Atiyah and Singer proved a theorem connecting this analytic index on some differential operators to a topological index on a manifold that generalized the winding number of a closed curve around a point. The famous Atiyah-Singer index theorem was a great achievement in the 20th century mathematics. We recommend Booss and Bleecker's book *Topology and Analysis* for an understanding of this theorem.

Definitions. For Banach spaces X and Y, $T \in L(X,Y)$ is a *Fredholm operator* iff ($\operatorname{ran} T$ is closed,) $\dim\ker T < \infty$ and $\operatorname{codim}\operatorname{ran} T < \infty$. For a Fredholm operator, the *index* of T is $\operatorname{ind} T = \dim\ker T - \operatorname{codim}\operatorname{ran} T$. In some literatures, the *cokernel* of T is defined to be $\operatorname{coker} T = Y/(\operatorname{ran} T)$ and in that case, $\operatorname{ind} T = \dim\ker T - \dim\operatorname{coker} T$.

Theorem. *If $Y/\operatorname{ran} T$ is finite dimensional as a vector space, then $\operatorname{ran} T$ is closed. So the condition $\operatorname{ran} T$ is closed is unnecessary in the definition of Fredhom operators.*

Proof. Let W be a finite dimensional vector subspace of Y such that $\operatorname{ran} T \cap W = \{0\}$ and $\operatorname{ran} T + W = Y$. Then $(X/\ker T) \oplus W$ is a Banach space. Define $f : (X/\ker T) \oplus W \to Y$ by $f([x], w) = Tx + w$. Since $\widehat{T} : X/\ker T \to \operatorname{ran} T$ is an isomorphism, $f = \widehat{T} \oplus I$ is bijective and continuous. Then f^{-1} is continuous. Since $(X/\ker T) \oplus \{0\}$ is complete, hence closed, $\operatorname{ran} T = f((X/\ker T) \oplus \{0\})$ is closed.

Examples. (1) If $T : X \to Y$ is invertible, then T is Fredholm with $\ker T = \{0\}$, $\operatorname{ran} T = Y$ and so $\operatorname{ind} T = 0$.

(2) If $T_0 : X_0 \to Y_0$ and $T_1 : X_1 \to Y_1$ are Fredholm, then $T_0 \oplus T_1 : X_0 \oplus X_1 \to Y_0 \oplus Y_1$ is Fredholm with $\ker(T_0 \oplus T_1) = (\ker T_0) \oplus (\ker T_1)$, $\operatorname{ran}(T_0 \oplus T_1) = (\operatorname{ran} T_0) \oplus (\operatorname{ran} T_1)$ and so $\operatorname{ind}(T_0 \oplus T_1) = \operatorname{ind} T_0 + \operatorname{ind} T_1$.

(3) If $K \in L(X)$ is compact and $c \neq 0$, then $K - cI$ is Fredholm and $\operatorname{ind}(K - cI) = 0$ by the Riesz-Fredholm lemma and theorem. (It is proved below that an operator is Fredholm with index 0 iff it is the sum of an invertible operator and a compact (in fact, finite rank) operator.)

(4) The unilateral shift S on ℓ^2 defined by $S(c_0, c_1, c_2, \ldots) = (0, c_0, c_1, c_2, \ldots)$ is Fredholm with $\operatorname{ind} S = \dim\ker S - \operatorname{codim}\operatorname{ran} S = 0 - 1 = -1$. The backward shift S^* on ℓ^2 defined by $S^*(c_0, c_1, c_2, \ldots) = (c_1, c_2, c_3, \ldots)$ is also Fredholm with $\operatorname{ind} S^* = \dim\ker S^* - \operatorname{codim}\operatorname{ran} S^* = 1 - 0 = 1$. (It is proved below that $\operatorname{ind} T^* = -\operatorname{ind} T$.) Also, S^n and $(S^*)^n$ are Fredholm with $\operatorname{ind}(S^n) = -n$ and $\operatorname{ind}((S^*)^n) = n$.

(5) If $T \in L(X,Y), \dim X < \infty$ and $\dim Y < \infty$, then T is Fredholm. Since $\operatorname{codim}\operatorname{ran} T = \dim(Y/\operatorname{ran} T) = \dim Y - \dim\operatorname{ran} T$ and $\dim\ker T + \dim\operatorname{ran} T = \dim X$, so $\operatorname{ind} T = \dim X - \dim Y$.

Theorem (Atkinson). *Let $T \in L(X,Y)$. The following are equivalent:*

(a) *T is Fredholm,*

(b) *there is $S \in L(Y,X)$ such that $I - TS$ and $I - ST$ are finite rank (S is called a Fredholm inverse of T).*

(c) *there are $S, S' \in L(Y,X)$ such that $I - TS$ and $I - S'T$ are compact.*

Proof. (a) $\Rightarrow$ (b) Since $\dim\ker T < \infty$, there exists a projection $P \in L(X)$ with $\operatorname{ran} P = \ker T$. Since $\operatorname{codim}\operatorname{ran} T < \infty$, there exists a projection $Q \in L(Y)$ with $\operatorname{ran} Q = \operatorname{ran} T$. Let $Z = \operatorname{ran}(I - P) = \ker P$. From $X = \operatorname{ran} P \oplus \operatorname{ran}(I - P) = \ker T \oplus Z$, we see $T_0 = T|_Z : Z \to \operatorname{ran} T$ is injective (as $\ker T \cap Z = \{0\}$) and surjective (as $T(X) = T(\ker T \oplus Z) = T(Z)$). By the inverse mapping theorem, T_0 is invertible.

Let $S = T_0^{-1}Q : Y \to Z \subseteq X$. We first check $QT = T_0(I - P)$. (For all $x \in X$, $Tx \in \operatorname{ran} Q$. So $QTx = Tx = T(Px + (I - P)x) = T_0\big((I - P)x\big)$.) So we have $ST = T_0^{-1}QT = T_0^{-1}T_0(I - P) = I - P$ and $TS = TT_0^{-1}Q = T_oT_o^{-1}Q = Q = I - (I - Q)$. Now $\dim\operatorname{ran}(I - ST) = \dim\operatorname{ran} P = \dim\ker T < \infty$ and $\dim\operatorname{ran}(I - TS) = \dim\operatorname{ran}(I - Q) = \operatorname{codim}\operatorname{ran} Q = \operatorname{codim}\operatorname{ran} T < \infty$.

(b) $\Rightarrow$ (c) Let $S' = S$. Finite rank operators are compact.

(c) $\Rightarrow$ (a) $TS = I + K$ for some compact operator $K \in L(Y)$. By the Riesz-Fredholm lemma, $\operatorname{ran} TS = \operatorname{ran}(I + K)$ is closed and $\operatorname{codim}\operatorname{ran} TS = \operatorname{codim}\operatorname{ran}(I + K) < \infty$. Also, since $\operatorname{ran} TS \subseteq \operatorname{ran} T \subseteq Y$, $\operatorname{codim}\operatorname{ran} T < \infty$. By the theorem following the definition of Fredhom operators, $\operatorname{ran} T$ is closed.

