5

BASIC FUNCTIONAL ANALYSIS

5.1 Normed vector space

5.1.1 Basic properties

Definition 5.1.1 (normed vector space). *A vector space equipped with a norm is a normed vector space. Let V be a vector space over the field $\mathbb{F}$. A norm of V is a function $\|\cdot\| : V \to \mathbb{R}^+$ that satisfies the following conditions:*

1. *Positive-definiteness: $\|v\| \geq 0, \forall v \in V$ with equality iff $v = 0$*
2. *Homogeneity: $\|aV\| = |a| \|v\|, \forall v \in V, \forall a \in \mathbb{F}$*
3. *Triangle inequality: $\|u + v\| \leq \|v\| + \|u\|, \forall u, v \in V$*

Normed vector space and metric space [Definition 2.1.1] are closely related. In factor, the norm function in a normed vector space can also be used a metric, as we show in the following.

Lemma 5.1.1 (normed as a metric). *Let $\|\cdot\|$ denote the norm of a normed vector space. Then we can define a metric d via*

$$d(x, y) = \|x - y\|.$$

Proof. We can verify $d(x, y)$ satisfies

1. $d(p, q) > 0$ if $p \neq q$;
2. $d(p, q) = 0$ if and only if $p = q$;
3. $d(p, q) = d(q, p)$;
4. (triangle inequality) we have

$$\begin{aligned} d(x, z) &= \|x - z\| \\ &= \|x - y + y - z\| \\ &\leq \|x - y\| + \|y - z\| \\ &= d(x, y) + d(y, z) \end{aligned}$$

□

Note 5.1.1 (inner product, norm, and metric).

- Let $\|\cdot\|$ denote the norm, we can define a metric via

$$d(x, y) = \|x - y\|.$$

Such metric will turn a normed vector space into more general metric space.

- Every inner product on a vector space induces a norm that can be used to measure the 'length' of the a vector [Lemma 5.3.1].
- Note that a norm can be simply defined using the inner product; but it can also be defined as other mapping as long as it satisfies above conditions.

Example 5.1.1.

- norm in $\mathbb{R}^n$ vector space includes 1-norm, ∞-norm, p-norm. They are given by

$$\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$$

- For vector space of matrix $M_{m\times n}(\mathbb{R})$, Frobenius norm
- Let $p \geq 1$ be a real number. The space $L^p([a,b])$ is normed vector space with L^p- norm given by

$$\|f\|_p = (\int_a^b |f(t)|^p \, dt)^{1/p}$$

Another useful triangle inequality associated with the norm function is the minus sign triangle inequality.

Lemma 5.1.2 (triangle inequality minus sign). *In a normed linear space*

$$\left| \|x\| - \|y\| \right| \leq \|x - y\|,$$

that is,

$$\|x\| - \|y\| \leq \|x - y\|, \|y\| - \|x\| \leq \|x - y\|$$

and

$$\|x\| - \|y\| \geq -\|x - y\|$$

for any two vectors x, y.

Proof. use triangle inequality for vector $x - y$ and y. □

Using the minus sign triangle inequality, we can show that every norm function must be a continuous function.

Theorem 5.1.1 (continuity of norm). *The norm as a function in a normed vector space E is a continuous function; that is, let $\{x_n\} \in E$ and we have*

$$\|x_n - x\| \to 0$$

implies

$$\big|\|x\| - \|x_n\|\big| = 0$$

Proof. Use triangle inequality

$$\big|\|x\| - \|x_n\|\big| \leq \|x - x_n\| \to 0.$$

□

5.1.2 Equivalence of norms

Definition 5.1.2 (equivalent norms). *A norm $\|\cdot\|_p$ on a vector space X is said to be **equivalent** to a norm $\|\cdot\|_q$ on X if there are positive numbers a and b such that for all $x \in X$, we have*

$$a\|x\|_q \leq \|x\|_p \leq b\|x\|_q.$$

Remark 5.1.1 (equivalent relation)**.** It can be easily showed that equivalent norms satisfies three properties: reflective, transitive and symmetric relation(we can scale p norm properly such that q norm will be squeezed instead).

Theorem 5.1.2 (equivalent norms for finite dimensional vector space). *[1, p. 75] On a **finite** dimension vector space X, any norm is equivalent to any other norm.*

Remark 5.1.2 (interpretations)**.**

- This theorem only holds for finite dimensional vector space, and might not hold for infinite dimensional space.
- This theorem implies that convergence or divergence of a sequence in a finite dimensional vector space does not depend on the particular choice of a norm on that space.

In the following, we list several useful norm inequality between different norms.

Lemma 5.1.3 (useful norm inequalities). *Let V be a n dimensional normed space. Then, for different norms, we have*

- $$\frac{1}{\sqrt{n}}\|x\|_1 \leq \|x\|_2 \leq \|x\|_1$$
- $$\frac{1}{\sqrt{n}}\|x\|_2 \leq \|A\|_\infty.$$

- *More generally, for* $0 < p < q$,

$$\|x\|_p \geq \|x\|_q .$$

Proof. (1)

$$\sqrt{x_1^2 + x_2^2 + ... + x_n^2} \leq \sum_{i=1}^{n} |x_i|$$

can be showed by squaring both sides. To show

$$\frac{1}{n}\|x\|_1^2 \leq \|x\|_2^2 ,$$

we use the fact that $\|x\|_1^2 = x^T J x/n, \|x\|_2^2 = x^T I x$, where J is the all 1 matrix and I is the identity matrix. It can be showed that

$$\frac{1}{n}J, I, (\frac{1}{n}J - I),$$

are all symmetric and idempotent(i.e. they are orthogonal projection matrix.). Then

$$x^T \frac{J}{n} x - x^T I x \geq 0$$

since an orthogonal projection matrix is semi-positive definite. Theorem 4.5.7.

(2)(3) need to use Holder's inequality. A good reference is link. □

5.2 Contraction mapping and fixed point theorems

5.2.1 Complete normed space (Banach space)

In many optimization algorithms, when seeking an optimal vector, we often construct a sequence of vectors, the desired optimal vector is the limit of such sequence. A normed vector space usually does not guarantee the existence of a limit if the sequence of a Cauchy sequence.

In the section, we first introduce the concept of complete normed space and then examine one of most powerful theoretical tool associated with such complete normed space - fixed point theorem.

Definition 5.2.1 (complete normed space and Banach space). *A normed vector space V is **complete** if the following two requirements are satisfied:*

- *Every Cauchy sequence of elements in V must be convergent.*
- *The limit of the Cauchy sequence must belong to V.*

*A complete normed space is called **Banach space**.*

A normed vector space and a Banach space for sequences have the following critical differences:

- In any normed vector space, a convergent sequence is automatically a Cauchy sequence.
- However, for some normed vector space, there exists nonconvergent Cauchy sequence.
- In a Banach space, a sequence is convergent if and only if it is Cauchy.

It can be showed that finite dimensional normed space is always complete.

Theorem 5.2.1 (closedness and completeness of finite dimensional normed space). *[1, p. 74]link Consider any finite dimensional subspace Y of a normed space X. It follows that*

- *If X is finite dimensional, then X is complete.*
- *Y is closed in X .*

In the following, we have listed a number of common Banach spaces [2, p. 48][3, p. 35]:

Theorem 5.2.2 (common Banach spaces).

- *The space $\mathbb{R}^n$ with Euclidean norm is a Banach space.*
- *The spaces $l^p(\mathbb{N}), 1 \leq p < \infty$ are Banach spaces.*
- *The space $L^p, L^p[0,1], 1 \leq p \leq \infty$ are Banach spaces.*
- *The space $C[a,b]$ of continuous functions on $[a,b]$ is a Banach space.*

Proof. (1) Note we use the Euclidean norm as the metric(Lemma 5.1.1], we can use the result that $\mathbb{R}^n$ is a complete metric space(Theorem 2.5.2], which implies that every Cauchy sequence in $\mathbb{R}^n$ converges to a limit in $\mathbb{R}^n$. Further $\mathbb{R}^n$ is a vector space and therefore, $\mathbb{R}^n$ is a Banach space. (2) We need to prove every Cauchy sequence converge to an element in $C[a,b]$. Let $\{x_n\}$ be a Cauchy sequence in $C[a,b]$. Let $t \in [a,b]$, then $|x_n(t) - x_m(t)| \leq \|x_n - x_m\| \to 0$, therefore, $\{x_n(t)\}$ converges pointwise to $x(t) \in \mathbb{R}$.(because $\mathbb{R}$ is complete, and every Cauchy sequence will converge) Now we show that the convergence is uniform:

$$|x_n(t) - x(t)| \leq |x_n(t) - x_m(t)| + |x_m(t) - x(t)| \leq \|x_n - x_m\| + |x_m(t) - x(t)| \to 0$$

since for a Cauchy sequence, any $m > N$, including $m = \infty$ is legitimate, therefore the first term is bounded and the second term can make to arbitrarily small.

Because uniform convergence preserve continuity, since $x_m(t)$ is continuous, therefore the limit $x(t)$ is continuous, therefore $x \in C[a,b]$. (3)(4) See reference. □

5.2.2 Contraction mapping

Contraction mapping in a normed space is a mapping that shrink distance.

Definition 5.2.2 (Contraction mapping and fixed point).

