Tutorial: Binary Star Masses from Newton's Laws

Steven Kenneth Kauffmann*

Abstract

By Newton's Third Law the gravitational forces on the two stars of a binary system are equal and opposite, so the acceleration of each star is inversely proportional to its mass, which yields the ratio of the two stars' masses. However the difference of the two stars' gravitational accelerations is proportional to the sum of their masses; it is also inversely proportional to the square of their vector separation, so that vector separation traces out an elliptical orbit. The orbit's period plus its major axis length yields the sum of the stars' masses. The complete orbit isn't required; five or more of the points which lie on an ellipse determine it, and the orbit sweeps out the area enclosed by the ellipse at a constant rate.

The Newtonian gravitational accelerations of the stars of a binary system

Newton's Law of Gravitational Force together with his Second Law of Motion imply the following coupled equations of motion for the two stars of a binary system,

$$m_1\ddot{\mathbf{r}}_1 = -Gm_1m_2(\mathbf{r}_1-\mathbf{r}_2)/|\mathbf{r}_1-\mathbf{r}_2|^3 \quad\text{and}\quad m_2\ddot{\mathbf{r}}_2 = -Gm_2m_1(\mathbf{r}_2-\mathbf{r}_1)/|\mathbf{r}_2-\mathbf{r}_1|^3, \tag{1a}$$

which, since the forces $-Gm_1m_2(\mathbf{r}_1-\mathbf{r}_2)/|\mathbf{r}_1-\mathbf{r}_2|^3$ and $-Gm_2m_1(\mathbf{r}_2-\mathbf{r}_1)/|\mathbf{r}_2-\mathbf{r}_1|^3$ are equal and opposite, is consonant with Newton's Third Law of Motion. Eq. (1a) immediately yields,

$$m_1\ddot{\mathbf{r}}_1 + m_2\ddot{\mathbf{r}}_2 = \mathbf{0} \;\Rightarrow\; \ddot{\mathbf{r}}_2 = -(m_1/m_2)\ddot{\mathbf{r}}_1, \tag{1b}$$

so the *ratio* of the masses of the two stars of a binary system can immediately be obtained from the magnitudes of their oppositely-directed accelerations $\ddot{\mathbf{r}}_1$ and $\ddot{\mathbf{r}}_1$.

Moreover, the two Eq. (1a) coupled equations of motion can *also be written*,

$$\ddot{\mathbf{r}}_1 = -Gm_2(\mathbf{r}_1-\mathbf{r}_2)/|\mathbf{r}_1-\mathbf{r}_2|^3 \quad\text{and}\quad \ddot{\mathbf{r}}_2 = -Gm_1(\mathbf{r}_2-\mathbf{r}_1)/|\mathbf{r}_2-\mathbf{r}_1|^3, \tag{1c}$$

which shows that the acceleration of each star is independent of its own mass, in accord with the gravitational principle of equivalence. *Subtracting the two equations given by* Eq. (1c) yields the equation of motion *for the vector separation* $\mathbf{r} = (\mathbf{r}_1-\mathbf{r}_2)$ *of the two stars*,

$$\ddot{\mathbf{r}} = -GM\mathbf{r}/|\mathbf{r}|^3,\ \text{where } \mathbf{r} \stackrel{\text{def}}{=} (\mathbf{r}_1-\mathbf{r}_2) \text{ and } M \stackrel{\text{def}}{=} (m_1+m_2), \tag{1d}$$

which purely for reasons of familiarity of terminology, can also conveniently be presented as,

$$m\ddot{\mathbf{r}} = -GmM\mathbf{r}/|\mathbf{r}|^3,\ \text{where } m \stackrel{\text{def}}{=} m_1m_2/(m_1+m_2) = m_1m_2/M \text{ has the name “reduced mass”.} \tag{1e}$$

Since in Eq. (1e), $mM = m_1m_2$ and $\mathbf{r} = (\mathbf{r}_1-\mathbf{r}_2)$, the Eq. (1e) "force" $-GmM\mathbf{r}/|\mathbf{r}|^3$ is equal to the Eq. (1a) force $-Gm_1m_2(\mathbf{r}_1-\mathbf{r}_2)/|\mathbf{r}_1-\mathbf{r}_2|^3$. The Eq. (1e) presentation of the Eq. (1d) vector-separation equation of motion sanctions the use of familiar terminology such as "force" for $-GmM\mathbf{r}/|\mathbf{r}|^3$, "angular momentum" $\mathbf{L}$ for $m(\mathbf{r}\times\dot{\mathbf{r}})$ and "energy" E for $m(|\dot{\mathbf{r}}|^2/2 - GM/|\mathbf{r}|)$.

