6 Approximation

Definition 6.1. Given a normed vector space \( Y \) of functions and its subspace \( X \subseteq Y \). A function \( \hat{\varphi} \in X \) is called the best approximation to \( f \in Y \) from \( X \) with respect to the norm \( \| \cdot \| \) iff
\[
\forall \varphi \in X, \quad \| f - \hat{\varphi} \| \leq \| f - \varphi \|. \tag{6.1}
\]

Remark 6.1. Given \( n \) linearly independent elements \( u_1, u_2, \ldots, u_n \in X \), we are interested in a particular type of best approximation \( \hat{\varphi} = \sum_{i=1}^{n} a_i u_i \). This is called the fundamental problem of linear approximation.

Definition 6.2 (\( L^p \) functions). Let \( p > 0 \). The class of functions \( f(x) \) which are measurable and for which \( |f(x)|^p \) is Lebesgue integrable over \([a, b]\) is known as \( L^p[a, b] \). If \( p = 1 \), the class is denoted by \( L[a, b] \).

Theorem 6.3. For a weight function \( \rho(x) \in L[a, b] \), define
\[
L^2_p[a, b] := \{ f(x) : \rho(x)|f(x)|^2 \in L[a, b] \}. \tag{6.2}
\]
Then \( L^2_p[a, b] \) is a vector space. If \( \int_a^b \rho(x)dx > 0 \) and \( \forall x \in [a, b] \), \( \rho(x) \geq 0 \), then \( L^2_p[a, b] \) with
\[
\langle u, v \rangle = \int_a^b \rho(t)u(t)v(t)dt \tag{6.3}
\]
is an inner product space over \( \mathbb{R} \) and \( L^2_p[a, b] \) with
\[
\|u\|_2 = \left( \int_a^b \rho(t)|u(t)|^2dt \right)^{\frac{1}{2}} \tag{6.4}
\]
is a normed vector space over \( \mathbb{R} \).

Proof. This follows from Definitions 0.56, 0.71, and 0.73, and 0.75.

Remark 6.2. \( |u(t)|^2 \) may be different from \( u^2(t) \) because the underlying field may be the complex numbers!

Definition 6.4. Setting the norm in (6.1) to that in (6.4) yields the least square approximation on \( L^2_p[a, b] \), one particular type of best approximation on which we focus.

6.1 Orthonormal systems

Definition 6.5. A subset \( S \) of an inner product space \( X \) is called orthonormal if
\[
\forall u, v \in S, \quad \langle u, v \rangle = \begin{cases} 0 & \text{if } u \neq v, \\ 1 & \text{if } u = v. \end{cases} \tag{6.5}
\]

Example 6.3. The unit vectors in \( \mathbb{R}^n \) are orthonormal.

Example 6.4. The Chebyshev polynomials of the first kind as in Definition (3.44) are orthogonal with respect to (6.3) where \( a = -1, b = 1, \rho = \frac{1}{\sqrt{1-x^2}} \). However, they do not satisfy the second case in (6.5).

Theorem 6.6. Any finite set of nonzero orthogonal elements \( u_1, u_2, \ldots, u_n \) is linearly independent.

Proof. by contradiction using Definitions 0.62 and 6.5.

Definition 6.7. The Gram-Schmidt process takes in a finite or infinite independent list \( (u_1, u_2, \ldots) \) and output two other lists \( (v_1, v_2, \ldots) \) and \( (u_1^*, u_2^*, \ldots) \) by
\[
v_{n+1} = u_{n+1} - \sum_{k=1}^{n} \langle u_{n+1}, u_k \rangle u_k^*, \tag{6.6}
\]
\[
u_{n+1}^* = v_{n+1}/\|v_{n+1}\|, \tag{6.7}
\]
with the recursion basis as \( v_1 = u_1, u_1^* = v_1/\|v_1\| \).

Theorem 6.8. For a finite or infinite independent list \( (u_1, u_2, \ldots) \), the Gram-Schmidt process yields constants
\[
a_{11} \quad a_{21} \quad a_{22} \\
a_{31} \quad a_{32} \quad a_{33} \\
\vdots
\]
such that \( a_{kk} = \frac{1}{\|v_k\|} > 0 \) and the elements \( u_1^*, u_2^*, \ldots \)
\[
u_1^* = a_{11} u_1 \\
u_2^* = a_{21} u_1 + a_{22} u_2 \\
u_3^* = a_{31} u_1 + a_{32} u_2 + a_{33} u_3 \\
\vdots
\]
are orthonormal.

