Concerning Simultaneous Maxima on the Penumbral Limiting Curve in Solar Eclipse Mapping

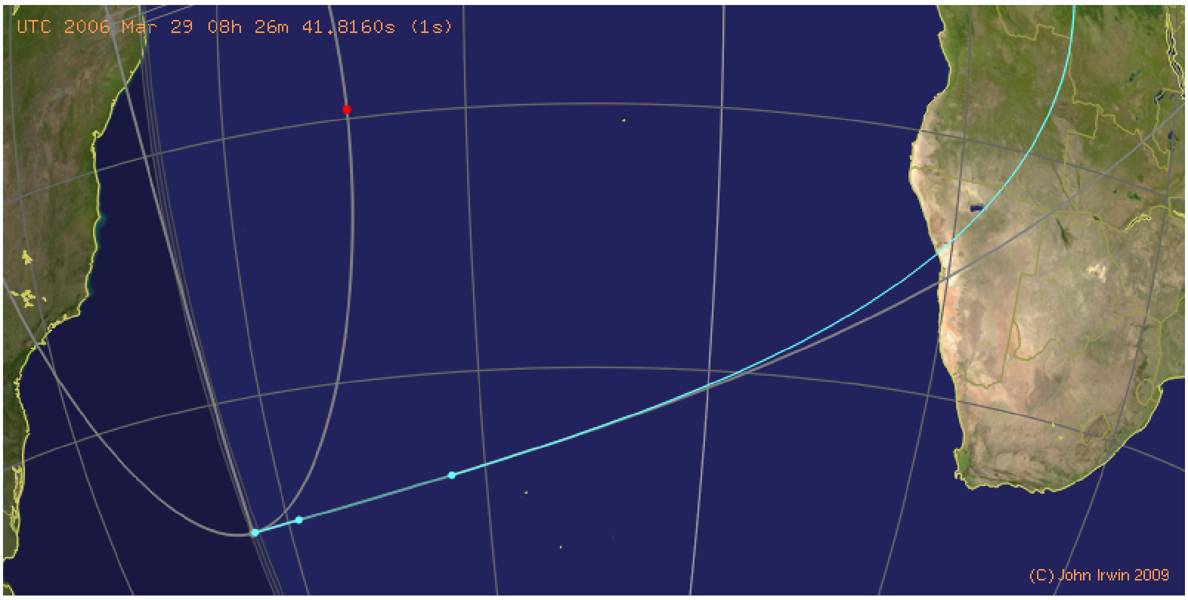
John Irwin

I would like to make some observations about Jean Meeus's "Strange Effect" [1], concerning the occurrence of two maximum eclipses at the same time on the limiting curve. This was also extended by Robert Nufer into lines of earliest/latest maximum eclipse [2], with which you may also be familiar. I make my observations because I do not believe they have been expressed before by anyone that I am aware of, in public anyway, even though the basis of my discovery (or realisation) is very simple. I have also come across an unexpected computational anomaly which someone may be able to shed some light on.

As is well known, the limiting curve is the locus of points at which the eclipse is beginning and ending at the same time, which means the curve is tangent to the outline of the shadow cone on the Earth's surface. It also means each point on the limiting curve is a maximum eclipse, though with zero magnitude. Now, from the discovery about the earliest and latest maximum eclipse, sometimes there are two points on the limiting curve which have maximum eclipse at the same time. So we can infer from this that there must be two points on the limiting curve that are tangent to the outline at the same time. In other words, the outline touches the limiting curve TWICE, rather than once as is usually the case.

This is a surprise because it is generally assumed that the outline curve only touches the limiting curve once, and that the first and last times it does this is at the extremes of the limiting curve. This is implicit in the Besselian method of computing the limiting curves using time as the independent variable (and we also know this method can sometimes fail at times approaching the extremes simply because this assumption is wrong).

