Chapter 4

Approximation Theory

4.1 Introduction

In this chapter, we consider approximation of, for example, $f \in C[a,b]$, where $C[a,b]$ represents the space of continuous functions on the interval $[a,b]$. We are interested in approximating $f(x)$ by an elementary function $p(x)$ on the interval $[a,b]$. For example, $p(x)$ could be a polynomial of degree n, a continuous piecewise linear function, a trigonometric polynomial, a rational function, or a linear combination of "nice" functions, that is, functions that are easy to use in numerical computation.

We study approximation of functions in the setting of normed vector spaces. We begin with a review of normed linear spaces, inner products, projections, and orthogonalization.

4.2 Norms, Projections, Inner Product Spaces, and Orthogonalization in Function Spaces

Recall the basic properties of normed vector spaces, which we review in Section 3.2 on page 88. Here, we will be using norms in the context of vector spaces without a finite-dimensional basis.

Example 4.1

Here, V is our vector space.

(a) $V = C[a,b]$. Two common norms are:

(a1) $\|v\|_\infty = \max_{a \le x \le b} (|v(x)|\rho(x))$,
where $\rho(x) > 0$ on $[a,b]$ and $\rho \in C[a,b]$. This is called the *Chebyshev*, *uniform*, or *max norm* with weight function $\rho(x)$.

(a2) $\|v\|_2 = \left(\int_a^b |v(x)|^2 \rho(x) dx \right)^{\frac{1}{2}}$, where $\rho(x) > 0$, $a < x < b$, $\rho \in C(a,b)$, and

$$\int_a^b \rho(x)dx < \infty.$$

This is called the L^2-*norm* or *least squares norm* with weight function $\rho(x)$.

(b) If $V = \mathbb{R}^n$, then $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ are norms for $\mathbb{R}^n$. (See Section 3.2 on page 88 and the subsequent pages.)

$\square$

4.2.1 Best Approximations

Now let W be a finite-dimensional subspace of V. A typical problem in approximation theory is: Given $v \in V$, find $w \in W$ such that the distance $\|v - w\|$ is least among all $w \in W$. Such a w is called a *best approximation in* W *to* v with respect to norm $\|\cdot\|$. (For example, $V = C[a,b]$, $\|\cdot\| = \|\cdot\|_\infty$, and $W = \{$*set of polynomials of degree* $\le n\}$.)
Question: Does such a w exist? We will prove that the answer to this is yes.

THEOREM 4.1
(Existence of a best approximation) *Let W be an $n+1$-dimensional subspace of a normed linear space V. Let u_0, u_1, $\ldots$, u_n be linearly independent elements of W. (Thus, $W = \mathrm{span}(u_0, u_1, u_2, \ldots, u_n)$.) Then there is a $p \in W$, i.e., $p = \sum_{j=0}^{n} \alpha_j u_j$ for a given $f \in V$, such that*

$$\|f - p\| = \Big\|f - \sum_{j=0}^{n} \alpha_j u_j\Big\| = \min_{\gamma_0,\gamma_1,\cdots\gamma_n} \Big\|f - \sum_{j=0}^{n} \gamma_j u_j\Big\|,$$

that is, $\|f - p\| \le \|f - q\|$ for all $q \in W$. (p is the best approximation to $f \in V$ with respect to norm $\|\cdot\|$.) This is illustrated schematically in Figure 4.1.

PROOF The proof is divided into two cases.

Case 1: Suppose that f is linearly dependent on $u_0, u_1, \ldots, u_n$. Then

$$f = \sum_{j=0}^{n} \alpha_j u_j = p,$$

and $\|f - p\| = 0$. That is, $f \in W$, so $p = f$.

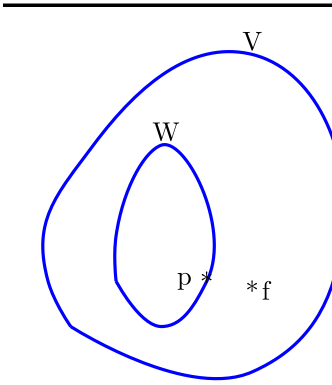

FIGURE 4.1: p is the best approximation to $f \in V$.

