MATH NOTES - DERIVATION OF THE FREE SHADOW METHOD (FSM)

There are several reasons to assume a conic section cross section for simple craters:

1. The currently prevalent paradigm that simple craters are parabolic (e. g. textbooks by de Pater and Lissauer, 2001; Melosh, 2011).
2. The results of Chappelow and Sharpton (2002), which showed that many simple craters can be reasonably approximated as either parabola-like or cone-like, and that the shapes of these two special cases can be analytically related to their cast shadows.
3. In real craters, these shadows vary considerably in shape, indicating that the crater shapes also vary significantly (Fig. 1). In fact they imply that real simple crater shapes occupy a continuum that includes cone-like, hyperbolic, parabolic, and elliptical shapes.
4. Personal observations of the ‘shadowfronts’ within natural simple craters, which show that these shadows generally consist of smooth, circular or elliptical arcs (Fig. 1), consistent with the results of Chappelow and Sharpton (2002).
5. Profiles of experimental (e. g. Oberbeck, 1971) and real (e. g. Ravine and Grieve, 1986) craters which appear, qualitatively, to have generally conic section shaped cross sections, though not necessarily parabolic ones (Fig. 2).
6. Low velocity impact experiments by DeVet and DeBruyn (2007) that, while well outside the “hypervelocity” impact regime, did result in craters with conic section shaped cross sections between the conical and parabolic limiting cases (i.e. they were hyperbolas).

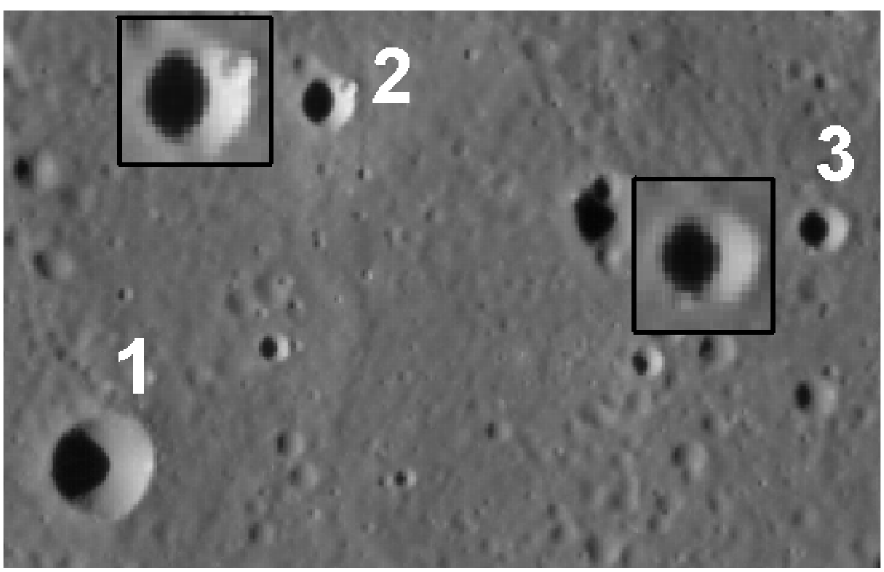

Figure 1: A clip of a MESSENGER image of Mercury that includes several simple impact craters. All are illuminated at the same solar elevation angle (16.2°), but the variation in their shadow shapes indicates that they have quite different shapes. The highly elliptical shadowfront in crater 1 demonstrates, by inspection, that it is hyperbolic, but nearly conical, in shape. The almost circular shadowfront in crater 2 shows that it is very close to parabolic (inset provided). Crater 3's shadow indicates that it too is hyperbolic, but is an intermediate case between craters 1 and 3. (Though difficult to see in this image, crater 2 also has a bright halo of ejecta similar to Linne, and thus is a very fresh crater).

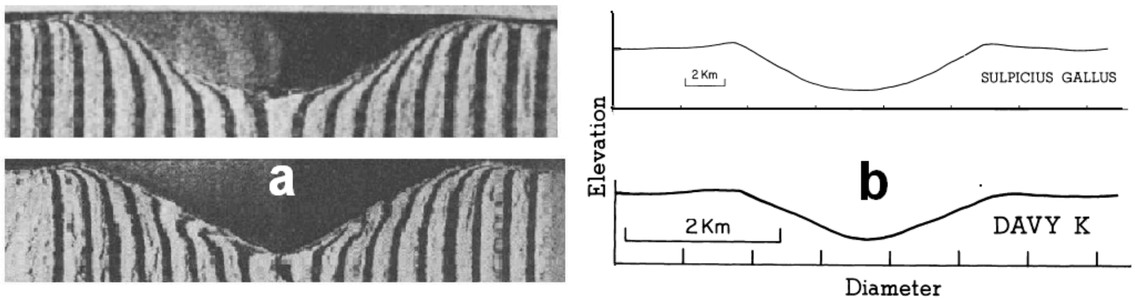

Figure 2: Examples of conic-section-like crater shapes taken from previous (a) experimental and (b) observational work. The upper profiles are parabola-like while the lower pair are cone-like. (a. stolen from Oberbeck, 1971, and b. swiped from Ravine and Grieve, 1986).

So, starting with the assumption that simple impact crater shapes can be approximated by conic sections....

Derivation of the shadow shape vs. crater shape relationship

The first thing to do is derive some kind of math relationship between shadow shape and crater shape, in terms of some parameters that describe each...

Consider a conic-section-shaped simple impact crater with radial symmetry about the vertical (z) axis, and illuminated by the Sun from above the positive x axis by solar elevation angle ϕ (Fig. 3). The illumination source is considered distant, so its rays may be considered parallel, and the crater rim is a flat horizontal circle. The shape of the shadow cast within it is determined only by the shape of the crater's interior, and not by any rim relief, so there must be a direct relationship between the crater's shape, and the shape of the shadow cast within it. The origin of coordinates is placed at the bottom of the crater, and the y-axis forms a right handed coordinate system with x and z.