In the next section we obtain *the elliptical locus* and also *the relation of the orbital period to the area enclosed by that locus* of the orbit described by the Eq. (1d) equation of motion *for the vector-separation* $\mathbf{r} = (\mathbf{r}_1-\mathbf{r}_2)$ *of the two stars.* This yields *the sum of the masses* $M = (m_1+m_2)$ *of the two stars* from the that orbit's period and major-axis length in conjunction with the universal gravitational constant G.

The sum of the two stars' masses from their vector-separation elliptical orbit

The Eq. (1d) vector-separation equation of motion $\ddot{\mathbf{r}} = -GM\mathbf{r}/|\mathbf{r}|^3$ yields "angular momentum" conservation,

$$d(\mathbf{L}/m)/dt = d(\mathbf{r}\times\dot{\mathbf{r}})/dt = (\dot{\mathbf{r}}\times\dot{\mathbf{r}}) + (\mathbf{r}\times\ddot{\mathbf{r}}) = (\dot{\mathbf{r}}\times\dot{\mathbf{r}}) - GM(\mathbf{r}\times\mathbf{r})/|\mathbf{r}|^3 = \mathbf{0}. \tag{2a}$$

Since $(\mathbf{L}/m) = (\mathbf{r}\times\dot{\mathbf{r}})$ *is a constant vector*, $\mathbf{r}(t)$ and $\dot{\mathbf{r}}(t)$ *are always confined to the plane perpendicular to that constant vector*, i.e., $\mathbf{r}(t)$ *is planar*. Thus $|\mathbf{r}\times\dot{\mathbf{r}}| = (|\mathbf{L}|/m)$ *alone* is relevant, so we define $L \stackrel{\text{def}}{=} |\mathbf{L}|$.

The Eq. (1d) vector-separation equation of motion $\ddot{\mathbf{r}} = -GM\mathbf{r}/|\mathbf{r}|^3$ also yields "energy" conservation,

$$\begin{gathered} d(E/m)/dt = d\big(|\dot{\mathbf{r}}|^2/2 - GM/|\mathbf{r}|\big)/dt = \\ d\big((\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})/2 - GM/(\mathbf{r}\cdot\mathbf{r})^{\frac{1}{2}}\big)/dt = (\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}}) + GM\big((\dot{\mathbf{r}}\cdot\mathbf{r})/(\mathbf{r}\cdot\mathbf{r})^{\frac{3}{2}}\big) = \dot{\mathbf{r}}\cdot\big(\ddot{\mathbf{r}} + GM\mathbf{r}/|\mathbf{r}|^3\big) = 0. \end{gathered} \tag{2b}$$

*Retired, American Physical Society Senior Life Member, E-mail: SKKauffmann@gmail.com

Having shown that the Eq. (1d) equation of motion $\ddot{\mathbf{r}} = -GM\mathbf{r}/|\mathbf{r}|^3$ for the vector-separation $\mathbf{r} = (\mathbf{r}_1 - \mathbf{r}_2)$ of the two stars implies the conservation relations $(\mathbf{r} \times \dot{\mathbf{r}}) = (\mathbf{L}/m)$ and $(|\dot{\mathbf{r}}|^2/2 - GM/|\mathbf{r}|) = (E/m)$, we would like *to solve these conservation relations* for the *locus* of that two-star vector-separation orbit, from which we in turn would like to obtain enough information to determine $M = (m_1 + m_2)$, the sum of the two star masses. Since we now know that this orbit is *planar*, the first thing we need to do is to express these conservation relations in *two-dimensional* polar coordinates, which have the following properties,

$$\begin{gathered}\mathbf{r} = (r\cos\theta,\, r\sin\theta), \quad |\mathbf{r}| = r, \\ \dot{\mathbf{r}} = (\dot{r}\cos\theta - r\dot{\theta}\sin\theta,\, \dot{r}\sin\theta + r\dot{\theta}\cos\theta), \quad |\dot{\mathbf{r}}|^2 = (\dot{r})^2 + r^2(\dot{\theta})^2, \quad |\mathbf{r}\times\dot{\mathbf{r}}| = r^2|\dot{\theta}|.\end{gathered} \tag{3a}$$

