Lectures on
Approximation Theory
(Math 443)

Lecturer
Husam L. Saad

#

Department of Mathematics, College of Science

Basrah University

Chapter 1

General Approximation Problem

1.1 Normed Linear Spaces

There are several reasons for studying approximation theory and method ranging from a need to represent functions in computer calculations to an interest in the mathematics of the subject (replace a complicated function by one which is simpler and more manageable). Although approximation algorithms are used throughout the science and in many industrial and commercial fields, and to find a simple function which gives a best fit to the experimental data. The problem of approximating a given function or a table of values by a class of simpler functions has been of great interest theoretically and practically. For instance we may approximate the solution of a differential equations by a function of a certain form that depends on adjustable parameters. Here the measure of goodness of the approximation is a scaler quantity that is derived from the residual occurs when the approximating function is substituted into the differential equation (this scaler quantity is called a norm, which is a convenient measure of the "error" in the approximation).

Definition 1.1. *A non empty set $\mathbb{F}$ is called a linear space over a field of real numbers $\mathbb{R}$ if and only if for all $A, B, C \in \mathbb{F}$ and for all real numbers r, s*

(1) $A + B = B + A$.

(2) $A + (B + C) = (A + B) + C$.

(3) There is a unique element 0 in $\mathbb{F}$ such that $A + 0 = A \quad \forall\, A \in \mathbb{F}$.

(4) For each $A \in \mathbb{F}$ there is a unique element $-A \in \mathbb{F}$ such that $A + (-A) = 0$.

(5) $r \cdot (A + B) = r \cdot A + r \cdot B$.

(6) $(r + s) \cdot A = r \cdot A + s \cdot A$.

(7) $(r \cdot s) \cdot A = r \cdot (s \cdot A)$.

(8) $1 \cdot A = A$.

Definition 1.2. *Let $\mathbb{F}$ be a linear space and let $\| \cdot \| : \mathbb{F} \to \mathbb{R}$ such that*

(1) $\|A\| > 0$ unless $A = 0$.

(2) $\|rA\| = |r|\|A\|$ where r is scaler.

(3) $\|A + B\| \leqslant \|A\| + \|B\|$.

Then $\| \cdot \|$ defines a norm on $\mathbb{F}$.

Definition 1.3. *A linear space $\mathbb{F}$ equipped with a norm is called a normed linear space.*

Definition 1.4. *A metric space is a nonempty set M of points together with a function $d : M \times M \to \mathbb{R}$ satisfying the following properties for all x, y and $z \in M$*

(1) $d(x, y) = 0$ if $x = y$.

(2) $d(x, y) > 0$ if $x \neq y$.

(3) $d(x, y) = d(y, x)$.

(4) $d(x,z) \leqslant d(x,y) + d(y,z)$.

Remark 1.1. *In a normed linear space the formula $d(x,y) = \|x-y\|$ defines a metric. i.e., a normed linear space becomes a metric space.*

Proof. H.W. □

Definition 1.5.

$$C[a,b] = \{f : f : [a,b] \to \mathbb{R}, f \text{ is continuous}\}.$$

$$\mathbb{R}^N = \{(x_1, x_2, \dots, x_N) : x_i \in \mathbb{R}, \text{for } i = 1, 2, \dots, N\}.$$

The three norms that are used most frequently are the p-norms, for $p = 1, 2$ and ∞. For finite p the p-norm in $C[a,b]$ is defined as

$$\|f\|_p = \left[\int_a^b |f(x)|^p dx\right]^{\frac{1}{p}} \quad 1 \leqslant p < \infty$$

and the p-norm in $\mathbb{R}^N$ as

$$\|f\|_p = \left[\sum_{i=1}^{N} |f(x_i)|^p\right]^{\frac{1}{p}} \quad 1 \leqslant p < \infty$$

where $f = (f(x_1), f(x_2), \dots, f(x_N))$. For $p = \infty$, the norms become,

$$\|f\|_\infty = \max_{a \leqslant x \leqslant b} |f(x)|$$

and

$$\|f\|_\infty = \max_{1 \leqslant i \leqslant N} |f(x_i)|$$

respectively.

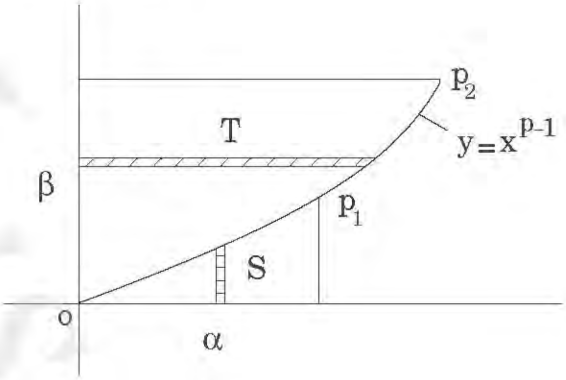
The ∞-norm is called the Chebyshev norm (sometimes called the uniform or minimax norm).

Theorem 1.1. (Holder inequality) *If $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\int_a^b |A(x)B(x)|dx \leqslant \left[\int_a^b |A(x)|^p dx\right]^{\frac{1}{p}} \cdot \left[\int_a^b |B(x)|^q dx\right]^{\frac{1}{q}},$$

where $A, B \in C[a,b]$.

