Solar sail dynamics in the three-body problem: homoclinic paths of points and orbits

Thomas J. Waters and Colin R. McInnes

Department of Mechanical Engineering, University of Strathclyde, Glasgow

Abstract

In this paper we consider the orbital dynamics of a solar sail in the Earth-Sun circular restricted 3-body problem. The equations of motion of the sail are given by a set of nonlinear autonomous ordinary differential equations, which are non-conservative due to the non-central nature of the force on the sail. We consider first the equilibria and linearisation of the system, then examine the nonlinear system paying particular attention to its periodic solutions and invariant manifolds. Interestingly, we find there are equilibria admitting homoclinic paths where the stable and unstable invariant manifolds are identical. What’s more, we find that periodic orbits about these equilibria also admit homoclinic paths; in fact the entire unstable invariant manifold winds off the periodic orbit, only to wind back onto it in the future. This unexpected result shows that periodic orbits may inherit the homoclinic nature of the point about which they are described.

1 Introduction

A solar sail is a novel type of spacecraft which uses the radiation pressure of photons reflecting off large sails as its impulse (see McInnes [13] for a detailed description). A natural setting for the solar sail is the circular restricted 3-body problem (CR3BP) where the Earth and the Sun are the primary bodies. This is partly because the 3-body problem more accurately describes solar system dynamics than the 2-body problem, but also because in the 3-body problem there are regions where the gravitational forces on the sail due to the primaries cancel each other, and hence the radiation pressure force on the sail plays a more dominant role. Also, the demands on sail efficiency would be less as the gravitational forces are less, and thus the applications of this analysis are more in the near-term.

A standard procedure in analysing a nonlinear system of ode’s is to find its equilibria or fixed points, linearise about these, and use the information from linear order to inform an analysis of the nonlinear system; this is the procedure.
we will follow in this paper. In particular we will direct our attention to the nonlinear system’s periodic orbits and invariant manifolds. There has been some work carried out on the dynamics of solar sails in the CR3BP. McInnes et al. [14] first described the surfaces of equilibrium points, and some possible uses of same. In Baoyin and McInnes [1] and McInnes [12], the authors describe periodic orbits about equilibrium points in the solar sail three body problem, however they consider only equilibrium points on the axis joining the primary masses, corresponding to artificial Lagrange points. Such orbits are analogous to the classical ‘halo’ orbits (where by classical we mean the particle is only acted upon by gravitational forces), which are well documented, for example Farquhar [4], Farquhar and Kamel [5], Breakwell and Brown [3], Richardson [15], Howell [9] and Thurman and Worfolk [18]. With regards to the invariant manifolds, there has been much analysis of the invariant manifolds of halo orbits in the classical problem for the sake of efficient transfer; for example the Genesis mission trajectory was designed using this technique (see Koon et al. [10]), and a ‘petit grand tour’ of Jovian moons has been proposed using a similar analysis (see Gómez et al. [6]). Some homoclinic paths for the classical triangular points have been found for large mass ratios (see for example Gómez et al. [7]), and homoclinic paths can exist for collinear points with particular mass ratios (see §9.9.2 of Szebehely [16]). Certain isolated homoclinic paths have been found for periodic orbits about the collinear Lagrange points (see for example Koon et al. [11]), however no periodic orbit whose invariant manifold is made up entirely of homoclinic paths has been found, to the best of our knowledge.

The structure of the paper is as follows: in the next section we will describe the setting of the problem and the equations of motion of the solar sail, as well as the equilibrium points. Section 3 considers the system linearised about equilibrium and the form of the linear solutions. In Section 4 we briefly describe the Lindstedt-Poincaré perturbation method used to find nonlinear approximations to periodic orbits, and in Section 5 we examine the invariant manifolds of equilibria and periodic orbits. We find a large variety in the position, inclination, amplitude and frequency of periodic solutions to the equations of motion, and unexpected homoclinic paths associated with equilibria and periodic orbits which have no analogue in the classical problem. Such results suggest that the solar sail CR3BP presents a rich and complex model, the intricate details of which are only beginning to become apparent.