Next $S'T = I + L$ for some compact operator $L \in L(X)$. Since $\dim\ker S'T = \dim\ker(I + L) < \infty$ and $\ker T \subseteq \ker S'T$, we get $\dim\ker T < \infty$. Therefore, T is Fredholm. $\square$

Definition. Let $K(X)$ be the set of all compact operators on X. By the properties of compact operators, we see $K(X)$ is a closed two-sided ideal in $L(X)$. Then $L(X)/K(X)$ is a Banach algebra with $[T][S] =$

$(T+K(X))(S+TK(X)) = TS+K(X) = [TS]$. (Let $K_n, L_n \in K(X)$ satisfy $\|T-K_n\| \to \|[T]\|$ and $\|S-L_n\| \to \|[S]\|$. Then

$$\begin{aligned}\|[TS]\| = \inf\{\|TS-K\| : K \in K(X)\} &\le \liminf_{n\to\infty} \|TS - TL_n - K_nS + K_nL_n\| \\ &\le \lim_{n\to\infty} \|T-K_n\|\|S-L_n\| = \|[T]\|\|[S]\|.)\end{aligned}$$

We called $L(X)/K(X)$ the <u>*Calkin algebra*</u> on X.

<u>Theorem (Properties of Fredholm Operators).</u> (a) *$T \in L(X)$ is Fredholm iff $[T] = T + K(X)$ is invertible in $L(X)/K(X)$.*

(b) *If $T \in L(X,Y)$ is Fredholm and $K \in L(X,Y)$ is compact, then $T+K$ is Fredholm.*

(c) *If $T \in L(X,Y)$ is Fredholm and $S \in L(Y,X)$ is a Fredholm inverse of T, then S is Fredholm.*

(d) *If $T \in L(X,Y)$ is Fredholm, then $T^* \in L(Y^*, X^*)$ is Fredholm with* $\operatorname{ind} T^* = -\operatorname{ind} T$.

<u>Proof.</u> (a) If T is Fredholm, then let S be a Fredholm inverse of T. We have $[T][S]-[I] = [TS-I] = [0] = [ST-I] = [S][T]-[I]$. So $[T][S] = [I] = [S][T]$. Conversely, if $[S] = [T]^{-1} \in L(X)/K(X)$, then $[I-TS] = [0] = [I-ST]$ implies $I-TS$ and $I-ST$ are compact. So T is Fredholm by Atkinson's theorem.

(b) By (b) of Atkinson's theorem, there is $S \in L(Y,X)$ such that $I-TS$ and $I-ST$ are finite rank. Then $I-(T+K)S = (I-TS)-KS$ and $I-S(T+K) = (I-ST)-SK$ are compact, which implies $T+K$ Fredholm by Atkinson's theorem.

(c) Observe that S has T as a Fredholm inverse. By Atkinson's theorem, S is Fredholm.

(d) By the closed range theorem, $\operatorname{ran} T$ closed implies $\operatorname{ran} T^*$ closed and w^*-closed. Since $\ker T$ and $Y/\operatorname{ran} T$ are finite dimensional, by the kernel-range relations and the duality theorem,

$$\dim\ker T^* = \dim(\operatorname{ran} T)^\perp = \dim(Y/\operatorname{ran} T)^* = \dim(Y/\operatorname{ran} T) = \operatorname{codim}\operatorname{ran} T < \infty,$$

$$\operatorname{codim}\operatorname{ran} T^* = \operatorname{codim}\overline{\operatorname{ran} T^*}^{w^*} = \operatorname{codim}(\ker T)^\perp = \dim(X^*/(\ker T)^\perp) = \dim(\ker T)^* = \dim\ker T < \infty.$$

Then $\operatorname{ind} T^* = \dim\ker T^* - \operatorname{codim}\operatorname{ran} T^* = \operatorname{codim}\operatorname{ran} T - \dim\ker T = -\operatorname{ind} T$.

<u>Lemma 1.</u> *If $T \in L(X,Y)$ is Fredholm and M is a closed subspace of X, then $T(M)$ is closed in Y.*

<u>Proof.</u> As $\dim\ker T < \infty$, it has a complementary subspace W so that $X = \ker T \oplus W$. Now $\ker T \cap W = \{0\}$ implies $T|_W$ is injective. Also $\operatorname{ran} T|_W = \operatorname{ran} T$ is closed, hence complete. By lower bound theorem, $T|_W$ is bounded below. Hence T maps closed subspaces of W to complete (hence closed) subspaces of Y.

If M is a closed subspace of X, then $M + \ker T$ is a closed subspaces of X because letting $\pi_M : X \to X/M$ be the quotient map, $\dim \pi_M(\ker T) \le \dim\ker T < \infty$ implies $\pi_M(\ker T)$ is closed in X/M and so $\pi_M^{-1}(\pi_M(\ker T)) = M + \ker T$ is closed. Next, $T(M+\ker T) = T((M+\ker T)\cap W)$ because every $x \in M + \ker T \subseteq X = \ker T \oplus W$ is of the form $t+w$, where $t \in \ker T$ and $w \in W$, so $x - t = w \in (M+\ker T)\cap W$ and $T(x) = T(w)$. Therefore, $T(M) = T(M+\ker T) = T\big((M+\ker T)\cap W\big)$ is a closed subspace of Y. ∎

<u>Lemma 2.</u> *If F is a subspace of X with finite codimension, E_0 is a subspace of X such that $E_0 \cap F = \{0\}$, then there is a closed subspace $E \supseteq E_0$ such that $E \oplus F = X$.*

<u>Proof.</u> For the quotient map $\pi : X \to X/F$, we have $\ker \pi = F$. Since $E_0 \cap F = \{0\}$, so $\pi|_{E_0}$ is injective. Take a basis $B = \{x_1, \ldots, x_i\}$ of E_0. Then $\pi(B)$ is a basis of $\pi(E_0)$. Since $\dim(X/F) < \infty$, we can extend $\pi(B)$ to a basis $W = \{x_1 + F, \ldots, x_n + F\}$ of X/F for some $n \ge i$. Let $E = \operatorname{span}\{x_1, \ldots, x_n\}$. Then E contains E_0. Now $\dim E < \infty$ implies E is complete, hence closed. Also, W linearly independent implies $E \cap F = \{0\}$ because $c_1x_1 + \cdots + c_nx_n \in E \cap F$ implies $c_1(x_1+F) + \cdots + c_n(x_n+F) = 0 + F$, which forces all $c_i = 0$ by linear independence of W. Also, $X/F = \operatorname{span} W$ implies $E + F = X$. Therefore, $E \oplus F = X$. ∎

Multiplication Theorem. *If $T \in L(X,Y)$ is Fredholm and $S \in L(Y,Z)$ is Fredholm, then ST is Fredholm with* $\operatorname{ind}(ST) = \operatorname{ind} S + \operatorname{ind} T$.

Proof. (*Due to Donald Sarason*) In the finite dimensional case (i.e. $\dim X, \dim Y, \dim Z < \infty$), by example 5, ST is Fredholm and $\operatorname{ind}(ST) = \dim X - \dim Z = \dim X - \dim Y + \dim Y - \dim Z = \operatorname{ind} S + \operatorname{ind} T$.

Otherwise, by lemma 1, $\operatorname{ran} ST = S(\operatorname{ran} T)$ is closed. Now $\dim\ker ST = \dim T^{-1}(\ker S) \le \dim\ker S + \dim\ker T < \infty$ and $\operatorname{codim}\operatorname{ran} ST = \operatorname{codim} S(\operatorname{ran} T) \le \operatorname{codim}\operatorname{ran} S + \operatorname{codim}\operatorname{ran} T < \infty$. So ST is Fredholm.