Proof. Let us study the curve op_1p_2.



$$y = x^{p-1} \quad \Rightarrow \quad x = y^{\frac{1}{p-1}}.$$

$$\left.\begin{array}{l} \frac{1}{p}+\frac{1}{q}=1. \quad \Rightarrow \quad \frac{q}{p}+1=q. \quad \Rightarrow \quad \frac{q}{p}=q-1. \\ \frac{1}{p}+\frac{1}{q}=1. \quad \Rightarrow \quad 1+\frac{p}{q}=p. \quad \Rightarrow \quad \frac{q}{p}=\frac{1}{p-1}. \end{array}\right\} \Rightarrow \quad \frac{1}{p-1}=q-1.$$

$$A_S = \int_0^\alpha x^{p-1}dx = \frac{x^p}{p}\Big|_0^\alpha = \frac{\alpha^p}{p}.$$

$$A_T = \int_0^\beta y^{q-1}dy = \frac{y^q}{q}\Big|_0^\beta = \frac{\beta^q}{q}.$$

Note that

$$\begin{aligned} \alpha\beta &\leqslant A_S + A_T \\ &\leqslant \frac{\alpha^p}{p} + \frac{\beta^q}{q}. \end{aligned} \tag{1.1}$$

Assume that $\alpha = \frac{|A(x)|}{\|A\|_p}$, $\beta = \frac{|B(x)|}{\|B\|_q}$ and substitute in (1.1) to get

$$\frac{|A(x)|}{\|A\|_p}\frac{|B(x)|}{\|B\|_q} \leqslant \frac{|A(x)|^p}{p\|A\|_p^p} + \frac{|B(x)|^q}{q\|B\|_q^q}.$$

$$\begin{aligned} \frac{1}{\|A\|_p\|B\|_q}\int_a^b |A(x)B(x)|dx &\leqslant \frac{1}{p\|A\|_p^p}\int_a^b |A(x)|^p dx + \frac{1}{q\|B\|_q^q}\int_a^b |B(x)|^q dx \\ &\leqslant \frac{1}{p\|A\|_p^p}\|A\|_p^p + \frac{1}{q\|B\|_q^q}\|B\|_q^q = \frac{1}{p}+\frac{1}{q} = 1. \end{aligned}$$

Hence

$$\int_a^b |A(x)B(x)|dx \leqslant \|A\|_p\|B\|_q = \left[\int_a^b |A(x)|^p dx\right]^{\frac{1}{p}} \cdot \left[\int_a^b |B(x)|^q dx\right]^{\frac{1}{q}}.$$

□

Remark 1.2. *When $p = q = 2$ the Holder inequality becomes*

$$\int_a^b |A(x)B(x)|dx \leqslant \left[\int_a^b |A(x)|^2 dx\right]^{\frac{1}{2}} \cdot \left[\int_a^b |B(x)|^2 dx\right]^{\frac{1}{2}}.$$

The above inequality called Cauchy-Schwartz inequality.

Theorem 1.2. (Minkowski inequality) *If $p \geqslant 1$ and $A, B \in C[a,b]$, then*

$$\left[\int_a^b \big[|A(x) + B(x)|\big]^p dx\right]^{\frac{1}{p}} \leqslant \left[\int_a^b |A(x)|^p dx\right]^{\frac{1}{p}} + \left[\int_a^b |B(x)|^p dx\right]^{\frac{1}{p}}.$$

Proof.

$$\begin{aligned}\big[|A(x) + B(x)|\big]^p &= \big[|A(x) + B(x)|\big] \cdot \big[|A(x) + B(x)|\big]^{p-1} \\ &\leqslant |A(x)|\big[|A(x) + B(x)|\big]^{p-1} + |B(x)|\big[|A(x) + B(x)|\big]^{p-1}. \qquad (1.2)\end{aligned}$$

Applying Holder inequality to every term on the right hand side of (1.2)

$$\begin{aligned}\int_a^b \big[|A(x) + B(x)|\big]^p dx &\leqslant \|A\|_p \left[\int_a^b \big[|A(x) + B(x)|\big]^{(p-1)q} dx\right]^{\frac{1}{q}} \\ &\quad + \|B\|_p \left[\int_a^b \big[|A(x) + B(x)|\big]^{(p-1)q} dx\right]^{\frac{1}{q}} \\ &\leqslant \left[\int_a^b \big[|A(x) + B(x)|\big]^{(p-1)q} dx\right]^{\frac{1}{q}} \Big[\|A\|_p + \|B\|_p\Big].\end{aligned}$$

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \Rightarrow \quad 1 + \frac{p}{q} = p \quad \Rightarrow \quad \frac{p}{q} = p - 1 \quad \Rightarrow \quad p = (p-1)q.$$

$$\int_a^b \big[|A(x) + B(x)|\big]^p dx \leqslant \left[\int_a^b \big[|A(x) + B(x)|\big]^p dx\right]^{\frac{1}{q}} \Big[\|A\|_p + \|B\|_p\Big].$$

Divide by $\left[\int_a^b \big[|A(x)+B(x)|\big]^p dx\right]^{\frac{1}{q}}$ and use the fact that $1-\frac{1}{q} = \frac{1}{p}$ to get the required result. □

Remark 1.3. *It will be noted that this method yields also the Holder inequality and Minkowski inequality for series. i.e., we have*

(1) *If* $p > 1$ *and* $\frac{1}{p} + \frac{1}{q} = 1$*, then*

$$\sum_{k=1}^{N} |a_k b_k| \leqslant \left[\sum_{k=1}^{N} |a_k|^p\right]^{\frac{1}{p}} \cdot \left[\sum_{k=1}^{N} |b_k|^q\right]^{\frac{1}{q}},$$

where $a_k, b_k \in \mathbb{R}$ *for* $k = 1, 2, \ldots, N$.

