Chapter 5. Best Approximation in $C[a, b]$
Least square approximation

Numerical Analysis 1. Winter Semester 2018-19

1 Best Approximation in $C[a, b]$

Let $f(x)$ be a given function that is continuous on some interval $[a, b]$. If $p(x)$ is an interpolation polynomial, then we are interested in measuring

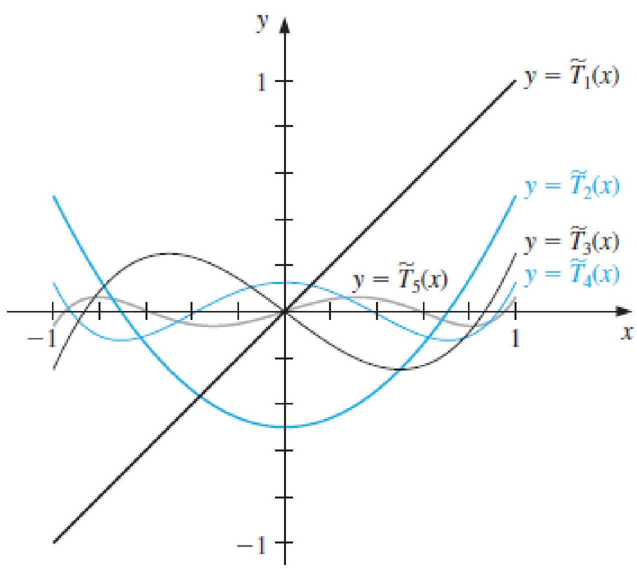
$$\min_{\deg p \leqslant n} E(p) := \min_{\deg p \leqslant n} \max_{a \leqslant x \leqslant b} |f(x) - p(x)| \, .$$

Chebyshev polynomials For $x \in [-1, 1]$, define $T_n(x) = \cos[n \arccos x]$, for each $n \geq 0$.
Three term recurrence

$$T_{n+1}(x) = 2xT_n - T_{n-1} \, .$$

Monic Chebyshev polynomials

$$\tilde{T}_n(x) = \frac{T_n(x)}{2^n} \, .$$



Theorem 1. *The Chebyshev polynomial $T_n(x)$ of degree $n \geq 1$ has n simple zeros in $[-1, 1]$ at*

$$x_k = \cos\Big(\frac{2k-1}{2n}\pi\Big), \text{ for each } k = 1, 2, \ldots, n.$$

Moreover, $T_n(x)$ assumes its absolute extrema at $x'_k = \cos\big(\frac{k\pi}{n}\big)$ with $T_n(x'_k) = (-1)^k$, for each $k = 0, 1, \ldots, n$.

Let $\widetilde{\prod}_n$ denote the set of all monic polynomials of degree n.

Theorem 2. *The polynomials of the form $\tilde{T}_n(x)$, when $n \geq 1$, have the property that*

$$\frac{1}{2n-1} = \max_{x\in[-1,1]} |\tilde{T}_n(x)| \leqslant \max_{x\in[-1,1]} |P_n(x)|, \textit{ for all } P_n(x) \in \widetilde{\prod}_n .$$

Moreover, equality can occur only if $P_n = \tilde{T}_n$.

Why are the Chebyshev nodes generally better than equally spaced nodes in polynomial interpolation? The answer lies in the term $\prod_{i=0}^{n}(x - x_i)$ that occurs in the error formula. If $x_i = \cos[(2i+1)\pi/(2n+2)]$, then

$$\prod_{i=0}^{n}(x - x_i) \leqslant \frac{1}{2^n} \text{ for all } x \in [-1, 1].$$

1.1 Error estimation

Corollary 1. *If $P(x)$ is the interpolating polynomial of degree at most n with nodes at the roots of $T_{n+1}(x)$, then*

$$\max_{x\in[-1,1]} |f(x) - P(x)| \leqslant \frac{1}{2n(n+1)!} \max_{x\in[-1,1]} |f^{(n+1)}(x)|, \textit{ for each } f \in C^{n+1}[-1, 1] .$$

2 Approximation in the least square sense

2.1 Discrete time case: Least Square Approximation Using Data Table

The wonderfully written motivation for least square approximation can be found from page 482-486, [4].

X	1	2	3	4	5	6	7	8	9	10
Y	1.3	3.5	4.2	5.0	7.0	8.8	10.1	12.5	13.0	15.6

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 in a similar manner. We choose the constants $a_0, a_1, \ldots, a_n$ to minimize the least squares error

$$E_2 = \sum_{i=1}^{m} (y_i - P_n(x_i))^2 = \sum_{i=1}^{m} y_i^2 - 2\sum_{i=1}^{m} P_n(x_i) y_i + \sum_{i=1}^{m} (P_n(x_i))^2$$

