Subject: Re: Find all points within a set of polygons
Posted by David Fanning on Wed, 30 Nov 2005 15:23:37 GMT
View Forum Message <> Reply to Message

Maarten writes:

> The single point version of the poly-tester, returns true if the point
> [pnt_x,pnt_y] lies within the polygon described by the (x,y) pairs in
> poly_x and poly_y:

You might want to try my INSIDE program. Just glancing at the
code, it looks a LOT faster than yours. :-)

http://www.dfanning.com/programs/inside.pro

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/

Subject: Re: Find all points within a set of polygons
Posted by Maarten[1] on Wed, 30 Nov 2005 15:43:38 GMT
View Forum Message <> Reply to Message

Thanks for the reply, that got rid of at least one loop. Funny how IDL
ends up doing things differently all the time: the sample C-code I
found denounces the angle method as horribly slow, and generally
recommends to use the number of border crossings on a half-line to
infinity (what I wrote above).

How I missed that file while searching your site I don't know, perhaps
adding a "Search Terms" field to the function header is a good idea (in
general this seems to classify as a "Point location" type of problem).

I'll report back when I've found a clever way of doing this for a lot
of boxes at once.
Maarten

Subject: Re: Find all points within a set of polygons
Posted by David Fanning on Wed, 30 Nov 2005 15:51:17 GMT
View Forum Message <> Reply to Message

Maarten writes:

> How I missed that file while searching your site I don't know, perhaps
> adding a "Search Terms" field to the function header is a good idea (in
> general this seems to classify as a "Point location" type of problem).

I don't know. It always comes up for me when I search for any
term I might put into a Search Term. Were you searching in French?
I don't do French. :-)

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/

Subject: Re: Find all points within a set of polygons
Posted by Maarten[1] on Wed, 30 Nov 2005 16:03:48 GMT
View Forum Message <> Reply to Message

David Fanning wrote:
> Maarten writes:
>
>> How I missed that file while searching your site I don't know, perhaps
>> adding a "Search Terms" field to the function header is a good idea (in
>> general this seems to classify as a "Point location" type of problem).
>
> I don't know. It always comes up for me when I search for any
> term I might put into a Search Term. Were you searching in French?

Nope.

> I don't do French. :-)

Neither do I, but I'm not a native speaker, so I might come up with
slightly different search terms anyway. But now that I've read the page
(and found out about the IDLanROI object), I really wonder how I missed
that page, as I think it does contain the terms I searched for.

In any case, I have enough material to keep me busy for a few days...

Maarten

Subject: Re: Find all points within a set of polygons
Posted by JD Smith on Wed, 30 Nov 2005 20:45:15 GMT
View Forum Message <> Reply to Message

On Wed, 30 Nov 2005 07:14:15 -0800, Maarten wrote:

> Hi,
>
> I'm trying to compare satellite images from two different polar
> sun-synchronous satellites. One dataset is high resolution (1 by 1 km,
> from MODIS/Aqua, if you're curious), the other has a lower spatial
> resolution (though still pretty high at about 15 by 15 km, from OMI/Aura).
>
> I'd like to attach a number to each MODIS pixel telling me in which "low
> resolution" box it is located - allowing me to continue with histogram and
> its reverse indices to extract relevant averages from the MODIS data.
>
> I have some code which tells me if a point falls within a specific box,
> but that would require a loop over all boxes and all pixels until a
> matching box is found - with ~ 99000 boxes in an orbit, things get tedious
> and slow very fast. I /think/ I can adapt this code to do either loop in
> one go, but not both loops at the same time - at least I have no idea
> where (no pun intended) to start.
>

Here's a good reference:

http://astronomy.swin.edu.au/~pbourke/geometry/insidepoly/

I believe Mark Hadfield has coded up one or more of these in his
Motley library, which at least vectorizes over the points to test.
Since all your polygons are the same length (4), it is possible to
simultaneously vectorize over both points and polygons (this would not
be true for arbitrary-sized polygons).

It sounds like you could use Solution 3 from the above reference for
convex polygons, i.e. see if a given point is to the same side of all
line segments in your polygons, by testing the sign of the quantity:

 (y-y0)(x1-x0)-(x-x0)(y1-y0)

for each polygon line segment, and concluding it's inside if all the
signs agree. How would this look vectorized? Something like the
function below. Notice it allows any arbitrary shaped array of x,y
input points, and just appends a dimension for the number of polygons.

However, similar to the "points inside a sphere" problem I just
commented on, you'll run out of memory fast using such a brute force
method. The algorithm takes several arrays of size

npts*npoly*mpoly*4, where npts is the number of points to test, npoly is the number of polygons to test against, and mpoly is the length of each polygon. You could probably test about 200 points at a time in 1GB memory against your 99000 polygons without hitting the memory limit (such a run takes about 15s on my system). Since a loop which processes 1GB/200 worth of memory at a time will not really feel the IDL loop penalty, you could just as well loop over those 200 points without a big loss of speed, or "chunk" them into 100 at a time or so. If you have multiple processors, it's more efficient to cast things into fairly large arrays.

For a smallish number of points and polygons, this brute force method will definitely be the fastest, thanks to the overhead of loops. However, for larger sized problems, just like for the spherical search case, you would do well to pre-reject points which definitely cannot be inside a given polygon. A similar method could be used: use HIST_2D on the points with a binsize similar to the typical size of each polygon. For each polygon, pre-compute the bounding-box in x and y, and from that compute which histogram bins lie within each bounding box. Then, perform inside/outside tests only on those points, and don't bother testing the rest. How much pre-selection to apply will depend on how many points vs. polygons, whether either set is fixed from calculation to calculation, etc.

