April 9, 2022

Chapter 2

P. 16 See [Halmos Section 11, Thms. 1 and 2, P. 17, Section 6, P. 9] and [Note 1]. In particular, the sum of arbitrary elements that belongs to the empty set is θ. As a consequence, the subspace spanned by $\emptyset$ is $\{\theta\}$.

P. 17

Lemma (P. 17)**.** *Let $\mathcal{V}$ be a collection of linear varieties in a vector space such that $\cap_{V\in\mathcal{V}}V \neq \emptyset$. Then the set $\cap_{V\in\mathcal{V}}V$ is a linear variety.*

Proof. Since $\cap_{V\in\mathcal{V}}V \neq \emptyset$, then let x be some vector in $\cap_{V\in\mathcal{V}}V$. Define a collection $\mathcal{M}$ as follows

$$\mathcal{M} = \{M : \exists V \in \mathcal{V} \text{ such that } V = x + M\}.$$

It is clear that $\mathcal{M}$ is a collection of subspaces so that $\cap_{M\in\mathcal{M}}M$ is a subspace (a proof of this follows from the arguments similar to the proof of [OVSM Prop. 1, P. 15]). We will show that $\cap_{V\in\mathcal{V}}V = x + \cap_{M\in\mathcal{M}}M$ by contrapositions.

Let $y \notin \cap_{V\in\mathcal{V}}V$, then there exists $V_0 \in \mathcal{V}$ such that $y \notin V_0$. Therefore, $\exists M_0 \in \mathcal{M}$ such that $V_0 = x + M_0$. In addition, $y - x \notin M_0$. Since $\cap_{M\in\mathcal{M}}M \subset M_0$, thus $y - x \notin \cap_{M\in\mathcal{M}}M$. As a result, $y \notin x + \cap_{M\in\mathcal{M}}M$.

Conversely, let $y \notin x + \cap_{M\in\mathcal{M}}M$. Then there exists $M_0 \in \mathcal{M}$ such that $y - x \notin M_0$. Since $M_0 \in \mathcal{M}$, then there exists $V_0 \in \mathcal{V}$ such that $V_0 = x + M_0$. As a result, $y \notin V_0$. Therefore, $y \notin \cap_{V\in\mathcal{V}}V$. □

P. 19

Lemma (P. 19)**.** *A set S is a linearly independent set if and only if all of its subsets with finitely many elements are linearly independent sets.*

Proof. The only if part follows the definition. For the if part, we prove its contraposition. Suppose that S is not a linearly independent set. Then there exists some $s \in S$ that is linear combination of vectors of S, which means $\exists s_1, s_2 \ldots, s_m \in S$ and scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$ with $m < \infty$ such that $s = \sum_{i=1}^{m} \alpha_i s_i$. Then the set $M = \{s, s_1, s_2 \ldots, s_m\}$ is finite subset and is not a linearly independent set. □

P. 24

Lemma (P. 24). *Let P be a subset of a normed linear space. Then $\mathring{P}$ is open.*

Proof. By definition, we have $\mathring{\mathring{P}} \subset \mathring{P}$. Let $x \in \mathring{P}$, then $\exists \varepsilon > 0$ such that $S(x, \varepsilon) \subset P$. We will show that $S(x, \varepsilon) \subset \mathring{P}$, which would implies that $x \in \mathring{\mathring{P}}$. For every $y \in S(x, \varepsilon)$, set $\delta = \varepsilon - \|x - y\|$. It is clear that $\delta > 0$ since $y \in S(x, \varepsilon)$, implying that $\|x - y\| < \varepsilon$. Then for all $z \in S(y, \delta)$, $\|z - x\| \leq \|z - y\| + \|y - x\| < \varepsilon$. Therefore, $S(y, \delta) \subset S(x, \varepsilon) \subset P$, namely $y \in \mathring{P}$. Since y is arbitrary, then every point in $S(x, \varepsilon)$ belongs to $\mathring{P}$, namely $S(x, \varepsilon) \subset \mathring{P}$. This means that x is an interior point of $\mathring{P}$, i.e., $x \in \mathring{\mathring{P}}$. Since x is arbitrary, then $\mathring{P} \subset \mathring{\mathring{P}}$. $\square$

P. 25

Lemma (1, P. 25). *Let P be a subset of a normed linear space. Then $\overline{\overline{P}} = \overline{P}$.*

Proof. By definition, we have that $\overline{P} \subset \overline{\overline{P}}$. To show that $\overline{\overline{P}} \subset \overline{P}$, let $x \in \overline{\overline{P}}$. Then $\forall \varepsilon > 0$, $\exists \bar{p} \in \overline{P}$ such that $\|x - \bar{p}\| < \varepsilon$. Since $\bar{p} \in \overline{P}$, for $\delta = \varepsilon - \|x - \bar{p}\|$, $\exists p \in P$ such that $\|p - \bar{p}\| < \delta$. Thus, we have that $\|x - p\| = \|x - \bar{p} + \bar{p} - p\| \leq \|x - \bar{p}\| + \|\bar{p} - p\| < \|x - \bar{p}\| + \varepsilon - \|x - \bar{p}\| = \varepsilon$. Thus, $p \in \overline{P}$. $\square$

We restate and prove [OVSM Prop. 1, P. 25] as the following two lemmas for clarity.

Lemma (2, P. 25). *Let P be a open subset of a normed linear space. Then $\tilde{P} = \{x : x \notin P\}$, its complement, is closed.*

Proof. We have that $x \in \overline{\tilde{P}}$ if and only if $\forall \varepsilon > 0$, $\exists p \in \tilde{P}$ such that $\|x - p\| < \varepsilon$, due to the definition of closure. This holds if and only if $\forall \varepsilon > 0$, $\exists p \notin P$ such that $\|x - p\| < \varepsilon$. This implies that $x \notin P$ as P is open. Therefore, $x \in \tilde{P}$, which means that $\overline{\tilde{P}} \subset \tilde{P}$. The converse relation is clear. $\square$

Lemma (3, P. 25). *Let C be a closed subset of a normed linear space. Then $\tilde{C} = \{x : x \notin C\}$, its complement, is open.*

Proof. We have that $x \in \tilde{C}$ if and only if $x \notin C$, which holds if and only if that $\exists \varepsilon > 0$ such that $S(x, \varepsilon) \cap C = \emptyset$ as C is closed. This holds if and only if $\exists \varepsilon > 0$ such that $S(x, \varepsilon) \subset \tilde{C}$. Thus, $\tilde{C}$ is open. $\square$

We restate and prove [OVSM Prop. 3, P. 25] as the following two lemmas for clarity.

Lemma (4, P. 25). *Let $\{P_i\}_{i=1}^m$ be a finite collection of closed subsets of a normed linear space. Then their union is closed.*

Proof. We prove that if P_1, P_2 are closed, then $\overline{P_1 \cup P_2} = P_1 \cup P_2$. Any finite collection can be proven by proving two at a step.

First, we have that $P_1 \cup P_2 \subset \overline{P_1 \cup P_2}$. To see it holds conversely, let $x \notin P_1 \cup P_2$, then $x \notin P_1$ and $x \notin P_2$. Since P_1, P_2 are closed, $\exists \varepsilon_1 > 0$, $\varepsilon_2 > 0$ such that $S(x, \varepsilon_1) \cap P_1 = \emptyset$ and $S(x, \varepsilon_2) \cap P_2 = \emptyset$. Define $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$, then we have $S(x, \varepsilon) \cap P_1 = S(x, \varepsilon) \cap P_2 = \emptyset$. Therefore, $S(x, \varepsilon) \cap (P_1 \cup P_2) = \emptyset$. Thus, $x \notin \overline{P_1 \cup P_2}$. □

Note that in the proof, the step $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ would fail if the collection of closed sets are inifinitely many. Also refer to [Note 2].

Lemma (5, P. 25)**.** *Let $\mathcal{P}$ be an arbitrary collection of closed subsets of a normed linear space. Then their intersection is closed.*

Proof. We denote the intersection of all elements in $\mathcal{P}$, i.e., $\cap_{P \in \mathcal{P}} P$, as Q. Clearly we have that $Q \subset \overline{Q}$. To show it holds conversely, for $x \notin Q$, then $\exists P' \in \mathcal{P}$ such that $x \notin P'$. Since P' is closed, then $\exists \varepsilon > 0$ such that $S(x, \varepsilon) \cap P' = \emptyset$. As $Q \subset P'$, we have that $S(x, \varepsilon) \cap Q = \emptyset$. Thus, $x \notin \overline{Q}$. □

P. 28 Note that a functional is a transformation from a vector space X to the space of real or complex scalars. It is possible that a functional maps a real vector space X to the space of complex scalars. See [Kreyszig P. 103].

P. 28 Let $T : X \mapsto Y$ be a linear operator. It is part of the definition of linear operator that both X and Y are vector spaces of the same field, namely both real or both complex. This is required in order to make the linearity of T well defined, [Kreyszig, Definition 2.6-1, P. 82]. As a result, for a linear functional f defined on vector space X with field K, its image is the scalars of K, [Kreyszig, Definition 2.8-1, P. 104].

P. 28 We prove the 'if' part via establishing the contraposition. Suppose T is not continuous at x_0. Then $\exists \varepsilon > 0$ such that $\forall \delta > 0$, $\exists x \in S(x_0, \delta)$ such that $\|T(x) - T(x_0)\| \geq 0$. Chose $\delta_n = 1/n$ and we can construct a sequence $\{x_n\}$ so that $x_n \to x_0$ yet $T(x_n) \not\to T(x_0)$.

P. 29 The Hölder inequality has the following direct implication.

Lemma (P. 29)**.** *Let $x = \{\xi_1, \xi_2, \dots\} \in l_p$ and $y = \{\eta_1, \eta_2, \dots\} \in l_q$, where $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $1/p + 1/q = 1$. Define scalar sequence $\{a_n\}$ with $a_n = \sum_{i=1}^{n} \xi_i \eta_i$. Then the sequence $\{a_n\}$ is convergent with a scalar limit.*

Proof. By Hölder inequality, we have $\sum_{i=1}^{\infty} |\xi_i \eta_i| < \infty$. Define $b_n = \sum_{i=1}^{n} |\xi_i \eta_i|$, then the sequence $\{b_n\}$ is real, monotonically increasing, and converges to a real number. As a result, it is also Cauchy. Therefore, for any $\varepsilon > 0$, there exists N such that $|b_n - b_m| < \varepsilon$ for all $m, n > N$. For all m, n, we have $|a_n - a_m| \leq |b_n - b_m|$. To show this inequality hold, we can first assume that

$m \leq n$, and prove it; then assume that $m > n$, and show it hold as well. To see it holds for $m \leq n$, we have

$$|a_n - a_m| = \left| \sum_{i=m}^{n} \xi_i \eta_i \right| \leq |b_n - b_m|.$$

Therefore, the sequence is $\{a_n\}$ is Cauchy, thus convergent. □

Note that a Cauchy sequence being convergent is true for both real and complex scalar sequences.