To get $\operatorname{ind}(ST) = \operatorname{ind} S + \operatorname{ind} T$, it suffices to decompose $X = X_0 \oplus X_1, Y = Y_0 \oplus Y_1, Z = Z_0 \oplus Z_1$ with $\dim X_0, \dim Y_0, \dim Z_0 < \infty$. Also, decompose $T = T|_{X_0} \oplus T|_{X_1}$, $S = S|_{Y_0} \oplus S|_{Y_1}$, where $T|_{X_i} : X_i \to Y_i$ and $S|_{Y_i} : Y_i \to Z_i$, with $T|_{X_1}$ and $S|_{Y_1}$ invertible.

Once these are done, we can finish as follow: $ST|_{X_i} = S|_{Y_i} \circ T|_{X_i} : X_i \to Z_i$ and $ST|_{X_1}$ is invertible. By examples 1 and 2, $\operatorname{ind} S = \operatorname{ind} S|_{Y_0}, \operatorname{ind} T = \operatorname{ind} T|_{X_0}$ and $\operatorname{ind}(ST) = \operatorname{ind}(ST|_{X_0})$. From the finite dimensional case, $\operatorname{ind}(ST|_{X_0}) = \operatorname{ind}(S|_{Y_0}) + \operatorname{ind}(T|_{X_0})$, which gives $\operatorname{ind}(ST) = \operatorname{ind} S + \operatorname{ind} T$.

Now we begin the decompositions. Let $\boxed{X_0 = \ker ST.}$ From above, $\dim X_0 < \infty$. So there is a closed subspace X_1 such that $X_0 \oplus X_1 = X$. By lemma 1, $\boxed{Y_1 = TX_1}$ is closed in Y. Since $\ker T \subseteq \ker ST = X_0$, so $\ker T \cap X_1 = \{0\}$ and $\boxed{T|_{X_1} : X_1 \to TX_1 = Y_1 \text{ is invertible.}}$ Now $\operatorname{ran} T = TX_0 \oplus TX_1$ and $\dim(\operatorname{ran} T/TX_1) = \dim TX_0 \le \dim X_0 < \infty$ imply

$$\operatorname{codim} Y_1 = \dim(Y/TX_1) = \dim(Y/\operatorname{ran} T) + \dim(\operatorname{ran} T/TX_1) \le \operatorname{codim}\operatorname{ran} T + \dim X_0 < \infty. \qquad (*)$$

Next $\ker S \cap Y_1 = \ker S \cap TX_1 = \{0\}$ because $Tx_1 \in \ker S$ for some $x_1 \in X_1$ implies $x_1 \in X_1 \cap X_0 = \{0\}$. By lemma 2, there is a closed subspace $Y_0 \supseteq \ker S$ such that $Y_0 \oplus Y_1 = Y$. Then $TX_0 = T(\ker ST) = T(T^{-1}(\ker S)) \subseteq \ker S \subseteq Y_0$, i.e. $\boxed{T|_{X_0} : X_0 \to Y_0.}$ Also $\dim Y_0 = \dim(Y/Y_1) = \operatorname{codim} Y_1 < \infty$ by (*). So we have $\boxed{T = T|_{X_0} \oplus T|_{X_1}.}$

By lemma 1, $\boxed{Z_1 = SY_1}$ is a closed subspace of Z. Since $Y = Y_0 \oplus Y_1$, $\ker S \subseteq Y_0$ and $\ker S \cap Y_1 = \{0\}$, so $\boxed{S|_{Y_1} : Y_1 \to SY_1 = Z_1 \text{ is invertible.}}$ As in (*) above, $\operatorname{codim} Z_1 \le \operatorname{codim}\operatorname{ran} S + \dim Y_0 < \infty$. (**)

Next $SY_0 \cap Z_1 = SY_0 \cap SY_1 = \{0\}$ (because $Sy_0 = Sy_1$ for $y_0 \in Y_0, y_1 \in Y_1$ implies $y_0 - y_1 \in \ker S \subseteq Y_0$, which implies $y_1 \in Y_0 \cap Y_1 = \{0\}$, then $Sy_0 = Sy_1 = 0$). By lemma 2, there is a closed subspace $Z_0 \supseteq SY_0$ such that $Z_0 \oplus Z_1 = Z$ and $\boxed{S|_{Y_0} : Y_0 \to Z_0}$. Also $\dim Z_0 = \operatorname{codim} Z_1 < \infty$ by (**). So $\boxed{S = S|_{Y_0} \oplus S|_{Y_1}.}$ ∎

Corollary. *If $T \in L(X,Y)$ is Fredholm and $S \in L(Y,X)$ is a Fredholm inverse of T, then* $\operatorname{ind}(S) = -\operatorname{ind}(T)$.

Proof. By property (c) of Fredholm operators, we know S is Fredholm. Now $I - ST = K$ for some compact operator $K \in L(X)$. By example 3 and muliplication theorem, $0 = \operatorname{ind}(I - K) = \operatorname{ind}(ST) = \operatorname{ind} S + \operatorname{ind} T$. So $\operatorname{ind} S = -\operatorname{ind} T$. ∎

Perturbation Theorem. *Let $T \in L(X,Y)$ be Fredholm. Then there is $\varepsilon > 0$ so that $T + A$ is Fredholm with* $\operatorname{ind}(T+A) = \operatorname{ind} T$, *where $A \in L(X,Y)$ with $\|A\| < \varepsilon$. (This implies the Fredholm operators form an open set in $L(X,Y)$ and the index is continuous and constant on each connected component of that set.)*

Proof. By Atkinson's theorem, there exists $S \in L(Y,X)$ such that $K = I - TS$ and $L = I - ST$ are finite rank. Let $\varepsilon = \|S\|^{-1}$. Let $A \in L(X,Y)$ satisfy $\|A\| < \varepsilon$. As $\|AS\| \le \|A\|\|S\| < 1$, $I + AS$ is invertible. Now

$$(T+A)S = I - K + AS = (I + AS) - K = \Big(I - K(I+AS)^{-1}\Big)(I + AS).$$

Solving for $K(I+AS)^{-1}$, we see $I - (T+A)\Big(S(I+AS)^{-1}\Big) = K(I+AS)^{-1}$ is compact. Similarly, $I + SA$ is invertible and $I - \Big((I+SA)^{-1}S\Big)(T+A) = (I+SA)^{-1}L$ is compact. So, by Atkinson's theorem, $T + A$ is Fredholm. The last equation is the same as $I - (I+SA)^{-1}L = (I+SA)^{-1}S(T+A)$. Taking index on both sides, by example 3, multiplication theorem and example 1, we get $0 = 0 + \operatorname{ind} S + \operatorname{ind}(T+A)$. By the corollary above, $\operatorname{ind}(T+A) = -\operatorname{ind} S = \operatorname{ind} T$. ∎

Corollary. *If* $T \in L(X,Y)$ *is Fredholm and* $K \in L(X,Y)$ *is compact, then* $\operatorname{ind}(T+K) = \operatorname{ind} T$.

Proof. Since $f(t) = \operatorname{ind}(T+tK)$ is a continuous function on $[0,1]$ with integer value, it is a constant function. In particular, $\operatorname{ind}(T+K) = f(1) = f(0) = \operatorname{ind}(T)$. $\square$

Theorem. *Let* $A \in L(X,Y)$. *The following are equivalent.*

(a) A *is Fredholm with* $\operatorname{ind} A = 0$,

(b) $A = C + F$, *where* C *is invertible in* $L(X,Y)$ *and* F *is finite rank in* $L(X,Y)$,

(c) $A = B + K$, *where* B *is invertible in* $L(X,Y)$ *and* K *is compact in* $L(X,Y)$.

