Subject: Re: Big arrays, reducing data
Posted by Paul Van Delst[1] on Wed, 21 Mar 2007 19:53:53 GMT
View Forum Message <> Reply to Message

Eric Hudson wrote:
> Hi,
>
> I have what I hope is an easy question and then probably a hard one.
>
> 1) I need to make some big arrays (ideally 16000^2 elements or more)
> but find I often get "unable to allocate memory" errors. Is there
> some way of determining (at run time) the largest array that I can
> make? In C, for example, I'd try to allocate the memory and check for
> whether it was allocated, then cut the array size if it wasn't. Is
> there an equivalent technique in IDL?

Too hard for me. I'll tackle the easy question below..... :o)

> 2) The reason I want to make these big arrays is that I have sets of
> on the order of 20,000 data curves (50-200 pts each). I'd like to
> reduce these to a set of "common curves" -- around 100 averaged/
> extracted/smoothed curves which are representative of the larger set.
> The curves are complex -- I don't have anything to fit to them -- and
> they are noisy. But I get the feeling that if I handed someone a
> stack of 20,000 of them and said "sort these into groups which are
> similar" that they'd be able to do it. The question is, is there a
> good way to do this programmatically?

Eigenvector decomposition/reconstruction maybe? We do that sort of thing for radiometric spectra and atmospheric profiles by decomposing a dataset (sometimes statically, sometimes pseudo-dynamically) and then reconstructing data with less than the full number of eigenvectors.

Of course, it depends a lot on how variable your data is (e.g. you may be tossing away information not just "noise"[*]) and if there is sufficient redundancy (e.g. if you need combinations of 19000 eigenvectors to reconstruct your original 20000 curves, there may be no point).

This sort of analysis is very easy to do in IDL so it shouldn't take too long to test it to see if it's applicable.

cheers,

paulv

[*] I used quotes since one person's noise is another's signal.

--

Paul van Delst Ride lots.
CIMSS @ NOAA/NCEP/EMC Eddy Merckx