P. 30

Lemma (P. 30)**.** *Let p and q be positive real numbers, and $x = \{\xi_1, \xi_2, \dots\} \in l_p$ and $y = \{\eta_1, \eta_2, \dots\} \in l_q$. Denote their truncated sequences as x_n and y_n respectively, namely*

$$x_n = \{\xi_1, \xi_2, \dots, \xi_{n-1}, \xi_n, 0, 0, \dots\}.$$

Then $|\xi_i|^p \|y\|_q^q = |\eta_i|^q \|x\|_p^p$ for $i = 1, 2, \dots$ if and only if $|\xi_i|^p \|y_n\|_q^q = |\eta_i|^q \|x_n\|_p^p$ for $n = 1, 2, \dots$ and $i = 1, \dots, n$.

Proof. For the only if part, we assume that x and y are nonzero, as otherwise it holds trivially. We note that by adding $|\xi_i|^p \|y\|_q^q = |\eta_i|^q \|x\|_p^p$ from 1 to n, we have

$$\|y\|_q^q \sum_{i=1}^{n} |\xi_i|^p = \|x\|_p^p \sum_{i=1}^{n} |\eta_i|^q,$$

which is $\|y\|_q^q \|x_n\|_p^p = \|x\|_p^p \|y_n\|_q^q$. After multiplying $|\xi_i|^p \|y\|_q^q$ with $\|x\|_p^p \|y_n\|_q^q / (\|x\|_p^p \|y\|_q^q)$, and multiplying $|\eta_i|^q \|x\|_p^p$ with $\|x_n\|_p^p \|y\|_q^q / (\|x\|_p^p \|y\|_q^q)$, we have the desired results.

For the if part, if is clear that $\|x_n\|_p^p \to \|x\|_p^p$, and $\|y_n\|_q^q \to \|y\|_q^q$ as $n \to \infty$. Thus, for every fixed i, since $|\xi_i|^p \|y_n\|_q^q = |\eta_i|^q \|x_n\|_p^p$, taking limit on both sides get the desired results. □

P. 31

Lemma (P. 31)**.** *Let $\{a_n\}$ and $\{b_n\}$ be two sequences such that $b_n \geq a_n$, $\sum_{n=1}^{\infty} a_n = a < \infty$, $\sum_{n=1}^{\infty} b_n = b < \infty$. Then $a = b$ if and only if $a_n = b_n$ for all n.*

Proof. The if part is obvious. For the only if part, assume $\exists \ell$ such that $a_\ell \neq b_\ell$. Since $a_\ell \leq b_\ell$, then $a_\ell < b_\ell$. Then we have

$$b - a = b_\ell - a_\ell + \sum_{n \neq \ell} (b_n - a_n) > 0.$$

□

Regarding the equality condition for the Minkowski inequality, we prove the case where both x and y are nonzero. For the if part, since $y = \alpha x$ where $\alpha > 0$, then for all n, we have

$$\sum_{i=1}^{n} |\xi_i + \eta_i|^p = \sum_{i=1}^{n} |\xi_i + \eta_i|^{p-1}|\xi_i| + \sum_{i=1}^{n} |\xi_i + \eta_i|^{p-1}|\eta_i|. \tag{1}$$

In addition, denote $\zeta_i = (\xi_i + \eta_i)^{p-1}$ and define $w = \{\zeta_1, \zeta_2, \dots\}$. It is clear from the proof that $w \in l_q$. The truncated sequences of x and w are denoted as x_n and w_n respectively. Then for every fixed n and $i = 1, \dots, n$, we have

$$\begin{aligned}\|w_n\|_q^q|\xi_i|^p =& \Big(\sum_{j=1}^{n} |\xi_j + \eta_j|^{(p-1)q}\Big)|\xi_i|^p = \Big(\sum_{j=1}^{n} |\xi_j + \eta_j|^p\Big)|\xi_i|^p = \Big(\sum_{j=1}^{n} |\xi_j|^p\Big)(1+\alpha)^p|\xi_i|^p \\ =& \|x_n\|_p^p(1+\alpha)^p|\xi_i|^p = \|x_n\|_p^p|\xi_i + \eta_i|^p = \|x_n\|_p^p|\xi_i + \eta_i|^{(p-1)q} \\ =& \|x_n\|_p^p|\zeta_i|^q.\end{aligned}$$

The relation between w_n and y_n follows in view of $y = \alpha x$. Therefore, for every n, applying the Hölder inequality and in view of the equality condition holds, we have

$$\sum_{i=1}^{n} |\zeta_i||\xi_i| = \|w_n\|_q\|x_n\|_p, \quad \sum_{i=1}^{n} |\zeta_i||\eta_i| = \|w_n\|_q\|y_n\|_p.$$

Combining with Eq. (1), we have for all n the equality

$$\sum_{i=1}^{n} |\xi_i + \eta_i|^p = \|w_n\|_q\|x_n\|_p + \|w_n\|_q\|y_n\|_p.$$

Knowing that both sides converges, taking limits as $n \to \infty$, we have

$$\|x + y\|_p^p = \|w\|_q(\|x\|_p + \|y\|_p). \tag{2}$$

Note that

$$\|w\|_q = \Big(\sum_{j=1}^{\infty} |\xi_j + \eta_j|^{(p-1)q}\Big)^{1/q} = \Big(\sum_{j=1}^{\infty} |\xi_j + \eta_j|^p\Big)^{1/q} = \|x + y\|_p^{p/q}, \tag{3}$$

then $\|x+y\|_p^p/\|w\|_q = \|x+y\|_p^p/\|x+y\|_p^{p/q} = \|x+y\|_p$ since $(pq-p)/q = 1$.

For the only if part, again we assume that both x and y are nonzero. Since $\|x + y\|_p = \|x\|_p + \|y\|_p$, then by Eq. (3), we have Eq. (2) hold. Since we also have

$$\|x+y\|_p^p = \sum_{i=1}^{\infty} |\xi_i+\eta_i|^p \leq \sum_{i=1}^{\infty} |\xi_i+\eta_i|^{p-1}|\xi_i| + \sum_{i=1}^{\infty} |\xi_i+\eta_i|^{p-1}|\eta_i| \leq \|w\|_q(\|x\|_p+\|y\|_p),$$

where the first inequality is established via first triangular inequality and then taking limit, and the second due to Hölder inequality. Therefore, we have the two inequalities are also equalities. Now we first focus on the first equality. Applying [Lemma P. 31] with $a_i = |\xi_i + \eta_i|^p$ and $b_i = |\xi_i + \eta_i|^{p-1}(|\xi_i| + |\eta_i|)$, we have

$$|\xi_i + \eta_i| = |\xi_i| + |\eta_i|, \quad i = 1, 2, \ldots. \tag{4}$$

For the second equality, since

$$\sum_{i=1}^{\infty} |\xi_i + \eta_i|^{p-1}|\xi_i| \leq \|w\|_q \|x\|_p, \quad \sum_{i=1}^{\infty} |\xi_i + \eta_i|^{p-1}|\eta_i| \leq \|w\|_q \|y\|_p,$$

therefore, we have

$$\sum_{i=1}^{\infty} |\xi_i + \eta_i|^{p-1}|\xi_i| = \|w\|_q \|x\|_p, \quad \sum_{i=1}^{\infty} |\xi_i + \eta_i|^{p-1}|\eta_i| = \|w\|_q \|y\|_p.$$

In view of the equality condition for Hölder inequality, we have

$$|\xi_i|^p \|w\|_q = |\zeta_i|^q \|x\|_p, \quad |\eta_i|^p \|w\|_q = |\zeta_i|^q \|y\|_p, \quad i = 1, 2, \ldots.$$

In view of x and y being nonzero, $\|x + y\|_p = \|x\|_p + \|y\|_p \neq 0$ and Eq. (3), we have

$$|\eta_i|^p = \frac{\|y\|_p}{\|x\|_p} |\xi_i|^p, \quad i = 1, 2, \ldots.$$

Equivalently, we have for all i, $|\eta_i| = \alpha|\xi_i|$, where $\alpha = \left(\frac{\|y\|_p}{\|x\|_p}\right)^{1/p}$. Combining with Eq. (4), we have $y = \alpha x$.

P. 38

Lemma (1, P. 38)**.** *Let X and Y be normed linear spaces. Consider the produce space $X \times Y$ with a norm $\|(x, y)\| = \max\{\|x\|, \|y\|\}$. Then if X or Y is incomplete, then so is the product space.*

Proof. Suppose that X is incomplete. Then there exists $\{x_n\} \subset X$ that is Cauchy but not convergent. Therefore, the sequence $\{(x_n, \theta)\}$ belongs to $X \times Y$. It is Cauchy, but not convergent. Thus, $X \times Y$ is incomplete. □

P. 38

Lemma (2, P. 38)**.** *In a normed linear space, if a subset is complete, then it is closed.*

Proof. Since a convergent sequence is always Cauchy with the limit being unique, and the subset is complete, then the subset is closed. □

See also the discussion in [Note 5]. Note that in the context of metric space, the term "subspace" has different meaning as it is used in [OVSM]. The term "subspace" in a metric space context means a subset equipped with the induced metric of the overall metric space (cf. [Kreyszig P. 4]), as opposed to the defintion of "subspace" given in [OVSM Definition P. 14] in a vector space setting. Therefore, the term "subspace" in [Note 5] should be interpreted as a "subset" if using the terminology of [OVSM].