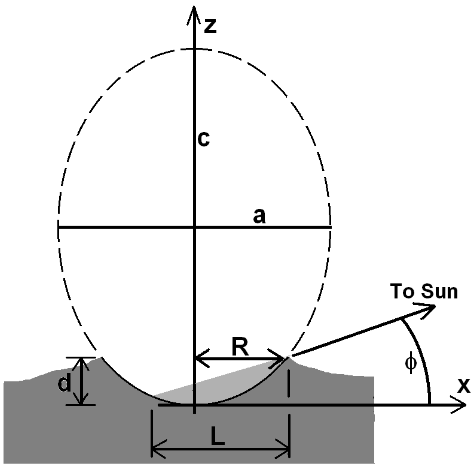

Figure 3: A vertical cross-section, along the illumination (x) axis, through the center of a crater with an elliptical cross-sectional shape, illuminated at solar elevation angle ϕ above the x-axis, which casts shadow of length L. Its depth is d, radius is R, and a and c are the semi-axes of the crater shape ellipse.

Relating Crater Shape to Shadow Shape-

For simplicity, in what follows I will use the case of an ellipsoidal crater shape (Fig. 3) as an example, since the geometric meanings of various quantities are clearer for this case; e. g. this is equivalent to restricting a and c to be real, positive quantities with simple meanings (axes of the ellipse). However the algebra below is completely general; it applies to all conic sections of revolution about the z-axis which may reasonably represent a simple crater, which in turn correspond to other possible values for a and c (e. g. negative, imaginary, and infinite values).

So, I will begin by considering a crater whose cross section has the form of an ellipse. In z-cylindrical coordinates, the equation for the ellipsoid representing the crater's interior surface is:

$$\frac{r^2}{a^2} + \frac{(z-c)^2}{c^2} = 1 \qquad [1]$$

where a and c are the radial and vertical semi-axes of the ellipsoid, respectively, which is centered at $(r, z) = (0, c)$, and with the origin of coordinates located at its bottom (Fig. 3). Because it is an ellipsoid, a^2 and c^2 must be positive numbers, and therefore a and c must be real. Furthermore, since only the lower half of the ellipsoid is of interest, c must be positive. (However, note again that equation [1] and what follows is fully general; other conic sections are simply represented by different forms of a and c.) Only the part of the ellipsoid for which $z \le d$, where d is the crater depth, represents the interior surface of the impact crater (Fig. 3). Thus a, c, and d are three parameters which completely describe the crater's approximating conic-sectional shape - a and c describe the shape of the ellipse, and d tells us what part of the ellipse is relevant.

A final parameter describing the crater in 3D is the radius, R. However R is not an independent parameter since a, c, and d together fix R. To see why, simply evaluate [1] on the crater rim, where $z = d$, and $r = R$. This yields the auxiliary equation:

$$\frac{R^2}{a^2} + \frac{(d-c)^2}{c^2} = 1 \qquad [2]$$

which will be useful later. Thus only three parameters - in this case *a*, *c*, and *d* - are sufficient to fully define the crater shape conicoid of revolution.

Next, "solving" [1] for *z* results in:

$$z = c \pm c\sqrt{1 - \frac{r^2}{a^2}},$$

Since *z*(*r*) represents a crater, we are interested here only in the lower half of the ellipsoid, which corresponds to the minus sign above, thus:

$$z(r) = c\left(1 - \sqrt{1 - \frac{r^2}{a^2}}\right) \qquad [3]$$

which relates the elevation of any point on the crater's interior surface to its radial position. Note that, by definition, z must be *positive*.

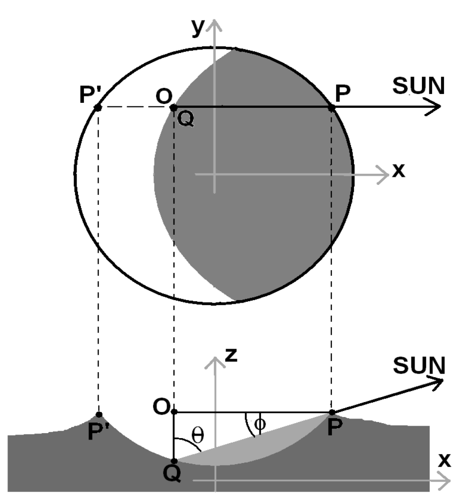

Figure 4: A vertical cross section PP' through a crater, parallel to the illumination direction. Plan view (top) and side view (bottom) of cross-section PP'. θ is the solar incidence angle relative to the vertical. Note that the cross section is NOT through the center of the crater, but is arbitrary.

Next, consider a solar light-ray which passes just over the crater rim at point *P* and is incident on the crater interior at point *Q*, which must by definition lie on the shadow-front (Fig. 4; note that points *O*, *P* and *Q* are *not* on the shadow symmetry axis, but in the plane defined by PP'). With point *O*, points *P* and *Q* form a vertical right triangle *OPQ*, and by definition of the tangent function:

$$\tan\theta = \frac{x_P - x_O}{z_O - z_Q}.$$

But $z_O = z_P = d$ and $x_O = x_Q$, and since P is on the crater rim (which is a circle), $x_P = +\sqrt{R^2 - y_P^2} = +\sqrt{R^2 - y_Q^2}$ (note that x_P must be positive for any point P that casts a shadow inside the crater). Substituting these for z_O, x_O, and x_P in the above equation:

$$\tan\theta = \frac{\left(R^2 - y_Q^2\right)^{1/2} - x_Q}{d - z_Q}.$$

Next, dropping the Q subscript (which denotes the point on the shadow boundary) and rearranging this to solve for z:

$$z = d - \frac{\left(R^2 - y^2\right)^{1/2} - x}{\tan\theta} \qquad [4].$$

This equation gives the elevation of any point *on the shadow-front* as a function of position, in rectangular coordinates. Since these points must *also* lie on the surface of the crater, we can equate the expressions for z given by [3] and by [4]:

$$d - \frac{\left(R^2 - y^2\right)^{1/2} - x}{\tan\theta} = c\left(1 - \sqrt{1 - \frac{x^2 + y^2}{a^2}}\right) \qquad [5]$$

where everything is now expressed in rectangular coordinates. This is a *mess* but it does have analytical solutions (as below).

