Analytical solution of the Colombo top problem

J. Haponiak · S. Breiter · D. Vokrouhlický

1 Introduction

Only about 60% of the Moon surface can be seen from the Earth. The first successful attempt to explain this fact was made by Cassini (1693), who borrowed the kinematic model of the 'triple Earth motion' from Copernicus (1543, Lib. 1, Cap. 11), and applied it to the Moon with some important amendments. Retaining the postulate of the fixed angle between the rotation axis and the orbital plane, Cassini postulated the equality of orbital period and the sidereal rotation period, which protected the far side from being seen from the Earth. By

S. Breiter
Astronomical Observatory Institute, Faculty of Physics, Adam Mickiewicz University, Śloneczna 36, 61-286 Poznan, Poland
E-mail: breiter@amu.edu.pl

J. Haponiak
Astronomical Observatory Institute, Faculty of Physics, Adam Mickiewicz University, Śloneczna 36, 61-286 Poznan, Poland
E-mail: jacek.haponiak@amu.edu.pl

D. Vokrouhlický
Institute of Astronomy, Charles University, Prague, V Holešovičkách 2, 180 00 Prague 8, Czech Republic
E-mail: vokrouhl@cesnet.cz
additionally postulating that the rotation axis, the ecliptic pole, and the lunar orbit pole remain coplanar, Cassini suppressed the possibility of revealing the complete polar caps over one lunar axis precession cycle. Two centuries later, Tisserand (1891) rephrased these postulates as laws. The three Cassini laws of Tisserand state that: 1) rotation and orbital periods are equal, 2) the rotation axis has a constant inclination to the ecliptic, and 3) the three axes are coplanar. The second law differs from the original Cassini’s statement, but in view of the third law, the difference is unimportant.

With the advent of the Newtonian dynamics, the question arose if the Cassini’s model is consistent with equations of motion. This was a part of the problem issued by the French Académie Royale des Sciences for the Prize of 1764. In his prize dissertation and in a later work, Lagrange (1764, 1780) demonstrated that the state described by Cassini is an equilibrium of the associated differential system, and studied small librations in its vicinity.

The work of Colombo (1966) brought a new understanding of the Cassini laws and motion near the stationary configuration which they describe in the specific case of the Moon. In particular, Colombo demonstrated that the second and the third laws are conceptually independent from the first law and themselves serve as a basis of an interesting dynamical problem which describes the long-term evolution of the spin axis of an arbitrary rotating body. He thus dropped from his analysis the assumption of the direct spin-orbit resonance, but kept the assumption of general precession of the orbital plane due to perturbations (either caused by the oblate central body, or by other masses in the system). He showed that the long-term dynamics of the spin axis can be described by a simple, one dimensional problem assuming the orbital node performs a uniform precession and the inclination remains constant. Its stationary points represent generalizations of the Cassini second and third laws.

It was soon understood that the Colombo problem is a very suitable starting point for analysis of the obliquity evolution of terrestrial planets, even when the orbital node and inclination undergo more complex evolution. A fascinating example are studies of Mars obliquity variations in relation to this planet’s past paleoclimate, starting with Ward (1973, 1974). Tides or internal process may additionally change some of the system’s parameters, a situation relevant to all terrestrial planet studies, including the Moon – see Peale (1974), Ward (1975, 1982) or Ward and de Campli (1979), to mention just few examples of early works. Later studies of Laskar and colleagues made a masterful use of detailed knowledge of planetary long-term dynamics and its implications on secular evolution of their spin axes (e.g. Laskar and Robutel, 1993; Laskar et al., 1993; Correia and Laskar, 2001, and many other with more technical details). Following earlier hints, mentioned already in Harris and Ward (1982), applications to giant planets were also developed in the past two decades (e.g. Ward and Hamilton, 2004; Hamilton and Ward, 2004; Ward and Canup, 2006; Boué et al., 2009; Vokrouhlický and Nesvorný, 2015; Brasser and Les, 2015; Rogoszinski and Hamilton, 2020).

Beyond planets and satellites, studies of secular spin evolution of asteroids flourished recently, especially after Vokrouhlický et al. (2003) applied it to explain space parallelism of spin axes of large members in the Koronis family (see also Vokrouhlický et al. (2006)). Further applications include spin states of exoplanets (e.g. Atobe et al. 2004; Atobe and Ida, 2007; Saillenfest et al. 2019), or artificial satellites and space debris (e.g. Efimov et al. 2018).

Taken altogether, we note that the backbone of all these studies is the basic Colombo model. Interestingly, a systematic mathematical treatment of this elegant Hamiltonian problem has not been significantly advanced beyond the state of art dating back to Henrard and Murigande (1987), and Henrard (1993). These classical works focused on the aspects directly related with the probability of capture into different phase space zones when a slow parameter evolution drives the system across the homoclinic orbits, hence they paid no attention to regular
orbits. One can also recognize the Colombo top in an anonymous quadratic Hamiltonian treated by Lanchares and Elipe (1995), who focused on the qualitative study of its parametric bifurcations.

The main reason to seek for the complete analytical solution of the Colombo top is its significance for further studies of more realistic, perturbed problems. Be it analytical perturbation techniques, or numerical splitting methods, the knowledge of explicit time dependence of the Colombo top motion is crucial. The present work is divided in two principal parts: Sect. 2 explores the problem using purely geometric and algebraic tools, whereas the integration of equations of motion is considered in Sect. 3. In other words, Sect. 2 concerns integral curves, that become time-dependent trajectories in Sect. 3.

The formulation the problem is given in Sect. 2.1, where we introduce two sets of variables: traditional \( x, y, z \), and shifted \( X, Y, Z \). Throughout the text we switch between the two sets, depending on convenience. In Sect. 2.2 the geometric construction of the integral curves is shown; intersections of the curves with the \( x = X = 0 \) meridians (their extremities in \( z \)) are found in Sect. 2.3, expressed in terms of the energy constant and of the two parameters \( a, b \). Some of these can be critical points (the Cassini states), which allows the expression in terms of the parameters only – given in Sect. 2.4. The information gained allows to partition the phase space and distinguish three types of the Colombo top problem.

In Sect. 3 we first provide a universal solution in terms of the Weierstrass elliptic function \( \wp \) (Sect. 3.1), comparing various formulations of the same result. But since the Weierstrass function behaves differently in various domains of the phase space, we reformulate the solution in terms of the more common Jacobi elliptic functions (Sect. 3.2). Finally, the specific trajectories that admit solutions in terms of elementary functions are presented in Sect. 3.3. The closing Sect. 4 summarizes the results and their implications.

2 The Colombo top model

2.1 Equations of motion

The basic assumptions leading to the Colombo top problem involve a rigid body on an orbit around some distant primary. The orbital motion might be called Keplerian, but with one notable addition: the orbital plane rotates uniformly around some fixed axis in the inertial space with the angular rate \( \mu \). The body is assumed to rotate in the lowest energy state, namely about the shortest principal axis of its inertia tensor. Following the work of Colombo (1966), we assume the the rotation period is not in resonance with the revolution period about the primary (see, e.g., Peale 1969 for generalizations to this situation). Considering the quadrupole torque due to the gravitational field of the primary, and averaging over both orbital revolution about the center and rotation cycle about the spin axis, one easily realizes that the total angular momentum of rotation (or, equivalently, the angular velocity \( \omega \) of rotation) is conserved. The whole dynamical problem then reduces to the analysis of the motion of the unit vector \( r \) of rotation pole in space.

In order to introduce fundamental astronomical parameters of relevance, let us for a moment assume \( \mu = 0 \), i.e. the orbital plane is fixed in the inertial space. In this simple case, \( r \) performs regular precession about the fixed direction \( K \) of the orbital angular momentum with a frequency

\[
\mu_r = -\alpha \cos \varepsilon, \tag{1}
\]

where \( \alpha \) is the precession constant and \( \varepsilon \) is the obliquity, namely the angle between \( r \) and \( K \) \((\cos \varepsilon = r \cdot K)\). Note the minus sign in Eq. (1) which indicates polar regression in the inertial
space. In this model, $\varepsilon$ stays constant and

$$\alpha = \frac{3}{2} \frac{n^2}{\omega} \frac{E_b}{(1 - e^2)^{3/2}},$$

(2)

where $n$ is the orbital mean motion about the primary, $\omega$ the angular rotational frequency, $E_b$ the dynamical ellipticity and $e$ the orbital eccentricity. The dynamical ellipticity expresses degree of non-sphericity of the body and it is defined using principal moments $A \leq B \leq C$ of the inertia tensor as

$$E_b = \frac{C - (A + B)/2}{C}.$$  

(3)

Things become more interesting in the case where the orbital plane about the primary is not constant. As mentioned above, the Colombo top model describes the situation when it performs a uniform precession in the inertial space. In particular, $K$ revolves uniformly on a cone about a fixed direction $K'$ in space, such that the orbital inclination $I$ with respect to the reference plane normal to $K'$ is constant. The magnitude of the precession rate is $\mu$, though most often the orbital plane performs again regression in the inertial space (assuming $I < 90^\circ$). The interest and complexity of this model revolves about a possibility of a resonance between the two precession frequencies $-\mu$ and $\mu_r$. In order to describe it using a simple Hamiltonian model, Colombo (1966) observed it is useful to refer to the reference frame following precession of the orbital plane, thus representing $r = (\sin \varepsilon \cos (h - \pi/2), \sin \varepsilon \sin (h - \pi/2), \cos \varepsilon)$. Here, $h - \pi/2$ is a longitude reckoned from the ascending node of the orbital plane and $\varepsilon$ is a colatitude measured from $K$ as above. It is actually an advantage to introduce $H = \cos \varepsilon$ (and $\sin \varepsilon = \sqrt{1 - H^2}$). This is because in terms of the symplectic variables $(h, H)$, the Colombo top is a one degree of freedom, conservative problem with a Hamiltonian function

$$\mathcal{H}_c(h, H) = -\frac{(H - b)^2}{2} - a \sqrt{1 - H^2} \cos h = E_H,$$

(4)

and two new nondimensional parameters are defined as

$$a = \frac{\mu}{\alpha} \sin I, \quad b = \frac{\mu}{\alpha} \cos I.$$  

(5)

As observed by Henrard and Murigande (1987), the discussion can be confined to non-negative constants $a$ and $b$ thanks to the symmetries $(a, h, H) \rightarrow (-a, h + \pi, H)$, and $(b, h, H) \rightarrow (-b, -h, -H)$ admitted by $\mathcal{H}_c$.

