MAP 2220 – FUNDAMENTOS DE ANÁLISE NUMÉRICA

2º Semestre - 2017

Prof. Dr. Luis Carlos de Castro Santos

lsantos@ime.usp.br/lccs13@yahoo.com

8 Approximation Theory 497

Polynomial Least Squares

The general problem of approximating a set of data, $\{(x_i, y_i) \mid i = 1, 2, \ldots, m\}$, with an algebraic polynomial

$$P_n(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

of degree $n < m - 1$, using the least squares procedure is handled similarly. We choose the constants $a_0, a_1, \ldots, a_n$ to minimize the least squares error $E = E_2(a_0, a_1, \ldots, a_n)$, where

$$\begin{aligned} E &= \sum_{i=1}^{m} (y_i - P_n(x_i))^2 \\ &= \sum_{i=1}^{m} y_i^2 - 2\sum_{i=1}^{m} P_n(x_i)y_i + \sum_{i=1}^{m} (P_n(x_i))^2 \end{aligned}$$

$$= \sum_{i=1}^{m} y_i^2 - 2\sum_{i=1}^{m}\left(\sum_{j=0}^{n} a_j x_i^j\right) y_i + \sum_{i=1}^{m}\left(\sum_{j=0}^{n} a_j x_i^j\right)^2$$

$$= \sum_{i=1}^{m} y_i^2 - 2\sum_{j=0}^{n} a_j \left(\sum_{i=1}^{m} y_i x_i^j\right) + \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k \left(\sum_{i=1}^{m} x_i^{j+k}\right).$$

As in the linear case, for E to be minimized it is necessary that $\partial E/\partial a_j = 0$, for each $j = 0, 1, \ldots, n$. Thus, for each j, we must have

$$0 = \frac{\partial E}{\partial a_j} = -2\sum_{i=1}^{m} y_i x_i^j + 2\sum_{k=0}^{n} a_k \sum_{i=1}^{m} x_i^{j+k}.$$

This gives $n + 1$ **normal equations** in the $n + 1$ unknowns a_j. These are

$$\sum_{k=0}^{n} a_k \sum_{i=1}^{m} x_i^{j+k} = \sum_{i=1}^{m} y_i x_i^j, \quad \text{for each } j = 0, 1, \ldots, n. \tag{8.3}$$

It is helpful to write the equations as follows:

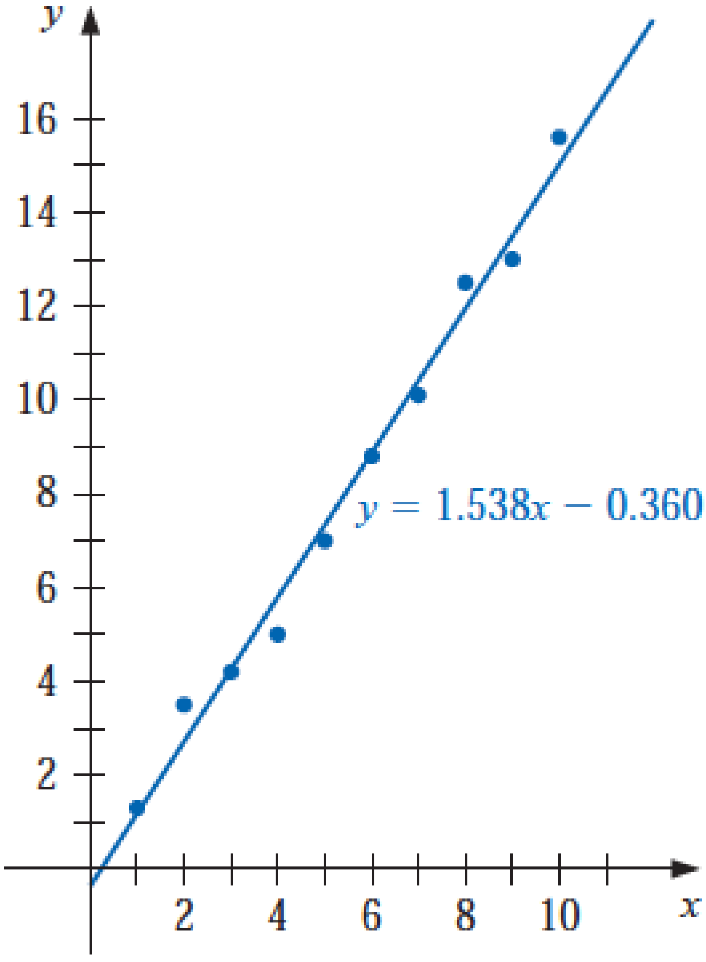
$$a_0 \sum_{i=1}^{m} x_i^0 + a_1 \sum_{i=1}^{m} x_i^1 + a_2 \sum_{i=1}^{m} x_i^2 + \cdots + a_n \sum_{i=1}^{m} x_i^n = \sum_{i=1}^{m} y_i x_i^0,$$