(2) *If* $p \geqslant 1$ *and* $a_k, b_k \in \mathbb{R}$ *for* $k = 1, 2, \ldots, N$*, then*

$$\left[\sum_{k=1}^{N} \left[|a_k + b_k|\right]^p\right]^{\frac{1}{p}} \leqslant \left[\sum_{k=1}^{N} |a_k|^p\right]^{\frac{1}{p}} + \left[\sum_{k=1}^{N} |b_k|^p\right]^{\frac{1}{p}}.$$

Proof. H.W. □

Examples of Normed Linear Spaces

Example 1.1. $C[a, b]$ *with the p-norm*

$$\|f\|_p = \left[\int_a^b |f(x)|^p dx\right]^{\frac{1}{p}} \quad 1 \leqslant p < \infty$$

is a normed linear space over a field $\mathbb{R}$ *with respect to operations addition and standard multiplication which is defined as follows:*

(1) $(f + g)(x) = f(x) + g(x)$ *for all* $f, g \in C[a, b]$.

(2) $(r \cdot f)(x) = r \cdot f(x)$ *for all* $r \in \mathbb{R}$ *and for all* $f \in C[a, b]$.

Proof.

i. $C[a, b]$ is a linear space.

(1) $(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x)$.

(2) $(f + (g + h))(x) = f(x) + (g + h)(x) = f(x) + (g(x) + h(x)) = (f(x) + g(x)) + h(x) = (f + g)(x) + h(x) = ((f + g) + h)(x)$.

(3) $(f+O)(x) = f(x). \Rightarrow f(x)+O(x) = f(x). \Rightarrow O(x) = 0 \quad \forall\, x \in [a,b]$.
i.e., the identity element is the function $O : [a,b] \to \mathbb{R}$ which is defined by $O(x) = 0 \quad \forall\, x \in [a,b]$.

(4) $(f + (-f))(x) = O(x). \Rightarrow (-f)(x) = -f(x)$. i.e., the inverse element is the function $-f : [a,b] \to \mathbb{R}$

(5) $(r \cdot (f+g))(x) = r \cdot (f+g)(x) = r \cdot (f(x) + g(x)) = r \cdot f(x) + r \cdot g(x) = (r \cdot f)(x) + (r \cdot g)(x)$.

(6) $((r+s) \cdot f)(x) = (r+s) \cdot f(x) = r \cdot f(x) + s \cdot f(x) = (r \cdot f)(x) + (s \cdot f)(x) = (r \cdot f + s \cdot f)(x)$.

(7) $((r \cdot s) \cdot f)(x) = (r \cdot s) \cdot f(x) = r \cdot (s \cdot f(x)) = r \cdot (s \cdot f)(x) = (r \cdot (s \cdot f))(x)$.

(8) $(1 \cdot f)(x) = 1 \cdot f(x) = f(x)$.

ii. The p-norm, $1 \leqslant p < \infty$, defines a norm on $C[a,b]$.

(1) $\|f\|_p > 0$ unless $f = 0$.

$$\|f\|_p = \left[\int_a^b |f(x)|^p dx\right]^{\frac{1}{p}}.$$

if $f(x) = 0 \Rightarrow \|f\|_p = 0$.
if $f(x) \neq 0 \Rightarrow \|f\|_p > 0$.

(2) $\|rf\|_p = |r|\|f\|_p$ where r is scaler.

$$\|rf\|_p = \left[\int_a^b |rf(x)|^p dx\right]^{\frac{1}{p}} = |r|\left[\int_a^b |f(x)|^p dx\right]^{\frac{1}{p}} = |r|\|f\|_p.$$

(3) By Minkowski inequality we get $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$.

□

Example 1.2. $\mathbb{R}^N$ *with the p-norm*

$$\|f\|_p = \left[\sum_{i=1}^{N} |f(x_i)|^p\right]^{\frac{1}{p}} \quad 1 \leqslant p < \infty$$

is a normed linear space over a field $\mathbb{R}$ *with respect to operations addition and standard multiplication which is defined as follows:*

(1) $(x_1, x_2, \ldots, x_N) + (y_1, y_2, \ldots, y_N) = (x_1 + y_1, x_2 + y_2, \ldots, x_N + y_N)$ *for all* $(x_1, x_2, \ldots, x_N)$, $(y_1, y_2, \ldots, y_N) \in \mathbb{R}^N$.

(2) $r(x_1, x_2, \ldots, x_N) = (rx_1, rx_2, \ldots, rx_N)$ *for all* $r \in \mathbb{R}$ *and for all* $(x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$.

Proof. H.W. □

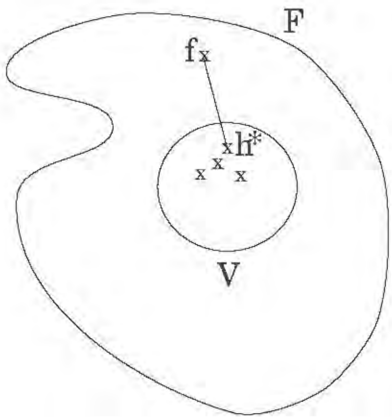
1.2 The Problem of Best Approximation

Let $\mathbb{F}$ be a normed linear space over the field $\mathbb{R}$ and let $\|f\|$ denote the norm of f. Let V be a subset of $\mathbb{F}$, then the general problem of best approximation may be defined in the following terms.

Definition 1.6. *Given a point* f *and a subset* V *in a normed linear space* $\mathbb{F}$. *A best approximation to* f *from* V *is an element* $h^* \in V$ *of minimum distance from* f.
i.e., given $f \in \mathbb{F}$, $f \notin V$, *find* $h^* \in V$ *such that*

$$\|f - h^*\| \leqslant \|f - h\| \ \forall\ h \in V.$$

We call h^* *a best approximation to* f *with respect to* V *and norm* $\|\cdot\|$.



