Chapter 1: Preliminaries and Error Analysis

Peter W. White
white@tarleton.edu

Department of Mathematics
Tarleton State University

Fall 2018 / Numerical Analysis

Overview

We All Remember Calculus

- Derivatives: limit definition, sum and difference rule, product rule, quotient rule, power rule and chain rule.
- Derivatives: instantaneous rates of change, related rates, differentials, slopes of tangents.
- Integrals: Riemann Sums, anti-derivatives, definite integrals, improper integrals.
- Taylor Series: Taylor polynomials, remainder formula, radius of convergence.

Taylor Series

Theorem 1

Let I be an open interval centered at a. If $f \in C^{n+1}(I)$, then

$$f(x) = P_n(x) + R_n(x)$$

where

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

is the n^{th}-degree ***Taylor Polynomial*** *centered at a and*

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

for some ξ between x and a is the ***remainder****.*

Example: Taylor Series

Example 2

Find the Taylor Polynomial centered at $a = 1$ that approximates $f(x) = \ln(x)$ accurate to three decimal digits on the interval $\left(\frac{1}{2}, \frac{3}{2}\right)$.

Example 3

Establish a theoretical error bound for $P_3(x)$ centered at $a = 0$ in approximating $f(x) = \sin(x)$ on the interval $\left(-\frac{\pi}{6}, \frac{\pi}{6}\right)$.

What Is Numerical Analysis?

Definition 4

Numerical Analysis is the branch of mathematics that deals with the development and use of numerical methods for approximating the solutions of problems.

Definition 5

A **numerical method** is a complete and unambiguous set of procedures for the solution of a problem, together with computable error estimates.

Homework

Homework assignment section 1.1, due: TBA

Overview

Review: Prerequisite Mathematics

Round-off Errors and Computer Arithmetic

Algorithms and Convergence

Introduction to Matlab

What is Round-off Error

Definition 6

Round-off Error is the difference between an approximation of a number used in computation and its exact (correct) value.

Example 7

π is irrational. Thus, π cannot be written in decimal form. But we can approximate π out to say six decimal places by 3.141593 with a **round-off error** bounded by 5×10^{-7}.

Absolute and Relative Error

Definition 8

Let p be an exact number and $\hat{p}$ be an approximation to p. Then the **absolute error** in the approximation is given by $|p - \hat{p}|$ and the **relative error** is given by $\left|\frac{p - \hat{p}}{p}\right|$.

Example 9

Find the absolute and relative error when $p = 1,005,862$ and $\hat{p} = 1,000,000$.

Binary Machine Numbers

Definition 10

A binary number system represents a signed floating-point number as $(-1)^s 2^{c-d}(1+f)$ where s is the **sign bit**, c is the n-bit **characteristic**, f is the m-bit **mantissa** and $d = 2^{n-1} - 1$. The binary number with all bits set to 0 is **zero**.

Example 11

In a one byte floating-point number system with a 3 bit characteristic and a 4 bit mantissa, then the binary number 01001011 ($s = 0$, $c = 100$, and $f = 1011$) is 3.375 in bas 10. What are the largest and smallest positive numbers that can be represented in this system?

Underflow and Overflow

- **Underflow** - a number smaller than the smallest representable number is set to zero.
- **Overflow** - a number larger than the largest representable number causes problems.

Repeated subtraction or division could cause underflow.
Repeated addition or multiplication could cause overflow.

Addition (in base 10)

Example 12

Add the following three numbers using a “3-digit” calculater: 1.08, 9.59×10^{-3}, and 9.03×10^{-3}.

Significant Digits

Definition 13

The number $\hat{p}$ is said to approximate p to t **significant digits** if t is the largest non-negative integer for which

$$\left|\frac{p-\hat{p}}{p}\right| < 5 \times 10^{-t}.$$

Example 14

Let $p = \pi$ and $\hat{p} = 3.141$, then

$$\left|\frac{p-\hat{p}}{p}\right| \approx 5.07 \times 10^{-4} < 5 \times 10^{-3}.$$

So $t = 3$.

Significant Digits II

Example 15

$p = 0.d_1 d_2 d_3 ... d_k d_{k+1} d_{k+2} ... \times 10^n$
$\hat{p} = 0.d_1 d_2 d_3 ... d_k \times 10^n$
then $|p - \hat{p}| = 0.d_{k+1} d_{k+2} d_{k+3} \times 10^{n-k}$ and

$$\left| \frac{p - \hat{p}}{p} \right| = \frac{0.d_{k+1} d_{k+2} d_{k+3} \times 10^{n-k}}{0.d_1 d_2 d_3 \times 10^n}$$

Since the minimum value of d_1 is 1 and the maximum value of d_{k+1} is 9, then

$$\left| \frac{p - \hat{p}}{p} \right| \leq \frac{1}{0.1} \times 10^{-k} = 10^{-k+1}.$$

Similarly, k digit rounding produces a relative error bounded by $0.5 \times 10^{-k+1}$. So $k - 1$ significant digits.