Proof. (a)$\Rightarrow$(b) If $\operatorname{ind} A = 0$, then $\dim\ker A = \operatorname{codim}\operatorname{ran} A < \infty$. Let Z be a complementary subspace of $\ker A$ in X. Let W be a complementary subspace of $\operatorname{ran} A$ in Y. Let $P \in L(X)$ be a projection such that $\operatorname{ran} P = \ker A$ is finite dimensional. Since $\dim W = \operatorname{codim}\operatorname{ran} A = \dim\ker A < \infty$, there is an invertible operator $T : \ker A \to W$.

Now $A + TP$ is injective because $(A+TP)(x) = 0$ implies $Ax = -TPx \in \operatorname{ran} A \cap W = \{0\}$. Then $Ax = 0$ implies $x \in \ker A = \operatorname{ran} P$ so that $Px = x$ and $Tx = TPx = -Ax = 0$. Since T is invertible, $x = 0$.

For surjectivity of $A + TP$, first observe that $X = \ker A \oplus Z$ implies $\operatorname{ran} A = A(X) = A(Z)$. Next, P is the projection onto $\ker A$ implies $P(Z) = \{0\}$. Also, $TP(\ker A) = TP(\operatorname{ran} P) = T(\operatorname{ran} P) = T(\ker A) = W$. Then, $A + TP$ is surjective since $(A+TP)(\ker A \oplus Z) = TP(\ker A) \oplus A(Z) = W \oplus \operatorname{ran} A = Y$.

Hence, $A + TP$ is invertible. Since $\dim W < \infty$, TP is finite rank. Then $A = (A+TP) - TP$ satisfies the required conditions.

(b)$\Rightarrow$(c) Finite rank implies compactness.

(c)$\Rightarrow$(a) $B + K$ is Fredholm follows by example 1 and property (b) of Fredholm operators. Also, by example 1 and the last corollary, $\operatorname{ind}(B+K) = \operatorname{ind}(B) = 0$. Alternatively, $\operatorname{ind} A = \operatorname{ind}(B+K) = \operatorname{ind} B(I + B^{-1}K) = \operatorname{ind} B + \operatorname{ind}(I + B^{-1}K) = 0$ by the multiplication theorem, examples 1 and 3. $\square$

Chapter 6. Basic Operator Facts on Hilbert Spaces.

§1. **Adjoints.** Throughout this chapter H, H_1, H_2 will denote Hilbert spaces over $\mathbb{C}$. The inner product on H will be denoted by $(\ ,\ )$. For every $y \in H$, the linear functional $f_y(x) = (x, y)$ is in H^*. Recall that the Riesz representation theorem asserted that there is a bijection from H onto H^* given by $y \mapsto f_y$. For all $y, y' \in H$, it satisfies $\|y\| = \|f_y\|$, $f_{y+y'} = f_y + f_{y'}$. It may seem H^* is isometric isomorphic to H. Unfortunately, for all $c \in \mathbb{K}$ and $y \in H$, $f_{cy} = \overline{c} f_y$. Keeping this in mind, we say there is a *conjugate-linear* isometric isomorphism between H and H^*. By a slight abuse of meaning, it is popular to write $H^* = H$, where f_y is identified with y. Alternatively, we can consider $H^* = H_{twin}$, where cx in H_{twin} is $\overline{c}x$ in H. In particular, H is reflexive so that the weak and weak-star topologies coincide.

Now for every $T \in L(H_1, H_2)$ and $y \in H_2$, the function $g(x) = (Tx, y)$ is in H_1^*. By the Riesz representation theorem, there exists a unique $w \in H_1$ such that $g(x) = (x, w)$. Define the *adjoint* of $T \in L(H_1, H_2)$ to be $T^* \in L(H_2, H_1)$ given by $T^*y = w$. So $(Tx, y) = (x, T^*y)$ for all $x \in H_1, y \in H_2$. Taking conjugate on both sides, also $(y, Tx) = (T^*y, x)$.

Remarks. (1) For $S, T \in L(H_1, H_2)$, $S = T$ if and only if for all $y \in H_1, x \in H_2$, $(x, Sy) = (x, Ty)$, which is clear if we set $x = Sy - Ty$ and get $\|Sy - Ty\|^2 = 0$. Hence $T^{**} = T$ as $(x, T^{**}y) = (T^*x, y) = (x, Ty)$.

(2) For $T, S \in L(H_1, H_2)$ and $c \in \mathbb{C}$, we have $(T + S)^* = T^* + S^*$ and$(cT)^* = \overline{c}T^*$ because $(x, (T + S)^*y) = ((T + S)x, y) = (Tx, y) + (Sx, y) = (x, T^*y) + (x, S^*y) = (x, (T^* + S^*)y)$ and $(x, (cT)^*y) = (cTx, y) = c(Tx, y) = c(x, T^*y) = (x, \overline{c}T^*y)$.

(3) For $T_0 \in L(H_0, H_1)$ and $T_1 \in L(H_1, H_2)$, we have $(T_1T_0)^* = T_0^*T_1^*$ because $(x, (T_1T_0)^*y) = (T_1T_0x, y) = (T_0x, T_1^*y) = (x, T_0^*T_1^*y)$. Also, T is invertible if (and only if) T^* is invertible with $(T^*)^{-1} = (T^{-1})^*$.

In general, facts about Banach spaces also apply to Hilbert spaces and in some places where adjoints will be needed, we need to do conjugations. For example, $(T - cI)^* = T^* - \overline{c}I$. So $\sigma(T^*) = \{\overline{c} : c \in \sigma(T)\}$ because $c \notin \sigma(T)$ iff $T - cI$ is invertible iff $(T - cI)^* = T^* - \overline{c}I$ is invertible iff $\overline{c} \notin \sigma(T^*)$.

Definitions. (1) An *involution* on a Banach algebra B is a map from B to B sending every $x \in B$ to some $x^* \in B$ such that for every $a, b \in B$ and $c \in \mathbb{K}$, $a^{**} = a$, $(a + b)^* = a^* + b^*$, $(ca)^* = \overline{c}a^*$ and $(ab)^* = b^*a^*$.

(2) A *C^*-algebra* is a Banach algebra B with an involution such that for every $x \in B$, we have $\|x^*x\| = \|x\|^2$. (Then $\|x^*\| = \|x\|$ because $\|x\|^2 = \|x^*x\| \le \|x^*\|\|x\|$ implies $\|x\| \le \|x^*\|$ and from this, $\|x^*\| \le \|x^{**}\| = \|x\|$. Also, the involution operation is continuous since $x_n \to x \iff \|x_n^* - x^*\| = \|x_n - x\| \to 0 \iff x_n^* \to x^*$.)

Theorem. *For $T \in L(H_1, H_2)$, we have $\|T^*T\| = \|T\|^2$. (So $L(H)$ is a C^*-algebra with adjoint as involution.) Also, $H_1 = \ker T \oplus \overline{\operatorname{ran} T^*}$ and $H_2 = \ker T^* \oplus \overline{\operatorname{ran} T}$.*

Proof. Since $\|T^*\| = \|T\|$, so $\|T^*T\| \le \|T^*\|\|T\| = \|T\|^2$. Conversely, for $\|x\| \le 1$, $\|Tx\|^2 = (Tx, Tx) = (x, T^*Tx) \le \|T^*Tx\|\|x\| \le \|T^*T\|\|x\|^2 \le \|T^*T\|$, which implies $\|T\|^2 \le \|T^*T\|$. The last statement follows from $H = V \oplus V^\perp$ in a Hilbert space H with a subspace V and the formulas $(\ker T)^\perp = \overline{\operatorname{ran} T^*}$ and $(\ker T^*)^\perp = \overline{\operatorname{ran} T}$. ∎

In computations later, we will need to know if two operators are equal. The concept of numerical range (in particular part (1) of the following theorem) will be useful in such situation.

