Chapter 3: Approximation Theory

Peter W. White

white@tarleton.edu

Department of Mathematics
Tarleton State University

Summer 2015 / Numerical Analysis

Overview

l_p-norms and Error functions

The l_p norm of an n-vector $\vec{v}$ is

$$||\vec{v}||_p = \left(\sum_{k=1}^{n} |v_k|^p \right)^{1/p}.$$

Let $S = \{(x_k, y_k)\}_{k=1}^{n}$ be a set of discrete data points derived from an unknown function f and let g be a function with parameters $\{a_j\}_{j=1}^{m}$. We say that g approximates f (or the data set) with l_p **error** of

$$E_p(a_1, a_2, ..., a_m) = \sum_{k=1}^{n} |y_k - g(x_k)|^p$$

Examples

Example 1

Suppose the approximating function is $g(x) = a_1 e^{a_2 x} + a_3$. Then the l_1 error would be

$$E_1(a_1, a_2, a_3) = \sum_{k=1}^{n} |y_k - (a_1 e^{a_2 x_k} + a_3)|.$$

Finding the best fit parameters in an absolute sense would require minimizing the l_∞ error:

$$E_\infty(a_1, a_2, a_3) = \max_{1 \le k \le n} \{|y_k - (a_1 e^{a_2 x_k} + a_3)|\}.$$

Both of these error functions lead to difficult minimization problems.

Square Error

The Euclidean or **Square error** function,

$$E \equiv E_2(a_1, ..., a_m) = \sum_{k=1}^{n} (y_k - g(x_k))^2,$$

is the commonly used error function because of it's convenient minimization properties and the following:

Theorem 2

(From Analysis) For all $p, q \in \mathbb{Z}^+ \cup \{\infty\}$ *and all* $n \in \mathbb{Z}^+$*, there exist* $m, M \in \mathbb{R}^+$ *such that for all* $\vec{v} \in \mathbb{R}^n$*,*

$$m||\vec{v}||_p \leq ||\vec{v}||_q \leq M||\vec{v}||_p.$$

Linear Least-Squares Approximation

Suppose that $\{\phi_j\}_{j=1}^{m}$ are a set of “basis” functions and $g(x) = a_1\phi_1(x) + a_2\phi_2(x) + ... + a_m\phi_m(x)$. Then g approximates a data set S with square error given above. To minimize this error we solve the system

$$\frac{\partial E}{\partial a_j} = 0, \ 1 \leq j \leq m$$

of linear equations.

Example 3

Straight-line, linear least-squares approximation uses $g(x) = a_1 x + a_2$ as the approximating function.

Overview

Function Norms

Suppose $f \in C[a, b]$, then the L_p norm of f is given by

$$||f||_p = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p} .$$

Suppose g is an approximating function for f. Then the **L_p error** between f and g is

$$E_p = \int_a^b |f(x) - g(x)|^p \, dx$$

where g (and then E_p) may depend on parameters a_1, a_2, ..., a_m.

Square Error for Functions

As with the discrete case, the common error that is used for functions is the **square error**, when $p = 2$.

Example 4

Suppose g is an m^{th} degree polynomial. Then the square error is

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

To minimize this error over the parameter space we solve the linear system of **normal equations**:

$$\frac{\partial E}{\partial a_k} = 0, \; k = 0, 1, ..., m.$$

Hilbert Matrix

The above system of normal equations leads to:

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{m+1} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{m+2} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdots & \frac{1}{m+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m+1} & \frac{1}{m+2} & \frac{1}{m+3} & \cdots & \frac{1}{2m+1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \int_a^b f(x)\,dx \\ \int_a^b xf(x)\,dx \\ \int_a^b x^2 f(x)\,dx \\ \vdots \\ \int_a^b x^m f(x)\,dx \end{bmatrix}$$

The coefficient matrix is called a **Hilbert matrix**, which is a classic example for demonstrating round-off error difficulties.

Linear Independence of Functions

Definition 5

A set of functions $\{f_1, f_2, ..., f_n\}$ is said to be **linearly independent** on $[a, b]$ if

$$a_1 f_f(x) + a_2 f_2(x) + \cdots + a_n f_n(x) = 0, \ \forall x \in [a, b]$$

$$\Leftrightarrow a_1 = a_2 = \cdots = a_n = 0.$$

Otherwise the set of functions is **linearly dependent**.