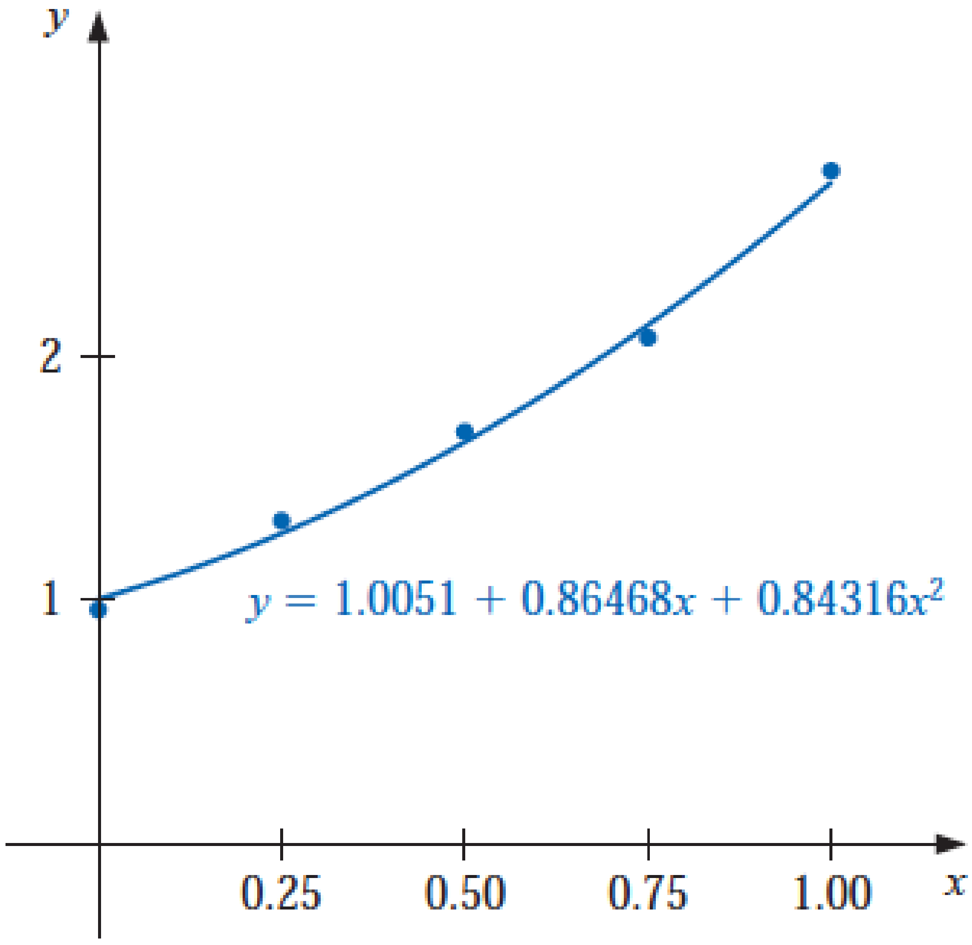
$$a_0 \sum_{i=1}^{m} x_i^1 + a_1 \sum_{i=1}^{m} x_i^2 + a_2 \sum_{i=1}^{m} x_i^3 + \cdots + a_n \sum_{i=1}^{m} x_i^{n+1} = \sum_{i=1}^{m} y_i x_i^1,$$

$$\vdots$$

$$a_0 \sum_{i=1}^{m} x_i^n + a_1 \sum_{i=1}^{m} x_i^{n+1} + a_2 \sum_{i=1}^{m} x_i^{n+2} + \cdots + a_n \sum_{i=1}^{m} x_i^{2n} = \sum_{i=1}^{m} y_i x_i^n.$$

These *normal equations* have a unique solution provided that the x_i are distinct (see Exercise 14).


y
16
14
12
10
8
6
4
2
2
4
6
8
10
x
y = 1.538x − 0.360


y
2
1
0.25
0.50
0.75
1.00
x
y = 1.0051 + 0.86468x + 0.84316x²

8.2 Orthogonal Polynomials and Least Squares Approximation

The previous section considered the problem of least squares approximation to fit a collection of data. The other approximation problem mentioned in the introduction concerns the approximation of functions.

Suppose $f \in C[a, b]$ and that a polynomial $P_n(x)$ of degree at most n is required that will minimize the error

$$\int_a^b [f(x) - P_n(x)]^2 \, dx.$$

To determine a least squares approximating polynomial; that is, a polynomial to minimize this expression, let