2 Equations of motion in the rotating frame

We follow the conventions set out in McInnes [13]. We consider a rotating coordinate system in which the primary masses are fixed on the $x$-axis with the origin at the centre of mass, the $z$-axis is the axis of rotation and the $y$-
axis completes the triad. We chose our units to set the gravitational constant, the sum of the primary masses, the distance between the primaries, and the magnitude of the angular velocity of the rotating frame to be unity. We shall denote by \( \mu = 3 \times 10^{-6} \) the dimensionless mass of the smaller body \( m_2 \), the Earth, and therefore the mass of the larger body \( m_1 \), the Sun, is given by \( 1 - \mu \) (see Figure 1).

Denoting by \( r, r_1 \) and \( r_2 \) the position of the sail w.r.t. the origin, \( m_1 \) and \( m_2 \) respectively, the solar sail’s equations of motion in the rotating frame are

\[
\frac{d^2 r}{dt^2} + 2\omega \times \frac{dr}{dt} = a - \omega \times (\omega \times r) - \nabla V = F, \tag{1}
\]

with \( \omega = \hat{z} \) and \( V = -[(1 - \mu)/r_1 + \mu/r_2] \) where \( r_i = |r_i| \). These differ from the classical equations of motion in the CR3BP by the radiation pressure acceleration term

\[
a = \beta \frac{(1 - \mu)}{r_1^2} (\hat{r}_1, n)^2 n, \tag{2}
\]

where \( \beta \) is the sail lightness number, and is the ratio of the solar radiation pressure acceleration to the solar gravitational acceleration. Here \( n \) is the unit normal of the sail and describes the sail’s orientation. We define \( n \) in terms of two angles \( \gamma \) and \( \phi \) w.r.t. the rotating coordinate frame,

\[
n = (\cos(\gamma) \cos(\phi), \cos(\gamma) \sin(\phi), \sin(\gamma)), \tag{3}
\]

where \( \gamma, \phi \) are the angles the normal makes with the \( x-y \) and \( x-z \) plane respectively (see Figure 1).

Equilibria are given by the zeroes of \( F \) in (1). We find a 3-parameter family of equilibria, as described in McInnes [13]. These are found by specifying the lightness number \( \beta \) and the sail angles \( \gamma \) and \( \phi \), and solving \( F = 0 \) for one of

![Fig. 1. The rotating coordinate frame and the sail position therein. The angles \( \gamma \) and \( \phi \) which the sail normal makes with respect to the rotating frame are also shown.](image)
Fig. 2. Surfaces of equilibrium points in the $x_e$-$z_e$ parameter space. Each curve is specified by a constant value of $\beta$, and the position of the equilibrium point along the curve is given by $\gamma$. The grey shaded regions denote areas where equilibrium is not possible.

two possible equilibria, one on the $L_1$ side and another on the $L_2$ side of the Earth. For reasons we shall describe in the next section we will take $\phi = 0$ so the equilibrium (and sail normal) is in the $x$-$z$ plane. In Figure 2 we show some of the equilibria near the Earth for low $\beta$ values. Practically speaking, while a $\beta$ value of about $0.3 - 0.4$ is considered within the realm of possibility of current engineering, to put the analysis in this paper well within the near-term we will consider very modest $\beta$ values of about 0.05.

3 Linearised system

We linearise about the equilibrium point (in the $x$-$z$ plane) by making the transformation $r \rightarrow r_e + \delta r$, Taylor expanding $F$ about $r_e$, and neglecting the terms quadratic in $\delta r$. We assume the orientation of the sail will remain fixed under perturbation of the sail position, in which case $\gamma, \phi$ and $\beta$ are constants. Letting $\delta r = (\delta x, \delta y, \delta z)^T$ and $X = (\delta r, \delta \dot{r})^T$, our linear system is $\dot{X} = AX$ with

$$A = \begin{pmatrix} 0 & I \\ \Omega & M \end{pmatrix}, \quad M = \begin{pmatrix} a & 0 & b \\ 0 & c & 0 \\ d & 0 & e \end{pmatrix}, \quad \Omega = \begin{pmatrix} 0 & 2 & 0 \\ -2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (4)$$
where a dot denotes differentiation w.r.t. \( t \),
\[
a = (\partial_x F^x)|_e, \quad b = (\partial_y F^y)|_e, \quad c = (\partial_y F^y)|_e, \quad d = (\partial_z F^z)|_e, \quad e = (\partial_z F^z)|_e,
\]
and \( b \neq d \). Here \( F^a \) denotes the \( a \)-th component of \( F \), and \( M \) is sparse due to \( y_e = 0 \).