P. 40 Let f be a real-valued functional defined on a normed space X. Its limit superior at x_0 is denoted by $\limsup_{x\to x_0} f(x)$ and defined as (according to [Ex. 49 in P. 50 and Ex. 50 in P.51, Royden])

$$\limsup_{x\to x_0} f(x) = \inf_{\delta>0} \Big\{ \sup \big\{ f(x) : x \in S(x_0,\delta) \setminus \{x_0\} \big\} \Big\}, \tag{5}$$

where $S(x_0,\delta) = \{x : \|x - x_0\| < \delta\}$ ([P. 24, OVSM]). Also, compare this with the incorrect definition of limit superior in [P. xvi, OVSM].

Lemma (P. 40)**.** *Let f be a real-valued functional defined on a normed space X. Given a point $x_0 \in X$, the following three statements are equivalent:*

(a) $\forall \varepsilon > 0$, $\exists \delta > 0$, such that $f(x) - f(x_0) < \varepsilon$ holds $\forall x \in S(x_0,\delta)$.

(b) The inequality $\limsup_{x\to x_0} f(x) \le f(x_0)$ holds.

(c) For all sequences $\{x_n\}$ such that $x_n \to x_0$, $\limsup_{n\to\infty} f(x_n) \le f(x_0)$.

Proof. We will first show that (a) and (b) are equivalent, and then show that (a) and (c) are equivalent.

To show that (b) implies (a), we introduce a function $g : [0,\infty) \mapsto \mathbb{R} \cup \{\infty\}$ defined as

$$g(\delta) = \begin{cases} \sup \big\{ f(x) : x \in S(x_0,\delta) \setminus \{x_0\} \big\}, & \text{if } \delta > 0, \\ \limsup_{x\to x_0} f(x), & \text{if } \delta = 0. \end{cases}$$

If $g(0) = -\infty$, then for some large n so that $f(x_0) > -n$, $\exists \delta_n > 0$ such that $g(\delta_n) < -n$. In view of the definition of g, we have $f(x) - f(x_0) < \varepsilon$ for all $x \in S(x_0,\delta_n)$ and $\varepsilon > 0$. Note that $f(x_0) \in \mathbb{R}$ in view of the definition of real-valued functional. If $g(0) > -\infty$, then $g(0) \in \mathbb{R}$. In view of the definition (5), we have $g(\delta) \ge g(0)$ $\forall \delta \ge 0$. In addition, for all $\varepsilon > 0$, there exists some $\delta > 0$, such that $g(\delta) < g(0) + \varepsilon$. Since $g(0) \le f(x_0)$, we then have $g(\delta) - f(x_0) < \epsilon$. In view of the definition of g, we obtain the statement (a).

To show that (a) implies (b), we prove the contraproposition. Suppose that the inequality $\limsup_{x\to x_0} f(x) > f(x_0)$ holds, namely $g(0) > f(x_0)$. If $g(0) = \infty$, then according to (5), we have $g(\delta) = \infty$ for all δ. This means for any δ, the functional $f(x)$ is unbounded above in $S(x_0,\delta) \setminus \{x_0\}$, thus the statement (a) is false. Otherwise, we have $g(0) \in \mathbb{R}$. Then $\exists \bar{\delta} > 0$ such that $g(\bar{\delta}) \in \mathbb{R}$. Since g is decreasing as δ decreases, we have $g(\delta) \in \mathbb{R}$ for all $\delta \in [0,\bar{\delta}]$ (since $g(\delta) \ge g(0) > f(x_0) > -\infty$ for all δ). Denote $\varepsilon = \frac{g(0)-f(x_0)}{2}$ and $\varepsilon > 0$ according to assumption. For all $\delta \in (0,\bar{\delta}]$, we have

$$\infty > g(\delta) = \sup \big\{ f(x) : x \in S(x_0,\delta) \setminus \{x_0\} \big\} \ge g(0) > f(x_0).$$

Consider the given ε introduced above. Due to the definition of supremum, for all $\delta \in (0, \bar{\delta}]$, there exists some $x \in S(x_0, \delta) \setminus \{x_0\}$ such that $f(x) > g(\delta) - \varepsilon \geq g(0) - \varepsilon = f(x_0) + \varepsilon$. For $\delta > \bar{\delta}$, we have $S(x_0, \bar{\delta}) \setminus \{x_0\} \subset S(x_0, \delta) \setminus \{x_0\}$, so for the given ε we can find $x \in S(x_0, \delta) \setminus \{x_0\}$ for all $\delta > 0$ such that $f(x) > f(x_0) + \varepsilon$.

To show that (a) implies (c), we prove its contraproposition. Suppose there exists a sequence $\{x_n\}$ such that $x_n \to x_0$, and $\limsup_{n\to\infty} f(x_n) > f(x_0)$. Denote $\bar{f} = \limsup_{n\to\infty} f(x_n)$. If $\bar{f} < \infty$, then denote $\varepsilon = \frac{\bar{f} - f(x_0)}{2}$. In addition, we have

$$\sup_{n \geq \ell} f(x_n) \geq \bar{f}, \quad \forall \ell. \tag{6}$$

Since $x_n \to x_0$, thus for all $\delta > 0$, $\exists \bar{\ell}$ such that $\{x_n\}_{n \geq \bar{\ell}} \subset S(x_0, \delta)$. By (6), we have $\sup_{n \geq \bar{\ell}} f(x_n) \geq \bar{f}$. By the definition of supremum, there exists $\bar{n} \geq \bar{\ell}$ such that $f(x_{\bar{n}}) > \bar{f} - \varepsilon = f(x_0) + \varepsilon$. The case where $\bar{f} = \infty$

To show that (c) implies (a), we also prove its contraproposition. Assume $\exists \varepsilon > 0$ such that $\forall \delta > 0$, $\exists x_\delta \in S(x_0, \delta)$ so that $f(x_\delta) - f(x_0) \geq \varepsilon$. Therefore, there exists a sequence $\{x_n\}$ so that $x_n \in S(x_0, 1/n)$ and $f(x_n) - f(x_0) > \varepsilon$ for all x_n. Since $\|x_n - x_0\| \in [0, 1/n)$, then

$$0 \geq \liminf_{n\to\infty} \|x_n - x_0\| \leq \limsup_{n\to\infty} \|x_n - x_0\| \leq \lim_{n\to\infty} \frac{1}{n} = 0 \implies \lim_{n\to\infty} x_n = x_0.$$

Since $f(x_n) - f(x_0) > \varepsilon$ for all x_n, then $\sup_{n \geq \ell} f(x_n) \geq f(x_0) + \varepsilon$ for all ℓ. As a result, we have $\limsup_{n\to\infty} f(x_n) \geq f(x_0) + \varepsilon > f(x_0)$. □

Note that when establishing (a) implies (b), we show that the statement (a) is false by identifying some x so that $f(x) > f(x_0) + \varepsilon$. However, comparing this with the inequality

$$f(x) - f(x_0) < \varepsilon \tag{7}$$

in the statement (a), the strict negation of (7) would be $f(x) \geq f(x_0) + \varepsilon$. The difference here is due to that the following two statements are equivalent

(a) $\forall \varepsilon > 0$, $\exists \delta > 0$, such that $f(x) - f(x_0) < \varepsilon$ holds $\forall x \in S(x_0, \delta)$.

(a') $\forall \varepsilon > 0$, $\exists \delta > 0$, such that $f(x) - f(x_0) \leq \varepsilon$ holds $\forall x \in S(x_0, \delta)$.

Also refer to [Note 3] for the equality issue in the ϵ-δ definition.

P. 40 In the Example 1, we have

$$\begin{aligned} |f(x)| &= \left| \int_0^{\frac{1}{2}} x(t)dt - \int_{\frac{1}{2}}^1 x(t)dt \right| \leq \left| \int_0^{\frac{1}{2}} x(t)dt \right| + \left| \int_{\frac{1}{2}}^1 x(t)dt \right| \\ &\leq \int_0^{\frac{1}{2}} |x(t)|dt + \int_{\frac{1}{2}}^1 |x(t)|dt \leq \|x\|. \end{aligned}$$

Suppose there exists $x^* \in C[0,1]$ such that $|f(x^*)| = \|x^*\|$, then the last inequality is in fact an equality. Suppose $\|x^*\| = 1$, then we have

$$\int_0^{\frac{1}{2}} |x^*(t)|dt + \int_{\frac{1}{2}}^1 |x^*(t)|dt = 1. \tag{8}$$

We will show that this implies two consequences: (1) $\forall t \in [0,1]$, $|x^*(t)| = 1$; (2) $\forall t_1, t_2 \in [0,1]$, $x(t_1) = x(t_2)$. To show the first one, suppose otherwise. Then $\exists \hat{t}$ such that $|x^*(\hat{t})| < 1$. Then by continuity, there exists some $\delta > 0$ such that for all $t \in [0,1] \cap (\hat{t} - \delta, \hat{t} + \delta)$, $|x^*(t)| < 1$, which contradicts (8). To show that (2) is true, assume otherwise. Then there exists some $t_1, t_2 \in [0,1]$ such that $x^*(t_1) \neq x^*(t_2)$. Without loss of generality, suppose that $t_1 < t_2$ and $x^*(t_1) = -1$. Since $x^* \in C[0,1]$, then by intermediate value theorem [Prop. 19, P. 48, Royden], there exists some $\bar{t} \in (t_1, t_2)$ such that $x^*(t) = 0$. Then for some $\varepsilon \in (0,1)$, $\exists \delta > 0$ such that $|x^*(t)| < 1$ $\forall t \in (\bar{t} - \delta, \bar{t} + \delta)$. Again, this contradicts (8). So the above two properties hold. However, this means that

$$|f(x^*)| = \left| \int_0^{\frac{1}{2}} x^*(t)dt - \int_{\frac{1}{2}}^1 x^*(t)dt \right| = 0,$$

which is a contradiction.

P. 42 Note that $\forall m \in M$, $-m \in M$, and $\exists y \in [x]$ such that $y = x + m$. Conversely, $\forall y \in [x]$, $\exists m \in M$ such that $y = x + m$. Therefore,

$$\|[x]\| = \inf_{m \in M} \|x + m\| = \inf_{m \in M} \|x - m\| = \inf_{y \in [x]} \|y\|.$$

Suppose that $\|[x]\| = 0$ for some $[x] \neq \theta$. Then For every $\varepsilon > 0$, $\exists m \in M$, such that $\|x - m\| < \varepsilon$. As a result, x is a closure point of M. In view of M being closed, we have $[x] = \theta$, which is a contradiction.

