Weak Convergence

Omar Khalil, Ruikun Luo, Anthony Della Pella, Xianan Huang

December 06, 2016

Definition 1. *A sequence $\{x_n\}$ in a normed linear vector space X is said to converge weakly to $x \in X$ if for every $x^* \in X^*$ we have $x^*(x_n) \to x^*(x)$. In this case we write $x_n \to x$ weakly.*

Definition 2. *(Strong Convergence) In a normed linear space an infinite sequence of vectors $\{x_n\}$ is said to converge to a vector x if the sequence $\{\|x - x_n\|\}$ of real numbers converges to zero. In this case, $x_n \to x$.*

Proposition 1. *If $x_n \to x$ strongly, then $x_n \to x$ weakly.*

Proof. Since X^* is the space of bounded linear functionals with $|x^*(x)| \leq \|x^*\|\|x\|$, then

$$|x^*(x_n) - x^*(x)| \leq \|x^*\|\|x_n - x\| \to 0.$$

□

Example 1. *In $X = \ell_2$ consider the element $x_n = \{0, 0, \ldots, 0, 1, 0, \ldots\}$ with 1 in the $n-$th position. Let $x^* = \{\eta_1, \eta_2, \ldots\} \in \ell_2 = X^*$. Since the ℓ_p space is finite, i.e.*

$$\sum_{n=1}^{\infty} |\eta_n|^p < \infty$$

with $p = 2$, then the norm of an element $x^* \in X^*$

$$\|x^*\| = \sqrt{\sum_{n=1}^{\infty} |\eta_n|^2} < \infty.$$

This implies that $\eta_n \to 0$ as $n \to \infty$. We have $x^*(x_n) = \langle x^*, x_n \rangle = \eta_n \to 0$ as $n \to \infty$ and $x_n \to 0$ weakly.

However, $x_n \not\to 0$ strongly.

Proof. We prove by contradiction. Let $x_n \to x$ with $x = 0$. By the definition of strong convergence, $\|x_n - 0\| \to 0$ as $n \to \infty$. This implies that $\|x_n\| \to 0$ as $n \to \infty$, which is a contradiction since $\|x_n\| = 1$. □

Definition 3. *A sequence $\{x_n^*\}$ in X^* is said to converge weak-star (or weak*) to the element x^* if for every $x \in X$, $x_n^*(x) \to x^*(x)$. In this case, we write $x_n^* \to x^*$ weak*.*

We give an example to demonstrate that weak* convergence does not imply weak convergence in X^*.

Example 2. *Let $X = c_0$ be the space of infinite sequences $x = \{\zeta_i\}_{i=1}^{\infty}$ of real numbers convergent to zero, with norm on c_0 being $\|x\| = \max_i |\zeta_i|$. Then $X^* = \ell_1$, $X^{**} = \ell_\infty$, where X^{**} is the dual of X^*. In $X^* = \ell_1$, let $x_n^* = \{0, 0, 0, \ldots, 0, 1, 0, 0, \ldots\}$ with 1 in the $n-th$ position. Then $x_n^*(x)$ picks the nth sequence ζ_n and*

$$x_n^*(x) = \zeta_n \to 0 \quad as \quad n \to \infty$$

since $\zeta_n \in c_0$, the space of infinite sequences convergent to zero. Thus, $x_n^*(x) \to 0$ as $n \to \infty$ and $x_n^* \to 0$ weak.*

However, for $x^{**} = \{1, 1, 1, \ldots\}$, $x^{**}(x_n^*) = 1 \neq 0$. Hence, $x^{**}(x_n^*) \not\to 0$ as $n \to \infty$ and $x_n^* \not\to 0$ weakly.

Definition 4. *A set $K \subset X^*$ is said to be weak* compact if every infinite sequence from K contains a weak* convergent subsequence.*

Remind 1. *A set D is said to be dense in a normed space X if for each element $x \in X$ and each $\epsilon > 0$ there exists $d \in D$ with $||x - d|| < \epsilon$.*

Remind 2. *A normed space is separable if it contains a countable dense set.*

Theorem 1. *(Alaoglu) Let X be a real normed linear vector space. The closed unit ball in X^* is weak* compact.*

Proof: This theorem is to say every finite sequence $\{x_n^*\}$ in the closed unit ball (which means that $||x_n^*|| \leq 1$) contains a weak* convergent subsequence. Thus we need to find such subsequence. Here we try to prove this theorem in the case that X is separable(although X^* need not be separable.).

