Introduction to Functional Analysis

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

Subsection 1

Zorn's Lemma

Partially Ordered Sets

Definition (Partially Ordered Set)

A **partially ordered set** is a set M on which there is defined a partial ordering, that is, a binary relation which is written $\leq$ satisfying:

(PO1) $a \leq a$, for every $a \in M$; (**Reflexivity**)

(PO2) If $a \leq b$ and $b \leq a$, the $a = b$; (**Antisymmetry**)

(PO3) If $a \leq b$ and $b \leq c$, then $a \leq c$. (**Transitivity**)

- Elements a and b for which neither $a \leq b$ nor $b \leq a$ holds are called **incomparable elements**.
- In contrast, two elements a and b are called **comparable elements** if they satisfy $a \leq b$ or $b \leq a$ (or both).

Chains and Bounds

Definition (Totally Ordered Set or Chain)

A **totally ordered set** or **chain** is a partially ordered set such that every two elements of the set are comparable. In other words, a chain is a partially ordered set that has no incomparable elements.

Definition (Upper Bound and Maximal Element)

An **upper bound** of a subset W of a partially ordered set M is an element $u \in M$, such that $x \leq u$, for every $x \in W$. (Depending on M and W, such a u may or may not exist.)
A **maximal element** of M is an $m \in M$, such that $m \leq x$ implies $m = x$. (Again, M may or may not have maximal elements.)

- Note that a maximal element need not be an upper bound.

Examples of Partial Orderings

- **Real numbers**: Let M be the set of all real numbers and let $x \le y$ have its usual meaning. M is totally ordered. M has no maximal elements.
- **Power set**: Let $\mathscr{P}(X)$ be the power set (set of all subsets) of a given set X and let $A \le B$ mean $A \subseteq B$, that is, A is a subset of B. Then $\mathscr{P}(X)$ is partially ordered. The only maximal element of $\mathscr{P}(X)$ is X.
- ***n*-tuples of numbers**: Let M be the set of all ordered n-tuples $x = (\xi_1, \ldots, \xi_n)$, $y = (\eta_1, \ldots, \eta_n)$, ..., of real numbers and let $x \le y$ mean $\xi_j \le \eta_j$, for every $j = 1, \ldots, n$, where $\xi_j \le \eta_j$ has its usual meaning. This defines a partial ordering on M.
- **Positive integers**: Let $M = \mathbb{N}$, the set of all positive integers. Let $m \le n$ mean that m divides n. This defines a partial ordering on $\mathbb{N}$.

Zorn's Lemma

Zorn's Lemma

Let $M \neq \emptyset$ be a partially ordered set. Suppose that every chain $C \subseteq M$ has an upper bound. Then M has at least one maximal element.

- Zorn's Lemma can be derived from the **Axiom of Choice**:

  For any given set E, there exists a mapping c ("**choice function**") from the power set $\mathscr{P}(E)$ into E, such that if $B \subseteq E$, $B \neq \emptyset$, then $c(B) \in B$.
- Conversely, the Axiom of Choice follows from Zorn's Lemma.
- So Zorn's Lemma and the Axiom of Choice can be regarded as equivalent axioms.

Application: Hamel Bases

Hamel Basis

Every vector space $X \neq \{0\}$ has a Hamel basis.

- Let M be the set of all linearly independent subsets of X. Since $X \neq \{0\}$, it has an element $x \neq 0$ and $\{x\} \in M$, so that $M \neq \emptyset$. Set inclusion defines a partial ordering on M. Every chain $C \subseteq M$ has an upper bound, namely, the union of all subsets of X which are elements of C. By Zorn's Lemma, M has a maximal element B.

  Claim: B is a Hamel basis for X.

  Let $Y = \text{span} B$. Then Y is a subspace of X. Moreover, $Y = X$: Otherwise, $B \cup \{z\}$, $z \in X$, $z \notin Y$, would be a linearly independent set containing B as a proper subset. And this would contradict the maximality of B.

Application: Total Orthonormal Sets

Total Orthonormal Set

In every Hilbert space $H \neq \{0\}$ there exists a total orthonormal set.

- Let M be the set of all orthonormal subsets of H. Since $H \neq \{0\}$, it has an element $x \neq 0$, and an orthonormal subset of H is $\{y\}$, where $y = \frac{1}{\|x\|}x$. Hence $M \neq \emptyset$. Set inclusion defines a partial ordering on M. Every chain $C \subseteq M$ has an upper bound, namely, the union of all subsets of X which are elements of C. By Zorn's Lemma, M has a maximal element F.

  Claim: F is total in H.

  Suppose that this is false. Then, by a previous theorem, there exists a nonzero $z \in H$, such that $z \perp F$. Hence $F_1 = F \cup \{e\}$, where $e = \frac{1}{\|z\|}z$ is orthonormal, and F is a proper subset of F_1. This contradicts the maximality of F.

Subsection 2

Hahn-Banach Theorem

Sublinear Functionals

- A **sublinear functional** is a real-valued functional p on a vector space X which is:
  - **subadditive**, that is,
    $$p(x+y) \le p(x) + p(y), \text{ for all } x, y \in X;$$
  - **positive-homogeneous**, that is,
    $$p(\alpha x) = \alpha p(x), \text{ for all } \alpha \ge 0 \text{ in } \mathbb{R} \text{ and } x \in X.$$
- Note that the norm on a normed space is such a functional.

Idea of the Hahn-Banach Theorem

- In an **extension problem** one considers a mathematical object (e.g., a mapping) defined on a subset Z of a given set X and the goal is to extend the object from Z to the entire set X in such a way that certain basic properties of the object continue to hold for the extended object.
- In the Hahn-Banach theorem, the object to be extended is a linear functional f which is defined on a subspace Z of a vector space X and has a certain boundedness property which will be formulated in terms of a sublinear functional.
  - We assume that the functional f to be extended is majorized on Z by such a functional p defined on X.
  - We extend f from Z to X without losing the linearity and the majorization, so that the extended functional $\widetilde{f}$ on X is still linear and still majorized by p.

Hahn-Banach Theorem (Extension of Linear Functionals)

Hahn-Banach Theorem (Extension of Linear Functionals)

Let X be a real vector space and p a sublinear functional on X. Furthermore, let f be a linear functional which is defined on a subspace Z of X and satisfies $f(x) \le p(x)$, for all $x \in Z$. Then f has a linear extension $\widetilde{f}$ from Z to X satisfying $\widetilde{f}(x) \le p(x)$, for all $x \in X$, that is, $\widetilde{f}$ is a linear functional on X, satisfies $\widetilde{f}(x) \le p(x)$ on X and $\widetilde{f}(x) = f(x)$, for every $x \in Z$.

- Proceeding stepwise, we shall prove:
  - (a) The set E of all linear extensions g of f satisfying $g(x) \le p(x)$ on their domain $\mathscr{D}(g)$ can be partially ordered and Zorn's Lemma yields a maximal element $\widetilde{f}$ of E.
  - (b) $\widetilde{f}$ is defined on the entire space X.
  - (c) An auxiliary relation which was used in (b).

Proof of Part (a)

(a) Let E be the set of all linear extensions g of f which satisfy the condition $g(x) \le p(x)$, for all $x \in \mathscr{D}(g)$.

Clearly, $E \ne \emptyset$ since $f \in E$.

On E we can define a partial ordering by $g \le h$ iff h is an extension of g, i.e., by definition, $\mathscr{D}(h) \supseteq \mathscr{D}(g)$ and $h(x) = g(x)$, for every $x \in \mathscr{D}(g)$.
For a chain $C \subseteq E$, we define $\widehat{g}$ by $\widehat{g}(x) = g(x)$, if $x \in \mathscr{D}(g)$, $g \in C$.

- $\widehat{g}$ is a linear functional, the domain being $\mathscr{D}(\widehat{g}) = \bigcup_{g \in C} \mathscr{D}(g)$, which is a vector space since C is a chain.
- The definition of $\widehat{g}$ is unambiguous: For an $x \in \mathscr{D}(g_1) \cap \mathscr{D}(g_2)$, with $g_1, g_2 \in C$, we have $g_1(x) = g_2(x)$, since C is a chain, so that $g_1 \le g_2$ or $g_2 \le g_1$.
- Clearly, $g \le \widehat{g}$ for all $g \in C$. Hence $\widehat{g}$ is an upper bound of C.

Since $C \subseteq E$ was arbitrary, by Zorn's Lemma, E has a maximal element $\widetilde{f}$.

By the definition of E, this is a linear extension of f, which satisfies $\widetilde{f}(x) \le p(x)$, for all $x \in \mathscr{D}(\widetilde{f})$.

Proof of Part (b)

(b) We show that $\mathscr{D}(\widetilde{f})$ is all of X.

Suppose that this is false. Then we can choose a $y_1 \in X - \mathscr{D}(\widetilde{f})$. Consider the subspace Y_1 of X spanned by $\mathscr{D}(\widetilde{f})$ and y_1. Note that $y_1 \neq 0$ since $0 \in \mathscr{D}(\widetilde{f})$. Any $x \in Y_1$ can be written $x = y + \alpha y_1$, $y \in \mathscr{D}(\widetilde{f})$. This representation is unique: In fact, $y + \alpha y_1 = \widetilde{y} + \beta y_1$ with $\widetilde{y} \in \mathscr{D}(\widetilde{f})$ implies $y - \widetilde{y} = (\beta - \alpha) y_1$, where $y - \widetilde{y} \in \mathscr{D}(\widetilde{f})$, whereas $y_1 \notin \mathscr{D}(\widetilde{f})$. The only solution is $y - \widetilde{y} = 0$ and $\beta - \alpha = 0$. This means uniqueness.

A functional g_1 on Y_1 is defined by

$$g_1(y + \alpha y_1) = \widetilde{f}(y) + \alpha c,$$

where c is any real constant. It is not difficult to see that g_1 is linear. Furthermore, for $\alpha = 0$, we have $g_1(y) = \widetilde{f}(y)$. Hence g_1 is a proper extension of $\widetilde{f}$, i.e., an extension such that $\mathscr{D}(\widetilde{f})$ is a proper subset of $\mathscr{D}(g_1)$. Consequently, if we show $g_1 \in E$ by showing that $g_1(x) \leq p(x)$, contradicting the maximality of $\widetilde{f}$, we get $\mathscr{D}(\widetilde{f}) = X$ is true.

Proof of Part (c)

(c) It now suffices to show that g_1, with a suitable c in $g_1(y+\alpha y_1)=\widetilde{f}(y)+\alpha c$, satisfies $g_1(x)\le p(x)$, for all $x\in\mathscr{D}(g_1)$.

We consider any y and z in $\mathscr{D}(\widetilde{f})$. We obtain $\widetilde{f}(y)-\widetilde{f}(z)=\widetilde{f}(y-z)\le p(y-z)=p(y+y_1-y_1-z)\le p(y+y_1)+p(-y_1-z)$.

Binging the last term to the left and the term $\widetilde{f}(y)$ to the right,

$$-p(-y_1-z)-\widetilde{f}(z)\le p(y+y_1)-\widetilde{f}(y),$$

where y_1 is fixed. Since y does not appear on the left and z not on the right, the inequality continues to hold if we take the supremum over $z\in\mathscr{D}(\widetilde{f})$ on the left (call it m_0) and the infimum over $y\in\mathscr{D}(\widetilde{f})$ on the right, call it m_1. Then $m_0\le m_1$ and for a c, with $m_0\le c\le m_1$, we have

$$\begin{aligned}-p(-y_1-z)-\widetilde{f}(z)&\le c, &&\text{for all } z\in\mathscr{D}(\widetilde{f}),\\ c&\le p(y+y_1)-\widetilde{f}(y), &&\text{for all } y\in\mathscr{D}(\widetilde{f}).\end{aligned}$$

Proof of Part (c) (Cont'd)

- We prove $g_1(x) \le p(x)$, for all $x \in \mathscr{D}(g_1)$, first for negative α in $g_1(y+\alpha y_1) = \widetilde{f}(y) + \alpha c$ and then for positive α.
  - For $\alpha < 0$ we use $-p(-y_1 - z) - \widetilde{f}(z) \le c$, with z replaced by $\frac{1}{\alpha}y$, that is, $-p(-y_1 - \frac{1}{\alpha}y) - \widetilde{f}(\frac{1}{\alpha}y) \le c$. Multiplication by $-\alpha > 0$ gives $\alpha p(-y_1 - \frac{1}{\alpha}y) + \widetilde{f}(y) \le -\alpha c$. From this and $g_1(y+\alpha y_1) = \widetilde{f}(y) + \alpha c$, using $y + \alpha y_1 = x$, we obtain

$$g_1(x) = \widetilde{f}(y) + \alpha c \le -\alpha p\left(-y_1 - \frac{1}{\alpha}y\right) = p(\alpha y_1 + y) = p(x).$$

  - For $\alpha = 0$, we have $x \in \mathscr{D}(\widetilde{f})$ and nothing to prove.
  - For $\alpha > 0$ we use $c \le p(y + y_1) - \widetilde{f}(y)$, with y replaced by $\frac{1}{\alpha}y$ to get $c \le p(\frac{1}{\alpha}y + y_1) - \widetilde{f}(\frac{1}{\alpha}y)$. Multiplication by $\alpha > 0$ gives

$$\alpha c \le \alpha p\left(\frac{1}{\alpha}y + y_1\right) - \widetilde{f}(y) = p(x) - \widetilde{f}(y).$$

  But $g_1(y + \alpha y_1) = \widetilde{f}(y) + \alpha c$. So $g_1(x) = \widetilde{f}(y) + \alpha c \le p(x)$.

Necessity of Zorn's Lemma

- If, in $g_1(y+\alpha y_1) = \widetilde{f}(y) + \alpha c$, we take f instead of $\widetilde{f}$, we obtain, for each real c, a linear extension g_1 of f to the subspace Z_1 spanned by $\mathscr{D}(f) \cup \{y_1\}$. We can choose c so that $g_1(x) \le p(x)$, for all $x \in Z_1$, as may be seen from part (c) of the proof with $\widetilde{f}$ replaced by f.

  If $X = Z_1$, we are done.

  If $X \neq Z_1$, we may take a $y_2 \in X - Z_1$ and repeat the process to extend f to Z_2 spanned by Z_1 and y_2, etc.

  This gives a sequence of subspaces Z_j, each containing the preceding, and such that f can be extended linearly from one to the next and the extension g_j satisfies $g_j(x) \le p(x)$, for all $x \in Z_j$.
  - If $X = \bigcup_{j=1}^{n} Z_j$, we are done after n steps.
  - If $X = \bigcup_{j=1}^{\infty} Z_j$, we can use ordinary induction.
  - If X has no such representation, we do need Zorn's lemma in the proof presented here.

Subsection 3

Hahn-Banach Theorem for Complex and Normed Spaces

Hahn-Banach Theorem (Generalized)

Hahn-Banach Theorem (Generalized)

Let X be a real or complex vector space and p a real-valued functional on X which is subadditive, i.e., for all $x, y \in X$, $p(x+y) \leq p(x) + p(y)$, and for every scalar α satisfies $p(\alpha x) = |\alpha| p(x)$. Furthermore, let f be a linear functional which is defined on a subspace Z of X and satisfies $|f(x)| \leq p(x)$, for all $x \in Z$. Then f has a linear extension $\widetilde{f}$ from Z to X satisfying $|\widetilde{f}(x)| \leq p(x)$, for all $x \in X$.