Thus expressed in two-dimensional polar coordinates, $|\mathbf{r} \times \dot{\mathbf{r}}| = (L/m)$ becomes,

$$r^2|\dot{\theta}| = (L/m), \tag{3b}$$

and in those coordinates $(|\dot{\mathbf{r}}|^2/2 - GM/|\mathbf{r}|) = (E/m)$ becomes,

$$\big((\dot{r})^2 + r^2(\dot{\theta})^2\big)/2 - GM/r = (E/m) \tag{3c}$$

We *won't try to solve* Eqs. (3c) and (3b) for $r(t)$ and $\theta(t)$. We intend to use Eqs. (3c) and (3b) *to obtain only the orbit's locus* $r(\theta)$. Therefore we *first* insert the relation $(\dot{r})^2 = (dr/d\theta)^2(\dot{\theta})^2$ into Eq. (3c), and *only then* do we use Eq. (3b) to eliminate $|\dot{\theta}|$ from Eq. (3c), *which produces the following locus differential equation*,

$$(L/m)^2 r^{-4}\big((dr/d\theta)^2 + r^2\big)/2 - GM/r = (E/m). \tag{3d}$$

The disquieting factor r^{-4} in Eq. (3d) is eliminated upon changing the dependent variable from r to $u = (1/r)$ because $dr/d\theta = -u^{-2}(du/d\theta)$; after that change of dependent variable, Eq. (3d) reads,

$$(L/m)^2\big((du/d\theta)^2 + u^2\big)/2 - GMu = (E/m) \tag{3e}$$

We know that Eq. (1d) *can be satisfied by circular orbits*. Indeed Eq. (3e) can *also* be satisfied by the simple circular *constant locus* $u(\theta) = 1/\rho_0$ if $(L/m)^2$ and (E/m) are adjusted to accommodate it. But circular orbits *clearly aren't general solutions of either* Eq. (1d) *or* Eq. (3e). In the limit of *zero gravitational force*, namely *when M is put to zero*, Eq. (1d) is obviously satisfied *by straight-line trajectories* (Newton's First Law of Motion). In two-dimensional *rectangular* coordinates, *a straight-line locus* has the form,

$$ax - by = \big(a^2 + b^2\big)^{\frac{1}{2}}\rho_0,$$

which is changed to *polar* coordinates *by substituting* $(r\cos\theta)$ *for* x *and* $(r\sin\theta)$ *for* y, with the result,

$$r\Big[\cos\theta\Big(a/\big(a^2+b^2\big)^{\frac{1}{2}}\Big) - \sin\theta\Big(b/\big(a^2+b^2\big)^{\frac{1}{2}}\Big)\Big] = \rho_0.$$

Of course *there is an angle δ such that* $\cos\delta = a/(a^2+b^2)^{\frac{1}{2}}$ *and* $\sin\delta = b/(a^2+b^2)^{\frac{1}{2}}$. That, plus the fact that $u = 1/r$, enables *this straight-line locus in two dimensions* to be written,

$$u(\theta) = \cos(\theta + \delta)/\rho_0. \tag{3f}$$

For *the particular zero-gravitational-force case that* $M = 0$, substitution of the Eq. (3f) *straight-line locus* into *the locus relation of* Eq. (3e), which is *consistent with the dynamics of the Newtonian* Eq. (1d), yields,

$$(L/(m\rho_0))^2/2 = (E/m). \tag{3g}$$

Eq. (3g) requires *nonnegative energy* E, which is of course *entirely expected* of a straight-line locus under the circumstance that the gravitational *force* has been put to *zero*. We have now exhibited *two special solutions* of the Eq. (3e) *locus relation*, which itself *is consistent with the gravitational Newtonian dynamics of* Eq. (1d). These two special locus solutions of Eq. (3e) are the constant *circular locus* $u(\theta) = 1/\rho_0$, which is entirely compatible with gravitational force and negative energy, and the Eq. (3f) *straight-line locus* $u(\theta) = \cos(\theta+\delta)/\rho_0$ which is progeny of *zero gravitational force* and absolutely requires *nonnegative energy*.