Since the equations of motion

$$\dot{h} = \frac{\partial \mathcal{H}_c}{\partial H} = b - H + \frac{aH \cos h}{\sqrt{1 - H^2}}, \quad \dot{H} = -a \frac{\partial \mathcal{H}_c}{\partial h} = -a \sqrt{1 - H^2} \sin h,$$

(6)

are singular at $H^2 = 1$, it is better to use the Cartesian coordinates of the unit momentum vector

$$r = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sqrt{1 - H^2} \sin h \\ -\sqrt{1 - H^2} \cos h \\ H \end{pmatrix}.$$  

(7)

Then, similarly to Henrard and Murigande (1987) we obtain the Hamiltonian function

$$\mathcal{H}_{BM}(x, y, z) = -\frac{(z - b)^2}{2} + ay = E_H,$$

(8)
that generates equations of motion
\[
\dot{x} = (z - b)(y + a) + ab, \\
\dot{y} = -(z - b)x, \\
\dot{z} = -ax.
\] (9)

For the sake of minor simplification, let us propose the shifted variables
\[
X = x, \quad Y = y + a, \quad Z = z - b, \quad \mathbf{R} = (X, Y, Z)^T = \mathbf{r} + (0, a, -b)^T.
\] (10)

Adding a constant \(a^2\) to the Hamiltonian (5) we simplify it to
\[
\mathcal{H} = -\frac{Z^2}{2} + aY = E,
\] (11)
with the new energy constant
\[
E = E_H + a^2.
\] (12)

Equations of motion for the shifted variables are
\[
\dot{X} = ZY + ab, \\
\dot{Y} = -ZX, \\
\dot{Z} = -aX.
\] (13)

Notably, when \(a = 0\), the problem is simplified to the symmetric free top, with constant \(Z\) and uniform rotation of \(\mathbf{r}\) around the third axis, with the frequency \(\sqrt{-2E} = |Z|\).

2.2 Geometric interpretation

It is customary to represent the integral curves of Eqs. (9) as the intersections of two surfaces:

- **S1** – a sphere \(x^2 + y^2 + z^2 = 1\),
- **S2** – a parabolic cylinder \((z - b)^2 - 2a(y + a) + 2E = 0\), implied by the energy integral.

The symmetry plane of S2 is \(z = b\), and its vertex line is parallel to the \(x\)-axis, passing through \(y = (E - a^2)/a\).

Observe that combining S1 and S2 we can also obtain another surface (see Fig. 1):

**S3** – a paraboloid of revolution \(x^2 + (y + a)^2 + 2bz - (2E - a^2 + b^2 + 1) = 0\).

Actually, the S3-based function
\[
\mathcal{H}_a(\mathbf{r}) = \frac{x^2}{2} + \frac{(y + a)^2}{2} + bz = E' = E + \frac{1}{2} - a^2 + \frac{b^2}{2},
\] (14)

is an alternative Hamiltonian of the Colombo top, leading to the same equations of motion (9) as the Hamiltonian (8).

The paraboloid S3 has the symmetry axis parallel to the \(z\)-axis, and passing through the points \(x = 0, y = -a\). Its vertex is located at \(z = E'/b\). The advantage of S3 appears when discussing the limit of \(a = 0\). Then, the paraboloid does not change the shape and the intersections of S1 and S3 are circles because of coincidence of the symmetry axes. Contrarily to this, setting \(a = 0\) in S2 results in degeneracy: the cylinder breaks in two parallel planes. On the other hand, \(b = 0\) turns the S3 paraboloid into a (circular) cylinder, whereas S2 retains...
its shape. Thus, S2 and S3 (or $\mathcal{H}$ and $\mathcal{H}_a$) can be considered complementary, although the regularity at $a = 0$ seems to us more favorable (e.g. if some perturbation approach is based upon the small inclination assumption).

An integral curve can be represented as a parametric curve in a number of ways. Of course, the best is to solve Eqs. (9) and find $r(t)$. But before we accomplish it (and, actually, in order to do it) let us consider parameterizations, where two coordinates are expressed in terms of the third. The most straightforward is to solve the system of S1 and S2 equations, using $z$ as a parameter variable, which leads to

$$r(z) = \left( \pm \sqrt{\frac{1 - z^2 - y(z)^2}{(z-b)^2 + 2E - 2a^2}} \right) \frac{z}{z},$$

or

$$R(Z) = \left( \pm \sqrt{\frac{1 - (Z + b)^2 - (Y(Z) - a)^2}{Z^2 + 2E}} \right) \frac{Z}{Z}.$$  \hspace{1cm} (15)

All three invariant surfaces S1, S2, S3 intersect the plane $x = 0$, which is their common plane of symmetry. The integral curves, as the lines of intersection of S1, S2, and S3 also pass through $x = 0$ and are symmetric with respect to this plane. Moreover, both $y$ and $z$ coordinates of a given integral curve attain their local extremes at $x = 0$, according to equations of motion (13).

2.3 Intersections with the plane $x = 0$

To find the intersection points of an integral curve with the plane $x = 0$ for some specified energy $E$, it is enough to find either $y$ or $z$ coordinate. Knowing one of them, one might recover the other from the relation $y^2 + z^2 = 1$. But this involves the ambiguity of sign, thus it is better to find $z$, and then use the parametric equation (15) for $y(z)$. To benefit from minor simplifications, we first find $Z$ and then $Y(Z)$.

According to Eq. (15), the relation between $X$ and $Z$ is

$$X = \pm \frac{1}{2a} \sqrt{W(Z)},$$  \hspace{1cm} (16)
where $W(Z)$ is a polynomial of degree 4

$$W(Z) = -Z^4 - 4EZ^2 - 8a^2bZ - 4(E - a^2)^2 + 4a^2(1 - b^2).$$ (17)

Note the absence of the cubic term thanks to the use of the shifted variable $Z$.

Solving the quartic equation $W(Z) = 0$ is a tedious task; the details can be found in Appendix A. Briefly, the four roots $Z_j$ are given in terms of the three roots $e_i$ of the reduced cubic resolvent equation (additionally modified to match the standard cubic polynomial appearing in the Weierstrass elliptic integrals). Depending on the parameters $a$ and $b$, selecting some energy value $E$, the number of meridian intersection points is 4, 3, 2, 1, or 0.

Four intersection points mean that there are two distinct integral curves with the same energy, each intersecting $x = 0$ in two points:

$$z_1 = b - \sqrt{e_1 - \frac{2}{3}E} - \sqrt{e_2 - \frac{2}{3}E} - \sqrt{e_3 - \frac{2}{3}E},$$
$$z_2 = b - \sqrt{e_1 - \frac{2}{3}E} + \sqrt{e_2 - \frac{2}{3}E} + \sqrt{e_3 - \frac{2}{3}E},$$ (18)

for a lower curve, and

$$z_3 = b + \sqrt{e_1 - \frac{2}{3}E} - \sqrt{e_2 - \frac{2}{3}E} + \sqrt{e_3 - \frac{2}{3}E},$$
$$z_4 = b + \sqrt{e_1 - \frac{2}{3}E} + \sqrt{e_2 - \frac{2}{3}E} - \sqrt{e_3 - \frac{2}{3}E},$$ (19)

for the upper curve, with $e_j$ defined in Eq. (141) as functions of $E$, $a$, and $b$, such that $z_1 < z_2 < z_3 < z_4$.

Three intersection points mean that either one of the two curves contracts to a single point, or two curves share the same intersection point. These situations are distinguished by the sign of invariant $g_3$ from Eq. (135).

– If $g_3 > 0$, then the upper curve shrinks into a point with

$$z_{34} = b + \sqrt{e_1 - \frac{2}{3}E}. \quad \text{(20)}$$

The remaining intersection points of the lower curve are

$$z_1 = 2b - z_{34} - 2\sqrt{e_{23} - \frac{2}{3}E}, \quad z_2 = 2b - z_{34} + 2\sqrt{e_{23} - \frac{2}{3}E}. \quad \text{(21)}$$

The expressions of $e_1$ and $e_{23}$ are given in Eq. (145).