Together with [2], this transcendental equation now relates the x and y coordinates of each point on the curve of the shadow-front cast inside an ellipsoidal crater, defined by a, c, and d, and illuminated at incidence angle θ from the +x direction. Its solutions, $y(x)$, define the boundaries of the shadow inside the crater as viewed from the zenith. Doing some rearranging of [5]:

Divide through by c:

$$\frac{d}{c} - \frac{\left(R^2 - y^2\right)^{1/2} - x}{c\tan\theta} = 1 - \sqrt{1 - \frac{x^2 + y^2}{a^2}}$$

move the 1 over to the left:

$$\frac{d}{c} - \frac{\left(R^2 - y^2\right)^{1/2} - x}{c\tan\theta} - 1 = -\sqrt{1 - \frac{x^2 + y^2}{a^2}}$$

and put the left side over common denominators:

$$\frac{d\tan\theta}{c\tan\theta}-\frac{\left(R^2-y^2\right)^{1/2}-x}{c\tan\theta}-\frac{c\tan\theta}{c\tan\theta}=-\sqrt{1-\frac{x^2+y^2}{a^2}}\ .$$

Combining terms on the left:

$$\frac{(d-c)\tan\theta-\left(R^2-y^2\right)^{1/2}+x}{c\tan\theta}=-\sqrt{1-\frac{x^2+y^2}{a^2}}$$

Next, squaring both sides:

$$\left(\frac{(d-c)\tan\theta-\left(R^2-y^2\right)^{1/2}+x}{c\tan\theta}\right)^2 = 1-\frac{x^2+y^2}{a^2}$$

and pulling a fast one by introducing mutually cancelling + and – R^2/a^2 terms on the right side:

$$\left(\frac{(d-c)\tan\theta-\left(R^2-y^2\right)^{1/2}+x}{c\tan\theta}\right)^2=1-\frac{R^2}{a^2}+\frac{R^2}{a^2}-\frac{x^2}{a^2}-\frac{y^2}{a^2}=1-\frac{x^2+R^2}{a^2}+\frac{R^2-y^2}{a^2}$$

Now let: $p(y) = (R^2–y^2)^{1/2}$. Then:

$$\left(\frac{(d-c)\tan\theta-p(y)+x}{c\tan\theta}\right)^2=1-\frac{x^2+R^2}{a^2}+\frac{p^2(y)}{a^2} \qquad [6]$$

This equation is quadratic in $p(y)$. Hmmm...I know what to do with quadratic equations! Next, for convenience, let:

$$K1=1-\frac{x^2+R^2}{a^2} \qquad K3=\frac{1}{c\tan\theta}$$

$$K2=\frac{(d-c)\tan\theta+x}{c\tan\theta} \qquad K4=\frac{1}{a^2}$$

All of these *K*-terms contain NO *y* parts. Substituting these, [6] becomes:

$$\left(K2-K3p(y)\right)^2=K1+K4p^2(y)$$

Squaring out the left side:

$$K2^2 - 2K2K3p(y) + K3^2 p^2(y) = K1 + K4p^2(y)$$

and collecting it into a more familiar quadratic form:

$$\left(K3^2 - K4\right)p^2(y) - 2K2K3p(y) + \left(K2^2 - K1\right) = 0$$

which as mentioned is quadratic in $p(y)$. In terms of the quadratic formula, the solutions for $p(y)$ are:

$$p(y) = -\frac{B}{2A} \pm \frac{\sqrt{B^2 - 4AC}}{2A}$$ where A, B and C are:

$A = K3^2 - K4$ $\quad\quad$ $B = -K2K3$ $\quad\quad$ $C = K2^2 - K1$

The leading term on the right ($-B/2A$) is: $\dfrac{2K2K3}{2\left(K3^2 - K4\right)}$ which equals:

$$\frac{\cancel{2}\dfrac{(d-c)\tan\theta + x}{c^2\tan^2\theta}}{\cancel{2}\left(\dfrac{1}{c^2\tan^2\theta} - \dfrac{1}{a^2}\right)} = \frac{\dfrac{(d-c)\tan\theta + x}{\cancel{c^2\tan^2\theta}}}{\dfrac{a^2 - c^2\tan^2\theta}{a^2\cancel{c^2\tan^2\theta}}} =$$

So: $$\boxed{-B/2A = a^2\frac{(d-c)\tan\theta + x}{a^2 - c^2\tan^2\theta}}$$ [7]

Next, the $B^2 - 4AC$ thingy under the square root is :

$$4K2^2K3^2 - 4\left(K3^2 - K4\right)\left(K2^2 - K1\right)$$

Expanding the product:

$$\cancel{4K2^2K3^2} - 4\left(\cancel{K2^2K3^2} - K1K3^2 - K2^2K4 + K1K4\right) =$$

$$4\left(K1K3^2 + K2^2K4 - K1K4\right) = 4\left(K1\left(K3^2 - K4\right) + K2^2K4\right)$$

Putting in the definitions of the Ks:

$$= 4\left[\underbrace{\left(1 - \frac{x^2 + R^2}{a^2}\right)\left(\frac{1}{c^2\tan^2\theta} - \frac{1}{a^2}\right) + \frac{1}{a^2}\left(\frac{(d-c)\tan\theta + x}{c\tan\theta}\right)^2}\right]$$

Leaving off the leading '4' for now, and strategically, partially expanding both terms:

$$\left(1-\frac{R^2}{a^2}\right)\left(\frac{1}{c^2\tan^2\theta}-\frac{1}{a^2}\right)-\frac{x^2}{a^2}\left(\frac{1}{c^2\tan^2\theta}-\frac{1}{a^2}\right)+\frac{1}{a^2c^2\tan^2\theta}\left((d-c)^2\tan^2\theta+2x(d-c)\tan\theta+x^2\right)$$