Proof. Apply the Gram-Schmidt process in Definition 6.7. It is clear that \( u_{n+1}^* \) is normal. To show \( u_{n+1}^* \) is orthogonal to \( u_n^*, u_{n-1}^*, \ldots, u_1^* \), we have
\[
\langle v_{n+1}, u_j^* \rangle = \left\langle u_{n+1} - \sum_{k=1}^{n} \langle u_{n+1}, u_k \rangle u_k^*, u_j^* \right\rangle
\]
\[
= \left\langle u_{n+1}, u_j^* \right\rangle - \sum_{k=1}^{n} \langle u_{n+1}, u_k \rangle \left( u_k^* \cdot u_j^* \right)
\]
\[
= \left\langle u_{n+1}, u_j^* \right\rangle - \left\langle u_{n+1}, u_j^* \right\rangle = 0.
\]

Finally, \( a_{kk} = \frac{1}{\|v_k\|} \) follows directly from (6.7).

Corollary 6.9. We can find constants
\[
b_{11} \quad b_{21} \quad b_{22} \\
b_{31} \quad b_{32} \quad b_{33} \\
\vdots
\]
such that \( b_{ii} > 0 \) and
\[
u_1 = b_{11} u_1^* \\
u_2 = b_{21} u_1^* + b_{22} u_2^* \\
u_3 = b_{31} u_1^* + b_{32} u_2^* + b_{33} u_3^* \\
\vdots
\]
Proof. A lower-triangular matrix with positive diagonal elements is invertible.

**Corollary 6.10.** In Theorem 6.8, we have \( \langle u_i, u_i \rangle = 0 \) for each \( i = 1, 2, \ldots, n - 1. \)

Proof. By Corollary 6.9, each \( u_i \) can be expressed as

\[
u_i = \sum_{k=1}^{i} b_{ik} u_k.
\]

Inner product the above equation with \( u_i \), apply the orthogonal conditions, and we reach the conclusion.

**Definition 6.11.** Using the Gram-Schmidt orthonormalizing process with the inner product (6.3), we obtain from the independent list of monomials \( (1, x, x^2, \ldots) \) the following classic orthonormal polynomials:

<table>
<thead>
<tr>
<th>Polynomial Type</th>
<th>( a )</th>
<th>( b )</th>
<th>( \rho(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chebyshev polynomials of the first kind</td>
<td>-1</td>
<td>1</td>
<td>( \frac{1}{\sqrt{1-x^2}} )</td>
</tr>
<tr>
<td>Chebyshev polynomials of the second kind</td>
<td>-1</td>
<td>1</td>
<td>( \sqrt{1-x^2} )</td>
</tr>
<tr>
<td>Legendre polynomials</td>
<td>-1</td>
<td>1</td>
<td>( (1-x^2)^{\frac{1}{2}} )</td>
</tr>
<tr>
<td>Laguerre polynomials</td>
<td>0</td>
<td>+\infty</td>
<td>( x^\alpha e^{-x} )</td>
</tr>
<tr>
<td>Hermite polynomials</td>
<td>-\infty</td>
<td>+\infty</td>
<td>( e^{-x^2} )</td>
</tr>
</tbody>
</table>

where \( \alpha, \beta > -1 \) for Jacobi polynomials and \( \alpha > -1 \) for Laguerre polynomials.

**Example 6.5.** We compute the first 3 Legendre polynomials using the Gram-Schmidt process.

\( u_1 = 1, v_1 = 1, \|v_1\|^2 = \int_{-1}^{1} \|x\| = 2, u_1^* = \frac{1}{\sqrt{2}}. \)

\( u_2 = x, v_2 = x - \left( x, \frac{1}{\sqrt{2}} \right) \frac{1}{\sqrt{2}} = x, \|v_2\|^2 = \frac{2}{3}. \)

\( u_2^* = \sqrt{\frac{3}{2}} x. \)

\( u_3 = x^2 - \left( x^2, \sqrt{\frac{3}{2}} x \right) \sqrt{\frac{3}{2}} x - \left( x^2, \frac{1}{\sqrt{2}} \right) \frac{1}{\sqrt{2}} = x^2 - \frac{1}{3}. \)

\( \|v_3\|^2 = \int_{-1}^{1} \left( x^2 - \frac{1}{3} \right) dx = \frac{2}{15} \sqrt{10}, \)

\( u_3^* = \frac{3}{4} \sqrt{10} \left( x^2 - \frac{1}{3} \right). \)