– If $g_3 < 0$, then the upper and the lower curves meet at

$$z_{23} = b + \sqrt{e_3 - \frac{2}{3}E}, \quad \text{(22)}$$

with the remaining intersection points at

$$z_1 = 2b - z_{23} - 2\sqrt{e_{23} - \frac{2}{3}E}, \quad z_4 = 2b - z_{23} + 2\sqrt{e_{23} - \frac{2}{3}E}. \quad \text{(23)}$$
If \( e_1 = e_2 = e_3 \), then \( z_2 \) from Eq. (21) becomes equal to \( z_{34} \) and only two intersection points remain

\[
\begin{align*}
  z_1 &= b - 3\sqrt{a^2 b}, \\
  z_{234} &= b + \sqrt{a^2 b},
\end{align*}
\] (24)

with \( a^2 + b^2 = 1 \). This case requires a unique value of energy with a simple expression (154).

A more generic situation with two intersection points occurs when \( E \) defines only one trajectory, with \( z_1 < z_2 \) given by

\[
\begin{align*}
  z_1 &= b - \sqrt{e_1 - \frac{2}{3}E - \sqrt{e_1 - 2E + \sqrt{\nu}}}, \\
  z_2 &= b - \sqrt{e_1 - \frac{2}{3}E + \sqrt{e_1 - 2E + \sqrt{\nu}}}, \\
  \nu &= \left( e_1 + \frac{4}{3}E \right)^2 + e_2^2,
\end{align*}
\] (25)

where \( e_1 \) and \( e_2 \) are defined in Eqs. (142) and (144). Two other roots of \( W(z - b) = 0 \), i.e. \( z_3 = Z_3 + b \) and \( z_4 = Z_3 \) are complex, so they do not define the intersections – see Eqs. (151).

Finally, one intersection point means that for the energy \( E \) there is only one integral curve that contracted to a point with

\[
\begin{align*}
  z_{12} &= b - \sqrt{e_1 - \frac{2}{3}E},
\end{align*}
\] (26)

and \( e_1 \) given by Eq. (145).

In further discussion, the point of intersection of a regular curve with the \( x = 0 \) plane will be named a turning point, unless it is an equilibrium from the dynamical point of view.

2.4 Cassini states and homoclinic orbits

Finding the Cassini states can be approached from different points of view. Geometrically, they are the points of tangency of the surfaces \( S_1, S_2, \) and \( S_3 \). Algebraically, they are the multiple roots of \( x(z, E) = 0 \), and \( x(y, E) = 0 \). Dynamically, they are the fixed points of equations of motion (9) or, equivalently, the local extremes and saddle points of the Hamiltonian on a sphere \( S_1 \).

In principle, the multiple roots have been found in Section 2.3. But they are given in terms of \( E \), which is an implicit function of \( a \) and \( b \) as a root of \( \Delta(E, a, b) = 0 \), where \( \Delta \) is the discriminant of \( W(Z) \), defined in Eq. (136). Hence, one possible way is to find the roots of \( \Delta = 0 \), which is a quartic equation in \( E \), and substitute them into \( z_{12}, z_{23}, \) or \( z_{34} \) from Sect. 2.3. We already did it for \( z_{234} \). The alternative is to find the stationary points \( (x, y, z) \) of the Hamiltonian and use them to evaluate the energy by the substitution into \( H(x, y, z) = E \). The latter path is more convenient and was recently taken by Saillenfest et al. (2019).

Discussing the Cassini states as the multiple roots of \( W(Z) = 0 \), we should include the information, that they are also the roots of its derivative \( W'(Z) = 0 \). Since two polynomials have a common root if and only if their resultant equals zero, we ask about the solutions of \( R(W(Z), W'(Z)) = 0 \). If we treat \( W(Z) \) and \( W'(Z) \) as polynomials in \( Z \), the result is simply the discriminant \( \Delta \) from Eq. (156), multiplied by a constant (negative) factor. Yet, we can
also treat \( W(Z) \) and \( W'(Z) \) as polynomials of \( E \) with degrees 2 and 1, respectively, whose coefficients depend on \( Z \). So, using

\[
W(Z) = U_1(E) = -4E^2 + 4(2a^2 - Z^2)E - Z^4 - 4a^2(2bZ + a^2 + b^2 - 1), \\
W'(Z) = U_2(E) = -8ZE - 4Z^2 - 8a^2b,
\]

we obtain the resultant \( R(U_1(E), U_2(E)) = 256a^2w(Z) \), where

\[
w(z) = -Z^4 - 2bZ^3 + 3pZ^2 - 2a^2bZ - a^3b^2,
\]

and

\[
\rho = \frac{1 - a^2 - b^2}{3}.
\]

For easier reference to Saillenfest et al\( (2019) \), we abandon the shifted variables and solve \( w(z - b) = 0 \), which is the same equation as the one they used. However, thanks to considering the Cartesian variables on a sphere, we do not have to warn about any loss of states \( C \), the problem on a sphere, there are only two generic situations: either there are two Cassini states \( C_2 \) and \( C_3 \), or there are four of them: \( C_1, C_2, C_3 \) and \( C_4 \). Let us call the former case ‘the Colombo top problem of type II’, and the latter - ‘type IV’. The special case when there are three Cassini states will be called ‘the Colombo top problem of type III’ (see Fig. 2).

In the following subsections we briefly review each type, providing the coordinates of the Cassini states \( C_j \) as the functions of \( a \) and \( b \) parameters. To avoid confusion with the \( z \) expressions of Sect. 2.3, we label the coordinates of \( C_j \) as \( z_j^* \), and \( y_j^* \) (of course, \( x_j^* = 0 \) for all the points). We only provide \( z_j^* \) that allows to find \( y_j^* \) from Eq. (31) and then the energy \( E_j \) of the Cassini state is

\[
E_j = -\frac{(z_j^* - b)^2}{2} - \frac{a^3b}{z_j^* - b}.
\]

Substituting this energy value into an appropriate \( z_1, z_2, z_3, \) or \( z_4 \) from Sect. 2.3, should result in returning to the expression of \( z_j^* \) in terms of \( a \) and \( b \).

### 2.4.1 Type II

When \( a^2 + b^2 > 1 \), the dynamics on a sphere is relatively simple. There are two stable equilibria (elliptic points): \( C_3 \) at the lower (\( z < 0 \)) hemisphere and \( C_2 \) at the upper hemisphere, located at

\[
z_3^* = \frac{b}{2} - \sqrt{A_2 - \sqrt{B_2}}, \quad z_2^* = \frac{b}{2} - \sqrt{A_2 + \sqrt{B_2}},
\]

where

\[
A_2 = -P_f^2 + \frac{1}{2} \left( \sqrt{g_3^* + \sqrt{-\Delta'}} + 3\sqrt{g_3^* - \sqrt{-\Delta'}} \right), \\
B_2 = -3P_f^2 - A_2 + 2\sqrt{A_2^2 + 3P_f^2 (P_f^2 + A_2)} - \frac{3p^2}{16},
\]

with

\[
p = \frac{1 - a^2 - b^2}{3},
\]

and \( g_3^* = \frac{1}{2} \left( \frac{3b - a^2 + b^2 - 1}{(3a^2 + b^2 - 1)^{1/2}} + \frac{3b + a^2 + b^2 - 1}{(3a^2 + b^2 - 1)^{1/2}} \right) \).
Three types of the Colombo top established by the values of $a$ and $b$, shown at the $x \geq 0$ hemisphere of $S_1$. The discriminant $\Delta$ \eqref{eq:136} is positive in the shaded area, negative in white zones, and null on the thick curves $\Gamma$ or at the Cassini states $C$. Exemplary parameters values are: $a = b = 0.5$ (type II), $a = (3/4)^{3/2}, b = (7/16)^{3/2}$ (type III), and $a = b = 0.2$ (type IV).

and, according to Eqs. \eqref{eq:164} and \eqref{eq:163}

\[
\sqrt{-\Delta'} = \frac{ab}{4} \sqrt{a^2 b^2 - \rho^3}, \quad g'_3 = \frac{a^2 b^2}{4} - \left(\frac{\rho}{2}\right)^3, \quad P'_2 = -\frac{\rho}{2} - \frac{b^2}{4}.
\]

We can identify the Cassini states coordinates as $z_3^* = z_{12}(E_3)$, and $z_2^* = z_{12}(E_2)$, where $z_{12}$ is given by Eq. \eqref{eq:26}.

The energy values in the type II problem are bounded by $E_3 \leq E \leq E_2$. All trajectories with energy $E_3 < E < E_2$ are simple periodic curves (one for each value of $E$), oscillating in $z$ between $z_1$ and $z_2$, as given by Eqs. \eqref{eq:25}. Extending and modifying the domains labeling of \cite{Henrard1987}, let us label the entire sphere surface (with the two Cassini states excluded) as $D_{23}$ (Fig. 2, left).

\subsection*{2.4.2 Type IV}

When $a^2 + b^2 < 1$, the flow is shaped by the presence of three elliptic fixed points $C_1$, $C_2$, $C_3$, and a hyperbolic point $C_4$ (Fig. 2 right). The latter is accompanied by two homoclinic orbits – upper $I_1$ and lower $I_2$. Thus the sphere is first partitioned into three domains: $D_1$ between $C_1$ and $I_1$, $D_2$ between $C_2$, $I_1$, $I_2$, and the third area bounded by $C_3$ and $I_2$. But the last of the three domains is further divided by the curve whose energy equals that of $C_1$ – the thick dashed curve $I_3$ in Fig. 2. The subdomains $D_3$ and $D_4$ may look similar from the geometrical point of view, but note the fact that each curve in $D_3$ has a companion in $D_1$ with the same energy, which is not the case in $D_3$.