Operations and Errors

Do the mathematical operations +, -, $\times$, and $\div$ magnify error?
Suppose x and y are exact values represented in a machine by $x + \Delta x$ and $y + \Delta y$. Let $S = x + y$, $M = x - y$, $T = xy$, and $Q = \dfrac{x}{y}$. In the machine these are represented by $S + \Delta S$, $M + \Delta M$, $T + \Delta T$, and $Q + \Delta Q$ where

$$\begin{aligned}
\Delta S &\approx dS = dx + dy \approx \Delta x + \Delta y, \\
\Delta M &\approx dM = dx - dy \approx \Delta x - \Delta y, \\
\Delta T &\approx dT = xdy + ydx \approx x\Delta y + y\Delta x, \\
\Delta Q &\approx dQ = \frac{ydx - xdy}{y^2} \approx \frac{\Delta x}{y} - \frac{x\Delta y}{y^2}.
\end{aligned}$$

Operations and Errors (Answer)

ΔS and ΔM are of the same size as Δx and Δy, so addition and subtraction do not exacerbate errors in representation of the components of these operations.

However, ΔT and ΔQ also depend on the size of x and y, so multiplication and division can magnify round-off errors in representation.

Homework

Homework assignment section 1.2, due: TBA

Overview

Definitions

Definition 16

Algorithm: unambiguous procedure or steps to follow in a specified order.

Pseudo-code: a description of an algorithm including inputs and the form of the output.

Loops: steps to be repeated with possible changes to variables based on a counting index.

Conditional statements: If ... then ... else ...

Condition controlled loops: loop who's termination is based on a condition (While ... do ...), (Repeat ... until ...).

Example of Pseudo-code

Example 17

Generate pseudo-code for the the following problem: Given a function $f(x)$ is integrable on the interval $[a, b]$ approximate the definite integral of $f(x)$, over $[a, b]$, using the Midpoint Method on N uniform subintervals.

Characterizing Algorithms

One criteria imposed on an algorithm (whenever possible) is that small changes in the initial input produce correspondingly small changes in the final results. In this case the algorithm is said to be **stable**; otherwise it is **unstable**. Some algorithms are stable only for certain choices of input and are called **conditionally stable**.

Growth in Error

Definition 18

Suppose that $E_0 > 0$ denotes an error introduced at some stage in the calculations and E_n represents the magnitude of the error after n subsequent operations.

- If $E_n \approx CnE_0$, where C is a constant independent of n, then the growth of error is said to be **linear**.
- If $E_n \approx C^n E_0$, for some $C > 1$, then the error is called **exponential**.

Linear growth of error is usually unavoidable, and when C and E_0 are small, the results are generally acceptable. (Stable)

Exponential growth of error is bad! (Unstable)

Iterative Techniques

Example 19

Approximate the definite integral of $f(x)$, over the interval $[a, b]$, using the Midpoint Method (as above) for $N = 1, 2, 4, 8, ...$ until the difference between successive iterations differ by no more than 10^{-4}.

Rate of Convergence

Definition 20

Suppose $\{\beta_n\}_{n=1}^{\infty}$ is a sequence known to converge to zero and $\{\alpha_n\}_{n=1}^{\infty}$ converges to a number α. If a positive constant K exists with

$$|\alpha_n - \alpha| \leq K|\beta_n|, \text{ for large } n,$$

then we say that $\{\alpha_n\}_{n=1}^{\infty}$ converges to α with **rate** (or **order**) **of convergence** $O(\beta_n)$ and we write $\alpha_n \to \alpha + O(\beta_n)$.

Rate of Convergence Example

Example 21

$x - \sin(x) \to 0$ as $x \to 0$. Find the rate of convergence.

Homework

Homework assignment section 1.3, due: TBA

Overview

Command Line and Basic Constructs

Example 22

Graph $\sin(x)$ and $\cos(x)$ for $0 \leq x \leq 2\pi$ using different styles and various partitions in Matlab.

M-files: Functions and Scripts

Example 23

Generate a program that has as input constants a, b and $n \geq 1$ and produces as output the Midpoint Method approximation on n subintervals to $\int_a^b f(x)\,dx$ where the function f is given by a specified M-file.

Then generate a script that calls this program with $n = 2^k$ for $k = 0, 1, 2, ..., 10$ and graphs the approximations to $\int_0^3 (1 + \sin(3x^2))\,dx$ vs. k.