Let $\{x_n^*\}$ be an infinite sequence in X^* such that $||x_n^*|| \leq 1$. Let $\{x_k\}$ be a sequence in X which is the dense set in the separable space X.

The sequence $\{x_n^*(x_1)\}$ of real numbers is bounded ($\because |x_n^*(x_1)| \leq ||x_n^*||||x_1|| \leq ||x_1||$) and thus contains a convergent subsequence which we denote as $\{x_{n1}^*(x_1)\}$(Bolzano-Weierstrass). Similarly, $\{x_{n1}^*(x_2)\}$ is also bounded and contains a convergent subsequence $\{x_{n2}^*(x_2)\}$. Continuing in this fashion to extract convergent subsequencs $\{x_{nk}^*(x_k)\}$, we then form the diagonal sequence $\{x_{nn}^*\}$ in X^* which is a subsequence of $\{x_n^*\}$. Note that as the way we build $\{x_{nn}^*\}$, we have $\{x_{nn}^*(x_k)\}$ converges. Thus the sequence $\{x_{nn}^*\}$ converges on the dense subset $\{x_k\}$ of X.

Next we need to prove that $\{x_{nn}^*\}$ converges weak* to an element $x^* \in X^*$.

Fix $x \in X$ and $\epsilon > 0$. Then for any n, m, k, we have

$$|x_{nn}^*(x) - x_{mm}^*(x)| \leq |x_{nn}^*(x) - x_{nn}^*(x_k)| + |x_{nn}^*(x_k) - x_{mm}^*(x_k)| + |x_{mm}^*(x_k) - x_{mm}^*(x)|$$

For $|x_{nn}^*(x) - x_{nn}^*(x_k)| = ||x_{nn}^*(x - x_k)| \leq ||x_{nn}^*||||x - x_k|| \leq ||x - x_k||$. Similarly, we have $|x_{mm}^*(x_k) - x_{mm}^*(x)| \leq ||x - x_k||$. As $\{x_k\}$ is in the dense subset of X. Thus choose k so that $||x_k - x|| < \frac{\epsilon}{3}$. As $\{x_{nn}^*(x_k)\}$ converges, $\{x_{nn}^*(x_k)\}$ is Cauchy. Thus there exists N such that for all $n, m > N$, we have $|x_{nn}^*(x_k) - x_{mm}^*(x_k)| < \frac{\epsilon}{3}$. Thus we have $|x_{nn}^*(x) - x_{mm}^*(x)| < \epsilon$. Then we know that $\{x_{nn}^*(x)\}$ is Cauchy and converges to a real number.

Now we define a functional f such that $f(x) = \lim_n x^*_{nn}(x)$. We need to prove that this functional f is in $X*$ which means that it is linear and bounded.

Let $x_1, x_2 \in X$ and $\alpha_1, \alpha_2 \in \mathcal{R}$. We have: $f(\alpha_1 x_1 + \alpha_2 x_2) = \lim_n x^*_{nn}(\alpha_1 x_1 + \alpha_2 x_2) = \lim_n x^*_{nn}(\alpha_1 x_1) + x^*_{nn}(\alpha_2 x_2) = \lim_n x^*_{nn}(\alpha_1 x_1) + \lim_n x^*_{nn}(\alpha_2 x_2) = f(\alpha_1 x_1) + f(\alpha_2 x_2)$. Thus f is linear.

To prove that f is bounded, we have

$$\begin{aligned}
||f|| &= \sup_{x \neq 0} \frac{|f(x)|}{||x||} \\
&= \sup_{x \neq 0} \frac{|\lim x^*_{nn}(x)|}{||x||} \\
&= \sup_{x \neq 0} \frac{\lim |x^*_{nn}(x)|}{||x||} \\
&\leq \sup_{x \neq 0} \frac{\lim ||x^*_{nn}||||x||}{||x||} \\
&\leq \sup_{x \neq 0} \frac{\lim ||x||}{||x||} \\
&= 1
\end{aligned}$$

Thus f is linear and bounded. Denote f as x^*. We have $x^*_{nn}(x)$ converges to $x^*(x)$ for all x. Thus $\{x^*_{nn}\}$ converges weak* and the closed unit ball in X^* is weak* compact.