The key difference between this analysis and the classical orbits about the collinear Lagrange points is the term \( d \neq 0 \), which appears precisely because we are linearising about an equilibrium point with \( z_e \neq 0 \). This means we cannot decouple the \( z \)-equation. While this initially seems to make the problem more complicated, we can use this to our advantage, as will be made clear.

The characteristic equation of the Jacobian \( A \) is bi-cubic (whose corresponding cubic equation has real roots); this means the eigenvalues of \( A \) are either in pairs of pure imaginary conjugates or real and of opposite sign. Thus equilibria in the \( x-z \) plane will have the dynamical structure of centres and saddles, akin to the classical collinear Lagrange points; in fact we find the linear spectrum to be typically
\[
\{\lambda_1i, -\lambda_1i, \lambda_2i, -\lambda_2i, \lambda_r, -\lambda_r\}.
\]
(5)

Had we chosen \( \phi \neq 0 \) for an equilibrium out of the \( x-z \) plane, then \( M \) given in (4) would be full and the characteristic equation would not have been bi-cubic; thus the dynamical structure of equilibria out of the \( x-z \) plane will be stable/unstable spirals and saddles. In fact, equilibria in the \( x-y \) plane (\( \gamma = 0, \phi \neq 0 \)) will be stable spirals crossed with saddles for \( \phi < 0 \ (y_e < 0) \) and unstable spirals crossed with saddles for \( \phi > 0 \ (y_e > 0) \). In this sense the \( x \)-axis is a bifurcation surface.

If we label the eigenvectors associated with \( \lambda_a \) \( (a = 1, 2) \) as \( u_a + w_a i \), and the eigenvectors associated with \( \lambda_r, -\lambda_r \) as \( v_1, v_2 \), then the general solution of the linear system (4) is [2]
\[
X(t) = \cos(\lambda_1 t) \left[ A u_1 + B w_1 \right] + \sin(\lambda_1 t) \left[ B u_1 - A w_1 \right] + \cos(\lambda_2 t) \left[ C u_2 + D w_2 \right] + \sin(\lambda_2 t) \left[ D u_2 - C w_2 \right] + E e^{\lambda_r t} v_1 + F e^{-\lambda_r t} v_2.
\]
(6)

We see that due to the coupling of the \( z \)-equation in the linear system, which in turn is due to our choice of \( z_e \neq 0 \), the linear order solution naturally contains periodic solutions in both linear frequencies. By setting \( E = F = 0 \) we may switch off the real modes, and by setting either \( A = B = 0 \) or \( C = D = 0 \) we have periodic solutions in the frequency of our choice.

We note that we may also consider combinations of both frequencies in the linear solutions, a generalisation of a family of classical curves known as ‘roses’.
While these trajectories will not in general close, by choosing an equilibrium point whose linear frequencies are in ratio we may find multiply periodic orbits, however this raises some stability issues. We will not pursue such orbits in this paper.

4 High-order approximations to periodic orbits

The linear solutions given in the previous section will only closely approximate the motion of the sail given in (1) for small amplitudes. For larger amplitude periodic orbits, we have two options: nonlinear analytical approximations or numerical continuation of initial data. While numerical continuation is the most straightforward and can be continued beyond the region where the linear term dominates, it is a slow process and needs to done for each equilibrium point of the 2-parameter family separately. The nonlinear analytical approximations will quickly provide converging initial data for large amplitude orbits at arbitrary equilibria, and so is to be preferred in the region where the linear term dominates. To calculate high-order approximations to periodic orbits we use the method of Linstedt-Poincaré. This procedure is well known and is described in the literature, for example [15,18], thus we will only outline the relevant issues for this model.