Example 6

Is $\{1, x, x^2\}$ linearly independent? How about $\{1, \cos(2x), \cos^2(x)\}$?

Orthogonal Functions and L_2 Inner Product

Definition 7

Let $f, g \in C[a, b]$. The L_2 **inner product** of f and g is given by

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

Note that the L_2 norm of f is then $||f||_2 = \sqrt{\langle f, f \rangle}$.

Definition 8

Two functions f and g, both in $C[a, b]$, are said to be **Orthogonal** if $\langle f, g \rangle = 0$.

Weighted Inner Products

Definition 9

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.

The purpose of a weight function is to assign more importance to approximations on certain portions of the interval.

Definition 10

For f and g in $C[a, b]$ and w a weight function on $[a, b]$,

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

is a weighted inner product.

Weighted Errors

The error function associated with a weighted inner product is

$$E(a_0, \ldots, a_m) = \int_a^b w(x)\,(f(x) - g(x))^2 \; dx$$

where the approximating function g depends on the parameters a_k.

Example 11

Suppose $\{\phi_0, \phi_1, \ldots, \phi_m\}$ 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 want to find the best fit approximation

$$g(x) = \sum_{k=0}^{m} a_k \phi_k(x).$$

Example Continued

That is, we wish to minimize the above error. This leads to a system of normal equations of the form

$$\langle f, \phi_j \rangle_w = \sum_{k=0}^{m} a_k \langle \phi_k, \phi_j \rangle_w .$$

If we can choose the functions in $\{\phi_0, \phi_1, ...\phi_m\}$ to be pairwise orthogonal (with respect to the weight w), then the minimizing parameters would be given by

$$a_k = \frac{\langle f, \phi_k \rangle_w}{\langle \phi_k, \phi_k \rangle_w} .$$

Sine and Cosine

Example 12

Show that $\{1, \sin(kx), \cos(kx)\}_{k=1}^{m}$ form an orthogonal set of functions with respect to $w(x) \equiv 1$ on $[-\pi, \pi]$.

Example 13

Find an orthogonal set of polynomials that span the space of third degree polynomials with respect to $w(x) \equiv 1$ on $[-1, 1]$. This uses a **Gram-Schmidt process**. These polynomials are the first four **Legendre Polynomials**.

Homework

Homework assignment 5, due: TBA

Overview

Why?

Polynomials are good in that:

- Any continuous function can be approximated on a closed interval to within an arbitrary tolerance.
- Polynomials are easy to evaluate at arbitrary values.
- The derivatives and integrals of polynomials exist and are easy to determine.

Polynomials do tend to oscillate dramatically. So in discrete approximations, the approximating polynomial may have small l_2 error even though the L_2 error between the polynomial and the underlying function is large. Polynomials do not do well with discontinuities, especially singularities.

Rational Functions

A **rational function** of degree N has the form

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

where

$$p(x) = p0 + p_1 x + \ldots + p_n x^n$$

and

$$q(x) = q_0 + q_1 x + \ldots + q_m x^m$$

with $n + m = N$.

Rational functions often do a better job of approximating functions (with the same effort) as polynomials, and can include discontinuities. Note that, without loss of generality, we may set $q_0 = 1$.

Padé Approximation

This is an extension of Taylor polynomials to rational functions. It chooses parameters so that $f^{(k)}(0) = r^{(k)}(0)$ for $k = 0, 1, ..., N$. When $n = N$ and $m = 0$, the Padé approximation is simply the N^{th} degree Maclaurin polynomial.

Suppose $f(x)$ has a Maclaurin expansion: $f(x) = \sum a_i x^i$. Then

$$f(x) - r(x) = \frac{f(x)q(x) - p(x)}{q(x)} = \frac{\sum_{i=0}^{\infty} a_i x^i \sum_{i=0}^{m} q_i x^i - \sum_{i=0}^{n} p_i x^i}{q(x)}.$$

Minding the p's and q's

The objective is to find the constants p_i and q_i so that

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

This means $f - r$ has a root of multiplicity $N + 1$ at $x = 0$. That is, the numerator of $f(x) - r(x)$ has no non-zero terms of degree less than $N + 1$. So,

$$\sum_{i=0}^{k} a_i q_{k-i} - p_k = 0$$

where we set $p_i = 0$ for $i = n + 1, n + 2, ..., N$ and $q_i = 0$ for $i = m + 1, m + 2, ...N$.