Equations \eqref{eq:171} from Appendix B can be used as they are, but here we add an alternative form, based upon the transformation \eqref{eq:129}.

\[
\begin{align*}
\bar{z}_1^* &= \frac{b}{2} + \sqrt{A_4} + \sqrt{B_4} - \sqrt{B_4'}, \quad \bar{z}_2^* = \frac{b}{2} + \sqrt{A_4} + \sqrt{B_4} - 2 \sqrt{B_4'}, \\
\bar{z}_3^* &= \frac{b}{2} - \sqrt{A_4} + \sqrt{B_4} + \sqrt{B_4'}, \quad \bar{z}_4^* = \frac{b}{2} - \sqrt{A_4} + \sqrt{B_4} + 2 \sqrt{B_4'}.
\end{align*}
\]
where
$$A_4 = \frac{b^2}{4} + \rho \cos^2 \frac{\phi_4}{3}, \quad B_4 = \frac{b^2}{4} + \rho \cos^2 \frac{\phi_4 - \pi}{3}, \quad B'_4 = \frac{b^2}{4} + \rho \cos^2 \frac{\phi_4 + \pi}{3},$$
and
$$\cos \phi_4 = abp^{-\frac{1}{2}}. \quad (38)$$

From $A_4 < B_4 < B'_4$ we infer $z^*_1 < 0 < z^*_2 < b < z^*_4 < z^*_1$. The associated $y$ coordinates satisfy inequalities $y^*_1 < y^*_1 < -a < y^*_2 < 0 < y^*_2$.

The energy values are $E_3 < E_1 < E_4 < 0 < E_2$. Referring to Sect. 2.3, we identify $z^*_1 = z_{14}(E_1)$, $z^*_2 = z_{23}(E_2)$, $z^*_3 = z_{12}(E_1)$, and $z^*_4 = z_{22}(E_4)$. The unstable equilibrium energy $E_4$ is of special importance, because it serves to determine the turning points of the separatrices from Eq. (23): $z_4(E_4)$ for $\Gamma_4$, and $z_1(E_4)$ for $\Gamma_2$.

For the energy values $E_1 < E < E_4$, when $\Delta > 0$, each $E$ refers to two periodic curves: one in $D_4$ with the turning points $(z_1, z_2)$ given by (18), and one in $D_1$ with $(z_3, z_4)$ given by Eq. (19). An energy in $E_3 < E < E_1$ defines only one periodic trajectory in $D_1$, and each $E_4 < E < E_2$ defines one periodic curve in $D_2$. Since $\Delta < 0$ in both cases, the turning points are given by $(z_1, z_2)$ from Eq. (35).

2.4.3 Type III

The case of $a^* + b^* = 1$ is specific, but it has to be included to understand the bifurcation between the two neighboring types. Increasing $a$ and/or $b$ from the type IV, we observe that the two Cassini states $C_1$ and $C_3$ merge into a single point $C_{14}$ of neutral stability, the homoclinic orbit $I_{14}$ contracts to a point, and the separatrix $I_3$ merges with the specific periodic orbit $I_3$ into $I_2$. Thus the domains $D_1$ and $D_4$ disappear. In course of transition from type III to type II, the division between $D_2$ and $D_3$ disappears, hence we merge them into a single $D_{23}$. Lanchares and Elipe (1995) refer to this transition under the name of teardrop bifurcation.

The coordinates of the Cassini states can be obtained from Eq. (174) and are expressible in terms of $a$ or $b$ alone, resulting in
$$z^*_1 = \sqrt{1 - A_3} = b^*, \quad y^*_1 = -\sqrt{A_3} = -a^*, \quad (39)$$
and
$$z^*_3 = -z^*_1 A_3 + \sqrt{B_3}, \quad z^*_2 = -z^*_1 A_3 - \sqrt{B_3}, \quad (40)$$
where
$$A_3 = 1 - b^* = a^*, \quad B_3 = A_3^2 - A_3 + 1. \quad (41)$$

The expression of the energy at $C_{14}$ is so simple, that we write it explicitly
$$E_{14} = -\frac{3}{2} \sqrt{ab^2} = -\frac{3}{2} \frac{(z^*_1)^2 (y^*_1)^4}{2} = -\frac{3A_3^2 (1 - A_3)}{2}. \quad (42)$$
It is useful also in finding the turning point of $I_{23}$ at $x = 0$, which is
$$z_m = z^*_1 (1 - 4A_3), \quad y_m = -y^*_1 (3 - 4A_3). \quad (43)$$

For regular trajectories with $E_3 < E < E_2$, and $E \neq E_{14}$, their turning points are given by $z_1$ and $z_2$ from Eq. (25), which also means $z_1 < z_2$. 
3 Analytical solution

3.1 Weierstrass form

3.1.1 General framework

Equations of motion (9) or (13) admit an exact analytical solution in terms of the elliptic functions. Actually, the solution hinges upon the fact that the variable \( Z \) can be considered separately from the remaining two. To see it, we can differentiate \( \dot{Z} = -aX \), substitute \( \dot{X} \) from the first of Eqs. (13), and use the energy integral (11) to eliminate \( Y \), obtaining

\[
\ddot{Z} = -\frac{1}{2}Z^3 - EZ - a^2 b. \tag{44}
\]

Equation (44) defines a 1 degree of freedom, conservative system with the potential

\[
V_z(Z) = \frac{Z^4}{8} + \frac{E}{2}Z^2 + a^2 bZ, \tag{45}
\]

and the energy integral

\[
\frac{Z^2}{2} + V_z(Z) = E_z = \text{const.} \tag{46}
\]

But the potential \( V_z(Z) \) and the energy \( E_z \) are closely related to the polynomial \( W(Z) \):

\[
E_z - V_z(Z) = \frac{1}{8}W(Z), \quad \text{with} \quad 2E_z = \frac{W(0)}{4} = -(E - a^2)^2 + a^2 (1 - b^2). \tag{47}
\]

Thus, whether we use Eq. (46), or the squared Eqs. (13) and (16), the outcome is \( 4\dot{Z}^2 = W(Z) \), amenable to the separation of variables method.

Since \( W(Z) \) is a quartic polynomial, finding \( Z(t) \) amounts to inverting the elliptic integral

\[
\sigma_z \int_{Z_0}^{Z} \frac{dZ}{\sqrt{W(Z)}} = \frac{1}{2} \int_{t_0}^{t} \sigma_x dt, \tag{48}
\]

where \( \sigma_z = \text{sgn}\dot{Z} = -\text{sgn}X \), and the right-hand side evaluates to

\[
\tau_0 = \frac{t - t_0}{2}. \tag{49}
\]

The integral to the left of (48) should be a monotonous function of \( Z \) to allow the inversion (solving for \( Z(t) \)). In the admissible range of \( Z \) between two turning points (or one turning point and an unstable equilibrium) the sign of \( Z \) is constant, which allows to establish its value from the initial condition and to pull \( \sigma_z \) out of the integrand.

The integral to the left of Eq. (48) is an elliptic integral and it can be reduced to the Weierstrass normal form by an appropriate transformation \( Z \rightarrow s \), so that

\[
\sigma_z \int_{Z_0}^{Z} \frac{dZ}{\sqrt{W(Z)}} = \int_{s}^{\infty} \frac{dv}{\sqrt{S(s)}} = \tau_0, \tag{50}
\]

where \( S(s) = 4s^3 - g_2s - g_3 \), is the cubic polynomial of the Weierstrass resolvent (133) for \( W(Z) = 0 \), with the invariants \( g_2, g_3 \) defined in Appendix A.2 – Eqs. (134) and (135). Note
that the initial value \( Z_0 \) is always mapped to \( s \to \infty \), regardless of the ordering of the integration limits \( Z_0 \) and \( Z \), which explains the presence of \( \sigma \) in the forthcoming transformations.

Solving the rightmost part of Eq. (50) amounts to the substitution of the Weierstrass elliptic function

\[
\sigma = \wp(\tau_0; g_2, g_3), \quad \text{hence } \sigma = \wp(\tau_0; g_2, g_3) d\tau_0.
\]

In all further instances we will use the abbreviated notation for the Weierstrass \( \wp \) function

\[
\wp(u) = \wp(u; g_2, g_3),
\]

hence \( d\sigma = \wp'(\tau_0; g_2, g_3) d\tau_0 \).

(51)

(52)

(53)

As expected from Eq. (50).

Once the solution for the first half-period is found, its continuation can be investigated by the substitution into Eq. (44) which is free from the restrictions imposed by the inversion procedure.

3.1.2 Initial conditions at turning point

Let us begin with the easiest situation, when the initial condition is \( Z_0 = Z_j = z_j - b \), and \( t_0 = t_j \) is the epoch of crossing the turning point \( X_0 = 0 \), i.e. \( \dot{Z} = 0 \). Then the integral to the left of (50) is reduced to the Weierstrass normal form in variable \( s \) through the rational transformation (Enneper, 1890; Bianchi, 1901)

\[
Z = Z_j + \frac{1}{4} W'(Z_j) s - \frac{1}{24} W''(Z_j), \quad s = \frac{W'(Z_j)}{4(Z_j)} + \frac{1}{24} W''(Z_j).
\]

(54)

Replacing the subscript 0 with \( j \) in the formulae of Sect. 3.1.1, we find

\[
Z = Z_j + \frac{1}{4} W'(Z_j) \tau_j - \frac{1}{24} W''(Z_j),
\]

where

\[
\tau_j = \frac{t - t_j}{2}.
\]

(55)

(56)

Thanks to the symmetry of trajectory with respect to the turning point, the solution does not depend on \( \sigma \), and remains valid for all values of \( \tau_j \).