Below I show an example of this phenomenon for the 2006 March 29 eclipse in the region of the extreme point PS1. The situation is depicted for the time TT:08:27:47 (UTC:08:26:41.8). The penumbral outline is bright-cyan at points where the eclipse is beginning, and dark-cyan where it is ending. I have also accurately calculated points on the outline where the eclipse changes between beginning and ending, and plotted them as bright-cyan points. For reference, the various curves of the eclipse map are included in the background, which have been calculated from bessel elements. The red point represents the location of the penumbra's first contact with Earth and the beginnings of the central track can be seen towards the top left hand corner of the image.



It is interesting to follow the time development of these contact points, and I have created an animation for the 2006 March 29 eclipse. You can find the animation file at:

```
http://www.jir1667.plus.com/simax/2006-Mar-29-double-contact-1.hav
```

You may not be familiar with the HAV format. It's a high-quality lossless video format, which I prefer over AVI. You can pick up a free HAV player from:

```
http://gromada.com/imagen/
```

I think you will require some patience to view this animation. The appearance of the contact points is sudden and transitory, but I also wanted to encompass a range of times before and after this event to show the outline when it is distinct from the limiting curve. If patience is not your forte, then you can use the slider on the HAV player to control the animation yourself.

I should make it clear at this stage that the data for the penumbral outline has been calculated, not from bessel elements, but from my custom cone projection method. With this method I can analyse the dynamics of each part of the outline curve directly from the planetary ephemerides. This is how I came to think about how the outline interacts with the limiting curve and understand its connection with the eclipse maxima, but I'm sure you could use bessel elements and get the same results (the effect is real, after all). However, I find I cannot do the calculation accurately enough with besselian methods, hence the computational anomaly; I will get to this later.

I've estimated that the outline first makes contact with the limiting curve at:

time = 08:27:26.8 TT = 08:26:21.9 UT1 = 08:26:21.6 UTC

ephLongW = 23.4208, lat = -36.7853 degrees.

This should correspond with Nufer's PS0, though it is not quite the same as the time he quotes (UT1:08:26:27), something I will also comment upon later in relation to the computational anomaly.

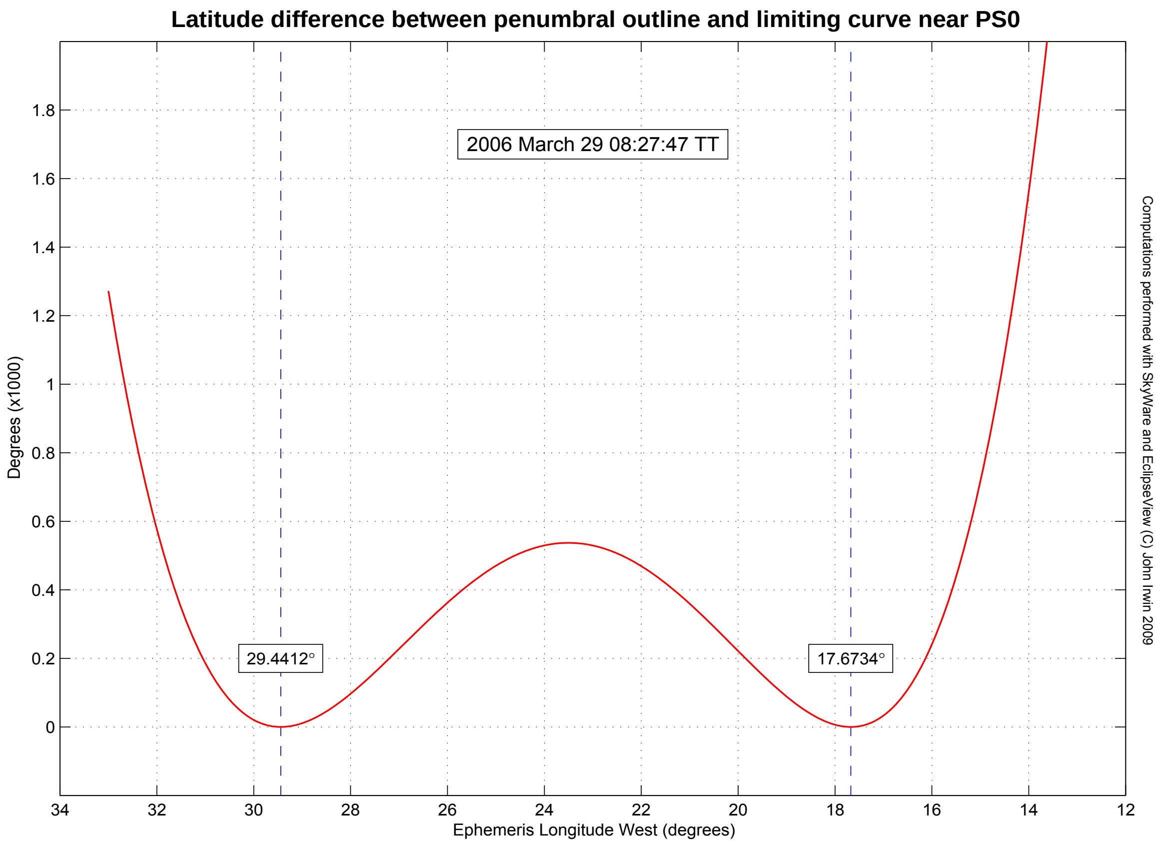
After this, contact with the limiting curve splits into two, one moving east and one moving west. Between the points the outline is dark indicating the eclipse is ending, while outside the outline is bright, showing the eclipse is beginning. Eventually, the westward moving point hits the terminator region at PS1 and disappears, leaving the single eastward moving point to make its familiar journey towards the other extreme of the limiting curve.