Case 2: Suppose that $f \notin W$, so f is linearly independent of u_0, u_1, $\ldots$, u_n. Let

$$E = \Big\{ w \in V : w = \sum_{j=0}^{n} z_j u_j + z_{n+1} f \Big\},$$

where $z_j \in \mathbb{R}$ for $j = 0, 1, \ldots, n+1$. Then E is an $(n+2)$-dimensional subspace of V. Let $z = (z_0, z_1, \ldots, z_{n+1})^T \in \mathbb{R}^{n+2}$, where $z_0, z_1, \ldots, z_{n+1}$ are given. Let $\|\cdot\|_*$ be defined on $\mathbb{R}^{n+2}$ by

$$\|z\|_* = \Big\| \sum_{j=0}^{n} z_j u_j + z_{n+1} f \Big\|.$$

Then,

1. $\|z\|_* \geq 0$ and $\|z\|_* = 0$ if and only if $z = 0$, since u_0, u_1, $\ldots$, u_n, f are independent,

2. $\|\lambda z\|_* = |\lambda| \|z\|_*$, and

3. $\|z_1 + z_2\|_* \leq \|z_1\|_* + \|z_2\|_*$,

because $\|\cdot\|$ has these properties. Thus, $\|\cdot\|_*$ is a norm on $\mathbb{R}^{n+2}$, i.e., $\|\cdot\|_* : \mathbb{R}^{n+2} \to \mathbb{R}$. Now, define p to be such that

$$\|f - p\| = \Big\| f - \sum_{j=0}^{n} \alpha_j u_j \Big\| = \min_{\gamma_0, \gamma_1, \ldots, \gamma_n} \Big\| f - \sum_{j=0}^{n} \gamma_j u_j \Big\|,$$

assuming that p exists. But

$$\|f - p\| = \Big\| f - \sum_{j=0}^{n} \alpha_j u_j \Big\| = \|e - a\|_*,$$

where $e = (0, 0, \ldots, 0, 1)^T$ and $a = (\alpha_0, \alpha_1, \ldots, \alpha_n, 0)^T$. Clearly,

$$\|e - a\|_* = \min_{z \in G} \|e - z\|_*,$$

where $G = \{z \in \mathbb{R}^{n+2} : z = (z_0, z_1, \ldots, z_n, 0)^T\}$. Thus, if $a \in \mathbb{R}^{n+2}$ exists such that $\|e - a\|_* = \min_{z \in G} \|e - z\|_*$, then

$$p = \sum_{j=0}^{n} \alpha_j u_j \in V$$

exists. (We have reduced the problem of finding $p \in W$ to finding $a \in \mathbb{R}^{n+2}$.)
Question: Can we find a ?

Define $H = \{z \in G : \|z - e\|_* \leq \|e\|_*\}$. Then,

(a) H is not empty, since $0 = (0, 0, \ldots, 0)^T \in H$.

(b) H is bounded in $\mathbb{R}^{n+2}$ since if $z \in H$, then

$$\|z\|_* \leq \|z - e\|_* + \|e\|_* \leq 2\|e\|_*.$$

Therefore, z has an infimum[1] $\mu = \inf_{z \in H} \|e - z\|_*$. Let

$$z^{(t)} = (z_0^{(t)}, z_1^{(t)}, \ldots z_n^{(t)}, 0) \in H \subset \mathbb{R}^{n+2}$$

be a sequence of vectors such that $\|e - z^{(t)}\|_* \to \mu$ as $t \to \infty$. By the Bolzano–Weierstrass Theorem (i.e. every bounded sequence in $\mathbb{R}^m$ has a convergent subsequence), a subsequence of $\{z^{(t)}\}$ has a limit point $\widehat{z}$. Thus, $a = \widehat{z}$. This, in turns, implies existence of $p \in W$. ▯

Example 4.2
Find the best approximation $p \in P^0$ to $f(x) = e^x \in C[0, 1]$ for the $\|\cdot\|_\infty$ and $\|\cdot\|_2$ norms. (Thus, $W = P^0 \subset V = C[0, 1]$.)