We let $\varepsilon$ be a perturbation parameter and expand each coordinate as $x \to x_e + \varepsilon x_1 + \varepsilon^2 x_2 + \ldots$ etc. We rescale the time coordinate $\tau = \omega t$ with $\omega = 1 + \varepsilon \omega_1 + \ldots$, and group together the powers of $\varepsilon$ in the high-order Taylor expansion of $F$. We choose our linear solution to be

$$x_1 = k A_y \cos(\lambda \tau + \xi), \quad y_1 = A_y \sin(\lambda \tau + \xi), \quad z_1 = m A_y \cos(\lambda \tau + \xi),$$

(7)

where $\lambda$ can be $\lambda_1$ or $\lambda_2$, $k, m$ are given in terms of components of the eigenvectors and $A_y, \xi$ are free parameters (we have set the $y$-amplitude to be the free parameter to avoid singularities in the coefficients when the orbits pass through vertical or horizontal). We use these linear solutions to build up nonlinear approximations to periodic orbits one order at a time in the following way:

At each order of $\varepsilon$, the system to be solved will be

$$x_n'' - 2y_n' - ax_n - bz_n = g_1(x_{n-1}, y_{n-1}, z_{n-1}, x_{n-2}, \ldots)$$

$$y_n'' + 2x_n' - cy_n = g_2(x_{n-1}, y_{n-1}, z_{n-1}, x_{n-2}, \ldots)$$

$$z_n'' - dx_n - ez_n = g_3(x_{n-1}, y_{n-1}, z_{n-1}, x_{n-2}, \ldots),$$

(8)

where prime denotes differentiation w.r.t. $\tau$. The left hand side is the same form as the linear system (4), and on the right hand side the previous orders'
solutions act as forcing terms. We use the freedom in $\omega_n$ to switch off the resonant or secular terms in the inhomogeneous part, that is those components on the right hand side of the form (7), and what remains is a series of trigonometric subharmonics up to order $n$.

In calculating the solution at $n$th order, we find two sets of solutions depending on whether $n$ is even or odd. When $n$ is even, the $n$th order solutions have the form (letting $T = \lambda \tau + \xi$)

$$
x_n = p_{n0} + p_{n2} \cos(2T) + \ldots + p_{nn} \cos(nT),
$$

$$
y_n = q_{n2} \sin(2T) + \ldots + q_{nn} \sin(nT),
$$

$$
z_n = s_{n0} + s_{n2} \cos(2T) + \ldots + s_{nn} \cos(nT),
$$

(9)

with $\omega_{n-1} = 0$. When $n$ is odd, the solutions at $n$th order have the form

$$
x_n = p_{n3} \cos(3T) + \ldots + p_{nn} \cos(nT),
$$

$$
y_n = q_{n1} \sin(T) + q_{n3} \sin(3T) + \ldots + q_{nn} \sin(nT),
$$

$$
z_n = s_{n1} \cos(T) + s_{n3} \cos(3T) + \ldots + s_{nn} \cos(nT),
$$

(10)

and $\omega_{n-1}$ solves

$$
\frac{2\lambda \beta_{n1}}{(e + \lambda^2)} + \frac{b \gamma_{n1}}{(e + \lambda^2)} - \alpha_{n1} = 0.
$$

(11)

Here $\alpha_{nj}, \beta_{nj}$ and $\gamma_{nj}$ are the coefficients of the $\cos, \sin$ and $\cos$ terms in the functions $g_1, g_2$ and $g_3$ respectively at order $n$ given in (8), and the coefficients $p_{nj}, q_{nj}$ and $s_{nj}$ are given by

$$
-(a + j^2 \lambda^2) p_{nj} - 2j \lambda q_{nj} - bs_{nj} - \alpha_{nj} = 0,
$$

$$
q_{nj} = \frac{-2j \lambda p_{nj} - \beta_{nj}}{(e + j^2 \lambda^2)}, \quad s_{nj} = \frac{-dp_{nj} - \gamma_{nj}}{(e + j^2 \lambda^2)},
$$

(12)

with the exception of $q_{n0} = 0$ and $p_{n1} = 0$.

With these high order approximations, we may find approximate initial data from which to integrate the system of equations (1). However these will not evolve to exactly periodic trajectories, as they are only approximations to periodic solutions. Thus we must define a differential corrector with which to adjust the initial data so as to close the orbit. As differential correctors are discussed in much detail in the literature, particularly relating to the classical halo orbits (see for example [3,15,18]), we will not describe them here.

