Numerical Analysis

Lecture 09: Approximation Theory

Instructor: Prof. Guanding Yu
Zhejiang University

Approximation Theory

Introduction

Approximation theory involves two general types of problems.

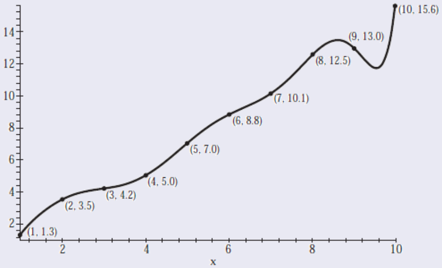
1. Fitting functions to given data and finding the "best" function in a certain class to represent the data.
2. When a function is given explicitly, but we wish to find a "simpler" type of function, such as a polynomial, to approximate values of the given function.

Outline

Discrete Least Squares Approximation

Example: Curve Fitting with Noise

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



Discrete Least Squares Approximation

Example: Curve Fitting with Noise

Assume that the model is $y = a_1 x + a_0$, we need to determine a_0 and a_1 based on the observations.

- Minimax problem

$$E_\infty(a_0, a_1) = \max_{1 \leq i \leq 10} \{|y_i - (a_1 x_i + a_0)|\}$$

- Absolute deviation

$$E_1(a_0, a_1) = \sum_{i=1}^{10} |y_i - (a_1 x_i + a_0)|$$

- Linear Least Squares

$$E_2(a_0, a_1) = \sum_{i=1}^{10} [y_i - (a_1 x_i + a_0)]^2$$

Linear Least Squares

Linear Least Squares Problem

The general problem of fitting the best least squares line to a collection of data $\{(x_i, y_i)\}_{i=1}^m$ involves minimizing the total error

$$E \equiv E_2(a_0, a_1) = \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2$$

with respect to the parameters a_0 and a_1.

Linear Least Squares

Linear Least Squares Problem

$$\{a_0^*, a_1^*\} = \underset{a_0,a_1}{\operatorname{argmin}} \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2$$

Solution: (1/3)

$$\frac{\partial E}{\partial a_0} = 0, \quad \text{and} \quad \frac{\partial E}{\partial a_1} = 0$$

That is

$$0 = \frac{\partial}{\partial a_0} \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2 = 2 \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))(-1)$$

$$0 = \frac{\partial}{\partial a_1} \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2 = 2 \sum_{i=1}^{m} (y_i - (a_1 x_i + a_0))(-x_i)$$

Linear Least Squares

Linear Least Squares Problem

$$\{a_0^*, a_1^*\} = \underset{a_0, a_1}{\operatorname{argmin}} \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2$$

Solution: (2/3)

These equations simplify to the normal equations:

$$a_0 m + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i$$

and

$$a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$$

Linear Least Squares

Solution: (3/3)

The solution to this system of equation is

$$a_0 = \frac{\displaystyle\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} x_i y_i \sum_{i=1}^{m} x_i}{\displaystyle m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}$$

and

$$a_1 = \frac{\displaystyle m\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{\displaystyle m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}$$

Linear Least Squares

Linear Least Squares Problem

$$\{a_0^*, a_1^*\} = \underset{a_0, a_1}{\operatorname{argmin}} \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2$$

$$\mathbf{a}^* = \underset{\mathbf{a}}{\operatorname{argmin}} ||\mathbf{y} - \mathbf{X}\mathbf{a}||_2^2$$

$$0 = -\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{a})$$

Normal equation:

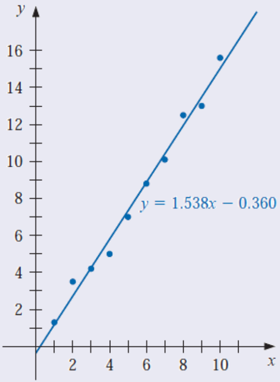
$$\mathbf{X}^T\mathbf{X}\mathbf{a} = \mathbf{X}^T\mathbf{y}$$

$$\Rightarrow \mathbf{a} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

Linear Least Squares

Example: Curve Fitting with Noise

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



Polynomial Least Squares

Polynomial Least Squares Problem

The general problem of approximating a set of data, $\{(x_i, y_i)\}_{i=1}^{m}$, 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 $n < m - 1$. The constants $a_0, a_1, \cdots, a_n$ are determined by minimizing the least squares error

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

Polynomial Least Squares

Polynomial Least Squares Problem

$$\{a_i^*\}_{i=0}^n = \operatorname*{argmin}_{\{a_i\}_{i=0}^n} \sum_{i=1}^{m} (y_i - P_n(x_i))^2$$