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = \sum_{k=0}^{n} a_k x^k,$$

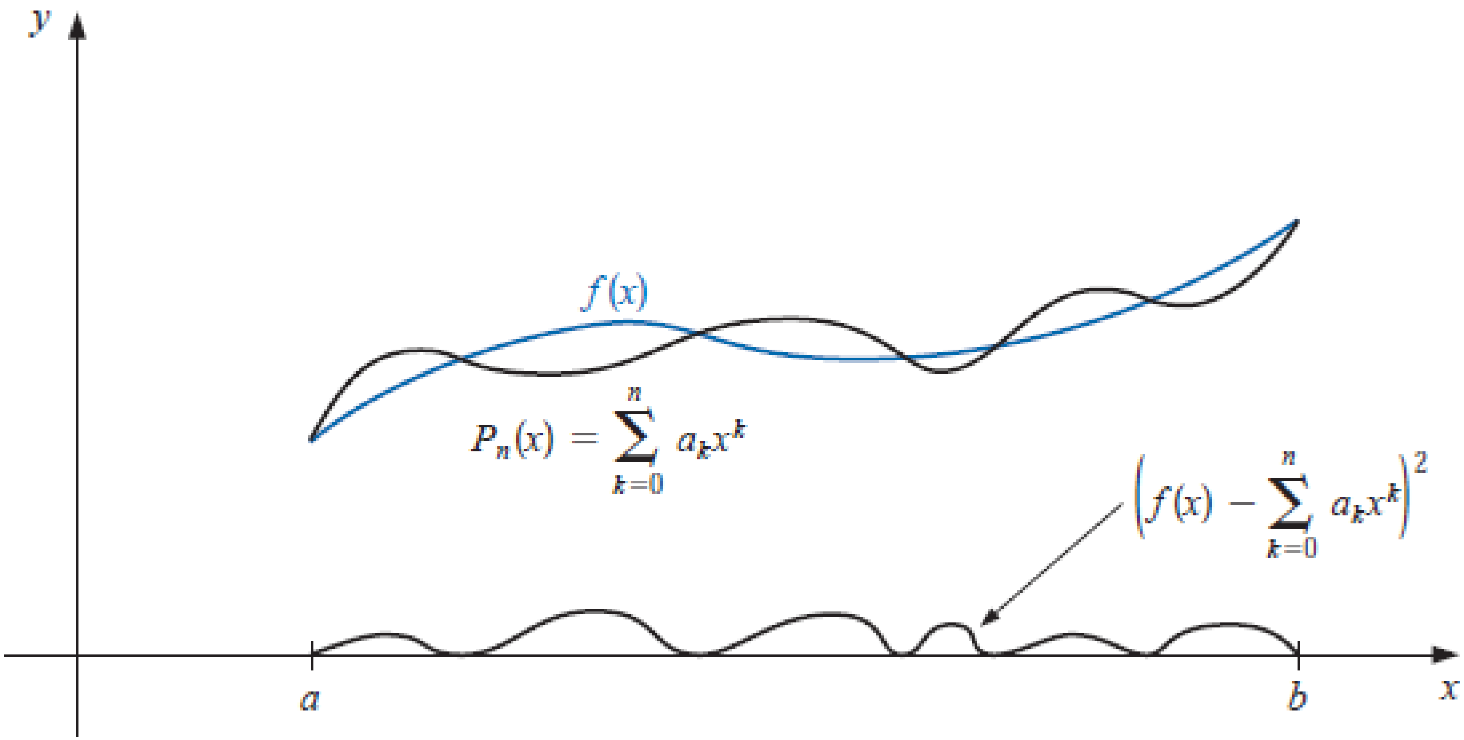
and define, as shown in Figure 8.6,

$$E \equiv E_2(a_0, a_1, \ldots, a_n) = \int_a^b \left(f(x) - \sum_{k=0}^{n} a_k x^k \right)^2 dx.$$

The problem is to find real coefficients $a_0, a_1, \ldots, a_n$ that will minimize E. A necessary condition for the numbers $a_0, a_1, \ldots, a_n$ to minimize E is that

$$\frac{\partial E}{\partial a_j} = 0, \quad \text{for each } j = 0, 1, \ldots, n.$$

Figure 8.6



Since

$$E = \int_a^b [f(x)]^2\,dx - 2\sum_{k=0}^{n} a_k \int_a^b x^k f(x)\,dx + \int_a^b \left(\sum_{k=0}^{n} a_k x^k\right)^2 dx,$$

we have

$$\frac{\partial E}{\partial a_j} = -2\int_a^b x^j f(x)\,dx + 2\sum_{k=0}^{n} a_k \int_a^b x^{j+k}\,dx.$$

Hence, to find $P_n(x)$, the $(n+1)$ linear **normal equations**

$$\sum_{k=0}^{n} a_k \int_a^b x^{j+k}\,dx = \int_a^b x^j f(x)\,dx, \quad \text{for each } j = 0, 1, \ldots, n, \tag{8.6}$$

must be solved for the $(n+1)$ unknowns a_j. The normal equations always have a unique solution provided that $f \in C[a,b]$. (See Exercise 15.)

Example 1

Find the least squares approximating polynomial of degree 2 for the function $f(x) = \sin \pi x$ on the interval [0, 1].

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = \sum_{k=0}^{n} a_k x^k,$$

$$\sum_{k=0}^{n} a_k \int_a^b x^{j+k}\, dx = \int_a^b x^j f(x)\, dx, \quad \text{for each } j = 0, 1, \ldots, n,$$

Solution The normal equations for $P_2(x) = a_2x^2 + a_1x + a_0$ are

$$a_0 \int_0^1 1\,dx + a_1 \int_0^1 x\,dx + a_2 \int_0^1 x^2\,dx = \int_0^1 \sin \pi x\,dx,$$

$$a_0 \int_0^1 x\,dx + a_1 \int_0^1 x^2\,dx + a_2 \int_0^1 x^3\,dx = \int_0^1 x \sin \pi x\,dx,$$

$$a_0 \int_0^1 x^2\,dx + a_1 \int_0^1 x^3\,dx + a_2 \int_0^1 x^4\,dx = \int_0^1 x^2 \sin \pi x\,dx.$$

Performing the integration yields

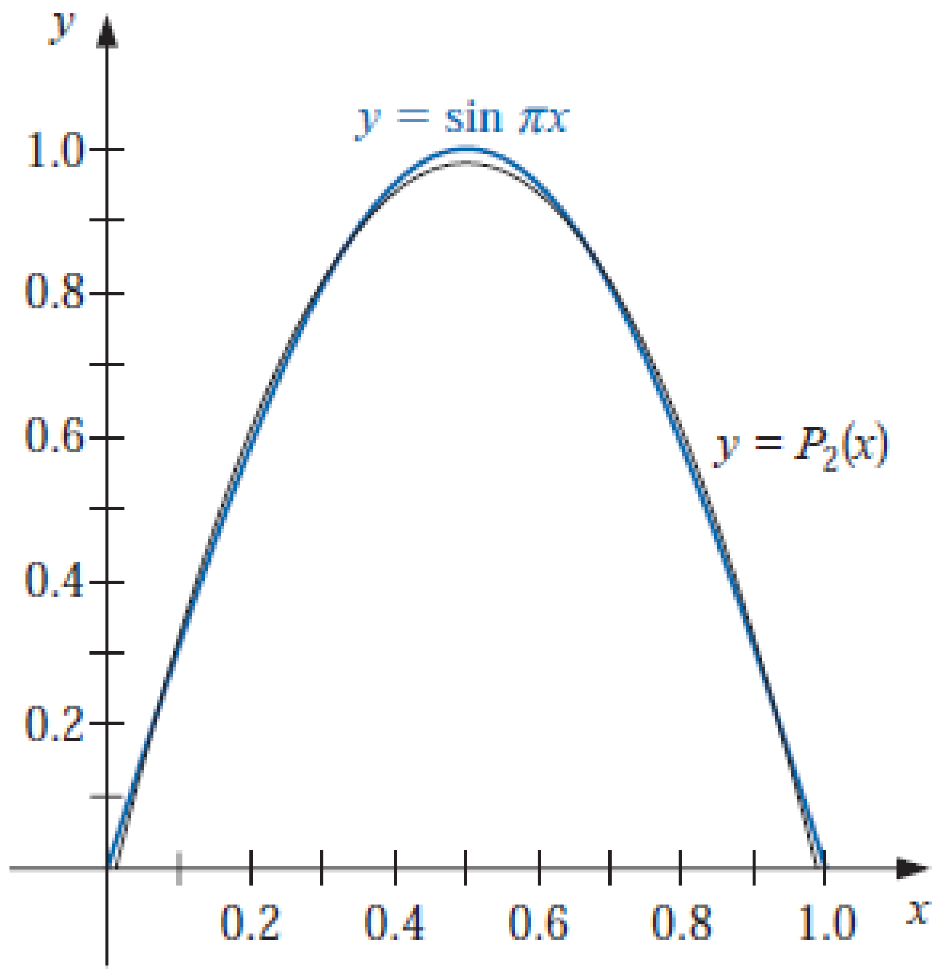
$$a_0 + \frac{1}{2}a_1 + \frac{1}{3}a_2 = \frac{2}{\pi}, \quad \frac{1}{2}a_0 + \frac{1}{3}a_1 + \frac{1}{4}a_2 = \frac{1}{\pi}, \quad \frac{1}{3}a_0 + \frac{1}{4}a_1 + \frac{1}{5}a_2 = \frac{\pi^2 - 4}{\pi^3}.$$