Definitions. The *numerical range* of $T \in L(H)$ is $W(T) = \{(Tx, x) : \|x\| = 1\}$. The *numerical radius* of T is $\sup\{|(Tx, x)| : \|x\| = 1\}$.

Theorem. *Let $T, T_0, T_1 \in L(H)$.*

(1) $T = 0$ iff $W(T) = \{0\}$, i.e. $(Tx, x) = 0$ for all $x \in H$. $T_0 = T_1$ iff $(T_0x, x) = (T_1x, x)$ for all $x \in H$.

(2) $\sigma(T) \subseteq \overline{W(T)}$ and if the distance from c to $\overline{W(T)}$ is $d > 0$, then $\|(T - cI)^{-1}\| \le 1/d$.

Proof. (1) $T = 0$ implies $W(T) = \{0\}$ is trivial. For the converse, $W(T) = \{0\}$ implies for all $w \in H$, $(Tw, w) = 0$. For every $x \in H$, let $y = Tx$, then

$$\|Tx\|^2 = (Tx, y) = \frac{1}{4}\Big(\big(T(x+y), x+y\big) - \big(T(x-y), x-y\big) + i\big(T(x+iy), x+iy\big) - i\big(T(x-iy), x-iy\big)\Big) = 0.$$

(2) Let $c \notin \overline{W(T)}$. Then the distance from c to $\overline{W(T)}$ is $d > 0$. For $\|x\| = 1$, $\|(T - cI)x\| \geq |((T - cI)x, x)| = |(Tx, x) - c| \geq d > 0$ implies $T - cI$ is bounded below. By the lower bound theorem, $T - cI$ is injective and has closed range. Assume $\operatorname{ran}(T - cI)$ is not dense. Then $\ker(T^* - \overline{c}I) = (\operatorname{ran}(T - cI))^\perp \neq \{0\}$. So there is $\|v\| = 1$ such that $T^*v = \overline{c}v$. Then $c = (v, \overline{c}v) = (v, T^*v) = (Tv, v) \in W(T)$, a contradiction. Hence $\operatorname{ran}(T - cI)$ is dense. So $T - cI$ is invertible and $c \notin \sigma(T)$. From $T - cI$ bounded below, for $\|x\| = 1$, let $y = (T - cI)^{-1}x$, then $\|(T - cI)^{-1}x\| = \|y\| \leq \|(T - cI)(y)\|/d = \|x\|/d = 1/d$. So $\|(T - cI)^{-1}\| \leq 1/d$. ∎

<u>Exercise.</u> Prove the Toeplitz-Hausdorff theorem that asserts for every Hilbert space H and $T \in L(H)$, $W(T)$ is convex.

Recall the projection theorem asserts that for every closed subspace M of H, every $x \in H$ has a unique decomposition $x = y + z$, where $y \in M$ (is the closest point to x in M) and $z \in M^\perp$. The function $P_M : H \to M$ defined by $P_M(x) = y$ is a projection since $P_M^2 x = P_M y = y = P_M x$. Its kernel $M^\perp$ and its range M are orthogonal. If $M \neq \{0\}$, then $\|P_M\| = 1$. Note $P_{M^\perp} = I - P_M$ and $\ker P_M = M^\perp = \operatorname{ran} P_{M^\perp} = \operatorname{ran}(I - P_M)$.

<u>Definition.</u> A projection $P \in L(H)$ is <u>*orthogonal*</u> iff $\ker P \perp \operatorname{ran} P$. In that case, $P = P_M$, where $M = \operatorname{ran} P$.

<u>Theorem.</u> *For a nonzero projection P,* (a) *P is orthogonal,* (b) *$P^* = P$ and* (c) *$\|P\| = 1$ are equivalent.*

<u>Proof.</u> (a) $\Rightarrow$ (b) P is orthogonal implies $\operatorname{ran} P \perp \operatorname{ran}(I - P)$. So, for all $x \in H$, $0 = (Px, (I - P)x) = ((I - P^*)Px, x)$. So $W((I - P^*)P) = \{0\}$. Hence, $(I - P^*)P = 0$, i.e. $P = P^*P$. So $P^* = (P^*P)^* = P^*P^{**} = P^*P = P$.

(b) $\Rightarrow$ (c) $P^* = P$ implies $\|Px\|^2 = (Px, Px) = (P^*Px, x) = (P^2x, x) = (Px, x) \leq \|Px\|\|x\|$. So $\|Px\| \leq \|x\|$ with equality if $x \in \operatorname{ran} P$. Thus, $\|P\| = 1$.

(c) $\Rightarrow$ (a) Assume P is not orthogonal. Then there is $x \in \operatorname{ran} P$, $y \in \ker P$ such that $\|x\| = 1 = \|y\|$ and $(x, y) \neq 0$. Replacing x by $e^{i\theta}x$, we may assume $(x, y) = -t < 0$. Take $z = x + ty$. Then $\|z\|^2 = \|x\|^2 + 2t(x, y) + t^2\|y\|^2 = 1 - t^2 < 1 = \|x\|^2 = \|Pz\|^2$, which implies $\|P\| \neq 1$, contradiction. ∎

<u>Remark.</u> For an orthogonal projection P, in the last proof we saw $(Px, x) = \|Px\|^2$. This is useful.

<u>Theorem (Sum of Orthogonal Projections).</u> *Let E, F be orthogonal projections with ranges Y, Z, respectively. The following are equivalent:*

(a) $Y \perp Z$, (b) $E(Z) = \{0\}$, (c) $EF = 0$, (d) $F(Y) = \{0\}$ *and* (e) $FE = 0$.

Also $E + F$ is an orthogonal projection iff $Y \perp Z$, in which case $\operatorname{ran}(E + F) = Y + Z$ *is the closed linear span of $Y \cup Z$.*

<u>Proof.</u> $Y \perp Z \Leftrightarrow Z \subseteq Y^\perp = \ker E \Leftrightarrow E(Z) = E(\operatorname{ran} F) = \{0\} \Leftrightarrow E(Fx) = 0$ for all $x \in H \Leftrightarrow EF = 0$. Similarly $Z \perp Y \Leftrightarrow F(Y) = \{0\} \Leftrightarrow FE = 0$.

If $Y \perp Z$, then $(E+F)^2 = E^2 + EF + FE + F^2 = E + 0 + 0 + F = E + F$ and $(E+F)^* = E^* + F^* = E + F$, so $E + F$ is an orthogonal projection.

Conversely, $E + F$ is an orthogonal projection implies $\|E + F\| = 1$. So for $x \in Y = \operatorname{ran} E$,

$$\|x\|^2 \geq \|(E + F)x\|^2 = ((E + F)x, x) = (Ex, x) + (Fx, x) = \|Ex\|^2 + \|Fx\|^2 = \|x\|^2 + \|Fx\|^2.$$

So $F(Y) = \{0\}$, which is equivalent to $Y \perp Z$.

Finally, in case $E + F$ is an orthogonal projection, let $M = \overline{\operatorname{span}(Y \cup Z)}$. Since $(E+F)|_Y = E|_Y + 0 = I$ and similarly $(E + F)|_Z = I$, we have $(E + F)|_{\overline{\operatorname{span}(Y \cup Z)}} = I$. Then $M = \overline{\operatorname{span}(Y \cup Z)} \subseteq \operatorname{ran}(E + F) \subseteq Y + Z \subseteq M$. So $\operatorname{ran}(E + F) = Y + Z = M$. ∎

<u>Exercises.</u> Let E, F be orthogonal projections with ranges Y, Z, respectively.

