CS 450: Numerical Anlaysis[1]

Introduction to Scientific Computing

University of Illinois at Urbana-Champaign

[1] *These slides have been drafted by Edgar Solomonik as lecture templates and supplementary material for the book "Scientific Computing: An Introductory Survey" by Michael T. Heath (slides).*

Scientific Computing Applications and Context

- **Mathematical modelling for computational science** *Typical scientific computing problems are numerical solutions to PDEs*
  - *Newtonian dynamics: simulating particle systems in time*
  - *Fluid and air flow models for engineering*
  - *PDE-constrained numerical optimization: finding optimal configurations (used in engineering of control systems)*
  - *Quantum chemistry (electronic structure calculations): many-electron Schrödinger equation*
- **Linear algebra and computation**
  - *Linear algebra and numerical optimization are building blocks for machine learning methods and data analysis*
  - *Computer architecture, compilers, and parallel computing use numerical algorithms (matrix multiplication, Gaussian elimination) as benchmarks*

Example: Mechanics[2]

- Newton's laws provide incomplete particle-centric picture
- Physical systems can be described in terms of *degrees of freedom* (DoFs)
  - A piston moving up and down requires _____ DoFs
  - 1-particle system requires _____ DoFs
  - 2-particle system requires _____ DoFs
  - 2-particles at a fixed distance require _____ DoFs
- N-particle system *configuration* described by $3N$ DoFs

[2] *Variational Principles of Mechanics*, Cornelius Lanczos, Dover Books on Physics, 1949.

Course Structure

- **Complex numerical problems are generally reduced to simpler problems**

- **The course topics will follow this hierarchical structure**

Numerical Analysis

- **Numerical Problems involving Continuous Phenomena**:

- **Error Analysis**:

Sources of Error

Demo: Floating Point vs Program Logic

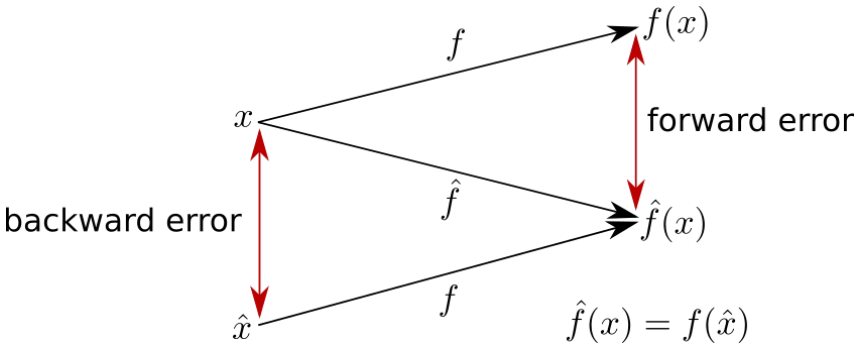
- **Representation of Numbers**:
- **Propagated Data Error**:
- **Computational Error =** $\hat{f}(x) - f(x)$ **= Truncation Error + Rounding Error**

Error Analysis

- **Forward Error**:

- **Backward Error**:

Visualization of Forward and Backward Error



Conditioning

Demo: *Conditioning of Evaluating tan*

- **Absolute Condition Number**:

- **(Relative) Condition Number**:

Posedness and Conditioning

- **What is the condition number of an ill-posed problem?**

Stability and Accuracy

- **Accuracy**:

- **Stability**:

Demo: Truncation vs Rounding

Error and Conditioning

- Two major sources of error: *roundoff* and *truncation* error.
  - roundoff error concerns floating point error due to finite precision
  - truncation error concerns error incurred due to algorithmic approximation, e.g. the representation of a function by a finite Taylor series

- To study the propagation of roundoff error in arithmetic we can use the notion of conditioning.