Definition 5. *A functional (possibly nonlinear) defined on a normed space X is said to be* weakly continuous at x_0 *if given $\epsilon > 0$ there is a $\delta > 0$ and a finite collection $\{x^*_1, \ldots, x^*_n\}$ from X^* such that $|f(x) - f(x_0)| < \epsilon$ for all x such that $|x^*_i(x - x_0)| < \delta$ for $i = 1, 2, \ldots, n$. Weak* continuity of a functional defined on X^* is defined analagously with the roles of X and X^* interchanged.*

Proposition 2. *If a functional $f : X^* \to \mathbb{R}$ is weak* continuous, then $x^*_i \to x$ in the weak* sense implies that $f(x^*_i) \to f(x)$.*

Theorem 2 (Heine-Borel)**.** *For a subset K of real numbers, the following are equivalent*

- *K is compact.*
- *K is closed and bounded.*

Theorem 3. *Let f be a weak* continuous real-valued functional on a weak* compact subset S of X^*. Then f is bounded on S and achieves its maximum on S.*

Proof. The theorem is proven true if we can show that the image of S under f is compact in $\mathbb{R}$ for any weak* continuous f. Let f be as such and suppose $\{f(x^*_i)\}$ is a sequence in $f(S)$ (i.e. $\{x^*_i\} \subseteq S$). Since S is weak* compact, we can find a weak* convergent subsequence $\{x^*_{i_j}\}$ of $\{x^*_i\}$, such that $x^*_{i_j} \to x$ in weak*. But, by assumption f is weak* continous, and so by the proposition above, $f(x^*_{i_j}) \to f(x)$. But, by weak* compactness of S, the limit x is also a member of S and so $\{f(x^*_i)\}$ has a convergent subsequence $\{f(x^*_{i_j}\}$ as desired. By the Heine Borel theorem above, $f(S)$ is a closed and bounded set of real numbers, and thus attains (by $f(S)$ closed) a maximum (by $f(S)$ bounded). □

Noting again that the notion of compactness in certain spaces is very restrictive, Theorem 3 provides us with hope when trying to optimize certain functionals. For example, it is known that in infinite dimensional Banach spaces, the unit ball is never compact. This could be troubling for instance in the following example.

Example 3. *Consider $L^2[a,b] = \{g \mid g : [a,b] \to \mathbb{R}$ and $\int_a^b |g|^2 < \infty\}$. Then, we can set up the following maximization problem:*

$$\begin{aligned} \underset{x^* \in L^2[a,b]^*}{\text{minimize}} \quad & \langle x, x^* \rangle \\ \text{subject to} \quad & ||x^*||_{L^2} \leq 1. \end{aligned} \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $L^2[a,b]$. Recall that L^2 is a Hilbert space, and so $L^2[a,b]^* = L^2[a,b]$. Moreover, note that $\langle x, x^* \rangle$ is a weak continuous functional on X^* (as strong continuity implies weak* continuity). By Alaoglu's theorem the unit ball $S = \{x^* \in L^{2^*} \mid ||x^*||_{L^2} \leq 1$ is compact, and so by Theorem 3 we can in fact find a solution to problem (1) above.*

Definition 6 (Notations in Measure Theory)**.** *Focus analysis on X :Normed Space*

σ-algebra Σ: set of subsets of X such that

$\phi \in \Sigma, X \in \Sigma; \forall A \in \Sigma \Rightarrow A^C \in \Sigma$

$A_i \in \Sigma, i = 1, \ldots, n \Rightarrow \bigcap A_i \in \Sigma, \bigcup A_i \in \Sigma$

Borel σ-algebra B_X: smallest σ-algebra containing all open subsets of X

Borel Measure μ on X: set function, $B_X \to [0, \infty)$

Measure is regular: $\forall \varepsilon > 0, \forall A \in \Sigma, \exists$open set O and closed set P such that $P \subset A \subset O, \mu(O - P) < \varepsilon$

Alternatively, $\forall A \in \Sigma, \begin{cases} \mu(A) = \inf\{\mu(O) | A \subset O, O \text{ open}\} \\ \mu(A) = \sup\{\mu(P) | A \supset P, P \text{ compact}\} \end{cases}$

Measure is finitely additive: $E_1, \ldots, E_n \in \Sigma$ are disjoint subsets of X, then $\mu\left(\bigcup E_i\right) = \sum \mu(E_i)$

Measure is countably additive: countable sequence $E_1, \ldots, E_n, \ldots \in \Sigma$ are pairwise disjoint subset of X, then one have $\mu\left(\bigcup E_i\right) = \sum \mu(E_i)$

Measure bounded: bounded with respect to total variance norm

Total variation norm:

Upper variation: $\bar{W}(\mu, E) = \sup(\mu(A) | A \in \Sigma, A \subset E)$

Lower variation: $\underline{W}(\mu, E) = \inf(\mu(A) | A \in \Sigma, A \subset E)$

Absolute variation: $|\mu|(E) = \bar{W}(\mu, E) + |\underline{W}(\mu, E)|$

Total variation norm: $\|\mu\| = |\mu|(X)$

Definition 7 (Convergence in Distribution)**.** *Let $\{\mu_n\}_{n \in N}$ be the sequence of probability measures on $(X, B_X), \mu_n$ converges in distribution (converge weakly) to a probability μ on (X, B_X) if $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu, \forall f \in C_b(X)$*

The definition of convergence in distribution is equivalent to weak* convergence of probability measure. Borel probability measure is bounded countabily additive measure. Typically, countably additive can imply finitely additive, but not vice versa. However, if defined in metric space and Borel σ-algebra, regulatity and finite additivity can imple countably additive. In this way, in order to show the equivalence, we only need to show that the dual space of $C_b(X)$ is the space of regular bounded finitely additive measure $rba(X)$.

Theorem 4. *The dual space of bounded functions $B(X)$ on normed space X with Borel σ-algebra is the space of bounded finitely additive measure $ba(X)$*

Theorem 5. *Measure μ is finitely additive if for disjoint sets A and B*

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

Theorem 6. *The dual space of continuous bounded functions $C_b(X)$ on normed space X with Borel σ-algebra is the space of regular bounded finitely additive measure $rba(X)$ [1]*

$$x^*(f) = \int_X f(x)d\mu(x)$$

Proof. First prove $\forall \mu \in rba(X)$ defines a linear functional $x^*(f) = \int_X f(x)d\mu(x)$

$\because f(x)$ is bounded continuous function defined on X

$\therefore$ Range set of $f(x)$ is a bounded set, can be covered with a finite set of open sets $G_1, \ldots, G_n$ with diameter less than $\varepsilon > 0$

Define disjoint sets A_i from G_i: $A_1 = G_1, A_j = G_j - \bigcup_{i=1}^{j-1} G_i$

For each A_i, select a point $\alpha_i \in A_i$, if $A_i = \phi, \alpha_i = 0$

Since f is continuous, $B_j = f^{-1}(A_j) \subset X$, in the domain of μ, in this way, a ε-approximation of f can be defined with the indicator function $\mathcal{X}_{B_j}$ of $B_j$by

$$f_\varepsilon = \sum \alpha_j \mathcal{X}_{B_j}$$

Since the measure is regular and finitely additive, the integral is defined as

$$\int f_\varepsilon d\mu = \sum \alpha_j \mu(B_j)$$

$\because |\alpha_j| \leqslant \sup_{x \in X} |f(x)| = \|f\|_\infty$

$\therefore \left|\int f_\varepsilon d\mu\right| = |\sum \alpha_j \mu(B_j)| \leqslant \|f\|_\infty |\mu|$

$\therefore \forall \mu \in rba(X)$ defines a linear functional $x^*(f) = \int_X f(x)d\mu(x)$

Next prove $\forall x^* \in C_b(X)^*$, $x^*(f)$ can be represented with $x^*(f) = \int_X f(x)d\mu(x)$ with $\mu \in rba(X)$

$\because C_b(X)$ is a subspace of space of bounded function $B(X)$

$\therefore$Apply Hahn-Banach theorem extension form Corollary 1 and Theorem 4

$\exists \lambda \in ba(X)$ such that $x^*(f) = \int_X f(x)d\lambda(x)$

The goal is to find $\mu \in rba(X)$ such that $\int_X f(x)d\mu(x) = \int_X f(x)d\lambda(x) \quad \forall f \in C_d(X)$

Denote F as general closed subset, G as general open subset, E as general subset of X. Define set function μ_1 and μ_2 as

$$\mu_1(F) = \inf_{G \supset F} \lambda(G), \mu_2(E) = \sup_{F \subset E} \mu_1(F)$$

By definition, μ_1 and μ_2 are nonnegative and nondecreasing. Let G_1 be open set and F_1 be closed set, if $G \supset (F_1 - G_1)$, then $(G \cup G_1) \supset F_1$, $\lambda(G \cup G_1) \leqslant \lambda(G) + \lambda(G_1)$. Since G is arbitary open set containing $F_1 - G_1$, we have

$$\mu_1(F_1) \leqslant \lambda(G_1) + \mu_1(F_1 - G_1)$$

By allowing G_1 range over all open sets containing $F \cap F_1$,

$$\mu_1(F_1) \leqslant \mu_1(F \cap F_1) + \mu_2(F_1 - F)$$

Let F_1 range over all closed subsets of E which is arbitary subset of X

$$\mu_2(E) \leqslant \mu_2(F \cap E) + \mu_2(E - F) = \mu_2(E \cap F) + \mu_2(E \cap F^c)$$