Polynomial Least Squares

Polynomial Least Squares Problem

$$\{a_i^*\}_{i=0}^n = \underset{\{a_i\}_{i=0}^n}{\operatorname{argmin}} \sum_{i=1}^m (y_i - P_n(x_i))^2$$

$$\begin{aligned}
E &= \sum_{i=1}^m (y_i - P_n(x_i))^2 \\
&= \sum_{i=1}^m y_i^2 - 2\sum_{i=1}^m P_n(x_i) y_i + \sum_{i=1}^m (P_n(x_i))^2 \\
&= \sum_{i=1}^m y_i^2 - 2\sum_{i=1}^m \left(\sum_{j=0}^n a_j x_i^j\right) y_i + \sum_{i=1}^m \left(\sum_{j=0}^n a_j x_i^j\right)^2 \\
&= \sum_{i=1}^m y_i^2 - 2\sum_{j=0}^n a_j \left(\sum_{i=1}^m y_i x_i^j\right) + \sum_{j=0}^n \sum_{k=0}^n a_j a_k \left(\sum_{i=1}^m x_i^{j+k}\right)
\end{aligned}$$

Polynomial Least Squares

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

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

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

It is helpful to write the equations as follows:

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

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

$$\cdots$$

Polynomial Least Squares

Example: Nonlinear Curve Fitting with Noise

Fit the data in the following table with the discrete least squares polynomial of degree at most 2.

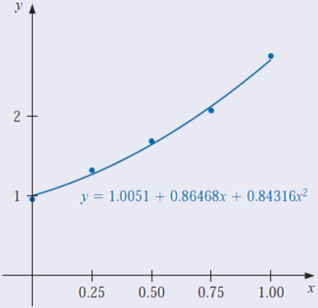
i	x_i	y_i
1	0	1.0000
2	0.25	1.2840
3	0.50	1.6487
4	0.75	2.1170
5	1.00	2.7183

Polynomial Least Squares

Example2: Nonlinear Curve Fitting with Noise

For this problem, $n = 2$, $m = 5$, and the three normal equations are:

$$\begin{aligned} 5a_0 + 2.5a_1 + 1.875a_2 &= 8.7680, \\ 2.5a_0 + 1.875a_1 + 1.5625a_2 &= 5.4514, \\ 1.875a_0 + 1.5625a_1 + 1.3828a_2 &= 4.4015. \end{aligned}$$



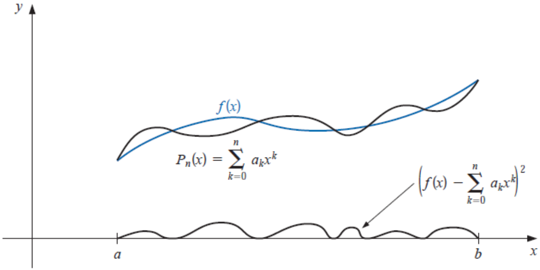
Outline

Approximation of Functions

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

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



Approximation of Functions

Approximation of Functions

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

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

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

Approximation of Functions

Example: Approximation of Functions

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

Approximation of Functions

Example: Approximation of Functions

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

Solution:(1/2)

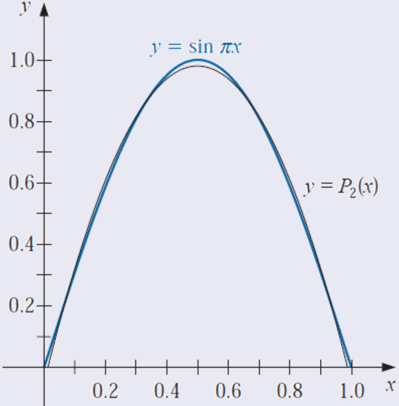
The normal equations for $P_2(x) = a_2x^2 + a_1x + a_0$ are

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

Solution:(2/2)

Consequently, the least squares polynomial approximation of degree 2 for $f(x) = sin\pi x$ on $[0, 1]$ is

$$P_2(x) = -4.12251x^2 + 4.12251x - 0.050465.$$



Approximation of Functions

Disadvantages

1. The coefficients $a_0, a_1, \cdots, a_n$ 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.
2. 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 $P_{n+1}(x)$.

Linearly Independent Functions

Motivation

Computationally efficient, and once $P_n(x)$ is known, it is easy to determine $P_{n+1}(x)$.

Linearly Independent Functions

Definition: Linearly Independent

The set of functions $\{\phi_0, \cdots \phi_n\}$ 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.

Linearly Independent Functions

Theorem

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

Proof (1/2)

Let $c_0, \cdots, c_n$ be real numbers for which

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

- The polynomial $P(x)$ vanishes on $[a, b]$, so it must be the zero polynomial, and the coefficients of all the powers of x are zero.
- In particular, the coefficient of x_n is zero. But $c_n\phi_n(x)$ is the only term in $P(x)$ that contains x_n, so we must have $c_n = 0$.