These three equations in three unknowns can be solved to obtain

$$a_0 = \frac{12\pi^2 - 120}{\pi^3} \approx -0.050465 \quad \text{and} \quad a_1 = -a_2 = \frac{720 - 60\pi^2}{\pi^3} \approx 4.12251.$$

Consequently, the least squares polynomial approximation of degree 2 for $f(x) = \sin \pi x$ on $[0, 1]$ is $P_2(x) = -4.12251x^2 + 4.12251x - 0.050465$. (See Figure 8.7.) ■

Figure 8.7



Example 1 illustrates a difficulty in obtaining a least squares polynomial approximation. An $(n + 1) \times (n + 1)$ linear system for the unknowns $a_0, \ldots, a_n$ must be solved, and the coefficients in the linear system are of the form

$$\int_a^b x^{j+k}\, dx = \frac{b^{j+k+1} - a^{j+k+1}}{j + k + 1},$$

a linear system that does not have an easily computed numerical solution. The matrix in the linear system is known as a **Hilbert matrix**, which is a classic example for demonstrating round-off error difficulties. (See Exercise 11 of Section 7.5.)

Another disadvantage is similar to the situation that occurred when the Lagrange polynomials were first introduced in Section 3.1. The calculations that were performed in obtaining the best nth-degree polynomial, $P_n(x)$, do not lessen the amount of work required to obtain $P_{n+1}(x)$, the polynomial of next higher degree.

Linearly Independent Functions

A different technique to obtain least squares approximations will now be considered. This turns out to be computationally efficient, and once $P_n(x)$ is known, it is easy to determine $P_{n+1}(x)$. To facilitate the discussion, we need some new concepts.

Definition 8.1

The set of functions $\{\phi_0, \ldots, \phi_n\}$ is said to be **linearly independent** on $[a, b]$ if, whenever

$$c_0\phi_0(x) + c_1\phi_1(x) + \cdots + c_n\phi_n(x) = 0, \quad \text{for all } x \in [a, b],$$

we have $c_0 = c_1 = \cdots = c_n = 0$. Otherwise the set of functions is said to be **linearly dependent**. ■

Theorem 8.2

Suppose that, for each $j = 0, 1, \ldots, n$, $\phi_j(x)$ is a polynomial of degree j. Then $\{\phi_0, \ldots, \phi_n\}$ is linearly independent on any interval $[a, b]$. ■

Example 2

Let $\phi_0(x) = 2, \phi_1(x) = x - 3$, and $\phi_2(x) = x^2 + 2x + 7$, and $Q(x) = a_0 + a_1 x + a_2 x^2$. Show that there exist constants c_0, c_1, and c_2 such that $Q(x) = c_0\phi_0(x) + c_1\phi_1(x) + c_2\phi_2(x)$.

Solution By Theorem 8.2, $\{\phi_0, \phi_1, \phi_2\}$ is linearly independent on any interval $[a, b]$. First note that

$$1 = \frac{1}{2}\phi_0(x), \quad x = \phi_1(x) + 3 = \phi_1(x) + \frac{3}{2}\phi_0(x),$$

and

$$x^2 = \phi_2(x) - 2x - 7 = \phi_2(x) - 2\left[\phi_1(x) + \frac{3}{2}\phi_0(x)\right] - 7\left[\frac{1}{2}\phi_0(x)\right]$$
$$= \phi_2(x) - 2\phi_1(x) - \frac{13}{2}\phi_0(x).$$

Hence

$$Q(x) = a_0\left[\frac{1}{2}\phi_0(x)\right] + a_1\left[\phi_1(x) + \frac{3}{2}\phi_0(x)\right] + a_2\left[\phi_2(x) - 2\phi_1(x) - \frac{13}{2}\phi_0(x)\right]$$
$$= \left(\frac{1}{2}a_0 + \frac{3}{2}a_1 - \frac{13}{2}a_2\right)\phi_0(x) + [a_1 - 2a_2]\phi_1(x) + a_2\phi_2(x).$$

■

The situation illustrated in Example 2 holds in a much more general setting. Let $\prod_n$ denote the **set of all polynomials of degree at most n**. The following result is used extensively in many applications of linear algebra. Its proof is considered in Exercise 13.

Theorem 8.3

Suppose that $\{\phi_0(x), \phi_1(x), \ldots, \phi_n(x)\}$ is a collection of linearly independent polynomials in $\prod_n$. Then any polynomial in $\prod_n$ can be written uniquely as a linear combination of $\phi_0(x)$, $\phi_1(x), \ldots, \phi_n(x)$. ■

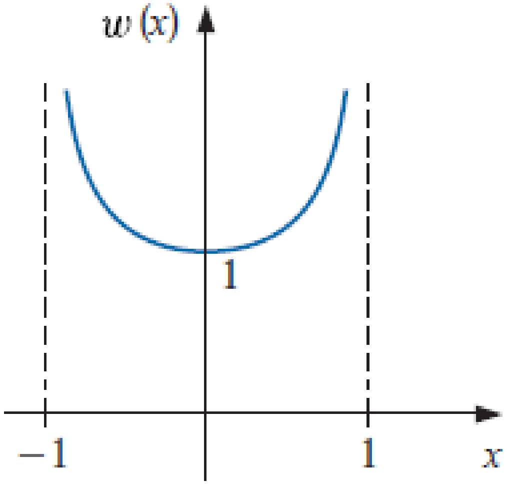
Orthogonal Functions

To discuss general function approximation requires the introduction of the notions of weight functions and orthogonality.