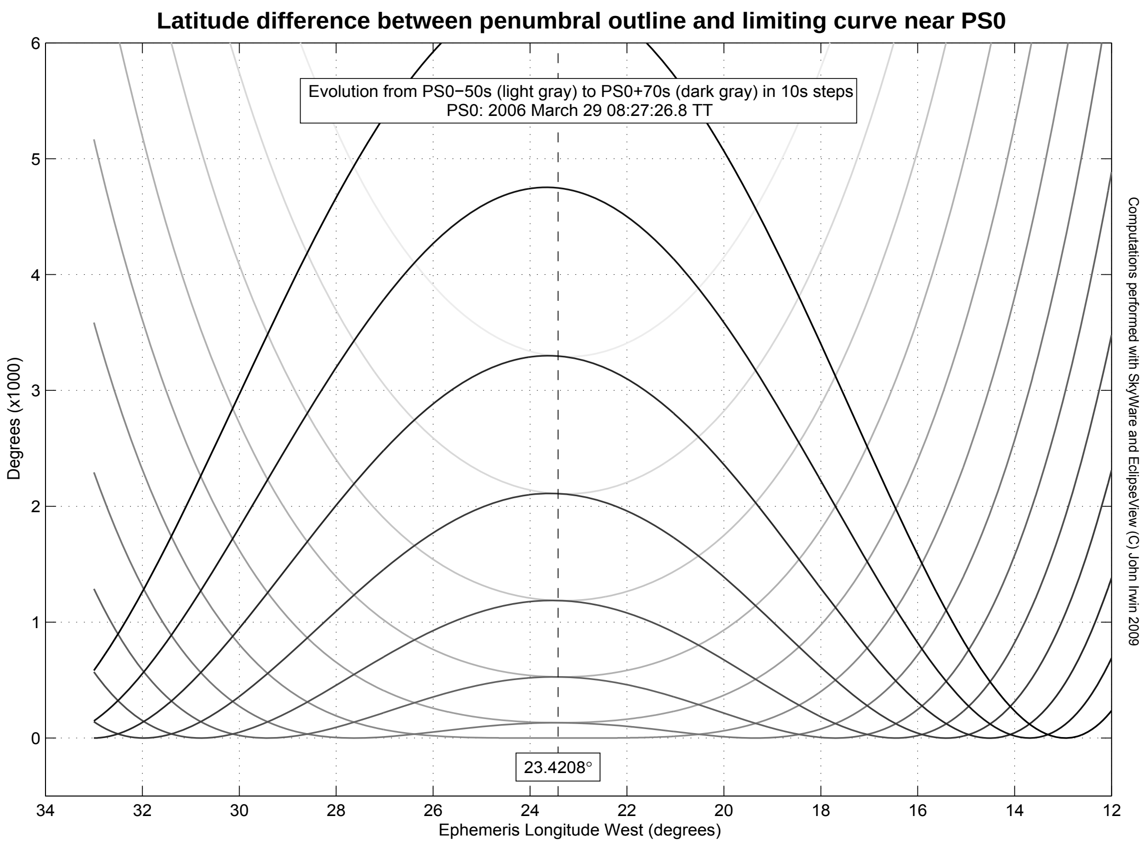
In general, we can say the outline first touches the limiting curve at the time of earliest maximum eclipse, not at the sunrise extreme point; and last touches it at the time of latest maximum eclipse, not the sunset extreme. In fact, we could define new contact points to add to P1, P2, etc. as part of the general circumstances of the eclipse. Following Nufer's notation, I would suggest PN0, PN1, PN2, PN3 for the northern limiting curve of the penumbra (in time order, where PN0 may equal PN1, and PN3 may equal PN2), and PS0, PS1, PS2, PS3 for the southern limit. Knowing the time of PN0, PN3, PS0, PS3 is certainly very useful if you want to calculate the whole of the limiting curve in the time domain. How that is done is another matter, however.