Linearly Independent Functions

Proof (2/2)

Hence

$$P(x) = \sum_{j=0}^{n-1} c_j \phi_j(x).$$

In this representation of $P(x)$, the only term that contains a power of x^{n-1} is $c_{n-1}\phi_{n-1}(x)$, so this term must also be zero and

$$P(x) = \sum_{j=0}^{n-2} c_j \phi_j(x).$$

In like manner, the remaining constants $c_{n-2}, c_{n-3}, \cdots, c_1, c_0$ are all zero, which implies that $\{\phi_0, \phi_1, \cdots, \phi_n\}$ is linearly independent on $[a, b]$.

Linearly Independent Functions

Theorem

Suppose that $\{\phi_0(x), \phi_1(x), \cdots, \phi_n(x)\}$ *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* $\phi_0(x), \phi_1(x), \cdots, \phi_n(x)$.

Orthogonal Functions

Definition: Weight function

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

Example

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

Orthogonal Functions

Definition: Orthogonal Set of Functions

$\{\phi_0(x), \phi_1(x), \cdots, \phi_n(x)\}$ 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 $j = 0, 1, \cdots, n$, the set is said to be orthonormal.

Orthogonal Functions

Theorem

If $\{\phi_0(x), \phi_1(x), \cdots, \phi_n(x)\}$ *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 the weight function* w *is*

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

where, for each $j = 0, 1, \cdots, n$*,*

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

Orthogonal Functions

Proof

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

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

$$\int_a^b w(x) f(x) \phi_j(x) dx = \sum_{k=0}^{n} a_k \int_a^b w(x) \phi_k(x) \phi_j(x) dx = a_j \alpha_j.$$

Orthogonal Functions

Theorem: Gram-Schmidt Process

The set of polynomial functions $\{\phi_0(x), \phi_1(x), \cdots, \phi_n(x)\}$ 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

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

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

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

Orthogonal Functions

Corollary

For any $n > 0$, the set of polynomial functions $\{\phi_0, \cdots, \phi_n\}$ constructed in the Gram-Schmidt process 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 $k < n$.

Orthogonal Functions

Proof:

- For each $k = 0, 1, \cdots, n$, $\phi_k(x)$ is a polynomial of degree k, which implies that $\{\phi_0(x), \cdots, \phi_n(x)\}$ is a linearly independent set.
- Let $Q_k(x)$ be a polynomial of degree $k < n$. There exist numbers $c_0, \cdots, c_k$ such that

$$Q_k(x) = \sum_{j=0}^{k} c_j \phi_j(x).$$

Because ϕ_n is orthogonal to ϕ_j for each $j = 0, 1, \cdots, k$, we have

$$\int_a^b w(x) Q_k(x) \phi_n(x) dx = \sum_{j=0}^{k} c_j \int_a^b w(x) \phi_j(x) \phi_n(x) dx = \sum_{j=0}^{k} c_j \cdot 0 = 0.$$

Orthogonal Functions

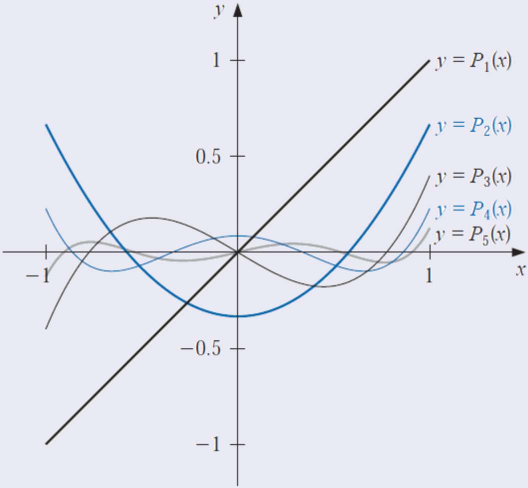
Legendre Polynomials

The set of **Legendre polynomials**, $\{P_n(x)\}$, is orthogonal on $[-1, 1]$ with respect to the weight function $w(x) \equiv 1$.

$$\begin{aligned}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= x^2 - \frac{1}{3} \\
P_3(x) &= x^3 - \frac{3}{5}x \\
P_4(x) &= x^4 - \frac{6}{7}x^2 + \frac{3}{35} \\
P_5(x) &= x^5 - \frac{10}{9}x^3 + \frac{5}{21}x
\end{aligned}$$

Orthogonal Functions

Example: Legendre Polynomials



Remarks

Advantages of Polynomials Approximation

- There are a sufficient number of polynomials to approximate any continuous function on a closed interval to within an arbitrary tolerance;
- Polynomials are easily evaluated at arbitrary values;
- The derivatives and integrals of polynomials exist and are easily determined.

Disadvantages

- Oscillation. This often causes error bounds in polynomial approximation to significantly exceed the average approximation error.

