Subject: Re: remove duplicates WITHOUT sorting
Posted by Craig Markwardt on Mon, 12 Feb 2007 17:23:23 GMT
View Forum Message <> Reply to Message

"rpertaub@gmail.com" <rpertaub@gmail.com> writes:
> Hello All,
> I have an array of 2xn which is a list of coordinates (x,y positions)
> and I need to remove any duplicates in this array. Since they are x,y
> coordinates I cannot really sort the array, hence cannot use the Uniq
> function. Is there another way of doing this?
> My array is similar to:
> [2 30
> 4 5
> 3 9
> 51 19
> 12 15
> 3 9
> 11 22
> 32 33
> 14 25]
>
> As you see index 4,5 is same as index 10,11. How do i rewrite this
> array without duplicates? Is there a general way of doing this, as
> some of my coordinates (to make matters even more complicated) are not
> exactly the same, but are close (3,9) and (3.5,9) - But I think I can
> figure this out with some if...then statements. Not sure how to remove
> them though. Any help is appreciated!

If your array is small enough, I would make an array of distances
between all the points and focus on the points which are closer than
your desired tolerance. Something like,

```
x = reform(data(0,*))
y = reform(data(1,*))
;; Form into square arrays, replicate in x and y direction
xx = x # (y*0 + 1)
yy = (x*0 + 1) # y
;; Compute difference in X and Y coordinate
dxx = xx - transpose(xx)
dyy = yy - transpose(yy)
;; Compute distance
drr = sqrt(dxx*dxx + dyy*dyy)
```

The matrix is symmetric, so if you are really smart you can reduce
your work by 50%. This starts to get ugly for more than a few
thousand points. In which case you would indeed need to sort them
into (overlapping) bands or squares, and work on one chunk at a time.

Good luck!
Craig

--
---------- ----------
Craig B. Markwardt, Ph.D. EMAIL: craigmnet@REMOVEcow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
---------- ----------