Most of the approximation problems that we consider, and which are of particular interest in practice are of two cases.

(1) Continuous approximation where f and V are in $C[a,b]$.

(2) Discrete approximation where f and V are in $\mathbb{R}^N$.

Remark 1.4. *The Chebyshev norm provides the foundation of much of the approximation theory, the next theorem shows that, if $h \in V$ approximates $f \in \mathbb{F}$ such that $\|E\|_\infty$ is small, where $E = f - h$, then $\|E\|_1$ and $\|E\|_2$ are small too (at least for $b-a$ not too large).*

Theorem 1.3. *For all E in $C[a,b]$ the inequalities*

$$\|E\|_1 \leqslant (b-a)^{\frac{1}{2}}\|E\|_2 \leqslant (b-a)\|E\|_\infty$$

hold.

Proof.

$$\begin{aligned}
\|E\|_1 &= \int_a^b |E(x)|dx = \int_a^b |1||E(x)|dx \\
&\leqslant \left[\int_a^b |1|^2 dx\right]^{\frac{1}{2}} \left[\int_a^b |E(x)|^2 dx\right]^{\frac{1}{2}} \quad \text{(By Cauchy-Schwartz inequality)} \\
&\leqslant (b-a)^{\frac{1}{2}}\|E\|_2.
\end{aligned}$$

Hence

$$\|E\|_1 \leqslant (b-a)^{\frac{1}{2}}\|E\|_2. \tag{1.3}$$

$$|E(x)| \leqslant \max_{a \leqslant x \leqslant b} |E(x)| = \|E\|_\infty.$$

$$\begin{aligned}
\|E\|_2 &= \left[\int_a^b |E(x)|^2 dx\right]^{\frac{1}{2}} \\
&\leqslant \left[\int_a^b \|E\|_\infty^2 dx\right]^{\frac{1}{2}} \\
&\leqslant \|E\|_\infty (b-a)^{\frac{1}{2}}.
\end{aligned}$$

Hence

$$(b-a)^{\frac{1}{2}}\|E\|_2 \leqslant (b-a)\|E\|_\infty. \tag{1.4}$$

from the equations (1.3) and (1.4) we get

$$\|E\|_1 \leqslant (b-a)^{\frac{1}{2}}\|E\|_2 \leqslant (b-a)\|E\|_\infty.$$

□

Remark 1.5. *The converse statement may not be true. i.e., it is not always possible to reduce the* $\|E\|_\infty$ *by making* $\|E\|_1$ *or* $\|E\|_2$ *small, as we see in the following example.*

Example 1.3. *Let* $f(x) = 1$, $h(x) = x^\lambda$, λ *is a positive parameter,* $0 \leqslant x \leqslant 1$.

Solution. $E = f - h = 1 - x^\lambda$.

$$\|E\|_1 = \int_a^b |E(x)|dx = \int_0^1 |1 - x^\lambda| dx.$$

$$0 \leqslant x \leqslant 1 \quad \Rightarrow \quad 0 \leqslant x^\lambda \leqslant 1 \quad \Rightarrow \quad 0 \geqslant -x^\lambda \geqslant -1 \quad \Rightarrow \quad 0 \leqslant 1 - x^\lambda \leqslant 1.$$

$$|x| = \begin{cases} x, & \text{if } x \geqslant 0; \\ -x, & \text{if } x < 0. \end{cases}$$

Hence

$$\|E\|_1 = \int_0^1 (1 - x^\lambda)dx = x - \frac{x^{\lambda+1}}{\lambda+1}\Big|_0^1 = (1 - \frac{1}{\lambda+1}) - (0-0) = \frac{\lambda}{\lambda+1}.$$

$$\begin{aligned}
\|E\|_2^2 &= \int_a^b |E(x)|^2 dx = \int_0^1 |1 - x^\lambda|^2 dx = \int_0^1 (1 - x^\lambda)^2 dx = \int_0^1 (1 - 2x^\lambda + x^{2\lambda})dx \\
&= x - 2\frac{x^{\lambda+1}}{\lambda+1} + \frac{x^{2\lambda+1}}{2\lambda+1}\Big|_0^1 = (1 - \frac{2}{\lambda+1} + \frac{1}{2\lambda+1}) - (0 - 0 + 0) \\
&= \frac{(\lambda+1)(2\lambda+1) - 2(2\lambda+1) + (\lambda+1)}{(\lambda+1)(2\lambda+1)} = \frac{2\lambda^2 + 3\lambda + 1 - 4\lambda - 2 + \lambda + 1}{(\lambda+1)(2\lambda+1)} \\
&= \frac{2\lambda^2}{(\lambda+1)(2\lambda+1)}.
\end{aligned}$$

Hence

$$\|E\|_2^2 = \frac{2\lambda^2}{(\lambda+1)(2\lambda+1)}. \quad \Rightarrow \quad \|E\|_2 = \left[\frac{2\lambda^2}{(\lambda+1)(2\lambda+1)}\right]^{\frac{1}{2}}.$$

$$\|E\|_\infty = \max_{a\leqslant x\leqslant b} |E(x)| = \max_{0\leqslant x\leqslant 1} |1 - x^\lambda| = 1.$$

if $\lambda \to 0$, then $\|E\|_1 \to 0$ and $\|E\|_2 \to 0$, but $\|E\|_\infty$ remains 1.

Theorem 1.4. *For all E in $\mathbb{R}^N$ the inequalities*

$$\|E\|_1 \leqslant N^{\frac{1}{2}}\|E\|_2 \leqslant N\|E\|_\infty$$

hold.