Given the initial conditions \( X(t_j) = 0, Y(t_j) = Y_j, Z(t_j) = Z_j \), and knowing \( Z(t) \), we may complete the solution for \( R(t) \). The two missing variables come from the energy integral (11), with

\[
E = \frac{Z^2}{2} + aY = -\frac{Z_j^2}{2} + aY_j.
\]

(57)
and from the third of the equations of motion (12), i.e. \(X = -Z/a\), so

\[
X = - \left( \frac{Z - Z_j}{2a} \right)^2 \frac{\varphi_j}{ab + Y_j Z_j},
\]

\[
Y = Y_j + \frac{Z^2 - Z_j^2}{2a},
\]

\[
Z = Z_j - \frac{2a^2 b + 2EZ_j + Z_j^3}{\varphi_j + \frac{1}{2} E + \frac{1}{2} Z_j^2} = Z_j - \frac{6a (ab + Y_j Z_j)}{3 \varphi_j + aY_j + Z_j^2}.
\]

The coordinates on the unit sphere are, as usually, \(x = X\), \(y = Y - a\), and \(z = Z + b\).

### 3.1.3 Arbitrary initial conditions: Weierstrass–Biermann form

If the initial conditions are not at the turning point, i.e. \(Z_0 \neq Z_f\), the transformation \(Z \mapsto s\) is more cumbersome than (54). Whittaker and Watson (1927) quote two alternative forms of the final solution: one due to Weierstrass, published by Biermann (1865), and one by Mordell (1915). Actually, there is yet another, formally elegant form – the secondo metodo d’inversione of Bianchi (1901), but the relation of its constants to initial conditions is rather awkward, so we do not consider it here.

The Weierstrass-Biermann form results from the substitution (Enneper, 1890)

\[
Z = Z_0 + \frac{\sigma \sqrt{W_0} \sqrt{S(s)} + \frac{1}{2} W_1 (s - \frac{1}{12} W_2) + \frac{1}{5} W_0 W_3}{2 \left( s - \frac{1}{12} W_2 \right)^2 - \frac{1}{2} W_0 W_4},
\]

\[
s = \frac{\sqrt{W_0} \sqrt{W(Z)} + \frac{1}{2} W_1 (Z - Z_0) + \frac{1}{5} W_2 (Z - Z_0)^2}{2 (Z - Z_0)^2},
\]

where \(\sigma\) is the sign of \(Z - Z_0\), and

\[
W_k = \frac{1}{k!} \left[ \frac{d^k W(Z)}{dz^k} \right] Z_0 ,
\]

are the Taylor coefficients in

\[
W(Z) = \sum_{k=0}^4 W_k (Z - Z_0)^k.
\]

After expressing \(E\) in terms of the initial conditions, they take the form

\[
W_0 = 4a^2 X_0^2, \quad W_1 = -8a (ab + Y_0 Z_0), \quad W_2 = -4 (aY_0 + Z_0^2), \quad W_3 = -4Z_0, \quad W_4 = -1.
\]

Substituting \(s = \varphi(\tau_0)\), and recalling that \(\sqrt{S(s)} = -\varphi'(\tau_0)\) over the first half-period, we obtain

\[
Z = Z_0 + \frac{\sigma \sqrt{W_0} \varphi'(\tau_0) + \frac{1}{2} W_1 (\varphi(\tau_0) - \frac{1}{12} W_2) + \frac{1}{5} W_0 W_3}{2 (\varphi(\tau_0) - \frac{1}{12} W_2)^2 - \frac{1}{2} W_0 W_4}.
\]

Then, accounting for Eqs. (63) and the fact that \(\sqrt{W_0} = 2a |X_0| = -2\sigma a X_0\),

\[
Z = Z_0 + a \left[ X_0 \varphi'(\tau_0) - 2 (ab + Y_0 Z_0) \left[ \varphi(\tau_0) + \frac{1}{5} (aY_0 + Z_0^2) \right] - \frac{2aX_0^2 Z_0}{\left[ \varphi(\tau_0) + \frac{1}{5} (aY_0 + Z_0^2) \right]^2 + a^2 X_0^2} .
\]

\[
(65)
\]
The substitution into Eq. (43) with the initial condition \( Z_0 = -aX_0 \), shows that the formula (65) is actually valid for all values of \( \tau_0 \), so the initial restriction to the first half-period can be abolished.

Like before, the remaining two variables \( Y, X \) are found from the energy integral and from the equations of motion, respectively. Thus for \( Y \) we have simply

\[
Y = Y_0 + \frac{Z_0^2 - Z_j^2}{2a},
\]

whereas \( X = -Z_j/a \), requires the differentiation and some manipulations leading to

\[
X = -X_0 \left( 3\varphi(\tau_0) - \frac{1}{2}g^2 \right) + (ab + Y_0 Z_0) \varphi'(\tau_0) + \frac{(Z - Z_0)}{a} \left( [\varphi(\tau_0) + \frac{1}{2} (aY_0 + Z_0^2)] \varphi'(\tau_0) \right),
\]

\[
[\varphi(\tau_0) + \frac{1}{2} (aY_0 + Z_0^2)]^2 + a^2X_j^2,
\]

where \( 0 \leq s \leq g^2/2 \).

### 3.1.4 Arbitrary initial conditions: Safford form

If \( X_0 = 0 \), so \( \tau_j = \tau_j \), the above solution simplifies to Eqs. (58) in a straightforward manner. On the other hand, as pointed out by Safford (1919), the general solution (64) can be derived from the particular solution (65) by assuming \( \tau_j = \tau_j + \phi_j \), and making use of the addition theorem for the Weierstrass function \( \varphi \). Given the initial conditions \( X_0, Y_0, Z_0 \) at \( t = \tau_0 \), we can find the turning point coordinates \( Z_j, Y_j \) for this trajectory using the formulae of Sect. 2.3.

Then, according to Safford (1919), the phase \( \phi_j \) is defined through Eq. (66) at \( Z = Z_j \), giving \( s_j = \varphi(\phi_j) \). Thus, as an alternative to using Eqs. (65) and (66) for arbitrary initial conditions, one can first compute \( \phi_j = \varphi^{-1}(s_j) \), and then apply the particular solution (58) with \( \tau_j = \tau_j + \phi_j \).

But Safford (1919) cared solely about the reduction of the integrand form, so he paid no attention to the problem that \( \varphi(\phi_j) \) is uniquely invertible only in the domain \( 0 \leq \phi_j \leq \omega_1 \), i.e. within the first half-period. In order to properly place the phase in the full period range \( -\omega_1 \leq \phi_j \leq \omega_1 \), one needs the information about the sign of the derivative \( \varphi'(\phi_j) \). To this end, we take a slightly different approach, that actually leads to a simpler expression for \( s_j \).

Substituting \( \tau_j = \phi_j \), and \( R = R_0 \) in Eqs. (58), we can solve them to find

\[
\varphi(\phi_j) = -2a \frac{ab + Y_j Z_j}{Z_0 - Z_j} - \frac{aY_j + Z_j^2}{3}.
\]

\[
\varphi'(\phi_j) = -X_0 \left( ab + Y_j Z_j \right) \left[ \frac{2a}{Z_0 - Z_j} \right]^2,
\]

and this set allows the unique determination of the phase. Then, we can apply Eqs. (58) to obtain

\[
X = -\left( \frac{Z - Z_j}{2a} \right)^2 \frac{\varphi'(\tau_0 + \phi_j)}{ab + Y_j Z_j}, \quad Z = Z_j - \frac{6a (ab + Y_j Z_j)}{3 \varphi(\tau_0 + \phi_j) + aY_j + Z_j^2},
\]

with \( Y_j \) and \( X \) derived from the energy integral.

Compared to the Weierstrass-Biermann solution, we gain the simplicity at the expense of pre-computing the turning point coordinates and the phase for the given (arbitrary) initial conditions.
3.1.5 Arbitrary initial conditions: Mordell form

The inversion formula of [Mordell (1915)] was formulated in the language of homogeneous binary forms; in order to apply it to the Colombo top problem, let us translate it to the univariate polynomials framework. The link is simple: the quartic polynomial \( W(Z) \) from Eq. (122), and the quartic binary form

\[
V(\xi, \eta) = a_0 \xi^4 + 4a_1 \xi^3 \eta + 6a_2 \xi^2 \eta^2 + 4a_3 \xi \eta^3 + a_4 \eta^4,
\]  

(71)

can be matched by

\[
W(Z) = V(Z, 1), \quad V(\xi, \eta) = \eta^4 W(\xi \eta^{-1}).
\]

(72)

Thus, after the substitution \( Z = \xi \eta^{-1} \), Eq. (50) is equivalent to

\[
\sigma \int_{(\xi_0, \eta_0)} \left( \int_{\sqrt{V(\xi, \eta)}} \frac{\eta \, d\xi - \xi \, d\eta}{\sqrt{V(\xi, \eta)}} \right) = \int_0^\infty \frac{dx}{\sqrt{s(s)}} = \tau_0,
\]

(73)

where the definite integral in \( Z \) has been replaced by a path-independent line integral. Skipping the intermediate steps described in [Mordell (1914, 1915)], the inversion of (73) results in

\[
\xi = -\sigma \xi_0 \sqrt{V_0} \phi' (\tau_0) + \frac{1}{2} \frac{\partial V_0}{\partial \eta_0} \phi (\tau_0) + \frac{1}{2} \frac{\partial \tilde{h}_0}{\partial \eta_0},
\]

\[
\eta = -\sigma \eta_0 \sqrt{V_0} \phi' (\tau_0) - \frac{1}{2} \frac{\partial V_0}{\partial \xi_0} \phi (\tau_0) - \frac{1}{2} \frac{\partial \tilde{h}_0}{\partial \xi_0},
\]

(74)

where \( V_0 = V(\xi_0, \eta_0) \), and \( \tilde{h}_0 \) is the Hessian covariant of \( V \).