(a) **Real vector space** If X is real, then $|f(x)| \leq p(x)$ implies $f(x) \leq p(x)$, for all $x \in Z$. Hence, by the Hahn-Banach Theorem, there is a linear extension $\widetilde{f}$ from Z to X, such that $\widetilde{f}(x) \leq p(x)$, for all $x \in X$. From this and the hypothesis, we obtain

$$-\widetilde{f}(x) = \widetilde{f}(-x) \leq p(-x) = |-1| p(x) = p(x).$$

That is, $\widetilde{f}(x) \geq -p(x)$. With $\widetilde{f}(x) \leq p(x)$, this yields the conclusion.

Hahn-Banach Theorem (The Complex Case)

(b) **Complex vector space** Let X be complex. Then Z is a complex vector space, too. Hence f is complex-valued, and we can write

$$f(x) = f_1(x) + if_2(x), \quad x \in Z,$$

where f_1 and f_2 are real-valued. For a moment we regard X and Z as real vector spaces and denote them by X_r and Z_r, respectively. This simply means that we restrict multiplication by scalars to real numbers (instead of complex numbers). Since f is linear on Z and f_1 and f_2 are real-valued, f_1 and f_2 are linear functionals on Z_r. Also $f_1(x) \le |f(x)|$ because the real part of a complex number cannot exceed the absolute value. Hence by $|f(x)| \le p(x)$, we get $f_1(x) \le p(x)$, for all $x \in Z_r$. By the Hahn-Banach Theorem, there is a linear extension $\widetilde{f}_1$ of f_1 from Z_r to X_r, such that $\widetilde{f}_1(x) \le p(x)$, for all $x \in X_r$.

Hahn-Banach Theorem (The Complex Case Cont'd)

- Considering Z and using $f = f_1 + if_2$, we have, for every $x \in Z$,

$$i[f_1(x) + if_2(x)] = if(x) = f(ix) = f_1(ix) + if_2(ix).$$

Equating the real parts, $f_2(x) = -f_1(ix)$, for all $x \in Z$. For all $x \in X$, we set

$$\widetilde{f}(x) = \widetilde{f}_1(x) - i\widetilde{f}_1(x), \quad x \in X.$$

Then $\widetilde{f}(x) = f(x)$ on Z. This shows that $\widetilde{f}$ is an extension of f from Z to X. We must now prove that:

(i) $\widetilde{f}$ is a linear functional on the *complex* vector space X;
(ii) $\widetilde{f}$ satisfies $|\widetilde{f}(x)| \le p(x)$ on X.

Hahn-Banach Theorem (Proving (i) and (ii))

$$\widetilde{f}(x) = \widetilde{f}_1(x) - i\widetilde{f}_1(x), \quad x \in X.$$

(i) Using $\widetilde{f}(x) = \widetilde{f}_1(x) - i\widetilde{f}_1(ix)$, $x \in X$, and the linearity of $\widetilde{f}_1$ on the real vector space X_r, we get

$$\begin{aligned} \widetilde{f}((a+ib)x) &= \widetilde{f}_1(ax+ibx) - i\widetilde{f}_1(iax - bx) \\ &= a\widetilde{f}_1(x) + b\widetilde{f}_1(ix) - i[a\widetilde{f}_1(ix) - b\widetilde{f}_1(x)] \\ &= (a+ib)[\widetilde{f}_1(x) - i\widetilde{f}_1(ix)] \\ &= (a+ib)\widetilde{f}(x). \end{aligned}$$

(ii) For any x, such that $\widetilde{f}(x) = 0$, (ii) holds since $p(x) \geq 0$. Let x be such that $\widetilde{f}(x) \neq 0$. Then we can write, using the polar form of complex quantities, $\widetilde{f}(x) = |\widetilde{f}(x)|e^{i\theta}$. Thus, $|\widetilde{f}(x)| = \widetilde{f}(x)e^{-i\theta} = \widetilde{f}(e^{-i\theta}x)$. Since $|\widetilde{f}(x)|$ is real, the last expression is real. So it is equal to its real part. Hence $|\widetilde{f}(x)| = \widetilde{f}(e^{-i\theta}x) = \widetilde{f}_1(e^{-i\theta}x) \leq p(e^{-i\theta}x) = |e^{-i\theta}|p(x) = p(x)$.

Hahn-Banach Theorem (Normed Spaces)

Hahn-Banach Theorem (Normed Spaces)

Let f be a bounded linear functional on a subspace Z of a normed space X. Then there exists a bounded linear functional $\widetilde{f}$ on X which is an extension of f to X and has the same norm, $\|\widetilde{f}\|_X = \|f\|_Z$, where

$$\|\widetilde{f}\|_X = \sup_{\substack{x\in X\\ \|x\|=1}} |\widetilde{f}(x)|, \quad \|f\|_Z = \sup_{\substack{x\in Z\\ \|x\|=1}} |f(x)|$$

(and $\|f\|_Z = 0$ in the trivial case $Z = \{0\}$).

- If $Z = \{0\}$, then $\widetilde{f} = 0$. The extension is $\widetilde{f} = 0$.
  Let $Z \neq \{0\}$. To use the theorem, we must first discover a suitable p. For all $x \in Z$, we have $|f(x)| \leq \|f\|_Z \|x\|$. This is of the right form, where $p(x) = \|f\|_Z \|x\|$. We see that p is defined on all of X.
  - p satisfies subadditivity on X:
    $p(x+y) = \|f\|_Z \|x+y\| \leq \|f\|_Z (\|x\| + \|y\|) = p(x) + p(y)$.
  - p satisfies the scalar property on X:
    $p(\alpha x) = \|f\|_Z \|\alpha x\| = |\alpha| \|f\|_Z \|x\| = |\alpha| p(x)$.

Hahn-Banach Theorem (Normed Spaces Cont'd)

- Hence, we can now apply the theorem and conclude that there exists a linear functional $\widetilde{f}$ on X which is an extension of f and satisfies

$$|\widetilde{f}(x)| \le p(x) = \|f\|_Z \|x\|, \quad x \in X.$$

  - Taking the supremum over all $x \in X$ of norm 1, we obtain the inequality

$$\|\widetilde{f}\|_X = \sup_{\substack{x \in X \\ \|x\|=1}} |\widetilde{f}(x)| \le \|f\|_Z.$$

  - Since under an extension the norm cannot decrease, we also have

$$\|\widetilde{f}\|_X \ge \|f\|_Z.$$

Together we obtain $\|\widetilde{f}\|_X = \|f\|_Z$.

The Case of Hilbert Spaces

- If Z is a closed subspace of a Hilbert space $X = H$, then f has a Riesz representation, say,
$$f(x) = \langle x, z \rangle, \quad z \in Z,$$
where $\|z\| = \|f\|$.
- Since the inner product is defined on all of H, this gives at once a linear extension $\widetilde{f}$ of f from Z to H.
- Moreover, $\widetilde{f}$ has the same norm as f because $\|\widetilde{f}\| = \|z\| = \|f\|$ by a preceding theorem.
- Hence in this case the extension is immediate.

Bounded Linear Functionals

Theorem (Bounded Linear Functionals)

Let X be a normed space and let $x_0 \neq 0$ be any element of X. Then there exists a bounded linear functional $\widetilde{f}$ on X such that $\|\widetilde{f}\| = 1$, $\widetilde{f}(x_0) = \|x_0\|$.

- We consider the subspace Z of X consisting of all elements $x = \alpha x_0$ where α is a scalar. On Z we define a linear functional f by $f(x) = f(\alpha x_0) = \alpha \|x_0\|$. f is bounded and has norm $\|f\| = 1$ because

$$|f(x)| = |f(\alpha x_0)| = |\alpha| \|x_0\| = \|\alpha x_0\| = \|x\|.$$

The theorem implies that f has a linear extension $\widetilde{f}$ from Z to X, of norm $\|\widetilde{f}\| = \|f\| = 1$. Thus, $\widetilde{f}(x_0) = f(x_0) = \|x_0\|$.

Norm, Zero Vector

Corollary (Norm, Zero Vector)

For every x in a normed space X, we have

$$\|x\| = \sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|}.$$

Hence if x_0 is such that $f(x_0) = 0$, for all $f \in X'$, then $x_0 = 0$.

- From the preceding theorem, we have, writing x for x_0,

$$\sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|} \geq \frac{|\widetilde{f}(x)|}{\|\widetilde{f}\|} = \frac{\|x\|}{1} = \|x\|.$$

But $|f(x)| \leq \|f\| \|x\|$. So, we also obtain

$$\sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|} \leq \|x\|.$$

Subsection 4

Bounded Linear Functionals on $C[a,b]$

Vector Space of Functions of Bounded Variation

- A function w defined on $[a,b]$ is said to be of **bounded variation** on $[a,b]$ if its **total variation** $\text{Var}(w)$ on $[a,b]$ is finite, where

$$\text{Var}(w) = \sup \sum_{j=1}^{n} |w(t_j) - w(t_{j-1})|,$$

  the supremum being taken over all **partitions** $a = t_0 < t_1 < \cdots < t_n = b$ of the interval $[a,b]$; both $n \in \mathbb{N}$ and the choice of values $t_1, \ldots, t_{n-1}$ in $[a,b]$ satisfying the inequalities are arbitrary.
- All functions of bounded variation on $[a,b]$ form a vector space.
- A norm on this space is given by

$$\|w\| = |w(a)| + \text{Var}(w).$$

- The normed space thus defined is denoted by $\text{BV}[a,b]$, where BV suggests "bounded variation".

The Riemann-Stieltjes Integral

- Let $x \in C[a,b]$ and $w \in \mathrm{BV}[a,b]$.
- Let P_n be any partition of $[a,b]$ given by $a = t_0 < t_1 < \cdots < t_n = b$ and denote by $\eta(P_n)$ the length of a largest interval $[t_{j-1}, t_j]$, that is,
$$\eta(P_n) = \max(t_1 - t_0, \ldots, t_n - t_{n-1}).$$
- For every partition P_n of $[a,b]$, we consider the sum
$$s(P_n) = \sum_{j=1}^{n} x(t_j)[w(t_j) - w(t_{j-1})].$$
- There exists a number $\mathscr{I}$ with the property that for every $\varepsilon > 0$, there is a $\delta > 0$, such that
$$\eta(P_n) < \delta \quad \text{implies} \quad |\mathscr{I} - s(P_n)| < \varepsilon.$$
- $\mathscr{I}$ is called the **Riemann-Stieltjes integral** of x over $[a,b]$ with respect to w and is denoted by $\int_a^b x(t)dw(t)$.
- We can obtain $\int_a^b x(t)dw(t)$ as the limit of the sums $s(P_n)$, for a sequence (P_n) of partitions of $[a,b]$ satisfying $\eta(P_n) \stackrel{n\to\infty}{\longrightarrow} 0$.

Two Special Cases

- For $w(t) = t$, the integral $\int_a^b x(t)dw(t) = \int_a^b x(t)dt$ is the familiar Riemann integral of x over $[a,b]$.
- If x is continuous on $[a,b]$ and w has a derivative which is integrable on $[a,b]$, then

$$\int_a^b x(t)dw(t) = \int_a^b x(t)w'(t)dt,$$

where the prime denotes differentiation with respect to t.

Linearity Properties and an Upper Bound

- The integral $\int_a^b x(t)dw(t)$ depends linearly on $x \in C[a,b]$, that is, for all $x_1, x_2 \in C[a,b]$ and scalars α and β we have

$$\int_a^b [\alpha x_1(t) + \beta x_2(t)]dw(t) = \alpha \int_a^b x_1(t)dw(t) + \beta \int_a^b x_2(t)dw(t).$$

- The integral also depends linearly on $w \in \mathrm{BV}[a,b]$, i.e., for all $w_1, w_2 \in \mathrm{BV}[a,b]$ and scalars γ and δ we have

$$\int_a^b x(t)d(\gamma w_1 + \delta w_2)(t) = \gamma \int_a^b x(t)dw_1(t) + \delta \int_a^b x(t)dw_2(t).$$

- It also holds that

$$\left| \int_a^b x(t)dw(t) \right| \le \max_{t \in [a,b]} |x(t)| \mathrm{Var}(w).$$

This generalizes the formula $\left| \int_a^b x(t)dt \right| \le \max_{t \in [a,b]} |x(t)|(b-a)$.

Riesz's Theorem for Functionals on $C[a,b]$

Riesz's Theorem (Functionals on $C[a,b]$)

Every bounded linear functional f on $C[a,b]$ can be represented by a Riemann-Stieltjes integral $f(x) = \int_a^b x(t)dw(t)$. where w is of bounded variation on $[a,b]$ and has the total variation $\text{Var}(w) = \|f\|$.

- From the Hahn-Banach theorem for normed spaces we see that f has an extension $\widetilde{f}$ from $C[a,b]$ to the normed space $B[a,b]$ consisting of all bounded functions on $[a,b]$ with norm defined by $\|x\| = \sup_{t\in J}|x(t)|$, $J = [a,b]$. Furthermore, by that theorem, the linear functional $\widetilde{f}$ is bounded and has the same norm as f, that is, $\|\widetilde{f}\| = \|f\|$. To define the function w, consider the **characteristic function** of the interval $[a,t]$: $x_t = \begin{cases} 1, & \text{if } a \le x \le t \\ 0, & \text{if } t < x \le b \end{cases}$. Using x_t and the functional $\widetilde{f}$, we define w on $[a,b]$ by $w(a) = 0$, $w(t) = \widetilde{f}(x_t)$, $t \in (a,b]$. We show that w is of bounded variation and $\text{Var}(w) \le \|f\|$.

Proof: Bounded Variation of w on $[a,b]$

- For a complex quantity we can use the polar form: Setting $\theta = \arg\zeta$, we may write $\zeta = |\zeta| e(\zeta)$, where $e(\zeta) = \begin{cases} 1, & \text{if } \zeta = 0 \\ e^{i\theta}, & \text{if } \zeta \neq 0 \end{cases}$.

  We see that if $\zeta \neq 0$, then $|\zeta| = \frac{\zeta}{e^{i\theta}} = \zeta e^{-i\theta}$. Hence, for any ζ, zero or not, we have $|\zeta| = \zeta \overline{e(\zeta)}$. Set $\varepsilon_j = \overline{e(w(t_j) - w(t_{j-1}))}$ and $x_{t_j} = x_j$. Then

$$\begin{array}{rcl} \sum_{j=1}^n |w(t_j) - w(t_{j-1})| & = & |\widetilde{f}(x_1)| + \sum_{j=2}^n |\widetilde{f}(x_j) - \widetilde{f}(x_{j-1})| \\ & = & \varepsilon_1 \widetilde{f}(x_1) + \sum_{j=2}^n \varepsilon_j [\widetilde{f}(x_j) - \widetilde{f}(x_{j-1})] \\ & = & \widetilde{f}(\varepsilon_1 x_1 + \sum_{j=2}^n \varepsilon_j [x_j - x_{j-1}]) \\ & \leq & \|\widetilde{f}\| \left\| \varepsilon_1 x_1 + \sum_{j=2}^n \varepsilon_j [x_j - x_{j-1}] \right\|. \end{array}$$

  On the right, $\|\widetilde{f}\| = \|f\|$ and the other factor $\|\cdots\|$ equals 1 because $|\varepsilon_j| = 1$ and from the definition of the x_j's we see that for each $t \in [a,b]$ only one of the terms $x_1, x_2 - x_1, \ldots$ is not zero (and of norm 1). On the left we now take the supremum over all partitions of $[a,b]$. Then we have $\text{Var}(w) \leq \|f\|$.