JD

```
;+
; NAME:
;
;  POINT_IN_POLY_SET
;
; PURPOSE:
;
;  Determine whether a group of points are inside a set of 2D, convex
;  polygons of the same length.
;
; CALLING SEQUENCE:
;
;  res=point_in_poly_set(x,y,px,py)
;
; INPUTS:
;
;  x,y: The points to test, arrays of any format.  The result will be
;    an array of the same size, with a final trailing dimension of
;    length the number of polygons.
;
;  px,py: The polygons to test against, arranged as arrays of size
```

```
;    nxm, where n is the number of polygons, and m is their length
;    (which must be the same).
;
; OUTPUTS:
;
;  res: A tri-valued array the same size as the input x,y vectors,
;    with an additional trailing dimension with length n, the number
;    of polygons passed.  It indicates, for each point and each
;    polygon, whether the point was inside (1), outside (-1), or on
;    (0) the given polygon.
;
; RESTRICTIONS:
;
;  The polygons must be convex.
;
; PROCEDURE:
;
;  From solution 3 of:
;  http://astronomy.swin.edu.au/~pbourke/geometry/insidepoly/
;
; MODIFICATION HISTORY:
;
;
;   2005-11-30 (J.D. Smith): Writen
;-
 ;#####################################################################
####################
;
; LICENSE
;
;  Copyright (C) 2005 J.D. Smith
;
;  This file is free software; you can redistribute it and/or modify
;  it under the terms of the GNU General Public License as published
;  by the Free Software Foundation; either version 2, or (at your
;  option) any later version.
;
;  This file is distributed in the hope that it will be useful, but
;  WITHOUT ANY WARRANTY; without even the implied warranty of
;  MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the GNU
;  General Public License for more details.
;
;  You should have received a copy of the GNU General Public License
;  along with this file; see the file COPYING.  If not, write to the
;  Free Software Foundation, Inc., 59 Temple Place - Suite 330,
;  Boston, MA 02111-1307, USA.
;
 ;#####################################################################
```

```
####################

function point_in_poly_set,x,y,px,py
  spol=size(px,/DIMENSIONS)
  if n_elements(spol) ne 2 then message,'Polygons must be nxm array'
  spts=size(x,/DIMENSIONS)
  m=spol[0]

  pxdiff=shift(px,1,0)-px & pydiff=shift(py,1,0)-py

  rs=[1,1,spts]               ;reform size
  ts=[spol,spts]              ;working target size

   in_test=(rebin(reform(y,rs),ts,/SAMPLE)-rebin(py,ts,/SAMPLE) )* $
         rebin(pxdiff,ts,/SAMPLE) - $
         (rebin(reform(x,rs),ts,/SAMPLE)-rebin(px,ts,/SAMPLE))* $
         rebin(pydiff,ts,/SAMPLE)

  ;; For each point and polygon, test for all negative or all positive
  ;; (inside polygon), mixed (outside) or any exactly 0.0 (on polygon)
  sign=1 - 2*(in_test lt 0.0) - (temporary(in_test) eq 0.0)

  ;; Inside/Outside polygon test: -m, or m: inside, otherwise: outside or on
  out=long(total(sign,1,/INTEGER))

  ;; On polygon test: on one or more polygon line segments, and to the
  ;; same side of all the others
  on=(product(sign,1,/INTEGER) eq 0LL) AND $
     (abs(out) eq long(total(temporary(sign) ne 0,1,/INTEGER)))
  out=temporary(out)/m eq 0L

  if size(out,/N_DIMENSIONS) le 1 then return,1-2*out+on
  return,transpose(1-2*out+on,[1+indgen(n_elements(spts)),0])
end
```

Subject: Re: Find all points within a set of polygons
Posted by Maarten[1] on Thu, 08 Dec 2005 15:53:24 GMT

View Forum Message <> Reply to Message

Hi,

JD Smith wrote:

> Here's a good reference:
>
> http://astronomy.swin.edu.au/~pbourke/geometry/insidepoly/

Thanks, that helped, but David's pointer to IDL's own IDLanROI object is
what I use now.

> I believe Mark Hadfield has coded up one or more of these in his
> Motley library, which at least vectorizes over the points to test.
> Since all your polygons are the same length (4), it is possible to
> simultaneously vectorize over both points and polygons (this would not
> be true for arbitrary-sized polygons).

[snip]

> However, similar to the "points inside a sphere" problem I just
> commented on, you'll run out of memory fast using such a brute force
> method.

I decided to apply some prior knowledge to the problem and limit the number of points I have to search, reducing the memory requirements significantly. The trick? The points are geolocation points from MODIS,

and are neatly arranged. What is more important: I have a set of 1x1 km

gelocations and another set of 5x5 km geolocations. Now I search the coarse array first, take those indices and make a generous estimate of the
indices I can expect in the fine array. I then create a subset from the fine
arrays and search those instead, searching on the order of 800 points (selecting up to 500), instead of searching 2.7 million. In the end the

speedup is a factor of ~ 22, and that is without vectorizing the code.

> For a smallish number of points and polygons, this brute force method
> will definitely be the fastest, thanks to the overhead of loops.
> However, for larger sized problems, just like for the spherical search
> case, you would do well to pre-reject points which definitely cannot
> be inside a given polygon.

I could probably speed things up by further pre-selection, but for now I
think I have enough other things to do in each loop to get away with what
I have now.

Thanks,

Maarten