Ryan Brosnahan

Homework Set 1

1. Write .000304569432608 in normalized form in base 10 (scientiﬁc notation) with a precision of 9 digits using rounding.

2. Construct a ﬂoating point system (by specifying the mantissa precision t and how many binary digits m should be used in the exponent) for which machine epsilon is less than 10−22 and the largest positive machine number is greater than 10300. Many answers are possible, but try to keep t and m as small as possible and still satisfy the conditions.

Using standard 64 bit double precision:

Exponent is 11 bit with range (-1022, 1023);

Mantissa is 52 bit; Thus, the smallest increment this system can handle is

Let’s try 70 bit:

If the system has to round 2 different numbers within an interval of this size, they could be indistinguishable.

3. Let y = 77302 be approximated by ˆy = 77000, and let z = .00023 be approximated by ˆz = .0002. For each pair, compute the absolute error and the relative error. Make a one or two sentence observation about the results – how does the magnitude of the numbers involved in the calculation aﬀect the absolute error and the relative error?

The absolute error of y seems like a big number, but the base numbers are big too so without context we cannot know if this would be detrimental to our goals.

This relative error is pretty small, in most situations < 0.4% noise is acceptable.

The absolute error of z seems like a small number, but the base numbers are also small so without context we cannot know if this would be detrimental to our goals.

This relative error is very large, in most situations > 13% noise is not acceptable.

4. Consider the calculation (A + B)/C on a machine that uses 6 digit rounding. Let A = 1.24854E0, B = 1.40032E − 5, and C = 1.20668E − 1. Perform the calculation as it would be done on the machine. Report your answer in normalized form.

Rounds to 1.24855

5. Solve by hand using 4-digit arithmetic with rounding using the most accurate formulas available. Then compute the roots more accurately (on your calculator, to 8 or 9 digits). If cancellation error is present in this problem, identify the exact spot where it occurs.

Some loss from rounding:

Cancellation error:

Some loss from rounding:

Some loss from rounding:

From wolframalpha:

It looks like 2-3% off, not bad for only 4-digit.

6. MATLAB BASICS – plotting: Write a script ﬁle that deﬁnes the inline function and plots it over the interval [−5, 5] with a resolution of 0.01. Cut and paste the script ﬁle and the plot into a Word document on one page.

%question 6

tic

x = -5:0.01:5;

y = (x+3) ./ ( (x.^2) + 5);

toc

figure(1)

plot(x,y)

title('question 6')

axis([-7 7 -.5 1])



7. MATLAB BASICS – function ﬁles. Write a function ﬁle, called fcnhw1n7.m, that

accepts 3 inputs, x, y and z, and outputs the 2 quantities

Write a script ﬁle, called hw1n7.m, that calls the function twice, ﬁrst with (x, y, z) = (1, 2, 3) and then with (x, y, z) = (−4, 1, 7) and prints the results with at least 10 signiﬁcant digits. Submit the code listings and the numerical results.

%question 7

function [a, b] = fcnhw1n7(x, y, z)

a = (x^2) / y + 3\*z;

b = ( exp (x) \* sin(y) )/ sqrt((x^2) + 1);

end

%question 7 cont

[a b] = fcnhw1n7(1, 2, 3)

[a b] = fcnhw1n7(-4, 1, 7)

>> hw1n7

a =

9.500000000000000e+00

b =

1.747774691014265e+00

a =

37

b =

3.737978138937480e-03