Proof: The Integration Formula

- We show $f(x) = \int_a^b x(t)dw(t)$, for $x \in [a,b]$.
  For every partition P_n, we define a function, denoted simply by z_n (instead of $z(P_n)$ or z_{P_n}), keeping in mind that z_n depends on P_n, not merely on n. The defining formula is

$$z_n = x(t_0)x_1 + \sum_{j=2}^{n} x(t_{j-1})[x_j - x_{j-1}].$$

  Then $z_n \in B[a,b]$. By the definition of w,

$$\begin{aligned} \widetilde{f}(z_n) &= x(t_0)\widetilde{f}(x_1) + \textstyle\sum_{j=2}^{n} x(t_{j-1})[\widetilde{f}(x_j) - \widetilde{f}(x_{j-1})] \\ &= x(t_0)w(t_1) + \textstyle\sum_{j=2}^{n} x(t_{j-1})[w(t_j) - w(t_{j-1})] \\ &= \textstyle\sum_{j=1}^{n} x(t_{j-1})[w(t_j) - w(t_{j-1})], \end{aligned}$$

  where the last equality follows from $w(t_0) = w(a) = 0$. Choose a (P_n), such that $\eta(P_n) \to 0$. As $n \to \infty$, the sum on the right approaches $\int_a^b x(t)dw(t)$. So it suffices to show $\widetilde{f}(z_n) \to \widetilde{f}(x)$, which equals $f(x)$, since $x \in C[a,b]$.

Proof: $\widetilde{f}(z_n) \to \widetilde{f}(x)$

- We prove that $\widetilde{f}(z_n) \to \widetilde{f}(x)$.

  Recall

  $$z_n = x(t_0)x_1 + \sum_{j=2}^{n} x(t_{j-1})[x_j - x_{j-1}].$$

  By the definition of x_t, this yields $z_n(a) = x(a) \cdot 1$, since the sum is zero at $t = a$. Hence, $z_n(a) - x(a) = 0$.

  Furthermore, if $t_{j-1} < t \leq t_j$, then we obtain $z_n(t) = x(t_{j-1}) \cdot 1$. It follows that for those t, $|z_n(t) - x(t)| = |x(t_{j-1}) - x(t)|$.

  Consequently, if $\eta(P_n) \to 0$, then $\|z_n - x\| \to 0$ because x is continuous on $[a,b]$, hence uniformly continuous on $[a,b]$, since $[a,b]$ is compact.

  The continuity of $\widetilde{f}$ now implies that $\widetilde{f}(z_n) \to \widetilde{f}(x)$.

Proof: $\mathrm{Var}(w) = \|f\|$

- Finally, we show $\mathrm{Var}(w) = \|f\|$.

  Recall $f(x) = \int_a^b x(t)dw(t)$ and $\left|\int_a^b x(t)dt\right| \le \max_{t\in[a,b]} |x(t)|\mathrm{Var}(w)$.

  So
  $$|f(x)| \le \max_{t\in[a,b]} |x(t)|\mathrm{Var}(w) = \|x\|\mathrm{Var}(w).$$

  Taking the supremum over all $x \in C[a,b]$ of norm one, we obtain $\|f\| \le \mathrm{Var}(w)$. Combining with $\mathrm{Var}(w) \le \|f\|$, this yields $\mathrm{Var}(w) = \|f\|$.
- We note that w in the theorem is not unique, but can be made unique by imposing the normalizing conditions that:
  - w be zero at a: $w(a) = 0$;
  - w continuous from the right: $w(t+0) = w(t)$, $a < t < b$.

Subsection 5

Adjoint Operator

The Adjoint Operator

- Consider a bounded linear operator $T : X \to Y$, where X and Y are normed spaces.
- Let g be any bounded linear functional on Y.
- Setting $y = Tx$, we obtain a functional on X, call it f:

$$f(x) = g(Tx), \quad x \in X.$$

  - f is linear since g and T are linear.
  - f is bounded because $|f(x)| = |g(Tx)| \le \|g\| \|Tx\| \le \|g\| \|T\| \|x\|$. Taking the supremum over all $x \in X$ of norm one, we obtain $\|f\| \le \|g\| \|T\|$.
- Thus, $f \in X'$, where X' is the dual space of X.
- For variable $g \in Y'$, $f(x) = g(Tx)$ defines an operator from Y' into X', which is called the **adjoint operator** of T and is denoted by $T^\times$:

$$X \xrightarrow{T} Y, \qquad X' \xleftarrow{T^\times} Y'.$$

The Adjoint Operator and its Norm

Definition (Adjoint operator $T^\times$)

Let $T:X\to Y$ be a bounded linear operator, where X and Y are normed spaces. The **adjoint operator** $T^\times: Y'\to X'$ is defined by $f(x)=(T^\times g)(x)$ $=g(Tx)$, $g\in Y'$, where X' and Y' are the dual spaces of X and Y.

Theorem (Norm of the Adjoint Operator)

The adjoint operator $T^\times$ is linear and bounded, and $\|T^\times\|=\|T\|$.

- The operator $T^\times$ is linear since its domain Y' is a vector space and we have

$$\begin{aligned}(T^\times(\alpha g_1+\beta g_2))(x) &= (\alpha g_1+\beta g_2)(Tx)\\ &= \alpha g_1(Tx)+\beta g_2(Tx)\\ &= \alpha(T^\times g_1)(x)+\beta(T^\times g_2)(x).\end{aligned}$$

From $f(x)=(T^\times g)(x)=g(Tx)$, we have $f=T^\times g$. By $\|f\|\le\|g\|\|T\|$, $\|T^\times g\|=\|f\|\le\|g\|\|T\|$. Taking the supremum over all $g\in Y'$ of norm one, we obtain $\|T^\times\|\le\|T\|$. So we need to see that $\|T^\times\|\ge\|T\|$.

The Adjoint Operator and its Norm

- By the Hahn-Banach Theorem, for every nonzero $x_0 \in X$, there is a $g_0 \in Y'$, such that $\|g_0\| = 1$ and $g_0(Tx_0) = \|Tx_0\|$. By the definition of the adjoint, $g_0(Tx_0) = (T^\times g_0)(x_0)$. Writing $f_0 = T^\times g_0$, we thus obtain

$$\|Tx_0\| = g_0(Tx_0) = f_0(x_0) \le \|f_0\|\|x_0\| = \|T^\times g_0\|\|x_0\| \le \|T^\times\|\|g_0\|\|x_0\|.$$

  Since $\|g_0\| = 1$, we get, for every $x_0 \in X$, $\|Tx_0\| \le \|T^\times\|\|x_0\|$. This includes $x_0 = 0$ since $T0 = 0$. But always $\|Tx_0\| \le \|T\|\|x_0\|$, and here $c = \|T\|$ is the smallest constant c, such that $\|Tx_0\| \le c\|x_0\|$ holds, for all $x_0 \in X$. Hence, $\|T^\times\|$ cannot be smaller than $\|T\|$, that is, we must have $\|T^\times\| \ge \|T\|$.

The Special Case of Matrices

- In n-dimensional Euclidean space $\mathbb{R}^n$, a linear operator $T:\mathbb{R}^n\to\mathbb{R}^n$ can be represented by matrices, where a matrix $T_E=(\tau_{jk})$ depends on the choice of a basis $E=\{e_1,\ldots,e_n\}$ for $\mathbb{R}^n$, whose elements are arranged in some order which is kept fixed.
- We choose a basis E, regard $x=(\xi_1,\ldots,\xi_n)$, $y=(\eta_1,\ldots,\eta_n)$ as column vectors and employ the usual notation for matrix multiplication: $y=T_Ex$ or in components $\eta_j=\sum_{k=1}^n\tau_{jk}\xi_k$.
- Let $F=\{f_1,\ldots,f_n\}$ be the dual basis of E.
- This is a basis for $\mathbb{R}^{n\prime}$ which is also Euclidean n-space.
- Then every $g\in\mathbb{R}^{n\prime}$ has a representation $g=\alpha_1f_1+\cdots+\alpha_nf_n$.

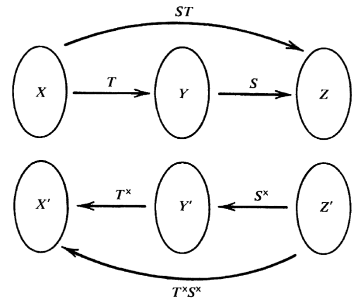
The Special Case of Matrices (Cont'd)

- By the definition of the dual basis, we have $f_j(y) = f_j(\sum \eta_k e_k) = \eta_j$.
- Hence we obtain $g(y) = g(T_E x) = \sum_{j=1}^n \alpha_j \eta_j = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \tau_{jk} \xi_k$.
- Interchanging the order of summation, $g(T_E x) = \sum_{k=1}^n \beta_k \xi_k$, where $\beta_k = \sum_{j=1}^n \tau_{jk} \alpha_j$.
- We may regard this as the definition of a functional f on X in terms of g: $f(x) = g(T_E x) = \sum_{k=1}^n \beta_k \xi_k$.
- Remembering the definition of the adjoint operator, we can write this $f = T_E^\times g$ or in components $\beta_k = \sum_{j=1}^n \tau_{jk} \alpha_j$.
- Noting that in β_k we sum with respect to the first subscript, i.e., over all elements of a column of T_E, we have:

  If T is represented by a matrix T_E, then the adjoint operator $T^\times$ is represented by the transpose of T_E.

Properties of Adjoints

- If $S, T \in B(X, Y)$, then
  - $(S + T)^{\times} = S^{\times} + T^{\times}$;
  - $(\alpha T)^{\times} = \alpha T^{\times}$.
- Let X, Y, Z be normed spaces and $T \in B(X, Y)$ and $S \in B(Y, Z)$. Then, for the adjoint of the product ST we have $(ST)^{\times} = T^{\times} S^{\times}$:



- If $T \in B(X, Y)$ and T^{-1} exists and $T^{-1} \in B(Y, X)$, then $(T^{\times})^{-1}$ also exists, $(T^{\times})^{-1} \in B(X', Y')$ and $(T^{\times})^{-1} = (T^{-1})^{\times}$.

The Operators A_1 and A_2

- Let $T : X \to Y$ be a bounded linear operator from a Hilbert space $X = H_1$ to a Hilbert space $Y = H_2$.
- In this case we first have $\begin{array}{ccc} H_1 & \xrightarrow{T} & H_2 \\ H_1' & \xleftarrow[T^\times]{} & H_2' \end{array}$ where the adjoint $T^\times$ is defined by $T^\times g = f$, $f(x) = g(Tx)$, for $f \in H_1', g \in H_2'$.
- Since f and g are functionals on Hilbert spaces, they have Riesz representations, say, $f(x) = \langle x, x_0 \rangle$, $x_0 \in H_1$, and $g(y) = \langle y, y_0 \rangle$, $y_0 \in H_2$.
- We also know that x_0 and y_0 are uniquely determined by f and g, respectively.
- Thus, we get operators

$$\begin{array}{ll} A_1 : H_1' \to H_1; & A_1 f = x_0, \\ A_2 : H_2' \to H_2; & A_2 g = y_0. \end{array}$$

Properties of the Operators A_1 and A_2

- We know that A_1 and A_2 are bijective and isometric since $\|A_1 f\| = \|x_0\| = \|f\|$, and similarly for A_2.
- Furthermore, the operators A_1 and A_2 are conjugate linear: If we write $f_1(x) = \langle x, x_1 \rangle$ and $f_2(x) = \langle x, x_2 \rangle$, we have, for all x and scalars α, β,

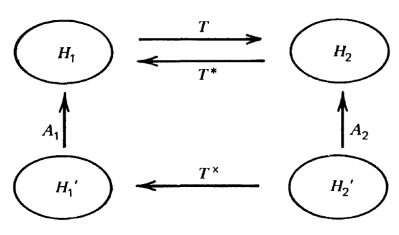
$$\begin{array}{rcl} (\alpha f_1 + \beta f_2)(x) & = & \alpha f_1(x) + \beta f_2(x) \\ & = & \alpha \langle x, x_1 \rangle + \beta \langle x, x_2 \rangle \\ & = & \langle x, \overline{\alpha} x_1 + \overline{\beta} x_2 \rangle. \end{array}$$

  By the definition of A_1, $A_1(\alpha f_1 + \beta f_2) = \overline{\alpha} A_1 f_1 + \overline{\beta} A_1 f_2$. So A_1 is conjugate linear.
- For A_2 the proof is similar.

Relation Between Adjoint and Hilbert-Adjoint

- Composition gives the operator

$$\begin{array}{rcl} T^* = A_1 T^\times A_2^{-1} : H_2 & \to & H_1; \\ T^* y_0 & = & x_0. \end{array}$$



T^* is linear since it involves two conjugate linear mappings, in addition to the linear operator $T^\times$.

We show T^* is indeed the Hilbert-adjoint operator of T:

$$\langle Tx, y_0 \rangle = g(Tx) = f(x) = \langle x, x_0 \rangle = \langle x, T^* y_0 \rangle.$$

- $T^* = A_1 T^\times A_2^{-1} : H_2 \to H_1;\ T^* y_0 = x_0$ represents the Hilbert-adjoint operator T^* of a linear operator T on a Hilbert space in terms of the adjoint operator $T^\times$ of T.
- $\|T^*\| = \|T\|$ follows from $\|T^\times\| = \|T\|$ and the isometry of A_1 and A_2.

Differences Between $T^\times$ and T^*

- Differences between the adjoint operator $T^\times$ of $T : X \to Y$ and the Hilbert-adjoint operator T^* of $T : H_1 \to H_2$, where X, Y are normed spaces and H_1, H_2 are Hilbert spaces:
  - $T^\times$ is defined on the dual of the space which contains the range of T; T^* is defined directly on the space which contains the range of T.
  - For $T^\times$ we have
  $$(\alpha T)^\times = \alpha T^\times;$$
  For T^* we have
  $$(\alpha T)^* = \overline{\alpha} T^*.$$
  - In the finite dimensional case:
    - $T^\times$ is represented by the transpose of the matrix representing T;
    - T^* is represented by the complex conjugate transpose of that matrix.

Subsection 6

Reflexive Spaces

Review of Algebraic Reflexivity

- Recall that a vector space X is said to be **algebraically reflexive** if the canonical mapping $C: X \to X^{**}$ is surjective.
- $X^{**} = (X^*)^*$ is the second algebraic dual space of X and the mapping C is defined by $x \mapsto g_x$, where

$$g_x(f) = f(x), \quad f \in X^* \text{ variable},$$

  i.e., for any $x \in X$, the image is the linear functional g_x defined as above.
- If X is finite dimensional, then X is algebraically reflexive.

The (Normed Space) Dual

- We consider a normed space X, its dual space X' and the dual space $(X')'$, of X'.
- The space $(X')'$ is denoted by X'' and is called the **second dual space** of X (or **bidual space** of X).
- We define a functional g_x on X' by choosing a fixed $x \in X$ and setting

$$g_x(f) = f(x), \quad f \in X' \text{ variable.}$$

- As contrasted to algebraic duality, f here is bounded.

Lemma (Norm of g_x)

For every fixed x in a normed space X, the functional g_x is a bounded linear functional on X', so that $g_x \in X''$, and has the norm $\|g_x\| = \|x\|$.

