This document contains all the math behind this simulator.

Contents

1 Basic Orbital Mechanics

We don't simulate the full n-body problem here, only the "patched-conic" approximation, which can be solved in closed-form.

The most basic orbit is a conic section: an ellipse, parabola, or hyperbola. There is also one additional possibility: if the secondary body is moving straight up or down from the primary, it is a "radial" orbit.

1.1 Orbital Parameters

Here are the physical parameters of an orbit we consider in this document:

- Gravitational parameter (μ): this represents the strength of the primary body's attraction, measured in $\mathrm{m}^3/\mathrm{s}^2$. It's more often seen as the mass of the primary times G, the gravitational constant. But it turns out, μ is the only thing we can directly measure, so we know that to more precision.

- Semi-major axis (a): The major axis of an ellipse is the longest distance across the it; the semi-major axis is half that. For a parabola, this is infinite, and for a hyperbola, it's negative.

- Eccentricity (e): This represents how "squished" the orbit is. A circle has an eccentricity of zero, an ellipse's is between zero and one, a parabola has exactly one, and a hyperbola's is more than one. A radial orbit also has exactly one, which will prove to be frustrating later.

- Specific orbital energy (ε): The total energy of the system, divided by the reduced mass. For our purposes, since the primary body is significantly heavier than the secondary, the reduced mass is just the mass of the secondary. Like the total energy, this is conserved.
- Specific angular momentum (h): The angular momentum of the system, divided by the reduced mass. Also conserved.
- Periapsis (r_p): The point in the secondary's orbit where it is closest to the primary.
- Apoapsis (r_a): The point in the secondary's orbit where it is furthest away. If the orbit is open, no such point exists.
- Semilatus rectum (ℓ): The latus rectum is the chord through the primary body perpendicular to the focal axis. The semilatus rectum is half of that.

Here are some relations between the parameters above:

$$\varepsilon = -\frac{\mu}{2a} \qquad \ell = \frac{h^2}{\mu} \qquad h^2 = \mu a(1-e^2) \qquad e^2 - 2\varepsilon\frac{h^2}{\mu^2} = 1$$

$$v^2 = \mu\left(\frac{2}{r} - \frac{1}{a}\right) \quad \text{(vis-viva equation)}$$

$$r_p = a(1-e) = \frac{h^2}{\mu(1+e)} \qquad v_p = \frac{h}{r_p} = \frac{\mu(1+e)}{h} \qquad v_p^2 = \frac{h^2}{r_p^2} = \frac{\mu}{a}\frac{1+e}{1-e}$$

$$r = \frac{h^2/\mu}{1+e\cos\theta} = a(1-e\cos E) = a(1-e\cosh H)$$

For a circular orbit:

$$\varepsilon = -\frac{\mu}{2r} \quad e = 0 \quad h = \sqrt{\mu r} \quad \ell = r \quad v = \frac{h}{r}$$

For a parabolic orbit:

$$\varepsilon = 0 \quad e = 1 \quad h = \text{anything} \quad r_p = \frac{h^2}{2\mu} = \frac{\ell}{2} \quad v_p = \frac{2\mu}{h}$$

TODO keplerian parameters?

2 Parameterizing the Orbit

The previous section dealt with the properties that the orbit, as a whole, has. But to simulate the trajectory of an orbiting body, we need to get into the dynamics. Starting from a known position and velocity, we want to travel along the orbit over a period of time, and compute our final state.

It turns out it will be easiest to do this by introducing a new auxiliary variable s, which will behave like a modified time variable. This section goes through the derivation of s, the equations governing position, velocity, and time, and the relationship of s to the classical anomalies.

2.1 Regularizing the Equation of Motion

From Newton's law of gravitation, the equation of motion for a body orbiting a stationary mass is

$$\ddot{\boldsymbol{x}} + \frac{\mu}{r^3}\boldsymbol{x} = 0 \tag{2.1.1}$$

We introduce a new variable s such that $\mathrm{d}s/\mathrm{d}t = 1/r$, and $s = 0$ at t_0. Since dots are used to represent derivatives with respect to t, we'll use primes to represent derivatives with respect to s. From the chain rule, we can derive the following relations for any state function f, which will be used repeatedly:

$$\dot{f} = \frac{1}{r}f' \qquad \ddot{f} = \frac{rf'' - f'r'}{r^3} \tag{2.1.2}$$

Using this, we rewrite (2.1.1) in terms of s:

$$\boldsymbol{x}'' - \frac{r'}{r}\boldsymbol{x}' + \frac{\mu}{r}\boldsymbol{x} = 0 \tag{2.1.3}$$

This is more complicated than before, but it can be simplified by introducing some conserved physical quantities. In particular, consider the energy and the eccentricity vector. The former is well-known, and can be computed from the kinetic plus potential energy. However, it will be convenient to use the quantity $\beta = -2\varepsilon$:

$$\varepsilon = \frac{v^2}{2} - \frac{\mu}{r} \qquad \beta = \frac{2\mu}{r} - v^2 \tag{2.1.4}$$

TODO should i use alpha?

