The Shape of a Newtonian Orbit

In Newtonian mechanics, the path of a satellite orbiting a massive body is a conic section. If the orbit is "closed" -- the satellite goes all the way around without escaping -- the path is an ellipse. If the satellite has just barely enough energy to escape, then it follows a parabolic path. If it has more than enough energy to escape, then its "orbit" is a hyperbola. I learned those things "by rote" a long, long time before I ever saw a proof of them; I was thrilled when I could first verify that they're really true. On this page we'll prove those things. Unfortunately, the proof is rather lengthy, with a lot of niggling algebra; deriving the Lorentz transforms, by comparison, is quite a bit simpler!

We'll be making some (very) slight use of Lagrangian mechanics, and we'll need to use a slightly clever substitution in order to integrate a differential equation, but aside from that it's mostly going to be a straightforward "equation grind".

(I'm going to refer to the thing being orbited as "the star" throughout this page.)

The Goal: Equations of motion in Cartesian coordinates

We want to end up with an equation relating y and x, which we can check against the equations for the conic sections, to see if the orbits really are what we think they should be.

Statement of the Problem

For simplicity we'll be doing most of the work in polar coordinates. We're going to start with the basics and work our way up. The velocity and kinetic energy in polar coordinates are easily seen to be:

(1)
$$\begin{aligned} \vec{\mathbf{v}} &= v_\theta\hat{\theta} + v_r\hat{r} \\ &= r\dot{\theta}\hat{\theta} + \dot{r}\hat{r} \\ T &= \tfrac{1}{2}m(v_\theta^2 + v_r^2) \\ &= \tfrac{1}{2}m(r^2\dot{\theta}^2 + \dot{r}^2) \end{aligned}$$

Potential energy is, of course,

(2) $$V = -m\tfrac{K}{r} \qquad \{\, where\ K = GM > 0 \,\}$$

and the Lagrangian is

(3) $$\mathscr{L} = \tfrac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2) + \tfrac{Km}{r}$$

Note, again, that we've taken K to be positive here. We now find the partials,

(4a) $\frac{\partial \mathscr{L}}{\partial \theta} = 0$

(4b) $\frac{\partial \mathscr{L}}{\partial \dot{\theta}} = mr^2\dot{\theta} \equiv L$

Since (4a) is *zero*, and the time derivative of (4b) is equal to (4a), (4b) must be a constant, which we've named ***L*** (it is, of course, the angular momentum). We'll use that later to eliminate θ from the equation for ***r***, and we won't explicitly calculate the time derivative of (4b).

(4c) $\frac{\partial \mathscr{L}}{\partial r} = m(r\dot{\theta}^2 - \frac{K}{r^2})$

(4d) $\frac{d}{dt}\frac{\partial \mathscr{L}}{\partial \dot{r}} = \frac{d}{dt}(m\dot{r}) = m\ddot{r}$

Putting together (4d) and (4c), and turning (4b) around, we get the equations of motion we'll need to solve:

(5a) $\ddot{r} = r\dot{\theta}^2 - \frac{K}{r^2}$

(5b) $\dot{\theta} = \frac{L}{mr^2}$

Solving the Equations in Polar Coordinates

Ideally we'd like to solve for ***r*** and θ as functions of time, but in this case that seems hard -- and in order to determine the shapes of the orbits, we don't need it. All we really need is to solve for ***r*** in terms of θ, which we can do if we can eliminate time from equations (5).

We're going to concentrate on (5a). The terms in 1/r are a little nasty, so the first thing we'll do is try to get rid of them. To do that we'll substitute ***u*** for ***r***:

(6)
$$\begin{aligned} r &= \frac{1}{u} \\ \dot{r} &= -\frac{1}{u^2}\dot{u} \\ \ddot{r} &= \frac{1}{u^3}(2\dot{u}^2 - u\ddot{u}) \end{aligned}$$

With ***r*** eliminated, and after multiplying through by u^3, we obtain:

(7) $2\dot{u}^2 - u\ddot{u} = u^2\dot{\theta}^2 - Ku^5$

This doesn't look much better, but let's move on and see where we get. We've got time derivatives of ***u*** which we'd like to get rid of. We'll expand them using the chain rule, as follows:

$$(8)\quad \begin{aligned} \dot{u} &= \frac{du}{d\theta}\dot{\theta} \\ \ddot{u} &= \frac{d^2u}{d\theta^2}\dot{\theta}^2 + \frac{du}{d\theta}\ddot{\theta} \end{aligned}$$