- Linearity of g_x is known. For the norm we have
$$\|g_x\| = \sup_{\substack{f \in X' \\ f \neq 0}} \frac{|g_x(f)|}{\|f\|} = \sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|} = \|x\|.$$

The Canonical Mapping

- To every $x \in X$, there corresponds a unique bounded linear functional $g_x \in X''$ given by $g_x(f) = f(x)$.
  This defines a mapping $C : X \to X''$; $x \mapsto g_x$, called the **canonical mapping** of X into X''.

Lemma (Canonical Mapping)

The canonical mapping C is an isomorphism of the normed space X onto the normed space $\mathscr{R}(C)$, the range of C.

- For linearity of C,
  $$g_{\alpha x+\beta y}(f) = f(\alpha x+\beta y) = \alpha f(x)+\beta f(y) = \alpha g_x(f)+\beta g_y(f).$$
  In particular, $g_x - g_y = g_{x-y}$. Since $\|g_x\| = \|x\|$, $\|g_x - g_y\| = \|g_{x-y}\| = \|x-y\|$. Thus, C is isometric, i.e., it preserves the norm.
  Isometry implies injectivity.
  Hence C is bijective, regarded as a mapping onto its range.

Embeddability and Reflexivity

- X is said to be **embeddable** in a normed space Z if X is isomorphic with a subspace of Z.
- Isomorphism refers to isomorphisms of normed spaces, that is, vector space isomorphisms which preserve norm.
- By the lemma, X is embeddable in X'', and C is also called the **canonical embedding** of X into X''.
- In general, C will not be surjective, so that the range $\mathscr{R}(C)$ will be a proper subspace of X''.

Definition (Reflexivity)

A normed space X is said to be **reflexive** if

$$\mathscr{R}(C) = X'',$$

where C is the canonical mapping $C : X \to X''$; $x \mapsto g_x$, with $g_x(f) = f(x)$, $f \in X'$.

Reflexivity Implies Completeness

- If X is reflexive, it is isomorphic (hence isometric) with X''.

  The converse does not generally hold (R.C. James 1950, 1951).
- Completeness does not imply reflexivity.

Theorem (Completeness)

If a normed space X is reflexive, it is complete (hence a Banach space).

- Since X'' is the dual space of X', it is complete by a previous theorem. Reflexivity of X means that $\mathscr{R}(C) = X''$. Completeness of X now follows from that of X''.

Examples

- $\mathbb{R}^n$ is reflexive: This follows directly from preceding work.
- If $\dim X < \infty$, then every linear functional on X is bounded, so that $X' = X^*$ and algebraic reflexivity of X thus implies:

Theorem (Finite Dimension)

Every finite dimensional normed space is reflexive.

- ℓ^p with $1 < p < +\infty$ is reflexive: This also follows from previous work.
- $L^p[a,b]$, with $1 < p < +\infty$, is reflexive.
- $C[a,b]$, ℓ^1, $L^1[a,b]$. ℓ^∞ are nonreflexive spaces.
- The subspaces c and c_0 of ℓ^∞, where c is the space of all convergent sequences of scalars and c_0 is the space of all sequences of scalars converging to zero, are also nonreflexive.

Reflexivity of Hilbert Spaces

Theorem (Hilbert Space)

Every Hilbert space H is reflexive.

- We shall prove surjectivity of the canonical mapping $C: H \to H''$ by showing that, for every $g \in H''$, there is an $x \in H$, such that $g = Cx$.
  As a preparation we define $A: H' \to H$ by $Af = z$, where z is given by the Riesz representation $f(x) = \langle x, z \rangle$. We know that A is bijective and isometric. A is conjugate linear. Now H' is complete and a Hilbert space with inner product defined by $\langle f_1, f_2 \rangle_1 = \langle Af_2, Af_1 \rangle$. Note the order of f_1, f_2 on both sides. (IP1) to (IP4) are readily verified; e.g., for (IP2), $\langle \alpha f_1, f_2 \rangle_1 = \langle Af_2, A(\alpha f_1) \rangle = \langle Af_2, \overline{\alpha} Af_1 \rangle = \alpha \langle f_1, f_2 \rangle_1$.
  Let $g \in H''$ be arbitrary. Let its Riesz representation be $g(f) = \langle f, f_0 \rangle_1 = \langle Af_0, Af \rangle$. Recall that $f(x) = \langle x, z \rangle$, where $z = Af$. Writing $Af_0 = x$, we thus have $\langle Af_0, Af \rangle = \langle x, z \rangle = f(x)$. Together, $g(f) = f(x)$, that is, $g = Cx$, by the definition of C. Since $g \in H''$ was arbitrary, C is surjective, so that H is reflexive.

Separability and Reflexivity

- We next show that separability of X' implies separability of X (the converse not being generally true).
- Hence, if a normed space X is reflexive, X'' is isomorphic with X, so that, in this case, separability of X implies separability of X'', and, by the aforementioned upcoming result, the space X' is also separable.
- These results imply:

  A separable normed space X with a nonseparable dual space X' cannot be reflexive.

  Example: ℓ^1 is not reflexive.

  ℓ^1 is separable, as seen before. $\ell^{1'} = \ell^\infty$ is not separable. It follows that ℓ^1 cannot be reflexive.

Existence of a Functional

Lemma (Existence of a Functional)

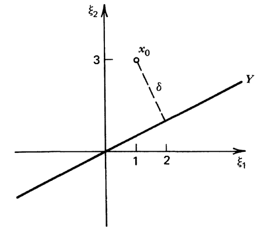
Let Y be a proper closed subspace of a normed space X. Let $x_0 \in X - Y$ be arbitrary and $\delta = \inf_{\tilde{y} \in Y} \|\tilde{y} - x_0\|$ the distance from x_0 to Y. Then, there exists an $\tilde{f} \in X'$, such that

$$\|\tilde{f}\| = 1, \quad \tilde{f}(y) = 0, \text{ for all } y \in Y, \quad \tilde{f}(x_0) = \delta.$$

- We consider the subspace $Z \subseteq X$ spanned by Y and x_0 and:
  - Define on Z a bounded linear functional f by

    $$f(z) = f(y + \alpha x_0) = \alpha\delta, \quad \text{for all } y \in Y;$$

  - Show that f satisfies the conditions;
  - Extend f to X.



Proof of the Existence of a Functional

- Every $z \in Z = \text{span}(Y \cup \{x_0\})$ has a unique representation $z = y + \alpha x_0$, with $y \in Y$. This is used to define

$$f(z) = f(y + \alpha x_0) = \alpha\delta.$$

  - Linearity of f is readily seen.
  - Since Y is closed, $\delta > 0$, so that $f \neq 0$.
  - $\alpha = 0$ gives $f(y) = 0$, for all $y \in Y$.
  - For $\alpha = 1$ and $y = 0$, we have $f(x_0) = \delta$.

  We show that f is bounded. $\alpha = 0$ gives $f(z) = 0$. Let $\alpha \neq 0$. Using the definition of δ and noting that $-\frac{1}{\alpha} y \in Y$, we obtain

$$|f(z)| = |\alpha|\delta = |\alpha| \inf_{\tilde{y} \in Y} \|\tilde{y} - x_0\| \leq |\alpha| \left\| -\frac{1}{\alpha} y - x_0 \right\| = \|y + \alpha x_0\|,$$

  i.e., $|f(z)| \leq \|z\|$. Hence f is bounded and $\|f\| \leq 1$.

Proof of the Existence of a Functional (Cont'd)

- We show that $\|f\| \geq 1$: By the definition of an infimum, Y contains a sequence (y_n), such that $\|y_n - x_0\| \to \delta$. Let $z_n = y_n - x_0$. Then we have $f(z_n) = -\delta$. Also

$$\|f\| = \sup_{\substack{z \in Z \\ z \neq 0}} \frac{|f(z)|}{\|z\|} \geq \frac{|f(z_n)|}{\|z_n\|} = \frac{\delta}{\|z_n\|} \to \frac{\delta}{\delta} = 1.$$

Hence $\|f\| \geq 1$.

By the Hahn-Banach theorem for normed spaces, we can extend f to X without increasing the norm.

Separability Theorem

Theorem (Separability)

If the dual space X' of a normed space X is separable, then X itself is separable.

- We assume that X' is separable. Then the unit sphere

$$U' = \{f : \|f\| = 1\} \subseteq X'$$

also contains a countable dense subset, say, (f_n). Since $f_n \in U'$, we have $\|f_n\| = \sup_{\|x\|=1} |f_n(x)| = 1$. By the definition of a supremum we can find points $x_n \in X$ of norm 1 such that $|f_n(x_n)| \geq \frac{1}{2}$. Let Y be the closure of span(x_n). Then Y is separable because Y has a countable dense subset, namely, the set of all linear combinations of the x_n's with coefficients whose real and imaginary parts are rational.

We show that $Y = X$.

Separability Theorem (Cont'd)

- We show that $Y = X$.

  Suppose $Y \neq X$. Then, since Y is closed, by the Lemma on the Existence of a Functional, there exists an $\widetilde{f} \in X'$ with $\|\widetilde{f}\| = 1$ and $\widetilde{f}(y) = 0$, for all $y \in Y$. Since $x_n \in Y$, we have $\widetilde{f}(x_n) = 0$ and for all n,

  $$\frac{1}{2} \leq |f_n(x_n)| = |f_n(x_n) - \widetilde{f}(x_n)| = |(f_n - \widetilde{f})(x_n)| \leq \|f_n - \widetilde{f}\| \|x_n\|,$$

  where $\|x_n\| = 1$. Hence $\|f_n - \widetilde{f}\| \geq \frac{1}{2}$.

  This contradicts the assumption that (f_n) is dense in U' because, as $\|\widetilde{f}\| = 1$, $\widetilde{f}$ is itself in U'.

Subsection 7

Category and Uniform Boundedness Theorems

Cornerstone Theorems of Functional Analysis

- The corner stones of functional analysis are:
  - The Hahn-Banach Theorem;
  - The Uniform Boundedness Theorem;
  - The Open Mapping Theorem;
  - The Closed Graph Theorem.
- Unlike the Hahn-Banach, the other three require completeness.
- We shall obtain all three other theorems from a common source, the so-called *Baire's Category Theorem*.

Category

- Each of the concepts needed for Baire's Category Theorem has two names, a new name and an old one given in parentheses.

Definition (Category)

A subset M of a metric space X is said to be:

(a) **rare** (or **nowhere dense**) in X if its closure $\overline{M}$ has no interior points;

(b) **meager** (or **of the first category**) in X if M is the union of countably many sets each of which is rare in X;

(c) **nonmeager** (or **of the second category**) in X if M is not meager in X.

Baire's Category Theorem for Complete Metric Spaces

Baire's Category Theorem (Complete Metric Spaces)

If a metric space $X \neq \emptyset$ is complete, it is nonmeager in itself. Hence, if $X \neq \emptyset$ is complete and $X = \bigcup_{k=1}^{\infty} A_k$, A_k closed, then, at least one A_k contains a nonempty open subset.

- The idea is simple: Suppose the complete metric space $X \neq \emptyset$ were meager in itself. Then $X = \bigcup_{k=1}^{\infty} M_k$, with each M_k rare in X. We shall construct a Cauchy sequence (p_k) whose limit p (which exists by completeness) is in no M_k, thereby contradicting $X = \bigcup_{k=1}^{\infty} M_k$.

  By assumption, M_1 is rare in X. By definition, M_1 does not contain a nonempty open set. But X does (for instance, X itself). This implies $\overline{M}_1 \neq X$. Hence, the complement $\overline{M}_1^c = X - \overline{M}_1$ of $\overline{M}_1$ is nonempty and open. We may thus choose a point p_1 in $\overline{M}_1^c$ and an open ball about it, say, $B_1 = B(p_1; \varepsilon_1) \subseteq \overline{M}_1^c$, $\varepsilon_1 < \frac{1}{2}$.

Baire's Category Theorem (Cont'd)

- By assumption, M_2 is rare in X. So $\overline{M}_2$ does not contain a nonempty open set. Hence it does not contain the open ball $B(p_1; \frac{1}{2}\varepsilon_1)$. This implies that $\overline{M}_2^c \cap B(p_1; \frac{1}{2}\varepsilon_1)$ is nonempty and open. Thus, we may choose an open ball $B_2 = B(p_2; \varepsilon_2) \subseteq \overline{M}_2^c \cap B(p_1; \frac{1}{2}\varepsilon_1)$, with $\varepsilon_2 < \frac{1}{2}\varepsilon_1$.
  By induction, we get a sequence of balls $B_k = B(p_k; \varepsilon_k)$, with $\varepsilon_k < \frac{1}{2^k}$, such that $B_k \cap M_k = \emptyset$ and $B_{k+1} \subseteq B(p_k; \frac{1}{2}\varepsilon_k) \subseteq B_k$, $k = 1, 2, \ldots$.
  Since $\varepsilon_k < \frac{1}{2^k}$, the sequence (p_k) is Cauchy and converges, say, $p_k \to p \in X$ because X is complete. Now, for all m, $n > m$, we have $B_n \subseteq B(p_m; \frac{1}{2}\varepsilon_m)$. So $d(p_m, p) \le d(p_m, p_n) + d(p_n, p) < \frac{1}{2}\varepsilon_m + d(p_n, p) \overset{n\to\infty}{\longrightarrow} \frac{1}{2}\varepsilon_m$. Hence $p \in B_m$, for all m. Since $B_m \subseteq \overline{M}_m^c$, we get $p \notin M_m$, for all m. So $p \notin \bigcup M_m = X$, a contradiction.
- We note that the converse of Baire's theorem is not generally true, i.e., there exists an incomplete normed space which is nonmeager in itself.

Uniform Boundedness Theorem

- We use Baire's Theorem to obtain the Uniform Boundedness Theorem.

Uniform Boundedness Theorem

Let (T_n) be a sequence of bounded linear operators $T_n : X \to Y$ from a Banach space X into a normed space Y such that $(\|T_n x\|)$ is bounded for every $x \in X$, say, $\|T_n x\| \leq c_x$, for all $n = 1, 2, \ldots$, where c_x is a real number. Then the sequence of the norms $\|T_n\|$ is bounded, that is, there is a c, such that $\|T_n\| \leq c$, $n = 1, 2, \ldots$.

- For every $k \in \mathbb{N}$, let $A_k \subseteq X$ be the set of all x, such that $\|T_n x\| \leq k$, for all n.

  A_k is closed: Let $x \in \overline{A}_k$. Then, there is a sequence (x_j) in A_k converging to x. This means that, for every fixed n, $\|T_n x_j\| \leq k$. We obtain $\|T_n x\| \leq k$, because T_n is continuous and so is the norm. Hence $x \in A_k$, and A_k is closed.

Uniform Boundedness Theorem (Cont'd)

- By hypothesis, each $x \in X$ belongs to some A_k. Hence $X = \bigcup_{k=1}^{\infty} A_k$. Since X is complete, Baire's Theorem implies that some A_k contains an open ball, say, $B_0 = B(x_0; r) \subseteq A_{k_0}$. Let $x \in X$ be arbitrary, nonzero. We set

$$z = x_0 + \gamma x, \quad \gamma = \frac{r}{2\|x\|}.$$

Then $\|z - x_0\| < r$, so that $z \in B_0$. Since $B_0 \subseteq A_{k_0}$, $\|T_n z\| \le k_0$, for all n. Also $\|T_n x_0\| \le k_0$, since $x_0 \in B_0$. Since $x = \frac{1}{\gamma}(z - x_0)$, for all n,

$$\|T_n x\| = \frac{1}{\gamma}\|T_n(z - x_0)\| \le \frac{1}{\gamma}(\|T_n z\| + \|T_n x_0\|) \le \frac{4}{r}\|x\| k_0.$$

Hence, for all n, $\|T_n\| = \sup_{\|x\|=1} \|T_n x\| \le \frac{4}{r} k_0$. This is the conclusion with $c = \frac{4}{r} k_0$.