Let F_1 and F_2 be disjoint closed sets, there exists disjoint open neighborhoods G_1 and G_2. G is an arbitary neighborhood of $F_1 \cup F_2$, then $\lambda(G) \geqslant \lambda(G \cap G_1) + \lambda(G \cap G_2)$, thus

$$\mu_1(F_1 \cup F_2) \geqslant \mu_1(F_1) + \mu_1(F_2)$$

Let F and E be arbitary subsets of X with F closed. Let F_1 range over all closed subsets of $E \cap F$, F_2 range over all closed subsets of $E \cap F^c$

$$\mu_2(E) \geqslant \mu_2(E \cap F) + \mu_2(E \cap F^c)$$

Restrict μ to μ_2, since $\mu(E) = \mu(E \cap F) + \mu(E \cap F^c)$, by Theorem 5, μ is finitely additive. By definition, $\mu_2(F) = \mu_1(F) = \mu(F)$, thus regular. Let F_3 be largest closed subset of X, $\mu(X) = \mu_1(F_3) \leqslant \lambda(X)$, thus μ is bounded, $\mu \in rba(X)$. Since μ and λ are bounded, they can be scaled such that $\mu(X) = \lambda(X)$. Remaining to show $\int_X f(x)d\mu(x) = \int_X f(x)d\lambda(x)$, $\forall f \in C_b(X)$

Since f is continuous and bounded. WLOG let $0 \leqslant f(x) \leqslant 1$

Since f is continuous, $\forall \varepsilon > 0$, $\exists$disjoint paritions $E_1, \ldots, E_n$ of X, $a_i = \inf_{x \in E_i}(f(x))$

$$\sum_{i=1}^{n} a_i \mu(E_i) + \varepsilon \geqslant \int_X f(x)d\mu(x)$$

Since μ is regular, i.e.$\forall \varepsilon > 0, \forall A \in \Sigma, \exists$open set O and closed set P such that $P \subset A \subset O, \mu(O - P) < \varepsilon$ Let $F_i, i = 1, \ldots, n$ be closed subsets of E_i such that

$$\sum_{i=1}^{n} a_i \mu(F_i) + 2\varepsilon \geqslant \int_X f(x)d\mu(x)$$

Since f is continuous, there are open sets $G_i, i = 1, \ldots, n; G_i \supset F_i$ such that

$$b_i = \inf_{x \in G_i} f(x) \geqslant a_i - \frac{\varepsilon}{n\|\mu\|}, \text{hence} \quad \sum_{i=1}^{n} b_i \mu(G_i) + 3\varepsilon \geqslant \int_X f(x)d\mu(x)$$

For a open set G containing F, $\mu(F) \leqslant \lambda(G)$. Since μ is regular, $\mu(G) \leqslant \lambda(G)$, therefore

$$\int_X f(x)d\mu(x) \leqslant \sum_{i=1}^{n} b_i \mu(G_i) \leqslant \sum_{i=1}^{n} b_i \lambda(G_i) \leqslant \int_X f(x)d\lambda(x)$$

Since $f \in C_d(X), 0 \leqslant f \leqslant 1 \Rightarrow 1 - f \in C_d(X)$, therefore

$$\int_X (1-f)(x)d\mu(x) \leqslant \int_X (1-f)(x)d\lambda(x)$$

Since it's scaled such that $\mu(X) = \lambda(X)$

$$\int_X (-f)(x)d\mu(x) \leqslant \int_X (-f)(x)d\lambda(x) \Rightarrow \int_X f(x)d\mu(x) \geqslant \int_X f(x)d\lambda(x)$$

Thus $\forall f \in C_d(X)$, $\exists \mu \in rba(X)$ such that $\int_X f(x)d\mu(x) = \int_X f(x)d\lambda(x)$ $\square$

References

[1] N. Dunford and J. T. Schwartz, "Linear operators, vol. i," *Interscience, New York*, vol. 1963, 1958.

[2] D. G. Luenberger, *Optimization by vector space methods.* John Wiley & Sons, 1969.