Substituting (8) into (7) leads to this:

$$(9)\quad 2\left(\frac{du}{d\theta}\right)^2\dot{\theta}^2 - u\left[\frac{d^2u}{d\theta^2}\dot{\theta}^2 + \frac{du}{d\theta}\ddot{\theta}\right] = u^2\dot{\theta}^2 - Ku^5$$

Now we finally want to get rid of the explicit θ terms, which we do by using:

$$(10)\quad \begin{aligned} \dot{\theta} &= \frac{L}{m}u^2 \\ \ddot{\theta} &= \frac{2L}{m}u\frac{du}{d\theta}\dot{\theta} \\ &= \frac{2L^2}{m^2}u^3\frac{du}{d\theta} \end{aligned}$$

Substituting (10) into (9) leads to this:

$$(11)\quad \begin{aligned} &2\left(\frac{du}{d\theta}\frac{L}{m}\right)^2u^4 - u\left[\frac{d^2u}{d\theta^2}\left(\frac{L}{m}\right)^2u^4 + 2\left(\frac{du}{d\theta}\frac{L}{m}\right)^2u^3\right] \\ &\qquad = u^2\left(\frac{L}{m}\right)^2u^4 - Ku^5 \end{aligned}$$

Collecting terms we finally arrive at:

$$(12)\quad \frac{d^2u}{d\theta^2} = -\left(u - \frac{m^2K}{L^2}\right)$$

That's certainly solvable! Just to make it even easier, we substitute:

$$(13)\quad \begin{aligned} w &= u - \frac{m^2K}{L^2} \\ u &= w + \frac{m^2K}{L^2} \\ \frac{du}{d\theta} &= \frac{dw}{d\theta} \end{aligned}$$

which leads to:

$$(14)\quad \frac{d^2w}{d\theta^2} = -w$$

which has the obvious solution

(15) $$w = A\cos(\theta + \theta_0)$$

We observe that θ_0 represents a rotation of the axes. Changing the sign of ***A*** flips the figure relative to the axes. Since we don't care about the orientation, we set θ_0 to *zero* and assume ***A*** is nonnegative. As we can see in equation (16), this forces the "long" axis of the figure to be aligned with the ***x*** axis, with the "far" side of the orbit extending to the *left*. The *magnitude* of ***A*** is not so arbitrary, of course -- it depends on the initial conditions, including the radial velocity; changing the magnitude of ***A*** would change the shape of the orbit.

Substituting ***r*** back in finally leads us to:

(16) $$\boxed{r = \frac{1}{A\cos\theta + \frac{m^2K}{L^2}}}$$

What does the Polar Coordinate form of the solution tell us?

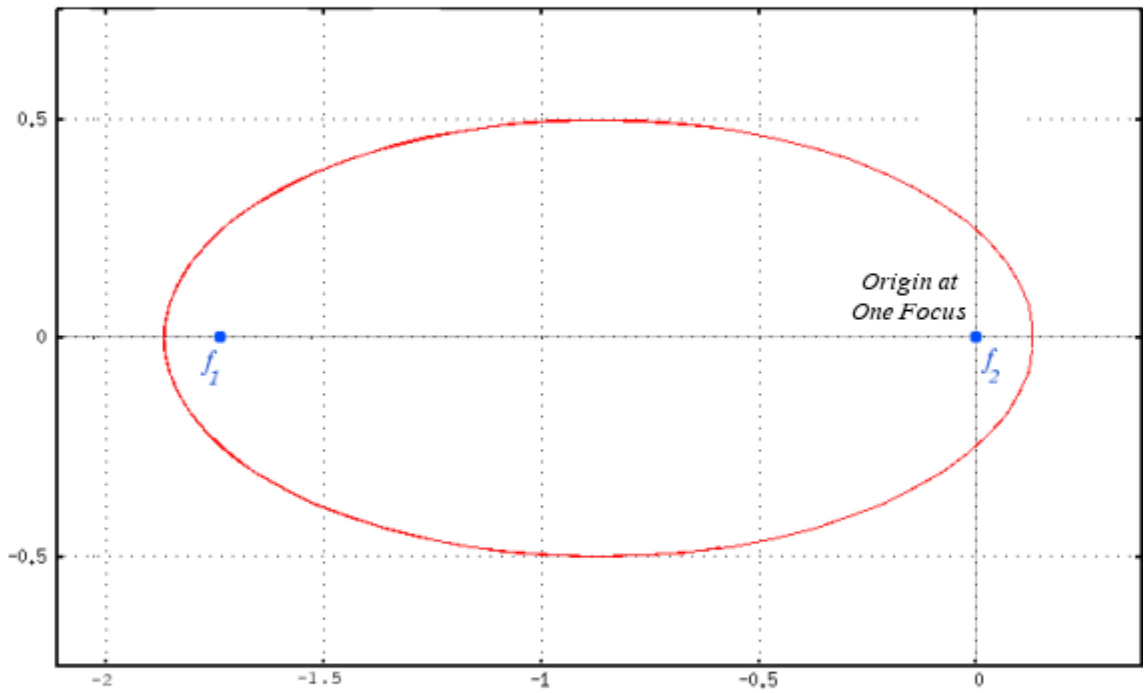
We can see quite a lot from (16). There are three separate cases (recall that we have assumed $A \geq 0$).