You may not be convinced the outline is really touching the limiting curve at more than one point at a time. The limiting curve around PS0 is very flat, as is the outline curve, and if plotted together they will look virtually coincident (on a small scale, anyway). But, to put some numbers to the situation, I have attempted to quantify the distance between the curves for the 2006 March 29 eclipse.

To do this I have calculated the difference in geodetic latitude between the curves, extending in longitude from just inside PS1 to a little beyond PS0. This should be adequate for this eclipse as the curves do not vary much in latitude around PS0/PS1, but that's not strictly required as the distance, however it is measured, should reduce to zero at the contact points. Again, I have used my cone projection method to generate the data for both the outline and limiting curves.

I have done this calculation for TT:08:27:47 when there are two simultaneous maximum eclipses, and is between the times of PS0 and PS1. The resulting curve is shown at the top of the next page. As you can see there are two minima at zero latitude difference, proving clearly that the outline touches the limiting curve at two, and only two, locations at this time. Also note the scale of this effect; the height of the central peak is only 0.00055 degrees.


Latitude difference between penumbral outline and limiting curve near PS0
2006 March 29 08:27:47 TT
29.4412°
17.6734°
Degrees (x1000)
Ephemeris Longitude West (degrees)
Computations performed with SkyWare and EclipseView (C) John Irwin 2009


Latitude difference between penumbral outline and limiting curve near PS0
Evolution from PS0−50s (light gray) to PS0+70s (dark gray) in 10s steps
PS0: 2006 March 29 08:27:26.8 TT
23.4208°
Degrees (x1000)
Ephemeris Longitude West (degrees)
Computations performed with SkyWare and EclipseView (C) John Irwin 2009

I have also generated an additional plot, included on the previous page, which attempts to show how the latitude difference evolves in time. Curves are shown at 10 second intervals from 50 seconds before PS0 to 70 seconds after. The longitude of PS0 is indicated by the central dotted line.

I'm afraid push came to shove and I was forced to create an animation of this graphical evolution, which you can find at:

`http://www.jir1667.plus.com/simax/2006-Mar-29-double-contact-2.hav`

Two things to note about this animation. The curves all have a western longitude limit at 33.0W; this is arbitrary, but is necessary so that the outline and limiting curves have a common set of longitudes from which a latitude difference can be determined. The longitude of PS1, at about 33.2W, is just beyond the longitude limit, which brings me to the second point. The time of PS1, at TT:08:28:18.8, is close to PS0+52s, at which time the westward moving contact point disappears into PS1, as we have seen from the first animation.