**6.2 Fourier (or orthogonal) expansions**

**Definition 6.12.** Let \( (u_1^*, u_2^*, \ldots) \) be a finite or infinite orthonormal list. The *orthogonal expansion* or Fourier expansion for an arbitrary \( w \) is the series

\[
w \sim \sum_{n} \langle w, u_n^* \rangle u_n^*.
\]

where the constants \( \langle w, u_n^* \rangle \) are known as the *Fourier coefficients* of \( w \) and the term \( \langle w, u_n^* \rangle u_n^* \) the projection of \( w \) on \( u_n^* \). The error of the Fourier expansion of \( w \) with respect to \( (u_1^*, u_2^*, \ldots) \) is simply \( \sum_{n} \langle w, u_n^* \rangle u_n^* - w. \)

**Remark 6.6.** We do not write “=” in (6.10) because for an infinite sequence the series might not converge.

**Example 6.7.** With the Euclidean inner product in Definition 0.72, we select orthonormal vectors in \( \mathbb{R}^3 \) as

\( u_1^* = (1, 0, 0)^T, u_2^* = (0, 1, 0)^T, u_3^* = (0, 0, 1)^T. \)

For the vector \( w = (a, b, c)^T \), the Fourier coefficients are

\[ \langle w, u_1^* \rangle = a, \langle w, u_2^* \rangle = b, \langle w, u_3^* \rangle = c, \]

and the projections of \( w \) onto \( u_1^* \) and \( u_2^* \) are

\[ \langle w, u_1^* \rangle u_1^* = (a, 0, 0)^T, \langle w, u_2^* \rangle u_2^* = (0, b, 0)^T. \]

The Fourier expansion of \( w \) is

\[ w \sim \sum_i \langle w, u_i^* \rangle u_i^* = \sum_i \langle w, u_i \rangle u_i. \]

with the error of Fourier expansion as 0; see Theorem 6.13.

**Remark 6.8.** For an orthogonal list of vectors \( u_i \), we can construct an orthonormal list by \( u_i^* = u_i / \|u_i\| \) to arrive at the Fourier expansion

\[ w \sim \sum_i \langle w, u_i^* \rangle u_i^* = \sum_i \langle w, u_i \rangle u_i. \]

**Example 6.9.** With the following orthonormal list

\[ 1, \sin x, \cos x, \frac{1}{\sqrt{2\pi}}, \frac{\sin(nx)}{\sqrt{\pi}}, \frac{\cos(nx)}{\sqrt{\pi}}, \ldots, \]

in \( L^2([-\pi, \pi]) \), we obtain from Definition 6.12 the Fourier series of a function \( f(x) \),

\[ f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{+\infty} \left( a_k \cos kx + b_k \sin kx \right). \]

where

\[ a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx dx, \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx dx. \]

**Theorem 6.13.** Let \( u_1, u_2, \ldots, u_n \) be linearly independent and let \( u_i^* \) be the \( u_i \)'s orthonormalized by the Gram-Schmidt process. If \( w = \sum_{i=1}^{n} a_i u_i \), then

\[ w = \sum_{i=1}^{n} \langle w, u_i^* \rangle u_i^*. \]

i.e. \( w \) is equal to its Fourier expansion.

Proof. By the condition \( w = \sum_{i=1}^{n} a_i u_i \) and Corollary 6.9, we can express \( w \) as a linear combination of \( u_i^* \)s.

\[ w = \sum_{i=1}^{n} c_i u_i^*. \]

Then the orthogonality of \( u_i^* \) implies

\[ \forall k = 1, 2, \ldots, n, \quad \langle u_i^*, w \rangle = c_i, \]

which completes the proof.
Theorem 6.14 (Minimum properties of Fourier expansions). Let $u_1^*, u_2^*, \ldots$ be an orthonormal system and let $w$ be arbitrary. Then

$$\left\| w - \sum_{i=1}^{N} (w, u_i^*) u_i^* \right\| \leq \left\| w - \sum_{i=1}^{N} a_i u_i^* \right\|,$$  
(6.14)

for any selection of constants $a_1, a_2, \ldots, a_N$.

Proof. With the shorthand notation $\sum_i = \sum_{i=1}^{N}$, we deduce from Definition 0.71 and properties of inner products

$$\left\| w - \sum_i a_i u_i^* \right\|^2 = \left\langle w, w \right\rangle - \left\langle w, \sum_i a_i u_i^* \right\rangle - \left\langle \sum_i a_i u_i^*, w \right\rangle \geq \left\langle \sum_i a_i u_i^*, \sum_i a_i u_i^* \right\rangle - \left\langle \sum_i a_i u_i^*, w \right\rangle = \left\langle w, w \right\rangle - \sum_i |a_i|^2 - \sum_i \left\langle u_i^*, w \right\rangle \left\langle u_i^*, u_i^* \right\rangle$$

$$= ||w||^2 - \sum_i |\langle w, u_i^* \rangle|^2 - \sum_i |a_i - \langle w, u_i^* \rangle|^2,$$  
(6.15)

where $|\cdot|$ denotes the modulus of a complex number. The first two terms are independent of $a_i$. Therefore $\|w - \sum_i a_i u_i^*\|^2$ is minimized only when $a_i = \langle w, u_i^* \rangle$. □