Definition 8.4

An integrable function w is called a **weight function** on the interval I if $w(x) \geq 0$, for all x in I, but $w(x) \not\equiv 0$ on any subinterval of I. ■

Figure 8.8



The purpose of a weight function is to assign varying degrees of importance to approximations on certain portions of the interval. For example, the weight function

$$w(x) = \frac{1}{\sqrt{1-x^2}}$$

places less emphasis near the center of the interval $(-1, 1)$ and more emphasis when $|x|$ is near 1 (see Figure 8.8). This weight function is used in the next section.

Suppose $\{\phi_0, \phi_1, \ldots, \phi_n\}$ is a set of linearly independent functions on $[a, b]$ and w is a weight function for $[a, b]$. Given $f \in C[a, b]$, we seek a linear combination

$$P(x) = \sum_{k=0}^{n} a_k \phi_k(x)$$

to minimize the error

$$E = E(a_0, \ldots, a_n) = \int_a^b w(x)\left[f(x) - \sum_{k=0}^{n} a_k \phi_k(x)\right]^2 dx.$$

This problem reduces to the situation considered at the beginning of this section in the special case when $w(x) \equiv 1$ and $\phi_k(x) = x^k$, for each $k = 0, 1, \ldots, n$.

The normal equations associated with this problem are derived from the fact that for each $j = 0, 1, \ldots, n$,

$$0 = \frac{\partial E}{\partial a_j} = 2\int_a^b w(x)\left[f(x) - \sum_{k=0}^{n} a_k \phi_k(x)\right]\phi_j(x)\, dx.$$

The system of normal equations can be written

$$\int_a^b w(x) f(x) \phi_j(x)\, dx = \sum_{k=0}^{n} a_k \int_a^b w(x)\phi_k(x)\phi_j(x)\, dx, \quad \text{for } j = 0, 1, \ldots, n.$$

If the functions $\phi_0, \phi_1, \ldots, \phi_n$ can be chosen so that

$$\int_a^b w(x)\phi_k(x)\phi_j(x)\,dx = \begin{cases} 0, & \text{when } j \neq k, \\ \alpha_j > 0, & \text{when } j = k, \end{cases} \tag{8.7}$$

then the normal equations will reduce to

$$\int_a^b w(x)f(x)\phi_j(x)\,dx = a_j \int_a^b w(x)[\phi_j(x)]^2\,dx = a_j\alpha_j,$$

for each $j = 0, 1, \ldots, n$. These are easily solved to give

$$a_j = \frac{1}{\alpha_j}\int_a^b w(x)f(x)\phi_j(x)\,dx.$$

Hence the least squares approximation problem is greatly simplified when the functions $\phi_0, \phi_1, \ldots, \phi_n$ are chosen to satisfy the *orthogonality* condition in Eq. (8.7). The remainder of this section is devoted to studying collections of this type.

Definition 8.5

$\{\phi_0, \phi_1, \ldots, \phi_n\}$ is said to be an **orthogonal set of functions** for the interval $[a, b]$ with respect to the weight function w if

$$\int_a^b w(x)\phi_k(x)\phi_j(x)\,dx = \begin{cases} 0, & \text{when } j \neq k, \\ \alpha_j > 0, & \text{when } j = k. \end{cases}$$

If, in addition, $\alpha_j = 1$ for each $j = 0, 1, \ldots, n$, the set is said to be **orthonormal**. ■

This definition, together with the remarks preceding it, produces the following theorem.

Theorem 8.6

If $\{\phi_0, \ldots, \phi_n\}$ is an orthogonal set of functions on an interval $[a, b]$ with respect to the weight function w, then the least squares approximation to f on $[a, b]$ with respect to w is

$$P(x) = \sum_{j=0}^{n} a_j \phi_j(x),$$

where, for each $j = 0, 1, \ldots, n$,

$$a_j = \frac{\int_a^b w(x)\phi_j(x) f(x)\,dx}{\int_a^b w(x)[\phi_j(x)]^2\,dx} = \frac{1}{\alpha_j}\int_a^b w(x)\phi_j(x) f(x)\,dx.$$ ■

Gram-Schmidt process

Theorem 8.7

The set of polynomial functions $\{\phi_0, \phi_1, \ldots, \phi_n\}$ defined in the following way is orthogonal on $[a, b]$ with respect to the weight function w.

$$\phi_0(x) \equiv 1, \quad \phi_1(x) = x - B_1, \quad \text{for each } x \text{ in } [a, b],$$

where

$$B_1 = \frac{\int_a^b x w(x)[\phi_0(x)]^2\, dx}{\int_a^b w(x)[\phi_0(x)]^2\, dx},$$

and when $k \geq 2$,

$$\phi_k(x) = (x - B_k)\phi_{k-1}(x) - C_k\phi_{k-2}(x), \quad \text{for each } x \text{ in } [a, b],$$

where

$$B_k = \frac{\int_a^b x w(x)[\phi_{k-1}(x)]^2\, dx}{\int_a^b w(x)[\phi_{k-1}(x)]^2\, dx}$$

and

$$C_k = \frac{\int_a^b x w(x)\phi_{k-1}(x)\phi_{k-2}(x)\, dx}{\int_a^b w(x)[\phi_{k-2}(x)]^2\, dx}.$$

■

Theorem 8.7 provides a recursive procedure for constructing a set of orthogonal polynomials. The proof of this theorem follows by applying mathematical induction to the degree of the polynomial $\phi_n(x)$.

Corollary 8.8

For any $n > 0$, the set of polynomial functions $\{\phi_0, \ldots, \phi_n\}$ given in Theorem 8.7 is linearly independent on $[a, b]$ and

$$\int_a^b w(x)\phi_n(x)Q_k(x)\,dx = 0,$$

for any polynomial $Q_k(x)$ of degree $k < n$. ■

Illustration

The set of **Legendre polynomials**, $\{P_n(x)\}$, is orthogonal on $[-1, 1]$ with respect to the weight function $w(x) \equiv 1$. The classical definition of the Legendre polynomials requires that $P_n(1) = 1$ for each n, and a recursive relation is used to generate the polynomials when $n \geq 2$. This normalization will not be needed in our discussion, and the least squares approximating polynomials generated in either case are essentially the same.