Example

Example 14

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

Solution: expand the following and collect terms, setting coefficients of x^j to zero for $j = 0, 1, ..., 5$.

$$\left(1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + ...\right)\left(1 + q_1 x + q_2 x^2\right)$$

$$-\left(p_0 + p_1 x + p_2 x^2 + p_3 x^3\right).$$

To get $p_0 = 1$, $p_1 = -\frac{3}{5}$, $p_2 = \frac{3}{20}$, $p_3 = -\frac{1}{60}$, $q_1 = \frac{2}{5}$, and $q_2 = \frac{1}{20}$.

Overview

Euler's Formula

$$
\begin{aligned}
e^{ix} &= 1 + (ix) + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \ldots \\
&= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots\right) \\
&= \cos(x) + i\sin(x),
\end{aligned}
$$

where $i = \sqrt{-1}$. So we can write

$$a_k \cos(kx) + b_k \sin(kx) = c_k e^{kx},$$

where a_k and b_k are real, and c_k is complex.

Trigonometric Polynomials

Given $2m$, evenly spaced, data points $\{x_j, y_j\}$ from a function f, we can transform the data in a linear way so that it is assumed that $x_j = -\pi + (j/m)\pi$ for $j = 0, 1, 2, \ldots, 2m-1$. Then we can find a_k and b_k so that

$$S_m(x) = \frac{a_0 + a_m \cos(mx)}{2} + \sum_{k=1}^{m-1} (a_k \cos(kx) + b_k \sin(kx))$$

interpolates the transformed data. That is, when

$$a_k = \frac{1}{m} \sum_{j=0}^{2m-1} y_j \cos(kx_j) \text{ and } b_k = \frac{1}{m} \sum_{j=0}^{2m-1} y_j \sin(kx),$$

the l_2 error between $S_m(x)$ and the data is zero.

Using Symmetry

Euler's formula gives us $S_m(x) = \frac{1}{m}\sum_{k=0}^{2m-1} c_k e^{ikx}$ where $c_k = \sum_{k=0}^{2m-1} y_j e^{ik\pi j/m}$. From this we have
$a_k + ib_k = \frac{(-1)^k}{m} c_k$.
Suppose $m = 2^p$ for some positive integer p, then for $k = 0, 1, 2, ..., m-1$, we have

$$c_k + c_{m+k} = \sum_{j=0}^{2m-1} y_j e^{ik\pi j/m}(1 + e^{ij\pi}).$$

But $1 + e^{ij\pi} = 2$ if j is even and zero if j is odd, so there are only m nonzero terms in the sum and we can write

$$c_k + c_{m+k} = 2\sum_{l=0}^{m-1} y_{2l} e^{ik\pi(2l)/m} = 2\sum_{l=0}^{m-1} y_{2l} e^{ik\pi l/(m/2)}.$$

Similarly,

$$c_k - c_{m+k} = 2e^{ik\pi/m} \sum_{l=0}^{m-1} y_{2l+1} e^{ik\pi l/(m/2)}.$$

Note that these two relationships allow us to calculate all of the c_k's but the sums now require $2m^2 + m$ complex multiplications instead of $(2m)^2$ multiplications calculating the coefficients directly.

These sums have the same form as the sum for calculating the c_k's directly except we replace m with $m/2$. Thus, we can repeat the process (another $p - 1$ times) to further reduce the number of complex multiplications to $3m + m\log_2(m) = O(m\log_2(m))$. If $m = 1024$, that is about 13,300 complex multiplications instead of about 4,200,000 using the direct method.

Example

Example 15

Play with Matlab to investigate FFT using functions like $\sin(nx)$ and $\cos(nx)$ for various values of $n \in \mathbb{Z}^+$. Then try $f_1(x) = 1 - x^2$ and $f_2(x) = x^3$ on $[-1, 1]$