Corollary 6.15. Let $(u_1, u_2, \ldots, u_n)$ be an independent list. The fundamental problem of linearly approximating an arbitrary vector $w$ is solved by the best approximation $\hat{\varphi} = \sum_k \langle w, u_k^* \rangle u_k$ where $u_k$’s are the $u_k$’s orthonormalized by the Gram-Schmidt process.

$$\left\| w - \hat{\varphi} \right\|^2 := \min_{a_k} \left\| w - \sum_{k=1}^{n} a_k u_k \right\|^2 = \|w\|^2 - \sum_{k=1}^{n} \left\langle w, u_k^* \right\rangle^2.$$

(6.16)

Proof. This follows directly from (6.15). □

Corollary 6.16 (Bessel inequality). If $u_1^*, u_2^*, \ldots, u_N^*$ are orthonormal, then, for an arbitrary $w$,

$$\sum_{i=1}^{N} \left| \langle w, u_i^* \rangle \right|^2 \leq \|w\|^2.$$  
(6.17)

Proof. This follows directly from Corollary 6.15 and the real positivity of a norm. □

Corollary 6.17. The Gram-Schmidt process in Definition 6.7 satisfies

$$\forall n \in \mathbb{N}^+, \left\| v_{n+1} \right\|^2 = \|u_{n+1}\|^2 - \sum_{k=1}^{n} \left| \langle u_{n+1}, u_k^* \rangle \right|^2.$$  
(6.18)

Proof. By (6.6), each $v_{n+1}$ can be regarded as the error of Fourier expansion of $u_{n+1}$ with respect to the orthonormal list $(u_1^*, u_2^*, \ldots, u_n^*)$. In Corollary 6.15, identifying $w$ with $u_{n+1}$ completes the proof. □

Remark 6.10. The Fourier expansion helps our understanding of the Gram-Schmidt process in two aspects:

- Definition 6.7 is a special type of Fourier expansion; this helps the memorization.
- As a complement to Definition 6.7 and Theorem 6.8, Corollary 6.17 reveals additional relations between vectors $u_k, v_k, \text{and } u_k^*$.

Example 6.11. The least square approximation problem of $e^x$ on $[−1, 1]$ with a linear and quadratic polynomial with the weight function $\rho = 1$.

First note that the problem is equivalent to

$$\min_{a_i} \int_{-1}^{1} \left( e^x - \sum_{i=0}^{n} a_i x^i \right)^2 \, dx$$  
(6.19)

for $n = 1, 2$. Use the Legendre polynomials derived in Example 6.5:

$$u_1^* = \frac{1}{\sqrt{2}}, \quad u_2^* = \sqrt{\frac{3}{2}} x, \quad u_3^* = \frac{1}{\sqrt{10}}(3x^2 - 1).$$

The Fourier coefficients of $e^x$ are

$$b_0 = \int_{-1}^{1} \frac{1}{\sqrt{2}} e^x \, dx = \frac{1}{\sqrt{2}} \left( e^1 - 1 \right),$$
$$b_1 = \int_{-1}^{1} \sqrt{\frac{3}{2}} e^x x \, dx = \sqrt{6} e^{-1},$$
$$b_2 = \int_{-1}^{1} \frac{1}{\sqrt{10}}(3x^2 - 1) e^x \, dx = \frac{\sqrt{10}}{2} \left( e - \frac{7}{e} \right).$$

The minimizing polynomials are thus

$$\hat{\varphi}_n = \left\{ \begin{array}{ll}
\frac{1}{2} (e^2 - 1) + \frac{e}{2} x & n = 1; \\
\frac{1}{2} (e^2 - 1) x + \frac{2}{3} x & n = 2.
\end{array} \right.$$  
(6.20)

This illustrates the permanence advantage of orthonormal systems in solving least square approximation: Fourier coefficients for the best approximation of a lower degree can be reused in constructing that of a higher degree.

6.3 The normal equations

Remark 6.12. With an orthonormal list, we can easily find the best linear approximation of a given function. If the list is not orthonormal but only independent, one way to the best approximation is via the normal equations.