Proof. H.W. □

Many question of mathematical interest arise in a natural way from the general best approximation problem (Definition 1.6). For example we may ask the following questions:

(1) Does a best approximation exists?

(2) Is a best approximation unique?

(3) How can a best approximation be characterized?

(4) How can a best approximation be computed?

While we shall refer to these questions, in this lectures the attention will be restricted to the Chebyshev norm as a measure of error.

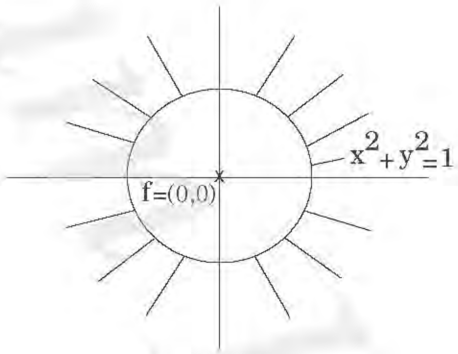
1.3 Existence

We can investigate an example with regard to question (1).

Example 1.4. *Let* $\mathbb{F} = \mathbb{R}^2$, $V = \{(x, y) : x^2 + y^2 > 1\}$, $f = (0, 0)$, $\|\cdot\| = \|\cdot\|_2$. *Discuss existence of best approximation.*

Solution. The problem of determining the point in V which is nearest to $(0, 0)$ has no solution. i.e., there is no best approximation to f from V.

If $V_1 = \{(x, y) : x^2 + y^2 \geqslant 1\}$ then all points that satisfy $x^2 + y^2 = 1$ is a best approximation to $f = (0, 0)$. i.e., the best approximation to f from V_1 is exists and not unique.



Definition 1.7. *A sequence* $\{x_n\}$ *in a normed linear space is said to converge to a point* x^* *and we write* $x_n \to x^*$ *if* $\|x_n - x^*\| \to 0$ *as* $n \to \infty$.

Definition 1.8. *We say that* $\delta = \inf X$ *if there exists a sequence* $\{x_n\}_{n=1}^{\infty} \in X$ *such that* $x_n \to \delta$ *as* $n \to \infty$.

Definition 1.9. *An element* $h^* \in V$ *satisfying* $\|f - h^*\| = \inf_{h \in V} \|f - h\|$ *is called a best approximation of* f *with respect to* V.

Definition 1.10. *A subset* V *of* $\mathbb{F}$ *is said to be compact if every sequence of points in* V *has a subsequence which is converge to a point of* V.

Theorem 1.5. *Let* V *be a compact subset of* $\mathbb{F}$, *then there exists* $h^* \in V$ *such that*

$$\|f - h^*\| \leqslant \|f - h\| \quad \forall\ h \in V.$$

Proof. Let $\delta = \inf_{h \in V} \|f - h\|$. We want to show that there exists $h^* \in V$ such that $\|f-h^*\| = \delta$. From the definition of infimum there exists a sequence of points $\{h_n\}_{n=1}^{\infty} \in V$ such that $\|f - h_n\| \to \delta$ as $n \to \infty$. Since V is compact, it follows that there exists a subsequence of $\{h_n\}_{n=1}^{\infty}$ converging to $h^* \in V$.

$$f - h^* = (f - h_n) + (h_n - h^*).$$
$$\|f - h^*\| \leqslant \|f - h_n\| + \|h_n - h^*\|.$$

when $n \to \infty$ we get

$$\|f - h^*\| \leqslant \delta. \tag{1.5}$$

Note that

$$\delta = \inf_{h \in V} \|f - h\| \leqslant \|f - h\| \quad \forall\, h \in V.$$

Since $h^* \in V$ we get

$$\delta \leqslant \|f - h^*\|. \tag{1.6}$$

From (1.5) and (1.6) we get

$$\|f - h^*\| = \delta.$$

i.e.,

$$\|f - h^*\| \leqslant \|f - h\| \quad \forall\, h \in V.$$

$\therefore$ h^* is a best approximation of f. □

Remark 1.6. *Compactness of V is a sufficient condition for a best approximation to exist and not necessary.*

Example 1.5. *Let $\mathbb{F} = \mathbb{R}$, $V = (-\infty, 1]$, $\|\cdot\|_1 = |\cdot|$. Discuss existence of best approximation.*

Solution. Note that V is not compact set, but there exists a best approximation to any point $f \in \mathbb{R}$.

Theorem 1.6. *Let V be a finite dimensional subspace of a normed linear space $\mathbb{F}$, then there exists a best approximation in V to any point of $\mathbb{F}$.*

Proof. Let V be such a subspace and let $f \in \mathbb{F}$ be the prescribed point. Then if h_0 is an arbitrary point of V, the point sought lies in the set

$$\{h \in V : \|f - h\| \leqslant \|f - h_0\|\}.$$

This set is closed and bounded and thus compact, then by Theorem 1.5 there exists a best approximation in V to $f \in \mathbb{F}$. □

Remark 1.7. *It is not possible to drop the finite dimensional requirement of the above theorem.*

Example 1.6. *Let $\mathbb{F} = C[0, \frac{1}{2}]$ with the ∞-norm, V = the space of polynomials of any degree.*

Solution. Let $f = \dfrac{1}{1-x}$, $h(x) = 1 + x + x^2 + \cdots + x^n \in V$.

$$\begin{aligned}\|f - h\| &= \max_{a \leqslant x \leqslant b} |f(x) - h(x)| \\ &= \max_{0 \leqslant x \leqslant \frac{1}{2}} \left|\frac{1}{1-x} - (1 + x + x^2 + \cdots + x^n)\right|.\end{aligned}$$

Thus any best approximation say h^* would satisfy $\|f - h^*\| = 0$ which implies $h^* = \dfrac{1}{1-x}$. This impossible and so no best approximation exists.