The latter is more obscure, but it's also a conserved quantity, and can be expressed as:

$$\boldsymbol{e} = \frac{\boldsymbol{v} \times \boldsymbol{h}}{\mu} - \frac{\boldsymbol{x}}{r} = \left(\frac{v^2}{\mu} - \frac{1}{r}\right)\boldsymbol{x} - \frac{\boldsymbol{x} \cdot \boldsymbol{v}}{\mu}\boldsymbol{v} \tag{2.1.5}$$

Multiplying this by μ, substituting in β, and using the identity $\boldsymbol{x} \cdot \boldsymbol{v} = r\dot{r}$, we get

$$\mu\boldsymbol{e} = \left(\frac{\mu}{r} - \beta\right)\boldsymbol{x} - r\dot{r}\boldsymbol{v} \tag{2.1.6}$$

Now, applying (2.1.2) to $\dot{r}$ and $\boldsymbol{v} = \dot{\boldsymbol{x}}$, we get:

$$\mu\boldsymbol{e} = \left(\frac{\mu}{r} - \beta\right)\boldsymbol{x} - \frac{r'}{r}\boldsymbol{x} \tag{2.1.7}$$

This now has an $(r'/r)\boldsymbol{x}$ term, which we isolate and plug into (2.1.3), giving us our regularized equation of motion:

$$\boldsymbol{x}'' + \beta\boldsymbol{x} = -\mu\boldsymbol{e} \tag{2.1.8}$$

This form of the equation, without any rs in the denominator, will be easier to solve, and afterwards, we'll delve into the meaning of s.

2.2 Separation into f and g Functions

The independent variable $\boldsymbol{x}$ is a vector, so we want to express it in a convenient basis, and see if we can separate this vector-valued equation into some scalar-valued ones. Since the initial position and velocity span the orbital plane, they are a convenient basis to choose. There are thus some functions $f(s)$ and $g(s)$ such that

$$\boldsymbol{x}(s) = f(s)\boldsymbol{x}_0 + g(s)\boldsymbol{v}_0 \tag{2.2.1}$$

Conveniently, we also know the coefficients of $\mu\boldsymbol{e}$ in this basis, from (2.1.6). Applying this basis to our equation of motion (2.1.8), we get:

$$\begin{gathered}
\boldsymbol{x}'' + \beta\boldsymbol{x} = -\mu\boldsymbol{e} \\
(f''\boldsymbol{x}_0 + g''\boldsymbol{v}_0) + \beta(f\boldsymbol{x}_0 + g\boldsymbol{v}_0) = -\left[\left(\frac{\mu}{r_0} - \beta\right)\boldsymbol{x}_0 - r_0\dot{r}_0\boldsymbol{v}_0\right] \\
(f'' + \beta f)\boldsymbol{x}_0 + (g'' + \beta g)\boldsymbol{v}_0 = \left(\beta - \frac{\mu}{r_0}\right)\boldsymbol{x}_0 + r_0\dot{r}_0\boldsymbol{v}_0
\end{gathered}$$

Since $\boldsymbol{x}_0$ and $\boldsymbol{v}_0$ are independent (in the general case), we can separately equate the coefficients on each basis vector, giving us two uncoupled, second-order, inhomogeneous, linear differential equations.

$$f'' + \beta f = \beta - \frac{\mu}{r_0} \qquad g'' + \beta g = r_0\dot{r}_0 \tag{2.2.2}$$

Solving these will give us the functions $f(s)$ and $g(s)$, which will give us $\boldsymbol{x}(s)$.

2.3 Stumpff Functions

To solve the inhomogeneous diffeqs in (2.2.2), we first have to solve the associated homogeneous equations. Fortunately, both f and g have the same equation, which has well-known solutions, the form of which depend on the sign of β:

$$f(s) = g(s) = \begin{cases} A\cos(\sqrt{\beta}s) + B\sin(\sqrt{\beta}s) & \beta > 0 \\ A\cosh(\sqrt{-\beta}s) + B\sinh(\sqrt{-\beta}s) & \beta < 0 \\ A + Bs & \beta = 0 \end{cases} \tag{2.3.1}$$

If we were only dealing with a specific β, then this would be enough; either we get cosines and sines (and thus, ellipses), or we get hyperbolic trig functions (hyperbolas), or when β is exactly zero, we get a linear equation. However, it would be nice to find an approach that is universal in β, for a few reasons:

1. We don't have to do any casework in subsequent proofs
2. We don't have to do any casework in the code
3. The transition from $\beta > 0$ to $\beta < 0$ is smooth and physically realizable; there **should** be some equations valid in both regimes!