Theorem 6.18. Let $u_1, u_2, \ldots, u_n \in X$ be linearly independent and let $u_k^*$ be the $u_k$’s orthonormalized by the Gram-Schmidt process. Then, for any element $w$,

$$\forall j = 1, 2, \ldots, n, \quad \left( w - \sum_{k=1}^{n} \langle w, u_k^* \rangle u_k^* \right) \perp u_j^*.$$  
(6.21)

where “$\perp$” denotes orthogonality.
Proof. Take the inner product of the two vectors and apply the conditions on orthonormal systems.

Remark 6.13. \( w \) may or may not be in \( X \).

Corollary 6.19. Let \( u_1, u_2, \ldots, u_n \in X \) be linearly independent. If \( \varphi = \sum_{k=1}^{n} a_k u_k \) is the best linear approximant to \( w \), then

\[
\forall j = 1, 2, \ldots, n, \quad \left( w - \sum_{k=1}^{n} a_k u_k \right) \perp u_j, \tag{6.22}
\]

Proof. Since \( \varphi = \sum_{k=1}^{n} a_k u_k \) is the best linear approximant to \( w \), Theorem 6.14 implies that

\[
\sum_{k=1}^{n} a_k u_k = \sum_{k=1}^{n} \langle w, u_k^* \rangle u_k.
\]

Corollary 6.9 and Theorem 6.18 complete the proof.

Definition 6.20. Let \( u_1, u_2, \ldots, u_n \) be a sequence of elements in an inner product space. The \( n \times n \) matrix

\[
G = G(u_1, u_2, \ldots, u_n) = (\langle u_i, u_j \rangle)
\]

is the Gram matrix of \( u_1, u_2, \ldots, u_n \). Its determinant

\[
g = g(u_1, u_2, \ldots, u_n) = \det(\langle u_i, u_j \rangle) \tag{6.24}
\]

is the Gram determinant of the elements \( u_i \)’s.

Lemma 6.21. Let \( w_i = \sum_{j=1}^{n} a_{ij} u_j \) for \( i = 1, 2, \ldots, n \). Let \( A = (a_{ij}) \) and its conjugate transpose \( A^H = (\bar{a}_{ji}) \). Then we have

\[
G(w_1, w_2, \ldots, w_n) = AG(u_1, u_2, \ldots, u_n)A^H \tag{6.25}
\]

and

\[
g(w_1, w_2, \ldots, w_n) = |\det A|^2 g(u_1, u_2, \ldots, u_n) \tag{6.26}
\]

Proof. The inner product of \( u_i \) and \( w_j \) yields

\[
\begin{bmatrix}
\langle u_1, u_1 \rangle & \langle u_1, u_2 \rangle & \cdots & \langle u_1, u_n \rangle \\
\langle u_2, u_1 \rangle & \langle u_2, u_2 \rangle & \cdots & \langle u_2, u_n \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle u_n, u_1 \rangle & \langle u_n, u_2 \rangle & \cdots & \langle u_n, u_n \rangle
\end{bmatrix}
= G(u_1, u_2, \ldots, u_n)A^H,
\]

Therefore (6.25) holds since

\[
G(w_1, w_2, \ldots, w_n) = \begin{bmatrix}
\langle w_1, w_1 \rangle & \langle w_1, w_2 \rangle & \cdots & \langle w_1, w_n \rangle \\
\langle w_2, w_1 \rangle & \langle w_2, w_2 \rangle & \cdots & \langle w_2, w_n \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle w_n, w_1 \rangle & \langle w_n, w_2 \rangle & \cdots & \langle w_n, w_n \rangle
\end{bmatrix}
= AG(u_1, u_2, \ldots, u_n)A^H.
\]

Then (6.26) follows from taking the determinant of (6.25) and applying the identities \( \det(AB) = \det(A) \det(B) \) and \( \det(A) = \det(A^T) \).

Theorem 6.22. For nonzero elements \( u_1, u_2, \ldots, u_n \),

\[
0 \leq g(u_1, u_2, \ldots, u_n) \leq \prod_{k=1}^{n} \| u_k \|^2, \tag{6.27}
\]

where the lower equality holds if and only if \( u_1, u_2, \ldots, u_n \) are linearly dependent and the upper equality holds if and only if they are orthogonal.