1.4 Uniqueness

We can investigate an example with regard to question (2).

Example 1.7. $\mathbb{F} = \mathbb{R}^2$, $V = \{(1, y) : y \in \mathbb{R}\}$, $f = (0, 0)$, $\|\cdot\| = \|\cdot\|_\infty$. *Discuss existence and uniqueness of best approximation.*

Solution.
$$\|f-h\|_\infty = \|(0,0)-(1,y)\| = \|(-1,-y)\|_\infty = \max\{1,|y|\}$$
$$= \begin{cases} 1, & |y| \leqslant 1; \\ >1, & |y|>1. \end{cases}$$
$$|y| \leqslant 1 \quad \Rightarrow \quad -1 \leqslant y \leqslant 1.$$

Hence
$$\|f-h\|_\infty = \begin{cases} 1, & -1 \leqslant y \leqslant 1; \\ >1, & y<-1 \text{ or } y>1. \end{cases}$$

$\therefore$ any point $(1,y)$ such that $-1 \leqslant y \leqslant 1$ is a best approximation to $f=(0,0)$.

i.e., the best approximation to $f=(0,0)$ exists and not unique.

To discuss the uniqueness of best approximation we need to define a convex set.

Definition 1.11. *A set V of a linear space $\mathbb{F}$ is convex if $\forall\ x,y \in V$ implies that $\lambda x+(1-\lambda)y \in V$ for all $0 \leqslant \lambda \leqslant 1$.*

Geometrically: A set is convex if all line segments joining pairs of points in the set also belongs to the set.

Theorem 1.7. *If $f \in \mathbb{F}$ and V is a subspace of $\mathbb{F}$, then the set of best approximation to f from V, call it V^*, is convex.*

Proof. Let $h_1^*, h_2^* \in V^*$, we want to proof that $\lambda h_1^* + (1-\lambda)h_2^* \in V^*$.

$$h_1^* \in V^* \quad \Rightarrow \quad \|f-h_1^*\| \leqslant \|f-h\| \quad \forall\ h \in V.$$
$$h_2^* \in V^* \quad \Rightarrow \quad \|f-h_2^*\| \leqslant \|f-h\| \quad \forall\ h \in V.$$

$$\begin{aligned}\|f-(\lambda h_1^*+(1-\lambda)h_2^*)\| &= \|(\lambda f+(1-\lambda)f)-(\lambda h_1^*+(1-\lambda)h_2^*)\| \\ &= \|\lambda(f-h_1^*)+(1-\lambda)(f-h_2^*)\| \\ &\leqslant \lambda\|f-h_1^*\|+(1-\lambda)\|f-h_2^*\| \\ &\leqslant \lambda\|f-h\|+(1-\lambda)\|f-h\| \quad \forall\ h \in V \\ &\leqslant \|f-h\| \quad \forall\ h \in V.\end{aligned}$$

$\therefore$ $\lambda h_1^*+(1-\lambda)h_2^*$ is a best approximation to f.

Hence $\lambda h_1^*+(1-\lambda)h_2^* \in V^*$. □

Remark 1.8. *This theorem has a consequence that if there are two distinct best approximation to f there are infinitely many. i.e., the set of best approximation consists either on one element or infinitely many.*

Definition 1.12. *We say that $\mathbb{F}$ is a strictly convex normed linear space if $f_1 \neq f_2$, $\|f_1\| = r$, $\|f_2\| = r$, then $\|\lambda f_1 + (1-\lambda) f_2\| < r$ for all λ satisfying $0 < \lambda < 1$.*

Example 1.8. *Let $\mathbb{F} = C[0,1]$, $f_1(x) = 2x$, $f_2(x) = 3x^2$, $\|\cdot\| = \|\cdot\|_1$. Show that $\|\cdot\|_1$ is not strictly convex.*

Solution. Note that $f_1 \neq f_2$.

$$\|f_1\|_1 = \int_a^b |f_1(x)|dx = \int_0^1 |2x|dx = \int_0^1 2x dx = x^2\Big|_0^1 = 1.$$
$$\|f_2\|_1 = \int_a^b |f_2(x)|dx = \int_0^1 |3x^2|dx = \int_0^1 3x^2 dx = x^3\Big|_0^1 = 1.$$

Take $\lambda = \frac{1}{2}$.

$$\begin{aligned}\|\lambda f_1 + (1-\lambda) f_2\|_1 &= \|\frac{1}{2}(2x) + \frac{1}{2}(3x^2)\|_1 = \|x + \frac{3}{2}x^2\|_1 = \int_0^1 |x + \frac{3}{2}x^2|dx \\ &= \int_0^1 (x + \frac{3}{2}x^2)dx = \frac{x^2}{2} + \frac{x^3}{2}\Big|_0^1 = 1.\end{aligned}$$

Hence

$$\|\lambda f_1 + (1-\lambda) f_2\|_1 = 1.$$

$\therefore$ $\|\cdot\|_1$ is not strictly convex.

Theorem 1.8. *In a strictly convex normed linear space $\mathbb{F}$. A finite dimensional subspace V contains a unique best approximation to any point $f \in \mathbb{F}$.*

Proof. There exists a best approximation to f from V because V is a finite dimensional subspace.

Suppose that h_1 and h_2 are two distinct best approximation to f.

Take $\lambda = \frac{1}{2}$.