We can do so with the help of *Stumpff functions*. The kth Stumpff function is defined by the following series, which converges absolutely for all x:

$$c_k(x) = \sum_{i=0}^{\infty} \frac{(-x)^i}{(k+2i)!} = \frac{1}{k!} - \frac{x}{(k+2)!} + \frac{x^2}{(k+4)!} - \frac{x^3}{(k+6)!} + \cdots \tag{2.3.2}$$

With some thought, the series for c_0 and c_1 can be recognized as related to trigonometric series, and are expressible in a nice-ish form.

$$c_0(x) = \begin{cases} \cos(\sqrt{x}) & x > 0 \\ \cosh(\sqrt{-x}) & x < 0 \\ 1 & x = 0 \end{cases} \tag{2.3.3}$$

$$c_1(x) = \begin{cases} \sin(\sqrt{x})/\sqrt{x} & x > 0 \\ \sinh(\sqrt{-x})/\sqrt{-x} & x < 0 \\ 1 & x = 0 \end{cases} \tag{2.3.4}$$

For $k \geq 2$, we can use the following recurrence relation between c_k and c_{k+2}:

$$c_k(x) = \frac{1}{k!} - xc_{k+2}(x) \tag{2.3.5}$$

We'll make use of c_2 and c_3 later, so let's expand them now:

$$c_2(x) = \begin{cases} (1 - \cos(\sqrt{x}))/x & x > 0 \\ (1 - \cosh(\sqrt{-x}))/x & x < 0 \\ 1/2 & x = 0 \end{cases} \tag{2.3.6}$$

$$c_3(x) = \begin{cases} (\sqrt{x} - \sin(\sqrt{x}))/\sqrt{x^3} & x > 0 \\ (\sqrt{x} - \sinh(\sqrt{-x}))/\sqrt{-x^3} & x < 0 \\ 1/6 & x = 0 \end{cases} \tag{2.3.7}$$

There's a strong similarity between c_0 and c_1, and the solutions to our diffeq in (2.3.1). We can rewrite the general solution in terms of these functions (note that the $\sqrt{\pm\beta}$ can get folded into the constant, but the s cannot):

$$f(s) = g(s) = Ac_0(\beta s^2) + Bsc_1(\beta s^2) \tag{2.3.8}$$

TODO show any two of them work?

A notation that is somewhat cleaner is to introduce the two-variable functions $G_k(\beta, s)$:

$$G_k(\beta, s) = s^k c_k(\beta s^2) \tag{2.3.9}$$

Like the c_k, these are smooth functions, defined everywhere, and they also form a basis for the solution space:

$$f(s) = g(s) = AG_0(\beta, s) + BG_1(\beta, s) \tag{2.3.10}$$

We can expand a few of them to get an idea of what they look like:

$$G_0(\beta, s) = c_0(\beta s^2) = \begin{cases} \cos(s\sqrt{\beta}) & \beta > 0 \\ \cosh(s\sqrt{-\beta}) & \beta < 0 \\ 1 & \beta = 0 \end{cases} \tag{2.3.11}$$

$$G_1(\beta, s) = sc_1(\beta s^2) = \begin{cases} \sin(s\sqrt{\beta})/\sqrt{\beta} & \beta > 0 \\ \sinh(s\sqrt{-\beta})/\sqrt{-\beta} & \beta < 0 \\ s & \beta = 0 \end{cases} \tag{2.3.12}$$

$$G_2(\beta, s) = s^2 c_2(\beta s^2) = \begin{cases} (1 - \cos(s\sqrt{\beta}))/\beta & \beta > 0 \\ (1 - \cosh(s\sqrt{-\beta}))/\beta & \beta < 0 \\ s^2/2 & \beta = 0 \end{cases} \tag{2.3.13}$$

The G_k have some nice properties, which will be useful in later derivations.

$$G_k(\beta, s) = \frac{s^k}{k!} - \beta G_{k+2}(\beta, s) \qquad \frac{\mathrm{d}}{\mathrm{d}s} G_{k+1} = G_k \qquad G_k(\beta, 0) = 0 \text{ for } k > 0 \qquad G_0(0) = 1$$

One special case to consider is $\frac{\mathrm{d}}{\mathrm{d}s} G_0$. Applying the first property, then the second, we get that:

$$\frac{\mathrm{d}}{\mathrm{d}s} G_0 = \frac{\mathrm{d}}{\mathrm{d}s}(1 - \beta G_2) = -\beta G_1$$

TODO separate section on computing these, where we expand, talk about catastrophic cancellation, and chebyshev

Aside: This may seem like sweeping the casework under the rug, but it isn't. The reason that this is acceptable is that $G_k(\beta, s)$ is an analytic function in ***both*** *variables, and varying β lets us smoothly transition between the cases. The casework here is only about how we* ***choose to express it*** *in elementary functions.*

2.4 Solving for Initial Conditions

Now that we have the solution set for the associated homogeneous diffeq of (2.2.2), we need to find some solution to the inhomogeneous equation. Conveniently, constants will work here:

$$f(s) = 1 - \frac{\mu}{\beta r_0} \qquad g(s) = \frac{r_0 \dot{r}_0}{\beta} \tag{2.4.1}$$

So, our full solutions for f and g are of the form:

$$f(s) = \left(1 - \frac{\mu}{\beta r_0}\right) + AG_0(\beta, s) + BG_1(\beta, s) \tag{2.4.2}$$

$$g(s) = \frac{r_0 \dot{r}_0}{\beta} + CG_0(\beta, s) + DG_1(\beta, s) \tag{2.4.3}$$

for some constants A, B, C, and D, determined by the initial conditions.