Chapter 3

P. 48 Note that there is a subtle point. We need to show that for $x = \{\xi_1, \xi_2 \ldots\}$ and $y = \{\eta_1, \eta_2 \ldots\}$ that belong to the (real) l_2 space, the value $\sum_{i=1}^{\infty} \xi_i \eta_i$ is well defined. That is, if we denote $a_n = \sum_{i=1}^{n} \xi_i \eta_i$, then we need to show that the sequence $\{a_n\}$ is convergent. This is established by [Lemma P. 29]. Therefore, $a_n \to (x \,|\, y)$. Since $|a_n| \leq |b_n| = b_n$, taking limit on both sides as both $\{a_n\}$ and $\{b_n\}$ converge, by the continuity of $|\cdot|$, we have $|(x \,|\, y)| \leq \sum_{i=1}^{\infty} |\xi_i \eta_i| \leq \|x\| \|y\|$.

P. 50 Note that the logically correct statement shall be that if there exists $m_0 \in M$ such that m_0 is the minimizing vector, then for all $m_1 \in M$ such that m_1 is the minimizing vector, m_1 is unique. This statement is logically equivalent to the following one: for all $m_0 \in M$ such that m_0 is the minimizing vector, m_0 is unique.

Symbolically, the first statement above has the structure

$$\exists x P_1(x) \implies \forall y\big(P_1(y) \implies P_2(y)\big),$$

which is logically equivalent to

$$\forall x\big(P_1(x) \implies P_2(x)\big),$$

cf. [Note 4]. On the other hand, the statement

$$\exists x P_0(x) \implies \forall y\big(P_1(y) \implies P_2(y)\big),$$

can be implied, but not equivalent to

$$\forall x\big(P_1(x) \implies P_2(x)\big).$$

This is because that it is possible that the statement $P_0(x)$ is false for all x, thus the condition is never fulfilled.

P. 50

Lemma (P. 50)**.** *Let M be a subspace in a pre-Hilbert space. Then for every vector x in its orthogonal complement $M^\perp$, there exists a unique vector $m_0 \in M$ such that $\|x - m_0\| \le \|x - m\|$. Moreover, $m_0 \equiv \theta$ for all $x \in M^\perp$.*

Proof. The proof is obtained by noting that $x \perp M$ implies $(x - \theta) \perp M$, and by applying [OVSM Thm. 1, P. 50]. □

P. 52 By examing the proof arguments, we see that [OVSM Prop. 1 (1), (2), (3) and (4), P. 52] hold in a pre-Hilbert space setting. This is stated in [Berberian Thm. 1, P. 59, Thm. 2, P. 60].

P. 52

Lemma (1, P. 52)**.** *Let N be a subspace in a normed space. Then $\overline{N}$ is a subspace.*

Proof. For $x, y \in \overline{N}$, there exists sequences $\{x_n\}, \{y_n\} \subset N$ such that $x_n \to x$ and $y_n \to y$. Since N is a subspace, then for any scalars α, β, we have $\alpha x_n + \beta y_n \in N$ and $\alpha x_n + \beta y_n \to \alpha x + \beta y$, which means that $\overline{N}$ is a subspace. □

Lemma (2, P. 52)**.** *Let S be a subset in a pre-Hilbert space. Then $S^\perp = ([S])^\perp$.*

Proof. Since $S \subset [S]$, we have $([S])^\perp \subset S^\perp$ by [OVSM, Prop. 1 (3), P. 52]. Conversely, if $x \in S^\perp$, then $x \perp y$ for all $y \in S$. Since for all $z \in [S]$, $z = \sum_{i=1}^m \alpha_i y_i$ for some scalars α_i and $y_i \in S$ and $m < \infty$, then we have $x \perp z$. Therefore, $S^\perp \subset ([S])^\perp$. □

Lemma (3, P. 52)**.** *Let S be a subset in a pre-Hilbert space. Then $S^\perp = (\overline{S})^\perp$.*

Proof. Since $S \subset \overline{S}$, we have $(\overline{S})^{\perp} \subset S^{\perp}$ by [OVSM, Prop. 1 (3), P. 52]. Conversely, if $x \in S^{\perp}$, then $x \perp y$ for all $y \in S$. Since for all $z \in \overline{S}$, there exists $\{z_n\} \subset S$ such that $z_n \to z$, and $(x \,|\, z_n) = 0$ for all n. By continuity of inner product, we have $x \perp z$, which means $S^{\perp} \subset (\overline{S})^{\perp}$. □

Lemma (4, P. 52)**.** *Let S be a subset of a Hilbert space. Then $S^{\perp\perp} = \overline{[S]}$.*

Proof. Denote $M = \overline{[S]}$. Then by [Lemma 1, P. 52], M is a closed subspace. By [OVSM, Thm. 1, P. 53], we have $M = M^{\perp\perp}$. We will show that $S^{\perp} = M^{\perp}$. By applying [Lemma 2, P. 52], we have $S^{\perp} = ([S])^{\perp}$. Then applying [Lemma 3, P. 52] with $N = [S]$, we have the desired result. □

In the proof of [Lemma 4, P. 52], since we rely on [OVSM, Thm. 1, P. 53], the proof of which relies on the projection theorem of Hilbert space for existence, so [Lemma 4, P. 52] holds true only for Hilbert space, but not for pre-Hilbert space.

Lemma (5, P. 52)**.** *Let S be a subset in a pre-Hilbert space. Then $S \cap S^{\perp} \subset \{\theta\}$.*

Proof. For all $s \in S \cap S^{\perp}$, it holds that $(s \,|\, s) = 0$. The only element that fulfills the property is θ. Therefore, $S \cap S^{\perp} = \{\theta\}$ or $S \cap S^{\perp} = \emptyset$. □

P. 53 Here in this note we use the notation $\oplus$ slightly different from the definition given in [OVSM P. 53]. Let M and N be two subspaces in a vector space X. Then we write $M + N = M \oplus N$ if every vector $x \in M + N$ has unique representation $x = m + n$ where $m \in M$ and $n \in N$. Therefore, the difference here is that we allow the case where $M + N \neq X$.

Lemma (1, P. 53)**.** *Let M and N be two subspaces in a vector space. Then $M + N = M \oplus N$ if and only if $M \cap N = \{\theta\}$.*

Proof. If $x \in M \cap N$ and $x \neq \theta$. Then $2x \in M \cap N$ since M and N are subspaces. As a result, $2x = 2x + \theta = x + x$, where $2x, x \in M$, and $\theta, x \in N$.

Conversely, if $x \in M + N$ such that $x = m_1 + n_1 = m_2 + n_2$ with $m_1, m_2 \in M$ and $n_1, n_2 \in N$ such that $m_1 \neq m_2$ and $n_1 \neq n_2$. As a result, we have $m_1 - m_2 = n_2 - n_1$. However, $m_1 - m_2 \in M$, and $n_2 - n_1 \in N$. Thus, $m_1 - m_2 \in M \cap N$ and $m_1 - m_2 \neq \theta$. □

P. 53

Lemma (2, P. 53)**.** *Let M be a finite-dimensional subspace of a pre-Hilbert space X. Then $X = M \oplus M^{\perp}$, and $M = M^{\perp\perp}$.*

Proof. Let $x \in X$, and consider the space generated by $\{x\} \cup M$. This is a finite-dimensional subspace and is thus complete. We denote this space as H. Then by projection theorem, there exits a unique $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$

for all $m \in M$, and $n_0 = x - m_0 \in M^\perp$. Therefore, $x = m_0 + n_0$ with $m_0 \in M$ and $n_0 \in M^\perp$.

To show the representation is unique, suppose that $x = m_1 + n_1$ with $m_1 \in M$ and $n_1 \in M^\perp$. Then we have $\theta = m_0 - m_1 + n_0 - n_1$ with $m_0 - m_1 \in M$ and $n_0 - n_1 \in M^\perp$. By Pythagorean theorem, $\|\theta\|^2 = \|m_0 - m_1\|^2 + \|n_0 - n_1\|^2$. This implies that $m_0 = m_1$ and $n_0 = n_1$.

To show that $M = M^{\perp\perp}$, we only need to show that $M^{\perp\perp} \subset M$ in view of [OVSM Prop. 1.2, P. 52]. Let $x \in M^{\perp\perp}$, then by the first part of the lemma, we have $x = m+n$ with $m \in M$ and $n \in M^\perp$. Since $M \subset M^{\perp\perp}$, then $m \in M^{\perp\perp}$. As a result, we have $n = x - m \in M^{\perp\perp}$ since $M^{\perp\perp}$ is a subspace. However, $n \in M^\perp$. Thus, $n \perp n$, which means $n = \theta$. Therefore, $x = m \in M$. □

P. 55 Note that the results of this section hold true in a pre-Hilbert setting. To see this, let $y_1, y_2, \ldots, y_n$, as well as x, be elements of a pre-Hilbert space. Denote as $\hat{H}$ the space generated by the set $\{y_1, y_2, \ldots, y_n, x\}$. Then, by [OVSM, Thm. 2, P. 38], we have that the space $\hat{H}$ is a Hilbert space. In addition, by [OVSM, Thm. 2, P. 38] as well as [Lemma 2, P. 38], we have the subspace M generated by vectos $y_1, y_2, \ldots, y_n$ being closed. Then by applying the classical projection theorem on $\hat{H}$, we have the existence result.

P. 56 As is commented in [P. 55], [OVSM Prop. 1, P. 56] holds true in a pre-Hilbert setting.

P. 59 Note that in [OVSM Thm. 1, P. 59], the if part requires the space being Hilbert, while the only if part holds in a pre-Hilbert setting. In addition, the Bessel's inequality [OVSM Lemma 1, P. 59] holds true in a pre-Hilbert setting, cf. [Berberian Thm. 3, P. 45, Corollary, P. 46].