At this point it seems not unreasonable to *guess* that *the general solution* of the Eq. (3e) *gravitational locus relation* is an arbitrary *linear combination* of the constant *circular locus* $u(\theta) = 1/\rho_0$ with the Eq. (3f) *straight-line locus* $u(\theta) = \cos(\theta + \delta)/\rho_0$, namely,

$$u(\theta) = (1 - \beta\cos(\theta + \delta))/\rho_0. \tag{3h}$$

In the next paragraph we show that the Eq. (3h) circle/straight-line hybrid locus *indeed always satisfies the* Eq. (3e) *gravitational locus relation.* Moreover, it turns out that if the Eq. (3e) energy is *negative*, then β^2 *is constrained to be less than unity*, so a *bound* gravitational locus *is always more a circle than it is a straight line*. If the Eq. (3h) circle/straight-line hybrid gravitational locus is converted to rectangular coordinates and *inspected*, it is seen *to always be a conic section.* Of course $\beta = 0$ produces circular loci, but furthermore, $0 < \beta^2 < 1$ produces elliptical loci, $\beta^2 = 1$ produces parabolic loci and $\beta^2 > 1$ produces hyperbolic loci. Needless to say, $\beta^2 \to \infty$ corresponds to straight lines (which *technically can't exist unless* M *is put to zero in the* Eq. (3e) *gravitational locus relation*).

We now *insert* the Eq. (3h) circle/straight-line hybrid locus into the Eq. (3e) gravitational locus relation,

$$(L/(m\rho_0))^2\big(-2\beta\cos(\theta + \delta) + 1 + \beta^2\big)/2 - (GM/\rho_0)\big(-\beta\cos(\theta + \delta) + 1\big) = (E/m) \tag{3i}$$

Since in Eq. (3i) *the coefficient of* $\cos(\theta+\delta)$ *must vanish*, the *first* consequence of Eqs. (3h) and (3e) is that,

$$GM = \big((L/m)^2/\rho_0\big). \tag{3j}$$

Putting the Eq. (3j) result *back into* Eq. (3i) yields,

$$(E/m) = (L/(m\rho_0))^2\big(\beta^2 - 1\big)/2, \tag{3k}$$

from which we immediately see that the energy E of this gravitational system *is negative only if* $\beta^2 < 1$, namely only if the Eq. (3h) circle/straight-line hybrid locus is more a circle than it is a straight line. Further on we study the Eq. (3h) circle/straight-line hybrid orbit locus *in vastly clearer detail* by *converting* it to two-dimensional *rectangular coordinates*; it's behavior *is almost impossible to grasp in the form it is written in* Eq. (3h) *because large values of* $r(\theta)$ *produce entirely innocuous-looking small values of* $u(\theta) = 1/r(\theta)$. We shall see *that it is always a conic section*. Of course we are well-aware that it is a *circle* when $\beta = 0$, but *a better overview* furthermore shows that it is an *ellipse* when $0 < \beta^2 < 1$, a *parabola* when $\beta^2 = 1$ and a *hyperbola* when $\beta^2 > 1$. Therefore *it is little wonder* that the Eq. (3k) *energy* of this gravitational system *is negative only if* $\beta^2 < 1$, *namely only if its orbit locus is a circle or ellipse.*

Eq. (3j) *is the ticket to obtaining* $M = (m_1+m_2)$, the sum of the masses of the two stars, from the universal gravitational constant G *and particular inputs from the vector-separation orbit* $\mathbf{r}(t) = (\mathbf{r}_1(t) - \mathbf{r}_2(t))$ *of the two stars*. The *two ingredients* which enter into the *right side* of Eq. (3j) are the Eq. (3h) *orbit parameter* ρ_0 and the Eq. (3b) *conserved dynamical orbit entity* $(L/m) = r^2|\dot\theta|$, which, since the infinitesimal area $|dA|$ of the plane that corresponds to an infinitesimal angular arc $|d\theta|$ of the orbit is,