Since we defined s to be 0 at the initial time, and $\boldsymbol{x} = f(s)\boldsymbol{x}_0 + g(s)\boldsymbol{v}_0$, we get the initial conditions in terms of f and g:

$$f(0) = 1 \qquad g(0) = 0 \qquad \dot{f}(0) = 0 \qquad \dot{g}(0) = 1 \tag{2.4.4}$$

Plugging in $s = 0$ to equations (2.4.2) and (2.4.3), and seeing that G_1 vanishes, we can get values for A and C. To get the other two, we need to look at the time derivatives (not the s-derivatives!) of those equations:

$$\dot{f}(s) = \frac{1}{r}f'(s) = -\frac{\beta A}{r}G_1(\beta, s) + \frac{B}{r}G_0(\beta, s) \tag{2.4.5}$$

$$\dot{g}(s) = \frac{1}{r}g'(s) = -\frac{\beta C}{r}G_0(\beta, s) + \frac{D}{r}G_0(\beta, s) \tag{2.4.6}$$

Now we can plug in $s = 0$ to get the remaining two constants. The resulting values are:

$$A = \frac{\mu}{\beta r_0} \qquad B = 0 \qquad C = -\frac{r_0 \dot{r}_0}{\beta} \qquad D = r_0 \tag{2.4.7}$$

Plugging back into (2.4.2) and (2.4.3) gives us our complete solutions for f and g.

$$f(s) = \left(1 - \frac{\mu}{\beta r_0}\right) + \frac{\mu}{\beta r_0}G_0(\beta, s) \tag{2.4.8}$$

$$g(s) = \frac{r_0 \dot{r}_0}{\beta} - \frac{r_0 \dot{r}_0}{\beta}G_0(\beta, s) + r_0 G_1(\beta, s) \tag{2.4.9}$$

These solutions have a β in some denominators, and that could prove troublesome when $\beta = 0$. Fortunately, substituting $G_0 = 1 - \beta G_2$ lets us eliminate these denominators:

$$f(s) = 1 - \frac{\mu}{r_0}G_2(\beta, s) \qquad g(s) = r_0 G_1(\beta, s) + r_0 \dot{r}_0 G_2(\beta, s) \tag{2.4.10}$$

Together the equations in (2.4.10) give us the position for any value of s. To get the velocity, we take the time derivative:

$$\dot{f}(s) = -\frac{\mu}{r r_0}G_1(\beta, s) \qquad \dot{g}(s) = \frac{r_0}{r}G_0(\beta, s) + \frac{r_0}{r}\dot{r}_0 G_1(\beta, s) \tag{2.4.11}$$

At this point, we can now compute $\boldsymbol{x}$ and $\boldsymbol{v}$ at any given s. If we want a physical simulation, we next have to connect s to t.

2.5 Connecting Time and s

Given a value of t, how do we convert it to a value of s? Recall the defining relation $\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{1}{r}$. Rearranging, we get that $\mathrm{d}t = r\ \mathrm{d}s$, and if we know r in terms of s, we can integrate.

We could try to use the relation $r = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$, but that square root gets really nasty. Instead, start with the equation $r' = r\dot{r} = \boldsymbol{x} \cdot \dot{\boldsymbol{x}}$, and take the time derivative of both sides.

$$\frac{\mathrm{d}}{\mathrm{d}t}r' = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{x} \cdot \dot{\boldsymbol{x}}) = \boldsymbol{x} \cdot \ddot{\boldsymbol{x}} + \dot{\boldsymbol{x}} \cdot \dot{\boldsymbol{x}}$$

Then apply the equations of motion, and the definition of energy (2.1.4), and grind a little bit of algebra:

$$\frac{\mathrm{d}}{\mathrm{d}t}r' = \boldsymbol{x}\left(-\frac{\mu}{r^3}\boldsymbol{x}\right) + 2\left(\varepsilon + \frac{\mu}{r}\right) = -\frac{\mu}{r} - \beta + 2\frac{\mu}{r} = \frac{\mu}{r} - \beta$$

Rearranging one more time, we get a diffeq describing r:

$$r'' + \beta r = \mu \tag{2.5.1}$$

We've already seen how to solve diffeqs of this form; we know that the solution for $r(s)$ looks like $r(s) = A + BG_1(\beta, s) + CG_2(\beta, s)$. (We could use G_0 and G_1 here instead, but if we did, we'd have to clear β from the denominator, just like before.)

Evaluating this and its s-derivatives at $s = 0$, it's straightforward to find these constants:

$$A = r(0) = r_0 \qquad B = r'(0) = r_0 \dot{r}_0 \qquad C = r''(0) = \mu - \beta r_0 \tag{2.5.2}$$

This gives us a clean equation for $r(s)$:

$$r(s) = r_0 G_0(\beta, s) + r_0 \dot{r}_0 G_1(\beta, s) + \mu G_2(\beta, s) \tag{2.5.3}$$

Integrating this with respect to s bumps up the subscript on the G_i, giving us the desired link between s and t:

$$t = \int r(s)\ \mathrm{d}s = r_0 G_1(\beta, s) + r_0 \dot{r}_0 G_2(\beta, s) + \mu G_3(\beta, s) + C \tag{2.5.4}$$

Since we pegged s to zero at our initial conditions, we have that $C = t_0$, or the time at start. This gives our equation for time:

$$t - t_0 = r_0 G_1(\beta, s) + r_0 \dot{r}_0 G_2(\beta, s) + \mu G_3(\beta, s) \tag{2.5.5}$$

So if we want to advance the system by some s, we can update position and velocity using (2.4.10) and (2.4.11), and update time using (2.5.5).