This gives $n + 1$ **normal equations** in the $n + 1$ unknowns a_j,

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

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

1. The *least squares approximation (discrete time)* is the solution $\vec{a}$ to the **normal system** $A\vec{a} = b$, where

$$A = [a_{ij}], \quad a_{ij} =< P_i(X), P_j(X) >, \quad b_j =< Y, P_j > \ .$$

2. In case of orthorgonal vectors $\{P_i(X),\ i = 0, \ldots, n\}$, the solution is

$$\vec{a}_j = \frac{b_j}{a_{jj}} = \frac{< Y, P_j >}{< P_i(X), P_j(X) >} \, .$$

2.2 Singular Value Decomposition (SVD) and Moore-Penrose Pseudo Inverse

THEOREM 1 **SVD LEAST SQUARES THEOREM**

Let A be an $m \times n$ matrix of rank r. Let the SVD factorization be $A = UDV^T$. The least-squares solution of the system $Ax = b$ is $x_{LS} = \sum_{i=1}^{n} (\sigma_i^{-1} c_i) v_i$, where $c_i = u_i^T b$. If there exist many least-squares solutions to the given system, then the one of least 2-norm is x as described above.

EXAMPLE 1 Find the least-squares solution of this nonsquare system

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

using the singular value decomposition:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}\sqrt{6} & 0 & \frac{1}{3}\sqrt{3} \\ \frac{1}{6}\sqrt{6} & \frac{1}{2}\sqrt{2} & -\frac{1}{3}\sqrt{3} \\ \frac{1}{6}\sqrt{6} & -\frac{1}{2}\sqrt{2} & -\frac{1}{3}\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{bmatrix}$$

Solution We have $r = \text{rank}(A) = 2$ and the singular values $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$. This leads to

$$c_1 = u_1^T b = \begin{bmatrix} \frac{1}{3}\sqrt{6} & \frac{1}{6}\sqrt{6} & \frac{1}{6}\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \frac{1}{3}\sqrt{6}$$

and

$$c_2 = u_2^T b = \begin{bmatrix} 0 & -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \sqrt{2}$$

and

$$x_{LS} = (\sigma_1^{-1} c_1) v_1 + (\sigma_2^{-1} c_2) v_2 = \frac{1}{\sqrt{3}} \left(\frac{1}{3}\sqrt{6} \right) \begin{bmatrix} \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{bmatrix} + \sqrt{2} \begin{bmatrix} \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} \\ -\frac{2}{3} \end{bmatrix}$$

This solution is the same as that from the normal equations.

Its pseudo-inverse D^+ is defined to be of the same form, except that it is to be $n \times m$ and it has $1/\sigma_j$ on its diagonal. For example,

$$D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \qquad D^+ = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{2} \\ 0 & 0 \end{bmatrix}$$

If A is any $m \times n$ matrix and if UDV^T is one of its **singular value decompositions**, we define the **pseudo-inverse** of A to be

$$A^+ = VD^+U^T$$

We do not stop to prove that the pseudo-inverse of A is unique if we impose the order $\sigma_1 \geqq \sigma_2 \geqq \cdots$.

■ **THEOREM 2** **MINIMAL SOLUTION THEOREM**

Consider a system of linear equations $Ax = b$, in which A is an $m \times n$ matrix. The minimal solution of the system is A^+b.

2.3 Continuous time case: Least Square Approximation $L_2[a, b]$

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,

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

Example 1 illustrates the 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 7.4.)

Definition 1. *An integrable function w is called a* ***weight function*** *on the interval $\mathbb{I}$ if $w(x) \geq 0$, for all x in $\mathbb{I}$, but $w(x) \neq 0$ on any subinterval of $\mathbb{I}$.*

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 places less emphasis near the center of the interval $(-1, 1)$ and more emphasis when $|x|$ is near 1.

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

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

then the normal equations 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 $j = 0, 1, \ldots, n$, and are easily solved to give

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

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

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

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

Figure 1: [4], p. 499

2.3.1 The Legendre polynomials

They have many special properties, and they are widely used in numerical analysis and applied mathematics.

$$P_0(x) = 1$$

$$P_n(x) = \frac{1}{n!2^n} \cdot \frac{d^n}{dx^n}\left[\left(x^2 - 1\right)^n\right], \qquad n = 1, 2, \ldots$$

For example,

$$\begin{aligned} P_1(x) &= x \\ P_2(x) &= \tfrac{1}{2}\left(3x^2 - 1\right) \\ P_3(x) &= \tfrac{1}{2}\left(5x^3 - 3x\right) \\ P_4(x) &= \tfrac{1}{8}\left(35x^4 - 30x^2 + 3\right) \end{aligned}$$

Figure 2: [2], p. 181

Properties:

- Degree and normalization:

$$\deg P_n = n, \qquad P_n(1) = 1, \qquad n \geq 0$$

- Triple recursion relation:

$$P_{n+1}(x) = \frac{2n+1}{n+1} x P_n(x) - \frac{n}{n+1} P_{n-1}(x), \qquad n \geq 1$$

- Orthogonality and size:

$$(P_i, P_j) = \begin{cases} 0, & i \neq j \\ \dfrac{2}{2j+1}, & i = j \end{cases}$$

- Zeros:

All zeros of $P_n(x)$ are located in the interval $-1 < x < 1$, and moreover, all zeros are simple roots of $P_n(x)$.