(a) Find p that minimizes $\|e^x - p\|_\infty = \max_{0 \leq x \leq 1} |e^x - p|$.

In this case, $p = \frac{1}{2}(e + 1)$ and $\max_{0 \leq x \leq 1} |e^x - p| = \frac{1}{2}(e - 1)$.

(b) Find p that minimizes $\|e^x - p\|_2 = \left(\int_0^1 (e^x - p)^2 dx\right)^{\frac{1}{2}}$.

In this case, $p = e - 1$ and $\|e^x - p\|_2 = \frac{1}{2}(4e - e^2 - 3)^{\frac{1}{2}}$.

[1] A fundamental property of a finite-dimensional vector space is that every set that is bounded below has a greatest lower bound, or infimum.

$\square$

We will now see that approximation in inner product spaces is straightforward. We introduced the concept of inner product spaces in the context of matrix computations (i.e. finite-dimensional vector spaces) in Definition 3.19 on page 91. We review this same concept here, in the more general context of function spaces.

DEFINITION 4.1 *A real vector space V is a* real inner product space *if for each $u, v \in V$, a real number (u, v) can be defined with the properties:*

(i) $(u, u) \geq 0$ *for* $u \in V$ *with* $(u, u) = 0$ *if and only if* $u = 0$,

(ii) $(u, v) = (v, u)$, *and*

(iii) $(\alpha u + \beta v, w) = \alpha(u, w) + \beta(v, w)$ *for all* $u, v, w \in V$ *and* $\alpha, \beta \in \mathbb{R}$.

Example 4.3
(of inner product spaces)

(1) $V = C[a, b]$ with $(f, g) = \int_a^b \rho(x) f(x) g(x) dx$, where $\rho(x) > 0$ for $a \leq x \leq b$, and $\rho \in C[a, b]$.

(2) $\mathbb{R}^n$ with $(x, y) = x^T y$.

$\square$

Unless we specify otherwise, when we say "inner product space" or "normed linear space" in the remainder of this chapter, we will mean "real inner product space." Much of what is presented is also true in spaces over the complex numbers.

REMARK 4.1 Complex inner product spaces can be defined analogously, with the following modifications to properties (ii) and (iii) of Definition 4.1:

(ii)$'$ $(u, v) = \overline{(v, u)}$,

(iii)$'$ $(\alpha u + \beta v, w) = \overline{\alpha}(u, w) + \overline{\beta}(v, w)$ for all u, v, $w \in V$ and α, $\beta \in \mathbb{C}$.

Complex inner product spaces corresponding to Example 4.3 are

(1) $\mathcal{V} = f : [a, b] \to \mathbb{C}$, f continuous, with $(f, g) = \int_a^b \rho(x) \overline{f}(x) g(x) dx$, where $\rho(x) > 0$ for $a \leq x \leq b$, and $\rho \in C[a, b]$.

(2) $\mathbb{C}^n$ with $(z, w) = z^H w$, where z^H is the conjugate transpose of z.

We will work with complex inner product spaces when we study trigonometric approximation in §4.5. ▯

THEOREM 4.2

Any real inner product space V is a real normed linear space with norm defined by $\|v\| = (v, v)^{\frac{1}{2}}$.