$\Rightarrow \quad \frac{1}{2}h_1 + \frac{1}{2}h_2$ is also a best approximation (by convexity).

suppose that

$$\|f - h_1\| = \|f - h_2\| = \|f - (\frac{1}{2}h_1 + \frac{1}{2}h_2)\| = r.$$

Put $f_1 = f - h_1$, $f_2 = f - h_2$ and notice that $f_1 \neq f_2$ and $\|f_1\| = r$, $\|f_2\| = r$.

Since $\mathbb{F}$ is strictly convex normed linear space

$$\|\frac{1}{2}f_1 + \frac{1}{2}f_2\| < r. \tag{1.7}$$

$$\|\frac{1}{2}f_1 + \frac{1}{2}f_2\| = \|\frac{1}{2}(f - h_1) + \frac{1}{2}(f - h_2)\| = \|f - (\frac{1}{2}h_1 + \frac{1}{2}h_2)\| = r.$$

$\therefore \quad \|\frac{1}{2}f_1 + \frac{1}{2}f_2\| = r.$

This contradicts inequality (1.7).

$\therefore \quad h_1 = h_2.$

i.e., the best approximation is unique. □

Remark 1.9. *If f and g are in $C[a, b]$ or $\mathbb{R}^N$, then*

$$\|af + bg\|_2^2 = a^2\|f\|_2^2 + 2ab < f, g > +b^2\|g\|_2^2,$$

where a, b are constants.

Proof.
$$\begin{aligned}\|af + bg\|_2^2 &= < af + bg, af + bg > \\ &= a^2 < f, f > +ab < f, g > +ba < g, f > +b^2 < g, g > \\ &= a^2\|f\|_2^2 + 2ab < f, g > +b^2\|g\|_2^2.\end{aligned}$$

□

Theorem 1.9. *The normed linear space $C[a, b]$ or $\mathbb{R}^N$ with the 2-norm is strictly convex.*

Proof. Let f and g be any two distinct points of $C[a, b]$ or $\mathbb{R}^N$ such that $\|f\|_2 = \|g\|_2 = r$ and note that

$$\begin{aligned}\|\lambda f+(1-\lambda)g\|_2^2+\lambda(1-\lambda)\|f-g\|_2^2 &= \lambda^2\|f\|_2^2+2\lambda(1-\lambda)<f,g>+(1-\lambda)^2\|g\|_2^2\\ &\quad+\lambda(1-\lambda)\{\|f\|_2^2-2<f,g>+\|g\|_2^2\}\\ &=\lambda^2r^2+(1-\lambda)^2r^2+2\lambda(1-\lambda)r^2\\ &=r^2\{\lambda^2+(1-2\lambda+\lambda^2)+(2\lambda-2\lambda^2)\}\\ &=r^2.\end{aligned}$$

$\therefore \quad \|\lambda f+(1-\lambda)g\|_2^2+\lambda(1-\lambda)\|f-g\|_2^2=r^2.$

$f\neq g \quad \Rightarrow \quad f-g\neq 0 \quad \Rightarrow \quad \lambda(1-\lambda)\|f-g\|_2^2>0.$

$\Rightarrow \quad \|\lambda f+(1-\lambda)g\|_2^2<r^2.$

$\Rightarrow \quad \|\lambda f+(1-\lambda)g\|_2<r.$

$\therefore$ the normed linear space $C[a,b]$ or $\mathbb{R}^N$ with the 2-norm is strictly convex. □

Remark 1.10.

(1) *From Theorem* 1.8 *and* 1.9, *we obtain that any point* $f\in C[a,b]$ *or* $\mathbb{R}^N$ *has a unique best approximation in the finite dimensional subspace* $V\subset C[a,b]$ *or* $\mathbb{R}^N$ *with the* 2*-norm.*

(2) *It has been stated already that the* 1*-norm and* ∞*-norm in* $C[a,b]$ *and in* $\mathbb{R}^N$ *are not strictly convex. If we prove that the norms are not strictly convex, then Theorem* 1.8 *does not answer the uniqueness question, but if we can demonstrate that a best approximation from a linear subspace is not unique , then we may deduce from Theorem* 1.8 *that the norm is not strictly convex.*

We give examples of this kind. In each one there is a linear subspace V *and a point* f *such that the best approximation from* V *to* f *is not unique , where* V *and* f *contained in either* $C[a,b]$ *or in* $\mathbb{R}^N$ *and where the accuracy of the approximation is measured either by the* 1*-norm or by the* ∞*-norm.*

Example 1.9. *Let* $\mathbb{F}=C[-1,1]$, V*=one dimensional linear space that contains all functions of the form* $h(x)=\lambda x,\ -1\leqslant x\leqslant 1, -1\leqslant\lambda\leqslant 1,\ f(x)=1,\ \|\cdot\|=\|\cdot\|_1$. *Show that* $\|\cdot\|_1$ *is not strictly convex.*

Solution. $$\|f-h\|_1=\int_a^b |f(x)-h(x)|dx=\int_{-1}^1 |1-\lambda x|dx.$$

$$-1\leqslant x\leqslant 1,\ -1\leqslant \lambda\leqslant 1 \quad\Rightarrow\quad -1\leqslant -\lambda x\leqslant 1 \quad\Rightarrow\quad 0\leqslant 1-\lambda x\leqslant 2.$$

Hence

$$\|f-h\|_1=\int_{-1}^1(1-\lambda x)dx=x-\frac{\lambda}{2}x^2\Big|_{-1}^1=(1-\frac{\lambda}{2})-(-1-\frac{\lambda}{2})=2.$$

$\therefore\quad \|f-h\|_1=2 \quad \forall\, h\in V. \quad\Rightarrow\quad \min_{h\in V}\|f-h\|_1=2.$

Hence the best approximation is not unique. i.e., the 1-norm in $C[-1,1]$ is not strictly convex.