- *An operator/mapping $f : X \to X$ is called a* contraction mapping *if there exists a constant $\beta, 0 < \beta < 1$, such that*

$$\|f(x) - f(x_0)\| \leq \beta \|x - x_0\|.$$

- *x is a fixed point of $f : X \to X$ if $f(x) = x$.*

Contraction mappings need not be linear mapping. Besides using the definition to identify a contraction maps, we can also check:

- Continuity. Contraction mappings are continuous since $\|f(x) - f(x_0)\| \leq \|x - x_0\| \to 0$.
- We can usually use bounds on first-order derivative and Lipschitz constant to check whether a mapping is contraction mapping. For a continuous differentiable function

g defined on $[a,b]$, if $|g'(x)| < 1$, then g is a contraction mapping. To see this, we have $|g(x) - g(y)| \leq |g'(x')||x - y|$.

Example 5.2.1. Consider metric space $M = \mathbb{R}$ with metric $\|x\| = |x|$. Consider a mapping $T : M \to M$ to be $T(x) = kx, k \in \mathbb{R}, k < 1$. Then T is a contraction mapping since $\|T(x) - T(y)\| = |kx - ky| = |k|\,|x - y|$. And the fixed point is 0.

5.2.3 Banach fixed point theorem

Theorem 5.2.3 (Banach fixed point theorem). *[4][5, p. 122][6, p. 771]Let $f : X \to X$ be a contraction mapping with contraction parameter $\beta < 1$ defined on a closed subset X of a Banach space. Then there exists a unique fixed point x; That is, for any $x_0 \in X$, the sequence $\{x_n\}$ generated by $x_{n+1} = f(x_n)$ converges to x.*

Proof. Select an arbitrary $x_0 \in X$. Define the sequence $\{x_n\}$ by $x_{n+1} = f(x_n)$. Denote ρ as the distance metric function induced by the norm as $\rho(x,y) = \|x - y\|$.. We have

- $\{x_n\}$ is the Cauchy sequence:
  - $\rho(x_{n+1}, x_n) = \rho(f(x_n), f(x_{n-1})) \leq \beta^n \rho(x_1, x_0))$
  - Using triangle inequality, we have for any $m > 0$

$$\rho(x_n, x_{n+m}) \leq \rho(x_n, x_{n+1}) + \rho(x_{n+1}, x_{n+2}) ... + \rho(x_{n+m-1}, x_{n+m}) \leq \beta^n/(1-\beta)\rho(x_0, x_1) \to 0, n \to \infty$$

- $\{x_n\}$ converges to $x \in X$: since X is complete and closed.
- x is a fixed point: since f is a contraction, and thus uniformly continuous, therefore

$$f(x) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x$$

- x is unique. Suppose that $f(z) = z, f(x) = z$.Then

$$\rho(x,z) = \rho(f(x), f(z)) \leq \beta\rho(x,z)$$

  since $\beta < 1$, we can only have $\rho(x,z) = 0, x = z$.

□

Remark 5.2.1 (global convergence property). The convergence is ensured from any starting point in X. And we use $\rho(x,y) = \|x - y\|$.

Corollary 5.2.3.1 (The rate of convergence). *Let x be the fixed point of T. Then* $\rho(x_n, x) \leq \beta^n \rho(x_0, x)$

Proof. $\rho(x_n, x) = \rho(Tx_{n-1}, Tx) \leq \beta\rho(x_{n-1}, x) \leq \beta^2\rho(x_{n-2}, x)$ □

5.2.4 Applications in root finding

Let $g \in C^1[a, b]$ be such that $g(x) \in [a, b]$. In addition, suppose that g is globally Lipschitz with a constant $0 < k < 1$. Then for any initial number $p_0 \in [a, b]$, the sequence defined by

$$p_n = g(p_{n-1}), n \geq 1$$

converges to an unique fixed point $p^* \in [a, b]$

To see this, based on definition of Lipschitz continuity, we have

$$|g(p_{n-1}) - g(p^*)| \leq k|p_{n-1} - p^*|,$$

which gives

$$|p_n - p^*| \leq k|p_{n-1} - p^*| \implies |p_n - p^*| \leq k^n|p_0 - p^*| \to 0.$$

Example 5.2.2. Consider $f(x) = 0.5x - \exp(-x)$ defined on $\mathbb{R}_{++}$, it can be showed that its derivative is bounded by 1.

Remark 5.2.2. For nonlinear equations, if we can write the equation as $x = f(x)$ and show that $f(x)$ satisfies above conditions, then we can use fixed point iteration to solve the equation.

5.2.5 Application to numerical linear equations

In subsection 4.17.1, we explore a generic numerical iteration approach to solution of $Ax = b$.

The working mechanism is to decompose $A = M - N$ with M being nonsingular, then the iteration

$$x = M^{-1}(b + Nx).$$

can bring any initial starting vector x^0 towards the solution $x^* = A^{-1}b$ as along as the norm of $M^{-1}N$ is less than 1.

Let T be the operator such that $x = M^{-1}(b + Nx)$. Then $Tx^* = x^*$ indicting that x^* is the fixed point.

To show T is a contraction mapping, we have

$$\|Tx_1 - Tx_2\| = \left\|M^{-1}N(x_1 - x_2)\right\| \leq \left\|M^{-1}N\right\| \|x_1 - x_2\|$$

where $\left\|M^{-1}N\right\|$ is matrix norm of $M^{-1}N$, and we have used inequality based on matrix norm.

Since T is a contraction mapping and x^* is the fixed point of T, then the iteration will converge to x^*.

5.2.6 Applications to integral and differential equations

Theorem 5.2.4. *[7, p. 225]If A is a bounded linear operator on a Banach space E, and ϕ is an arbitrary element of E, for the operator equation $x = \alpha Ax + \phi = Tx$, fi $\|Af\| \leq k\|f\|, \forall f \in E$(or $\|A\| = k$) then $Tx = x$ has a unique solution whenever $|\alpha| k < 1$. Moreover, the solution is given as*

$$x = \phi + \alpha A\phi + \alpha^2 A^2 \phi ...$$

Proof:(1) **We need to show the operator** $Tx = \alpha Ax + \phi$ **is a contraction mapping**.

$$\|Tx_1 - Tx_2\| = \|\alpha A(x_1 - x_2)\| \leq |\alpha| \|A\| \|x_1 - x_2\| = |x| k \|x_1 - x_2\|$$

(2) Since T is a contraction mapping, we can generate the solution by iterating

$$x_{n+1} = Tx_n$$

with x_0 be an arbitrary element in E. We have

$$\begin{aligned}
x_0 &= f \\
x_1 &= \alpha Ax_0 + \phi \\
x_2 &= \alpha Ax_1 + \phi = \alpha^2 A^2 x_0 + \alpha A\phi + \phi \\
\ldots
\end{aligned}$$

Theorem 5.2.5 (Neumann series). *[7, p. 226]If A is a bounded linear operator in a Banach space E and $\|A\| < \lambda$, then*

$$A_\lambda = (A - \lambda I)^{-1} = -\sum_{i=0}^{\infty} \frac{A^n}{\lambda^{n+1}}$$

and A_λ is bounded as

$$\|A_\lambda\| \leq \frac{1}{|\lambda| - \|A\|}$$

Corollary 5.2.5.1. *If A is a bounded linear operator in a Banach space E and $\|A\| < 1$, then*

$$B = (I - A)^{-1} = \sum_{i=0}^{\infty} A^i$$

and B is bounded by

$$\|B\| \leq \frac{1}{1 - \|A\|}$$

Proof: (1) $(I-A)^{-1}(\sum_{i=0}^{\infty} A^i) = I - A^\infty = I$(since the infinite power of a contraction mapping is zero mapping in Banach space) (2) use the inequality

$$\|B\| = \left\| \sum_{i=0}^{\infty} A^i \right\| \leq \sum_{i=0}^{\infty} \|A\|^i = \frac{1}{1 - \|A\|}$$

Corollary 5.2.5.2. *If A is a bounded linear operator in a Banach space E and $\|A\| < 1$, then the equation*

$$x = x_0 + Ax$$

has a unique solution given by

$$x = \sum_{n=0}^{\infty} A^n x_0$$

Proof: $x = x_0 + Ax \Leftrightarrow (I - A)x = x_0$

Theorem 5.2.6 (Picard's existence and uniqueness theorem). *[7, p. 228] Consider the initial value problem for the ordinary differential equation*

$$\frac{dy}{dt} = f(t, y)$$

with the initial condition $y(t_0) = y_0$, where f is a continuous function in some closed domain

$$R = \{(t, y) : a \leq t \leq b, c \leq y \leq d\}$$

containing the point (t_0, y_0) in its interior. If f satisfies the Lipschitz condition

$$|f(t, y_1) - f(t, y_2)| \leq K|y_1 - y_2|$$

for $K > 0$ and all $t, y_1, y_2 \in R$, then there exist a unique solution of $y = \phi(x)$ defined in ***some neighborhood of*** t_0

Proof: define $S = C[t_0 - \epsilon, t_0 + \epsilon]$ to be Banach space with the sup-norm

$$\|\phi\| = \sup_x |\phi(x)|$$

Define operator T on S, such that $(Ty)(t) = y_0 + \int_{t_0}^{t} f(\tau, y) d\tau$

For any $\phi_1, \phi_2 \in S$, we have

$$\|T\phi_1 - T\phi_2\| = \sup_x \left| \int_{x_0}^{x} (f(t, \phi_1) - f(t, \phi_2)) dt \right| \leq K\epsilon |\phi_1 - \phi_2|$$

where ϵ is chosen such that $K\epsilon < 1$, then T is the contraction mapping, and therefore $Tx = x$ has a solution in S.