Proof. Suppose \( u_1, u_2, \ldots, u_n \) are linearly dependent. Then we can find constants \( c_1, c_2, \ldots, c_n \) such that \( \sum_{i=1}^{n} c_i u_i = 0 \) with at least one constant \( c_j \) being nonzero. Construct vectors

\[
w_k = \begin{bmatrix}
\sum_{i=1}^{n} c_i u_i = 0, & k = j; \\
0, & k \neq j.
\end{bmatrix}
\]

We have \( g(w_1, w_2, \ldots, w_n) = 0 \) because \( \langle w_j, w_k \rangle = 0 \) for each \( k \). By the Laplace theorem, we expand the determinant of \( C = (c_{ij}) \) according to minors of its \( j \)th row:

\[
det(C) = \det \begin{bmatrix}
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
c_1 & c_2 & \cdots & c_j & c_n \\
0 & 0 & \cdots & 0 & 1
\end{bmatrix}
= 0 + \cdots + 0 + c_j + 0 + \cdots + 0 = c_j \neq 0.
\]

Then Lemma 6.21 yields \( g(u_1, u_2, \ldots, u_n) = 0 \).

Now suppose \( u_1, u_2, \ldots, u_n \) are linearly independent. Theorem 6.8 yields constants \( a_{ij} \) such that \( a_{kk} > 0 \) and the following vectors are orthonormal:

\[
u_k^* = \sum_{i=1}^{k} a_{ik} u_i.
\]

Then Definition 6.20 implies \( g(u_1^*, u_2^*, \ldots, u_n^*) = 1 \). Also, we have \( \det(a_{ij}) = \prod_{i=1}^{n} a_{ii} \) because the matrix \( (a_{ij}) \) is triangular. It then follows from Lemma 6.21 that

\[
g(u_1, u_2, \ldots, u_n) = \prod_{i=k}^{n} \frac{1}{a_{kk}^*} > 0. \tag{6.28}
\]

The above arguments show that \( g(u_1, u_2, \ldots, u_n) = 0 \) if and only if \( u_1, u_2, \ldots, u_n \) are linearly dependent.

51
Suppose \( u_1, u_2, \ldots, u_n \) are orthogonal. By Definition 6.20, \( G(u_1, u_2, \ldots, u_n) \) is a diagonal matrix with \( \|u_k\|^2 \) on the diagonals. Hence the orthogonality of \( u_k \)'s implies
\[
g(u_1, u_2, \ldots, u_n) = \prod_{k=1}^n \|u_k\|^2. \tag{6.29}
\]
For the converse statement, suppose (6.29) holds. From Theorem 6.8 we know \( \frac{1}{\|u_k\|} = \|v_k\| \). Then (6.28) and (6.29) yield
\[
\forall k = 1, 2, \ldots, n, \quad \|u_k\|^2 = \|v_k\|^2. \tag{6.30}
\]
Then Corollary 6.17 and (6.30) imply
\[
\forall k = 1, 2, \ldots, n, \quad \sum_{j=1}^{k-1} \langle u_k, u_j^* \rangle^2 = 0,
\]
which further implies
\[
\forall k = 1, 2, \ldots, n, \forall j = 1, 2, \ldots, k-1, \quad \langle u_k, u_j^* \rangle = 0,
\]
which, together with Corollary 6.9, implies the orthogonality of \( u_k \)'s. Finally, we remark that the maximum of \( g(u_1, u_2, \ldots, u_n) \) is indeed \( \prod_{k=1}^n \|u_k\|^2 \) because of (6.28), \( \frac{1}{\|u_k\|} = \|v_k\| \), and Corollary 6.17.


Proof. Consider a nonzero vector \( x = (x_1, x_2, \ldots, x_n)^T \) with each \( x_i \) in the field that underlies \( \text{span}(u_1, \ldots, u_n) \). Let \( x^H \) denote the conjugate transpose of \( x \). By properties of inner products in Definition 0.71, we have
\[
xG(u_1, u_2, \ldots, u_n)x^H = \sum_{i=1}^n \sum_{j=1}^n x_i \langle u_i, u_j \rangle x_j = \sum_{i=1}^n \sum_{j=1}^n x_i u_i x_j u_j = \left\| \sum_{i=1}^n x_i u_i \right\|^2 \geq 0,
\]
where, by (NRM-2) in Theorem 0.78, the equality holds only when \( \sum_{i=1}^n x_i u_i = 0 \), i.e., the \( u_i \)'s are linearly dependent. Hence \( G(u_1, u_2, \ldots, u_n) \) is positive definite if and only if \( (u_1, u_2, \ldots, u_n) \) is linearly independent. From linear algebra, we know that the determinant of a positive definite matrix is positive. The above argument proves the first inequality in (6.27).