Now notice that the $-\frac{x^2}{a^2}\left(\frac{1}{c^2\tan^2\theta}\right)$ and the $+\frac{1}{a^2c^2\tan^2\theta}\left(x^2\right)$ terms cancel out (yay):

$$\left(1-\frac{R^2}{a^2}\right)\left(\frac{1}{c^2\tan^2\theta}-\frac{1}{a^2}\right)-\frac{x^2}{a^2}\left(\cancel{\frac{1}{c^2\tan^2\theta}}-\frac{1}{a^2}\right)+\frac{1}{a^2c^2\tan^2\theta}\left((d-c)^2\tan^2\theta+2x(d-c)\tan\theta+\cancel{x^2}\right)$$

leaving:

$$\left(1-\frac{R^2}{a^2}\right)\left(\frac{1}{c^2\tan^2\theta}-\frac{1}{a^2}\right)-\frac{x^2}{a^2}\left(-\frac{1}{a^2}\right)+\frac{1}{a^2c^2\tan^2\theta}\left((d-c)^2\tan^2\theta+2x(d-c)\tan\theta\right)$$

or, rearranging the last term:

$$\left(1-\frac{R^2}{a^2}\right)\left(\frac{1}{c^2\tan^2\theta}-\frac{1}{a^2}\right)+\frac{x^2}{a^4}+\frac{(d-c)^2}{a^2c^2}+\frac{2x(d-c)}{a^2c^2\tan\theta}$$

Now recall Eq. [2], and note that it can be arranged to read $1-\frac{R^2}{a^2}=\frac{(d-c)^2}{c^2}$. Putting this into the above equation:

$$\left(\frac{(d-c)^2}{c^2}\right)\left(\frac{1}{c^2\tan^2\theta}-\frac{1}{a^2}\right)+\frac{x^2}{a^4}+\frac{(d-c)^2}{a^2c^2}+\frac{2x(d-c)}{a^2c^2\tan\theta} \qquad \text{or:}$$

$$\left(\frac{(d-c)}{c^2\tan\theta}\right)^2-\cancel{\frac{(d-c)^2}{a^2c^2}}+\frac{x^2}{a^4}+\cancel{\frac{(d-c)^2}{a^2c^2}}+\frac{2x(d-c)}{a^2c^2\tan\theta}$$

and terms 2 and 4 cancel, leaving:

$$4\left[\left(\frac{(d-c)}{c^2\tan\theta}\right)^2+\frac{2x(d-c)}{a^2c^2\tan\theta}+\frac{x^2}{a^4}\right]$$

where the leading '4' has been restored. Now remember that this stuff equals the whole B^2-4AC thingy, and notice that it is a perfect square!!! Thus, taking the square root:

$$\pm\sqrt{B^2-4AC}=\pm 2\left[\frac{d-c}{c^2\tan\theta}+\frac{x}{a^2}\right]$$

Now:

$$A=\frac{1}{c^2\tan^2\theta}-\frac{1}{a^2}=\frac{a^2-c^2\tan^2\theta}{a^2c^2\tan^2\theta} \quad \text{so:}\quad \frac{1}{2A}=\frac{a^2c^2\tan^2\theta}{2\left(a^2-c^2\tan^2\theta\right)}$$

and:

$$\pm\frac{\sqrt{B^2-4AC}}{2A}=\pm 2\left[\frac{d-c}{c^2\tan\theta}+\frac{x}{a^2}\right]\left[\frac{a^2c^2\tan^2\theta}{2\left(a^2-c^2\tan^2\theta\right)}\right]$$

Putting the first term on the right over common denominators:

$$=\pm 2\left[\frac{(d-c)a^2\tan\theta}{a^2c^2\tan^2\theta}+\frac{xc^2\tan^2\theta}{a^2c^2\tan^2\theta}\right]\left[\frac{a^2c^2\tan^2\theta}{2\left(a^2-c^2\tan^2\theta\right)}\right]$$

and cancelling some terms:

$$=\pm \cancel{2}\left[\frac{(d-c)a^2\tan\theta}{\cancel{a^2c^2\tan^2\theta}}\cancel{+}\frac{xc^2\tan^2\theta}{\cancel{a^2c^2\tan^2\theta}}\right]\left[\frac{\cancel{a^2c^2\tan^2\theta}}{\cancel{2}\left(a^2-c^2\tan^2\theta\right)}\right]$$

(Wow, that was almost as good as having a good sneeze!)

$$\pm\frac{\sqrt{B^2-4AC}}{2A}=\pm\frac{\left(c^2\tan^2\theta\right)x+\left((d-c)a^2\tan\theta\right)}{a^2-c^2\tan^2\theta} \qquad [8]$$

FINALLY putting [7] and [8] together to form the solutions, $p(y)$:

$$p(y)=-\frac{B}{2A}\pm\frac{\sqrt{B^2-4AC}}{2A}=\frac{a^2(d-c)\tan\theta+a^2x}{a^2-c^2\tan^2\theta}\pm\frac{\left(c^2\tan^2\theta\right)x+\left((d-c)a^2\tan\theta\right)}{a^2-c^2\tan^2\theta}$$

and recalling the definition of $p(y)$:

$$+\sqrt{R^2-y^2}=\frac{a^2(d-c)\tan\theta+a^2x}{a^2-c^2\tan^2\theta}\pm\frac{\left(c^2\tan^2\theta\right)x+\left((d-c)a^2\tan\theta\right)}{a^2-c^2\tan^2\theta} \qquad [9]$$

which constitutes TWO solutions for $y(x)$. Lets start with the minus sign solution:

$$+\sqrt{R^2-y^2} = \frac{\cancel{a^2(d-c)\tan\theta} + a^2 x}{a^2-c^2\tan^2\theta} - \frac{(c^2\tan^2\theta)x + \cancel{((d-c)a^2\tan\theta)}}{a^2-c^2\tan^2\theta}$$

$$+\sqrt{R^2-y^2} = \frac{a^2 x}{a^2-c^2\tan^2\theta} - \frac{(c^2\tan^2\theta)x}{a^2-c^2\tan^2\theta} = \cancel{\frac{a^2-c^2\tan^2\theta}{a^2-c^2\tan^2\theta}}x$$

$+\sqrt{R^2-y^2} = x$ or:

$$\boxed{x^2+y^2=R^2}$$

This solution coincides with the crater rim circle and represents the segment of the rim that forms the *sunward* boundary of the shadow.