As a side note, we can use the expression for r to simplify the $\dot{g}$ expression in (2.4.11):

$$\dot{g}(s) = 1 - \frac{\mu}{r} G_2(\beta, s) \tag{2.5.6}$$

(TODO: can i do something similar for g?)

2.6 Anchoring to Zero

Given an initial position and velocity, we now know how to advance the state forward or backward by any s. This gives us a parameterization of the orbit, however, depending on where on the orbit we start out, we get a different parameterization. This is somewhat unsatisfying – what is there a way we can pick a canonical one?

Much like with propagation, the classical approach involves treating parabolic, elliptic, and hyperbolic orbits differently, but using s will allow us to treat them in a uniform manner.

A natural point, present on all orbits, is the periapsis. Other than on a circular orbit, this point is unique, so we'll use it for the initial conditions, at $s = 0$, and describe all other points on the orbit from there. Plugging in the radius at periapsis (r_p) into (2.4.10), and noting that velocity is perpendicular to the radius there, we get:

$$f(s) = 1 - \frac{\mu}{r_p} G_2(\beta, s) \qquad g(s) = r_p G_1(\beta, s) \tag{2.6.1}$$

This looks pretty good, but there's a small problem we have to worry about: collision orbits. For those orbits, $r_p = 0$ and $v_p = \infty$, and conversely, $f(s) = \infty$ and $g(s) = 0$. Fortunately, since $\boldsymbol{x}(s) = f(s)\boldsymbol{x}_p + g(s)\boldsymbol{v}_p$, these 0s and ∞s are paired up and resolve to something finite, so this isn't an irresolvable problem, but it does mean that we should not use r_p and v_p themselves as the parameterizing constants.

To get around this, we'll normalize our coordinate axes. Let's work in a coordinate system where x points towards periapsis, and y points along the velocity vector. Then the x coordinate is just $r_p f(s)$, and the y coordinate is $v_p g(s)$. If we expand the definitions of those, and make use of some of the classic orbital elements (ε, h, and e), then we arrive at coherent expressions for all orbits:

$$x = r_p f(s) = r_p - \mu G_2(\beta, s) = \frac{h^2}{\mu(1+e)} - \mu G_2(\beta, s) \tag{2.6.2}$$

$$y = v_p g(s) = r_p v_p G_1(\beta, s) = h G_1(\beta, s) \tag{2.6.3}$$

Following the same process on equation (2.4.11) (or just taking the derivative of the equations above), we get expressions for velocity:

$$\dot{x} = r_p \dot{f}(s) = -\frac{\mu}{r} G_1(\beta, s) \tag{2.6.4}$$

$$\dot{y} = v_p \dot{g}(s) = \frac{r_p v_p}{r} G_0(\beta, s) = \frac{h}{r} G_0(\beta, s) \tag{2.6.5}$$

For time, the coordinates don't matter, so we can just plug in the state at periapsis into equation (2.5.5):

$$t(s) - t_0 = r_p G_1(\beta, s) + \mu G_3(\beta, s) \tag{2.6.6}$$

By using the recurrence relation on G_k, we can simplify this a little further:

$$t(s) - t_0 = r_p G_1(\beta, s) + \mu G_3(\beta, s) = r_p s + (\mu - r_p \beta) G_3(\beta, s) = r_p s + \mu e G_3(\beta, s) \tag{2.6.7}$$

Taking the s derivative, we get an equation for the radius too:

$$r(s) = r_p + \mu e G_2(\beta, s) \tag{2.6.8}$$

TODO make sure you use these in the actual code!

2.7 s as Universal Anomaly

We've gotten a lot of mileage out of s, but what exactly is it? It describes how far along an orbit a satellite is, and can be used to propagate the orbit, so in that sense, it's very similar to time. But there's another variable it's even more closely related to: the eccentric anomaly.

TODO use χ to mean s since periapsis?

Elliptical Orbits

For elliptical orbits, the *true anomaly* is the actual angle that a satellite makes with its focal axis, and is denoted by θ. The *mean anomaly* M is an averaged measure of motion; if t is the time elapsed since periapsis, and T is the period of the orbit, $M = 2\pi t/T$.

The *eccentric anomaly* is somewhat of a hybrid of these two quantities. The important things for us to know are Kepler's equation, which relates it to the mean anomaly, and a relation to the true-anomaly:

$$M = E - e \sin E \tag{2.7.1}$$

$$\tan \frac{E}{2} = \sqrt{\frac{1-e}{1+e}} \tan \frac{\theta}{2} \tag{2.7.2}$$

To connect s to E, we will integrate $\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{1}{r}$, and miraculously, the r and t quantities will coincide into a single expression in E.