PROOF Clearly (i), (ii), and (iii) of Definition 3.13 (the definition of norm, on page 88) are satisfied, while (iv) follows from the Cauchy–Schwarz inequality

$$|(u, v)| \leq \|u\|\|v\| \quad \forall u, v \in V.$$

In particular,

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

▯

Example 4.4

(norms on inner product spaces)

(1) $V = C[a, b]$ is an inner product space with inner product

$$(f, g) = \int_a^b f(x)g(x)dx$$

and a normed linear space with the norm

$$\|f\| = \left(\int_a^b f^2(x)dx\right)^{\frac{1}{2}}.$$

The Cauchy–Schwarz inequality for this inner product has the form

$$\left|\int_a^b f(x)g(x)dx\right| \leq \left(\int_a^b f^2(x)dx\right)^{\frac{1}{2}} \left(\int_a^b g^2(x)dx\right)^{\frac{1}{2}}.$$

(2) $V = \mathbb{R}^n$ is an inner product space with inner product

$$(x, y) = \sum_{i=1}^n x_i y_i = x^T y$$

and a normed space with norm

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}.$$

The Cauchy–Schwarz inequality for this inner product is

$$\left|\sum_{i=1}^{n} x_i y_i\right| \le \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} y_i^2\right)^{\frac{1}{2}}.$$

▯

REMARK 4.2 The concept of a *Cauchy sequence* (Definition 2.3 on page 40) can be generalized to normed linear spaces. A sequence of vectors u_1, u_2, ... in a normed linear space is said to be a Cauchy sequence if, given any $\epsilon > 0$, there is an integer $N = N(\epsilon)$ such that $\|u_n - u_m\| < \epsilon$ for all $m, n \ge N$. It is easy to show that every convergent sequence is a Cauchy sequence. A Cauchy sequence, however, may not be convergent to an element of the space. If every Cauchy sequence in V is convergent, we say that V is complete.[2] A complete normed linear space is called a *Banach space*. A complete inner product space is called a *Hilbert space*. ▯

Example 4.5
(of Hilbert and Banach spaces)

(1) Let ℓ_2 = set of all real sequences $u = \{u_1, u_2, \ldots\} = \{u_i\}_{i=1}^{\infty}$ that satisfy $\sum_{i=1}^{n} |u_i|^2 < \infty$. Define

$$(u, v) = \sum_{j=1}^{\infty} u_j v_j.$$

Then ℓ_2 is an inner product space, and ℓ_2 can be shown to be complete. Thus, ℓ_2 is a Hilbert space.

(2) $C[a, b]$ with norm $\| \cdot \|_\infty$ is complete and is thus a Banach space.

▯

[2]Basically, this means that the space contains all of its limit points.

4.2.2 Best Approximations in Inner Product Spaces

Consider now the problem of finding the best approximation in an inner product space. Specifically, we wish to find $w \in W \subset V$ that is closest to a given $v \in V$, where V is an inner product space and W is finite-dimensional. Let $W = \text{span}(w_1, w_2, \ldots, w_n) \subset V$, where $\{w_i\}_{i=1}^n$ is a linearly independent set. We wish to find $w \in W$ such that $\|w - v\|^2 \le \|u - v\|^2$ for all $u \in W$. But

$$
\begin{aligned}
\|w - v\|^2 &= (w - v, w - v) \\
&= \left(\sum_{j=1}^{n} \alpha_j w_j - v, \sum_{k=1}^{n} \alpha_k w_k - v\right) \\
&= \sum_j \sum_k \alpha_j \alpha_k (w_j, w_k) - \sum_j \alpha_j (v, w_j) - \sum_k \alpha_k (v, w_k) + (v, v) \\
&= F(\alpha_1, \alpha_2, \ldots, \alpha_n),
\end{aligned}
$$

where $w = \sum_{j=1}^{n} \alpha_j w_j$.

Thus, the problem reduces to finding the minimum of F as a function of α_1, α_2, $\ldots$, α_n. Hence, setting $\partial F / \partial \alpha_\ell = 0$ gives

$$\sum_{j=1}^{n} \alpha_j (w_j, w_\ell) - (v, w_\ell) = 0,$$

or

$$\sum_{j=1}^{n} \alpha_j (w_j, w_\ell) = (v, w_\ell) \text{ for } \ell = 1, 2, \ldots, n. \tag{4.1}$$

Equations (4.1) represent a linear system that is positive definite and hence invertible, and thus can be solved for the α_j. In this case, the best approximation w is called the least-squares approximation. (The matrix corresponding to the system (4.1) is often called the *Gram matrix*.)