(1) Prove that EF is an orthogonal projection iff $EF = FE$, in which case, $\operatorname{ran} EF = Y \cap Z$.

(2) Prove that the following are equivalent: (a) $Y \subseteq Z$, (b) $FE = E$, (c) $EF = E$, (d) $\|Ex\| \le \|Fx\|$ for all $x \in H$ and (e) $E \le F$. Then prove that $F - E$ is an orthogonal projection iff $Y \subseteq Z$, in which case $\operatorname{ran}(F - E) = Z \cap Y^{\perp}$.

§2. **Normal Operators.** Next we will study an important class of operators.

Definitions. Let $T \in L(H)$.

(1) T is *normal* iff $T^*T = TT^*$ (iff $(T^*Tx, x) = (TT^*x, x)$ iff $(Tx, Tx) = (T^*x, T^*x)$ iff $\|Tx\| = \|T^*x\|$ for all $x \in H$). So $\ker T = \ker T^*$ and $\overline{\operatorname{ran} T} = (\ker T^*)^{\perp} = (\ker T)^{\perp} = \overline{\operatorname{ran} T^*}$.

(2) T is *self-adjoint* (or *Hermitian*) iff $T = T^*$ (iff $(Tx, x) = (T^*x, x) = (x, Tx) = \overline{(Tx, x)}$, i.e. $(Tx, x) \in \mathbb{R}$ for all $x \in H$).

(3) T is *positive* (and we write $T \ge 0$) iff $(Tx, x) \ge 0$ for all $x \in H$ (which implies $T^* = T$). For self-adjoint operators A and B, define $A \le B$ (or $B \ge A$) iff $B - A \ge 0$.

(4) T is an *isometry* iff $I = T^*T$ (iff $(x, x) = (T^*Tx, x) = (Tx, Tx)$ iff $\|Tx\| = \|x\|$ for all $x \in H$). T is an *co-isometry* iff $TT^* = I$ iff T^* is an isometry.

(5) T is *unitary* iff $TT^* = I = T^*T$. (By (4), it is equivalent to an invertible isometry.) If $\mathbb{K} = \mathbb{R}$, unitary operators are also called *orthogonal* operators.

Other than isometry and co-isometry, these are all normal operators. Also, for orthogonal projection P, since $(Px, x) = \|Px\|^2 \ge 0$, they are positive, hence normal. Now we begin to study normal operators.

Theorem (Properties of Normal Operators). *Let $T \in L(H)$ be normal.*

(1) *For every $c \in \mathbb{C}$, $T - cI$ is normal. If T is invertible, then T^{-1} is normal.*

(2) *Eigenvectors for different eigenvalues of T are orthogonal, i.e. if $a \ne b$, $Tx = ax$ and $Ty = by$, then $(x, y) = 0$.*

(3) *T is invertible iff T is right invertible iff T is bounded below iff T is left invertible.*

(4) $\sigma(T) = \sigma_{ap}(T)$.

(5) *The spectral radius and the numerical radius both equal $\|T\|$.*

Proof. (1) $(T - cI)(T - cI)^* = (T - cI)(T^* - \bar{c}I) = TT^* - cT^* - \bar{c}T + |c|^2 I = T^*T - cT^* - \bar{c}T + |c|^2 I = (T^* - \bar{c}I)(T - cI) = (T - cI)^*(T - cI)$. For T invertible, since $(T^{-1})^* = (T^*)^{-1}$, so $T^{-1}(T^{-1})^* = T^{-1}(T^*)^{-1} = (T^*T)^{-1} = (TT^*)^{-1} = (T^*)^{-1}T^{-1} = (T^{-1})^*T^{-1}$.

(2) For normal T, $Ty = by$ iff $T^*y = \bar{b}y$ since $\|(T - bI)y\| = \|(T - bI)^*y\| = \|(T^* - \bar{b}I)y\|$. Then $a(x, y) = (Tx, y) = (x, T^*y) = (x, \bar{b}y) = b(x, y)$ and $a \ne b$ imply $(x, y) = 0$.

(3) The left inverse theorem asserts that an operator in $L(H)$ is left invertible iff it is bounded below. Now T is right invertible $\Leftrightarrow$ T^* is left invertible $\Leftrightarrow$ T^* is bounded below $\Leftrightarrow$ T is bounded below $\Leftrightarrow$ T is left invertible. Finally T invertible $\Rightarrow$ T is right invertible $\Rightarrow$ T is left and right invertible $\Rightarrow$ T is invertible.

(4) By (1) and (3), $c \notin \sigma(T)$ iff $T - cI$ is invertible iff $T - cI$ is bounded below iff $c \notin \sigma_{ap}(T)$.

(5) $\|T^2\| = \|(T^2)^*T^2\|^{1/2} = \|(T^*T)^*(T^*T)\|^{1/2} = \|T^*T\| = \|T\|^2$. Iterating this, we get $\|T^{2^n}\| = \|T\|^{2^n}$. Therefore, $r(T) = \lim_{n\to\infty} \|T^{2^n}\|^{1/2^n} = \|T\|$.

Next, since $\sigma(T)$ is compact, there is $c \in \sigma(T)$ with $|c| = r(T) = \|T\|$. By (4), there are $x_n \in H$ such that $\|x_n\| = 1$ and $\|(T - cI)x_n\| \to 0$. Since $\|(T - cI)x_n\| \ge |((T - cI)x_n, x_n)| = |(Tx_n, x_n) - c|$, so $(Tx_n, x_n) \to c$. Hence $\|T\| = |c| = \lim_{n\to 0} |(Tx_n, x_n)| \le \sup\{|(Tx, x)| : \|x\| = 1\} \le \|T\|$ and the numerical radius of T is $\|T\|$. ∎

Remark. It is known that for a normal operator, the closure of the numerical range is the convex hull of the spectrum. See [H], pp. 116 and 318.

Theorem. (1) *If T is self-adjoint, then $\sigma(T) \subseteq \mathbb{R}$.*

(2) *If $T \geq 0$, then $\sigma(T) \subseteq [0, +\infty)$.*

(3) *If T is unitary, then $\sigma(T) \subseteq \{z : |z| = 1\}$.*

(4) *If T is normal and $c \notin \sigma(T)$, then $\|(T - cI)^{-1}\| = 1/d$, where $d = \min\{|z - c| : z \in \sigma(T)\}$.*

Proof. (1) Since $(Tx, x) \in \mathbb{R}$, we get $\sigma(T) \subseteq \overline{W(T)} \subseteq \mathbb{R}$.

(2) Since $(Tx, x) \geq 0$, we get $\sigma(T) \subseteq \overline{W(T)} \subseteq [0, +\infty)$.

(3) For $|c| \neq 1$, since $\|Tx\| = \|x\|$, we have $\|(T - cI)x\| \geq \big|\|Tx\| - \|cx\|\big| = \big|1 - |c|\big|\|x\|$. Then $T - cI$ is normal and bounded below, hence invertible by property (3) of normal operators. So $\sigma(T) \subseteq \{z : |z| = 1\}$.