Application: Space of Polynomials

- The normed space X of all polynomials with norm defined by

$$\|x\| = \max_j |\alpha_j|, \quad \alpha_0, \alpha_1, \ldots \text{ the coefficients of } x$$

is not complete.

We construct a sequence of bounded linear operators on X which satisfies $\|T_n x\| \le c_x$ but not $\|T_n\| \le c$, so that X cannot be complete.

Write $x \neq 0$ of degree N_x as $x(t) = \sum_{j=0}^{\infty} \alpha_j t^j$, $\alpha_j = 0$, for $j > N_x$.

As a sequence of operators on X, take $T_n = f_n$ defined by $T_n 0 = f_n(0) = 0$, $T_n x = f_n(x) = \alpha_0 + \alpha_1 + \cdots + \alpha_{n-1}$. f_n is linear. f_n is bounded since $|f_n(x)| = |\alpha_0 + \cdots + \alpha_{n-1}| \le n\|x\|$.

  - For each fixed $x \in X$, the sequence $(|f_n(x)|)$ satisfies $\|T_n x\| \le c_x$: A polynomial x of degree N_x has $N_x + 1$ coefficients. So $|f_n(x)| \le (N_x + 1)\max_j |\alpha_j| = c_x$.
  - We show there is no c such that $\|T_n\| = \|f_n\| \le c$, for all n. For f_n, we choose x defined by $x(t) = 1 + t + \cdots + t^n$. Then $\|f_n\| \ge \frac{|f_n(x)|}{\|x\|} = n$. So $(\|f_n\|)$ is unbounded.

Application: Fourier Series

- The **Fourier series** of a given periodic function x of period 2π is of the form
$$\frac{1}{2}a_0 + \sum_{m=1}^{\infty}(a_m \cos mt + b_m \sin mt)$$
with the Fourier coefficients of x given by the Euler formulas
$$a_m = \frac{1}{\pi}\int_0^{2\pi} x(t)\cos mt\, dt, \quad b_m = \frac{1}{\pi}\int_0^{2\pi} x(t)\sin mt\, dt.$$
It is well-known that the series may converge even at points where x is discontinuous. Thus, continuity is not necessary for convergence.
Surprising enough, continuity is not sufficient either:
Claim: There exist real-valued continuous functions whose Fourier series diverge at a given point t_0.
Let X be the normed space of all real-valued continuous functions of period 2π with norm defined by $\|x\| = \max|x(t)|$. X is a Banach space. We may take $t_0 = 0$, without restricting generality.

Fourier Series (Cont'd)

- To prove our statement, we shall apply the uniform boundedness theorem to $T_n = f_n$, where $f_n(x)$ is the value at $t = 0$ of the n-th partial sum of the Fourier series of x. Since for $t = 0$ the sine terms are zero and the cosine is one, we get

$$f_n(x) = \frac{1}{2}a_0 + \sum_{m=1}^{n} a_m = \frac{1}{\pi}\int_0^{2\pi} x(t)\left[\frac{1}{2} + \sum_{m=1}^{n} \cos mt\right] dt.$$

We want to determine the function represented by the sum under the integral sign. For this purpose we calculate

$$\begin{aligned} 2\sin\tfrac{1}{2}t\textstyle\sum_{m=1}^{n}\cos mt &= \textstyle\sum_{m=1}^{n} 2\sin\tfrac{1}{2}t\cos mt \\ &= \textstyle\sum_{m=1}^{n}[-\sin(m-\tfrac{1}{2})t + \sin(m+\tfrac{1}{2})t] \\ &= -\sin\tfrac{1}{2}t + \sin(n+\tfrac{1}{2})t. \end{aligned}$$

Dividing this by $\sin\frac{1}{2}t$ and adding 1 on both sides, we have $1 + 2\sum_{m=1}^{n}\cos mt = \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}$.

The Linear Functional f_n is Bounded

- Now the formula for $f_n(x)$ can be written in the simple form

$$f_n(x) = \frac{1}{2\pi}\int_0^{2\pi} x(t) q_n(t) dt, \quad q_n(t) = \frac{\sin\left(n+\frac{1}{2}\right)t}{\sin\frac{1}{2}t}.$$

Using this, we can show that the linear functional f_n is bounded: Using $\|x\| = \max|x(t)|$ and the preceding relations,

$$|f_n(x)| \leq \frac{1}{2\pi}\max|x(t)|\int_0^{2\pi}|q_n(t)|dt = \frac{\|x\|}{2\pi}\int_0^{2\pi}|q_n(t)|dt.$$

From this we see that f_n is bounded. Furthermore, by taking the supremum over all x of norm 1, we obtain $\|f_n\| \leq \frac{1}{2\pi}\int_0^{2\pi}|q_n(t)|dt$.

$\|f_n\| = \frac{1}{2\pi}\int_0^{2\pi}|q_n(t)|dt$

- We showed $\|f_n\| \le \frac{1}{2\pi}\int_0^{2\pi}|q_n(t)|dt$. Actually, the equality sign holds: To see this, first write $|q_n(t)| = y(t)q_n(t)$, where $y(t) = +1$ at every t at which $q_n(t) \ge 0$ and $y(t) = -1$ elsewhere. y is not continuous, but, for any given $\varepsilon > 0$, it may be modified to a continuous x of norm 1 such that, for this x, we have $\frac{1}{2\pi}|\int_0^{2\pi}[x(t)-y(t)]q_n(t)dt| < \varepsilon$. Writing this as two integrals, we obtain

$$\begin{aligned}\frac{1}{2\pi}\left|\int_0^{2\pi}x(t)q_n(t)dt - \int_0^{2\pi}y(t)q_n(t)dt\right| \\ = \left|f_n(x) - \frac{1}{2\pi}\int_0^{2\pi}|q_n(t)|dt\right| < \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary and $\|x\| = 1$, this proves

$$\|f_n\| = \frac{1}{2\pi}\int_0^{2\pi}|q_n(t)|dt.$$

The Sequence $(\|f_n\|)$ is Unbounded

- Substituting $q_n(t) = \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}$ into $\|f_n\| = \frac{1}{2\pi}\int_0^{2\pi}|q_n(t)|dt$, using the fact that $\left|\sin\frac{1}{2}t\right| < \frac{1}{2}t$ for $t \in (0, 2\pi]$ and substituting $v = (n+\frac{1}{2})t$,

$$\begin{aligned}\|f_n\| &= \frac{1}{2\pi}\int_0^{2\pi}\left|\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\right|dt > \frac{1}{\pi}\int_0^{2\pi}\frac{|\sin(n+\frac{1}{2})t|}{t}dt \\ &= \frac{1}{\pi}\int_0^{(2n+1)\pi}\frac{|\sin v|}{v}dv = \frac{1}{\pi}\sum_{k=0}^{2n}\int_{k\pi}^{(k+1)\pi}\frac{|\sin v|}{v}dv \\ &\geq \frac{1}{\pi}\sum_{k=0}^{2n}\frac{1}{(k+1)\pi}\int_{k\pi}^{(k+1)\pi}|\sin v|dv = \frac{2}{\pi^2}\sum_{k=0}^{2n}\frac{1}{k+1} \to \infty.\end{aligned}$$

  Hence $(\|f_n\|)$ is unbounded. Thus, with $T_n = f_n$, $\|T_n\| \leq c$ does not hold. Since X is complete, this implies that $\|T_n x\| \leq c_x$ cannot hold for all x. Hence, there must be an $x \in X$, such that $(|f_n(x)|)$ is unbounded. But, by the definition of the f_n's, this means that the Fourier series of that x diverges at $t = 0$.

Subsection 8

Strong and Weak Convergence

Introducing Weak Convergence

- In calculus one defines different types of convergence:
  - ordinary;
  - conditional;
  - absolute;
  - uniform.
- This yields greater flexibility in the theory and application of sequences and series.
- In functional analysis one has an even greater variety of possibilities that turn out to be of practical interest.
  - "Weak convergence" is a basic concept whose theory makes essential use, and is one of the major applications, of the uniform boundedness theorem.

Strong Convergence

- Convergence of sequences of elements in a normed space, as defined previously, will be called strong convergence, to distinguish it from "weak convergence" to be introduced on the following slide.

Definition (Strong Convergence)

A sequence (x_n) in a normed space X is said to be **strongly convergent** (or **convergent in the norm**) if there is an $x \in X$, such that

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

This is written $\lim_{n\to\infty} x_n = x$ or simply $x_n \to x$.
x is called the **strong limit** of (x_n), and we say that (x_n) **converges strongly to** x.

Weak Convergence

- Weak convergence is defined in terms of bounded linear functionals on X as follows:

Definition (Weak Convergence)

A sequence (x_n) in a normed space X is said to be **weakly convergent** if there is an $x \in X$, such that, for every $f \in X'$,

$$\lim_{n \to \infty} f(x_n) = f(x).$$

This is written $x_n \xrightarrow{w} x$ or $x_n \rightharpoonup x$.
The element x is called the **weak limit** of (x_n), and we say that (x_n) **converges weakly to** x.

- Note that weak convergence means convergence of the sequence of numbers $a_n = f(x_n)$, for every $f \in X'$.

The Weak Convergence Lemma

Lemma (Weak Convergence)

Let (x_n) be a weakly convergent sequence in a normed space X, say, $x_n \xrightarrow{w} x$. Then:

(a) The weak limit x of (x_n) is unique.

(b) Every subsequence of (x_n) converges weakly to x.

(c) The sequence $(\|x_n\|)$ is bounded.

(a) Suppose that $x_n \xrightarrow{w} x$ as well as $x_n \xrightarrow{w} y$. Then $f(x_n) \to f(x)$ as well as $f(x_n) \to f(y)$. Since $(f(x_n))$ is a sequence of numbers, its limit is unique. Hence $f(x) = f(y)$, that is, for every $f \in X'$, we have $f(x) - f(y) = f(x - y) = 0$. This implies $x - y = 0$ and shows that the weak limit is unique.

The Weak Convergence Lemma (Cont'd)

(b) follows from the fact that $(f(x_n))$ is a convergent sequence of numbers, so that every subsequence of $(f(x_n))$ converges and has the same limit as the sequence.

(c) Since $(f(x_n))$ is a convergent sequence of numbers, it is bounded, say, $|f(x_n)| \leq c_f$, for all n, where c_f is a constant depending on f but not on n. Using the canonical mapping $C : X \to X''$, we can define $g_n \in X''$ by $g_n(f) = f(x_n)$, for all $f \in X'$. (We write g_n instead of g_{x_n}.) Then for all n, $|g_n(f)| = |f(x_n)| \leq c_f$, that is, the sequence $(|g_n(f)|)$ is bounded for every $f \in X'$. Since X' is complete, by the Uniform Boundedness Theorem, $(\|g_n\|)$ is bounded. Now $\|g_n\| = \|x_n\|$, so that (c) is proved.

Relation Between Strong and Weak Convergence

- In finite dimensional normed spaces the distinction between strong and weak convergence disappears completely.

Theorem (Strong and Weak Convergence)

Let (x_n) be a sequence in a normed space X. Then:

(a) Strong convergence implies weak convergence with the same limit.

(b) The converse of (a) is not generally true.

(c) If $\dim X < \infty$, then weak convergence implies strong convergence.

(a) By definition, $x_n \to x$ means $\|x_n - x\| \to 0$. This implies, for all $f \in X'$,

$$|f(x_n) - f(x)| = |f(x_n - x)| \leq \|f\| \|x_n - x\| \to 0.$$

This shows that $x_n \xrightarrow{w} x$.

Relation Between Strong and Weak Convergence: Part (b)

(b) can be seen from an orthonormal sequence (e_n) in a Hilbert space H. In fact, every $f \in H'$ has a Riesz representation $f(x) = \langle x, z \rangle$. Hence, $f(e_n) = \langle e_n, z \rangle$. Now the Bessel inequality is $\sum_{n=1}^{\infty} |\langle e_n, z \rangle|^2 \le \|z\|^2$. Hence, the series on the left converges, so that its terms must approach zero as $n \to \infty$. This implies $f(e_n) = \langle e_n, z \rangle \to 0$. Since $f \in H'$ was arbitrary, we see that $e_n \xrightarrow{w} 0$. However, (e_n) does not converge strongly because, for all $m \neq n$,

$$\|e_m - e_n\|^2 = \langle e_m - e_n, e_m - e_n \rangle = 2.$$

Relation Between Strong and Weak Convergence: Part (c)

(c) Suppose that $x_n \xrightarrow{w} x$ and $\dim X = k$. Let $\{e_1, \ldots, e_k\}$ be any basis for X and, say, $x_n = \alpha_1^{(n)} e_1 + \cdots + \alpha_k^{(n)} e_k$ and $x = \alpha_1 e_1 + \cdots + \alpha_k e_k$. By assumption, $f(x_n) \to f(x)$, for every $f \in X'$. We take in particular $f_1, \ldots, f_k$ defined by $f_j(e_j) = 1$, $f_j(e_m) = 0$, for $m \neq j$. This is the dual basis of $\{e_1, \ldots, e_k\}$. Then $f_j(x_n) = \alpha_j^{(n)}$, $f_j(x) = \alpha_j$. Hence, $f_j(x_n) \to f_j(x)$ implies $\alpha_j^{(n)} \to \alpha_j$. From this we readily obtain

$$\|x_n - x\| = \left\| \sum_{j=1}^{k} (\alpha_j^{(n)} - \alpha_j) e_j \right\| \leq \sum_{j=1}^{k} |\alpha_j^{(n)} - \alpha_j| \|e_j\| \xrightarrow{n \to \infty} 0.$$

This shows that (x_n) converges strongly to x.

- It is interesting to note that there also exist infinite dimensional spaces such that strong and weak convergence are equivalent concepts.

The Weak Convergence Lemma

Lemma (Weak Convergence)

In a normed space X we have $x_n \xrightarrow{w} x$ if and only if:

(A) The sequence $(\|x_n\|)$ is bounded.

(B) For every element f of a total subset $M \subseteq X'$, we have $f(x_n) \to f(x)$.

- In the case of weak convergence, a preceding lemma gives (A), and (B) is trivial.

  Conversely, suppose that (A) and (B) hold. Let us consider any $f \in X'$ and show that $f(x_n) \to f(x)$, which is weak convergence, by definition.

  By (A), $\|x_n\| \le c$, for all n, and $\|x\| \le c$, where c is sufficiently large. Since M is total in X', for every $f \in X'$, there is a sequence (f_j) in $\operatorname{span} M$, such that $f_j \to f$. Hence, for any given $\varepsilon > 0$, we can find a j, such that $\|f_j - f\| < \frac{\varepsilon}{3c}$. Moreover, since $f_j \in \operatorname{span} M$, by (B), there is an N, such that, for all $n > N$, $|f_j(x_n) - f_j(x)| < \frac{\varepsilon}{3}$.