Floating Point Numbers

Demo: Picking apart a floating point number
Demo: Density of Floating Point Numbers

- **Scientific Notation**

- **Significand (Mantissa) and Exponent** Given x with s leading bits $x_0, \ldots, x_{s-1}$

Rounding Error

Demo: *Floating point and the Harmonic Series*
Demo: *Floating Point and the Series for the Exponential Function*

- **Maximum Relative Representation Error (Machine Epsilon)**

$1.0110(1) \times 2^{-3}$

significand

$\epsilon = \underset{\epsilon > 0}{\operatorname{argmin}} \; fl(1+\epsilon) = 1+\epsilon$

1.0000×2^{0}

$+\,0.0001 \times 2^{-1}$

1.0001

Rounding Error in Operations (I)

Demo: Catastrophic Cancellation
Activity: Cancellation in Standard Deviation Computation

- **Addition and Subtraction**

Cancellation – loss of sig. digits

$1.247 \quad (1+\epsilon)$

$-1.233 \quad (1+\epsilon)$

$= .014 \Rightarrow 1.4 \times 10^{-2} + \epsilon \cdot 1.247 + \epsilon 1.233$

$$\kappa_{add}(x,y) = \lim_{\delta x \to 0} \left| \frac{(x+\delta x)+y - (x+y)}{x+y} \right| \Big/ |\delta x| |x|$$

$$= \lim_{\delta x \to 0} \frac{\left| \frac{\delta x}{|x+y|} \right|}{|\delta x| / |x|} = \frac{|x|}{|x+y|}$$

$y \approx -x$ → ∞

$f(x) = x + y$

$$\kappa_f(x) = \frac{|f'(x)| \cdot |x|}{|f(x)|} = \frac{|x|}{|x+y|}$$

$$\kappa_{abs}(x) = |f'(x)|$$

(relative) $\kappa_f(x) = \kappa_{abs}(x) \Big/ \frac{|f(x)|}{|x|}$

Rounding Error in Operations (II)

Demo: *Polynomial Evaluation Floating Point*

- **Multiplication and Division**

Exceptional and Subnormal Numbers

- **Exceptional Numbers**

not normalized numbers ($1.0110\ldots \times 2^{\text{signal/exp}}$)

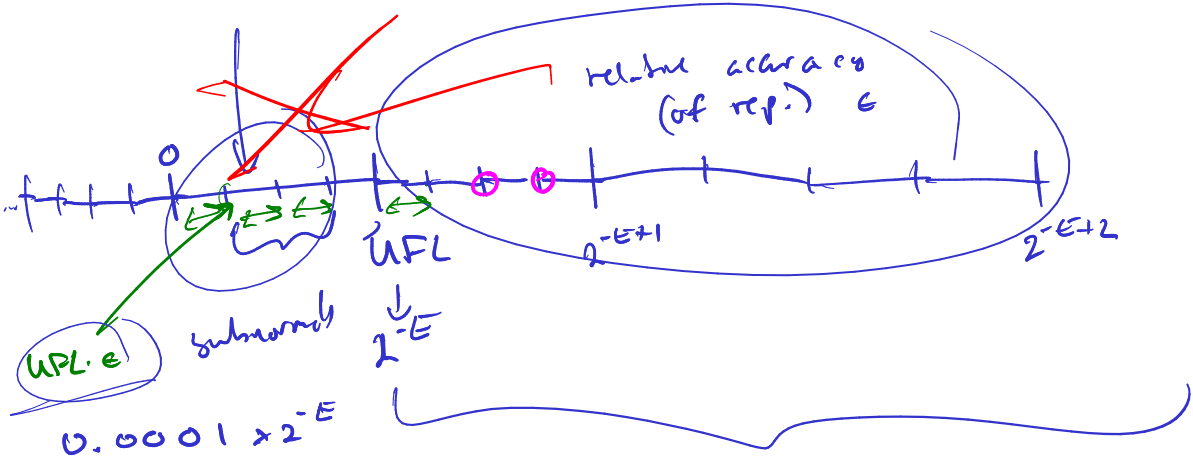
$0, -0, \infty, -\infty, NaN = 0/0 = \infty - \infty$

- **Subnormal (Denormal) Number Range**

$0.\underbrace{\text{signal/exp}} \times 10^{-1023}$ | UFL $1.0000\ldots0 \times 10^{-1023}$

smallest possible exponent

- **Gradual Underflow: Avoiding underflow in addition**



$fl(fl(x) - fl(y)) \neq 0$

$\Leftrightarrow fl(x) \neq fl(y)$

because of subnormal

Floating Point Number Line