From the correspondence rules (73) we derive

\[
\frac{\partial V_0}{\partial \xi_0} = \eta_0 W_1,
\]

\[
\frac{\partial V_0}{\partial \eta_0} = \eta_0^3 (4 \eta_0 W_0 - \xi_0 W_1),
\]

\[
\tilde{h}_0 = \frac{1}{144} \left( \frac{\partial^2 V_0}{\partial \xi^2_0} \frac{\partial^2 V_0}{\partial \eta^2_0} - \left( \frac{\partial^2 V_0}{\partial \xi_0 \partial \eta_0} \right)^2 \right) = \eta_0^4 \left( 8 W_0 W_2 - 3 W_1^2 \right),
\]

(75)

\[
\frac{\partial \tilde{h}_0}{\partial \xi_0} = \eta_0^3 \left( 6 \eta_0 W_0 W_2 - W_1 W_2 \right),
\]

\[
\frac{\partial \tilde{h}_0}{\partial \eta_0} = \frac{\eta_0^4}{12} \left( 8 \eta_0 W_0 W_2 - 3 \eta_0 W_1^2 + \xi_0 W_1 W_2 - 6 \xi_0 W_0 W_1 \right),
\]

where \( W_n \) are defined in Eq. (63). By letting \( Z = \xi / \eta \), and \( Z_0 = \xi_0 / \eta_0 \), we find that (74) and (75) amount to

\[
Z = \frac{\xi}{\eta} = Z_0 + \frac{48 W_0 \phi' (\tau_0) + 8 W_0 W_2 - 3 W_1^2}{24 \sigma \sqrt{W_0} \phi' (\tau_0) - 12 W_1 \phi (\tau_0) + W_1 W_2 - 6 W_0 W_1}.
\]

(76)

---

1. In this paper we use the Hessian covariant as defined by Janson (2011), which is the same as in Whittaker and Watson (1923). Its sign is opposite to the one originally applied by Mordell (1915).
Proceeding like in Sect. 3.1.3 we obtain the final form
\[ Z = Z_0 + 4\eta \left(\frac{X_0^2}{X_0^0} \left[ \varphi(\tau_0) - \frac{2}{7} (aY_0 + Z_0^2) \right] - (ab + Y_0Z_0) \right) \] \( \times \left[ \varphi(\tau_0) + \frac{2}{7} (aY_0 + Z_0^2) \right] + 2aX_0^2Z_0, \] (77)
looking different from the Weierstrass-Biermann solution (65), yet providing the same values of \( Z \). Notably, the Mordell solution involves only the first power of \( \varphi \).

In order to demonstrate the equivalence of the Weierstrass-Biermann and the Mordell solutions, one can multiply the numerator and the denominator in Eq. (65) by the factor
\[ (48xX_0\varphi(\tau_0) - (12\varphi(\tau_0) - W_2) W_1 - 6W_0W_3), \]
use the identity \((\varphi(\tau_0))^2 = S(\varphi(\tau_0))\), and substitute the expressions of the invariants in terms of \( W_\eta \)
\[ g_2 = \frac{W_2^2}{12} - \frac{W_2W_3}{4} + W_0W_4, \quad g_3 = \begin{vmatrix} W_4 & \frac{1}{2}W_3 & \frac{1}{2}W_2 & \frac{1}{2}W_1 \\ \frac{1}{2}W_3 & W_2 & \frac{1}{2}W_1 & W_0 \end{vmatrix}, \] (78)
by analogy with the definitions (134) and (135). The result of this procedure is the Mordell solution (77). Obviously, both (65) and (77) admit the same limit expression (58) when \( X_0 = 0 \) and \( \tau_0 = \tau_\eta \).

As usually, the solution for \( X \) and \( Y \) can be derived from the equations of motion and the energy integral, like in Sect. 3.1.3.

3.2 Solution in terms of the Jacobii elliptic functions

3.2.1 Weierstrass functions in terms of the Jacobii functions

Although the solution in terms of the Weierstrass \( \varphi \) function presented in Sect. 3.1 may look universal, its qualitative properties depend on the values of the invariants through the sign of the discriminant \( \Delta \). Indeed, the sign plays the central role in expressing the solution in terms of the Jacobian elliptic functions. In this section, only the generic, \( \Delta \neq 0 \) cases are to be discussed.

The basic relation between the Weierstrass and Jacobii functions is formally universal (Byrd and Friedman, 1971)
\[ \varphi(\tau) = e_3 + \left(\frac{\gamma_p}{\text{sn}(u_p, k_p)}\right)^2 = e_3 + \frac{1}{\gamma_p^2} \frac{1 + \text{dn}(2u_p, k_p)}{1 - \text{cn}(2u_p, k_p)}, \] (79)
\[ \varphi'(\tau) = -2 \left(\frac{\gamma_p}{\text{sn}(u_p, k_p)}\right)^3 \frac{\text{cn}(u_p, k_p) \text{dn}(u_p, k_p)}{\text{sn}(2u_p, k_p)}, \] (80)
where
\[ u_p = \gamma_p \tau, \quad \gamma_p = \sqrt{e_1 - e_3}, \quad k_p = \sqrt{\frac{e_2 - e_3}{e_1 - e_3}} \] (81)
But if we restrict considerations to the real arguments and moduli of the elliptic functions, the above expressions are valid only if the discriminant \( \Delta \) from Eq. (126) is positive.
When $\Delta < 0$, which means complex $e_2$ and $e_3$, the appropriate form is (Abramowitz and Stegun, 1972)

$$\varphi(\tau) = e_1 + \eta_0 \left( \frac{\operatorname{cn}(u_n, k_n)}{\operatorname{sn}(u_n, k_n) \operatorname{dn}(u_n, k_n)} \right)^2 = e_1 + \eta_0 \frac{1 + \operatorname{cn}(2u_n, k_n)}{1 - \operatorname{cn}(2u_n, k_n)}, \quad (82)$$

$$\varphi'(\tau) = -2 \left( \frac{\sqrt{k_n}}{\operatorname{sn}(u_n, k_n) \operatorname{dn}(u_n, k_n)} \right)^3 \operatorname{cn}(u_n, k_n) \left( 1 - \kappa_n^2 + \kappa_n^2 \operatorname{cn}^4(u_n, k_n) \right)$$

$$= -4\eta_0^3 \frac{\operatorname{sn}(2u_n, k_n) \operatorname{dn}(2u_n, k_n)}{(1 - \operatorname{cn}(2u_n, k_n))^2}, \quad (83)$$

where

$$u_n = \sqrt{\kappa_n} \tau, \quad \eta_0 = \gamma_p \sqrt{e_1 - e_2} = \frac{1}{2} \sqrt{12e_1^2 - g_2}, \quad k_n = \frac{1}{2} \sqrt{2 - \frac{3e_1}{\gamma_p}}. \quad (84)$$

The two cases are linked by the complex modulus transformation – see Byrd and Friedman (1971) formula 165.07.

Let the initial conditions at the epoch $t_0$ be ($X_0, Y_0, Z_0$). The equations relating the Jacobi and Weierstrass functions can be substituted in to any of the solution forms provided in Sect. 3.1. For the Weierstrass–Biemann or the Mordell form, it is enough to compute the energy $E = E(Y_0, Z_0)$ and the discriminant $\Delta$ to choose the appropriate set (79, 80) or (82, 83). Then, from the invariants $g_2, g_3$ the roots $e_1, e_2, e_3$ are found, which allows the computation of $R(t)$ or $r(t)$ for any epoch $t$.

Below, we discuss the Safford form from Sect. 3.1.4, which allows to use the initial conditions at any $t_0$, but requires the turning point coordinates ($X_j = 0, Y_j = y_j + a, Z_j = z_j - b$) as supplementary parameters. Having computed two appropriate turning points, such that either $Z_1 < Z_0 < Z_2$, or $Z_3 < Z_0 < Z_4$, we pick one of them as the reference point $Z_j$ to be used in Eq. (70). Then, after determining phase $\phi_j$ with respect to the turning point $Z_j$, the motion can be computed for any epoch $t$, using $\tau_j = t_0 + \phi_j$.