Using the Gram-Schmidt process with $P_0(x) \equiv 1$ gives

$$B_1 = \frac{\int_{-1}^{1} x\,dx}{\int_{-1}^{1} dx} = 0 \quad \text{and} \quad P_1(x) = (x - B_1)P_0(x) = x.$$

Also,

$$B_2 = \frac{\int_{-1}^{1} x^3\,dx}{\int_{-1}^{1} x^2\,dx} = 0 \quad \text{and} \quad C_2 = \frac{\int_{-1}^{1} x^2\,dx}{\int_{-1}^{1} 1\,dx} = \frac{1}{3},$$

so

$$P_2(x) = (x - B_2)P_1(x) - C_2P_0(x) = (x - 0)x - \frac{1}{3} \cdot 1 = x^2 - \frac{1}{3}.$$

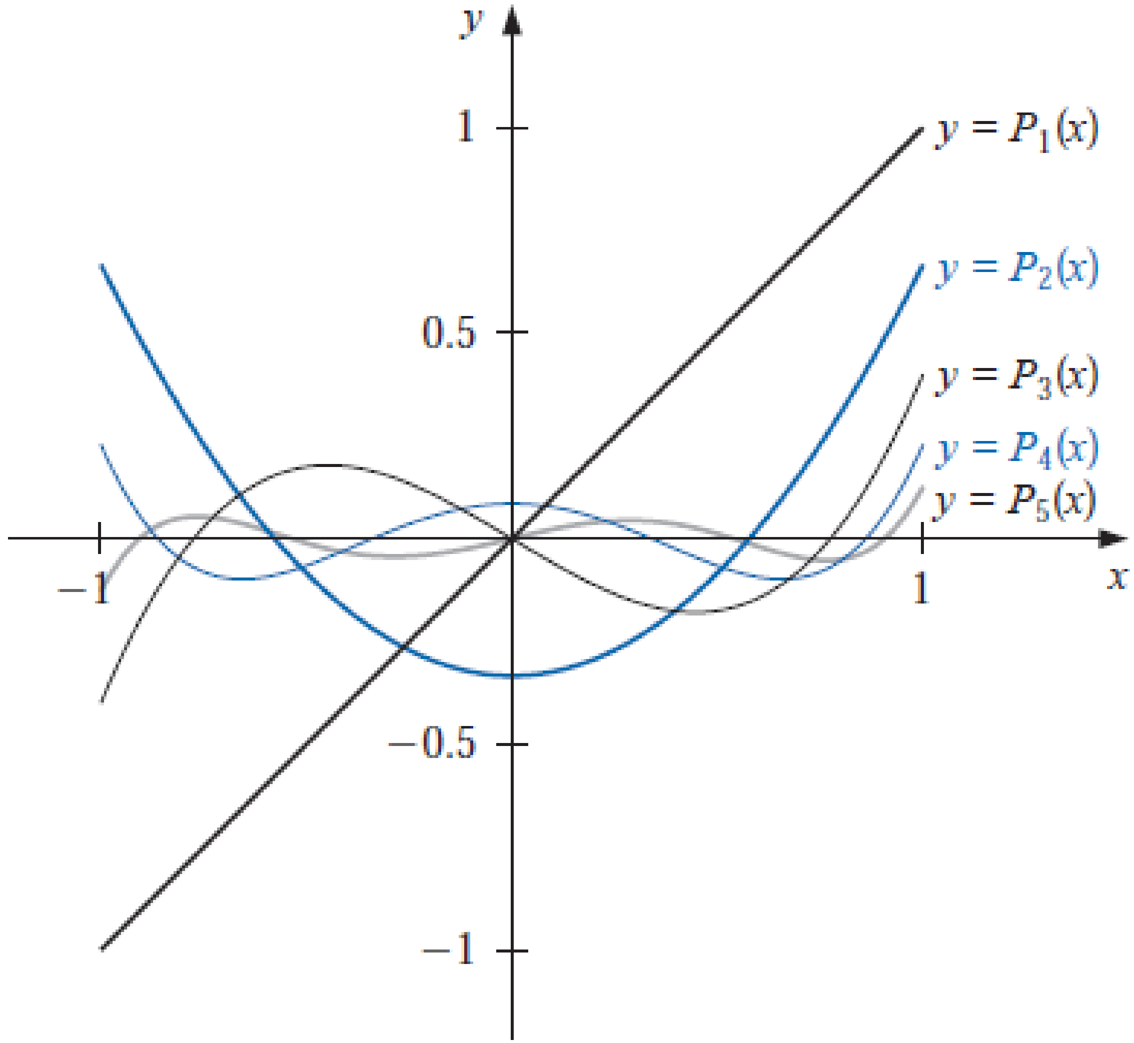
Thus

$$P_3(x) = xP_2(x) - \frac{4}{15}P_1(x) = x^3 - \frac{1}{3}x - \frac{4}{15}x = x^3 - \frac{3}{5}x.$$

The next two Legendre polynomials are

$$P_4(x) = x^4 - \frac{6}{7}x^2 + \frac{3}{35} \quad \text{and} \quad P_5(x) = x^5 - \frac{10}{9}x^3 + \frac{5}{21}x. \qquad \square$$

Figure 8.9



EXERCISE SET 8.2

2. Find the linear least squares polynomial approximation on the interval $[-1, 1]$ for the following functions.
   d. $f(x) = e^x$

4. Find the least squares polynomial approximation of degree 2 on the interval $[-1, 1]$ for the functions in Exercise 2

6. Compute the error E for the approximations in Exercise 4.

7. Use the Gram-Schmidt process to construct $\phi_0(x)$, $\phi_1(x)$, $\phi_2(x)$, and $\phi_3(x)$ for the following intervals.