First, we expand r using the polar form of an ellipse:

$$\Delta s = \int \frac{\mathrm{d}t}{r} = \int \frac{1 + e\cos\theta}{a(1-e^2)} \, \mathrm{d}t \tag{2.7.3}$$

Taking the derivative of (2.7.1), we can switch our integration variable to E:

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \frac{2\pi}{T} = \frac{\mathrm{d}E}{\mathrm{d}t} - e\cos E \frac{\mathrm{d}E}{\mathrm{d}t}$$

$$\frac{\mathrm{d}t}{\mathrm{d}E} = \frac{1 - e\cos E}{2\pi/T} \tag{2.7.4}$$

Plugging it into (2.7.3), we get an integrand without time:

$$\Delta s = \frac{T/2\pi}{a(1-e^2)} \int (1 + e\cos\theta)(1 - e\cos E) \, \mathrm{d}E \tag{2.7.5}$$

Lastly, we need to express $\cos\theta$ in terms of E. By applying several different half-angle formulas to (2.7.2), and grinding away, we can get

$$\cos\theta = \frac{\cos E - e}{1 - e\cos E} \tag{2.7.6}$$

which, when plugged back into the previous integral, gives

$$\Delta s = \frac{T/2\pi}{a(1-e^2)} \int (1-e^2) \, \mathrm{d}E = \frac{T}{2\pi a} \Delta E = \frac{\Delta E}{\sqrt{\beta}} \tag{2.7.7}$$

where the last equality comes from Kepler's third law.

Surprisingly, s is just a rescaled version of the eccentric anomaly! Since $s = 0$ and $E = 0$ at the periapsis, we can drop the Δ and simply state that

$$s = \frac{E}{\sqrt{\beta}} \tag{2.7.8}$$

Non-Elliptical Orbits

Similarly, hyperbolic orbits have the hyperbolic anomaly, denoted H, and related to the true anomaly by:

$$\tanh\frac{H}{2} = \sqrt{\frac{e-1}{e+1}}\tan\frac{\theta}{2} \tag{2.7.9}$$

If you do the same kind of integral, you get a similar relationship between hyperbolic anomaly and s:

$$s = \frac{H}{\sqrt{-\beta}} \tag{2.7.10}$$

Lastly, we look at the parabola. The polar form of a parabola is $r = \frac{h^2/\mu}{1+\cos\theta}$, and instead of Kepler's equation, we use Barker's equation:

$$t = \frac{h^3}{2\mu^2}(D + \frac{1}{3}D^3) \tag{2.7.11}$$

where $D = \tan\frac{\theta}{2}$ and h is the angular momentum.

Taking the derivative of (2.7.11), we can change our integral to be over D:

$$s = \frac{h}{2\mu}\int (1+\cos\theta)(1+D^2)\;\mathrm{d}D \tag{2.7.12}$$

Through double-angle identities, we rephrase $\cos\theta$ to be $(1-D^2)/(1+D^2)$, allowing us to simplify the integrand to a constant 2, and so

$$s = \frac{h}{\mu}D \tag{2.7.13}$$

In summary, s is a rescaled version of the eccentric, hyperbolic, and parabolic anomalies, and through these, can be related to the true anomaly.

3 Orbit Representations in Code

The space of orbits is five-dimensional, so we need five parameters to represent it. We reduce the problem first by specifying a rotation (3 degrees of freedom), that takes the x-axis to the focal axis, and the z axis to the orbital plane. Choosing the remaining two parameters is tricky.

There's a wide set of possible options we can pick from:

- semimajor axis
- eccentricity
- energy
- periapsis
- semilatus rectum

If we pick semimajor axis and eccentricity, then all parabolic orbits look the same: $a = \infty$ and $e = 1$. Even if we include energy, we have the same problem. The only things that distinguish parabolic orbits from each other are the periapsis and the semilatus rectum.

If we pick the periapsis, we need to pick energy or semimajor axis as the other parameter, since we need a way to distinguish radial orbits. The problem is that not all pairs of values are valid orbits. For a fixed periapsis r_p, the orbit with the least energy is a circular one. Therefore, we must have $\varepsilon \geq -\frac{\mu}{2r_p}$, which is a bit of an awkward constraint to have.

TODO i ended up choosing ℓ and $1/a$, but why can't I do r_p and a?

4 SOI Encounter and Escape

Kerbal Space Program simulates an n-body trajectory by patching together arcs from different one-body trajectories. When do we decide that the primary body is the sun, and when is it one of the planets? (Equivalently, when near a planet, do we use the planet or a moon?)

KSP does this by using "spheres of influence". When a satellite is in a body's sphere of influence (SOI), then KSP uses that body as the primary of the orbit. The SOIs form a hierarchy, where the SOI of each moon nests within the SOI of its planet, and the SOI of each of those nests within the SOI of the Sun (infinite).

When a body of mass m orbits a body of mass M, then KSP estimates the radius of its SOI as

$$r_{SOI} = a\left(\frac{m}{M}\right)^{2/5} \tag{4.0.1}$$

When planning a spacecraft's trajectory, we are nearly always interested in a path that crosses multiple spheres of influence. Therefore, we need to figure out when exactly we enter and exit the SOIs in our neighborhood.

4.1 Escape

Determining whether a ship ever escapes an SOI is straightforward: does the ship ever reach a radius of r_{SOI}?

Since we have formulas for the apoapsis, we can just compute it, and see if it's larger than r_{SOI}. If so, then we escape.

However, for accurate simulation, we also have to figure out when we escaped. To do that, we use equation (2.5.3), which relates r to s. With some root-finding techniques, we can determine this s, and then plug it into equation (2.5.5) to get the elapsed time.