Remark 5.2.3.

- The requirement on $f(t, y)$ can be summarized as continuous and 'slow'-changing(reflected by Lipschitz condition.)
- Other 'restrictive' condition could be $f(t, y)$ continuous and continuously differentiable on closed interval.

5.3 Inner product space and Hilbert space

5.3.1 Inner product space (pre-Hilbert space) and Hilbert space

5.3.1.1 *Foundations*

Definition 5.3.1 (inner product space). *An **inner product space** is a vector space V over $\mathbb{F}$ equipped with an additional structure called inner product that assigns each pair of vectors $v_1, v_2 \in V$ a number $\langle v_1, v_2 \rangle \in \mathbb{F}$, and its assignment satisfies*

1. *Conjugate symmetry:* $\langle v_1, v_2 \rangle = \overline{\langle v_2, v_1 \rangle}$
2. *Linearity in the first slot:* $\langle c_1 v_1 + c_2 v_2, v_3 \rangle = c_1 \langle v_1, v_2 \rangle + c_2 \langle v_2, v_3 \rangle$*, for all* $c_1, c_2 \in \mathbb{F}$ *(many physicists define linearity in the second slot)*
3. $\langle v, v \rangle \geq 0$ *and* $\langle v, v \rangle = 0$ *iff* $v = 0$

Remark 5.3.1. Inner product space is sometimes called **pre-Hilbert space**.

Remark 5.3.2. **The standard inner product on complex vector fields** is usually defined as

$$\langle u, v \rangle = u^T \overline{v}$$

by mathematicians or

$$\langle u, v \rangle = u^H v = \overline{u}^T v$$

by physicists. The result are conjugate to each other. The second is simplier in writing. We prefer the second.

Theorem 5.3.1 (Cauchy-Schwarz inequality). *[8][9, p. 100]In an inner product space S with induced norm, we have Cauchy-Schwarz inequality : for $x, y \in V$, we have*

$$\langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle$$

with equality if and only if $x = \lambda y$.

If use the inner product as the induced norm [Lemma 5.3.1], we can write the inequality as

$$|\langle x, y \rangle| \leq \|x\| \|y\| .$$

Proof. use the fact the $\langle x - \lambda y, x - \lambda y \rangle \geq 0$ is valid for all real λ,then require quadratic equation determinant to be less than 0. Specifically,

$$\langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle \geq 0, \forall \lambda.$$

The left reaches minimum when $\lambda = \langle x, y \rangle / \langle y, y \rangle$ with minimum value

$$\langle x, x \rangle - \frac{\langle x, y \rangle^2}{\langle y, y \rangle} \geq 0.$$

□

Corollary 5.3.1.1 (Cauchy-Schwarz inequality for finite real-valued terms). *[8, p. 120] Let $a_1, a_2, ..., a_n \in \mathbb{R}$ and $b_1, b_2, ..., b_n \in \mathbb{R}$. Then*

$$\left| \sum_{k=1}^{n} a_k b_k \right| \leq \sqrt{(\sum_{k=1}^{n} a_k^2) \sum_{k=1}^{n} a_k^2)}$$

Proof. Consider inner product space $\mathbb{R}^n$ with inner product defined by

$$\langle a, b \rangle = \sum_{i=1}^{n} a_i b_i.$$

□

Lemma 5.3.1 (inner product as induced norm). *[8, p. 290] [10, p. 172] When inner product is used as norm, we have*

$$\|u + v\| \leq \|u\| + \|v\|$$

Proof.

$$\begin{aligned}
\|u + v\|^2 &= \|u\|^2 + \|v\|^2 + \langle u, v \rangle + \langle v, u \rangle \\
&= \|u\|^2 + \|v\|^2 + 2Re(\langle u, v \rangle) \\
&= \|u\|^2 + \|v\|^2 + 2|\langle u, v \rangle| \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\|
\end{aligned}$$

where Cauchy-Schwarz inequality (Theorem 5.3.1]is used. □

Lemma 5.3.2 (parallelogram law). *[2, p. 64] Let V be an inner product space with induced norm. For all $v, w \in V$, we have*

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2).$$

Proof. Note that

$$\|v+w\|^2 = \langle v+w, v+w \rangle = \|v\|^2 + \|w\|^2 + 2\langle v, w \rangle,$$

and

$$\|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2\langle v, w \rangle.$$

□

Remark 5.3.3 (caution!). For general normed vector spaces, we usually do not have such parallelogram law, since these norms do not have an associated inner product space inducing it.

Lemma 5.3.3 (basic properties of inner product).

- *For fixed $u \in V$, $\langle v, u \rangle$ is linear map from U to $\mathbb{F}$ (for the second slot, it is not)*
- $\langle u, \lambda v \rangle = \bar{\lambda}\overline{\langle u, v \rangle}, \forall \lambda$
- $\langle u, v+w \rangle = \langle u, v \rangle + \langle u, w \rangle$

Proof. Proof of (2): $\langle u, \lambda v \rangle = \overline{\langle \lambda v, u \rangle} = \bar{\lambda}\overline{\langle v, u \rangle} = \bar{\lambda}\langle u, v \rangle$ □

Lemma 5.3.4 (continuity of inner product). *In an inner product space, if $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n \rangle \to \langle x, y \rangle$.*

Proof.

$$\begin{aligned}
|\langle x_n, y_n \rangle - \langle x, y \rangle| &= |\langle x_n, y_n \rangle - \langle x_n, y \rangle + \langle x_n, y \rangle - \langle x, y \rangle| \\
&\leq |\langle x_n, y_n \rangle - \langle x_n, y \rangle| + |\langle x_n, y \rangle - \langle x, y \rangle| \\
&= |\langle x_n, y_n - y \rangle| + |\langle x_n - x, y \rangle| \\
&\leq \|x_n\|\|y_n - y\| + \|y_n\|\|y_n - y\| \\
&\to 0
\end{aligned}$$

where we use Cauchy-Schwarz inequality (Theorem 5.3.1] and the fact that convergent sequence is bounded. □

5.3.2 Hilbert spaces

5.3.2.1 *Basics*

Definition 5.3.2 (Hilbert space). *[2, p. 65]A vector space V with an inner product is called a* ***Hilbert space*** *if the vector space V equipped with the induced norm (Lemma 5.3.1] is a Banach space.*

In the following, we listed common Hilbert spaces.

Theorem 5.3.2 ($\mathbb{R}^n$ is Hilbert space). *[2, p. 65]*

- *The $\mathbb{R}^n$ with the inner product is Hilbert spaces.*
- *The $l^2(\mathbb{N})$ is a Hilbert spaces.*

Proof. Note that this inner product induces a norm (Lemma 5.3.1]. Therefore, since $\mathbb{R}^n$ is a complete Banach space (Theorem 5.2.2], this inner product will make it a Hilbert space. □

Note 5.3.1. Note that only $l^p(\mathbb{N}), p = 2$ is Hilbert space. For any other p, $l^p(\mathbb{N})$ is not a Hilbert space. Suppose $l^p(\mathbb{N})$ is an Hilbert space. So it must satisfy the parallelogram law (Lemma 5.3.2]. That is, for all $u, v \in l^p(\mathbb{N})$:

$$2\|u\|_p^2 + 2\|v\|_p^2 = 2\|u+v\|_p^2 + 2\|u-v\|_p^2.$$

If we take $u = e_1 = (1, 0, ..., 0, ...)$ and $v = e_2 = (0, 1, 0, ..., 0, ...)$, we have

$$2 + 2 = 2^{2/p} + 2^{2/p},$$

which only holds when $p = 2$.

5.3.3 Orthogonal decomposition of Hilbert spaces

5.3.3.1 *Orthogonality*

Definition 5.3.3 (complementary subspaces, orthogonal complementary subspaces). *[11, pp. 392, 403][9, p. 108]*

- *Subspaces A, B of a Hilbert space H are said to be* ***complementary subspaces*** *if*

$$H = A + B, A \cap B = 0.$$

We can also denoted as

$$H = A \oplus B.$$

- *Let x, y be vectors in a Hilbert space H, we say x and y are* ***orthogonal*** *if $\langle x, y \rangle = 0$. We say subspaces A and B are orthogonal to each other if for every vector $x \in A, y \in B$, we have $\langle x, y \rangle = 0$.*
- *The* ***orthogonal complement*** *of a subspace A, denoted by $A^\perp$, is the set of vectors orthogonal to A, i.e.,*

$$A^\perp = \{x \in H | \langle x, y \rangle = 0 \forall y \in A\}.$$

Remark 5.3.4 (orthogonality does not implies complementary). **Caution! Two subspaces orthogonal to each other does not imply they are complementary to each other.** For example, any subspace is orthogonal to $\{0\}$, but they are not necessarily complementary to each other.