### 3.2.2 Motion in D1 and D4

The case $\Delta > 0$ occurs only in type IV, when the energy is bounded by $E_4 < E < E_1$; either in the domain $D_1$, where turning points $Z_1, Z_4$ are given by Eq. (19) with $Z_1 > Z_3 < Z_4$, or in the domain $D_4$, where the turning points are $Z_1, Z_2$ given by Eq. (18) and $Z_1 < Z_3 < Z_4$. Then, selecting any appropriate $Z_j$ as the reference point, we can use the expressions

$$X = \frac{A_p y_p \operatorname{sn}(u_p, k_p) \operatorname{cn}(u_p, k_p) \operatorname{dn}(u_p, k_p)}{(1 + B_p \operatorname{sn}^2(u_p, k_p))^2} \quad (85)$$

$$Y = Y_j + \frac{Z_j^2 - Z_j^2}{2a} = Y_j - \frac{A_p \operatorname{sn}^2(u_p, k_p)}{1 + B_p \operatorname{sn}^2(u_p, k_p)} \left( Z_j - \frac{a A_p \operatorname{sn}^2(u_p, k_p)}{1 + B_p \operatorname{sn}^2(u_p, k_p)} \right) \quad (86)$$

$$Z = Z_j - \frac{a A_p \operatorname{sn}^2(u_p, k_p)}{1 + B_p \operatorname{sn}^2(u_p, k_p)} \quad (87)$$

where

$$A_p = \frac{2(ab + Y_j Z_j)}{e_1 - e_3}, \quad B_p = \frac{a (Y_j + Z_j^2)}{e_1 - e_3}. \quad (88)$$
and the phase \( \phi_j \) in
\[
u_p = \gamma_p (t_0 + \phi_j),
\]
(89)
can be computed from the incomplete elliptic function of the first kind \( F \):
\[
\gamma_p \phi_j = F(v_j, k_p),
\]
(90)
\[
v_j = \text{am}(\gamma_p \phi_j, k_p) = \text{sgn}(A_p X_0) \arcsin \sqrt{\frac{Z_j - Z_0}{aA_p - (Z_j - Z_0)B_p}},
\]
(91)
The motion is periodic and the period \( P_t \) (with respect to time \( t \)) is given by the complete elliptic integral of the first kind \( K \)
\[
P_t = \frac{4K(k_p)}{\gamma_p}.
\]
(92)

### 3.2.3 Motion in \( D_2, D_3 \) and \( D_{23} \)

The common feature of trajectories in domains \( D_2 \), \( D_3 \) and \( D_{23} \) is the negative discriminant \( \Delta \). Thus, given the initial conditions and resulting energy, we pick \( Z_1 \) or \( Z_2 \) computed from Eqs. (25) as the reference point \( Z_j \) and compute, for any \( t \)
\[
X = \frac{A_n \sqrt{2(1 - B_n) \text{sn}(2\alpha_n, \kappa_n) \text{dn}(2\alpha_n, \kappa_n)}}{(1 - B_n \text{cn}(2\alpha_n, \kappa_n))^2},
\]
(93)
\[
Y = Y_j + \frac{Z_j^2 - Z_j^2}{2a} = Y_j - A_n \frac{1 - \text{cn}(2u_n, \kappa_n)}{1 - B_n \text{cn}(2u_n, \kappa_n)} \left( Z_j - \frac{aA_n}{2} \frac{1 - \text{cn}(2u_n, \kappa_n)}{1 - B_n \text{cn}(2u_n, \kappa_n)} \right),
\]
(94)
\[
Z = Z_j - aA_n \frac{1 - \text{cn}(2u_n, \kappa_n)}{1 - B_n \text{cn}(2u_n, \kappa_n)},
\]
(95)
where
\[
A_n = \frac{2(ab + Y_j Z_j)}{e_1 + \gamma + \frac{1}{2} (aY_j + Z_j^2)}, \quad B_n = 1 - \frac{2\gamma}{e_1 + \gamma + \frac{1}{2} (aY_j + Z_j^2)}.
\]
(96)
The phase in
\[
u_n = \sqrt{\frac{2}{\gamma_n}} (\tau_0 + \phi_j),
\]
(97)
results from
\[
2 \sqrt{\frac{2}{\gamma_n}} \phi_j = F(v_j, k_n),
\]
(98)
\[
v_j = \text{am}(2 \sqrt{\frac{2}{\gamma_n}} \phi_j, k_n) = \text{sgn}(A_n X_0) \arccos \left( \frac{aA_n + Z_0 - Z_j}{aA_n + (Z_0 - Z_j)B_n} \right).
\]
(99)
The related period is
\[
P_t = \frac{4K(k_n)}{\sqrt{\gamma_n}}
\]
(100)
The above formulation is universal, i.e. appropriate in types II, III, and IV, provided \( \Delta < 0 \).
3.3 Special cases

3.3.1 Reduction rules

By special cases we mean the ones where elliptic functions reduce to the elementary ones. They could be studied by taking limits of the Jacobi functions at $k_n$ or $k_p$ tending to 0 or 1. But it is more direct to observe that each of the special cases, be it the Cassini states, separatrices, or the special curve $\Gamma_1$, results from the reduction of the Weierstrass function $\wp(u, g_2, g_3)$ to the special case

$$\wp(u; 3, 1) = 1 + \frac{3}{2\tan^2\left(\sqrt{\frac{2}{3}}u\right)}.$$  (101)

Since $\Delta = 0$ implies $(g_2/3)^3 = g_3^2$, reduction to the above form is always possible by the homogeneity relations (Abramowitz and Stegun, 1972)

$$\wp(u; g_2, g_3) = \lambda^{-2} \wp(\lambda^{-1}u; \lambda^4g_2, \lambda^6g_3),$$  (102)

except for $g_2 = g_3 = 0$, when

$$\wp(u; 0, 0) = u^{-2}.$$  (103)

For the derivative $\wp'$, respective equations are obtained by straightforward differentiation.

Depending on the sign of $g_3$, two procedures are available. For $g_3 > 0$, the substitution $\lambda = g_3^{-\frac{1}{6}}$ leads straight to

$$\wp(u; g_2, g_3) = e_1\wp(\sqrt{e_1}u; 3, 1),$$  (104)

where $g_3 = e_1^3$, according to Eq. (145).

If $g_3 < 0$, two steps are taken. First, letting $\lambda = i$, we convert

$$\wp(u; g_2, g_3) = -\wp(iu; g_2, -g_3).$$  (105)

Then, with $\lambda = (-g_3)^{-\frac{1}{6}}$, we recover

$$\wp(u; g_2, g_3) = -2e_{12}\wp(i\sqrt{2e_{12}}u; 3, 1).$$  (106)

according to Eq. (146).

3.3.2 Cassini states $C_2$ and $C_3$

When discussing the Cassini states, we can use the simplest form (58) for $Z$. Although the time dependence at the Cassini states does vanish due to $ab + Y_0Z_0 = -W'(Z_0)/(8a) = 0$, and $X_0 = 0$, it remains of interest to inspect $\wp(\tau)$ in the numerator, because its period is the period of small oscillations around the stable equilibrium.

Given $a$ and $b$, one should first establish the problem type in order to compute the appropriate coordinate $Z_j^* = z_j^* - b$, or $Z_j^* = z_j^* - b$. These are Eqs. (33), (40), and (36) for the types II, III, and IV, respectively. Starting from this point, the procedure is common: $Z_j^*$ gives the energy of the Cassini state

$$E_j = -\frac{(Z_j^*)^3 + a^2b}{2Z_j^*}, \quad j = 2, 3,$$  (107)
which substituted in Eq. (134) or (135) gives the invariants $g_2(E_j)$ or $g_3(E_j)$ – both positive.

Resorting to Eqs. (104) and (101), we find that $Z$ in solution (58) depends on the squared tangent of $\sqrt{3e_1/8}$ which implies the period

$$P_j = 2\pi \sqrt{\frac{2}{3e_1(E_j)}}, \quad j = 2, 3,$$

where $e_3(E_j) = \sqrt{g_3} = \sqrt{g_2/2}$. The same result can be obtained by taking the limit of (100) at $k_n \to 0$.

3.3.3 Cassini state $C_4$ and homoclinic orbits $\Gamma_1, \Gamma_2$

The energy of the unstable Cassini state $C_4$ can be evaluated from Eq. (107) with $j = 4$, and $Z_4^*$ given by (36). However, we need it not for the Cassini state itself, but rather to describe the motion on homoclinic orbits having the energy $E_4$, which are $\Gamma_1$ and $\Gamma_2$. To this end, we will use the Safford form (70) with the reference points given by Eq. (23), namely: $Z_4$ for $\Gamma_1$, and $Z_1$ for $\Gamma_2$.

Since $g_3 < 0$, the reduction (106) leads to

$$X = \tilde{A} \sqrt{3e_{12}} \sinh(2\tilde{u}) / (1 + \tilde{B} \sinh^2 \tilde{u})^2,$$

$$Y = Y_j + \frac{Z_j^2 - Z_j^*}{2a} = Y_j - \frac{2\tilde{A} \sinh^2 \tilde{u}}{1 + \tilde{B} \sinh^2 \tilde{u}} \left( Z_j - \frac{a\tilde{A} \sinh^2 \tilde{u}}{1 + \tilde{B} \sinh^2 \tilde{u}} j \right),$$

$$Z = Z_j - \frac{2a\tilde{A} \sinh^2 \tilde{u}}{1 + \tilde{B} \sinh^2 \tilde{u}},$$

with the coefficients

$$\tilde{A} = \frac{ab + Y_j Z_j}{3e_{12}}, \quad \tilde{B} = \frac{e_{12} + \frac{1}{3}(aY_j + Z_j^2)}{3e_{12}},$$

where $e_{12}$ is given by Eq. (146). The argument

$$\tilde{u} = \sqrt{3e_{12}}(\tau_0 + \phi_j),$$

whose phase with respect to $Z_j$ at the initial epoch $t_0$ is given by

$$\phi_j = \text{sgn}(X_0\tilde{A}) \arcsinh \sqrt{\frac{Z_j - Z_0}{2a\tilde{A} - B(Z_j - Z_0)}}.$$

Note that the time rate of argument $\tilde{u}$ is the same at both the separatrices, and, as expected the solution tends to the Cassini state $C_4$ asymptotically at $t \to \pm\infty$. 