The Weak Convergence Lemma (Cont'd)

- We have

$$\|f_j - f\| < \frac{\varepsilon}{3c} \quad \text{and} \quad |f_j(x_n) - f_j(x)| < \frac{\varepsilon}{3}, \text{ all } n > N.$$

Using these two inequalities and applying the triangle inequality, we obtain for $n > N$,

$$\begin{array}{rcl} |f(x_n) - f(x)| & \leq & |f(x_n) - f_j(x_n)| + |f_j(x_n) - f_j(x)| + |f_j(x) - f(x)| \\ & < & \|f - f_j\| \|x_n\| + \frac{\varepsilon}{3} + \|f_j - f\| \|x\| \\ & < & \frac{\varepsilon}{3c} c + \frac{\varepsilon}{3} + \frac{\varepsilon}{3c} c = \varepsilon. \end{array}$$

Since $f \in X'$ was arbitrary, this shows that the sequence (x_n) converges weakly to x.

Examples

- **Hilbert Space**: In a Hilbert space, $x_n \xrightarrow{w} x$ if and only if $\langle x_n, z\rangle \to \langle x, z\rangle$, for all z in the space.
- **Space ℓ^p**: In ℓ^p, where $1 < p < +\infty$, we have $x_n \xrightarrow{w} x$ if and only if:
  (A) The sequence $(\|x_n\|)$ is bounded.
  (B) For every fixed j, $\xi_j^{(n)} \xrightarrow{n\to\infty} \xi_j$, where $x_n = (\xi_j^{(n)})$ and $x = (\xi_j)$.

  The dual space of ℓ^p is ℓ^q. A Schauder basis of ℓ^q is (e_n), where $e_n = (\delta_{nj})$ has 1 in the n-th place and zeros elsewhere. Span(e_n) is dense in ℓ^q, so the conclusion results from the Weak Convergence Lemma.

Subsection 9

Convergence of Sequences of Operators and Functionals

Sequences of Operators

- For sequences of elements in a normed space, strong and weak convergence as defined in the previous section are useful.
- For sequences of operators $T_n \in B(X, Y)$, three types of convergence turn out to be of theoretical as well as practical value:
  (1) Convergence in the norm on $B(X, Y)$;
  (2) Strong convergence of $(T_n x)$ in Y;
  (3) Weak convergence of $(T_n x)$ in Y.

Convergence of Sequences of Operators

Definition (Convergence of Sequences of Operators)

Let X and Y be normed spaces. A sequence (T_n) of operators $T_n \in B(X,Y)$ is said to be:

(1) **uniformly operator convergent** if (T_n) converges in the norm on $B(X,Y)$;

(2) **strongly operator convergent** if $(T_n x)$ converges strongly in Y, for every $x \in X$;

(3) **weakly operator convergent** if $(T_n x)$ converges weakly in Y, for every $x \in X$.

In formulas, this means that there is an operator $T : X \to Y$, such that:

(1) $\|T_n - T\| \to 0$;

(2) $\|T_n x - Tx\| \to 0$, for all $x \in X$;

(3) $|f(T_n x) - f(Tx)| \to 0$, for all $x \in X$ and all $f \in Y'$.

T is called the **uniform**, **strong** and **weak operator limit** of (T_n), resp.

Uniform versus Strong Operator Convergence

- It is not difficult to show that

$$\text{uniform} \Rightarrow \text{strong} \Rightarrow \text{weak operator convergence.}$$

- The converse is not generally true:
- Example (**Space** ℓ^2) In the space ℓ^2, we consider a sequence (T_n), where $T_n : \ell^2 \to \ell^2$ is defined by

$$T_n x = (0, 0, \ldots, 0, \xi_{n+1}, \xi_{n+2}, \ldots),$$

where, $x = (\xi_1, \xi_2, \ldots) \in \ell^2$.

This operator T_n is linear and bounded.

Clearly, (T_n) is strongly operator convergent to 0 since $T_n x \to 0 = 0x$.

However, (T_n) is not uniformly operator convergent since we have $\| T_n - 0 \| = \| T_n \| = 1$.

Strong versus Weak Operator Convergence

- Example (**Space** ℓ^2): Another sequence (T_n) of operators $T_n : \ell^2 \to \ell^2$ is defined by

$$T_n x = (\underbrace{0, 0, \ldots, 0}_{n \text{ zeros}}, \xi_1, \xi_2, \ldots),$$

where $x = (\xi_1, \xi_2, \ldots) \in \ell^2$.

This operator T_n is linear and bounded.

(T_n) is weakly operator convergent to 0 but not strongly:

  - Every bounded linear functional f on ℓ^2 has a Riesz representation $f(x) = \langle x, z \rangle = \sum_{j=1}^{\infty} \xi_j \overline{\zeta}_j$, where $z = (\zeta_j) \in \ell^2$.

Strong versus Weak Operator Convergence (Cont'd)

Hence, setting $j = n + k$ and using the definition of T_n, we have

$$f(T_n x) = \langle T_n x, z\rangle = \sum_{j=n+1}^{\infty} \xi_{j-n}\overline{\zeta}_j = \sum_{k=1}^{\infty} \xi_k \overline{\zeta}_{n+k}.$$

By the Cauchy-Schwarz inequality

$$|f(T_n x)|^2 = |\langle T_n x, z\rangle|^2 \leq \sum_{k=1}^{\infty} |\xi_k|^2 \sum_{m=n+1}^{\infty} |\zeta_m|^2.$$

The last series is the remainder of a convergent series. Hence, the right-hand side approaches 0 as $n \to \infty$. So $f(T_n x) \to 0 = f(0x)$ and (T_n) is weakly operator convergent to 0.

- (T_n) is not strongly operator convergent because for $x = (1, 0, 0, \ldots)$, we have, for all $m \neq n$, $\|T_m x - T_n x\| = \sqrt{1^2 + 1^2} = \sqrt{2}$.

The Case of Linear Functionals

- Linear functionals are linear operators (with range in the scalar field $\mathbb{R}$ or $\mathbb{C}$), so that the previous definitions apply immediately.
- In this case (2) and (3) become equivalent for the following reason:

  We had $T_n x \in Y$, but we now have $f_n(x) \in \mathbb{R}$ (or $\mathbb{C}$).

  Hence, convergence in (2) and (3) now takes place in the finite dimensional (one-dimensional) space $\mathbb{R}$ (or $\mathbb{C}$).

  So the equivalence of (2) and (3) follows from a preceding theorem.

Strong and Weak* Convergence

Definition (Strong and Weak* Convergence)

Let (f_n) be a sequence of bounded linear functionals on a normed space X. Then:

(a) **Strong convergence** of (f_n) means that there is an $f \in X'$, such that $\|f_n - f\| \to 0$. This is written $f_n \to f$.

(b) **Weak* convergence** of (f_n) means that there is an $f \in X'$, such that $f_n(x) \to f(x)$, for all $x \in X$. This is written $f_n \xrightarrow{w^*} f$.

f in (a) and (b) is called the **strong limit** and **weak* limit** of (f_n), respectively.

Properties of the Limit Operators

- Considering the limit operator $T: X \to Y$ of the sequence $T_n \in B(X, Y)$:
  - If the convergence is uniform, $T \in B(X, Y)$; otherwise, $\| T_n - T \|$ would not make sense.
  - If the convergence is strong or weak, T is still linear but may be unbounded if X is not complete.

Example: The space X of all sequences $x = (\xi_j)$ in ℓ^2 with only finitely many nonzero terms, taken with the metric on ℓ^2, is not complete.

A sequence of bounded linear operators T_n on X is defined by

$$T_n x = (\xi_1, 2\xi_2, 3\xi_3, \ldots, n\xi_n, \xi_{n+1}, \xi_{n+2}, \ldots),$$

so that $T_n x$ has terms $j\xi_j$ if $j \leq n$ and ξ_j if $j > n$.

This sequence (T_n) converges strongly to the unbounded linear operator T defined by $Tx = (\eta_j)$, where $\eta_j = j\xi_j$.

Strong Operator Convergence (Complete Domain)

- If X is complete, the situation illustrated by the example cannot occur:

Lemma (Strong Operator Convergence)

Let $T_n \in B(X,Y)$, where X is a Banach space and Y a normed space. If (T_n) is strongly operator convergent with limit T, then $T \in B(X,Y)$.

- Linearity of T follows readily from that of T_n. Since $T_n x \to Tx$, for every $x \in X$, the sequence $(T_n x)$ is bounded for every x. Since X is complete, $(\|T_n\|)$ is bounded by the Uniform Boundedness Theorem, say, $\|T_n\| \leq c$, for all n. Hence,

$$\|T_n x\| \leq \|T_n\| \|x\| \leq c\|x\|.$$

This implies $\|Tx\| \leq c\|x\|$.

Criterion of Strong Operator Convergence

Theorem (Strong Operator Convergence)

A sequence (T_n) of operators $T_n \in B(X, Y)$, where X and Y are Banach spaces, is strongly operator convergent if and only if:

(A) The sequence $(\|T_n\|)$ is bounded.

(B) The sequence $(T_n x)$ is Cauchy in Y, for every x in a total subset M of X.

- If $T_n x \to Tx$, for every $x \in X$, then (A) follows from the Uniform Boundedness Theorem (since X is complete), and (B) is trivial.

  Conversely, suppose that (A) and (B) hold, so that, say,

  $$\|T_n\| \le c, \text{ for all } n.$$

  We consider any $x \in X$ and show that $(T_n x)$ converges strongly in Y.

Criterion of Strong Operator Convergence (Cont'd)

- Let $\varepsilon > 0$ be given. Since span$M$ is dense in X, there is a $y \in \text{span}M$, such that $\|x - y\| < \frac{\varepsilon}{3c}$. Since $y \in \text{span}M$, the sequence $(T_n y)$ is Cauchy by (B). Hence, there is an N, such that $\|T_n y - T_m y\| < \frac{\varepsilon}{3}$, for all $m, n > N$. Using these two inequalities and applying the triangle inequality, we readily see that $(T_n x)$ is Cauchy in Y: For $m, n > N$,

$$\begin{array}{rcl} \|T_n x - T_m x\| & \leq & \|T_n x - T_n y\| + \|T_n y - T_m y\| + \|T_m y - T_m x\| \\ & < & \|T_n\| \|x - y\| + \frac{\varepsilon}{3} + \|T_m\| \|x - y\| \\ & < & c\frac{\varepsilon}{3c} + \frac{\varepsilon}{3} + c\frac{\varepsilon}{3c} = \varepsilon. \end{array}$$

  Since Y is complete, $(T_n x)$ converges in Y. Since $x \in X$ was arbitrary, this proves strong operator convergence of (T_n).

Criterion for Weak* Convergence of Functionals

Corollary (Functionals)

A sequence (f_n) of bounded linear functionals on a Banach space X is weak* convergent, the limit being a bounded linear functional on X, if and only if:

(A) The sequence $(\|f_n\|)$ is bounded.

(B) The sequence $(f_n(x))$ is Cauchy for every x in a total subset M of X.

Subsection 10

Application to Summability of Sequences

Summability Methods

- A divergent sequence has no limit in the usual sense.
- The theory of divergent sequences aims at associating with certain divergent sequences a "limit" in a generalized sense.
- A procedure for that purpose is called a **summability method**.

  Example: A divergent sequence $x = (\xi_k)$ being given, we may calculate the sequence $y = (\eta_n)$ of the arithmetic means

$$\eta_1 = \xi_1,\ \eta_2 = \frac{1}{2}(\xi_1 + \xi_2), \ldots, \eta_n = \frac{1}{n}(\xi_1 + \cdots + \xi_n), \ldots$$

  If y converges with limit η (in the usual sense), we say that x is **summable** by the present method and has the **generalized limit** η.

  For instance, if $x = (0,1,0,1,0,\ldots)$, then $y = (0, \frac{1}{2}, \frac{1}{3}, \frac{1}{2}, \frac{2}{5}, \ldots)$ and x has the generalized limit $\frac{1}{2}$.

Matrix Summability Methods

- A summability method is called a **matrix method** if it can be represented in the form $y = Ax$, where $x = (\xi_k)$ and $y = (\eta_n)$ are written as infinite column vectors and $A = (\alpha_{nk})$ is an infinite matrix.
- In the formula $y = Ax$ we used matrix multiplication, that is, y has the terms
$$\eta_n = \sum_{k=1}^{\infty} \alpha_{nk}\xi_k, \quad n = 1, 2, \ldots.$$
- The preceding example illustrates a matrix method.
- The method given here is briefly called an A-**method** because the corresponding matrix is denoted by A.
- If the series η_n converges, for all n, and $y = (\eta_n)$ converges in the usual sense, its limit is called the A-**limit** of x, and x is said to be A-**summable**.
- The set of all A-summable sequences is called the **range** of the A-method.

Regular Matrix Summability Methods

- An A-method is said to be **regular** (or **permanent**) if its range includes all convergent sequences and if, for every such sequence, the A-limit equals the usual limit, that is, if $\xi_k \to \xi$ implies $\eta_n \to \xi$.
- A method which is not applicable to certain convergent sequences or alters their limit would be of no practical use.

Toeplitz Limit Theorem (Regular Summability Methods)

An A-summability method with matrix $A = (\alpha_{nk})$ is regular if and only if:

(1) $\lim_{n\to\infty} \alpha_{nk} = 0$, for $k = 1, 2, \ldots$;

(2) $\lim_{n\to\infty} \sum_{k=1}^{\infty} \alpha_{nk} = 1$;

(3) $\sum_{k=1}^{\infty} |\alpha_{nk}| \leq \gamma$, for $n = 1, 2, \ldots$, where γ is a constant which does not depend on n.

Necessity of the Conditions

(a) Suppose that the A-method is regular. Let x_k have 1 as the k-th term and all other terms zero. For x_k we have $\eta_n = \alpha_{nk}$. Since x_k is convergent and has the limit 0, this shows that (1) must hold.

Furthermore, $x = (1, 1, 1, \ldots)$ has the limit 1. Now η_n equals the series in (2). Consequently, (2) must hold.

We prove that (3) is necessary for regularity: Let c be the Banach space of all convergent sequences with norm defined by $\|x\| = \sup_j |\xi_j|$. Linear functionals f_{nm} on c are defined by $f_{nm}(x) = \sum_{k=1}^{m} \alpha_{nk}\xi_k$, $m, n = 1, 2, \ldots$. Each f_{nm} is bounded since

$$|f_{nm}(x)| \le \sup_j |\xi_j| \sum_{k=1}^{m} |\alpha_{nk}| = \left(\sum_{k=1}^{m} |\alpha_{nk}|\right) \|x\|.$$

Regularity implies the convergence of the series η_n, for all $x \in C$.

Necessity of the Conditions (Cont'd)

- Hence, it defines linear functionals $f_1, f_2, \ldots$ on c given by $\eta_n = f_n(x) = \sum_{k=1}^{\infty} \alpha_{nk}\xi_k$, $n = 1, 2, \ldots$. Thus, $f_{nm}(x) \to f_n(x)$ as $m \to \infty$, for all $x \in c$. This is weak* convergence, and f_n is bounded. Also, $(f_n(x))$ converges for all $x \in c$, and $(\|f_n\|)$ is bounded, say, $\|f_n\| \leq \gamma$. For an arbitrary fixed $m \in \mathbb{N}$, define

$$\xi_k^{(n,m)} = \begin{cases} \frac{|\alpha_{nk}|}{\alpha_{nk}}, & \text{if } k \leq m \text{ and } \alpha_{nk} \neq 0 \\ 0, & \text{if } k > m \text{ or } \alpha_{nk} = 0 \end{cases}.$$

Then $x_{nm} = (\xi_k^{(n,m)}) \in c$. Also $\|x_{nm}\| = 1$, if $x_{nm} \neq 0$, and $\|x_{nm}\| = 0$, if $x_{nm} = 0$. Furthermore, for all m,

$$f_{nm}(x_{nm}) = \sum_{k=1}^{m} \alpha_{nk}\xi_k^{(n,m)} = \sum_{k=1}^{m} |\alpha_{nk}|.$$

Hence, $\sum_{k=1}^{m} |\alpha_{nk}| = f_{nm}(x_{nm}) \leq \|f_{nm}\|$ and $\sum_{k=1}^{\infty} |\alpha_{nk}| \leq \|f_n\|$. This shows that the series in (3) converges.