REMARK 4.3 Compare this with our explanation in Section 3.3.8.4 on page 139. In particular, in both (3.29) and (4.1), the system of equations is called the *normal equations*. The functions φ in Section 3.3.8.4 correspond to the vectors w here. The dot product in Section 3.3.8.4 is the finite dot product

$$(\varphi^{(i)}, \varphi^{(j)}) = \sum_{k=1}^{m} \varphi^{(i)}(t_k) \varphi^{(j)}(t_k).$$

The latter is a dot product only if $\left\{\varphi^{(i)}\right\}_{i=1}^{m}$ is "linearly independent on the finite set $\{t_k\}_{k=1}^{m}$." ▯

The following concept is closely related to finding best approximations in Hilbert spaces.

DEFINITION 4.2 *Let W be a finite-dimensional subspace of an inner product space V. An operator P that maps V into W such that $P^2 = P$ is called a* projection operator *from V into W.*

REMARK 4.4 Projections are another useful way of defining approximations. For example, $P : V \to W$ can be defined as

$$Pv = \sum_{k=1}^{n} \alpha_k w_k,$$

where the α_k's satisfy

$$\sum_{k=1}^{n} \alpha_k (w_k, w_\ell) = (v, w_\ell) \text{ for } \ell = 1, 2, \ldots, n$$

and $\{w_1, \ldots, w_n\}$ is a basis for W. In this example, P is a "least squares" projection operator. $\square$

We now revisit the concept of orthonormal sets of vectors, which we originally introduced on page 92 in conjunction with QR factorizations.

DEFINITION 4.3 *(a restatement of Definition 3.21) Let V be an inner product space. Two vectors u and v in V are called orthogonal if $(u, v) = 0$. A set of such vectors that are pairwise orthogonal is called orthonormal, provided $(u, u) = 1$ for every vector u in that set.*

Let $w_1, w_2, \ldots, w_m$ be an orthonormal set in V, i.e., $(w_i, w_j) = \delta_{ij}$. Let

$$M = \text{span}(w_1, w_2, \ldots, w_m)$$

be the subspace in V spanned by $w_1, w_2, \ldots, w_m$, i.e., given $v \in M$, $v = \sum\limits_{i=1}^{m} c_i w_i$. Define

$$M^\perp = \{v \in V : (v, w) = 0 \text{ for every } w \in M\},$$

i.e., the elements of $M^\perp$ are orthogonal to those of M.

DEFINITION 4.4 *$M^\perp$ is called the* orthogonal complement *of M in V.*

REMARK 4.5 $M^\perp$ is a subspace of V. That is,

(a) $0 \in M^{\perp}$,

(b) if $v_1, v_2 \in M^{\perp}$ then $v_1 + v_2 \in M^{\perp}$,

(c) if $v_1 \in M^{\perp}$ then $c_1 v_1 \in M^{\perp}$.

$\square$

REMARK 4.6 Given $u \in V$, we can associate with u two vectors Pu and Qu, with

$$u = Pu + Qu,$$

where

$$Pu = \sum_{k=1}^{m} (u, w_k) w_k \quad \text{and} \quad Qu = u - \sum_{k=1}^{m} (u, w_k) w_k.$$

Clearly, $Pu \in M$ and $Qu \in M^{\perp}$. $\square$

DEFINITION 4.5 *The vector Pu is the* orthogonal projection *of u onto M and Qu is the* perpendicular from u onto M.

We now have

PROPOSITION 4.1

(Projections and best approximation) Let V be an inner product space. Given $u \in V$, $\|u - Pu\| \le \|u - h\|$ for any $h \in M$, where $\|\cdot\| = (\cdot,\cdot)^{\frac{1}{2}}$. Thus, the vector in M closest to $u \in V$ is Pu, i.e., Pu is the best approximation in M to $u \in V$ with respect to norm $\|\cdot\|$.