(4) Observe that if S is invertible, then $0 \notin \sigma(S)$ and $\sigma(S^{-1}) = \{w^{-1} : w \in \sigma(S)\}$. This follows from the identity $-w^{-1}S^{-1}(S - wI) = S^{-1} - w^{-1}I$ and the fact that $-w^{-1}S^{-1}$ is invertible. From this, letting $S = T - cI$, we have $\sigma((T - cI)^{-1}) = \{w^{-1} : w \in \sigma(T - cI)\} = \{(z - c)^{-1} : \ z \in \sigma(T)\}$.

Finally, by property (1) of normal operators, $(T - cI)^{-1}$ is normal and so

$$\|(T - cI)^{-1}\| = r\big((T - cI)^{-1}\big) = \max\{|z - c|^{-1} : \ z \in \sigma(T)\} = 1/\min\{|z - c| : \ z \in \sigma(T)\}. \qquad \square$$

Theorem (Properties of Self-adjoint Operators). *If T is self-adjoint, then*

(1) *either $\|T\|$ or $-\|T\|$ is in $\sigma(T)$,*

(2) $\sup \sigma(T) = \sup W(T), \inf \sigma(T) = \inf W(T)$ *and so* $\sigma(T) \subseteq [\inf \sigma(T), \sup \sigma(T)] = [\inf W(T), \sup W(T)]$ *(in particular, $m = \inf W(T)$ and $M = \sup W(T)$ are in $\sigma(T) = \sigma_{ap}(T)$),*

(3) *$T \geq 0$ iff $\sigma(T) \subseteq [0, +\infty)$.*

Proof. (1) By property (5) of normal operators, $r(T) = \|T\|$. Since $\sigma(T) \subseteq \mathbb{R}$ and $\{z \in \mathbb{C} : |z| = r(T)\}$ intersects $\sigma(T)$, so either $\|T\|$ or $-\|T\|$ is in $\sigma(T)$.

(2) Let $M = \sup W(T) = \sup\{(Tx, x) : \|x\| = 1\}$ and $M' = \sup \sigma(T) = \sup\{c : c \in \sigma(T)\}$. Now $S = \|T\|I + T$ is positive (as $\big((\|T\|I + T)x, x\big) = \|T\|\|x\|^2 + (Tx, x) \geq 0)$ and self-adjoint. Now $W(S) = \{\|T\| + (Tx, x) : \|x\| = 1\} \subseteq [0, +\infty)$ and $\sigma(S) = \{\|T\| + c : c \in \sigma(T)\} \subseteq [0, +\infty)$. By property (5) of normal operators, the numerical radius of S and the spectral radius of S are equal. So $\|T\| + M = \|T\| + M'$. Hence $M = M'$. Applying a similar argument to $\|T\|I - T$, we see the infima are the same.

(3) The only-if direction follows from part (2) of the last theorem. For the if-direction, since $\sigma(T) \subseteq [0, +\infty)$, so by (2), $\inf W(T) = \inf \sigma(T) \geq 0$. Then $W(T) \subseteq [0, +\infty)$, which implies $T \geq 0$. $\square$

Definitions. Let $T \in L(H)$ and M be a subspace of H. We say M is *invariant* under T iff $T(M) \subseteq M$. Also, M *reduces* T iff $T(M) \subseteq M$ and $T(M^\perp) \subseteq M^\perp$.

Lemma. *Let $T \in L(H)$, M be a subspace of H and P be the orthogonal projection onto M.*

(1) *$T(M) \subseteq M$ iff $PTP = TP$ iff $T^*(M^\perp) \subseteq M^\perp$. $T(M^\perp) \subseteq M^\perp$ iff $PTP = PT$ iff $T^*(M) \subseteq M$.*

(2) *M reduces T iff $PT = TP$ iff M reduces T^*.*

Proof. (1) For $x \in H$, write $x = y + y'$, $Ty = z + z'$, where $y, z \in M$, $y', z' \in M^\perp$. We have $PTPx = PTy = z$ and $TPx = Ty = z + z'$. So $PTP = TP \Leftrightarrow TPx = z \in M$ for all $x \in H \Leftrightarrow T(M) = TP(H) \subseteq M$.

Next $I - P$ is the orthogonal projection onto $M^\perp$. So $T^*(M^\perp) \subseteq M^\perp$ iff $(I - P)T^*(I - P) = T^*(I - P)$ iff $PT^*P = PT^*$ iff $PTP = TP$ iff $T(M) \subseteq M$. The second statement is similar.

(2) follows from (1) by combining the two statements.

Remarks. For all $x, y \in M$, $(x, (T|_M)^*y) = (T|_M x, y) = (Tx, y) = (x, T^*y) = (x, T^*|_M y)$. So $(T|_M)^* = T^*|_M$. Similarly, $(T|_{M^\perp})^* = T^*|_{M^\perp}$.

Theorem (Properties of Normal Operators). *Let $T \in L(H)$ be normal and M a subspace of H.*

(6) *For every $c \in \mathbb{C}$, $\ker(T - cI)$ reduces T (and hence also T^*).*

(7) *If M reduces T, then $T|_M$, $T|_{M^\perp}$ and their adjoints are normal and $\|T\| = \max\{\|T|_M\|, \|T|_{M^\perp}\|\}$.*

Proof. (6) For $x \in \ker(T - cI)$, $(T - cI)Tx = T(T - cI)x = 0$ implies $Tx \in \ker(T - cI)$. Similarly, $(T - cI)T^*x = T^*(T - cI)x = 0$ implies $T^*x \in \ker(T - cI)$. By the lemma, $\ker(T - cI)$ reduces T and T^*.

(7) Using the remark, $T|_M(T|_M)^* = T|_M T^*|_M = (TT^*)|_M = (T^*T)|_M = T^*|_M T|_M = (T|_M)^* T|_M$. So $T|_M$ and $T^*|_M$ are normal. Since $M^\perp$ also reduces T, similarly $T|_{M^\perp}$ and $T^*|_{M^\perp}$ are normal.

Next, let $A = \max\{\|T|_M\|, \|T|_{M^\perp}\|\}$. Clearly $\|T|_M\|, \|T|_{M^\perp}\| \le \|T\|$. So $A \le \|T\|$. For the reverse inequality, write $x = y + z$, where $y \in M$ and $z \in M^\perp$. Then $\|x\|^2 = \|y\|^2 + \|z\|^2$. Since M reduces T, so $Ty \in M$, $Tz \in M^\perp$. Then $\|Tx\|^2 = \|Ty\|^2 + \|Tz\|^2 \le \|T|_M\|^2\|y\|^2 + \|T|_{M^\perp}\|^2\|z\|^2 \le A^2\|x\|^2$. So $\|T\| \le A$. ◻

Spectral Theorem for Compact Normal Operators. *Let $T \in L(H)$ be a compact normal operator. For an eigenvalue c of T, let P_c denote the orthogonal projection onto $H_c = \ker(T - cI)$. As $\sigma(T)$ is a countable set with 0 as the only possible accumulation point, let its nonzero elements be $c_1, c_2, c_3, \ldots$ arranged so that $|c_1| \ge |c_2| \ge |c_3| \ge \cdots$. Then $T = \sum_i c_i P_{c_i}$ (where the series converges in the norm of $L(H)$ if there are infinitely many terms) and H has an orthonormal basis consisting of eigenvectors of T.*

Proof. Since T is compact, the H_c's ($c \ne 0$) are finite dimensional by the Riesz-Fredholm lemma. Since T is normal, by property (2) of normal operators, the H_c's (for all $c \in \sigma(T)$) are pairwise orthogonal. By the theorem on sum of orthogonal projections, (*) $P_c P_{c'} = 0 = P_{c'} P_c$ if $c \ne c'$.