   Legendre polynomials, $\{P_n(x)\}$, is orthogonal on $[-1, 1]$

   $P_0(x) \equiv 1 \quad P_1(x) = x. \quad P_2(x) = x^2 - \frac{1}{3}. \quad P_3(x) = x^3 - \frac{3}{5}x.$

8. Repeat Exercise 2 using the results of Exercise 7.

9. Obtain the least squares approximation polynomial of degree 3 for the functions in Exercise 2 using the results of Exercise 7.

$$\sum_{k=0}^{n} a_k \int_a^b x^{j+k}\,dx = \int_a^b x^j f(x)\,dx,$$

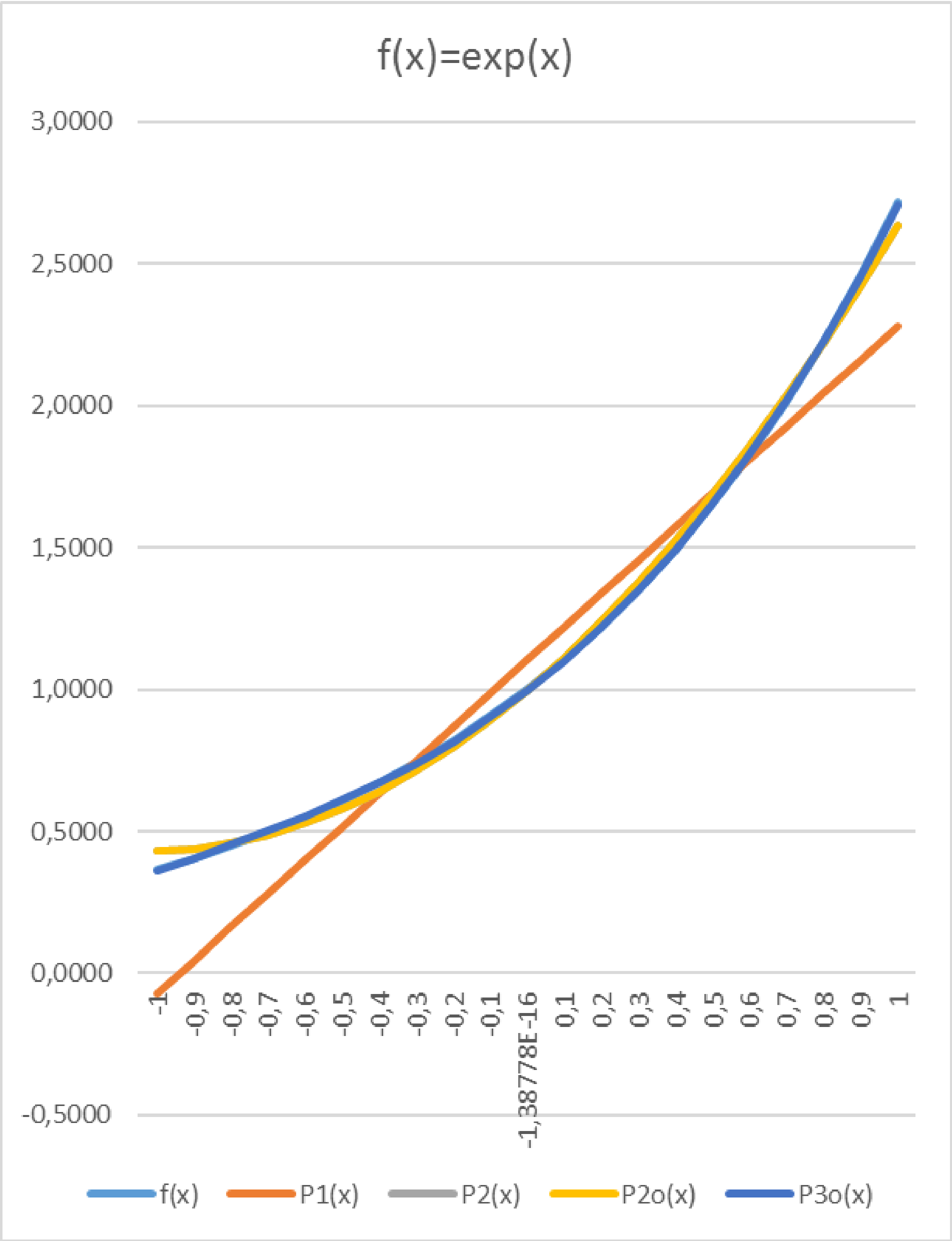
for each $j = 0, 1, \ldots, n$,

$$P(x) = \sum_{j=0}^{n} a_j \phi_j(x),$$

$$a_j = \frac{\int_a^b w(x)\phi_j(x) f(x)\,dx}{\int_a^b w(x)[\phi_j(x)]^2\,dx} = \frac{1}{\alpha_j}\int_a^b w(x)\phi_j(x) f(x)\,dx.$$