The much more interesting solution is the one that comes from the PLUS sign:

$$+\sqrt{R^2-y^2} = \frac{a^2(d-c)\tan\theta + a^2 x}{a^2-c^2\tan^2\theta} + \frac{(c^2\tan^2\theta)x + ((d-c)a^2\tan\theta)}{a^2-c^2\tan^2\theta}$$

$$+\sqrt{R^2-y^2} = \left(\frac{a^2+c^2\tan^2\theta}{a^2-c^2\tan^2\theta}\right)x + \left(\frac{2a^2(d-c)\tan\theta}{a^2-c^2\tan^2\theta}\right)$$

This can be rewritten:

$$+\sqrt{R^2-y^2} = \left[\frac{x+\dfrac{2a^2(d-c)\tan\theta}{a^2+c^2\tan^2\theta}}{\dfrac{a^2-c^2\tan^2\theta}{a^2+c^2\tan^2\theta}}\right]$$

and dividing through by *R*:

$$+\frac{1}{R}\sqrt{R^2-y^2} = \left[\frac{x+\dfrac{2a^2(d-c)\tan\theta}{a^2+c^2\tan^2\theta}}{\left(\dfrac{a^2-c^2\tan^2\theta}{a^2+c^2\tan^2\theta}\right)R}\right] \qquad [11]$$

When I now define:

$$\alpha = \pm\left(\frac{a^2 - c^2 \tan^2\theta}{a^2 + c^2 \tan^2\theta}\right)R \quad [12]$$

and:

$$x_C = \frac{-2a^2(d-c)\tan\theta}{a^2 + c^2 \tan^2\theta} \quad [13]$$

Eq. [11] becomes:

$$+\sqrt{1 - y^2/R^2} = \left(\frac{x - x_C}{\alpha}\right)$$

and squaring this:

$$1 - \frac{y^2}{R^2} = \left(\frac{x - x_C}{\alpha}\right)^2$$

or:

$$\left(\frac{x - x_C}{\alpha}\right)^2 + \frac{y^2}{R^2} = 1 \quad [14]$$

This second solution is the equation of an ellipse in the x-y plane, centered at $x = x_C$ and $y = 0$, and with semi-axes α in the x (illumination) direction and R in the y (transverse) direction (Fig. 5). ***Therefore, as viewed from the zenith, the shadowfront in a conic section shaped crater must itself be a segment of an ellipse, oriented with axes parallel to and transverse to the illumination direction, and with the transverse semi-axis equal to the radius of the crater.***

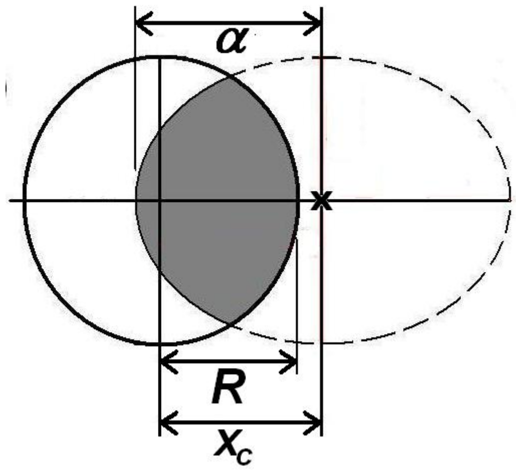

Figure 5: Definitions of the terms that define the crater shadow bounding ellipse. Illumination is from the right, the crater's rim is the solid black circle and it is hyperbolic in cross sectional shape (as will be shown below). The **X** marks the center of the shadow boundary ellipse and is displaced by distance x_c from the origin (center of crater) in the direction of the Sun.

This concludes the first stage of these notes. Eqs. [12] and [13] are the key results so far. They define the shape of the shadow that must be cast within a conic section shaped crater under the basic assumptions laid out above. Fig. 6 summarizes the general craterforms and their corresponding shadows. Next, the hard part....inverting [12] and [13]........

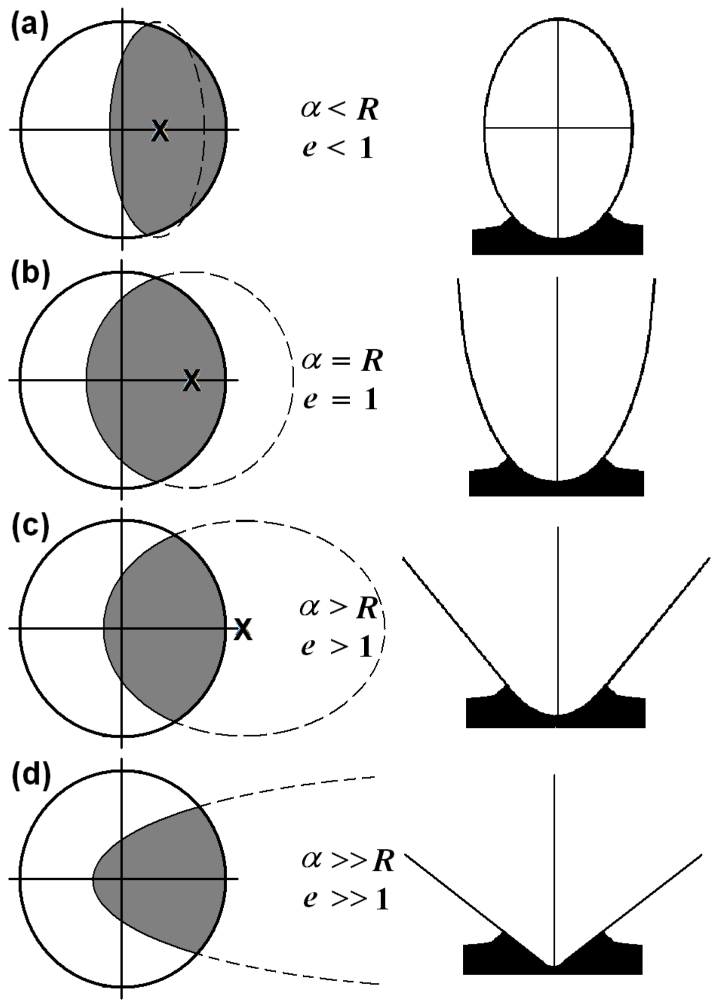

Figure 6: Crater shadows (left) and corresponding crater shapes. (a) Elliptical, (b) parabolic, (c) hyperbolic, and (d) conical. The **x**s mark the centers of the shadow boundary ellipses.