Sufficiency of the Conditions

- We prove that (1) to (3) is sufficient for regularity.

  We define a linear functional f on c by $f(x) = \xi = \lim_{k\to\infty} \xi_k$, where $x = (\xi_k) \in c$.

  Boundedness of f can be seen from $|f(x)| = |\xi| \le \sup_j |\xi_j| = \|x\|$.

  Let $M \subseteq c$ be the set of all sequences whose terms are equal from some term on, say, $x = (\xi_k)$, where $\xi_j = \xi_{j+1} = \xi_{j+2} = \cdots = \xi$, and j depends on x. Then $f(x) = \xi$, and we obtain

$$\eta_n = f_n(x) = \sum_{k=1}^{j-1} \alpha_{nk}\xi_k + \xi \sum_{k=j}^{\infty} \alpha_{nk} = \sum_{k=1}^{j-1} \alpha_{nk}(\xi_k - \xi) + \xi \sum_{k=1}^{\infty} \alpha_{nk}.$$

  Hence by (1) and (2), for all $n \in M$, $\eta_n = f_n(x) \to 0 + \xi \cdot 1 = \xi = f(x)$.

Sufficiency of the Conditions (Cont'd)

- We show next that the set M on which we have the convergence $\eta_n = f_n(x) \to f(x)$ is dense in c.

  Let $x = (\xi_k) \in c$ with $\xi_k \to \xi$. Then, for every $\varepsilon > 0$, there is an N, such that $|\xi_k - \xi| < \varepsilon$, for all $k \geq N$. Clearly, $\widetilde{x} = (\xi_1, \ldots, \xi_{N-1}, \xi, \xi, \ldots) \in M$ and $x - \widetilde{x} = (0, \ldots, 0, \xi_N - \xi, \xi_{N+1} - \xi, \ldots)$. It follows that $\|x - \widetilde{x}\| \leq \varepsilon$. Since $x \in c$ was arbitrary, this shows that M is dense in c.

  Finally, by (3), $|f_n(x)| \leq \|x\| \sum_{k=1}^{\infty} |\alpha_{nk}| \leq \gamma \|x\|$, for all $x \in c$ and all n.

  Hence, $\|f_n\| \leq \gamma$, that is, $(\|f_n\|)$ is bounded.

  Furthermore, $f(x_n) \to f(x)$ gives convergence for all x in a dense M.

  By a preceding corollary, this implies weak* convergence $f_n \xrightarrow{w^*} f$.

  Thus, we have shown that, if $\xi = \lim \xi_k$ exists, it follows that $\eta_n \to \xi$.

  By definition, this means regularity and the theorem is proved.

Subsection 11

Numerical Integration and Weak* Convergence

Integral Approximation Methods

- We consider **numerical integration**, that is, the problem of obtaining approximate values for a given integral $\int_a^b x(t)dt$.
- Various methods have been developed, e.g., the trapezoidal rule, Simpson's rule, Newton-Cotes and the Gauss methods.
- The common feature of those methods:
  - We choose points in $[a, b]$, called **nodes**;
  - We approximate the unknown value of the integral by a linear combination of the values of x at the nodes.
- The nodes and the coefficients of that linear combination depend on the method but not on the integrand x.
- The usefulness of a method is determined by its **accuracy**, and one may want the accuracy to increase as the number of nodes gets larger.
- We employ functional analysis to describe a general setting for those methods and consider the problem of convergence as the number of nodes increases.

The General Method

- We deal with continuous functions, so introduce the Banach space $X = C[a,b]$ of all continuous real-valued functions on $J = [a,b]$, with norm defined by $\|x\| = \max_{t \in J} |x(t)|$.
- Then $f(x) = \int_a^b x(t)dt$ defines a linear functional f on X.
- For each positive integer n, we choose $n+1$ real numbers (called **nodes**) $t_0^{(n)}, \ldots, t_n^{(n)}$, such that $a \le t_0^{(n)} < \cdots < t_n^{(n)} \le b$.
- Then we choose $n+1$ real numbers (called **coefficients**) $\alpha_0^{(n)}, \ldots, \alpha_n^{(n)}$.
- We define linear functionals f_n on X by setting

$$f_n(x) = \sum_{k=0}^{n} \alpha_k^{(n)} x(t_k^{(n)}), \quad n = 1, 2, \ldots.$$

The Norm of f_n

- Each f_n is bounded since $|x(t_k^{(n)})| \leq \|x\|$ by the definition of the norm.
- Consequently,

$$|f_n(x)| \leq \sum_{k=0}^{n} |\alpha_k^{(n)}||x(t_k^{(n)})| \leq \left(\sum_{k=0}^{n} |\alpha_k^{(n)}|\right) \|x\|.$$

Claim: f_n has norm $\|f_n\| = \sum_{k=0}^{n} |\alpha_k^{(n)}|$.

The preceding inequality shows that $\|f_n\| \leq \sum_{k=0}^{n} |\alpha_k^{(n)}|$.
Equality follows if we take an $x_0 \in X$, such that $|x_0(t)| \leq 1$ on J and

$$x_0(t_k^{(n)}) = \operatorname{sgn}\alpha_k^{(n)} = \begin{cases} 1, & \text{if } \alpha_k^{(n)} \geq 0 \\ -1, & \text{if } \alpha_k^{(n)} < 0 \end{cases}.$$

Then $\|x_0\| = 1$ and $f_n(x_0) = \sum_{k=0}^{n} \alpha_k^{(n)} \operatorname{sgn}\alpha_k^{(n)} = \sum_{k=0}^{n} |\alpha_k^{(n)}|$.

Convergence and Requirement on Polynomials

Definition (Convergence)

The numerical process of integration defined by $f_n(x) = \sum_{k=0}^{n} \alpha_k^{(n)} x(t_k^{(n)})$ is said to be **convergent** for an $x \in X$ if, for that x, $f_n(x) \stackrel{n\to\infty}{\longrightarrow} f(x)$, where f is defined by $f(x) = \int_a^b x(t)dt$.

- Given that exact integration of polynomials is easy, it is natural to make the following

Requirement

For every n, if x is a polynomial of degree $\leq n$, then $f_n(x) = f(x)$.

Simplifying the Requirement

- Since the f_n's are linear, it suffices to impose the preceding requirement for the $n+1$ powers defined by

$$x_0(t)=1,\ x_1(t)=t,\ \ldots,\ x_n(t)=t^n.$$

Then, for a polynomial of degree n given by $x(t)=\sum \beta_j t^j$, we get

$$f_n(x)=\sum_{j=0}^{n}\beta_j f_n(x_j)=\sum_{j=0}^{n}\beta_j f(x_j)=f(x).$$

We thus have the $n+1$ conditions $f_n(x_j)=f(x_j)$, $j=0,1,\ldots,n$.

Feasibility of the Requirement

- We show that $f_n(x_j) = f(x_j)$, $j = 0, 1, \ldots, n$, can be fulfilled.
- Since we have $2n+2$ parameters ($n+1$ nodes and $n+1$ coefficients), we can choose some of them in an arbitrary fashion.

  Claim: If we choose the $t_k^{(n)}$, we can determine the $\alpha_k^{(n)}$ uniquely.

  In $f_n(x_j) = f(x_j)$, we have $x_j(t_k^{(n)}) = (t_k^{(n)})^j$. So, for $j = 0, \ldots, n$, $\sum_{k=0}^{n} \alpha_k^{(n)} (t_k^{(n)})^j = \int_a^b t^j dt = \frac{1}{j+1}(b^{j+1} - a^{j+1})$. For each fixed n, this is a nonhomogeneous system of $n+1$ linear equations in the $n+1$ unknowns $\alpha_0^{(n)}, \ldots, \alpha_n^{(n)}$. A unique solution exists if:

  - The homogeneous system $\sum_{k=0}^{n} (t_k^{(n)})^j \gamma_k = 0$, $j = 0, \ldots, n$, has only the trivial solution $\gamma_0 = 0, \ldots, \gamma_n = 0$;
  - Equivalently, if the same holds for the system $\sum_{j=0}^{n} (t_k^{(n)})^j \gamma_j = 0$, $k = 0, \ldots, n$, whose coefficient matrix is the transpose of the coefficient matrix of the previous system.

  This holds, since $\sum_{j=0}^{n} \gamma_j t^j$, which is of degree n, being zero at the $n+1$ nodes, must be identically zero, i.e., $\gamma_j = 0$.

Weierstraß Approximation Theorem for Polynomials

Weierstraß Approximation Theorem (Polynomials)

The set W of all polynomials with real coefficients is dense in the real space $C[a,b]$. Hence, for every $x \in C[a,b]$ and, given $\varepsilon > 0$, there exists a polynomial p, such that $|x(t) - p(t)| < \varepsilon$, for all $t \in [a,b]$.

- Every $x \in C[a,b]$ is uniformly continuous on $J = [a,b]$ since J is compact. Hence, for any $\varepsilon > 0$, there is a y whose graph is an arc of a polygon such that $\max_{t \in J} |x(t) - y(t)| < \frac{\varepsilon}{3}$.

  Assume, first, that $x(a) = x(b)$ and $y(a) = y(b)$.

  Since y is piecewise linear and continuous, by applying integration by parts to the formulas for the Fourier coefficients a_m and b_m, we get bounds of the form

$$|a_0| < k, \quad |a_m| < \frac{k}{m^2}, \quad |b_m| < \frac{k}{m^2}.$$

Weierstraß Approximation Theorem (Cont'd)

- Hence for the Fourier series of y (representing the periodic extension of y, of period $b-a$), we have, writing $\kappa = \frac{2\pi}{b-a}$ for simplicity,

$$\left|a_0 + \sum_{m=1}^{\infty}(a_m \cos \kappa mt + b_m \sin \kappa mt)\right| \leq 2k(1 + \sum_{m=1}^{\infty} \frac{1}{m^2}) = 2k(1 + \frac{1}{6}\pi^2).$$

This shows that the series converges uniformly on J.

Consequently, for the n-th partial sum s_n, with sufficiently large n,

$$\max_{t \in J} |y(t) - s_n(t)| < \frac{\varepsilon}{3}.$$

The Taylor series of the cosine and sine functions in s_n also converge uniformly on J. So there is a polynomial p (obtained, for instance, from suitable partial sums of those series) such that $\max_{t \in J} |s_n(t) - p(t)| < \frac{\varepsilon}{3}$.

Weierstraß Approximation Theorem (Cont'd)

- Now we get
$$|x(t)-p(t)| \le |x(t)-y(t)| + |y(t)-s_n(t)| + |s_n(t)-p(t)|,$$
whence
$$\max_{t\in J}|x(t)-p(t)| < \varepsilon.$$
This takes care of every $x \in C[a,b]$, such that $x(a)=x(b)$.
Suppose, next, that $x(a) \ne x(b)$.
  - Take $u(t)=x(t)-\gamma(t-a)$, with γ such that $u(a)=u(b)$.
  - For u there is a polynomial q satisfying $|u(t)-q(t)|<\varepsilon$ on J.
  - $p(t)=q(t)+\gamma(t-a)$ satisfies $\max_{t\in J}|x(t)-p(t)|<\varepsilon$ (since $x-p=u-q$).

Since $\varepsilon>0$ was arbitrary, we have shown that W is dense in $C[a,b]$.

Pólya Convergence Theorem (Numerical Integration)

- We showed that for every choice of nodes $t_k^{(n)}$, $a \le t_0^{(n)} < \cdots < t_n^{(n)} \le b$, there are uniquely determined $\alpha_k^{(n)}$, such that $f_n(x) = f(x)$.
  Hence, the corresponding process is convergent for all polynomials.

Pólya Convergence Theorem (Numerical Integration)

A process of numerical integration $f_n(x) = \sum_{k=}^n \alpha_k^{(n)} x(t_k^{(n)})$ which satisfies $f_n(x) = f(x)$, for every n and polynomial x of degree not exceeding n, converges for all real-valued continuous functions on $[a,b]$ if and only if there is a number c, such that $\sum_{k=0}^n |\alpha_k^{(n)}| \le c$, for all n.

- The set W of all polynomials with real coefficients is dense in the real space $X = C[a,b]$, by the Weierstraß approximation theorem, and, for every $x \in W$, we have convergence by the Requirement. Since $\|f_n\| = \sum_{k=0}^n |\alpha_k^{(n)}|$, $(\|f_n\|)$ is bounded if and only if $\sum_{k=0}^n |\alpha_k^{(n)}| \le c$ holds for some real number c. The theorem now follows, since convergence $f_n(x) \to f(x)$, for all $x \in X$, is weak* convergence $f_n \xrightarrow{w^*} f$.

Steklov's Theorem for Numerical Integration

- In this theorem we may replace the polynomials by any other set which is dense in the real space $C[a,b]$.

Steklov's Theorem (Numerical Integration)

A process of numerical integration f_n which satisfies the Requirement and has nonnegative coefficients $\alpha_k^{(n)}$, converges for every continuous function.

- Suppose that the coefficients are all nonnegative.

  Taking $x=1$, we then have

$$\sum_{k=0}^{n} |\alpha_k^{(n)}| = \sum_{k=0}^{n} \alpha_k^{(n)} = f_n(1) = f(1) = \int_a^b dt = b - a.$$

  So $\sum_{k=0}^{n} |\alpha_k^{(n)}| \leq c$ holds.

Subsection 12

Open Mapping Theorem

Open Mapping

Definition (Open Mapping)

Let X and Y be metric spaces. Then $T:\mathscr{D}(T)\to Y$ with domain $\mathscr{D}(T)\subseteq X$ is called an **open mapping** if for every open set in $\mathscr{D}(T)$ the image is an open set in Y.

- If a mapping is not surjective, one must distinguish between the assertions that the mapping is open as a mapping from its domain
  - (a) into Y;
  - (b) onto its range.

  (b) is weaker than (a): For instance, if $X\subseteq Y$:
  - The mapping $x\mapsto x$ of X into Y is open if and only if X is an open subset of Y;
  - The mapping $x\mapsto x$ of X onto its range (which is X) is open in any case.

Open Mappings versus Continuous Mappings

- A continuous mapping $T: X \to Y$ has the property that for every open set in Y the inverse image is an open set in X.
- This does not imply that T maps open sets in X onto open sets in Y.

  Example: The mapping $\mathbb{R} \to \mathbb{R}$ given by $t \mapsto \sin t$ is continuous but maps $(0, 2\pi)$ onto $[-1, 1]$.

Open Unit Ball Lemma

Lemma (Open Unit Ball)

A bounded linear operator T from a Banach space X onto a Banach space Y has the property that the image $T(B_0)$ of the open unit ball $B_0 = B(0;1) \subseteq X$ contains an open ball about $0 \in Y$.