Theorem 5.3.3 (Orthogonality relation in inner product space). *[9, p. 108] Let V and W be* ***subsets*** *of inner product space S (not necessarily complete), then*

- *The orthogonal complementary subspace of V, denoted by $V^\perp$, is closed subspace of S.*
- $V \subset V^{\perp\perp}$
- *If $V \subset W$, then $W^\perp \subset V^\perp$*
- $V^{\perp\perp\perp} = V^\perp$
- *If $x \in V \cap V^\perp$, then $x = 0$*
- $\{0\}^\perp = S, S^\perp = \{0\}$

Proof. (1) Let $\{x_n\}$ be a convergent sequence in $V^\perp$, then $\langle x_n, v \rangle = 0$, for any $v \in V$. Then we can show

$$\lim_{n\to\infty} \langle x_n, v \rangle = \left\langle \lim_{n\to\infty} x_n, v \right\rangle = 0$$

which is due to the continuity of inner product. Therefore, $\lim_{n\to\infty} x_n$ is lying in $V^\perp$. And therefore $V^\perp$ is closed. (2) If $x \in V$, then $\langle x, y \rangle = 0, \forall y \in V^\perp$, therefore $x \in V^{\perp\perp}$. (3) Let $w' \in W^\perp$, then $w' \perp w, \forall w \in W \Rightarrow w' \perp v \in V$ since $V \subset W$, therefore $w' \in V^\perp$, therefore $W^\perp \subset V^\perp$. (4) From (2) $V^\perp \subset V^{\perp\perp\perp}$, from (3) $V^{\perp\perp\perp} \subset V^\perp$. (5) $\langle x, x \rangle = 0 \Rightarrow x = 0$ □

5.3.4 Projection and orthogonal decomposition

Theorem 5.3.4 (projection theorem in inner product space). *[3, p. 50] Let X be an inner product space in X, and let M be a subspace of X. It follows that*

- ***If there exists a vector*** *$m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|, \forall m \in M$, then m_0 is unique and $\langle x, x - m_0 \rangle = 0$.*

- *Moreover, a necessary and sufficient condition for $m_0 \in M$ being the unique minimizing vector in M is that*

$$(x - m_0) \perp M$$

Proof. (1) m_0 **is minimazing vector implies** $x - m_0$ **is orthogonal to** M. We prove by **contraposition**. We suppose $x - m_0$ is not orthogonal to M, then there is $m \in M$ such that $\langle x - m_0, m \rangle = \delta \neq 0$. WLOG, we let $\|m\| = 1$. Let $m_1 \in M$ be $m_1 = m_0 + \delta m$, then

$$\|x - m_1\|^2 = \|x - m_0\|^2 - \delta^2 < \|x - m_0\|^2$$

Therefore, if $x - m_0$ is not orthogonal to M, m_0 is not a minimizing vector. (2) $x - m_0$ is orthogonal to M implies m_0 is unique minimizer. For any $m \in M$,

$$\begin{aligned}\|x - m\|^2 &= \|x - m_0 + m_0 - m\|^2 \\ &= \|x - m_0\|^2 + \|m_0 - m\| \\ &< \|x - m_0\|^2, \forall m \neq m_0\end{aligned}$$

□

Theorem 5.3.5 (projection theorem in Hilbert space). *[3, p. 51] [2, p. 67] Let X be a Hilbert space, let x be a given element in X, and let M be a closed subspace of X.*

- *Then there exists a vector $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|, \forall m \in M$, then m_0 is unique and $\langle x, x - m_0 \rangle = 0$.*
- *Moreover, a necessary and sufficient condition for $m_0 \in M$ being the unique minimizing vector in M is that*

$$(x - m_0) \perp M.$$

Proof. **We only need to establish the existence of minimizing vector**, the rest has been proved in the above inner space projection theorem (Theorem 5.3.4]. If $x \in M$, then $m_0 = x$ and everything is settled. Let us assume $x \notin M$, and defined $\delta = \inf_{m \in M} \|x - m\|$.

Let $\{m_i\}$ be a sequence in M that $\|x - m_i\| \to \delta$. Then

$$\left\|m_i - m_j\right\|^2 = 2\|m_i - x\|^2 + 2\left\|m_j - x\right\|^2 - 4\left\|x - \frac{m_i + m_j}{2}\right\|^2 \to 0$$

as $i, j \to 0$ Because of M is closed subspace, then the Cauchy sequence has a limit in M, and the limit is the unique minimizing vector. □

Theorem 5.3.6 (Orthogonal decomposition in Hilbert space). *[2, p. 68] [3, p. 53]Let S be an arbitrary closed subspace of an complete Hilbert space V, then*

$$V = S \oplus S^{\perp}, S = S^{\perp\perp};$$

that is, any element $x \in V$ can be uniquely written $x = a + b, a \in S, b \in S^{\perp}$.

Proof. If $S = \{0\}$, then $S^{\perp} = H$ and the result holds. We therefore assume $S \neq \{0\}$. (1)From projection theorem in Hilbert space (Theorem 5.3.5], any element $x \in V$ has a unique element $u \in S$ satisfying $x \perp u \in S^{\perp}$. Then we can write $x = u + (x - u)$ uniquely. Because the decomposition is unique, therefore it is direct sum. (2) We have showed that $S \subset S^{\perp\perp}$ (Theorem 5.3.3]. To show $S = S^{\perp\perp}$, let $x \in S^{\perp\perp} \subset V$, then $x = a + b, a \in S, b \in S^{\perp}$(using (1), every $x \in V$ can be written as the sum of components in S and $S^{\perp}$). Since $a \in S, S \subseteq S^{\perp\perp}$, a is also in $S^{\perp\perp}$, then $x - a = b \in S^{\perp\perp}$, therefore $b = 0$(using Theorem 5.3.3, $b \in S^{\perp} \cap S^{\perp\perp} \implies b = 0$). Eventually, we have $x = a \in S, S = S^{\perp\perp}$. □

5.4 Approximations in Hilbert space

5.4.1 Approximation via projection

Theorem 5.4.1 (Hilbert space subspace approximation theorem). *Let $(S,\|\cdot\|]$ be a Hilbert space. Let $T = \{p_1, p_2, ..., p_m\} \subset S$ be a set of linearly independent vectors in S. Let $V = span(T)$. Given a vector $x \in S$, the element $\hat{x}$ in T such that*

$$\|x - \hat{x}\| \leq \|x - y\|, \forall y \in T$$

can found via the following procedure.

- *Let $\tilde{x} = \sum_{i=1}^m c_i p_i$ with unknown parameters $c_1, c_2, ..., c_m$. The coefficients $c_1, c_2, ..., c_n$ has to satisfy the orthogonality condition of*

$$\langle x - \hat{x}, \hat{x} \rangle = 0;$$

or equivalently,

$$\left\langle x - \sum_{i=1}^m c_i p_i, p_j \right\rangle = 0, \forall j \in \{1, 2, ..., m\}.$$

- *The orthogonal condition gives a system of equation as*

$$Gc = p, c, p \in \mathbb{R}^m, G \in \mathbb{R}^{m\times m},$$

where

$$G = \begin{bmatrix} \langle p_1, p_1 \rangle & \langle p_1, p_2 \rangle & \langle p_1, p_2 \rangle & \cdots & \langle p_1, p_m \rangle \\ \langle p_2, p_1 \rangle & \langle p_2, p_2 \rangle & \langle p_2, p_3 \rangle & \cdots & \langle p_2, p_m \rangle \\ \langle p_3, p_1 \rangle & \langle p_3, p_2 \rangle & \langle p_3, p_3 \rangle & \cdots & \langle p_3, p_m \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle p_m, p_1 \rangle & \langle p_m, p_2 \rangle & \langle p_m, p_3 \rangle & \cdots & \langle p_m, p_m \rangle \end{bmatrix}, p = \begin{bmatrix} \langle x, p_1 \rangle \\ \langle x, p_2 \rangle \\ \langle x, p_3 \rangle \\ \vdots \\ \langle x, p_n \rangle \end{bmatrix}$$

that is, $G_{i,j} = \left\langle p_i, p_j \right\rangle$, $c = [c_1, c_2, ..., c_m]^T$ and $p = [\langle x, p_1 \rangle, \langle x, p_2 \rangle, ..., \langle x, p_m \rangle]^T$.
*The matrix G is known as **Gram** matrix of the set T.*

- *If the element x is in T, then $\hat{x} = x$.*

Proof. □

Remark 5.4.1 (existence of minimizer). In a Hilbert space(also a normed linear space), any finite-dimensional subspace is closed (Theorem 5.2.1].This theorem and the projection theorem (Theorem 5.3.5] guarantees the existence of solution.

Theorem 5.4.2 (invertible condition for Gram matrix). *The Gram matrix G is always positive-semidefinite. It is positive-definite if and only if the vectors $p_1, p_2, ..., p_m$ are linearly independent.*

Proof. R can be written as $G = J^T J$, where J is matrix formed by the columns of p_is. Then, for any nonzero $x \in H$,

$$x^T G x = x^T J^T J x = \|Jx\|^2 \geq 0.$$

If J is nonsingular, then $Jx \neq 0$ and the equality sign will not hold. □

Theorem 5.4.3 (condition of finality). *[12, p. 54] Suppose we are given a linearly independent element $u_1, u_2, u_3, ...$ in a Hilbert space (a function space) V, and we want to minimize the value of*

$$\|f - \alpha_1 u_1 - \alpha_2 u_2 - ... - \alpha_n u_n\|^2$$

over the coefficients α. Then a sequential optimization and all-at-once optimization will give the same result ***if and only if the sequence is orthogonal****.*

Proof. directly from uniqueness of optimal coefficients(due to the full rank nature of Gram matrix) □

.