$$|dA| = \tfrac{1}{2}r(r|d\theta|) = r^2|d\theta|/2,$$

the area of the plane which the orbit sweeps out per unit time is,

$$|dA/dt| = r^2|\dot\theta|/2 = (L/m)/2. \tag{3l}$$

It was Johannes Kepler who first realized, in the course of studying the precise planetary-orbit observations of Tycho Brahe, *that* $|dA/dt|$ *is a conserved dynamical orbit entity*. Thus it has been *routine* for around 400 years for astronomers *to read off* $(L/m) = 2|dA/dt|$ *from orbital data*. Alternatively, *if the period* T *of the orbit has been observed*, then since $|dA/dt|$ is constant in time,

$$(L/m) = 2|dA/dt| = 2A/T, \tag{3m}$$

where A *is the area enclosed by the complete orbit*. In terms of *the two* Eq. (3h) *orbit parameters* ρ_0 *and* β, it turns out that when $\beta^2 < 1$, so that the complete orbit is an ellipse or a circle, its area A is,

$$A = \pi\,\rho_0^2/\big(1 - \beta^2\big)^{\frac{3}{2}} = \pi\,r_0R_0, \text{ where } r_0 = \rho_0/(1 - \beta^2)^{\frac{1}{2}} \text{ and } R_0 = \rho_0/(1 - \beta^2). \tag{3n}$$

This r_0 and R_0 are the half-lengths of, respectively, the minor and major ellipse axes. Thus from Eq. (3m),

$$(L/m) = (2\pi/T)\,\rho_0^2/\big(1 - \beta^2\big)^{\frac{3}{2}} = (2\pi/T)\,r_0R_0. \tag{3o}$$

Therefore from Eqs. (3j) and (3o) the sum of the masses of the two stars is given by,

$$(m_1 + m_2) = M = G^{-1}(2\pi/T)^2\big(\rho_0/\big(1 - \beta^2\big)\big)^3 = G^{-1}(2\pi/T)^2R_0^3. \tag{3p}$$

Is there a neat way, in terms of observations, to characterize r_0 and R_0? Perhaps the neatest way is to note that the perigee distance of the orbit, *the smallest distance between the two stars*, is $R_0 - (R_0^2 - r_0^2)^{\frac{1}{2}}$, whereas the apogee distance of the orbit, *the greatest distance between the two stars*, is $R_0 + (R_0^2 - r_0^2)^{\frac{1}{2}}$. Therefore R_0 *is the arithmetic mean of the greatest and smallest distances between the two stars* whereas r_0 *is the geometric mean of the greatest and smallest distances between the two stars*.

It may also be of interest that the two Eq. (3h) orbit parameters ρ_0 and β are related to r_0 and R_0 by $\rho_0 = r_0^2/R_0$, $\beta = (1 - (r_0/R_0)^2)^{\frac{1}{2}}$ and $(1 - \beta^2)^{\frac{1}{2}} = (r_0/R_0)$. It has already been pointed out in Eq. (3n) that $r_0 = \rho_0/(1 - \beta^2)^{\frac{1}{2}}$ and $R_0 = \rho_0/(1 - \beta^2)$.

We have pointed out that ever since Johannes Kepler's discovery that $|dA/dt|$ is a conserved dynamical orbit entity, astronomers routinely read off $(L/m) = 2|dA/dt|$ from orbital data. Alternatively, if the period T of the orbit has been observed, we have from Eq. (3o) that $(L/m) = (2\pi/T)\, r_0 R_0$. Likewise, for the basic parameters r_0 and R_0 of the ellipse, we have pointed out that R_0 is the arithmetic mean of the smallest and greatest distances between the two stars, whereas r_0 is the geometric mean of those two distances. Alternatively, just as a circle is in principle determined by three or more of the points which lie on it, an ellipse is in principle determined by five or more of the points which lie on it.

