Subject: Re: converting floats to doubles
Posted by Dick Jackson on Fri, 20 May 2005 20:17:57 GMT
View Forum Message <> Reply to Message

Hi,

"Benjamin Hornberger" <benjamin.hornberger@stonybrook.edu> wrote in message
news:428e3721_4@marge.ic.sunysb.edu...
> Hi computation gurus,
>
> is dblarr(n) equivalent in precision to double(fltarr(n))? I know that in
> a case like sqrt(dblarr(n)) vs. double(sqrt(fltarr(n))), they are not
> equivalent (the second version is not true double precision). But I
> thought when I start with whole numbers anyway, it might be the case.
>
> In other words, when a floating point number is converted to double, are
> the additional digits always set to zero, or is it possible that they
> aren't? I tried it out by printing some numbers, and it looks like they
> add only zeroes, but I would be happy if the experts could confirm.

You're right, it's good to be careful about these things, but indeed there
are a lot of integers that are precisely correct in Float (and even more in
Double). Empirically:

;; Run a loop until the Double version of an integer is not equal to
;; the Float version (this took several seconds to run)

IDL> for i=0D,1D9 do if i ne Double(Float(i)) then break

;; Variable 'i' (Double) has the first mismatch...

IDL> print,i,format='(F20.10)'
 16777217.0000000000
IDL> print,Float(i),format='(F20.10)'
 16777216.0000000000

Well, look at that, the number of good integer Floats is 256^3:
IDL> print,256L*256*256
    16777216
... which makes all kinds of sense, as a Float has 3 bytes for the mantissa
(or significand). For more, see:

http://en.wikipedia.org/wiki/Floating_point

Cheers,
--
-Dick

Dick Jackson / dick@d-jackson.com
D-Jackson Software Consulting / http://www.d-jackson.com
Calgary, Alberta, Canada / +1-403-242-7398 / Fax: 241-7392