5.4.2 Application examples

5.4.2.1 *Orthogonal projection and normal equations in $\mathbb{R}^n$*

Theorem 5.4.4. *Let H be the finite dimension Hilbert space $\mathbb{R}^n$, let y be a vector in H and let S be a closed subspace of H and $\dim(S) = k$. Further let U be the matrix with columns consisting of the k linearly independent basis of S. It follows that*

- *The unique solution to the minimization problem:*

$$\min_x \|y - Ux\|_2^2, x \in \mathbb{R}^k, Ux \in S$$

should satisfy the orthogonality condition

$$\langle U_i, y - Ux^* \rangle = 0, \forall i = 1, 2, ..., k.$$

The orthogonality condition can be summarized by

$$U^T U x^* = U^T y,$$

which is known as ***normal equation****, with solution*

$$x^* = (U^T U)^{-1} U^T y.$$

- *(single coefficient) the coefficient associated with the subspace spanned by column vector of U_i is*

$$x_i^* = \frac{U_i^T N_{-i}}{U_i^T N_{-i} U_i} y,$$

where $N_{-i} = (I - U_{-i}(U_{-i}^T U_{-i}]^{-1} U_{-i}^T] U_i$,$U_{-i}$is the matrix without column i.
- *From the perspective of projection, we can view Ux as the vector lying in the subspace S and $Ux = Py = U(U^T U)^{-1} U^T y$ where $P = U(U^T U)^{-1} U^T$ is the projection matrix. Moreover,*
  - *if U consists of orthonormal basis, then $P = UU^T$.*
  - *P is orthonormal projection operator since $P = P^T$.*

Proof. (1) (optimization method) The solution to the minimization can be obtained by expanding the norm and take first derivative to be zero. Note that this optimization problem is strictly convex and therefore the solution is unique.

$$\begin{aligned}&\langle Ux^*, y - x^* \rangle \\ &= (U(U^T U)^{-1} U^T y)^T (I - U(U^T U)^{-1} U^T) y = y^T (U(U^T U)^{-1} U^T - U(U^T U)^{-1} U^T) y\end{aligned}$$

We can alternatively use Hilbert space subspace projection theorem (Theorem 5.4.1] to prove. Note that the system equation is given by

$$Gc = Uy, x = Ux,$$

where G is the Gram matrix given by $G = U^T U$.

(2) Note that we can decompose our orthogonal projector into two orthogonal projectors that are orthogonal to each other (Theorem 4.5.8]

$$U(U^T U)^{-1} U^T = U_{-i}(U_{-i}^T U_{-i}]^{-1} U_{-i}^T + \frac{N_{-i} U_i U_i^T N_{-i}^T}{U_i^T N_{-i} U_i}.$$

The coefficient associated with vector U_i must be the same as the coefficient associated with $N_{-i}U_i$, which is orthogonal to the rest of the subspace U_{-i}. Therefore, the coefficient can be calculated via projection of y onto $N_{-i}U_i$,i.e.,

$$\frac{U_i^T N_{-i}}{U_i^T N_{-i} U_i} y$$

Note that the best approximate $\hat{y}$ can be written by

$$y = x_i U_i + x_{-i} U_{-i} = x_i N_{-i} U_i + x_i P_{-i} U_i + x_{-i} U_{-i},$$

Note that the subspace $N_{-i}U_i$ is orthogonal to the subspace of $[U_{-i}, P_{-i}U_i]$. Therefore, the coefficient of x_i is uniquely determined by projecting onto $N_{-i}U_i$. (3) since $P^2 = P$, P is a projection. (4) notice that $U^T U = I$. (5) use the theorem above that every orthogonal projection is self-adjoint. □

5.4.2.2 *Approximation by continuous polynomials*

Corollary 5.4.4.1 (approximation via polynomial). *Consider we want to approximate a function $f(x)$ in $C[a,b]$ using basis function $\{1, t, t^2, ..., t^{m-1}\}$. The Grammian matrix is given as*

$$R_{i,j} = \int_a^b t^{i+j} dt$$

*If $[a,b] = [0,1]$, R is known as **Hilbert matrix**. Let $p \in \mathbb{R}^m$,*

$$p_i = \int_a^b f(x) t^{i-1} dt, i = 1, 2, ..., m,$$

and solve $c = R^{-1}p$. Then we get the approximation for $f(x)$ given by

$$\hat{f}(x) = \sum_{i=1}^{m} c_i t^{i-1}.$$

Remark 5.4.2 (the Hilbert matrix is ill-conditioned). Note that the Hilbert matrix is ill-conditioned for large m, therefore it is not a good choice for approximation.

Lemma 5.4.1 (approximate discrete points using polynomial). *[9] Consider we have n training data $\{(x_i, y_i)\}, x_i \in \mathbb{R}^n, y_i \in \mathbb{R}$, then we want to approximate the vector $y = [y_1, y_2, ..., y_n]^T$ by using polynomial basis $\{1, t, t^2, ...\}$ sampled at values $x_1, x_2, ...x_n$ as our basis set $\{p_1, p_2, ...\}, p_i \in \mathbb{R}^n$.*

It follows that

- *if we have n polynomials given by $\{1, t, t^2, ..., t^{n-1}\}$, then denote $P = [p_1, ...p_n]$, and we have*
$$y = Pc \implies c = P^{-1}y.$$
That is, the n points $\{x_i, y_i\}$ can be exactly passed through by the function defined by
$$f(x) = \sum_{i=1}^{n} c_i x^{n-1}.$$
- *if we have $m < n$ polynomials given by $\{1, t, t^2, ..., t^{m-1}\}$, then denote $P = [p_1, ...p_m]$, and we have the least-square best approximation for y given by*
$$\hat{y} = Pc \implies c = (P^T P)^{-1} P^T y.$$

Proof. Note that the columns in P are linearly independent (Lemma 4.2.3]. In both cases, we have (Theorem 5.4.4]
$$c = (P^T P)^{-1} P^T y.$$
When $P \in \mathbb{R}^{n \times n}$,
$$(P^T P)^{-1} P^T = P^{-1} P^{-T} P^T = P^{-1}.$$
□

Remark 5.4.3 (connection to Lagrange polynomial approximation finite data set). This lemma achieve the same conclusion as in Lemma 3.5.2 that there exists a polynomial of degree $k-1$ passing k points.

5.4.2.3 *Legendre polynomial via Gram-Schmidt process*

The set of functions $\{1, t, t^2, ...\}$ defined on $[-1, 1]$ forms a linearly independent set. Let the inner product be
$$\langle f, g \rangle = \int_{-1}^{1} fg dt$$
By Gram-Schmidt process, we will get the Legendre polynomial.

5.5 Orthonormal systems

5.5.1 Basic definitions

Definition 5.5.1 (Orthogonal/orthonormal set). *[7, p. 101] Let V denotes an inner product space, a set S of non-zero elements $p_1, p_2, \ldots$ that are orthogonal/orthonormal to each other are called orthogonal/orthonormal system/set.*

Theorem 5.5.1. *Orthonormal/orthogonal systems in an inner product space are linearly independent.*

Definition 5.5.2 (orthonormal sequence). *[7, p. 101] A sequence of vectors in an inner product space that are also orthonormal system is call orthonormal sequence.*

5.5.2 Gram-Schmidt process

Given a set of linear independent vectors $T = \{p_1, p_2, \ldots, p_n\}$, we want to find a set of vectors $T' = \{q_1, q_2, \ldots, q_n\}$ such that

$$span\{q_1, q_2, \ldots, q_n\} = span\{p_1, p_2, \ldots p_n\}$$

We can obtained T' using the following process:

1. $q_1 = p_1 / \|p_1\|$
2. $q_i = p_i - \sum_{j=1}^{i} \left\langle p_i, q_j \right\rangle p_i, q_i = q_i / \|q_i\|, i = 2, 3., n$

5.5.3 Properties of orthonormal systems

Theorem 5.5.2 (Pythagorean formula). *If $x_1, x_2, \ldots, x_n$ are orthogonal vectors in an inner product space, then*

$$\left\| \sum_{i=1}^{n} x_k \right\|^2 = \sum_{k=1}^{n} \|x_k\|^2$$

Proof: directly expand the left hand side and use orthonormal properties.

Theorem 5.5.3 (Bessels' inequality). *[8, p. 305][7, p. 106] Given an orthonormal system and an element x in an inner product space S, we have*

$$\|x\|^2 \geq \sum_{i=1}^{n} \langle x, x_i \rangle^2$$

$$\|x\|^2 \geq \sum_{i=1}^{\infty} \langle x, x_i \rangle^2$$

and the series $\sum_{i=1}^{\infty} \langle x, x_i \rangle^2$ converge.

Proof. Let $c_i = \langle x, x_i \rangle$. Expand

$$\begin{aligned} 0 \leq \left\| x - \sum_i^{\infty} c_i x_i \right\|^2 &= \left\langle x - \sum_i^{\infty} c_i x_i, x - \sum_i^{\infty} c_i x_i \right\rangle \\ &= \langle x, x \rangle - 2\left\langle x, \sum_i^{\infty} c_i x_i \right\rangle + \left\langle \sum_i^{\infty} c_i x_i, \sum_i^{\infty} c_i x_i \right\rangle \\ &= \langle x, x \rangle - 2\sum_i^{\infty} c_i \langle x, x_i \rangle + \sum_i^{\infty} c_i^2 \\ &= \langle x, x \rangle - \sum_i^{\infty} \langle x, x_i \rangle^2 \end{aligned}$$

The convergence can directly from bounded monotone sequence. ☐

Remark 5.5.1. $\sum_{i=1}^{\infty} \langle x, x_i \rangle^2$ **might not converge to** $\|x\|^2$ **unless the orthornomal system is complete**.