For a positive definite matrix \( A \) satisfying \( A^H = A \), consider its block form,
\[
A = \begin{bmatrix} a_{11} & b^H \\ b & A_{n-1} \end{bmatrix}.
\]
Let \( I \) denote an identity matrix. Define
\[
P = \begin{bmatrix} 1 & 0^T \\ -A_{n-1}^{-1}b & I \end{bmatrix}.
\]
Since \( \det(P) = 1 \), the determinant of \( A \) is the same as
\[
P^H AP = \begin{bmatrix} a_{11} - b^H A_{n-1}^{-1}b & 0^T \\ 0 & A_{n-1} \end{bmatrix}.
\]
Without changing the determinant of \( A \), we can repeat the above process to diagonalize \( A \). Consequently, we have
\[
det(A) = \prod_{i=1}^n (a_{ii} - \alpha_i), \tag{6.31}
\]
where \( \alpha_i \) is the \( i \)th diagonal entry of \( A \), \( \alpha_n = 0 \), and the positive-definiteness of \( A \) implies that \( \alpha_i > 0 \) for \( i = 1, 2, \ldots, n - 1 \).

Apply the arguments in the previous paragraph to \( G(u_1, u_2, \ldots, u_n) \) and we have
\[
g(u_1, u_2, \ldots, u_n) \leq \prod_{k=1}^n \|u_k\|^2,
\]
where the equality holds only when all off-diagonal entries in \( G(u_1, u_2, \ldots, u_n) \) are zero, i.e. \( (u_1, u_j) = 0 \) for all \( i \neq j \), i.e. \( (u_1, u_2, \ldots, u_n) \) is orthogonal. Conversely, if \( (u_1, u_2, \ldots, u_n) \) is orthogonal, then \( G(u_1, u_2, \ldots, u_n) \) is clearly a diagonal matrix with \( \|u_k\|^2 \) on the diagonal.

Theorem 6.23. Let \( \bar{\phi} = \sum_{i=1}^n a_i u_i \) be the best approximation to \( w \) constructed from the independent list \( (u_1, u_2, \ldots, u_n) \). Then the coefficients
\[
a = [a_1, a_2, \ldots, a_n]^T
\]
are uniquely determined from the linear system of normal equations,
\[
G(u_1, u_2, \ldots, u_n)a = c, \tag{6.32}
\]
where \( c = [\langle w, u_1 \rangle, \langle w, u_2 \rangle, \ldots, \langle w, u_n \rangle]^T \).

Proof. Corollary 6.19 yields
\[
\langle w, u_j \rangle = \sum_{k=1}^n a_k \langle u_k, u_j \rangle,
\]
which is simply the \( j \)th equation of (6.32). The uniqueness of the coefficients follows from Theorems 6.22 and Cramer's rule.

Example 6.15. Solve Example 6.11 by normal equations.

To find the best approximation \( \bar{\phi} = a_0 + a_1 x + a_2 x^2 \) to \( e^x \) from the linearly independent list \( (1, x, x^2) \), we first construct the Gram matrix from (6.23), (6.3), and \( \rho = 1 \):
\[
G(1, x, x^2) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & x & x^2 \\ 1 & x^2 & x^3 \end{bmatrix} = \begin{bmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{5} \end{bmatrix}.
\]
We then calculate the vector
\[
c = \begin{bmatrix} e^x, 1 \\ e^x, x \\ e^x, x^2 \end{bmatrix} = \begin{bmatrix} e - 1/e \\ 2/e \\ e - 5/e \end{bmatrix}.
\]
The normal equations then yields
\[
a_0 = \frac{3(11 - e^2)}{4e}, \quad a_1 = \frac{3}{e}, \quad a_2 = \frac{15(e^2 - 7)}{4e}.
\]
With these values, it is easily verified that the best approximation \( \bar{\phi} = a_0 + a_1 x + a_2 x^2 \) equals that in (6.20).
Remark 6.16. Comparing Examples 6.15 and 6.11, we observe several advantages of the Fourier-expansion (FE) approach over the normal-equations (NE) approach:

- the number of calculating inner products in FE is only $O(n)$ while that in NE is $O(n^2)$;
- the coefficients in FE have the property of permanence while those in NE change for a different $n$ (verify this by solving the normal equations with $n = 2$);
- FE is well-conditioned while NE is ill-conditioned for large $n$.

### 6.4 Discrete least squares

**Definition 6.24.** Define a function $\lambda : \mathbb{R} \to \mathbb{R}$

$$
\lambda(t) = \begin{cases} 
0 & \text{if } t \in (-\infty, a), \\
\int_{a}^{b} \rho(\tau) d\tau & \text{if } t \in [a, b], \\
\int_{b}^{\infty} \rho(\tau) d\tau & \text{if } t \in (b, +\infty).
\end{cases}
$$

Then a corresponding **continuous measure** $d\lambda$ can be defined as

$$
d\lambda = \begin{cases} 
\rho(t) dt & \text{if } t \in [a, b], \\
0 & \text{otherwise}, \end{cases}
$$

where the **support** of the continuous measure $d\lambda$ is the interval $[a, b]$.