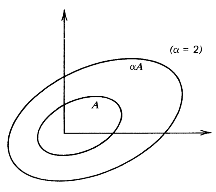
- Proceeding stepwise, we prove:
  - (a) The closure of the image of the open ball $B_1 = B(0;\frac{1}{2})$ contains an open ball B^*.
  - (b) $\overline{T(B_n)}$ contains an open ball V_n about $0 \in Y$, where $B_n = B(0;\frac{1}{2^n}) \subseteq X$.
  - (c) $T(B_0)$ contains an open ball about $0 \in Y$.

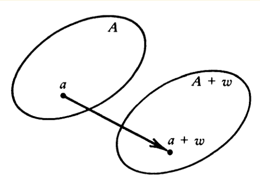
Proof of Open Unit Ball Part (a)

(a) In connection with subsets $A \subseteq X$ we shall write αA (α a scalar) and $A + w$ ($w \in X$) to mean

(1) $\alpha A = \{x \in X : x = \alpha a, a \in A\}$;

(2) $A + w = \{x \in X : x = a + w, a \in A\}$;





and, similarly, for subsets of Y.

- We consider the open ball $B_1 = B(0; \frac{1}{2}) \subseteq X$. Any fixed $x \in X$ is in kB_1 with real k sufficiently large ($k > 2\|x\|$). Hence $X = \bigcup_{k=1}^{\infty} kB_1$.

Since T is surjective and linear,

$$Y = T(X) = T\left(\bigcup_{k=1}^{\infty} kB_1\right) = \bigcup_{k=1}^{\infty} kT(B_1) = \bigcup_{k=1}^{\infty} \overline{kT(B_1)}.$$

Proof of Open Unit Ball Part (a) (Cont'd)

- $Y = T(X) = T(\bigcup_{k=1}^{\infty} kB_1) = \bigcup_{k=1}^{\infty} kT(B_1) = \bigcup_{k=1}^{\infty} \overline{kT(B_1)}$.

  Note that by taking closures we did not add further points to the union since that union was already the whole space Y.

  Since Y is complete, by Baire's Category, it is nonmeager in itself.

  Hence, we conclude that a $\overline{kT(B_1)}$ must contain some open ball.

  This implies that $\overline{T(B_1)}$ also contains an open ball, say,

  $$B^* = B(y_0; \varepsilon) \subseteq \overline{T(B_1)}.$$

  It follows that

  $$B^* - y_0 = B(0; \varepsilon) \subseteq \overline{T(B_1)} - y_0.$$

Proof of Open Unit Ball Part (b)

(b) We prove that $B^* - y_0 \subseteq \overline{T(B_0)}$, where B_0 is given in the theorem. This we do by showing that $\overline{T(B_1)} - y_0 \subseteq \overline{T(B_0)}$.

Let $y \in \overline{T(B_1)} - y_0$. Then $y + y_0 \in \overline{T(B_1)}$, and we remember that $y_0 \in \overline{T(B_1)}$, too. Thus, there are

$$u_n = Tw_n \in T(B_1), \text{ such that } u_n \to y + y_0,$$
$$v_n = Tz_n \in T(B_1), \text{ such that } v_n \to y_0.$$

Since $w_n, z_n \in B_1$ and B_1 has radius $\frac{1}{2}$, it follows that

$$\|w_n - z_n\| \leq \|w_n\| + \|z_n\| < \frac{1}{2} + \frac{1}{2} = 1.$$

So $w_n - z_n \in B_0$. From $T(w_n - z_n) = Tw_n - Tz_n = u_n - v_n \to y$, we see that $y \in \overline{T(B_0)}$. This proves $\overline{T(B_1)} - y_0 \subseteq \overline{T(B_0)}$.

From (a), we thus have $B^* - y_0 = B(0;\varepsilon) \subseteq \overline{T(B_0)}$. Let $B_n = B(0; \frac{1}{2^n}) \subseteq X$. Since T is linear, $\overline{T(B_n)} = \frac{1}{2^n}\overline{T(B_0)}$. Since $B^* - y_0 = B(0;\varepsilon) \subseteq \overline{T(B_0)}$, we thus obtain $V_n = B(0; \frac{\varepsilon}{2^n}) \subseteq \overline{T(B_n)}$.

Proof of Open Unit Ball Part (c)

(c) We prove $V_1 = B(0; \frac{\varepsilon}{2}) \subseteq T(B_0)$ by showing that $y \in V_1$ is in $T(B_0)$.

Let $y \in V_1$. From $V_n = B(0; \frac{\varepsilon}{2^n}) \subseteq \overline{T(B_n)}$, with $n = 1$, we have $V_1 \subseteq \overline{T(B_1)}$. Hence, $y \in \overline{T(B_1)}$. So, there exists $v \in T(B_1)$ close to y, say, $\|y - v\| < \frac{\varepsilon}{4}$. Now $v \in T(B_1)$ implies $v = Tx_1$, for some $x_1 \in B_1$. Hence, $\|y - Tx_1\| < \frac{\varepsilon}{4}$.

From this and (b), with $n = 2$, we see that $y - Tx_1 \in V_2 \subseteq \overline{T(B_2)}$. As before, there is an $x_2 \in B_2$, such that $\|(y - Tx_1) - Tx_2\| < \frac{\varepsilon}{8}$. Hence $y - Tx_1 - Tx_2 \in V_3 \subseteq \overline{T(B_3)}$, and so on. In the n-th step we can choose an $x_n \in B_n$, such that $\|y - \sum_{k=1}^n Tx_k\| < \frac{\varepsilon}{2^{n+1}}$, $n = 1, 2, \ldots$. Let $z_n = x_1 + \cdots + x_n$. Since $x_k \in B_k$, we have $\|x_k\| < \frac{1}{2^k}$. This yields for $n > m$, $\|z_n - z_m\| \le \sum_{k=m+1}^n \|x_k\| < \sum_{k=m+1}^\infty \frac{1}{2^k} \xrightarrow{m \to \infty} 0$. Hence (z_n) is Cauchy. (z_n) converges, say, $z_n \to x$, since X is complete. Also $x \in B_0$, since B_0 has radius 1 and $\sum_{k=1}^\infty \|x_k\| < \sum_{k=1}^\infty \frac{1}{2^k} = 1$. Since T is continuous, $Tz_n \to Tx$, and $\|y - \sum_{k=1}^n Tx_k\| < \frac{\varepsilon}{2^{n+1}}$ shows that $Tx = y$.

The Open Mapping Theorem

Open Mapping Theorem, Bounded Inverse Theorem

A bounded linear operator T from a Banach space X onto a Banach space Y is an open mapping. Hence if T is bijective, T^{-1} is continuous and thus bounded.

- We prove that for every open set $A \subseteq X$, the image $T(A)$ is open in Y. It suffices to show that, for every $y = Tx \in T(A)$, the set $T(A)$ contains an open ball about $y = Tx$.

  Let $y = Tx \in T(A)$. Since A is open, it contains an open ball with center x. Hence $A - x$ contains an open ball $B(0; r)$. Set $k = \frac{1}{r}$, so that $r = \frac{1}{k}$. Then $k(A - x)$ contains the open unit ball $B(0; 1)$. By the lemma, $T(k(A - x)) = k[T(A) - Tx]$ contains an open ball about 0, and so does $T(A) - Tx$. Hence $T(A)$ contains an open ball about $Tx = y$. Since $y \in T(A)$ was arbitrary, $T(A)$ is open.

  Finally, if $T^{-1} : Y \to X$ exists, it is continuous because T is open. Since T^{-1} is linear, it is bounded.

Subsection 13

Closed Linear Operators and Closed Graph Theorem

Closed Linear Operator

Definition (Closed Linear Operator)

Let X and Y be normed spaces and $T:\mathscr{D}(T)\to Y$ a linear operator with domain $\mathscr{D}(T)\subseteq X$. Then T is called a **closed linear operator** if its **graph**

$$\mathscr{G}(T)=\{(x,y):x\in\mathscr{D}(T),y=Tx\}$$

is closed in the normed space $X\times Y$, where the two algebraic operations of a vector space in $X\times Y$ are defined as usual, that is,

$$\begin{array}{rcl}(x_1,y_1)+(x_2,y_2) & = & (x_1+x_2,y_1+y_2),\\ \alpha(x,y) & = & (\alpha x,\alpha y), \quad \alpha \text{ a scalar,}\end{array}$$

and the norm on $X\times Y$ is defined by

$$\|(x,y)\|=\|x\|+\|y\|.$$

Closed Graph Theorem

Closed Graph Theorem

Let X and Y be Banach spaces and $T:\mathscr{D}(T)\to Y$ a closed linear operator, where $\mathscr{D}(T)\subseteq X$. Then, if $\mathscr{D}(T)$ is closed in X, the operator T is bounded.

- We first show that $X\times Y$ is complete. Let (z_n) be Cauchy in $X\times Y$, where $z_n=(x_n,y_n)$. Then, for every $\varepsilon>0$, there is an N, such that

$$\|z_n-z_m\|=\|x_n-x_m\|+\|y_n-y_m\|<\varepsilon,\quad m,n>N.$$

Hence (x_n) and (y_n) are Cauchy in X and Y, respectively, and converge, say, $x_n\to x$ and $y_n\to y$, because X and Y are complete. This implies that $z_n\to z=(x,y)$ since, by the inequality above with $m\to\infty$, we have $\|z_n-z\|\le\varepsilon$, for $n>N$. Since the Cauchy sequence (z_n) was arbitrary, $X\times Y$ is complete.

Closed Graph Theorem (Cont'd)

- By assumption, $\mathscr{G}(T)$ is closed in $X \times Y$ and $\mathscr{D}(T)$ is closed in X. Hence $\mathscr{G}(T)$ and $\mathscr{D}(T)$ are complete. We now consider the mapping

$$\begin{array}{rrcl} P: & \mathscr{G}(T) & \to & \mathscr{D}(T); \\ & (x, Tx) & \mapsto & x. \end{array}$$

  P is linear.

  P is bounded because $\|P(x, Tx)\| = \|x\| \leq \|x\| + \|Tx\| = \|(x, Tx)\|$.

  P is bijective: The inverse mapping is $P^{-1} : \mathscr{D}(T) \to \mathscr{G}(T)$; $x \mapsto (x, Tx)$.

  Since $\mathscr{G}(T)$ and $\mathscr{D}(T)$ are complete, we can apply the Bounded Inverse Theorem and see that P^{-1} is bounded, say, $\|(x, Tx)\| \leq b\|x\|$, for some b and all $x \in \mathscr{D}(T)$. Hence T is bounded, as for all $x \in \mathscr{D}(T)$,

$$\|Tx\| \leq \|Tx\| + \|x\| = \|(x, Tx)\| \leq b\|x\|.$$

Criterion for Closedness

Theorem (Closed Linear Operator)

Let $T : \mathscr{D}(T) \to Y$ be a linear operator, where $\mathscr{D}(T) \subseteq X$ and X and Y are normed spaces. Then T is closed if and only if it has the following property: If $x_n \to x$, where $x_n \in \mathscr{D}(T)$, and $Tx_n \to y$, then $x \in \mathscr{D}(T)$ and $Tx = y$.

- $\mathscr{G}(T)$ is closed if and only if $z = (x, y) \in \overline{\mathscr{G}(T)}$ implies $z \in \mathscr{G}(T)$.
  We know that:
  - $z = (x, y) \in \overline{\mathscr{G}(T)}$ if and only if there are $z_n = (x_n, Tx_n) \in \mathscr{G}(T)$, such that $z_n \to z$, i.e., $x_n \to x$, $Tx_n \to y$;
  - $z = (x, y) \in \mathscr{G}(T)$ if and only if $x \in \mathscr{D}(T)$ and $y = Tx$.

  Putting these together, we get the conclusion.

Closedness versus Continuity

- The following property of a bounded linear operator is different:

  > If a linear operator T is bounded and thus continuous, and if (x_n) is a sequence in $\mathscr{D}(T)$ which converges in $\mathscr{D}(T)$, then (Tx_n) also converges.

  This need not hold for a closed linear operator.
- However, if T is closed and two sequences (x_n) and $(\widetilde{x}_n)$ in the domain of T converge with the same limit and if the corresponding sequences (Tx_n) and $(T\widetilde{x}_n)$ both converge, then the latter have the same limit.

Example: Differential Operator

- Let $X = C[0,1]$ and $T : \mathscr{D}(T) \to X;\ x \mapsto x'$, where the prime denotes differentiation and $\mathscr{D}(T)$ is the subspace of functions $x \in X$ which have a continuous derivative. Then T is not bounded, but is closed.

  We see, using the sequence $x_n(t) = t^n$, that T is not bounded.

  We prove that T is closed by applying the preceding theorem.

  Let (x_n) in $\mathscr{D}(T)$ be such that both (x_n) and (Tx_n) converge, say, $x_n \to x$ and $Tx_n = x_n' \to y$. Since convergence in the norm of $C[0,1]$ is uniform convergence on $[0,1]$, from $x_n' \to y$, we have

$$\int_0^t y(\tau)d\tau = \int_0^t \lim_{n\to\infty} x_n'(\tau)d\tau = \lim_{n\to\infty} \int_0^t x_n'(\tau)d\tau = x(t) - x(0),$$

  i.e., $x(t) = x(0) + \int_0^t y(\tau)d\tau$. This shows that $x \in \mathscr{D}(T)$ and $x' = y$. Hence, T is closed.
- Note that in this example, $\mathscr{D}(T)$ is not closed in X, since T would then be bounded by the closed graph theorem.

Independence of Closedness and Boundedness

- Closedness does not imply boundedness of a linear operator. Conversely, boundedness does not imply closedness.

  The first statement is illustrated by the Differential Operator.

  The second one by the following example:

  Let $T : \mathscr{D}(T) \to \mathscr{D}(T) \subseteq X$ be the identity operator on $\mathscr{D}(T)$, where $\mathscr{D}(T)$ is a proper dense subspace of a normed space X.
  - It is trivial that T is linear and bounded.
  - However, T is not closed.
    This follows immediately from the preceding theorem, if we take an $x \in X - \mathscr{D}(T)$ and a sequence (x_n) in $\mathscr{D}(T)$ which converges to x.

The Closed Operator Lemma

Lemma (Closed Operator)

Let $T:\mathscr{D}(T)\to Y$ be a bounded linear operator with domain $\mathscr{D}(T)\subseteq X$, where X and Y are normed spaces. Then:

(a) If $\mathscr{D}(T)$ is a closed subset of X, then T is closed.

(b) If T is closed and Y is complete, then $\mathscr{D}(T)$ is a closed subset of X.

(a) If (x_n) is in $\mathscr{D}(T)$ and converges, say, $x_n\to x$, and is such that (Tx_n) also converges, then
- $x\in\overline{\mathscr{D}(T)}=\mathscr{D}(T)$ since $\mathscr{D}(T)$ is closed;
- $Tx_n\to Tx$, since T is continuous.

Hence T is closed.

(b) For $x\in\overline{\mathscr{D}(T)}$, there is a sequence (x_n) in $\mathscr{D}(T)$, such that $x_n\to x$. Since T is bounded, $\|Tx_n-Tx_m\|=\|T(x_n-x_m)\|\leq\|T\|\|x_n-x_m\|$. This shows that (Tx_n) is Cauchy. (Tx_n) converges, say, $Tx_n\to y\in Y$ because Y is complete. Since T is closed, $x\in\mathscr{D}(T)$ and $Tx=y$. Hence $\mathscr{D}(T)$ is closed.