AMSC/CMSC 466 Dr. Wolfe ASSIGNMENT #1 Due September 20, 2004

1. Using four digit arithmetic, add the following numbers, first in ascending order (from smallest to largest) and then in descending order. In doing so round off the partial sums to four significant figures. Compare your results with the correct sum $x = 0.107101023$ E5. ($En = 10^n$.)

0.1580 E0	0.4288 E1	0.7767 E3
0.2653 E0	0.6266 E2	0.7889 E3
0.2581 E1	0.7555 E2	0.8999 E4

2. In IEEE *single precision floating-point arithmetic* a non-zero machine number x is represented as

$$x = \pm 1.a_1 a_2 \dots a_{23} \cdot 2^e$$

where $a_i = 0$ or 1 and $-126 \le e \le 127$. Find the largest integer n such that $n!$ can be exactly represented as a machine number.

3. Ex.7, p.9, *Atkinson & Han.* . Also, use MATLAB to produce a picture like fig 1.3 on p.7.

4. Suppose a computer carries three decimal digits and rounds. If x and y are machine numbers, define the machine version of addition $x \oplus y$ to be the result of adding x and y and rounding to three digits. For example

$$49.3 \oplus 57.4 = 107.$$

Define machine multiplication $x \otimes y$ similarly. For example,

$$1.23 \otimes 4.86 = 5.98.$$

Constuct examples to show that, in general, the following statements are <u>not</u> true:
(a) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$,
(b) $(x \otimes y) \otimes z = x \otimes (y \otimes z)$,
(c) $x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$.

5. Ex.9(c), p.19, *Atkinson & Han.*

6. Ex.6(c),(d),(f), p.54, *Atkinson & Han.*

7. Ex.1, p.62, *Atkinson & Han.*

8. The **error function** is an important function in many branches of applied mathematics. It is defined by an integral.

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, dt.$$

The integral cannot be expressed in terms of more elementary functions. It is a library function in MATLAB. The Taylor series for the error function is

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!(2n+1)}.$$

The series converges for all x. Write a MATLAB function m-file to evaluate erf(x) using this series. Use as many terms in the series as are necessary so that the first neglected term does not alter the accumulated sum when it is added to it in floating point arithmetic. Since this is an alternating series, the error caused by truncating the infinite sum will then be less than the roundoff error. Investigate the effect of roundoff error by comparing the computed sum with the value given by the MATLAB function erf. Try $x = 0.5, 1.0, 2.0, 5.0, 7.0, 10.0$. Explain your results. *Hint:* The main loop of your program might look like this:

while $s \sim= so$
$so = s;$
$n = n + 1;$
$t = -xqt * t * (2.0 * n - 1.0)/(n * (2 * n + 1));$
$s = s + t;$
end

What values should be assigned to t, n, s and xqt before entering the loop ? Do not forget the factor $2/\sqrt{\pi}$.