With the high order approximations and the differential corrector described above, we may integrate the full nonlinear system of equations (1) to find large amplitude periodic orbits. We find there is a 4-parameter family of periodic orbits about equilibria in the $x$-$z$ plane: two parameters specifying the equilibrium point and two specifying the amplitude in each of the linear frequencies.
Fig. 3. A family of orbits with $\beta = 0.05$. Each orbit has the same amplitude and is about a different equilibrium point along the $\beta$ level curve shown in Figure 2, each equilibrium point being defined by a different $\gamma$ value. For reference the Earth (to scale) and $L_1$ are shown.

There is therefore much variety in the position, inclination, amplitude and period of the orbits we may describe, leading to a rich source of periodic orbits for various applications. To illustrate this, we may hold the amplitude fixed and allow $\gamma$ to vary, thus forming a tube of periodic orbits along the $\beta$ level curve. An example of this for one family of periodic orbits is shown in Figure 3.

5 Homoclinic paths

As a practical consideration it is useful to examine the transfer of a solar sail from Earth orbit to an equilibrium point. This is most easily accomplished by considering the equilibrium point’s invariant manifolds.

Letting $r = (x, y, z)^T$ and $X = (r, \dot{r})$, we write the system (1) as $\dot{X} = f(X)$ with equilibrium given by $f(X_e) = 0$. Let us denote by $E^u, E^s$ and $E^c$ the linear eigenspaces spanned by the unstable, stable and centre eigenvectors of the equilibrium point, that is those due to eigenvalues with positive, negative and zero real part respectively, and denoted $v_1, v_2$ and $u_a + i w_a$ in (6). From (5) we see that these eigenspaces are typically of dimension 1,1 and 4 respectively.
We will define the invariant manifolds $W^u, W^s$ and $W^c$ as those families of solutions to the non-linear system (1) which are tangential to $E^u, E^s$ and $E^c$ at the equilibrium point, with the property that

$$X(t_0) \in W^* \Rightarrow X(t) \in W^* \forall t,$$

that is, the manifolds are invariant under the flow.

These manifolds are of great practical importance as they characterise the dynamics of (1) away from the equilibrium point. For example, the centre manifold describes the surface on which periodic solutions exist, and the Lindstedt-Poincaré method can be seen as calculating high-order approximations to trajectories on the centre manifold. The stable and unstable manifolds characterise the flow onto and away from the equilibrium point, in particular the unstable mode dominating the flow and pushing solutions away from equilibrium in the direction of the unstable manifold.

The true stable and unstable manifolds tend onto and away from the equilibrium point asymptotically. To avoid having to integrate the equations of motion for infinite lengths of time we must therefore approximate the invariant manifolds by perturbing the initial data a small amount away from the equilibrium point in the direction of the linear eigenvectors, and then integrate the system in this direction forward and backward in time. If we find a portion of the approximation to the stable manifold passes close to the Earth then this provides us with a potential efficient trajectory on which to transfer from the Earth to the equilibrium point.

Interestingly however, we find that there are certain equilibria on the $L_2$ side of the Earth which admit homoclinic paths; we define a homoclinic path as a phase path which joins an equilibrium point to itself. This means the stable and unstable invariant manifolds intersect smoothly and thus are identical. Equilibrium points admitting homoclinic paths are not to be confused with homoclinic points, which are points (other than the equilibrium point) where the stable and unstable invariant manifolds intersect transversally. Homoclinic paths occur for specific parameter values and are not structurally stable (in fact they represent a homoclinic bifurcation), unlike homoclinic points. See, for example, Tabor [17] for an introduction to homoclinic points/paths.