Outline

Rational Function Approximation

Definition

A rational function r of degree N has the form

$$r(x) = \frac{p(x)}{q(x)},$$

where $p(x)$ and $q(x)$ are polynomials whose degrees sum to N.

Every polynomial is a rational function, so approximation by rational functions gives results that are no worse than approximation by polynomials.

Rational Function Approximation

Pade Approximation

Suppose r is a rational function of degree $N = n + m$ of the form

$$r(x) = \frac{p(x)}{q(x)} = \frac{p_0 + p_1 x + \cdots + p_n x^n}{q_0 + q_1 x + \cdots + q_m x^m},$$

that is used to approximate a function f on a closed interval I containing zero.

- The Pade approximation technique is the extension of Taylor polynomial approximation to rational functions.
- It chooses the $N + 1$ parameters $q_1, q_2, \cdots, q_m, p_0, p_1, \cdots, p_n$ so that $f^{(k)}(0) = r^{(k)}(0)$, for each $k = 0, 1, \cdots, N$.
- When $n = N$ and $m = 0$, the Pade approximation is simply the Nth Maclaurin polynomial.

Rational Function Approximation

The Pade Approximation Technique

- Consider the difference

$$f(x)-r(x)=f(x)-\frac{p(x)}{q(x)}=\frac{f(x)q(x)-p(x)}{q(x)}=\frac{f(x)\sum_{i=0}^{m}q_ix^i-\sum_{i=0}^{n}p_ix^i}{q(x)},$$

and suppose f has the Maclaurin series expansion $f(x)=\sum_{i=0}^{\infty}a_ix^i$. Then

$$f(x)-r(x)=\frac{\sum_{i=0}^{\infty}a_ix^i\sum_{i=0}^{m}q_ix^i-\sum_{i=0}^{n}p_ix^i}{q(x)},$$

- The object is to choose the constants $q_1, q_2, \cdots, q_m$ and $p_0, p1, \cdots, p_n$ so that

$$f^{(k)}(0)-r^{(k)}(0)=0, \quad \text{for each } k=0,1,\cdots,N.$$

Rational Function Approximation

The Pade Approximation Technique

We choose $q_1, q_2, \cdots, q_m$ and $p_0, p_1, \cdots, p_n$ so that the numerator

$$(a_0 + a_1 x + \cdots)(1 + q_1 x + \cdots + q_m x^m) - (p_0 + p_1 x + \cdots + p_n x^n),$$

has no terms of degree less than or equal to N.
Then, we can express the coefficient of x^k more compactly as

$$\left(\sum_{i=0}^{k} a_i q_{k-i} \right) - p_k.$$

The rational function for Pade approximation results from the solution of the $N + 1$ linear equations

$$\left(\sum_{i=0}^{k} a_i q_{k-i} \right) = p_k, \quad k = 0, 1, \cdots, N$$

in the $N + 1$ unknowns $q_1, q_2, \cdots, q_m, p_0, p_1, \cdots, p_n$.

Rational Function Approximation

Example:

The Maclaurin series expansion for e^{-x} is

$$\sum_{i=0}^{\infty} \frac{(-1)^i}{i!} x^i.$$

Find the Pade approximation to e^{-x} of degree 5 with $n = 3$ and $m = 2$.

Rational Function Approximation

Solution: (1/2)

To find the Pade approximation we need to choose p_0, p_1, p_2, p_3, q_1, and q_2 so that the coefficients of x^k for $k = 0, 1, \cdots, 5$ are 0 in the expression

$$\left(1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \cdots\right)(1 + q_1 x + q_2 x^2) - (p_0 + p_1 x + p_2 x^2 + p_3 x^3)$$

Expanding and collecting terms produces

$$\begin{aligned} x^5 &: -\frac{1}{120} + \frac{1}{24}q_1 - \frac{1}{6}q_2 = 0; & x^2 &: \frac{1}{2} - q_1 + q_2 = p_2; \\ x^4 &: \frac{1}{24} - \frac{1}{6}q_1 + \frac{1}{2}q_2 = 0; & x^1 &: -1 + q_1 = p_1; \\ x^3 &: -\frac{1}{6} + \frac{1}{2}q_1 - q_2 = p_3; & x^0 &: 1 = p_0. \end{aligned}$$

Rational Function Approximation

Solution: (2/2)

This gives

$$\left\{p_1 = -\frac{3}{5}, p_2 = \frac{3}{20}, p_3 = -\frac{1}{60}, q_1 = \frac{2}{5}, q_2 = \frac{1}{20}\right\}$$

So the Pade approximation is

$$r(x) = \frac{1 - \frac{3}{5}x + \frac{3}{20}x^2 - \frac{1}{60}x^3}{1 + \frac{2}{5}x + \frac{1}{20}x^2}.$$

Assignment

- Reading Assignment: Chap 8.1, 8.2, 8.4