Problem Set 1

Due: Fri Sept 1

Reading: LeVeque: Chapters 1, 9, E.1, E.2; MATLAB: Chapter 1.1, 1.7 (IEEE floating point)

Please label with APM 522, PS1, and your name; write up the problems neatly and in numerical order; and write (2) omitted if problem (2) is not attempted. No need to restate problem. Derive your answers! If you typeset the whole assignment, you may email me a PDF instead of turning in a hardcopy.

(1) Suppose a double precision floating point number is stored on a computer using 64 bits in the following way: sign 1 bit, exponent 8 bits, and mantissa 55 bits. A given real number \( r \) is written as

\[
r = \pm m 2^n
\]

where the mantissa \( m \) satisfies \( 1/2 \leq m < 1 \) and \( -128 \leq n \leq 127 \). Give the following numbers in both base 2 and base 10 scientific notation:

(a) What is the largest positive number \( \text{realmax} \) that can be stored?

(b) What is the smallest positive number \( \text{realmin} \) that can be stored?

(c) What is the machine epsilon \( \epsilon_M \) (take the leading 1 in \( m \) to be a phantom)?

For problems (2) and (3), write formulas for the leading error term as

\[
\frac{du}{dt} \approx \frac{u^{n+1} - u^n}{\Delta t} = u' + \frac{u''}{2} \Delta t + \cdots
\]

or

\[
\left( \frac{d^2f}{dx^2} \right)_i \approx \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta x^2} = f''_i + \frac{\Delta x^2}{12} f^{(4)}_i + \cdots
\]

(2) (a) Verify that the three-point central difference formulas for \( df/dx \) and \( d^2f/dx^2 \) are second-order accurate.
(b) Verify that the one-sided difference formula
\[ \frac{du}{dt} \approx \frac{u^{n+1} - u^n}{\Delta t} \]
is first-order accurate.

(3) Verify that the approximation
\[ \left( \frac{df}{dx} \right)_j = \frac{1}{\Delta x} \left[ -\frac{1}{12} f_{j+2} + \frac{2}{3} f_{j+1} - \frac{2}{3} f_{j-1} + \frac{1}{12} f_{j-2} \right] \]
is fourth-order accurate.

(4) Show that the general solution to the heat equation
\[ u(x,t) = \int_{-\infty}^{\infty} K(x - y, t) u_0(y) dy \]
where the fundamental solution or kernel
\[ K(x,t) = \frac{1}{\sqrt{4\pi\kappa t}} \exp \left\{ -\frac{x^2}{4\kappa t} \right\} \]
does in fact satisfy \( u_t = \kappa u_{xx} \) with \( u(x, t = 0) = u_0(x) \). Hint: First show that \( K \) satisfies the heat equation. Then argue that \( u \) satisfies the heat equation. Finally show that \( u \) satisfies the initial conditions \( u(x, t = 0) = u_0(x) \).

(5) Show that the TR method for the heat equation \( u_t = u_{xx} \)
\[ u_i^{n+1} = u_i^n + \frac{\Delta t}{2\Delta x^2} \left( u_{i+1}^n - 2u_i^n + u_{i-1}^n + u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1} \right) \]
is second-order accurate, using the definition of the local truncation error:
i.e., show that the global error \( \tau = c_1 \Delta t^2 + c_2 \Delta x^2 + \cdots \), where “\( \cdots \)” means higher order terms in \( \Delta t \) and \( \Delta x \). (Fill in the values of \( c_1 \) and \( c_2 \).) Hint: Use the facts that
\[ \frac{u(x + h) - 2u(x) + u(x - h)}{h^2} = u_{xx} + \frac{h^2}{12} u_{xxxx} + \cdots , \quad h \equiv \Delta x \]
and
\[ u_{xx}(t + \Delta t) = u_{xx} + \Delta t u_{xxt} + \frac{\Delta t^2}{2} u_{xxtt} + \cdots \]