We present some of these points in Figure 4 for different parameter values. As mentioned above, homoclinic paths have the equilibrium point as an asymptotic limit as $t \to \pm \infty$; in fact a homoclinic path can be seen as a periodic orbit of infinite period (hence the old-fashioned term ‘asymptotically periodic’). The trajectories presented here are therefore approximations to homoclinic paths and so they will not settle exactly on the equilibrium point; as the sail approaches the equilibrium the unstable mode will push the sail away.
We may also define the invariant manifolds of periodic orbits about equilibria in a manner analogous to above, using Floquet theory (see Grimshaw [8] for example). Let \( \Gamma(t) \) denote a periodic solution to (1), with period \( T \). By letting \( X = \Gamma + Y \), we may linearise the nonlinear system about this periodic solution, resulting in the variational equations

\[
\dot{Y} = \left. \frac{\partial f}{\partial X} \right|_{X=\Gamma} Y \equiv A(t) Y,
\]

(13)

where \( A(t+T) = A(t) \). This is a non-autonomous linear system with periodic coefficients. A well known result of Floquet theory is that for every fundamental solution matrix \( Y(t) \) of a system such as (13), there is a non-singular constant matrix \( B \) such that

\[
Y(T) = Y(0) B.
\]

(14)

Therefore the eigenvalues of \( B \) tell us about the linear orbital stability of the periodic orbit. Recasting the variational equations in terms of the state transition matrix (or principal fundamental matrix) \( \Phi = \partial X / \partial X(0) \), we have

\[
\dot{\Phi} = A(t) \Phi, \quad \Phi(0) = I,
\]

and thus we associate the matrix \( B \) given in (14) with \( \Phi(T) \), the monodromy matrix.

As the divergence of our original system vanishes, that is the trace of the Jacobian \( \sum \partial f^i / \partial X_i = 0 \) (see (4)), Liouville’s theorem (constancy of volume in phase space) applies. Thus we may say the following: the eigenvalues of the monodromy matrix occur in reciprocal pairs, two of them are unity, and stability of the periodic orbit is given by eigenvalues within the unit circle in the complex plane. As the saddle nature of the equilibrium point about which the periodic orbit is described will dominate the flow in the region in which the linear terms dominate (and thus the region where the Lindstedt-Poincaré
approximations are valid), it follows that all of the periodic orbits described in this paper will be unstable, that is all monodromy matrices will contain two real reciprocal eigenvalues one of which will be larger than one. Complex conjugate eigenvalues must be on the unit circle and this represents a marginal stability, akin to eigenvalues of the Jacobian on the imaginary axis. Thus we find the spectrum of the monodromy matrix of the periodic orbits described in this paper to be

\[ \{1, 1, \lambda_i, \bar{\lambda}_i, \lambda_r, 1/\lambda_r \}, \]  

where an overbar denotes complex conjugacy.

Analogous to equilibrium points, we may integrate points along the periodic orbit in the direction of the eigenvectors of the monodromy matrix and find the trajectories which wind onto and off of the periodic orbit, that is the stable and unstable invariant manifolds of the periodic orbit. We find that the periodic orbits inherit the homoclinic characteristics of the equilibrium point about which they are described, and we show in Figure 5 an example of this for one of the equilibrium points in Figure 4. We see that every trajectory in the invariant manifold of the periodic orbit is itself a homoclinic path, and this persists for varying amplitudes of the periodic orbit.

6 Conclusions

In this paper we have presented an initial analysis of the periodic orbits available to solar sails in the circular restricted three-body problem (CR3BP). We find there is a four parameter family of periodic orbits about equilibria in the \(x-z\) plane (two due to choice of equilibrium point and two due to amplitudes in the linear frequencies). We have further examined the invariant manifolds of equilibria and periodic orbits, and have found unexpected families of homoclinic paths. These homoclinic paths suggest the model contains an interesting degree of complexity, perhaps akin to the Rössler attractor. There, homoclinic paths can be seen as a limit of multiply periodic orbits and this suggests that the solar sail model will contain many families of periodic orbits which pass close to the equilibrium point. This limit-type structure is seen also in the triangular points of the classical three-body problem. A more detailed examination of the homoclinic and indeed heteroclinic paths contained in the problem should therefore be of interest.

We conclude that the the dynamics associated with solar sails in the CR3BP present a rich and complex topic for analysis, and this paper represents only a first step in unraveling the many details of this model.
Fig. 5. A triply-periodic homoclinic invariant manifold of a periodic orbit about the equilibrium point with parameter values $\beta = 0.05, \gamma = 0.7, \phi = 0$ (on the right in Figure 4). To give a sense of the flow the shading of the trajectory fades to grey with time.

References