Example 1.10. *Let* $\mathbb{F}=\mathbb{R}^N$, $V=\{h:h=(\lambda x_1,\lambda x_2,\ldots,\lambda x_N),-1\leqslant x_i\leqslant 1$ *for all* $i=1,2,\ldots,N,-1\leqslant\lambda\leqslant 1\}$, $f(x)=(1,1,\ldots,1)\in\mathbb{R}^N$, $\|\cdot\|=\|\cdot\|_1$. *Show that* $\|\cdot\|_1$ *is not strictly convex.*

Solution. divide the interval $[-1,1]$ by the points $-1=x_1<x<\cdots<x_N=1$ which are equally spaced.

$$h=\frac{b-a}{N-1}=\frac{1-(-1)}{N-1}=\frac{2}{N-1}.$$

$\therefore\quad x_{i+1}-x_i=\dfrac{2}{N-1}.$

$$x_i=x_1+(i-1)h=-1+\frac{2(i-1)}{N-1} \quad i=1,2,\ldots,N.$$

$$\|f-h\|_1=\sum_{i=1}^N|f(x_i)-h(x_i)|=\sum_{i=1}^N|1-\lambda x_i|.$$

$$-1\leqslant x_i\leqslant 1,\ -1\leqslant\lambda\leqslant 1 \quad\Rightarrow\quad -1\leqslant -\lambda x_i\leqslant 1 \quad\Rightarrow\quad 0\leqslant 1-\lambda x_i\leqslant 2$$

Hence

$$\|f-h\|_1=\sum_{i=1}^N(1-\lambda x_i)=\sum_{i=1}^N 1-\lambda\sum_{i=1}^N x_i=N-\lambda\sum_{i=1}^N x_i.$$

$$\sum_{i=1}^N x_i=\sum_{i=1}^N\Big[-1+\frac{2}{N-1}(i-1)\Big]=-N+\frac{2}{N-1}(1+2+\cdots+N-1)=-N+N=0.$$

$\therefore\quad \|f-h\|_1=N \quad \forall\, h\in V. \quad\Rightarrow\quad \min_{h\in V}\|f-h\|_1=N.$

Hence the best approximation is not unique. i.e., the 1-norm in $\mathbb{R}^N$ is not strictly convex.

Example 1.11. *Let* $\mathbb{F} = C[-1,1]$, V=*one dimensional linear space that contains all functions of the form* $h(x) = \lambda(1+x)$, $-1 \leqslant x \leqslant 1, 0 \leqslant \lambda \leqslant 1$, $f(x) = 1$, $\|\cdot\| = \|\cdot\|_\infty$. *Show that* $\|\cdot\|_\infty$ *is not strictly convex.*

Solution. $$\|f-h\|_\infty = \max_{a\leqslant x\leqslant b} |f(x) - h(x)| = \max_{-1\leqslant x\leqslant 1} |1 - \lambda(1+x)|.$$

$$-1 \leqslant x \leqslant 1 \quad \Rightarrow \quad 0 \leqslant x+1 \leqslant 2 \quad \Rightarrow \quad -2 \leqslant -2\lambda \leqslant -\lambda(x+1) \leqslant 0$$

$$\Rightarrow \quad -1 \leqslant 1 - \lambda(x+1) \leqslant 1$$

$\therefore \quad \|f-h\|_\infty = 1 \quad \forall\, h \in V. \quad \Rightarrow \quad \min_{h\in V} \|f-h\|_\infty = 1.$

Hence the best approximation is not unique. i.e., the ∞-norm in $C[-1,1]$ is not strictly convex.

Example 1.12. *Let* $\mathbb{F} = \mathbb{R}^N$, $V = \{h : h = (\lambda(1+x_1), \lambda(1+x_2), \dots, \lambda(1+x_N)), -1 \leqslant x_i \leqslant 1$ *for all* $i = 1, 2, \dots, N, 0 \leqslant \lambda \leqslant 1\}$, $f(x) = (1,1,\dots,1) \in \mathbb{R}^N$, $\|\cdot\| = \|\cdot\|_\infty$. *Show that* $\|\cdot\|_\infty$ *is not strictly convex.*

Solution. $$\|f-h\|_\infty = \max_{1\leqslant i\leqslant N} |f(x_i) - h(x_i)| = \max_{1\leqslant i\leqslant N} |1 - \lambda(1+x_i)|.$$

$$-1 \leqslant x_i \leqslant 1 \quad \Rightarrow \quad 0 \leqslant x_i+1 \leqslant 2 \quad \Rightarrow \quad -2 \leqslant -2\lambda \leqslant -\lambda(x_i+1) \leqslant 0$$

$$\Rightarrow \quad -1 \leqslant 1 - \lambda(x_i+1) \leqslant 1$$

$\therefore \quad \|f-h\|_\infty = 1 \quad \forall\, h \in V. \quad \Rightarrow \quad \min_{h\in V} \|f-h\|_\infty = 1.$

Hence the best approximation is not unique. i.e., the ∞-norm in $\mathbb{R}^N$ is not strictly convex.

Remark 1.11. *When the normed linear space is* $C[a,b]$ *and when the norm is either* 1*-norm or the* ∞*-norm and* V *is the space* $P_n(\mathbb{R})$ *then the best approximation is unique for all* f *in* $C[a,b]$ (*this statement is proved later*).

The purpose of the four examples, therefore is to show that if V *is a linear subspace of a normed linear space whose norm is not strictly convex, then the uniqueness of best approximation depends on properties of* V *and* f.