**Definition 6.25.** The **discrete measure** or the **Dirac measure** associated with the point set $\{t_1, t_2, \ldots, t_N\}$ is a measure $d\lambda$ that is nonzero only at the points $t_i$ and has the value $\rho_i$ there. The **support** of the discrete measure is the set $\{t_1, t_2, \ldots, t_N\}$.

**Definition 6.26.** The **Heaviside function** is the truncated power function with exponent 0,

$$
H(x) = x_+^0 = \begin{cases} 
1 & \text{if } x \geq 0, \\
0 & \text{if } x < 0.
\end{cases}
$$

**Remark 6.17.** The **Dirac Delta function**, $\delta(x)$, is roughly a generalized function that satisfies

$$
\delta(x) = \begin{cases} 
+\infty & x = 0, \\
0 & x \neq 0.
\end{cases}
$$

Note: the above definition of $\delta(x)$ is heuristic. A rigorous one should employ the concept of measures.

Useful properties of $\delta(x)$ include

$$
\int_{-\infty}^{+\infty} \delta(x) dx = 1, 
$$

$$
\int_{0}^{x} \delta(t) dt = H(x), 
$$

$$
\int_{-\infty}^{+\infty} f(t) \delta(t - t_0) dt = f(t_0).
$$

**Lemma 6.27.** For a function $u : \mathbb{R} \to \mathbb{R}$, define

$$
\lambda(t) = \sum_{i=1}^{N} \rho_i H(t - t_i),
$$

and we have

$$
\int_{\mathbb{R}} u(t) d\lambda = \sum_{i=1}^{N} \rho_i u(t_i).
$$

Proof. (6.40), (6.38), and (6.39) yield

$$
\int_{\mathbb{R}} u(t) d\lambda = \int_{\mathbb{R}} \sum_{i=1}^{N} \rho_i \delta(t - t_i) u(t) dt = \sum_{i=1}^{N} \rho_i u(t_i).
$$

□

**Remark 6.18.** By Definition 6.24 and Lemma 6.27, we can solve the discrete least square problem by reusing the procedures in Examples 6.11 and 6.15, but the definition of inner product should be switched to

$$
\langle u(t), v(t) \rangle = \sum_{i=1}^{N} \rho_i u(t_i) v(t_i),
$$

which follows directly from (6.41).

**Example 6.19.** Consider a table of sales record.

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>250</td>
<td>201</td>
<td>159</td>
<td>61</td>
<td>77</td>
<td>40</td>
</tr>
<tr>
<td>x</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>y</td>
<td>17</td>
<td>25</td>
<td>103</td>
<td>156</td>
<td>222</td>
<td>345</td>
</tr>
</tbody>
</table>

From the plot of the discrete data, it appears that a quadratic polynomial would be a good fit. Hence we formulate the least square problem as finding the coefficients of a quadratic polynomial to minimize the following error,

$$
\sum_{i=1}^{12} \left( y_i - \sum_{j=0}^{2} a_j x_j^2 \right)^2.
$$

Reusing the procedures in Example 6.15, we have

$$
G(1, x, x^2) = \begin{bmatrix} 
\langle 1, 1 \rangle & \langle 1, x \rangle & \langle 1, x^2 \rangle \\
\langle x, 1 \rangle & \langle x, x \rangle & \langle x, x^2 \rangle \\
\langle x^2, 1 \rangle & \langle x^2, x \rangle & \langle x^2, x^2 \rangle 
\end{bmatrix} = \begin{bmatrix} 
12 & 78 & 650 \\
78 & 650 & 6084 \\
650 & 6084 & 60710
\end{bmatrix},
$$

$$
c = \begin{bmatrix} 
\langle y, 1 \rangle \\
\langle y, x \rangle \\
\langle y, x^2 \rangle 
\end{bmatrix} = \begin{bmatrix} 
\sum_{i=1}^{12} y_i \\
\sum_{i=1}^{12} y_i x_i \\
\sum_{i=1}^{12} y_i x_i^2 
\end{bmatrix} = \begin{bmatrix} 
1662 \\
11392 \\
109750
\end{bmatrix}.
$$

Then the normal equations yield

$$
a = G^{-1} c = [386.00, -113.43, 9.04]^T.
$$

The corresponding polynomial is plotted in the figure.