Finally, we have pointed out that the Eq. (3h) representation of the two-dimensional locus of a gravitational orbit, namely $u(\theta) = (1 - \beta\cos(\theta + \delta))/\rho_0$, is *as opaque* as it is *simple*. We now *remedy* that by expressing it in rectangular coordinates. Since $u = 1/r$, $\cos(\theta + \delta) = \cos\theta\cos\delta - \sin\theta\sin\delta$, $x = r\cos\theta$ and $y = r\sin\theta$, then after we *multiply* the Eq. (3h) representation $1/r = (1 - \beta(\cos\theta\cos\delta - \sin\theta\sin\delta))/\rho_0$ of the gravitational-orbit locus *through by* $(r\rho_0)$, we readily express it in terms of x and y as,

$$\left(y^2 + x^2\right)^{\frac{1}{2}} = \rho_0 + \beta(x\cos\delta - y\sin\delta). \tag{4a}$$

We now *rotate the axes of our reference* (x, y) *coordinate system in order to effectively render the angle* δ *zero*, which allows us to deal *with the more symmetrically-oriented locus*,

$$\left(y^2 + x^2\right)^{\frac{1}{2}} = \rho_0 + \beta x. \tag{4b}$$

We square both sides of Eq. (4b) and reorganize the resulting terms; the upshot is the *conic-section locus*,

$$y^2 + (1 - \beta^2)\Big(x - \big(\beta\rho_0/(1 - \beta^2)\big)\Big)^2 = \rho_0^2/(1 - \beta^2). \tag{4c}$$

This conic section locus is a circle when $\beta = 0$, an ellipse when $0 < \beta^2 < 1$, a parabola when $\beta^2 = 1$ and a hyperbola when $\beta^2 > 1$. Thus this locus *only applies to a binary star system if* $\beta^2 < 1$. In *that case*, Eq. (4c) *yields the semi-minor axis length* $r_0 = \rho_0/(1 - \beta^2)^{\frac{1}{2}}$ *and the semi-major axis length* $R_0 = \rho_0/(1 - \beta^2)$, *which is in full accord with* Eq. (3n) *and the work subsequent to it.* Since $(1 - \beta^2) = (r_0/R_0)^2$, $\beta = (R_0^2 - r_0^2)^{\frac{1}{2}}/R_0$ and $\rho_0 = r_0^2/R_0$, *writing the* Eq. (4c) *ellipse orbit locus in terms of* r_0 *and* R_0 *produces*,

$$y^2 + (r_0/R_0)^2\Big(x - \big(R_0^2 - r_0^2\big)^{\frac{1}{2}}\Big)^2 = r_0^2. \tag{4d}$$

The perigee of the Eq. (4d) ellipse orbit locus occurs at the *coordinates* $(-R_0 + (R_0^2 - r_0^2)^{\frac{1}{2}}, 0)$, with the corresponding perigee *distance* being $R_0 - (R_0^2 - r_0^2)^{\frac{1}{2}}$, while the apogee of the Eq. (4d) ellipse orbit locus occurs at the *coordinates* $(R_0 + (R_0^2 - r_0^2)^{\frac{1}{2}}, 0)$, with the corresponding apogee *distance* being $R_0 + (R_0^2 - r_0^2)^{\frac{1}{2}}$. These perigee and apogee *distances* $R_0 - (R_0^2 - r_0^2)^{\frac{1}{2}}$ *and* $R_0 + (R_0^2 - r_0^2)^{\frac{1}{2}}$ *have been broached above*, where it was noted that R_0 *is their arithmetic mean whereas* r_0 *is their geometric mean.*

The *area* A of the Eq. (4d) ellipse orbit locus is,

$$\begin{aligned} A = 2r_0 \int_{-R_0 + \left(R_0^2 - r_0^2\right)^{\frac{1}{2}}}^{R_0 + \left(R_0^2 - r_0^2\right)^{\frac{1}{2}}} \left(1 - \left(\left(x - \left(R_0^2 - r_0^2\right)^{\frac{1}{2}}\right)/R_0\right)^2\right)^{\frac{1}{2}} dx = 2r_0 R_0 \int_{-1}^{1} \left(1 - u^2\right)^{\frac{1}{2}} du = \\ 2r_0 R_0 \int_{-\pi/2}^{\pi/2} \cos^2\theta\, d\theta = r_0 R_0 \int_{-\pi/2}^{\pi/2} (1 + \cos 2\theta) d\theta = r_0 R_0[\pi + (\sin\pi - \sin(-\pi))/2] = \pi\, r_0 R_0, \end{aligned} \tag{4e}$$

a result *which has been utilized above, starting with* Eq. (3n).