4.2 Encounter

For SOI encounter, things are hairier. Not only do we have to hit a moving target, instead of a fixed radius, there are multiple possible moons we can encounter, and we have to check for encounter with each of them separately.

TODO note: haven't actually done this yet but it seems promising!

To do this, we do what mathematicians call "interval analysis", and programmers call "bounding box computation". For a given time interval, we determine the possible range of x coordinates we could have, and same for y and z. This gives us a box that we know the ship is contained in. If we repeat the process for the moon, we have two boxes, and it's easy to test whether they intersect.

If they don't intersect, then we know that there is no possibility of encounter. If they do intersect, there could be an encounter, but not necessarily. We split the interval in half, and repeat the test. If we don't intersect in the first half, then we know the possible intersection lies in the second half, and repeat the process, until we TODO well, when do we stop? when we have one foot in and one foot out? that might not suffice to get the *first* intersection though

Bounding Boxes

To compute this bounding box, we need to know what kind of range the x coordinate could take on in an arbitrary time interval. First, we find the corresponding interval in s, using (2.5.5). Then, we could take equation (2.6.1), and dot it against the $\hat{x}$ unit vector.

However, for computational purposes, it will turn out to be easier to transform the basis vectors into the frame of the orbit, and then do all our computation in that frame. So instead, let's dot it against an arbitrary unit vector $\hat{u} = (u_x, u_y, u_z)$:

$$\boldsymbol{x} \cdot \hat{u} = (r_p - \mu G_2(\beta, s))u_x + hG_1(\beta, s)u_y \tag{4.2.1}$$

Our goal is to find the range of this expression over an arbitrary interval in s. It suffices to find out what range the G_k functions have over a particular interval. If G_k were monotonic, this would be easy, since we can just take the range between the endpoints. Instead, we have to find the intervals on which G_k *is* monotonic, and account for the maxima and minima in between.

Fortunately, we can exploit the similarity to trigonometric functions. Revisiting the expansion of G_1 (2.3.12), we split into cases based on β. When $\beta < 0$, it looks like $\sinh(-)$, so there are no extrema then. When $\beta = 0$, it looks like s, so there are no extrema then either. When $\beta < 0$, the function equals $\sin(s\sqrt{\beta})/\sqrt{\beta}$. In that regime, there is a maximum at every $\left(2\pi + \frac{1}{2}\right)\pi/\sqrt{\beta}$, with output $1/\sqrt{\beta}$, and its minima are similar.

For G_2, we do something similar. When $\beta < 0$, it looks like $(1 - \cosh(s\sqrt{-\beta}))/\beta$. There's only one extremum, a minimum, at the origin. When $\beta = 0$, we have the same situation. When $\beta > 0$, it looks like $(1 - \cos(s\sqrt{\beta}))/\beta$, and we're back to periodicity. The minima are at $2n\pi/\beta$, and take a value of 0; the maxima occur at $(2n+1)\pi/\beta$, and take the value $2/\beta$.

So, if we're evaluating our position over the interval $[s_0, s_1]$, then we check whether the interval contains an extremum. If not, we just take the range over the endpoints. Otherwise, we expand the range to include the extremum. Repeat this for each coordinate, and we get our bounding box!

5 Frame Transformation

Consider two orthogonal reference frames, A and B, which are displaced by a vector T_{BA} and related by a rotation R_{BA}. Additionally, frame B is moving w.r.t. frame A, with linear velocity V_{BA} and angular velocity (around its origin) Ω_{BA}.

TODO picture

5.1 Positions and Directions

Even though the two frames are not aligned, and some rotation R_{BA} is needed to convert between the coordinates, we're not going to really worry about it in this document. Vectors are physical objects, and have coordinate-independent existences, and so we will write them without specifying a basis.

For positions, we'll specify them as vector offsets from the origin of the frame. So, given a point, we can measure both r_A, the displacement from the origin of A, and r_B, the displacement from the origin of B. They are related by $r_A = r_B + T_{BA}$.

Of course, in the code itself, we'll need to make use of R_{BA} to convert between coordinates. Specifically, if R_{BA} actively transforms the basis vectors of A to those of B, it passively converts coordinates from B to A. As an example, the above identity, when expressed in A's coordinates, is $[r_A]_A = R_{BA}[r_B]_B + [T_{BA}]_A$, where $[-]_A$ is "this vector is expressed in A's coordinates". But we don't have to worry about that except in code.

5.2 Derivatives

These two frames move relative to one another, and so they will disagree on whether objects are moving or not. Let D_A mean "derivative in frame A", and likewise for D_B. However, they will agree on scalar quantities, and so for those we'll just use the usual d/dt.

First, we'd like to determine how the bases move. Let the basis vectors in frame A be $\hat{x}$, $\hat{y}$, and $\hat{z}$, and the ones in frame B be $\hat{X}$, $\hat{Y}$, and $\hat{Z}$. We know from basic geometry that if a vector u is fixed in frame B, then $D_A(u) = \Omega_{BA} \times u$. So in particular this is true of the basis vectors.