P. 60

Lemma (1, P. 60)**.** *Let S_1 and S_2 be two subspaces of a linear space, and S be their union, i.e., $S = S_1 \cup S_2$. Then $[S] = S_1 + S_2$.*

Proof. First, for every $s \in S_1 + S_2$, there exists $s_1 \in S_1$ and $s_2 \in S_2$ such that $s = s_1 + s_2$. As a result, s is linear combination of vectors in S, thus $s \in [S]$. Conversely, for $s \in [S]$, we have $s = \sum_{i=1}^n \alpha_i s_1^i + \sum_{i=1}^m \beta_i s_2^i$, where n, m are nonnegative integers, α_i, β_i are scalars, and $s_1^i \in S_1$ and $s_2^i \in S_2$. Since S_1 and S_2 are subspaces, then $\sum_{i=1}^n \alpha_i s_1^i \in S_1$ and $\sum_{i=1}^m \beta_i s_2^i \in S_2$. Therefore, $s \in S_1 + S_2$. □

Lemma (2, P. 60)**.** *Let M be a closed subspace in a normed linear space, and a be a nonzero vector of the normed linear space. Denote as S the union of $\{a\}$ and M, i.e., $S = \{a\} \cup M$. Then $[S] = \overline{[S]}$.*

Proof. Denote $A = [\{a\}]$, and $C = A \cup M$. We will first show that $[S] = [C]$. Note that $S \subset C$, so $[S] \subset [C]$. Conversely, for $x \in [C]$, it is a linear combination of elements in A and C. Since $A = [\{a\}]$, so x is linear combination of elements in S. Thus $[S] = [C]$. Then by [Lemma 1, P. 60], we have $[S] = A + M$.

If $a \in M$, then $S = [S] = M$ and M is closed by assumption, and the proof is done. Otherwise, denote

$$\delta = \inf_{m \in M} \|a + m\|.$$

Since $a \notin M$ and M is closed, we have $\delta > 0$. In addition, we also have

$$\inf_{m \in M} \|\alpha a + m\| \geq |t|\delta \tag{9}$$

hold for all scalars t. To see this, note that when $\alpha = 0$, it trivially holds. Otherwise, $\|\alpha a + m\| = |\alpha|\|a + m/\alpha\|$, and the inequality follows.

Let $s \in \overline{[S]}$, then there exists $\{s_n\} \subset [S]$ such that $s_n \to s$. Since $[S] = A + M$, we can write $s_n = \alpha_n a + m_n$, where α_n are scalars, and $\{m_n\} \subset M$. Since the sequence $\{s_n\}$ is convergent, it is Cauchy. In addition,

$$\|s_n - s_\ell\| = \|\alpha_n a + m_n - \alpha_\ell a - m_\ell\| = \|(\alpha_n - \alpha_\ell)a + (m_n - m_\ell)\| \geq |\alpha_n - \alpha_\ell|\delta,$$

where the last inequality is due to Eq. (9). As a result, the sequence of scalars $\{\alpha_n\}$ is also Cauchy, and we denote its limit as α. Denote $m = s - \alpha a$. Clearly $\alpha a \in A$, and s is a vector of the linear space. As a result, m is a vector of the linear space. We will show that $m_n \to m$, and then by closedness of M, it follows that $m \in M$ and $s = \alpha a + m \in A + M$. Indeed, we have

$$\begin{aligned}
\|m_n - m\| =& \|m_n - (s - \alpha a)\| = \|m_n - s + \alpha a + \alpha_n a - \alpha_n a\| \\
=& \|(m_n + \alpha_n a) - s + \alpha a - \alpha_n a\| = \|s_n - s + \alpha a - \alpha_n a\| \\
\leq& \|s_n - s\| + |\alpha_n - \alpha|\|a\|.
\end{aligned}$$

Therefore, $m_n \to m$, and by closedness of M, we have $m \in M$, and $s \in A + M = [S]$. □

With above results, one may be tempted to relax the completeness property of H in [OVSM Thm. 2, P. 60], by using the arguments discussed in [P. 55], namely considering the subspace generated by $\{x\} \cup M$ as the overall space. As is shown in [Lemma 2, P. 60], this subspace is closed with respect to H. However, since it is infinite dimensional, if the underlying space H is not complete, then the subspace generated by $\{x\} \cup M$ may not be complete either. Note that the proof of [OVSM Thm. 2, P. 60] relies on the if part of [OVSM Thm. 1, P. 59], which in turn relies on the completeness. Therefore, it is necessary to require the underlying space to be Hilbert.

On the other hand, it is possible for a incomplete normed linear space to have an infinite dimensional subspace that is complete. To see this, we consider the construction used in [Lemma 1, P. 38], where we have two spaces X and Y. Let X be [OVSM Example 2, P. 34], which is incomplete. Let $Y = l_1$, which is complete. In addition, denote $e_n \in Y$ as the sequence with nth element being 1, and 0 otherwise. So $\{e_n\}$ is orthonormal. Then the closed subspace generated by $\{e_n\}$ is Y. To see this, assume $y = \{\eta_1, \eta_2, \dots\} \in Y$ such that $y \perp e_n$ for all n. Then $\eta_n = 0$ for all n. Thus $y = \theta$. Then by [OVSM Lemma 1, P. 61], we reach above conclusion. Now consider the closed subspace generated by $\{(\theta, e_n)\}$. This space is $\{\theta\} \times Y$, which is complete.

P. 60 Note that the conclusion here is true in a pre-Hilbert setting. The reason is that for every vector x in M, the target to which the constructed Cauchy sequence converges is known (namely X), so the completeness condition is not needed. We state and prove the result below. The proof approach follows the proof of [Deitmar Thm. 2.3.3, P. 33].

Lemma (3, P. 60)**.** *Let $\{e_i\}$ be an orthonormal sequence in a pre-Hilbert space X, and denote as M the closed subspace generated by $\{e_i\}$, i.e., $M = \overline{[\{e_i\}]}$. Then for every $x \in M$, it can be expressed as the limit of an infinite series of the form $x = \sum_{i=1}^{\infty} \alpha_i e_i$. Moreover, the parameters are uniquely given as $\alpha_i = (x \,|\, e_i)$.*

Proof. Let x be some vector in the pre-Hilbert space. Denote as s_n the partial sum of infinite series, i.e., $s_n = \sum_{i=1}^{n} \alpha_i e_i$, where $\alpha_i = (x \,|\, e_i)$. Then by Bessel's inequality (which holds in a pre-Hilbert space setting, cf. [P. 59]), we have $\sum_{i=1}^{\infty} |\alpha_i|^2 \leq \|x\|^2$, namely, for every x, the sequence $\{\alpha_i\}$ is in l_2. Thus, we introduce a mapping $T : X \mapsto l_2$ as $T(x) = \{\alpha_i\}$. It can be seen that the mapping is linear. In view of Bessel's inequality, we have $\|T(x)\| \leq \|x\|$. (Note that the norm on $T(x)$ is l_2 norm, while the norm on x is the one on the pre-Hilbert space X.) Due to this inequality, we have that $x_n \to x$ implies $T(x_n) \to T(x)$. To see this, we note first that

$$\begin{aligned}
&\|T(x_n) - T(x)\| = \|T(x_n - x)\| \leq \|x_n - x\| \implies \\
&\limsup \|T(x_n) - T(x)\| = \limsup \|T(x_n - x)\| \leq \limsup \|x_n - x\| = \lim \|x_n - x\| = 0.
\end{aligned}$$

In addition, we have that $\liminf \|T(x_n) - T(x)\| \geq 0$. Therefore, $T(x_n) \to T(x)$. On the other hand, $\|T(s)\| = \|s\|$ for all $s \in [\{e_i\}]$. By continuity of norm, for all $x \in M$, $\|T(x)\| = \|x\|$. Again, for every $x \in M$, we note that $\limsup \|x - s_n\|^2 = \|x\|^2 - \sum_{i=1}^{\infty} |\alpha_i|^2 = \|x\|^2 - \|T(x)\|^2 = 0$. Therefore, $s_n \to x$.

To see the representation is unique, for every $\{\beta_i\} \in l_2$ such that $\hat{s}_n \to x$, where $\hat{s}_n = \sum_{i=1}^{n} \beta_i e_i$, we see that for every i, $(x - \hat{s}_n \,|\, e_i)$ becomes constant once $n \geq i$. Taking limit $\lim_{n\to\infty}(x - \hat{s}_n \,|\, e_i) = 0$, we have that $(x - \beta_i e_i \,|\, e_i) = 0$, which implies $\beta_i = (x \,|\, e_i)$. □

Next lemma is a special case of the above result. The proof is identical, and is thus neglected. Note that in [OVSM Definition, P. 60], complete orthonormal sequence is defined for Hilbert space. The following lemma stretch the definition a bit, and uses the same meaning for a sequence in a pre-Hilbert space.

Lemma (4, P. 60)**.** *Let $\{e_i\}$ be a complete orthonormal sequence in a pre-Hilbert space X. Then for every $x \in X$, it can be expressed as the limit of an infinite* *series of the form $x = \sum_{i=1}^{\infty} \alpha_i e_i$. Moreover, the parameters are uniquely given as $\alpha_i = (x \,|\, e_i)$.*

P. 61

Lemma (1, P. 61)**.** *Let $\{e_i\}$ be a complete orthonormal sequence in a pre-Hilbert space X. Then the only vector orthogonal to each of e_i's is the null vector.*

Proof. Let $x \in X$ be orthogonal to each of e_i's, i.e., $x \perp e_i$ for all i. By [Lemma 4, P. 60], we have that $x = \sum_{i=1}^{\infty} \alpha_i e_i$, where $\alpha_i = (x \,|\, e_i)$. Namely $s_n \to x$, where $s_n = \sum_{i=1}^{n} \alpha_i e_i$. Since $x \perp e_i$, then $s_n \equiv \theta$. Therefore, $x = \lim s_n = \theta$. □

Lemma (2, P. 61)**.** *Let $\{e_i\}$ be a orthonormal sequence in a Hilbert space H. The sequence $\{e_i\}$ is complete if and only if the only vector orthogonal to each of e_i's is the null vector.*

Proof. The only if part is given by [Lemma 1, P. 61]. We prove the if part here via proving its contraposition. Denote as M the closed subspace generated by $\{e_i\}$, i.e., $M = \overline{[\{e_i\}]}$. Assume $\tilde{M} = H \setminus M \neq \emptyset$. Then for all $x \in \tilde{M}$, we have $\delta = \inf_{y \in M} \|x - y\| > 0$, as otherwise x would be a closure point of M and thus belongs to M. Since M is closed and H is Hilbert, there exists $\hat{x} \in M$ such that $(x - \hat{x}) \perp M$, and $x - \hat{x} \neq \theta$. As a result, $(x - \hat{x}) \perp e_i$ for all i. □

Note that in view $x \notin M$, we have $\delta = \inf_{y \in M} \|x - y\| > 0$ (in fact, we also have $\inf_{y \in M} \|x - y\| = \inf_{y \in [\{e_i\}]} \|x - y\|$). This means that there exists $\{y_n\} \subset M$ such that $\lim_{n \to \infty} \|y_n - x\| = \delta$. As a result, we have $\{y_n\}$ is Cauchy. At this point, we cannot proceed if the space is pre-Hilbert. On the other hand, if H is Hilbert, then we know $\{y_n\}$ is convergent, and its limit $\hat{x}$ exists in M. Then by [OVSM Thm. 1, P. 50], $(x - \hat{x}) \perp e_i$ for all i.