- Basis: Every polynomial $p(x)$ of degree $\leq n$ can be written in the form

$$p(x) = \sum_{j=0}^{n} \beta_j P_j(x)$$

with the choice of $\beta_0, \beta_1, \ldots, \beta_n$ uniquely determined from $p(x)$.

The *least squares approximation (continuous time)* of degree n to $f(x)$ on $[-1, 1]$ is

$$\ell_n(x) = \sum_{j=0}^{n} \frac{< f, P_j >}{< P_j, P_j >} P_j(x) \; . \tag{1}$$

2.3.2 Least Squares Approximation with Weight

For various reasons, we generalize the concept of "average error" by considering the so-called *weighted least squares approximation*.

$$\min \frac{1}{c} \int_a^b w(x)[f(x) - p(x)]^2 dx, \quad c = \int_a^b w(x) dx \; .$$

The function $w(x)$ is assumed to satisfy the following assumptions:

i. $w(x) > 0$ for alll $x \in [a, b]$.

ii. For all integer n, $\int_a^b w(x)|x|^n dx < \infty$.

With these assumptions, formula (1) is still valid, with the new scalar product

$$< f, g >:= \int_a^b w(x) f(x) g(x) dx \; .$$

Example 1. *The Chebyshev polynomial* $\{T_n(x)\}$ *are orthogonal on* $(-1, 1)$ *w.r.t. the weight function* $w(x) = \dfrac{1}{\sqrt{1-x^2}}$.

3 What have not been covered?

In this script we have not considered the following topics.

i) Rational Function Approximation, see [4], Section 8.4.

ii) Trigonometric Polynomial Approximation, see [4], Section 8.5.

4 Survey of Methods and Software

In this chapter we considered approximating data and functions with elementary functions. The elementary functions used were polynomials, rational functions, and trigonometric polynomials. We considered two types of approximations, discrete and continuous. Discrete approximations arise when approximating a finite set of data with an elementary function. Continuous approximations are used when the function to be approximated is known.

Discrete least squares techniques are recommended when the function is specified by giving a set of data that may not exactly represent the function. Least squares fit of data can take the form of a linear or other polynomial approximation or even an exponential form. These approximations are computed by solving sets of normal equations, as given in Section 8.1.

If the data are periodic, a trigonometric least squares fit may be appropriate. Because of the orthonormality of the trigonometric basis functions, the least squares trigonometric approximation does not require the solution of a linear system. For large amounts of periodic data, interpolation by trigonometric polynomials is also recommended. An efficient method of computing the trigonometric interpolating polynomial is given by the fast Fourier transform.

When the function to be approximated can be evaluated at any required argument, the approximations seek to minimize an integral instead of a sum. The continuous least squares polynomial approximations were considered in Section 8.2. Efficient computation of least squares polynomials lead to orthonormal sets of polynomials, such as the Legendre and Chebyshev polynomials. Approximation by rational functions was studied in Section 8.4, where Pad´e approximation as a generalization of the Maclaurin polynomial and its extension to Chebyshev rational approximation were presented. Both methods allow a more uniform method of approximation than polynomials. Continuous least squares approximation by trigonometric functions was discussed in Section 8.5, especially as it relates to Fourier series.

References

[1] P. K. Anh. *Numerical Mathematics*. NXB DHQGHN, 2000.

[2] K. Atkinson and W. Han. *Elementary numerical analysis*. J. Wiley & Sons, 2004.

[3] M. Bollhöfer and V. Mehrmann. *Numerische Mathematik: Eine projektorientierte Einführung für Ingenieure, Mathematiker und Naturwissenschaftler*. vieweg studium; Grundkurs Mathematik. Vieweg+Teubner Verlag, 2013.

[4] R. Burden and J. Faires. *Numerical Analysis*. Cengage Learning, 2010.

[5] E. Cheney and D. Kincaid. *Numerical Mathematics and Computing*. Cengage Learning, 2007.

[6] W. Gautschi. *Numerical Analysis*. SpringerLink : Bücher. Birkhäuser Boston, 2011.

[7] G. Hämmerlin, L. Schumaker, and K. Hoffmann. *Numerical Mathematics*. Undergraduate Texts in Mathematics. Springer New York, 2012.

[8] A. Quarteroni, F. Saleri, and P. Gervasio. *Scientific Computing with MATLAB and Octave*. Texts in Computational Science and Engineering. Springer Berlin Heidelberg, 2014.

[9] J. Stoer, R. Bartels, W. Gautschi, R. Bulirsch, and C. Witzgall. *Introduction to Numerical Analysis*. Texts in Applied Mathematics. Springer New York, 2013.

[10] E. Süli and D. Mayers. *An Introduction to Numerical Analysis*. Cambridge University Press, 2003.