It is worth noting that if simultaneous maxima occur on one of the penumbral limiting curves, then two maxima will also occur on the umbral limiting curve on the same side of the central track (if it exists), though on a very much smaller scale. For the 2006 March 29 eclipse, the contact between the umbral outline and its southern limiting curve mimics that of the penumbra, but it lasts only 12.8 milliseconds. Of course, this is to be expected, based on Nufer’s curves of earliest/latest maximum eclipse. These curves seem to start/end on the points C1 or C2 (first and last contacts of the shadow axis with the Earth), so they must always intersect one of the umbral limiting curves. Although, whether they start/end exactly at C1 or C2 is uncertain.

Computational anomaly

I originally attempted to do the latitude difference calculation using bessel elements, in particular using Meeus's longitude method for calculating points on the limiting curve [3]. For a range of values of the position angle (Q) covering the area of interest around PS1/PS0, I calculated the longitude and latitude of points on the outline curve, and then for each point's longitude find the corresponding latitude on the limiting curve, from which the latitude difference follows.

However, I found that the limiting curve was being calculated slightly too high in latitude, resulting in a latitude difference which sometimes goes negative, although the overall shape of the difference curve still shows two minima as before. I am confident that the calculation of points on the outline curve is accurate (by comparison with the cone projection method), so the problem must lie in the limiting curve calculation. The extent of this negative offset is, however, not insignificant compared to the size of the latitude difference between the minima, so it certainly needed further investigation.

I am now beginning to understand why there is a problem. It seems that the algorithm does not consider variations of the radii of the shadow cones at the observer (L1' and L2', using Meeus's notation). So the longitude method is definitely not exact, it's an approximation because it assumes maximum eclipse occurs when the distance of the observer from the shadow axis is at a minimum. This is not strictly correct. Remember what we mean by "maximum eclipse": It means maximum magnitude. But the magnitude also depends on the radii of the shadow cones at the observer, as well as its distance from the shadow axis *m*, as in (L1' - *m*)/(L1' + L2').

The main effect of this neglect is to shift the time of maximum eclipse on the limiting curve by several seconds (if not more), compared to that I get from my cone projection method (which I trust to be accurate). This effect has actually been reported by Gossner, way back in 1955 [4], who provided a correction to the time of maximum eclipse which takes L1' and L2' variation into account. When I apply this correction to points along the limiting curve, I can match the times to a few milliseconds. So I think I am moving in the right direction.

As an interesting aside, it is worth asking what sort of maximum eclipse is assumed in the calculation of the limiting curve? Ostensibly, the limiting curve is defined by points where the eclipse is beginning and ending simultaneously, and we then assume that this is some sort of maximum eclipse. But if you look at the mathematics carefully [5], you will see that it is only a maximum in the sense that (L1' - *m*) is a maximum. This appears to be more accurate than assuming maximum occurs only when *m* is a minimum, because it considers variations of L1', but it appears that the difference between considering variations of L1' only, and both L1' and L2', is very small. So the limiting curve calculated using the standard time method (via the formula for tanQ) will produce accurate results. Note that Gossner's time correction is derived by varying both L1' and L2', so it will not be strictly consistent with the maxima along the limiting curve.