P. 64 The following result is the linear variety counter part of [OVSM Thm. 1, P. 50]. The proof is obtained via applying [OVSM Thm. 1, P. 50] and is thus neglected.

Lemma (P. 64)**.** *Let M be a subspace in a pre-Hilbert space X. Let x be a fixed element in X and V be a linear variety $x + M$. For some $x_0 \in V$, it has minimum norm among all elements in V if and only if x_0 is orthogonal to M. In addition, the vector $x_0 \in V$ that has minimum norm is unique if it exists.*

P. 65

Lemma (1, P. 65)**.** *Let X be a pre-Hilbert space and $\{y_1, y_2, \dots, y_n\}$ a set of linearly independent vectors in X. Let M be the (closed) subspace generated by y_i's. Then for every set of scalars $\{c_1, c_2, \dots, c_n\}$, there exists a vector $x \in M$ such that*

$$(x \,|\, y_i) = c_i, \quad i = 1, 2, \dots, n.$$

Proof. By [OVSM Prop. 1, P. 56], the Gram G of y_i's is invertable, then $x = \sum_{i=1}^{n} \alpha_i y_i$ fulfills the condition, where $\alpha = [\alpha_1 \;\; \alpha_2 \;\; \dots \;\; \alpha_n]'$, and $\alpha = (G')^{-1}[c_1 \; c_2 \; \cdots \; c_n]'$. □

Lemma (2, P. 65)**.** *Let X be a pre-Hilbert space, $\{y_1, y_2, \dots, y_n\}$ be a set of linearly independent vectors in X and M be the (closed) subspace generated by y_i's. Let $\{c_1, c_2, \dots, c_n\}$ be a set of given scalars and define set C as*

$$C = \{x \,:\, x \in X,\, (x \,|\, y_i) = c_i,\, i = 1, 2, \dots, n\}.$$

Then there exists $x_0 \in X$ such that

$$\|x_0\| = \inf_{y \in C} \|y\|.$$

Moreover, the minimizing vector is unique and belongs to the subspace M.

Proof. By [Lemma 1, P. 65], we know that $C \neq \emptyset$. Let $y \in C$ be some vector. Then by proving $C \subset (y + M^{\perp})$ and $(y + M^{\perp}) \subset C$, we have $C = y + M^{\perp}$. In addition, by [OVSM Prop. 1.2, P. 52], as well as [P. 52] in this note, we have $M \subset M^{\perp\perp}$.

By [OVSM Prop. 1, P. 56], the Gram G of y_i's is invertable. Consider the vector x defined as $x = \sum_{i=1}^{n} \alpha_i y_i$, where $\alpha = [\alpha_1 \;\; \alpha_2 \;\; \dots \;\; \alpha_n]'$, and $\alpha = (G')^{-1}[c_1 \; c_2 \; \cdots \; c_n]'$. Then one can see that $x \in C$. Moreover, $x \in M \subset M^{\perp\perp}$. By [Lemma, P. 64], x has minimum norm among all vectors in V, and $x \in M$. □

P. 67 The [OVSM Fig. 3.5, P. 67] illustrates why the result [Lemma 2, P. 65] holds true in a pre-Hilbert setting. The reason is that since the codimension n is finite, and we can construct within the finite dimensional subspace M a projection x_0 of x, thus, the minimum distance from x to $M^{\perp}$ is automatically resolved, with the minimum distance given as $\|x_0\|$.

P. 69 Given two subsets B and C, we define their difference $B - C$ as $B + (-C)$.

Lemma (1, P. 69)**.** *Let A, B and C be some subsets in a vector space such that $A = B + C$. Then $B \subset (A - C)$.*

Proof. When B or C is empty set, the relation holds true. So we consider the case where $B \neq \emptyset$ and $C \neq \emptyset$. Thus $A \neq \emptyset$. Let $b \in B$, then for some $c \in C$, we have $b + c \in A$. Thus $b \in (A - C)$. □

Note that the converse may not be true. Consider $B = [0, 1]$, and $C = (0, \infty)$. Then we have $A = C = (0, \infty)$. In addition, we have $A - C = \mathbb{R} \neq B$.

For the sum of two sets B and C, if C is a singleton $\{x\}$, we write $B + C$ as $x + B$.

Lemma (2, P. 69)**.** *Let x be a fixed vector, and A, B be some subsets in a vector space such that $A = x + B$. Then $B = A - x$.*

Proof. We have shown that $C \subset A - x$ by [Lemma 1, P. 69]. Let $b \in (A - x)$, then there exists $a \in A$ such that $b = a - x$. Due to the definition of A, given $a \in A$, there exists $\hat{b} \in B$ such that $\hat{b} + x = a$. Therefore, we have $\hat{b} = b$, thus $b \in B$. □

The next result can be verified through the definition of compactness.

Lemma (3, P. 69)**.** *Let K be a compact set in a normed linear space. Then for every scalar α, the set αK is compact.*

The next result can be verified by definition of convergence.

Lemma (4, P. 69)**.** *Let $\{\alpha_i\}$ be a scalar sequence. Then $\alpha_i \to 0$ if and only if $|\alpha_i| \to 0$.*

Lemma (5, P. 69)**.** *Let A, B and K be some subsets in a normed linear space such that K is compact and $A = B + K$. If B is closed, then A is closed.*

Proof. Let $a \in \overline{A}$. Then there exists $\{a_i\} \subset A$ such that $a_i = b_i + k_i$, $b_i \in B$, $k_i \in K$, and $a_i \to a$. Since K is compact, there exists a subsequence $\{k_{i_n}\} \subset \{k_i\}$ such that $k_{i_n} \to k \in K$. We will show that the subsequence $\{b_{i_n}\}$, with the same indices as $\{k_{i_n}\}$, is convergent. To see this, we note that $\{a_i\}$ is convergent. Therefore, the subsequence $\{a_{i_n}\}$ is convergent to a. Denote $b = a - k$. Since we have that

$$\|a_{i_n} - a\| = \|b_{i_n} + k_{i_n} - (b + k)\| \geq |\|b_{i_n} - b\| - \|k_{i_n} - k\||.$$

Taking limit superior and limit inferior on both sides, we have $|\|b_{i_n} - b\| - \|k_{i_n} - k\|| \to 0$. By [Lemma 4, P. 69], we have $\lim \|b_{i_n} - b\| = 0$. As a result, b is a closure point of B. Since B is closed, then we have $b \in B$. As a result, $b + k = a \in A$. □

Note that the converse may not be true. Consider $B = \mathbb{Q}$, the rational numbers, and $K = [0, 1]$. Then we have $A = \mathbb{R}$. In this case, A is closed, but B is not.

Lemma (6, P. 69)**.** *Let x be a fixed vector, and A, B be some subsets in a normed linear space such that $A = x + B$. Then A is closed if and only if B is closed.*

Proof. The if part is proven by noting that $\{x\}$ is compact, and then applying [Lemma 5, P. 69]. To see the only if part, we first note that $B = A - x$ by [Lemma 2, P. 69]. Since $\{-x\}$ is compact, then by [Lemma 5, P. 69], we see that set B, as the sum of closed set A and a compact set $\{-x\}$, is closed. □

P. 69 We carry out the detailed computation here. Since $y(t)-x_n(t) = \int_0^t [u(\tau)-u_n(\tau)]d\tau$, we have

$$|y(t) - x_n(t)|^2 = \left| \int_0^t 1 \cdot [u(\tau) - u_n(\tau)]d\tau \right|^2 = |(e\,|\,u - u_n)_{[0,t]}|^2,$$

where $e(\tau) \equiv 1$ for $\tau \in [0, t]$ and $(\cdot\,|\,\cdot)_{[0,t]}$ is the inner product for the space $L_2[0, t]$. (Note that in the expression $(e\,|\,u - u_n)_{[0,t]}$, the notation u in fact represents the restriction $u\big|_{[0,t]}$, so that it is an element of $L_2[0, t]$. The same applies to u_n. It should be clear from context whether u and u_n represents vectors in $L_2[0, t]$ or $L_2[0, T]$.) Then by Cauchy-Schwarz inequality for the space $L_2[0, t]$, we have

$$|(e\,|\,u - u_n)_{[0,t]}|^2 \leq \|e\|^2_{[0,t]}\|u - u_n\|^2_{[0,t]} = t \int_0^t |u(\tau) - u_n(\tau)|^2 d\tau,$$

where $\|\cdot\|_{[0,t]}$ is the 2-norm for $L_2[0, t]$. Since the right hand side of the equality above increases with t, therefore, by combining the relations together, we get

$$|y(t) - x_n(t)|^2 \leq T \int_0^T |u(\tau) - u_n(\tau)|^2 d\tau = T\|u - u_n\|^2, \quad \forall t \in [0, T],$$

where $\|\cdot\|$ represents the norm for $L_2[0, T]$. Then we have

$$\int_0^T |y(t) - x_n(t)|^2 dt \leq \int_0^T T\|u - u_n\|^2 dt = T^2\|u - u_n\|^2,$$

namely $\|y - x_n\| \leq T\|u - u_n\|$.

P. 69 Let X and U be two pre-Hilbert spaces with inner products $(\cdot\,|\,\cdot)_X$ and $(\cdot\,|\,\cdot)_U$, and norms $\|\cdot\|_X$ and $\|\cdot\|_U$ respectively. Then the default inner product $(\cdot\,|\,\cdot)$ for $X \times U$ is given as

$$\big((x_1, u_1)\big|\,(x_2, u_2)\big) = (x_1\,|\,x_2)_X + (u_1\,|\,u_2)_U,$$

which can be verified fulfilling the inner-product axioms. The norm $\|\cdot\|$ for $X \times U$ is similarly defined through

$$\|(x, u)\| = \sqrt{\|x\|^2_X + \|u\|^2_U}.$$

Therefore, if the sequence $\{(x_n, z_n)\} \subset X \times U$ converges to (x, u), then one can verify that the sequences $\{x_n\} \subset X$ and $\{u_n\} \subset U$ are both convergent with limits x and u respectively.