PROOF Let $a = h - Pu$, $a \in M$. Then $h = a + Pu$. Thus,

$$\|h - u\|^2 = \|a + Pu - u\|^2 = \|a + c\|^2,$$

where $c = Pu - u$. Continuing,

$$\|h - u\|^2 = (a + c, a + c) = (a, a) + (c, c),$$

since $c = -Qu \in M^{\perp}$ and $a \in M$, so $(a, c) = 0$. Therefore,

$$\|h - u\|^2 = \|a\|^2 + \|Pu - u\|^2,$$

so $\|Pu - u\| \le \|h - u\|$ for any $h \in M$. $\square$

REMARK 4.7 Notice how easy it is to find a best approximation in an inner product space from a finite-dimensional subspace with an orthonormal basis. $\square$

REMARK 4.8 Consider this proposition geometrically for $V = \mathbb{R}^3$ and

$$M = \text{span}(w_1, w_2) = \text{span}\left((1,0,0)^T), (0,1,0)^T\right).$$

(See Figure 4.2.) Notice that M is the xy-plane. The vector in M closest to u is

$$Pu = \sum_{k=1}^{2} (u, w_k) w_k = u_1 w_1 + u_2 w_2,$$

and $Qu = u_3(0,0,1)^T$. ▯

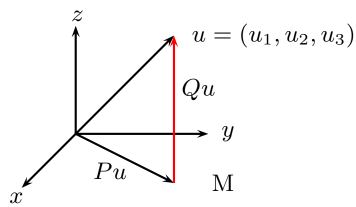

FIGURE 4.2: The projection is the best approximation. See Prop. 4.1.

REMARK 4.9 By Proposition 4.1, $\|u - Pu\| = \|Qu\|$ is the shortest distance from subspace M to u in V. ▯

4.2.3 The Gram–Schmidt Process (Formal Treatment)

Suppose now M has a basis $\{u_1, u_2, \ldots, u_m\}$ that is not orthonormal. Can we find an orthogonal basis $w_1, w_2, \ldots, w_n$? (Recall how easy it is to find a best approximation in M to V if M has an orthogonal basis.) This motivates the well-known Gram–Schmidt orthogonalization process.

In Section 3.3.8 (on page 138), we briefly introduced the Gram–Schmidt process in the context of QR-factorizations. Here, we carefully consider the process formally.

THEOREM 4.3

(Gram–Schmidt Process)

Let $u_1, u_2, \ldots, u_m$ be linearly independent vectors (elements) in inner product *space V. If*

$$\begin{aligned} v_1 &= u_1 \\ v_j &= u_j - \sum_{k=1}^{j-1} \frac{(u_j, v_k)v_k}{(v_k, v_k)} \qquad \text{for } j = 2, 3, \dots, m, \text{ and} \\ w_j &= \frac{v_j}{\|v_j\|} \qquad \text{for } j = 1, 2, \dots, m, \end{aligned}$$

then $v_1, v_2, \dots, v_m$ is an orthogonal system, and $w_1, w_2, \dots, w_m$ is an orthonormal system. Furthermore,

$$M = \text{span}(u_1, u_2, \dots, u_m) = \text{span}(v_1, v_2, \dots, v_m) = \text{span}(w_1, w_2, \dots, w_m).$$

PROOF (By induction) Suppose that

$$M_j = \text{span}(u_1, u_2, \dots, u_j) = \text{span}(v_1, v_2, \dots, v_j) = \text{span}(w_1, w_2, \dots, w_j),$$

where $v_1, \dots, v_j$ are orthogonal and $w_1, \dots, w_j$ are orthonormal. Now, for the induction step, assume that

$$v_{j+1} = u_{j+1} - \sum_{k=1}^{j} \frac{(u_{j+1}, v_k)v_k}{(v_k, v_k)}. \tag{4.2}$$