Remark 5.5.2. One important application of Bessel's inequality is to prove weak convergence or orthonormal sequence and Riemann-Lebesgue lemma.

Corollary 5.5.3.1. *[7, p. 185] Orthonormal sequences $\{x_n\}$ in inner product space E will weakly converge to 0. Moreover,* ***the convergence cannot be strong convergence****.*

Proof. Because of Bessel's inequality, we have

$$\sum_{i=1}^{\infty} \langle x, x_i \rangle^2 < \infty$$

which implies that $\langle x, x_i \rangle^2 \to 0$ as $i \to \infty$. **The convergence is not strong since** $\|x_i\|$ **= 1**. ☐

Theorem 5.5.4 (Riemann-Lebesgue Lemma for Fourier series). *[8, p. 306] Let $f \in \mathcal{R}[a, a+2\pi]$(Riemann integrable), then*

$$\lim_{n\to\infty} \int_a^{a+2\pi} f(t) \sin ntdt = 0 \lim_{n\to\infty} = \int_a^{a+2\pi} f(t) \cos ntdt$$

Theorem 5.5.5 (general cases). *[8, p. 306] For any orthonormal sequence $\phi_1, \phi_2, \phi_3, ...$ and any f in a Hilbert space V,*

$$\lim_{k\to\infty} \langle f, \phi_k \rangle = 0$$

Proof. directly from the convergence of $\sum_{k=1}^{\infty} \langle f, \phi_k \rangle^2$ due to Bessel's inequality. □

5.5.4 Orthonormal expansion in Hilbert space

Note:

- In finite n dimensional inner product space, given an orthonormal basis $\{e_1, ..., e_n\}$, every element can be written as

$$v = \langle v, e_1 \rangle e_1 + ... + \langle v, e_n \rangle e_n$$

- In infinite dimensional Hilbert space, given an **orthonormal sequence**, we have a series $s = \langle v, e_1 \rangle e_1 + ...$ We want to know (1) the condition for its convergence; and (2) whether it will converge to v(completeness property).

Theorem 5.5.6 (convergence condition of orthonormal series). *Let $\{x_n\}$ be an orthonormal sequence in a Hilbert space H and let $\{a_i\}$ be a sequence of complex numbers. Then the series $\sum_{i=1}^{\infty} a_n x_n$ converges if and only if $\sum_{i=1}^{\infty} a_i x_i < \infty$, and in that case*

$$\sum_{i=1}^{\infty} |a_n|^2 = \sum_{i=1}^{\infty} a_n x_n$$

Proof. (1) (forward part)Define $s_m = \sum_{i=1}^{m} a_n x_n$, we know that s_m will converge if s_m is a Cauchy sequence(since the **Hilbert space is complete**]. And

$$\|s_m - s_k\|^2 = \left\| \sum_{i=k}^{m} a_i x_i \right\| = \sum_{i=k}^{m} |a_n|^2$$

from Pythagorean formula. Note that the right-hand side if the Cauchy sequence of a convergent series $\sum_{i=1}^{\infty} |a_n|^2$; that is $\|s_m - s_k\|^2$ can be made arbitrarily small by choosing N and requiring $k > m > N$. Therefore s_m will converge. (2) (converse part) We define $p_m = \sum_{i=1}^{m} |a_n|^2$, we know that p_m will converge if p_m is Cauchy sequence(since the real line is complete). And

$$|s_m - s_k|^2 = \left\| \sum_{i=k}^{m} a_i x_i \right\| = \sum_{i=k}^{m} |a_n|^2$$

Note that the middle part can be made arbitrarily small by choosing N and requiring $k > m > N$. Therefore p_m will converge. □

Corollary 5.5.6.1 (convergence condition of orthonormal expansion). *Let $\{x_n\}$ be an orthonormal sequence in a Hilbert space H, let $x \in H$, then the orthonormal expansion*

$$\sum_{i=1}^{\infty} \langle x, x_n \rangle x_n$$

will converge.

Proof: directly from above theorem and Bessel's inequality.

5.5.5 Complete orthonormal system

Definition 5.5.3 (completeness). *[9, p. 187][7, p. 109]An orthonormal set $\{p_1, p_2, ...\}$ in a Hilbert space S is complete if*

$$x = \sum_{i=1}^{\infty} \langle x, p_i \rangle p_i$$

for every $x \in S$

Remark 5.5.3 (meaning of equality). It is important to note that here the equality means

$$\lim_{n\to\infty} \left\| x - \sum_{k=1}^{n} \langle x, x_k \rangle x_k \right\|$$

which **usually does not imply pointwise convergence in $L^p, 1 \leq p < \infty$ norm**.(unless ∞ norm is used)

An 'incomplete' orthonormal system
Consider the orthonormal system of $\{\sin(nx)/\sqrt{\pi}, n = 1, 2, ...\}$. To approximate the function $\cos(x)$, we have the series

$$\sum_{n=1}^{\infty} \langle \cos(x), \sin(nx) \rangle \sin(nx)$$

This series will converge(from above theorem), but will converge to 0, since $\langle \cos(x), \sin(nx) \rangle = 0$.

We can interpret an 'incomplete' orthonormal system as a linearly independent set that cannot span the vector space even if it has infinitely many terms.

Definition 5.5.4 (orthonormal basis). ***orthonormal basis*** *An orthonormal system B in an inner product space E is called an orthonormal basis if every $x \in E$ has a* ***unique*** *representation*

$$x = \sum_{n=1}^{\infty} a_n x_n, a_i \in \mathbb{C}, x_i \in B$$

Theorem 5.5.7 (uniqueness of representation). *Every complete orthonormal sequence in an inner product space E is an orthonormal basis; that is, it can uniquely represent elements in E.*

Proof. Let $x = \sum_i a_i x_i = \sum_i b_i x_i$, we have

$$0 = \sum_i (a_i - b_i) x_i \Rightarrow 0 = \left\| \sum_i (a_i - b_i) x_i \right\|^2 = \sum_i |a_i - b_i|^2$$

where we use Pythagorean formula. □

Theorem 5.5.8 (completeness criterion). *[9, p. 187][7, pp. 109–111]A set of orthonormal functions $\{p_i, i = 1, 2, ...\}$ is complete in a Hilbert space S with induced norm if any of the following equivalent statements holds:*

1. $\forall x \in S$,

$$x = \sum_{i=1}^{\infty} \langle x, p_i \rangle p_i$$

2. *for any* $\epsilon > 0$*, there is an* $N < \infty$ *such that* $\forall n \geq N$*:*

$$\left\| x - \sum_{i=1}^{N} \langle x, p_i \rangle p_i \right\| < \epsilon$$

3. *Parseval's equality holds:*

$$\|x\|^2 = \sum_{i=1}^{\infty} \langle x, p_i \rangle^2$$

4. *if* $\langle x, p_i \rangle = 0$ *for all* i*, then* $x = 0$*.*
5. *There is no nonzero function* $f \in S$ *for which the set* $\{p_i, i = 1, 2, ...\} \cup f$ *forms an orthogonal set.*

Remark 5.5.4.

- Note that in the infinite dimension Hilbert space, not every orthonormal function sequence is complete. **This is one fundamental difference compared to finite dimensional space:** every orthonormal basis set of size n can span the space. In infinite dimensional space, however, even thought two orthonormal set are both of size ∞, they are not equal in their express power.
- Note that if we want to use the completeness property to do approximation, then this approximating is in the mean square sense instead of pointwise.

5.5.5.1 *Weierstrass approximation theorem for polynomials*

Theorem 5.5.9. *[8, p. 321] Let* f *be continuous on* $[a, b]$ *and let* $\epsilon > 0$*. Then there exists a polynomial* p *such that*

$$|f(x) - p(x)| < \epsilon$$

Remark 5.5.5. This theorem also suggests that polynomials are complete on the Hilbert space of continuous functions on $[a, b]$.

5.5.5.2 *Examples of complete orthonormal function set*

For complete description and other exotic function set(Chebyshev polynomials, wavelets...)[9, p. 187][7, p. 112]

- Fourier Series:

$$p_n(x) = \frac{1}{\sqrt{2\pi}} e^{int}$$

- Discrete Fourier transform: In the vector space $\mathbb{R}^n$, each element is given by $x[i], i = 0, 1, ..., N-1$
$$p_k[t] = \frac{1}{\sqrt{N}} e^{i2\pi k/N}$$
- Legendre polynomials: For vector space $L^2[-1, 1]$,
$$p_0(t) = 1, p_1(t) = 1, p_2(t) = t^2 - 1/3...$$

5.6 Theory for trigonometric Fourier Series

5.6.1 Basic definitions

Definition 5.6.1 (Fourier series of a function). *A trigonometric series*

$$\frac{1}{2}a_0 + a_1\cos(x) + b_1\sin(x) + \ldots$$

in which

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx, n = 0, 1, 2, \ldots$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx, n = 1, 2, \ldots$$

is called the Fourier series of $f(x)$.

Remark 5.6.1. Here we define what a Fourier series of a function is; the existence of such Fourier series only require the existence of above definite integral to calculate a_n, b_n.

Remark 5.6.2. Given the existence of the Fourier series, another question we are concerned with is whether such Fourier series will converge to $f(x)$.