P. 69 The latter part of [OVSM Thm. 1, P. 69] is stated for the real Hilbert space case. The following lemma extends the theorem, which is applicable to complex vector space.

Lemma (7, P. 69). *Let x be a vector in a pre-Hilbert space X and let K be a closed convex subset contained in a finite-dimensional subspace M. Then there is a unique vector $k_0 \in K$ such that*

$$\|x - k_0\| \leq \|x - k\|$$

for all $k \in K$. Furthermore, a necessary and sufficient condition that k_0 be the unique vector is that $(x - k_0 \,|\, k - k_0) + (k - k_0 \,|\, x - k_0) \leq 0$ for all $k \in K$.

Proof. We consider the subspace generated by $\{x\} \cup M$. This subspace is finite dimensional thus complete. We denote the space as H, which is a Hilbert space itself. The existence and uniqueness part of the proof follows exactly the same approach as is used to prove [OVSM Thm. 1, P. 69] by focusing on the complete subspace H.

For the second part of the proof, denote as x_0 the minimizing vector. Assume to the contrary that there exists some k_1 such that $(x - k_0 \,|\, k_1 - k_0) + (k_1 - k_0 \,|\, x - k_0) = 2\varepsilon > 0$. Consider the vector $k_\alpha = k_0 + (1-\alpha)k_1$; $0 \leq \alpha \leq 1$. Since K is convex, then $k_\alpha \in K$ for all α. Also,

$$\begin{aligned}
\|x - k_\alpha\|^2 =& \|(1-\alpha)(x - k_0) + \alpha(x - k_1)\|^2 \\
=& \big((1-\alpha)(x - k_0) + \alpha(x - k_1) \,\big|\, (1-\alpha)(x - k_0) + \alpha(x - k_1)\big) \\
=& (1-\alpha)^2\|x - k_0\|^2 + \alpha(1-\alpha)(x - k_1 \,|\, x - k_0) + \\
& \alpha(1-\alpha)(x - k_0 \,|\, x - k_1) + \alpha^2\|x - k_1\|^2.
\end{aligned}$$

The quantity $\|x - k_\alpha\|^2$ is a differentiable function of $\alpha$with derivative at $\alpha = 0$ equal to

$$\begin{aligned}
\frac{d}{d\alpha}\|x - k_\alpha\|^2\big|_{\alpha=0} =& -2\|x - k_0\|^2 + (x - k_1 \,|\, x - k_0) + (x - k_0 \,|\, x - k_1) \\
=& -2(x - k_0 \,|\, x - k_0) + (x - k_1 \,|\, x - k_0) + (x - k_0 \,|\, x - k_1) \\
=& -(k_1 - k_0 \,|\, x - k_0) - (x - k_0 \,|\, k_0 - k_1) = -2\varepsilon < 0.
\end{aligned}$$

Thus for some small positive α, $\|x - k_\alpha\| < \|x - k_0\|$, which contradicts the minimizing property of k_0.

Conversely, suppose that $k_0 \in K$ such that $(x - k_0 \,|\, k - k_0) + (k - k_0 \,|\, x - k_0) \leq 0$ for all $k \in K$. Then for every $k \in K$ and $k \neq k_0$. we have

$$\begin{aligned}
\|x - k\|^2 =& |x - k_0 + k_0 - k\|^2 \\
=& \big((x - k_0) + (k_0 - k) \,\big|\, (x - k_0) + (k_0 - k)\big) \\
=& \|x - k_0\|^2 + (k_0 - k \,|\, x - k_0) + (x - k_0 \,|\, k_0 - k) + \|k_0 - k\|^2 \\
=& \|x - k_0\|^2 + \|k_0 - k\|^2 - (k - k_0 \,|\, x - k_0) - (x - k_0 \,|\, k - k_0) > \|x - k_0\|^2.
\end{aligned}$$

□

Chapter 4

P. 81 Suppose $\{y_1, y_2, \ldots, y_m\}$ is a collection of n-dimensional random vectors. Consider the Hilbert space $\mathcal{H}$ of n-dimensional random vectors consisting of all vectors whose components are linear combinations of the components of y_i's. Denote as e_ℓ the n-dimensional vector with ℓ-th element being 1 and 0 otherwise. Then we introduce n-dimensional random vectors v_{ij}^ℓ, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and $\ell = 1, 2, \ldots, n$ defined as $v_{ij}^\ell = y_{ij} e_\ell$. We can see that an arbitrary element in $\mathcal{H}$ can be described as linear combinations of v_{ij}^ℓ's, and the number of v_{ij}^ℓ's are mn^2.

In the context of minimum-variance unbiased estimation or minimum-variance estimation, as well as recursive estimation, given a collection of n-dimensional random vectors $\{y_1, y_2, \ldots, y_m\}$ and/or p-dimensional random vectors $\{x_1, x_2, \ldots, x_q\}$, we would like to estimate a p-dimensional random vector β. Since the estimation of the random vector β can be decomposed as estimating each of its element β_i independently, cf. [OVSM P. 85, P. 86, P. 88], it is better to interpret the underlying Hilbert space as one composed of random variables, which is denoted as H, instead of one composed of random vectors like $\mathcal{H}$. In this way, if both $\{y_1, y_2, \ldots, y_m\}$ and $\{x_1, x_2, \ldots, x_q\}$ are involved, the dimension of H can be at most $mn + pq$. The reason for the preference of H over $\mathcal{H}$ is that the involved random vectors may be of different dimensions. Thus, we regard all the random vectors involved as a compact form of describing several random variables simultaneously. As a result, the matrices involved are compact representation of describing several linear combinations simultaneously. For example, if we would like to find the projection $\hat{\beta}$ of β to $\{y_1, y_2, \ldots, y_m\}$, and suppose

$$\hat{\beta} = K_1 y_1 + K_2 y_2 + \cdots + K_m y_m,$$

where K_i's are $p \times n$ matrices. Then the i-th row of $K = [K_1\, K_2\, \ldots\, K_m]$ is linear coefficients for representing $\hat{\beta}_i$, and K is a compact representation to group together the coefficients of all $\hat{\beta}_i$'s.

P. 85 For this Hilbert space of random variables, the elements that are relevant to the estimation problem are $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$, and constant 1. Therefore, y_i is linear combination of the ε_i and 1. In addition, the constraint $E(\hat{\beta}_i) = \beta_i$ is in fact $(k_i' y \,|\, 1) = \beta_i$.

P. 92

Lemma (1, P. 92). *Let Y_1 and Y_2 be two subspaces in a vector space. Then $Y_1 + Y_2 = [Y_1 \cup Y_2]$.*

Proof. We first show that $[Y_1 \cup Y_2] \subset Y_1 + Y_2$. For $x \in [Y_1 \cup Y_2]$, there exists y_i, $i = 1, 2, \ldots, n$ with $n < \infty$ such that $x = \sum_{i=1}^n \alpha_i y_i$ for some scalars α_i. Define $I_\ell = \{i : y_i \in Y_\ell\}$, $\ell = 1, 2$. Then

$$x = \sum_{i \in I_1} \alpha_i y_i + \sum_{i \in I_2} \alpha_i y_i.$$

(For the special case where $I_\ell = \emptyset$, the corresponding sum shall be interpreted according to [P. 16].) Since Y_1 and Y_2 are subspaces, we have that $\sum_{i\in I_1} \alpha_i y_i \in Y_1$ and $\sum_{i\in I_2} \alpha_i y_i \in Y_2$.

Conversely, if $x \in Y_1 + Y_2$, then there exists $y_1 \in Y_1$ and $y_2 \in Y_2$ such that $x = y_1 + y_2$, where the right hand side is a linear combination of elements in $Y_1 \cup Y_2$. $\square$

Lemma (2, P. 92)**.** *Let Y_1 and Y_2 be two subspaces in a Hilbert space. In addition, Y_1 is closed. Then $Y_1 + Y_2 = Y_1 \oplus \tilde{Y}_2$, where $\tilde{Y}_2 = (Y_1 + Y_2) \cap Y_1^\perp$.*

Proof. First, we note that since $Y_1 + Y_2$ is a subspace according to [Lemma 1, P. 92], $\tilde{Y}_2$ is a subspace according to [OVSM Prop. 1, P. 15]. In addition, $\tilde{Y}_2 \subset Y_1^\perp$ implies that $Y_1 + \tilde{Y}_2 = Y_1 \oplus \tilde{Y}_2$ according to [Lemma 1, P. 53].

Let $x \in Y_1 + Y_2$. Then $x = y_1 + y_2$ with $y_1 \in Y_1$ and $y_2 \in Y_2$. By the projection theorem [OVSM Thm. 2, P. 51], $\exists \hat{y}_2 \in Y_1$ and $\tilde{y}_2 \in Y_1^\perp$ such that $y_2 = \hat{y}_2 + \tilde{y}_2$. In addition, we have $\tilde{y}_2 = -\hat{y}_2 + y_2 \in Y_1 + Y_2$ since $-\hat{y}_2 \in Y_1$. Therefore, $\tilde{y}_2 \in \tilde{Y}_2$. As a result, $x = (y_1 + \hat{y}_2) + \tilde{y}_2 \in Y_1 \oplus \tilde{Y}_2$.