MATH NOTES - DERIVATION OF THE FREE SHADOW METHOD (FSM)

Part 2 - Inverting the equations:

So far I have established:

$$\alpha = \pm\left(\frac{a^2 - c^2\tan^2\theta}{a^2 + c^2\tan^2\theta}\right)R \qquad [12]$$

$$x_c = -\frac{2a^2(d-c)\tan\theta}{a^2 + c^2\tan^2\theta} \qquad [13]$$

and the conditions, that (1) the shadow boundary ellipse must be oriented such that one (longitudinal) axis is coincident with the illumination axis, and (2) that the transverse semi-axis of the shadow boundary ellipse must be equal to the crater radius, *R*. Together, [12], [13], and these conditions fix the position, shape, and orientation of the shadow boundary ellipse, given a fixed crater shape.

Before I go further, some observations about these results:

1. The case $\alpha = R$ corresponds to a *circular* shadowfront, and a parabolic crater shape (see Chappelow&Sharpton, 2002). This special case represents the division between the hyperbolic crater shapes (for which $\alpha > R$; again also see Chappelow&Sharpton, 2002, the conical case), and the elliptical crater shapes (for which $\alpha < R$) for which there is no theoretical 'end member'. However, it is important to note that I have never seen a shadowed crater that doesn't fit between the conical and parabolic cases. Thus, for realistic simple craters, these two shapes do represent end members.

2. To visualize the parabolic special case, consider an upright ellipse and stretch it vertically to infinity to approach a parabola in shape (ie. let c -> ∞). Then, in Eq. [12] above, a^2 vanishes compared to c^2 as c^2 gets very large:

$$\alpha = \pm\left(\frac{\cancel{a^2} - c^2\tan^2\theta}{\cancel{a^2} + c^2\tan^2\theta}\right)R \cong \frac{-c^2\tan^2\theta}{c^2\tan^2\theta}R = \mp R$$ which means that the shadowfront is a

circular arc. Yup, that checks out...

3. The special case where $\alpha = 0$ corresponds to a shadowfront that is a straight line and marks the bound between convex shadowfronts (for which the shadow is contained within the shadowfront ellipse), and concave shadowfronts (for which the shadow is NOT inside the shadow ellipse). For convex shadowfronts α is positive, and it is negative when the shadow is concave. The concave shadowfront would also correspond to a very shallow, nearly flat-bottomed crater. I haven't observed a single *real*, simple crater that even approaches this case, however...

4. For α to be zero, we must have $a^2 - c^2\tan^2\theta = 0$, or $\tan\theta = \pm a/c$. I don't know if or what physical meaning this has beyond just a curiosity...

Inverting the Transcendental Equations for *a*, *c* and *d*

At this point we have equations for the shadow shape (α and x_c) in terms of the approximating conic section-oid (*a*, *c* and *d*) and the geometric parameters *R* and $\tan\theta$.

$$\alpha = \pm\left(\frac{a^2 - c^2 \tan^2\theta}{a^2 + c^2 \tan^2\theta}\right)R \qquad [12]$$

$$x_c = -\frac{2a^2(d-c)\tan\theta}{a^2 + c^2 \tan^2\theta} \qquad [13]$$

But we <u>need</u> the crater shape *a*, *c* and *d*, in terms of the measurable quantities α, x_c and *R* (and $\tan\theta$), since they are the unknowns; α and x_c are obtained by measuring the shadow, *R* by measuring the crater rim, and $\tan\theta$ is just the sun angle. Clearly we also need a 3rd equation, (since we have 3 unknowns and only 2 equations) if we hope to solve this mess. Recall that evaluating the equation for the crater's cross-sectional shape at $r = \pm R$ yields just such an (auxiliary) equation:

$$\frac{R^2}{a^2} + \frac{(d-c)^2}{c^2} = 1 \qquad [14]$$

which inter-relates the non-measurables *a*, *c* and *d*, reducing the unknowns to 2, or alternately, providing a 3rd equation.

A complication is the ± sign ambiguity in eqn. [12], but it can be dealt with. As a reminder, the two general crater shape cases, not counting parabolic and conical (which are special cases), and related stuff are:

Table 1

Elliptical (Upright) Crater	Hyperbolic Crater
a = real, a^2 = positive	a = imaginary, a^2 = negative
c = positive, c^2 = positive	c = negative, c^2 = positive
α = positive and $\alpha < R$	α = positive and $\alpha > R$

Now to "solve" (ie. invert) the eqn. set [12] - [14].

The first step is to carefully rearrange [12] to isolate a^2. Since it has a "±" in it, I will do this for each sign to ensure getting the right ± and ∓ signs in the result. (This turns out to be very inportant!)

The PLUS sign:

$$\alpha = +\left(\frac{a^2 - c^2 \tan^2\theta}{a^2 + c^2 \tan^2\theta}\right)R$$

$$\left(a^2 + c^2 \tan^2\theta\right)\alpha = \left(a^2 - c^2 \tan^2\theta\right)R$$

$$a^2\alpha + c^2 \tan^2\theta \cdot \alpha = a^2\alpha - c^2 \tan^2\theta \cdot \alpha$$

$$a^2(\alpha - R) = c^2 \tan^2\theta(-\alpha - R)$$

$$a^2 = -c^2 \tan^2\theta \left(\frac{\alpha + R}{\alpha - R}\right) \quad [15+]$$

The MINUS sign:

$$\alpha = -\left(\frac{a^2 - c^2 \tan^2\theta}{a^2 + c^2 \tan^2\theta}\right) R$$

$$\left(a^2 + c^2 \tan^2\theta\right)\alpha = -\left(a^2 - c^2 \tan^2\theta\right) R$$

$$a^2\alpha + c^2 \tan^2\theta \cdot \alpha = -a^2\alpha + c^2 \tan^2\theta \cdot \alpha$$

$$a^2(\alpha + R) = c^2 \tan^2\theta(-\alpha + R)$$

$$a^2(\alpha + R) = -c^2 \tan^2\theta(\alpha - R)$$

$$a^2 = -c^2 \tan^2\theta \left(\frac{\alpha - R}{\alpha + R}\right) \quad [15-]$$

Combining [15+] and [15−] give the combined expression:

$$\boxed{a^2 = -c^2 \tan^2\theta \left(\frac{\alpha \pm R}{\alpha \mp R}\right)} \quad [15]$$

(Note that the distinction between the $\pm$ and $\mp$ is important!)