Theorem 5.6.1. *[13, p. 470]Every uniformly convergent trigonometric series is a Fourier series; Let $f(x)$ be the limit of the trigonometric series, then $f(x)$ is continuous for all x, with period of 2π.*

Proof: By the orthogonal property of trigonometric property, we can verify that a_n, b_n are met by the definitions. Also because trigonometric function is continuous and has period 2π, then the limit function $f(x)$ is continuous and has period 2π since uniform convergence will preserve continuity.

Corollary 5.6.1.1. *If two trigonometric series converge uniformly for all x and have the same sum for all x, then the two series are equal in their coefficients.*

Proof: Let $f(x)$ be the limits, the coefficients are uniquely determined by the definite integral.

Remark 5.6.3. This corollary answer the question of uniqueness of Fourier series.

Theorem 5.6.2 (fundamental theorem, local convergence). *[13, p. 472] Let $f(x)$ be piece C^2 in the interval $-\pi \leq x \leq \pi$. Then the Fourier series of $f(x)$:*

$$\frac{1}{2}a_0 + a_1\cos(x) + b_1\sin(x) + ...$$

in which

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx, n = 0, 1, 2, ...$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx, n = 1, 2, ...$$

converges uniformly *to $f(x)$ whenever $f(x)$ is continuous inside the interval. The series converge to*

$$\frac{1}{2}[\lim_{x\to x_1-} f(x) + \lim_{x\to x_1+} f(x)]$$

at the point of discontinuity x_1 inside the interval, and to

$$\frac{1}{2}[\lim_{x\to \pi-} f(x) + \lim_{x\to \pi+} f(x)]$$

at $x = \pm\pi$

Remark 5.6.4. The most important aspect of this theorem is that for a piece-wise C^2 function, its Fourier series will converge uniformly to it.

Remark 5.6.5 (uniform convergence vs mean-square convergence).

- uniform convergence is a much stronger convergence than the mean-square convergence; We usualy call uniform convergence as local convergence, and mean-square convergence as global convergence; [14, p. 69]
- For approximation applications, we usually only concerns mean-square error, i.e., mean-square convergence;
- The completeness concept is defined based on mean-square convergence rather than local convergence.

Remark 5.6.6. [14, p. 83] **Continuity cannot guarantee the convergence of Fourier series**, except that other conditions are allowed

Lemma 5.6.1. *[14, p. 41] Suppose that f is continuous function on $[-\pi, \pi]$, and the Fourier coefficients satisfy $\sum_{i=1}^{\infty} a_i^2 + b_i^2 < \infty$, i.e., convergent. Then the Fourier series converges uniformly to f.*

Proof. the continuity guarantees the integrability. And then use M-test for uniformly convergence. □

Remark 5.6.7 (sufficient condition for pointwise convergence). [15] In mathematics, the Dirichlet conditions are sufficient conditions for a real-valued, periodic function $f(x)$ to be equal to the sum of its Fourier series at each point where f is continuous. The conditions are:

1. f(x) must be absolutely integrable over a period.
2. f(x) must have a finite number of extrema in any given bounded interval, i.e. there must be a finite number of maxima and minima in the interval.
3. f(x) must have a finite number of discontinuities in any given bounded interval, however the discontinuity cannot be infinite.

These three conditions are satisfied if f is a function of bounded variation over a period.

For a piece-wise smooth and periodic function f defined on the interval $[-\pi, \pi]$, then the Fourier series converges to $(f(x+) + f(x-))/2$.

5.6.2 Completeness of Fourier series

Remark 5.6.8 (completeness). The completeness theorem says that a continuous periodic function f equals its Fourier series. This means that the set of function $cos(nt), sin(nt)$ form a complete set of basis function, and you do not need any more functions to express any periodic functions as a linear combination.To prove this, we need to show the fourier series f_1 of any function f are equal at any point in the domain(we need to construct delta function to prove this).

Corollary 5.6.2.1. *The trigonometric set is complete in $CP[a, a + 2\pi]$ (the function space of continuous 2π periodic functions)[8]*

Corollary 5.6.2.2 (Fourier cosine series and sine series). *The Fourier cosine and sins series are complete for even and odd piecewise continuous functions defined on $[-\pi, \pi]$ respectively.*

Lemma 5.6.2 (best approximations). *[14, p. 78] Let S_N be the partial sum of the Fourier series of $f(x)$(assuming that the Fourier coefficient in S_N exists), then*

$$\int_{-\pi}^{\pi} \|f(x) - S_N(x)\| \, dx \leq \int_{-\pi}^{\pi} \left\| f(x) - \sum_{k=0}^{N} c_k \cos(kx) + d_k \sin(kx) \right\| dx$$

for any c_i and d_i.

Remark 5.6.9. Note that we did not impose any condition on f except that it is Riemann integrable when calculating the Fourier coefficients.

5.6.3 Complex representation

Lemma 5.6.3 (complex representation of Fourier series). *[16, p. 271] Consider a real-value Fourier series in the interval $[-\pi, \pi]$ given by*

$$f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \sin(kx).$$

It can also be represented by the following complex form:

$$f(x) = \sum_{k=-\infty}^{\infty} c_k \exp(ikx),$$

where

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \exp(-ikx) dx.$$

- $c_0 = \frac{1}{2}a_0$
- *for* $k \neq 0$,

$$c_k = a_k + ib_k, c_{-k} = a_k - ib_k.$$

- $c_k = \overline{c_{-k}}$.

Proof. Use

$$cos(kx) = \frac{e^{ikx} + e^{-ikx}}{2}, sin(kx) = \frac{e^{ikx} - e^{-ikx}}{2i}.$$

The match of $cos(kx)$ coefficients requires

$$(\frac{c_k}{2} + \frac{c_{-k}}{2}) = a_k.$$

The match of $sin(kx)$ coefficients requires

$$(\frac{c_k}{2} - \frac{c_{-k}}{2}) = ib_k.$$

We can solve

$$c_k = a_k + ib_k, c_{-k} = a_k - ib_k.$$

The rest is straight forward. □

5.7 Fourier transform

5.7.1 Definitions and basic concepts

Definition 5.7.1 (Fourier transform). *[16, p. 267] The Fourier transform of the function $f \in L^1(\mathbb{R})$ is given as*

$$\mathcal{F}[f] = \tilde{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} f(x) dx$$

and the inverse Fourier transform

$$\mathcal{F}^{-1}[\tilde{f}] = f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx} \tilde{f}(k) dk.$$

Usually we denote $\overline{f} = F$.

Remark 5.7.1.

- Because $\left|e^{ikx} f(x)\right| \leq |f(x)|$, limit the domain to $L^1(\mathbb{R})$ guarantees the existence of Fourier transform.
- [7, p. 194] Not all functions in $L^2(\mathbb{R})$ has the existence of Fourier transform.

Definition 5.7.2 (Fourier transform basis). *The Fourier transform basis function with frequency $k \in (-\infty, \infty)$ is defined as*

$$e_k(x) = 1/\sqrt{2\pi} \exp(ikx)$$

where $x \in (-\infty, \infty)$.

Define $\langle f, g \rangle = \int_{-\infty}^{\infty} f\overline{g} dx$, we can redefine Fourier transform as

$$\mathcal{F}[f](k) = \langle e_k, f \rangle, f(x) = \left\langle e_k, \langle e_k, f \rangle \right\rangle.$$

Lemma 5.7.1 (unitary property of basis function). *[16, p. 268]The basis function of the Fourier transform satisfies:*

$$\langle e_k, e_{k'} \rangle = \int_{-\infty}^{+\infty} e^{ikx} \overline{e^{ik'x}} dx = 2\pi\delta(k - k')$$

Some equivalent forms are:

$$\int_{-\infty}^{+\infty} e^{ikx} dx = 2\pi\delta(k)$$

or

$$\int_{-\infty}^{+\infty} e^{i(k-k')x} dx = 2\pi\delta(k-d')$$

Proof.

$$\lim_{k\to\infty} \frac{1}{2\pi} \int_{-k}^{k} \exp[ik(\eta - x)]dk = \lim_{k\to\infty} \frac{\sin K(\eta - x)}{\pi(\eta - x)} = \delta(\eta - x)$$

here we use the definition of δ function. □

Remark 5.7.2 (real inner product vs unitary inner product)**.** Note that the basis is unitary instead of orthonormal: $\int_{-\infty}^{+\infty} e^{ik_1x}e^{ik_2x}dx = 2\pi\delta(k_1 - (-k_2)) = 2\pi\delta(k_1 + k_2) \neq 2\pi\delta(k_1 - k_2)$,i.e., they are not orthogonal(in the sense of real inner product) for different k.