For every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $k > N$ implies $|c_k| < \varepsilon$, i.e. $\sigma(T) \setminus \{c_1, c_2, \ldots, c_N\} \subseteq B(0, \varepsilon)$. Let $M = \sum_{i=1}^{N} H_{c_i}$ and $T_N = \sum_{i=1}^{N} c_i P_{c_i}$. By (*), $T_N P_c = P_c T_N$. By property (6) of normal operators, $TP_c = P_c T$. Hence, $T - T_N$ is normal since T, T^* commute with P_c's and T_N, T_N^* are in the span of P_c's. Since H_c's are pairwise orthogonal, $P_M = \sum_{i=1}^{N} P_{c_i}$ by the theorem on sum of orthogonal projections. Then M reduces T and T_N. Note $T|_M = T_N|_M$ (as $T|_M(v_i) = c_i v_i = T_N|_M(v_i)$ for $v_i \in H_{c_i}$ with $i \le n$) and $T_N|_{M^\perp} = 0$ (as $v \in M^\perp$ implies $v \perp H_{c_i}$ for $i \le n$ and so these $P_{c_i}(v) = 0$). By property (7) of normal operators and last sentence,

$$\|T - T_N\| = \max\{\|T|_M - T_N|_M\|, \|T|_{M^\perp} - T_N|_{M^\perp}\|\} = \|T|_{M^\perp}\|.$$

By property (7) of normal operators and properties (f) of compact operators, $T|_{M^\perp}$ is also a compact normal operator. Now the eigenvalues of $T|_{M^\perp}$ are in $\sigma(T) \setminus \{c_1, c_2, \ldots, c_N\}$. By property (5) of normal operators, $\|T - T_N\| = \|T|_{M^\perp}\| = r(T|_{M^\perp}) < \varepsilon$. So T is the limit of T_N in the norm of $L(H)$, i.e. $T = \sum_i c_i P_{c_i}$.

Let H' be the closed linear span of the union of <u>all</u> H_c's, where $c \in \sigma(T)$. Since H_c's reduce T for all $c \in \sigma(T)$, so H' reduces T. By property (7) of normal operators, $T|_{H'^\perp}$ is compact normal and cannot have any nonzero eigenvalues by the definition of H'. So $\sigma(T|_{H'^\perp}) = \{0\}$ and $\|T|_{H'^\perp}\| = r(T|_{H'^\perp}) = 0$. Then $H'^\perp \subseteq \ker T = H_0$. By the definition of H', $H'^\perp \cap H_0 = \{0\}$. So $H'^\perp = \{0\}$, i.e. $H' = H$. Taking an orthonormal basis in every H_c ($c \in \sigma(T)$), their union is complete, hence is an orthonormal basis of H. ◻

Remark. The compact self-adjoint case of the spectral theorem is called the Hilbert-Schmidt theorem.

Simultaneous Diagonalization Therorem. *Let $T_1, T_2 \in L(H)$ be compact normal operators such that $T_1T_2 = T_2T_1$. Then H has an orthonormal basis B consisted of common eigenvectors of T_1 and T_2.*

(By the last theorem, the matrices of T_1, T_2 are diagonal. By induction, the same result also holds for finitely many pairwise commuting compact normal operators. In particular, this is true for commuting normal operators on finite dimensional vector spaces since all operators are finite rank, hence compact.)

<u>Proof.</u> Apply the spectral theorem to T_1. Then H is the closed linear span of <u>all</u> $H_c = \ker(T_1 - cI)$, where $c \in \sigma(T_1)$. From $x \in H_c$ implies $(T_1 - cI)T_2x = T_2(T_1 - cI)x = 0$, we get $T_2(H_c) \subseteq H_c$. Also, $H_c^\perp = \sum_{c' \neq c} H_{c'}$ is invariant under T_2. So H_c reduces T_2, hence $T_2|_{H_c}$ is normal. Applying the spectral theorem to $T_2|_{H_c}$, we get an orthonormal basis of H_c consisting of eigenvectors of T_2 (which are also eigenvectors of T_1 as they are in H_c). The union of these orthonormal bases of H_c is a desired orthonormal basis for H. $\square$

<u>Tensor Notations for Rank One Operators.</u> For $v, e \in H$, define the linear functional $e \otimes v$ on H by $(e \otimes v)(x) = (x, v)e$. If $v, e \neq 0$, then it is a rank one operator since its range is the span of $\{e\}$.

<u>Theorem.</u> *Every rank n operator $F \in L(H)$ is the sum of n rank one operators.*

<u>Proof.</u> Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of $\operatorname{ran} F$. Since $e_i \in \operatorname{ran} F$, $g_i(x) = (F(x), e_i)$ is a nonzero element of H^*. By the Riesz representation theorem, there is a nonzero $v_i \in H$ such that $g_i(x) = (x, v_i)$. Then $F(x) = \sum_{i=1}^{n} (F(x), e_i)e_i = \sum_{i=1}^{n} g_i(x)e_i = \sum_{i=1}^{n} (x, v_i)e_i$, i.e. $F = \sum_{i=1}^{n} e_i \otimes v_i$. $\square$

<u>Theorem.</u> *Let $T \in L(H)$ be a compact operator. Then there are countable orthonormal sets $\{e_i\}$ and $\{v_i\}$ in H and positive real numbers $\{c_i\}$ (converging to 0 if infinitely many i's) such that for all $x \in H$, $Tx = \sum_i c_i(x, v_i)e_i$. ($\sum_i c_i(e_i \otimes v_i)$ is called the <u>Schmidt representation</u> of T. The c_i's are called the <u>singular values</u> of T.) In particular, every compact operators on a Hilbert space is the limit of finite rank operators.*

<u>Proof.</u> Since T is compact, $S = T^*T$ is a positive compact operator. By the spectral theorem for compact normal operators, $S = \sum_{a \in \sigma(S) \setminus \{0\}} aP_a$. Let <u>sequence</u> $\{v_i\}$ be the union of the orthonormal bases of $\ker(S - aI)$ for all $a \in \sigma(S) \setminus \{0\}$. So every v_i is the eigenvector of some $a \in \sigma(S) \setminus \{0\} \subseteq (0, +\infty)$. Let $c_i = \sqrt{a}$ and let $e_i = (Tv_i)/c_i$. If $\sigma(S)$ is infinite, we may arrange the a's to go to 0, then c_i's will also go to 0.

For $i \neq j$, $(Tv_i, Tv_j) = (Sv_i, v_j) = a(v_i, v_j) = 0$, Also, $(Tv_i, Tv_i) = (Sv_i, v_i) = a(v_i, v_i) = c_i^2$, which implies $\|Tv_i\| = c_i$, so $\|e_i\| = 1$. Hence, $\{e_i\}$ is an orthonormal set.

For all $x \in H$, we now check $Tx = \sum_i c_i(x, v_i)e_i$. On $\operatorname{span}\{v_i\}$, it holds since $Tv_j = c_je_j = \sum_i c_i(v_j, v_i)e_i$. On $(\operatorname{span}\{v_i\})^\perp$, $x \in (\operatorname{span}\{v_i\})^\perp$ implies for all $a \in \sigma(S) \setminus \{0\}$, $x \in (\ker(S - aI))^\perp$ as $\ker(S - aI) \subseteq \operatorname{span}\{v_i\}$. Hence, all $P_ax = 0$ and so $Sx = 0$. Then $\|Tx\|^2 = (Sx, x) = 0 = \sum_i c_i(x, v_i)e_i)$. $\square$