First case:

(17a) $$A < \frac{m^2K}{L^2}$$

In case (17a) the radius oscillates around the value m^2K/L^2. The orbit is clearly a closed loop, and is in fact an ellipse with the star at one focus as shown in figure 1 (and as we'll prove later on). The extremes are at $\theta=0$ and $\theta=\pi$; at $\theta=+\pi/2$ and $\theta=-\pi/2$ the radius is exactly L^2/m^2K. If ***A*** is *zero* then there is no oscillation; the orbit is circular, with fixed radius m^2K/L^2.

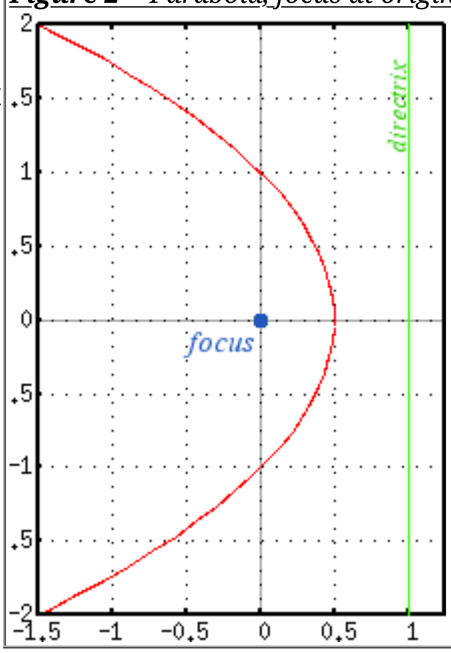
***Figure 1** -- An ellipse, long axis horizontal, with a focus at the origin:*



Second case:

(17b) $A = \frac{m^2 K}{L^2}$

Figure 2 -- Parabola, focus at origin:



In case (17b) the orbit goes *almost* all the way around, but there is a singularity at θ=π. The radius goes to infinity at that point on the orbit. What is this figure, which has infinite distance from the origin at exactly one angle? It's a parabola with the focus at the origin (where the star is located), as shown in figure 2, and as we'll actually prove later on.

Third case:

(17c) $A > \frac{m^2 K}{L^2}$

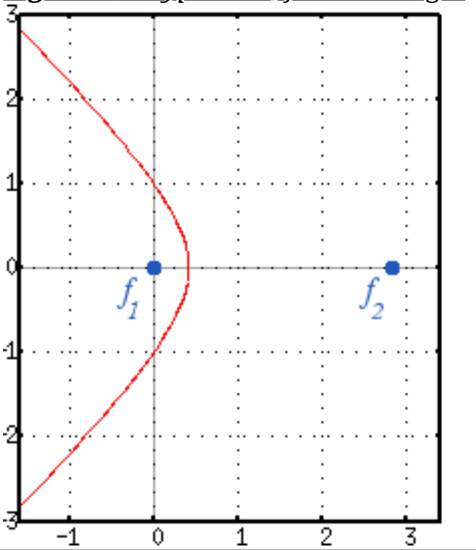
Finally, in case (17c) there is an entire forbidden sector; the radius is only positive and finite as long as

(18) $|\theta| < \arccos(-\frac{1}{A}\frac{m^2 K}{L^2})$

Note that the argument to arccos() in (18) is always *negative*. So, the "allowed" region for θ always includes, at the least, the region from +π/2 to -π/2, which is a half-circle. This makes sense: At arbitrarily high speed (and hence arbitrarily high ***L***) an object will zip through the solar system following what is essentially a straight line, and it will arrive and leave from points 180 degrees apart in the sky.

