

CSC349A Numerical Analysis

Lecture 2

Rich Little

University of Victoria

2018

Table of Contents I

1 MATLAB

2 Approximation and Roundoff Errors

- <https://matlab.engr.uvic.ca/student/>
- pgs. 27-39 - Elementary programming concepts and pseudocode
- pgs. 935-942 - Appendix B: Getting Started with MATLAB
- pgs. 43-47 - MATLAB and the free-falling example
- UVic MATLAB reference manual in ConneX - Lecture 2 folder (Note: outdated).
- MATLAB = *MAT*rix *LAB*oratory
- The matrix is the fundamental data structure.
- MATLAB is *interpreted* NOT compiled.

Table of Contents I

1 MATLAB

2 Approximation and Roundoff Errors

Types of error

- **Formulation or modeling error** arises because of incomplete mathematical models. For example the simplified parachutist model we described will never exactly predict the motion of a real parachutist.
- **Data uncertainty/inherent error** Noisy data due to inexact measurements or observation.
- **Truncation error (Chapter 4)** results from using inexact approximations instead of an exact mathematical procedure such as the errors between the Euler method numerical procedure for predicting velocity versus the analytical solution obtained through differential calculus.
- **Roundoff error (Chapter 3)** is due to the fixed, finite precision representations used to represent real/complex numbers in computers.

Ways to measure error

- **Significant digits** The digits of a number that can be used with confidence
- **Absolute error** Difference between actual value and approximation
- **Relative error** Normalized absolute error
- **True error** Error when actual value is known
- **Approximate error** Error when actual value is unknown

Absolute Error

If p denotes the true (exact) value of some quantity, and p^* denotes some approximation to p then the **absolute true error** is defined as:

$$|E_t| = |p - p^*| \quad (1)$$

The absolute error only makes sense if you have a sense of the magnitude of p , the quantity you are approximating.

Relative Error

A better approach is to introduce the concept of relative error, in which the magnitude of the exact value p is used to “normalize” the error. More specifically the **relative error** is defined as:

$$|\varepsilon_t| = \frac{|p - p^*|}{|p|} = \left|1 - \frac{p^*}{p}\right| \quad (2)$$

Significant Digits

The *significant digits* of a number are those that can be used with confidence. It is a discrete way of measuring error in approximations in contrast to the relative error which is a continuous way of measuring error in approximation. The relative error also indicates the number of correct significant digits in an approximation p^* . For example for $p = \pi = 3.14159265 \dots$:

approximations	number of significant digits	relative error
3.1	2	0.013
3.14	3	0.00051
3.141	4	0.00019
3.1415	5	0.000029

Relative error and significant digits

Frequently it is useful to relate the relative error to the number of correct significant digits. This can be done with heuristics like the following:

$$|\varepsilon_t| < 5 \times 10^{-n} \quad (3)$$

in which case p^* approximates p to **n significant digits**. Note that this is not an exact relationship and is a conservative estimate. It is possible that the approximation has more than n correct significant digits but not less.

When the true value is unknown

In order to calculate the relative or absolute error of a particular quantity we need to know the value of the approximation p^* as well as the true value p . In many cases we don't know the true value. In fact numerical methods in engineering are used when we don't know the true value otherwise we could just simply use it instead of computing a numeric approximation.

There are many algorithms in numerical methods that operate in an iterative function in which an estimate of the solution is used to compute a better estimate and the process is repeated until some convergence criterion is met. In this cases it is common to use our best current estimate of the true value (i.e the previous estimate) to compute the error compared to the current estimate.

Approximation to relative error

In this case of iterative algorithms we can use the following approximation to the relative error:

$$|\varepsilon_a| = \frac{|p_i - p_{i-1}|}{|p_i|} \quad (4)$$

Similarly to the imprecise conservative estimation of significant digits above we can use the following relation to estimate the number of significant digits for the relative error in iterative algorithms:

$$|\varepsilon_a| < 0.5 \times 10^{-n} \quad (5)$$

An example

Mathematicians like to represent functions by infinite series.
For example the exponential function can be computed using:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \quad (6)$$

As more terms are added to the series the approximation becomes better and better. This is called a McLaurin series expansion. Let $p_1 = 1$, $p_2 = 1 + x$, $p_3 = 1 + x + \frac{x^2}{2!}$ and so on.

Approximate at different precisions

We can use this series approximation at different levels of precision to compute the value of e^x for $x = 0.5$. We can use the true value $p = e^{0.5} = 1.648721..$ to compute the true relative error $|\varepsilon_t|$.

i	p_i	$ \varepsilon_t = \frac{ e^{0.5} - p_i }{ e^{0.5} }$	$ \varepsilon_a = \frac{ p_i - p_{i-1} }{ p_i }$
1	1	0.393	
2	1.5	0.0902	0.333
3	1.625	0.0144	0.0769
4	1.645833	0.00175	0.0127
5	1.6484375	0.000172	0.00158
6	1.6486979	0.0000142	0.000158

Note that the value of $|\varepsilon_t|$, which can only be computed if we know the true (exact) answer p , can be used to estimate the number of correct significant digits in each approximation. For example: For p_3 we have $0.0144 < 5 \times 10^{-2}$ so the approximation $p_3 = 1.625$ has at least 2 correct significant digits. For p_5 we have $0.000172 < 5 \times 10^{-4}$ and therefore $p_5 = 1.6484375$ has at least 4 correct significant digits.

What happens when true value is unknown ?

In practice, if a sequence of approximations to some unknown value is computed using an iterative algorithm, (we will use several such algorithms during this course), then the exact relative error $|\varepsilon_t|$ can not be computed. However, the relative error in each approximation p_i can be approximated by $|\varepsilon_a|$ and can be used to estimate conservatively the number of correct significant digits. For examples for $i=6$ in the table above we have $|\varepsilon_a| = 0.000158 < 0.5 \times 10^{-3}$ implying that $p_6 = 1.6486979$ has at least 3 correct significant digits (notice that in fact it has 4 which means this is a conservative estimate).