Next, rearrange [13] to isolate (*d*–*c*):

$$x_c = -\frac{2a^2(d - c)\tan\theta}{a^2 + c^2 \tan^2\theta} \quad [13]$$

And 'solving' for *d*–*c*:

$$d - c = -\frac{x_c}{2\tan\theta}\left(\frac{a^2 + c^2 \tan^2\theta}{a^2}\right)$$

or:

$$d - c = -\frac{x_c}{2\tan\theta}\left(1 + \frac{c^2 \tan^2\theta}{a^2}\right)$$

and substituting [15] into this for a^2:

$$d-c=-\frac{x_c}{2\tan\theta}\left(1+\frac{\cancel{c^2\tan^2\theta}}{-\cancel{c^2\tan^2\theta}\left(\alpha\pm R \big/ \alpha\mp R\right)}\right)$$

$$d-c=-\frac{x_c}{2\tan\theta}\left(1-\frac{\alpha\mp R}{\alpha\pm R}\right)$$

this numerator turns out to $= \pm 2R$

$$d-c=-\frac{x_c}{2\tan\theta}\left(\frac{(\alpha\pm R)-(\alpha\mp R)}{\alpha\pm R}\right)$$

this is from $-(\pm)$

$$d-c=-\frac{x_c}{2\tan\theta}\left(\frac{\pm 2R}{\alpha\pm R}\right)=\frac{\mp Rx_c}{(\alpha\pm R)\tan\theta} \qquad [16a]$$

or, squaring [16a]:

$$(d-c)^2=\frac{R^2x_c^2}{(\alpha\pm R)^2\tan^2\theta} \qquad [16b]$$

Now recall [14]:

$$\frac{R^2}{a^2}+\frac{(d-c)^2}{c^2}=1$$

and substitute [15] and [16b] into it:

$$a^2=-c^2\tan^2\theta\left(\frac{\alpha\pm R}{\alpha\mp R}\right) \qquad [15]$$

$$(d-c)^2=\frac{R^2x_c^2}{(\alpha\pm R)^2\tan^2\theta} \qquad [16b]$$

$$\frac{R^2}{-c^2\tan^2\theta\left(\frac{\alpha\pm R}{\alpha\mp R}\right)}+\frac{\left(\frac{R^2x_c^2}{(\alpha\pm R)^2\tan^2\theta}\right)}{c^2}=1$$

Multiplying through by c^2:

$$\frac{R^2}{-\tan^2\theta\left(\frac{\alpha\pm R}{\alpha\mp R}\right)} + \frac{R^2 x_c^2}{(\alpha\pm R)^2\tan^2\theta} = c^2$$

and cleaning up the algebra:

$$c^2 = \frac{R^2}{\tan^2\theta}\left(-\frac{\alpha\mp R}{\alpha\pm R} + \frac{x_c^2}{(\alpha\pm R)^2}\right)$$

this = $\alpha^2 - R^2$

Factoring out the denominator:

$$c^2 = \frac{R^2}{(\alpha\pm R)^2\tan^2\theta}\left(-(\alpha\mp R)(\alpha\pm R) + x_c^2\right)$$

$$c^2 = \left(\frac{R}{(\alpha\pm R)\tan\theta}\right)^2\left(R^2 - \alpha^2 + x_c^2\right)$$

Here note that, since c^2 MUST be a positive for both elliptical- AND hyperbolic-type craters (see Table 1) - c itself can be negative (and is, for hyperbolic type craters) but cannot be imaginary, for z-oriented conic section shapes, and since the leading term on the right is both **real** (because all of its contents are real**)** AND **squared** (and thus MUST be positive), it must be true that the quantity ($R^2 - \alpha^2 + x_c^2$) is *also* positive, for any physically realistic combination of R, α, and x_c (ie. that correspond to any real crater). Thus the quantity ($R^2 - \alpha^2 + x_c^2$) MUST be both positive and real for any "real" crater under consideration. If this term is ever negative, then either a basic assumption has been violated, or the measurement process of R, α, and x_c has been screwed up somehow.

Now, taking the square root to get c itself:

$$c = \pm\left(\frac{R}{(\alpha\pm R)\tan\theta}\sqrt{R^2 - \alpha^2 + x_c^2}\right)$$

Here the two ± signs are NOT related to each other, and thus this equation represents FOUR potential expessions for c!! However, we need to have c = positive for $\alpha < R$ and negative for $\alpha > R$, and thus a sign change in c when $\alpha = R$ (see Table 1). The sign change requirement means we must use the negative sign, and not the + between α and R in the denominator:

$$c = \pm\left(\frac{R}{(\alpha - R)\tan\theta} \sqrt{R^2 - \alpha^2 + x_c^2} \right)$$

and because we **still** need to have $c =$ *positive* for $\alpha < R$ and *negative* for $\alpha > R$ (Table 1), the remaining ± out front must *also* be a minus sign:

$$c = -\left(\frac{R}{(\alpha - R)\tan\theta} \sqrt{R^2 - \alpha^2 + x_c^2} \right)$$

or: $$\boxed{c = \left(\frac{R}{(R - \alpha)\tan\theta} \sqrt{R^2 - \alpha^2 + x_c^2} \right)} \qquad [17a]$$

and: $$\boxed{c^2 = \left(\frac{R}{(R - \alpha)\tan\theta} \right)^2 \left(R^2 - \alpha^2 + x_c^2 \right)} \qquad [17b]$$

The potential problem presented by the $R - \alpha$ term when $R = \alpha$ is discussed below.