Conversely, let $x = y_1 + \tilde{y}_2$ with $y_1 \in Y_1$ and $\tilde{y}_2 \in \tilde{Y}_2$. Then $\tilde{y}_2 = z_1 + z_2$ with $z_1 \in Y_1$ and $z_2 \in Y_2$. Consequently, we have that $x = (y_1 + z_1) + z_2$, which can be seen as an element in $Y_1 + Y_2$. $\square$

Lemma (3, P. 92)**.** *Let Y_1 and Y_2 be two subspaces in a pre-Hilbert space. In addition, Y_1 is finite-dimensional. Then $Y_1 + Y_2 = Y_1 \oplus \tilde{Y}_2$, where $\tilde{Y}_2 = (Y_1 + Y_2) \cap Y_1^\perp$.*

Proof. The proof is nearly identical to that of [Lemma 2, P. 92]. The difference is that when showing that $x \in Y_1 + Y_2$ implies $x \in Y_1 + \tilde{Y}_2$, we apply [Lemma 2, P. 53] in place of the projection theorem [OVSM Thm. 2, P. 51]. $\square$

Lemma (4, P. 92)**.** *Let Y_1 be a closed subspace in a Hilbert space. Let Y_2 be a subspace spanned by a finite set of vectors $\{y_1, y_2, \ldots, y_n\}$, i.e., $Y_2 = [\{y_1, y_2, \ldots, y_n\}]$. Denote as $\hat{y}_i$ the projection of y_i on Y_1, $i = 1, 2, \ldots, n$. Then $Y_1 + Y_2 = Y_1 \oplus \tilde{Y}_2$, where $\tilde{Y}_2 = [\{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n\}]$, with $\tilde{y}_i = \hat{y}_i - y_i$, $i = 1, 2, \ldots, n$.*

Proof. First, by proving that $\tilde{Y}_2 \subset Y_1^\perp$ and applying [Lemma 1, P. 53], we see that $Y_1 + \tilde{Y}_2 = Y_1 \oplus \tilde{Y}_2$.

For $x \in Y_1 + Y_2$, there exists $z_1 \in Y_1$ and $z_2 \in Y_2$ such that $x = z_1 + z_2$. According to the definition of Y_2, we have that $z_2 = \sum_{i=1}^n \alpha_i y_i = \sum_{i=1}^n \alpha_i \hat{y}_i + \sum_{i=1}^n \alpha_i \tilde{y}_i$ for some scalars α_i. Therefore, $x = \left(z_1 + \sum_{i=1}^n \alpha_i \hat{y}_i\right) + \sum_{i=1}^n \alpha_i \tilde{y}_i \in Y_1 \oplus \tilde{Y}_2$.

Conversely, if $x \in Y_1 \oplus \tilde{Y}_2$, then $x = z_1 + \tilde{z}_2$ for some $z_1 \in Y_1$ and $\tilde{z}_2 \in \tilde{Y}_2$. Since $\tilde{z}_2$ is linear combination of $\tilde{y}_i$'s, while $\tilde{y}_i = y_i - \hat{y}_i$, we can see that x can be written as sum of two vectors from Y_1 and Y_2. $\square$

Lemma (5, P. 92)**.** *Let Y_1 be a finite-dimensional subspace in a pre-Hilbert space. Let Y_2 be a subspace spanned by a finite set of vectors $\{y_1, y_2, \ldots, y_n\}$, i.e., $Y_2 = [\{y_1, y_2, \ldots, y_n\}]$. Then every y_i has a projection $\hat{y}_i$ on Y_1, $i = 1, 2, \ldots, n$. In addition, $Y_1 + Y_2 = Y_1 \oplus \tilde{Y}_2$, where $\tilde{Y}_2 = [\{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n\}]$, with $\tilde{y}_i = \hat{y}_i - y_i$, $i = 1, 2, \ldots, n$.*

Proof. The proof is identical as the one for [Lemma 4, P. 59], once we restreat the analysis within the subspace $Y_1 + Y_2$, which is finite dimensional since $Y_1 + Y_2 = [Y_1 \cup Y_2]$ according to [Lemma 1, P. 92], thus complete. $\square$

Lemma (6, P. 92)**.** *Let Y_1 and Y_2 be subspaces in a Hilbert space H such that Y_1 is closed, Y_2 is compact, and $Y_2 \subset Y_1^{\perp}$. Then we have $Y_1 + Y_2 = Y_1 \oplus Y_2$. Besides, for all $x \in H$, there exists a unique vector $\hat{x} \in Y_1 \oplus Y_2$ such that $\|x - \hat{x}\| \leq \|x - z\|$ for all $z \in Y_1 \oplus Y_2$. In addition, let the projection of x on Y_1 and Y_2 be $\hat{y}_1$ and $\hat{y}_2$ respectively, then $\hat{x} = \hat{y}_1 + \hat{y}_2$.*

Proof. To see that $Y_1 + Y_2 = Y_1 \oplus Y_2$, we note that $Y_1 \cap Y_2 \subset Y_1 \cap Y_1^{\perp} = \{\theta\}$, and $\theta \in Y_1 \cap Y_2$ since both are subspaces. Therefore, $Y_1 \cap Y_2 = \{\theta\}$ and $Y_1 + Y_2 = Y_1 \oplus Y_2$ according to [Lemma 1, P. 53].

Due to [Lemma 5, P. 69], $Y_1 + Y_2$ is closed. Thus, for every $x \in H$. the existence of the unique minimizing vector $\hat{x}$ is assured according to the projection theorem [OVSM Thm, 2, P. 51]. To see that $\hat{x} = \hat{y}_1 + \hat{y}_2$, we just need to show that $(x - \hat{y}_1 - \hat{y}_2) \perp z$ for all $z \in Y_1 + Y_2$. For every $z \in Y_1 + Y_2$, there exists $z_1 \in Y_1$ and $z_2 \in Y_2$ so that $z = z_1 + z_2$. Then we have

$$\begin{aligned}
(z \,|\, x - \hat{y}_1 - \hat{y}_2) = &(z_1 + z_2 \,|\, x - \hat{y}_1 - \hat{y}_2) \\
=&(z_1 \,|\, x - \hat{y}_1 - \hat{y}_2) + (z_2 \,|\, x - \hat{y}_1 - \hat{y}_2) \\
=&\overline{(x - \hat{y}_1 - \hat{y}_2 \,|\, z_1)} + \overline{(x - \hat{y}_1 - \hat{y}_2 \,|\, z_2)} \\
=&(z_1 \,|\, x - \hat{y}_1) - (z_1 \,|\, \hat{y}_2) + (z_2 \,|\, x - \hat{y}_2) - (z_2 \,|\, \hat{y}_1).
\end{aligned}$$

Since $x - \hat{y}_1 \perp Y_1$ and $z_1 \in Y_1$, we have $(z_1 \,|\, x - \hat{y}_1) = 0$. Similarly, $(z_2 \,|\, x - \hat{y}_2) = 0$. In addition, $Y_2 \subset Y_1^{\perp}$ implies that $(z_1 \,|\, \hat{y}_2) = (z_2 \,|\, \hat{y}_1) = 0$. $\square$

P. 93 Suppose the past measurement is composed of finite length random vector z and according to [OVSM Thm. 1, P. 87], we have $\hat{\beta} = E(\beta z')[E(zz')]^{-1}z$. Similarly, we have

$$\begin{aligned}
\hat{y} =&E(yz')[E(zz')]^{-1}z = E[(W\beta + \varepsilon)z'][E(zz')]^{-1}z \\
=&WE(\beta z')[E(zz')]^{-1}z + E(\varepsilon z')[E(zz')]^{-1}z.
\end{aligned}$$

Since ε has zero mean and is uncorrelated to the previous measurement, we have $\hat{y} = W\hat{\beta}$.

From abstract viewpoint, we have the following lemma.

Lemma (P. 93)**.** *Let M be a subspace in a per-Hilbert space X. Let y_1 and y_2 be some vectors in X and assume that there exist $\hat{y}_1 \in M$ and $\hat{y}_2 \in M$ satisfying $\|y_1 - \hat{y}_1\| \leq \|y_1 - z\|$ and $\|y_2 - \hat{y}_2\| \leq \|y_2 - z\|$ for all $z \in M$. For arbitrary scalars α_1 and α_2, define $y = \alpha_1 y_1 + \alpha_2 y_2$. Then the vector $\hat{y} = \alpha_1 \hat{y}_1 + \alpha_2 \hat{y}_2$ fulfills that $\|y - \hat{y}\| \leq \|y - z\|$ for all $z \in M$.*

Proof. The proof can be obtained by first showing that $y - (\alpha_1 \hat{y}_1 + \alpha_2 \hat{y}_2) = \alpha_1(y_1 - \hat{y}_1) + \alpha_2(y_2 - \hat{y}_2)$, thus orthogonal to M, according to the definition of $\hat{y}_1$ and $\hat{y}_2$. Then the second half of [OVSM Thm. 1, P. 50] assures the desired conclusion. □

From above theorem, we see that for $y = W\beta + \varepsilon$, the projection of $W\beta$ to the previous measurement is $W\hat{\beta}$. On the other hand, since ε is uncorrelated to the previous measurements and has zero mean, the elements of ε belongs to the orthogonal complement of the subspace spanned by previous measurements. Then according to [Lemma, P. 50], the projection of ε to previous measurement is an m-dimensional vector composed of θ's (note that the underlying (pre-)Hilbert space is one composed of random variables, instead of random vectors). Therefore, $\hat{y} = W\hat{\beta}$.

P. 97

Lemma (P. 97)**.** *Let M be a subspace in a per-Hilbert space X. Let x and y be some vectors in X and assume that there exist $\hat{x} \in M$ and $\hat{y} \in M$ satisfying $\|x - \hat{x}\| \leq \|x - z\|$ and $\|y - \hat{y}\| \leq \|y - z\|$ for all $z \in M$. Then we have that*

$$(x - \hat{x} \,|\, y - \hat{y}) = (x \,|\, y - \hat{y}) = (x - \hat{x} \,|\, y).$$

In particular, we have that

$$\|x - \hat{x}\|^2 = (x \,|\, x - \hat{x}) = (x - \hat{x} \,|\, x),$$

and $(x \,|\, x - \hat{x}) = (x - \hat{x} \,|\, x)$ is a real number.

Proof. By [OVSM Thm. 1, P. 50], we have $(\hat{x} \,|\, y - \hat{y}) = (\hat{y} \,|\, x - \hat{x}) = 0$. Therefore,

$$(x - \hat{x} \,|\, y - \hat{y}) = (x \,|\, y - \hat{y}) - (\hat{x} \,|\, y - \hat{y}).$$

Similarly, we have that

$$(x - \hat{x} \,|\, y - \hat{y}) = \overline{(y - \hat{y} \,|\, x - \hat{x})} = \overline{(y \,|\, x - \hat{x})} - \overline{(\hat{y} \,|\, x - \hat{x})} = (x - \hat{x} \,|\, y).$$

The rest part can be proven by carrying out the same computation. □