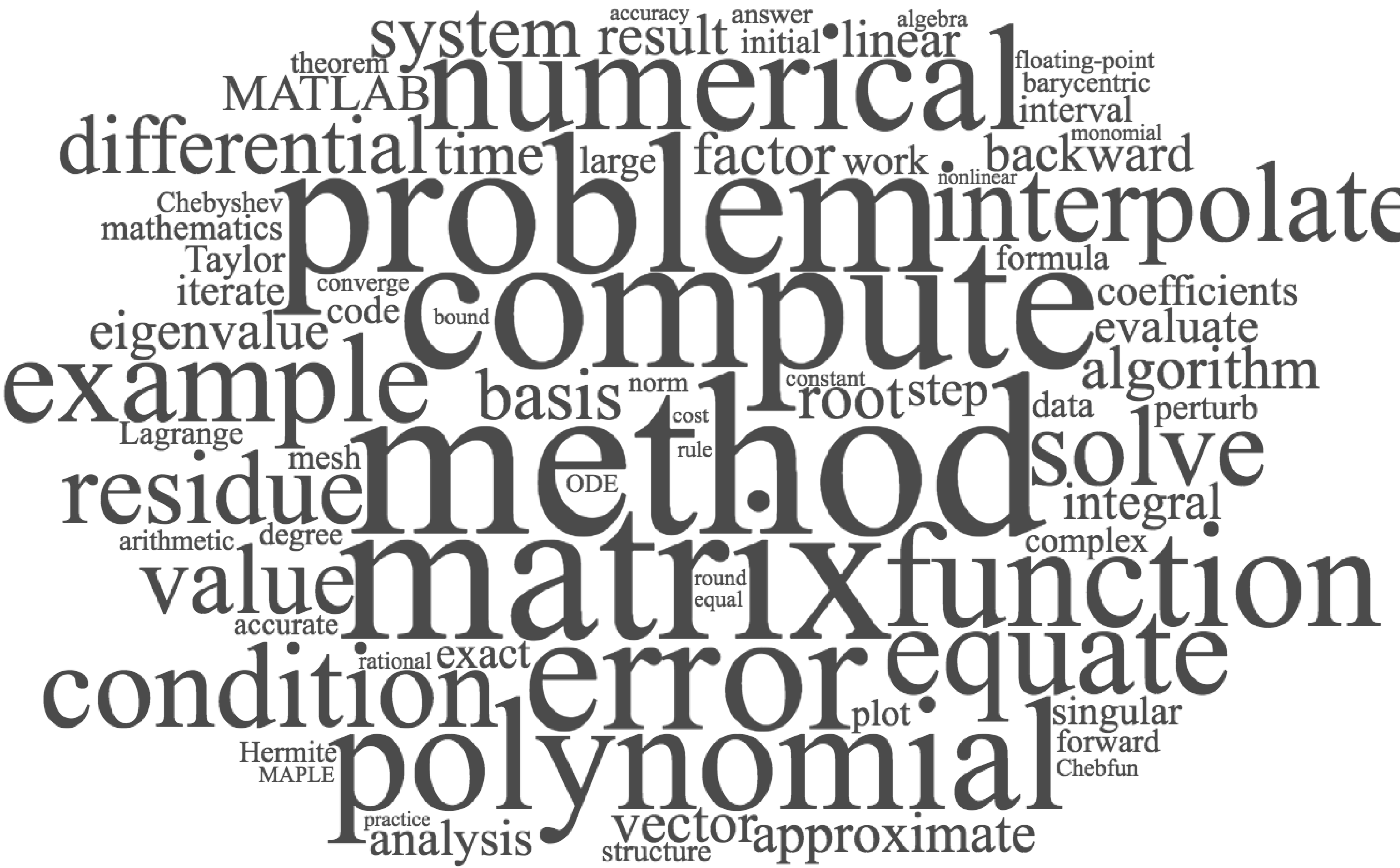
MAP2220

				orthogonal				
x	f(x)	P1(x)	P2(x)	P2o(x)	P3o(x)	df1	df2	df3
-1	0,3679	-0,0717	0,4296	0,4295	0,3586	0,1932	0,0038	0,0001
-0,9	0,4066	0,0458	0,4380	0,4379	0,4044	0,1301	0,0010	0,0000
-0,8	0,4493	0,1633	0,4571	0,4570	0,4513	0,0818	0,0001	0,0000
-0,7	0,4966	0,2809	0,4869	0,4868	0,5005	0,0465	0,0001	0,0000
-0,6	0,5488	0,3984	0,5274	0,5274	0,5529	0,0226	0,0005	0,0000
-0,5	0,6065	0,5159	0,5788	0,5787	0,6097	0,0082	0,0008	0,0000
-0,4	0,6703	0,6334	0,6408	0,6408	0,6720	0,0014	0,0009	0,0000
-0,3	0,7408	0,7509	0,7136	0,7136	0,7407	0,0001	0,0007	0,0000
-0,2	0,8187	0,8685	0,7971	0,7971	0,8169	0,0025	0,0005	0,0000
-0,1	0,9048	0,9860	0,8913	0,8913	0,9018	0,0066	0,0002	0,0000
-1,4E-16	1,0000	1,1035	0,9963	0,9963	0,9963	0,0107	0,0000	0,0000
0,1	1,1052	1,2210	1,1120	1,1120	1,1016	0,0134	0,0000	0,0000
0,2	1,2214	1,3385	1,2385	1,2385	1,2186	0,0137	0,0003	0,0000
0,3	1,3499	1,4561	1,3757	1,3757	1,3485	0,0113	0,0007	0,0000
0,4	1,4918	1,5736	1,5236	1,5236	1,4924	0,0067	0,0010	0,0000
0,5	1,6487	1,6911	1,6823	1,6822	1,6512	0,0018	0,0011	0,0000
0,6	1,8221	1,8086	1,8516	1,8516	1,8261	0,0002	0,0009	0,0000
0,7	2,0138	1,9261	2,0318	2,0317	2,0181	0,0077	0,0003	0,0000
0,8	2,2255	2,0437	2,2227	2,2226	2,2283	0,0331	0,0000	0,0000
0,9	2,4596	2,1612	2,4243	2,4242	2,4577	0,0891	0,0012	0,0000
1	2,7183	2,2787	2,6366	2,6365	2,7074	0,1932	0,0067	0,0001
						0,1361	0,0031	0,0001


f(x)=exp(x)
3,0000
2,5000
2,0000
1,5000
1,0000
0,5000
0,0000
-0,5000
-1
-0,9
-0,8
-0,7
-0,6
-0,5
-0,4
-0,3
-0,2
-0,1
-1,38778E-16
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
f(x)
P1(x)
P2(x)
P2o(x)
P3o(x)

8 Approximation Theory 497

MAP2220


accuracy
answer
algebra
system
result
initial
linear
theorem
floating-point
numerical
MATLAB
barycentric
interval
differential
monomial
time
large
factor
work
backward
nonlinear
problem
interpolate
Chebyshev
mathematics
Taylor
formula
compute
converge
iterate
coefficients
code
bound
eigenvalue
evaluate
algorithm
example
basis
norm
constant
root
step
data
perturb
cost
Lagrange
method
solve
rule
mesh
residue
ODE
integral
arithmetic
degree
complex
matrix
function
value
round
equal
accurate
equate
exact
rational
error
condition
plot
singular
Hermite
polynomial
forward
MAPLE
Chebfun
practice
vector
approximate
analysis
structure