Chapter 1 Key Terms & Assignments

Class #1

Section 1.1 Algorithms Read pages 1-16
Note there is a typo (error) in Equation 3 page 3;
it should have the last term as $\mathbf{ce^{-kt}}$, the minus sign is missing.

Key Terms:
Algorithm
Trapezoidal Rule
Absolute Error
Iterative method
Approximation of a square root

Exercises for Section 1.1: #2, 3, 4, 6, 15

For 2, 3, 4, 6, use your calculator and record the steps.
For 15 you can use MATLAB routine **bradie_sec1_1_num15.m**
Use command **bradie_sec1_1_num15(Nmax, 5E-3)** where you specify a value for **Nmax.**
In MATLAB type **help bradie_sec1_1_num15** for information.

In preparation for the next class Read pages 20-27

++

Class #2

Section 1.2 Convergence, covers Pages 20 - 27;
there are important concepts and terminology in this section.

Key Terms:
Rate of Convergence
Big O meaning & notation
Benchmark
Order of convergence
Asymptotic Error Constant
Taylor's Theorem

Exercises for Section 1.2: #1, 2a,b, 3, 6, 9, 13, 14, 15, 16, 17

For 2a, b use Taylor series expression with a few terms.
For 3 determine a table of values for x = 1. 0.1, 0.01, 0.001,
keeping as many digits as your calculator permits. Then use Taylor series expressions for the sine functions that appear.
For 6 use the contents of the sentence below the definition that is on Page 23.
For 9 set up a table like the one that appears in Example 1.7.
Exercises 13-15 are important for later theoretical work. You are responsible for reading and understanding the steps in the proofs of these exercises. You will be given the proofs.
You will be assigned one of 16 or 17.

In preparation for the next class Read pages 30-39

++

Class #3

Section 1.3 Floating Point Number Systems, covers pages 30 – 39
Make sure you know the definition of roundoff error. Also **read the article by Cleve Moler on Floating Point Numbers and Arithmetic in MATLAB** and **answer the questions on t he accompanying Floating Points sheet.**

Key Terms:
Discretization/Truncation Error
Roundoff error
Conditioning/ill conditioned
Stability
Absolute error
Relative error
Significant digits

Exercises for Section 1.3: #1a,b,e, 5, 6a,b, 7, 15, 16, 18

For #1 use your calculator and set up table with headings like the following.

	Chopping		Rounding	
y	$fl(y)$	error	$fl(y)$	error

For #5 List the type of calculator and use the following: Machine precision with rounding is $0.5*10^{1-k}$, the smallest number positive number is $(0.1)*10^{m}$, and the largest positive number is $(1-10^{-k})*10^{M}$.
For #6a, b just determine the number of significant digits in base 10.
For #7 you are told that the values have been rounded to the digits shown. So here you must determine an appropriate interval that contains the value of a parameter. For example in part (a) P = 0.750 atm, thus we must consider a range of values ±0.0005 from the given value ➔ 0.7495 < P < 0.7505. Do this for each parameter, then set up an interval that contains the desired expression.
For #15 In parts b and c relative change compares with part (a).
For #16 follow the ideas in #17 which will either be done in class or given to you in a handout.

In preparation for the next class Read pages 42 – 50.

++

Class #4

Section 1.4 Math on the Computer, covers pages 42 – 50.
In particular Example 1.11 reveals how floating point arithmetic impact solving linear system of equations.

Key Terms:
Floating point arithmetic
Normalized decimal form
Roundoff error
Pitfalls of floating point arithmetic

Exercises for Section 1.4: 1a,b, 2, 3, 7a, 12a,c,d

For #1 follow the ideas in the example below.
For #12 consider using trig identities, properties of logs, etc.

Example: 1c; compute the 4-digit rounding value of $\pi\ln(2) + \sqrt{10}\cos(22°)$.

Step 1. Find the "float values of each value:

fl(π) = fl(3.141592654) = 3.142
fl(ln(2)) = fl(0.6931471806) = 0.6931
fl($\sqrt{10}$) = fl(3.16227766) = 3.162
fl(cos(22°) = fl(0.9271838546) = 0.9272

Step 2. Compute the products in the floating point system.

fl(3.142 × 0.6931) = fl(2.1777202) = 2.178
fl(3.162 ×0.9272) = fl(2.9318064) = 2.932

Step 3. Compute the sum in floating point system.

fl(2.1278 + 2.932) = fl(5.110) = **5.110**

++