As θ approaches the forbidden values, the radius goes to infinity. We can turn that around, and say instead that as the radius goes to infinity, θ approaches a fixed value. Consequently, for points very far from the center, the satellite is essentially following a straight line away from the center in this case: the curve approaches an asymptote. This figure is a hyperbola, as shown in figure 3, and as we'll prove later on.

***Figure 3** -- Hyperbola, focus at origin:*



The Value of the Constant "A"

By choosing $\boldsymbol{A}$ to be nonnegative and choosing θ_0 to be zero, we've assured that the closest point in the orbit -- the perihelion -- will be on the X axis, when θ=0. If we call the radius at that point $\boldsymbol{r}_0$, then from (16), we have

(19) $$A = \frac{1}{r_0} - \frac{m^2K}{L^2}$$

We can use (19) to recast conditions (17) in terms of $\boldsymbol{r}_0$ and $\boldsymbol{L}$:

An elliptical orbit:

(20a) $$L < \sqrt{2r_0m^2K}$$

A parabolic orbit:

(20b) $$L = \sqrt{2r_0m^2K}$$

A hyperbolic orbit:

(20c) $$L > \sqrt{2r_0m^2K}$$

We can also see that the farthest point on an elliptical orbit -- the aphelion -- must be:

(21) $$r_1 = \frac{r_0}{2r_0\frac{m^2K}{L^2} - 1}$$

Returning to Cartesian coordinates, part I: Parabolic orbit

To finally prove that the orbits really are elliptical, parabolic, or hyperbolic, we need to recast them in a form we've previously shown represents those figures. That means we need to convert equation (16) back to Cartesian coordinates and see if we can extract each of the standard forms for the conic sections from it, depending on conditions (17).

First we observe that

(22)
$$\begin{aligned} \cos\theta &= \frac{x}{r} \\ r &= \sqrt{x^2 + y^2} \end{aligned}$$

To reduce clutter we introduce

(23) $$N \equiv \frac{m^2 K}{L^2}$$

Substituting (22) and (23) into (16) and cross-multiplying, we obtain:

(24) $$1 = Ax + N\sqrt{x^2 + y^2}$$

Shifting the last term to the left and the 1 to the right and then squaring, we obtain:

(25) $$x^2 + y^2 = \frac{A^2}{N^2}x^2 - \frac{2Ax}{N^2} + \frac{1}{N^2}$$

Collecting terms,

(26) $$y^2 + \left(1 - \frac{A^2}{N^2}\right)x^2 + \frac{2A}{N^2}x = \frac{1}{N^2}$$

If $\boldsymbol{A}^2=\boldsymbol{N}^2$, then looking back at condition (17b), we see that this should be a parabolic orbit. In that case the second term drops out, and, after substituting back from (23) and rearranging, we have:

(27) $$x - \frac{L^2}{2m^2K} = -\frac{m^2 K}{2L^2}y^2$$

If we shift the origin by substituting

(28) $$\zeta = x - \frac{L^2}{2m^2K}$$

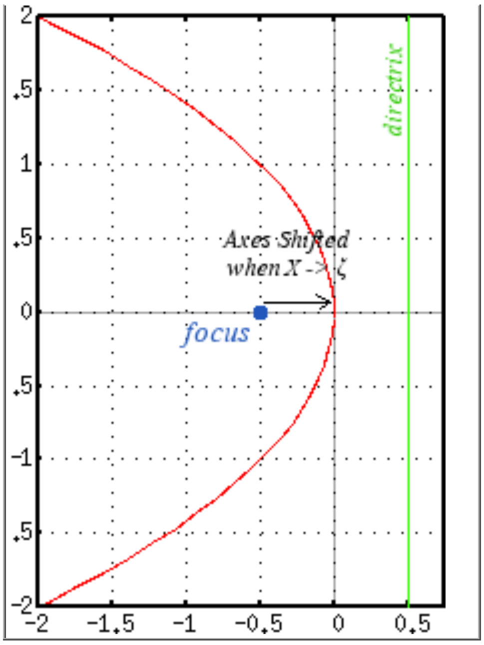
as shown in figure 4, we get the standard form

(29) $$\boxed{\zeta = -\frac{m^2 K}{2L^2}y^2}$$

Comparing this with equation (4) on our parabola focus page, we see that it's a parabola with the focus at

Figure 4 *-- Parabola, touching origin:*

(30)
$$focus = -\frac{L^2}{2m^2K}$$



Looking back at (28), we see that, before we shifted the axes, the focus of the parabola was indeed at the origin.

