Subject: Re: Inaccuracies
Posted by C R Shaw on Thu, 16 Nov 1995 08:00:00 GMT
View Forum Message <> Reply to Message

Andy Loughe wrote..

```
> Ok, I am sure this has been discussed before, but let
> me start this thread again. I wish to create a 15-element
> vector which contains the numbers -1.4 to 1.4 by an increment
> of 0.2 I also wish the sum of these elements to be zero
> (No, this isn't the new math). Here is what I tried...
> TRIAL #2
> ======
> IDL> a = dindgen(15)*(.2D)-1.4D
> IDL> print, total(a, /double)
> 4.4408921e-15
> Ok. this is better but not correct.
> And what are the values of a?
> IDL> print, a
    -1.4000000
                   -1.2000000
                                  -1.0000000
                                                 -0.80000000
>
     -0.60000000
                    -0.40000000
                                    -0.20000000
                                                   2.2204460e-16
>
                                    0.60000000
>
     0.20000000
                    0.40000000
                                                   0.80000000
                                   1.4000000
     1.0000000
                    1.2000000
>
> I seem to have lost a zero somewhere, and for me this matters!!!
> Maybe I am missing something here, but this kind of behavior
> makes IDL a bit problematical for scientific use. With only 15
> numbers and double precision arithmetic. I can't believe this
> would fail in FORTRAN or C!
```

I've had a similar problem with numerical accuracy. The problem, however, appears to be in the representation of the values using the IEEE (I think) floating point standards.

It is also not just limited to IDL programs. The program below is a C program to do the same calculations...

```
/*
PROGRAM: test.c
PURPOSE: Tests the numerical accuracy of C
HISTORY: Written by Carl Shaw, Nov 1995
*/
```

```
#include <stdio.h>
main()
double count, value, total;
for (count=0.; count < 15.; count++)
 value=count*0.2-1.4;
 printf("%g ", value);
 total=total+value;
printf("\nTotal = \%g \n", total);
And guess what the results are???
-1.4 -1.2 -1 -0.8 -0.6 -0.4 -0.2 2.22045e-16 0.2 0.4
0.6 0.8 1 1.2 1.4
Total = 4.44089e-15!
Sound familiar?
For most applications though, a value of
0.000000000000004
is close enough to zero!
Mail: Dept. of Pure and Applied Physics,
     Queen's University Belfast,
     University Road.
     Belfast BT7 1NN
E-mail: C.Shaw@QUB.ac.uk
Phone: (01232) 245133 Ext 3045
Fax: (01232) 438918
```