# Numerical Analysis 

Math 370 Fall 1998
MWF 11:30am - 12:25pm
(c) 1998 Ron Buckmire

## Class 3: Friday September 4

## SUMMARY Floating Point Arithmetic

CURRENT READING Burden \& Faires Sections 1.2 and 1.4

## Chopping versus Rounding

Let's consider the relative error involved in using chopping versus rounding in the representation of machine numbers with $k$ decimal digits.

Any positive real number $y$ can be normalized to be written in the form

$$
y=0 . d_{1} d_{2} d_{3} \cdots d_{k} d_{k+1} \cdots \times 10^{n}
$$

This is called a floating point number. It is clear that if one restricts the number of digits which can be used to represent the floating point number one will have to either perform chopping or rounding in order for the floating point number to "fit" the allotted memory of $k$ digits.

## Exercise

Write down an expression which represents $\epsilon_{\text {rel }}$, the relative error involved in representing $y$ as a decimal machine number $f l(y)$ (i.e. a floating point number using a fixed number of decimal digits)

## Example

(a) Write down what $f l(y)$ would look like if chopping were implemented
(b) Write down what $f l(y)$ would look like if rounding were implemented

## GROUPWORK

Show that the expression involving $k$ which gives you an upper bound for the relative error involved in using chopping arithmetic is $\epsilon_{r e l}=10^{-k+1}$

## Example

It can be shown that a bound for the relative error involved in using rounding arithmetic is half that for chopping, $\epsilon_{\text {rel }}=0.5 \times 10^{-k+1}=5 \times 10^{-k}$.

## GROUPWORK

Compute the following arithmetic operations using finite precision arithmetic.
Half the group should use five-digit chopping arithmetic and the other half five-digit ronding arithmetic.
Also compute the absolute and relative error from the "actual value."
What results do you expect?
(a) $\frac{1}{5} \oplus \frac{5}{7}=$
(b) $\frac{1}{5} \ominus \frac{5}{7}=$
(c) $\frac{1}{5} \otimes \frac{5}{7}=$
(d) $\frac{1}{5} \oslash \frac{5}{7}=$