Returning to Cartesian coordinates, part II: Non-parabolic orbit

For the non-parabolic case, $A^2 \neq N^2$ and the linear term in (26) doesn't vanish. We'll continue by completing the square in (26):

(31) $$y^2 + \left(\frac{N^2-A^2}{N^2}\right)\left(x + \frac{A}{N^2-A^2}\right)^2 = \frac{A^2}{N^2(N^2-A^2)} + \frac{1}{N^2}$$

Again, to put this in standard form, we shift the origin by the substitution

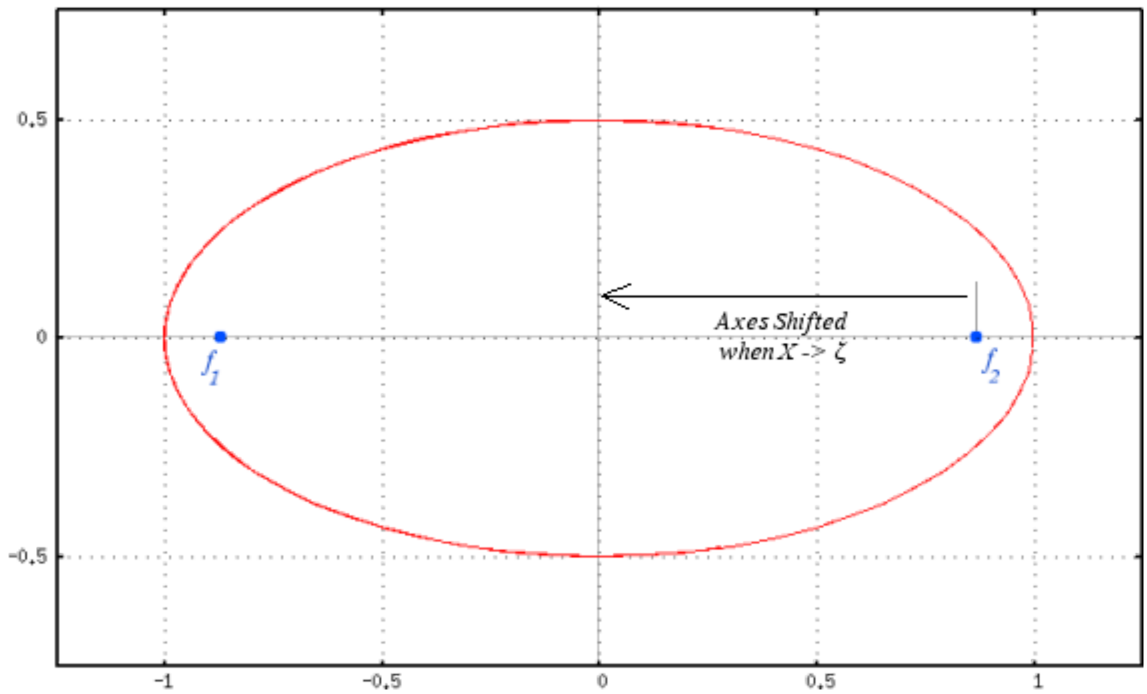
(32) $$\zeta = x + \frac{A}{N^2-A^2}$$

to obtain:

(33) $$\boxed{y^2 + \left(\frac{N^2-A^2}{N^2}\right)\zeta^2 = \frac{1}{N^2-A^2}}$$

Comparing with equation (8) on the ellipse focus page we see that, if $N^2>A^2$, this is indeed an ellipse with its long axis along the X axis, which is in agreement with condition (17a). Note that we shifted the figure to the *right* (and the axes to the *left*) in this case to center the origin between the foci, as shown in figure 5. We will check the exact location of the focus a little later.