Now, putting the correct signs in, as determined above, [15] becomes:

$$a^2 = -c^2 \tan^2\theta \left(\frac{\alpha - R}{\alpha + R} \right) \qquad [18]$$

Substituting [17b] into [18] for c^2:

$$a^2 = -\frac{R^2}{(\alpha - R)^2 \tan^2\theta} \left(R^2 - \alpha^2 + x_c^2 \right) \tan^2\theta \left(\frac{\alpha - R}{\alpha + R} \right)$$

and cancelling terms:

$$a^2 = -\frac{R^2}{(\alpha - R)^{\cancel{2}} \cancel{\tan^2\theta}} \left(R^2 - \alpha^2 + x_c^2 \right) \cancel{\tan^2\theta} \left(\frac{\cancel{\alpha - R}}{\alpha + R} \right)$$

this = $\alpha^2 - R^2$

$$a^2 = -\frac{R^2}{(\alpha - R)(\alpha + R)} \left(R^2 - \alpha^2 + x_c^2 \right)$$

$$\boxed{a^2 = \left(\frac{R^2 - \alpha^2 + x_c^2}{R^2 - \alpha^2} \right) R^2} \qquad [19]$$

Here note that, as expected (Table 1), a^2 is positive for elliptical craters (where $R > \alpha$), and negative for hyperbolic ones ($R < \alpha$), since R and $R^2 - \alpha^2 + x_c^2$ are necessarily positive quantities (see notes above in re $R^2 - \alpha^2 + x_c^2 > 0$), and therefore the $R^2 - \alpha^2$ term in the denominator rules this.

The next, and final equation we need is the one for the depth, d. To get this, we start with [16a], but with both correct signs in place:

$$d - c = \frac{Rx_c}{(\alpha - R)\tan\theta}$$

$$d = c + \frac{Rx_c}{(\alpha - R)\tan\theta} = c - \frac{Rx_c}{(R - \alpha)\tan\theta}$$

Substituting [17a] into this, for c yields:

$$d = \left(\frac{R}{(R - \alpha)\tan\theta} \sqrt{R^2 - \alpha^2 + x_c^2} \right) - \frac{Rx_c}{(R - \alpha)\tan\theta}$$

$$\boxed{d = \frac{R}{(R - \alpha)\tan\theta} \left(\sqrt{R^2 - \alpha^2 + x_c^2} - x_c \right)} \qquad [20]$$

Collecting all the results in one place:

$$\boxed{a^2 = \left(\frac{R^2 - \alpha^2 + x_c^2}{R^2 - \alpha^2} \right) R^2} \qquad [21]$$

$$\boxed{c^2 = \left(\frac{R^2 - \alpha^2 + x_c^2}{(R - \alpha)^2 \tan^2\theta} \right) R^2} \qquad [22]$$

$$\boxed{c = + \left(\frac{\sqrt{R^2 - \alpha^2 + x_c^2}}{(R - \alpha)\tan\theta} \right) R} \qquad [23]$$

$$d = \left(\frac{\sqrt{R^2 - \alpha^2 + x_c^2} - x_c}{(R - \alpha)\tan\theta} \right) R \qquad [24]$$

(Note that no expression for *a* is needed, since only a^2 and not *a* itself appears in any of this math. In fact it would be just an added complication for no gain.)

The three quantities a^2, *c*, and *d* are sufficient to fully define the crater-approximating conic section, Given measurements of *R*, α, and x_c from an image of a crater, calculation of a^2, *c*, and *d* via Eqs. [21]-[24] is sufficient to fully define the approximating conic section (Fig. 6 and Table 1 qualitatively summarize the relationships between crater shapes, crater shape parameters, and shadow shape parameters, and shadow shapes). However there are other possible choices of the crater shape parameters, and *a*, *c*, and *d* have certain drawbacks. First, they are rather non-intuitive except for the elliptical case; the crater shape cannot be easily visualized simply by inspection of these values (Table 1). Second, all three grow very large for nearly parabolic craters, and are infinite (Eqs. [21]-[24]) for exactly parabolic ones (for which $\alpha = R$), which happens to fall right in the middle of our geometry-range of interest, between hyperbolic and elliptical craters. Third, as a consequence, experience has shown that *a* and *c* can be extremely sensitive to even very small variations in the crater shape, again especially for nearly parabolic craters. This can result in large ranges of values for *a* and *c*, even for repeated measurements of the same crater.

Fortunately there is an alternative set that largely avoids all of these difficulties. The first two of these are the diameter, *D*, which is obtained by direct measurement of *R*, and the eccentricity, *e*, of the crater shape, defined as $\sqrt{1-a^2/c^2}$. Substituting [21] and [22] into this definition yields:

$$e = \sqrt{1 - \left(\frac{R-\alpha}{R+\alpha} \right) \tan^2\theta} \qquad [25].$$

This quantity is free of singularities and varies smoothly for the conic sections of interest here. It is also readily interpretable, in terms of crater shapes and their corresponding shadow shapes: $e < 1$ for an ellipse, $e = 1$ for a parabola, and $e > 1$ for a hyperbola (Fig. 6, Table 1). As the crater shape approaches the limiting, conical case, *e* approaches infinity, and on the other end it approaches unity for nearly parabolic craters.

The third and final parameter to fix the crater shape is still the depth, *d*, but when $\alpha = R$ (i.e. the crater is parabolic), Eq. [24] gives an undefined (zero divided by zero) result for the depth. In fact all of the equations give infinite results, as noted above. However, in reality, this is a rare special case and is easy to recognize and handle. When this occurs, the eccentricity is, by definition, equal to unity, and the depth can be calculated from the equation for the depths of exactly parabolic craters given by Chappelow and Sharpton (2002) using just the shadow length:

$$d = \frac{D}{4(1 - L/D)\tan\theta} = \frac{R}{(2 - L/R)\tan\theta} \qquad [26]$$

The shadow length can be measured directly from the image, or calculated from:

$$L = \alpha - (x_c - R) = \alpha + R - x_c$$

(see Fig. 5).

And this concludes the derivation of the equations needed to use the FSM, at least in theory. In reality it is quite impractical to apply "manually"; a computer program is needed to, for example, carry out the measurements of R, α, and x_c. The program fits an ellipse to a set of user-defined points on the shadowfront and a circle to similarly defined points on the rim and gets R, α, and x_c from these. But that is a different subject......