We have $v_{j+1} \notin M_j$ and $v_{j+1} \neq 0$ since u_{j+1} is independent of $u_1, u_2, \dots, u_j$. Also, $(v_{j+1}, v_\ell) = 0$ for $\ell = 1, 2, \dots, j$ (by simply plugging into (4.2)) and $(w_{j+1}, w_{j+1}) = 1$ since $w_{j+1} = v_{j+1}/\|v_{j+1}\|$. Thus, $v_1, \dots, v_{j+1}$ are orthogonal and $w_1, w_2, \dots, w_{j+1}$ are orthonormal. By construction,

$$\text{span}(w_1, w_2, \dots, w_{j+1}) = \text{span}(v_1, v_2, \dots, v_{j+1}) \subset \text{span}(u_1, u_2 \dots, u_{j+1}).$$

(Notice that by the induction hypothesis, each v_k is a linear combination of $u_1, u_2, \dots, u_j$ for $1 \leq k \leq j$.) Furthermore, we have

$$u_{j+1} = v_{j+1} + \sum_{k=1}^{j} \frac{(u_{j+1}, v_k)}{(v_k, v_k)} v_k.$$

Hence, $u_{j+1} \in \text{span}\ (v_1, v_2, \dots, v_{j+1})$. Therefore,

$$\begin{aligned} M_{j+1} &= \text{span}(u_1, u_2, \dots, u_{j+1}) = \text{span}(v_1, v_2, \dots, v_{j+1}) \\ &= \text{span}(w_1, w_2, \dots, w_{j+1}) \end{aligned}$$

for $j = 0, 1, \dots, m-1$. ∎

Example 4.6
(Two important cases)

(1) *Legendre Polynomials*
Let $V = C[-1,1]$, $M = \operatorname{span}(1, x, x^2)$,

$$(f,g) = \int_{-1}^{1} f(x)g(x)dx \text{ for } f,g \in V, \text{ and } \|f\| = (f,f)^{\frac{1}{2}}.$$

Notice that $p \in M$ has the form $p(x) = a + bx + cx^2$. Using the Gram–Schmidt process,

$$\begin{aligned}
v_1 &= 1,\\
w_1 &= \frac{1}{\left(\int_{-1}^{1} dx\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{2}},\\
v_2 &= x - \frac{1}{2}\int_{-1}^{1} x dx = x,\\
w_2 &= \frac{v_2}{\|v_2\|} = \frac{x}{\sqrt{2/3}},\\
v_3 &= x^2 - \frac{1}{2}\int_{-1}^{1} x^2 dx - \frac{x}{2/3}\int_{-1}^{1} x^3 dx = x^2 - 1/3,\\
w_3 &= \frac{x^2 - 1/3}{\|x^2 - 1/3\|} = \frac{x^2 - 1/3}{\sqrt{8/45}}.
\end{aligned}$$

Thus,

$$M = \operatorname{span}\left(\frac{1}{\sqrt{2}}, \frac{x}{\sqrt{2/3}}, \frac{x^2 - 1/3}{\sqrt{8/45}}\right).$$

(Note: w_1, w_2, and w_3 are the first three well-known *Legendre polynomials*. The entire sequence of Legendre polynomials is formed similarly.)

Now, we will find the best approximation to $f(x) = e^x \in V$ in M relative to the inner product

$$(u,v) = \int_{-1}^{1} u(x)v(x)dx.$$

We need

$$Pf = \sum_{k=1}^{3} (f, w_k)w_k.$$

Substituting the values of w_k we have just computed, we find that $(f, w_1) \approx 1.661985$, $(f, w_2) \approx 0.9011169$, and $(f, w_3) \approx 0.226302$. Thus,

$$\begin{aligned}
e^x &\approx \frac{1.661985}{\sqrt{2}} + \frac{0.9011169x}{\sqrt{2/3}} + 0.226302\left(\frac{x^2 - 1/3}{\sqrt{8/45}}\right)\\
&\approx .996293 + 1.103638x + 0.536722x^2.
\end{aligned}$$