Remark 5.7.3. This property is the key of inverse Fourier transform to hold:

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-ikx} F(k)dk \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-ikx} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x')e^{ikx'}dx'dk \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ik(x-x')}dx'dk = f(x) \end{aligned}$$

Corollary 5.7.0.1 (Fourier transform pair of 1 and δ). *We have*

$$\int_{-\infty}^{\infty} 1e^{ikx}dx = 2\pi\delta(k)$$

and

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} 2\pi\delta(k)e^{-ikx}dk = 1$$

Lemma 5.7.2 (basic properties). *[16, p. 269] Suppose the Fourier transform of f exists, then we have*

- $\mathcal{F}^2[f] = \mathcal{F}[\mathcal{F}[f]] = f(-x)$
- $\mathcal{F}^{-1}[f] = \tilde{f}(-k)$

- *For a real function, $\overline{\tilde{f}}(k) = \tilde{f}(-k)$.*

Proof. (3)

$$\overline{\tilde{f}(k)} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \overline{e^{ikx} f(x)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx} f(x) dx = \tilde{f}(-k)$$

□

Lemma 5.7.3 (Parseval equality). *[16, p. 269] [16, p. 269]Let f be an integrable function, then*

$$\|f\|_2^2 = \left\|\tilde{f}\right\|_2^2$$

more generally,

$$\langle f, g \rangle = \left\langle \tilde{f}, \tilde{g} \right\rangle$$

Proof. (informal)

$$\begin{aligned}
\|f\|^2 &= \int_{-\infty}^{\infty} f(x) \overline{f}(x) dx \\
\|f\|^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-ikx} F(k) dk \int_{-\infty}^{\infty} e^{ikx} \overline{F}(k') dk' dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(k) dk \int_{-\infty}^{\infty} \overline{F}(k') 2\pi (k - k') dk' \\
&= \int_{-\infty}^{\infty} F(k) \overline{F}(k') dk = \|F\|^2
\end{aligned}$$

□

5.7.2 Convolution theorem

Definition 5.7.3. *The convolution of f and g is defined as*

$$f * g(z) = \int_{-\infty}^{\infty} f(x) g(z - x) dx = \int_{-\infty}^{\infty} f(z - x) g(x) dx.$$

Remark 5.7.4 (equivalence of two forms).

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)g(z-x)dx &= \int_{\infty}^{-\infty} f(z-y)g(y)d(z-y) \\ &= -\int_{\infty}^{-\infty} f(z-y)g(y)d(y) \\ &= \int_{-\infty}^{\infty} f(z-y)g(y)dy \\ &= \int_{-\infty}^{\infty} f(z-x)g(x)dx\end{aligned}$$

Lemma 5.7.4 (convolution theorem). *[16, p. 271]* $\mathcal{F}[f*g] = \sqrt{2\pi}\tilde{f}\tilde{g}$ *and* $\mathcal{F}^{-1}[f*g] = 1/\sqrt{2\pi}(f*g)(x)$

Proof.

$$\begin{aligned}f*g(z) &= \int_{-\infty}^{\infty} f(x)g(z-x)dx \\ \implies \mathcal{F}[f*g(z)] &= \int_{-\infty}^{\infty} e^{ikz}\int_{-\infty}^{\infty} f(x)g(z-x)dxdz \\ &= \int_{-\infty}^{\infty} e^{ikx}\int_{-\infty}^{\infty} f(x)e^{ik(z-x)}g(z-x)dxdz \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ikx}\int_{-\infty}^{\infty} f(x)e^{iky}g(y)dxdy \\ &= \sqrt{2\pi}\mathcal{F}[f]\mathcal{F}[g]\end{aligned}$$

□

5.7.3 Fourier transform and Fourier series

Lemma 5.7.5. *[16, p. 271] Consider a real-value function in* $\mathbb{R}$ *given by*

$$f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k\cos(kx) + b_k\sin(kx) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}.$$

It follows that

- *the Fourier series of $f(x)$* ***on the interval*** $[-\pi, \pi]$ *is*

$$\sum_{k=-\infty}^{\infty} c_k e^{ikx},$$

where

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\exp(-ikx)dx.$$

- *the Fourier transform of $f(x)$ is given by*

$$\tilde{f}(k) \triangleq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ikx} f(x)dx = \sqrt{2\pi}\sum_{k'=-\infty}^{\infty} c_{k'}\delta(k+k').$$

Proof. (1)

$$\sum_{k'=-\infty}^{\infty} c_k e^{ikx} = \sum_{k'=-\infty}^{\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\exp(-ik'x)\exp(ikx)dx = f(x).$$

(2)

$$\tilde{f}(k) \triangleq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ikx} f(x)dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ikx}\sum_{k=-\infty}^{\infty} c_k e^{ikx}dx = \sqrt{2\pi}\sum_{k'=-\infty}^{\infty} c_{k'}\delta(k+k').$$

□

Remark 5.7.5.

- Note that Fourier series only applies to periodic function defined on finite intervals.
- Fourier transform can apply to function defined on $\mathbb{R}$ as long as the integral exists.
- For a periodic function, the Fourier transform will generate spikes.

5.7.4 Discrete Fourier transform

5.7.4.1 *Properties*

Definition 5.7.4. *Given an positive integer N, let $w = \exp(2\pi i/N)/\sqrt{N}$, a Fourier transform matrix F is given by*

$$F_{jk} = w^{jk}, j = 0, 1, ..., N-1; k = 0, 1, ..., N-1$$

Lemma 5.7.6. *Basic properties of Fourier transform matrix:*

- $w^{jN} = 1/\sqrt{N} = w^0, j = 0, 1, ..., N-1$
- *F is symmetric*
- $\sum_{j=0}^{N-1} w^{mj} = \delta(m)\sqrt{N}$
- *columns of F form an orthonormal basis, i.e., each column has unit 1 and different columns are orthogonal.*
- *F is symmetric.*
- $w^{N+k} = w^k, w^{N/2+k} = -w^k$.

Proof. (1) and (2) are easy. (3) use

$$(w^0 + w^1 + ... + w^{N-1})(1-w) = 1 - w^N = 1 - 1 = 0$$

if $w \neq 1$, then $w^0 + w^1 + ... + w^{N-1} = 0$ We now prove (4). Unit length:

$$1/\sqrt{N}^2 \sum_{j=0}^{N-1} w^{mj}\overline{w^{mj}} = 1/\sqrt{N}^2 N = 1$$

where we use $\left|w^{jm}\right| = 1$. Orthonormal:

$$1/\sqrt{N}^2 \sum_{j=0}^{N-1} w^{mj}\overline{w^{nj}} = 1/\sqrt{N}^2 \sum_{j=0}^{N-1} \exp((m-n)ji2\pi/N) = \delta(m-n)$$

where we have use the property in (3). □

Theorem 5.7.1. *The columns of F is a complete orthonormal set in space of complex vectors in* $\mathbb{C}^N$.

Proof. Because $\mathbb{C}^N$ has dimension of N and columns in F is linearly independent and have length N, therefore F is a basis of $\mathbb{C}^N$. □

Lemma 5.7.7. *The conjugate of matrix F has the same property of F*

Proof. this can be directly verified. □

Note: Because F and $\overline{F}$ have the same property above, there is no essential differences in using F or $\overline{F}$ in performing discrete Fourier transform.

Lemma 5.7.8. *Let F be the Fourier transform matrix, let $\overline{F}$ be the conjugate matrix, then*

$$F\overline{F} = FF^H = I$$

That is, F is unitary.

Proof. directly from the property of F (F is symmetric). $\square$

5.8 Notes on bibliography

For functional analysis, see [1][17][18].

For comprehensive treatment on both theory and applications of linear algebra and functional analysis on signal processing, see [9].

For treatment on contraction mapping and fixed point theorem with applications, see [5][19][7].

For linear operator theory, see [20].

For treatment on Hilbert space and application, see [7].

BIBLIOGRAPHY

1. Kreyszig, E. *Introductory functional analysis with applications* (wiley New York, 1989).
2. Christensen, O. *Functions, spaces, and expansions: mathematical tools in physics and engineering* (Springer Science & Business Media, 2010).
3. Luenberger, D. G. *Optimization by vector space methods* (John Wiley & Sons, 1969).
4. Carter, M. *Foundations of mathematical economics* (MIT Press, 2001).
5. Griffel, D. H. *Applied functional analysis* (Courier Corporation, 2002).
6. Bertsekas, D. *Nonlinear programming* ISBN: 9781886529007 (Athena Scientific, 2016).
7. Debnath, L. & Mikusiński, P. *Hilbert spaces with applications* (Academic press, 2005).
8. Johnsonbaugh, R. & Pfaffenberger, W. Foundations of mathematical analysis (2010).
9. Moon, T. K. S. & Wynn, C. *Mathematical methods and algorithms for signal processing* **621.39: 51 MON** (2000).
10. Axler, S. J. *Linear algebra done right* (Springer).
11. Meyer, C. D. *Matrix analysis and applied linear algebra* (Siam, 2000).
12. Davis, H. F. *Fourier series and orthogonal functions* (Courier Corporation, 1989).
13. Kaplan, W. *Advanced calculus* (Addison Wesley Publishing Company, 1973).
14. Stein, E. & Shakarchi, R. *Fourier Analysis: An Introduction* ISBN: 9781400831234 (Princeton University Press, 2011).
15. Wikipedia. *Dirichlet conditions — Wikipedia, The Free Encyclopedia* [Online; accessed 17-May-2016]. 2016.
16. Prosperetti, A. Advanced mathematics for applications. *SIAM Review Vol. 55, Issue 2 (June 2013)* **55,** 403 (2013).
17. Lax, P. *Functional Analysis* ISBN: 9781118626740 (Wiley, 2014).
18. Duren, P. L. *Invitation to classical analysis* (American Mathematical Soc., 2012).
19. Ok, E. A. *Real analysis with economic applications* (Princeton University Press, 2007).
20. Naylor, A. W. & Sell, G. R. *Linear operator theory in engineering and science* (Springer Science & Business Media, 2000).

Part II

MATHEMATICAL OPTIMIZATION METHODS