***Figure 5** -- Ellipse, origin shifted so it's centered between the foci:*

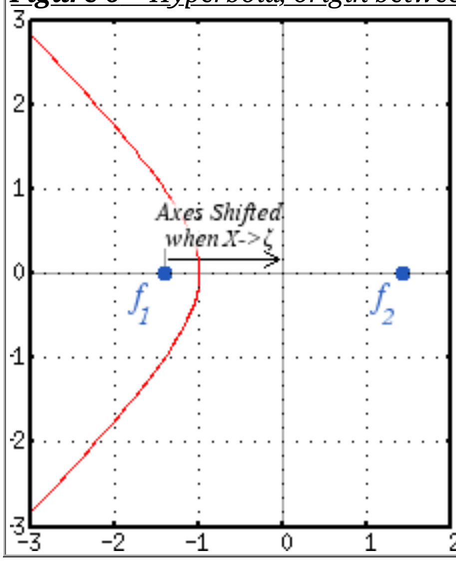


If, on the other hand, $N^2 < A^2$, it's surely a hyperbola, but it's not in standard form, because the right hand side is *negative* in this case. We divide through by the coefficient on ζ to obtain,

$$(34) \quad \zeta^2 - \left(\frac{N^2}{A^2-N^2}\right) y^2 = \frac{N^2}{(A^2-N^2)^2}$$

Comparing with (the *different*) equation (8) on the hyperbola focus page (I need better equation numbers!!), we see that this is a hyperbola in standard form with its axis along the X axis. This is in agreement with condition (17c), as was to be shown. Note that in this case we had a focus at the origin, and a second one off to the *right* (opposite side from the ellipse case). We shifted the origin to the *right* to center it between the foci, as shown in figure 6.

***Figure 6** -- Hyperbola, origin between the foci:*



Confirming that the focus really is in the middle of the star is a little messier for the ellipse and hyperbola, and we'll do that in the next section.

Where's the Focus of the Elliptical Orbit?

The standard forms of the ellipse and hyperbola equations place the origin midway between the two foci. We asserted that, in (16) (which was the polar coordinate form of the solution), one focus was located at the origin. If that was true, then from (32), we see that, in our final form, the foci *should* be at $\pm\frac{A}{N^2-A^2}$. We need to confirm that they really are. We shall start with the elliptical orbit.

To start with, we have too many "r" variables running around. We'll rewrite the ellipse equation here, representing the sum of the two distances from a point on the ellipse to the two foci as $2h$ rather than $2r$:

(35) $$\left(1-\frac{f^2}{h^2}\right)x^2+y^2 = h^2-f^2$$

Comparing (33) and (35) we equate the second terms:

(36) $$1-\frac{f^2}{h^2} = \frac{N^2-A^2}{N^2}$$
$$f^2 = \left[1-\frac{N^2-A^2}{N^2}\right]h^2$$

And we equate the right hand sides:

(37) $$h^2 = f^2+\frac{1}{N^2-A^2}$$

Substituting (37) into (36) we obtain:

(38) $$f^2 = \frac{N^2}{N^2-A^2}\left[1-\frac{N^2-A^2}{N^2}\right]\frac{1}{N^2-A^2}$$

Multiplying out, we obtain:

(39) $$f^2 = \frac{N^2}{(N^2-A^2)^2}-\frac{1}{N^2-A^2}$$

Sum the fractions:

(40) $$f^2 = \frac{A^2}{(N^2-A^2)^2}$$

Take square roots and we're done:

(41) $$f = \pm\frac{A}{N^2-A^2}$$

That takes care of the elliptical case. We still need to do the hyperbolic case.

Where's the Focus of the Hyperbolic Orbit?

As we did for the ellipse equation, we'll rewrite the equation for a hyperbola here, renaming "***r***" to "***h***":

(42)
$$y^2 - \left(\frac{1}{\gamma^2-1}\right) x^2 = \frac{1}{4}h^2$$
$$\gamma = \frac{2f}{h}$$

Comparing with (34), we equate the second terms:

(43)
$$\frac{1}{\gamma^2-1} = \frac{N^2}{A^2-N^2}$$

Dividing the top and bottom of the right side by N^2, we see that we must have:

(44)
$$\gamma = \frac{2f}{h} = \frac{A}{N}$$

We also equate the right hand sides of (42) and (34):

(45)
$$\frac{1}{4}h^2 = \frac{N^2}{(A^2-N^2)^2}$$
$$h = \pm\frac{2N}{A^2-N^2}$$

Combining (45) with (44) we obtain:

(46)
$$f = \frac{1}{2}\frac{A}{N} \cdot h = \pm\frac{A}{A^2-N^2}$$

which was to be shown.

Page created on 11/18/06; minor corrections, and flipped the sign on K (to make it positive) on 11/26/06