For comparison, the Taylor series about 0 gives $e^x \approx 1 + x + x^2/2$. See the following table.

x	Pf	Taylor series	e^x
-1	0.429377	0.500	0.367879
-1/2	0.578655	0.625	0.606531
0	0.996293	1.000	1.000000
1/2	1.682293	1.625	1.648721
1	2.636654	2.500	2.718282

(2) *Fourier Series*
Let $V = C[-\pi, \pi]$ with inner product

$$(f, g) = \int_{-\pi}^{\pi} f(x)g(x)dx \text{ and } \|f\| = (f, f)^{1/2}.$$

Using

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \cos \ell x \cos mx dx = \delta_{\ell m}, \qquad \frac{1}{\pi}\int_{-\pi}^{\pi} \sin \ell x \sin mx dx = \delta_{\ell m},$$

and

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \cos \ell x \sin mx dx = 0,$$

it is readily seen that

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \ldots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin nx}{\sqrt{\pi}}$$

are orthogonal. Let $M = \text{span}\left(\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \ldots, \frac{\sin nx}{\sqrt{\pi}}\right)$ and let $f \in V$. Then the best approximation in M to $f(x)$ is $w(x)$, where

$$w(x) = \frac{a_0}{\sqrt{2\pi}} + \sum_{\ell=1}^{n}\left(a_\ell \frac{\cos \ell x}{\sqrt{\pi}} + b_\ell \frac{\sin \ell x}{\sqrt{\pi}}\right),$$

and

$$a_0 = \left(f, \frac{1}{\sqrt{2\pi}}\right), \text{ and } a_\ell = \left(f, \frac{\cos \ell x}{\sqrt{\pi}}\right), \; b_\ell = \left(f, \frac{\sin \ell x}{\sqrt{\pi}}\right),$$

for $\ell = 1, 2, \ldots, n$.

□

Before continuing, it is worthwhile to summarize our study so far: If $W \subset V$ is a finite-dimensional subspace of a vector space V consisting of "nice" (easy to work with) elements, a fundamental problem in approximation theory is

"given $v \in V$, find $w \in W$ close to v."

We saw the following.

(a) If V is a normed linear space, by Theorem 4.1, there exists a $w \in W$ such that $\|w - v\| \leq \|u - v\|$ for all $u \in W$. (However, as we will see, w may not be easy to find.)

(b) If V is an inner product space, then the best approximation is easy to find, especially if an orthonormal basis is applied.

In the remainder of this chapter, we consider specific vector spaces V, such as $C[0, 1]$, along with various choices for spaces W of nice functions such as polynomials, trigonometric functions, and continuous piecewise polynomials. Furthermore, we consider approximations in different norms or normed linear spaces.

4.3 Polynomial Approximation

The simplest choice for W is a set of polynomials. Polynomials are easy to work with and understand. Furthermore, the following Weierstrass Approximation Theorem tells us that polynomials can provide good approximations.

4.3.1 The Weierstrass Approximation Theorem

The Weierstrass approximation theorem is

THEOREM 4.4
Given $f \in C[a, b]$ and $\epsilon > 0$ there exists a polynomial $p(x)$ such that

$$\|p - f\|_\infty = \max_{a \leq x \leq b} |f(x) - p(x)| < \epsilon.$$

REMARK 4.10 This result is somewhat surprising, because f is only required to be continuous, and polynomials are smooth. For example, the graph of $f(x)$ may be as in Figure 4.3, but a polynomial $p(x)$ can be found such that $\|f - p\|_\infty < \epsilon$, no matter how small we choose ϵ. Basically, we can make the change of direction of a polynomial's graph be arbitrarily abrupt by taking the degree of the polynomial to be sufficiently high. However, as we will see, the theorem is not practical for actually finding $p(x)$. ▯

PROOF Let $x = (b - a)t + a$ for $t \in [0, 1]$. Then

$$h(t) = f\left((b - a)t + a\right)$$