Moving on, it is here that Gossner's correction also explains why there is a difference between my estimate of the time of PS0 (from cone projection) and Nufer's estimate which I believe is obtained from the calculation of local circumstances based again on the assumption that maximum eclipse occurs when an observer is at a minimum distance from the shadow axis. This applies generally to the calculation of local circumstances using the "standard" method, as illustrated with the following data for the southern limiting curve of the 2006 March 29 eclipse:

TT:08:30:00.0	08:30:05.9	-5.9
TT:09:00:00.0	09:00:04.7	-4.7
TT:09:30:00.0	09:30:01.3	-1.3
TT:10:00:00.0	09:59:57.1	+2.9
TT:10:30:00.0	10:29:53.1	+6.9
TT:11:00:00.0	10:59:50.3	+9.7
TT:11:30:00.0	11:29:49.8	+10.2

For each of the times shown in the left column, I've calculated the point on the limiting curve, and then used this location to calculate the time of maximum eclipse from the local circumstances, shown in the middle column. The last column is the difference in seconds. However, it should be emphasised that the magnitude of the time discrepancy depends on the observer's distance from the shadow axis, which is why it is large on the penumbral limiting curves and small along the central track; on the central line it is exactly zero, of course.

As an example of how well Gossner's correction works, take the case for TT:11:30:00 (2006 March 29 eclipse again). At this time, my cone projection method locates the southern limit at:

ephLongW = 296.009217, lat = +17.491721 degrees.

Using this longitude in Meeus's method gives:

TT = 11:29:49.7735 (+10.2265 seconds)

lat = +17.491809 (-0.000088 degrees)

The calculation for the local circumstances at this longitude and Meeus's latitude gives a maximum eclipse at:

TT = 11:29:49.7734 (+10.2266 seconds)

So we see that the local circumstances and longitude methods are consistent, but consistently wrong!

Incorporating Gossner's time correction in Meeus's longitude method gives:

TT = 11:29:59.9949 (+0.0051 seconds)

lat = +17.491641 (+0.000080 degrees)

So the time is much improved, but the latitude is still refusing to cooperate at the level of accuracy I require (though admittedly accurate enough for most general purposes).

Therefore, it might be worth trying to make a corresponding correction to the calculation of the latitude of the limiting curve in the longitude method to see if this makes any significant difference.

The gist of Meeus's algorithm for calculating the incremental latitude correction is presented for the case of the central line (page 18), and then amended slightly for the limiting curves (page 22). I have checked the accuracy of the algorithm for the central line and found it to be very high, both in time and latitude, so there may be something not quite right about the amendment made for incorporating the shadow cone radius. However, it is not clear to me how this method works, in particular what the quantities "W" and "Q" represent and how the expressions for them are derived.

Until I understand that I am unable to make progress on solving this problem.

One other thing that has always puzzled me about this algorithm is why the correction +W/Q applies to the geodetic latitude, rather than the geocentric latitude. The reason for this is that I have noticed in the expression for Q, the factor multiplying "b" is just the partial of xi with respect to the geocentric latitude (with a change of sign), and the factor multiplying "a" is the partial of eta with respect to the geocentric latitude. So I would have thought that the resulting correction W/Q applies to the geocentric latitude. But, of course, in practice the converged latitude is clearly geodetic. It certainly works for the central line, and I have even assumed it is geocentric, converted the converged latitude to geodetic at the end, and found the results to be exactly the same. Very odd, but I'm sure when I finally understand what's going on it will all become clear.

Unfortunately, the originator of the algorithm is not in a position to elucidate the longitude method, so I am opening this up to anyone else who may be able to point me in the right direction. Please get in touch if you think you can help. Thank you.

John Irwin

24 August 2009

References

[1] Meeus J, Mathematical Astronomy Morsels IV, pp 87-93 (2007)

[2] Nufer R, Some Remarks on Solar Eclipses: The Curves "Earliest maximum eclipse" and "Latest maxiumum eclipse" (2007) (http://robertnufer.ch/99_diverses/Earliest_and_Latest_Maximum_Eclipse.pdf)

[3] Meeus J, Elements of Solar Eclipses 1951-2200 (1989)

[4] Gossner S D, "A Correction to the Time of Maximum Obscuration in Solar Eclipse," *Astronomical Journal*, **60**, pp 383-384 (November 1955)

[5] Explanatory Supplement to The Astronomical Ephemeris and The American Ephemeris And Nautical Almanac (1961), Third Impression (1974), p222