Now consider some vector, possibly time-dependent, $u(t)$. In frame A, it has coordinates $(x, y, z)_A$, where $v = x\hat{x} + y\hat{y} + z\hat{z}$, and in the B frame, it has similar coordinates $(X, Y, Z)_B$.

Let's take the derivative in the A frame. On one hand, it's simply $\frac{\mathrm{d}x}{\mathrm{d}t}\hat{x} + \frac{\mathrm{d}y}{\mathrm{d}t}\hat{y} + \frac{\mathrm{d}z}{\mathrm{d}t}\hat{z}$. But if we expand it in B-coordinates, we get:

$$\begin{aligned}
D_A(u) &= D_A(X\hat{X} + Y\hat{Y} + Z\hat{Z}) \\
&= \frac{\mathrm{d}X}{\mathrm{d}t}\hat{X} + XD_A(\hat{X}) + \frac{\mathrm{d}Y}{\mathrm{d}t}\hat{Y} + YD_A(\hat{X}) + \frac{\mathrm{d}Z}{\mathrm{d}t}\hat{Z} + ZD_A(\hat{Z}) \\
&= \left(\frac{\mathrm{d}X}{\mathrm{d}t}\hat{X} + \frac{\mathrm{d}Y}{\mathrm{d}t}\hat{Y} + \frac{\mathrm{d}Z}{\mathrm{d}t}\hat{Z}\right) \\
&\qquad + \Big(X(\Omega_{BA} \times \hat{X}) + Y(\Omega_{BA} \times \hat{Y}) + Z(\Omega_{BA} \times \hat{Z})\Big) \\
&= D_B(u) + (\Omega_{BA} \times u)
\end{aligned}$$

TODO figure out how to number only the last of these

5.3 Velocities

Now that we know how derivatives work, we're equipped to talk about velocities. As mentioned above, if an object has position r_A in frame A, and r_B in frame B, then the two vectors are related by $r_A = r_B + T_{BA}$. Taking the derivative in the A frame, we have:

$$D_A(r_A) = D_A(r_B) + D_A(T_{BA}) = D_B(r_B) + (\Omega_{BA} \times r_B) + D_A(T_{BA})$$

The $D_A(r_A)$ and $D_B(r_B)$ terms are the component-wise derivatives of position in their respective frames. By definition, these are the velocities v_A and v_B. The last term is the velocity (as measured by A) of B's origin, in other words, V_{BA}. (Note that there is an asymmetry here; $D_B(T_{BA})$ is *not* V_{BA}, because there's extra velocity that A's origin acquires from B's rotation!) So our final result is:

$$v_A = v_B + (\Omega_{BA} \times r_B) + V_{BA}$$

5.4 Composition and Inverses

Let's now try to figure out the inverses by flipping some labels. Since $r_A = r_B + T_{BA}$, it should be the case that $r_B = r_A + T_{AB}$, which would imply that $T_{AB} = -T_{BA}$. Similarly, $D_A(u) = D_B(u) + \Omega_{BA} \times u$ implies that $\Omega_{AB} = \Omega_{BA}$.

The case for velocity is a little trickier, but nothing horrendous. Starting by flipping the labels, and isolating V_{AB}:

$$\begin{aligned}
v_B &= v_A + (\Omega_{AB} \times r_A) + V_{AB} \\
V_{AB} &= v_B - v_A - (\Omega_{AB} \times r_A) \\
V_{AB} &= -v_A + v_B + (\Omega_{BA} \times (r_B + T_{BA})) \\
V_{AB} &= (v_B + (\Omega_{BA} \times r_B)) - v_A + (\Omega_{BA} \times T_{BA}) \\
V_{AB} &= -V_{BA} + (\Omega_{BA} \times T_{BA})
\end{aligned}$$

Physically, what this means is that the velocity of A's origin, w.r.t. B, is the negative of the velocity we saw before (as expected), but there's also an additional term that comes from the rotation Ω.

For what it's worth, we could derive this a second way. We know that $D_A(T_{BA}) = V_{BA}$, and so we can use what we already know about derivatives:

$$V_{AB} = D_B(T_{AB}) = -D_B(T_{BA}) = -D_A(T_{BA}) - (\Omega_{AB} \times T_{BA})$$

$$= -V_{BA} + (\Omega_{BA} \times T_{BA})$$

Composition employs some similar tricks. Clearly $T_{CA} = T_{CB} + T_{BA}$ and with some thought, one can justify $\Omega_{CA} = \Omega_{CB} + \Omega BA$. As usual, it's the velocity that needs some work.

We can use the fact that $V_{CA} = D_A(T_{CA})$ to derive the correct velocity:

$$\begin{aligned} V_{CA} &= D_A(T_{CA}) \\ &= D_A(T_{CB}) + D_A(T_{BA}) \\ &= D_B(T_{CB}) + (\Omega_{BA} \times T_{CB}) + D_A(T_{BA}) \\ &= V_{CB} + V_{BA} + (\Omega_{BA} \times T_{CB}) \end{aligned}$$

Another way to read this result is as the conversion of the velocity of C's origin from B's frame to A's frame. (Take the usual velocity transformation from B to A, but with $v_B = V_{CB}$ and $r_B = T_{CB}$).

6 Root-Finding Methods

7 Computation of Stumpff Functions

TODO catastrophic cancellation and chebyshev