3.3.4 Cassini state $C_1$ and special orbit $\Gamma_3$

The Cassini state $C_1$, being a stable equilibrium with $g_3 > 0$, is characterized by the period of small oscillations given directly by the formula (108) with the energy $E_1$ evaluated at $Y_1$, $Z_2$ deduced from Eq. (36). What makes the difference, compared to $C_2$ or $C_3$, is the presence of another trajectory having the energy $E_1$ – the special curve $\Gamma_3$. 

The motion along $\Gamma_3$ can be described using the Safford form solution (70) with respect to the turning points $Z_1$ or $Z_2$, given by Eq. (21). Then, performing the reduction (104), we obtain the solution in terms of trigonometric functions

\[
X = \frac{\bar{A} \gamma_3 \sin 2 \bar{u}}{(1 + \bar{B} \sin^2 \bar{u})^2},
\]

\[
Y_j = Y_j + \frac{Z^2 - Z_2^2}{2a} = Y_j - \frac{\bar{A} \sin^2 \bar{u}}{1 + \bar{B} \sin^2 \bar{u}} \left( Z_j - \frac{a \bar{A} \sin^2 \bar{u}}{2(1 + \bar{B} \sin^2 \bar{u})} \right),
\]

\[
Z = Z_j - \frac{a \bar{A} \sin^2 \bar{u}}{1 + \bar{B} \sin^2 \bar{u}},
\]

where

\[
\bar{A} = \frac{4(ab + Y_j Z_j)}{3e_1}, \quad \bar{B} = \frac{2}{3} \left( aY_j + Z_j^2 \right) - e_1.
\]

The argument $\bar{u}$ is

\[
\bar{u} = \gamma_3 (\tau_0 + \phi_j),
\]

where $\gamma_3 = \sqrt{3e_1/2}$, as in Eq. (81) for $e_3 = -e_1/2$, hence the time period of the solution is the same as that of small oscillations around $C_1$, i.e. $P_1$ given by Eq. (108). The phase $\phi_j$ can be computed from

\[
\gamma_3 \phi_j = \text{sgn}(\bar{A}X_0) \arcsin \sqrt{\frac{Z_j - Z_0}{aA - (Z_j - Z_0)\bar{B}}},
\]

The above solution can be obtained either from (86) with $k_p = 0$, and $e_3 = -e_1/2$, or – in a different form – from (93) with $k_n = 0$, and $g_2 = 3e_1^2$.

3.3.5 Cassini state $C_{14}$ and special orbit $\Gamma_{23}$

The most degenerate case occurs in type III, where $C_{14}$ is the cusp (parabolic) equilibrium with energy $E_{14}$ given by Eq. (42). The homoclinic curve $\Gamma_{23}$ with this energy can be parameterized using rational functions of time, as indicated by the reduction formula (103). For the sake of using the Safford form, we introduce

\[
u_{23} = t - t_0 + \tau_{23}, \quad \tau_{23} = t_0 - t_m,
\]

where $\tau_{23}$ is the time interval between the initial epoch $t_0$ and the epoch of crossing the reference turning point with coordinates $Y_m = y_m + a$, and $Z_m = z_m - b$, as given by Eq. (43).
Then, with $\beta = \sqrt{a^3 b^2}$,

$$X = -\frac{8abu_{23}}{1 + \beta u_{23}^2} Z,$$  \hspace{1cm} (118)

$$Y = Y_m + \frac{4abu_{23}^2}{1 + \beta u_{23}^2} \left( Z_m + \frac{2a^2 b u_{23}^2}{1 + \beta u_{23}^2} \right),$$  \hspace{1cm} (119)

$$Z = Z_m + \frac{4a^2 b u_{23}^2}{1 + \beta u_{23}^2}.$$  \hspace{1cm} (120)

The time offset $\tau_{23}$ is given by

$$\tau_{23} = -\text{sgn} X_0 \sqrt{\frac{Z_0 - Z_m}{4a^2 b - \beta (Z_0 - Z_m)}}.$$  \hspace{1cm} (121)

### 4 Conclusions

It is common to describe the integral curves of the Colombo top problem as an intersection of a parabolic cylinder and a unit sphere, the latter being centered at the origin of the $x, y, z$ coordinate system. We have proposed a new point of view: the curves can be tracked along the intersection of two out of the three invariant surfaces: a parabolic cylinder, a sphere, and a paraboloid of revolution. This thread has been merely signaled, but it can be of possible interest when designing geometric integrators for the numerical treatment of the Colombo top motion. Moreover, using the shifted coordinates $X, Y, Z$, one introduces the symmetry to the parabolic cylinder on the paraboloid (at the expense of having an off-centered sphere) which does simplify a number of expressions given in this work.

When partitioning the phase space of the Colombo top problem, we have completed the landscape, well known from earlier works, with an interesting but hitherto overlooked feature: the trajectory $\Gamma_3$ which is unique by being periodic, yet expressible in terms of elementary functions of time. Its presence calls for the distinction of $D_3$ and $D_4$ domains even though qualitatively they look similar. It also adds to a better understanding of the parametric bifurcation associated with the passage from type II, to type IV.

The analytical expressions for the turning points of the Colombo top trajectories as functions of energy, given in Sect. 2.3, had not been reported so far. The expressions for the location of the Cassini states from Sect. 2.4 depend only on parameters $a, b$. Up to some rearrangement of terms, they are similar to those of Saillenfest et al (2019) in type II or III. For type IV, when $a^2 + b^2 < 1$, the Cardano form provided by Saillenfest et al (2019) is formally correct, but it gives real values only as the sums of two complex conjugates (casus irreducibilis of the resolvent cubic). In the present work we have preferred to use expressions based on the purely real trigonometric form whenever the quartic has a positive discriminant.

The differential equation for $Z$, with its right-hand side proportional to the square root of the degree 4 polynomial, is not a novelty in celestial mechanics. The same form pops up while discussing the Second Fundamental Model of resonance (Henrard and Lemaître, 1983). Its solution in terms of the Weierstrass elliptic function has always been given either in the simplified form of Eq. (55), as in Ferraz-Mello (2007), or in the Biermann-Weierstrass form (Nesvorný and Vokrouhlický, 2016). We have taken an opportunity to recall other possibilities (Safford and Mordell forms) than can be of use in other applications as well.
We hope that the results of the present work will facilitate the study of perturbed Colombo top problems. They should be useful either as the kernel of analytical perturbation procedures, or as a building block of numerical integrators based upon composition methods.

Acknowledgements The work of DV was funded by the Czech Science Foundation (grant 18-06083S).

Compliance with ethical standards

Conflict of interest The authors S. Breiter, J. Haponiak and D. Vokrouhlický declare that they have no conflict of interest. The article is partially based upon the doctoral thesis prepared by J. Haponiak, but it includes the results obtained independently by the co-authors.

References

Abramowitz M, Stegun IA (1972) Handbook of Mathematical Functions, Applied Mathematics Series, vol 55. National Bureau of Standards, Washington DC

Atobe K, Ida S (2007) Obliquity evolution of extrasolar terrestrial planets. Icarus 188(1):1–17, DOI 10.1016/j.icarus.2006.11.022, \texttt{arXiv:astrop-ph/0611669}

Atobe K, Ida S, Ito T (2004) Obliquity variations of terrestrial planets in habitable zones. Icarus 168(2):223–236, DOI 10.1016/j.icarus.2003.11.017

Bianchi L (1901) Lezioni Sulla Teoria delle Funzioni di Variabile Complessa e delle Funzioni Ellittiche. Enrico Spoerri, Pisa

Biermann WGA (1865) Problemata Quaedam Mechanica Functionum Ellipticarum Ope Soluta. Carl Schulze, Berlin

Boué G, Laskar J, Kuchynka P (2009) Speed limit on Neptune migration imposed by Saturn tilting. Astrophys J Lett 702(1):L19–L22, DOI 10.1088/0004-637X/702/1/L19, \texttt{arXiv:astrop-ph/0909.0332}

Brasser R, Lee MH (2015) Tilting Saturn without tilting Jupiter: constraints on giant planet migration. Astron J 150(5):157, DOI 10.1088/0004-6256/150/5/157, \texttt{arXiv:astrop-ph/1509.06834}

Brizard AJ (2015) Notes on the Weierstrass Elliptic Function. arXiv e-prints \texttt{arXiv:math-ph/1510.07818}

Byrd PF, Friedman MD (1971) Handbook of Elliptic Integrals for Engineers and Scientists. Springer-Verlag, Berlin, Heidelberg, New York

Cassini GD (1693) De l’origine et du progrès de l’Astronomie, et de son usage dans la Geographie et dans la Navigation. In: Recueil d’observations faites en plusieurs voyages par ordre de Sa Majesté pour perfectionner l’astronomie et la geographie, l’Imprimerie Royale, Paris, pp 1–43

Colombo G (1966) Cassini’s second and third laws. Astron J 71:891–896, DOI 10.1086/109983

Copernicus N (1543) De Revolutionibus Orbium Coelestium. Johannes Petreius, Nuremberg

Correia ACM, Laskar J (2001) The four final rotation states of Venus. Nature 411(6839):767–770, DOI 10.1038/35081000

Efimov S, Priykin D, Sidorenko V (2018) Long-term attitude dynamics of space debris in Sun-synchronous orbits: Cassini cycles and chaotic stabilization. Celest Mech Dyn Astron 130(10):62, DOI 10.1007/s10569-018-9854-4, \texttt{arXiv:physics.space-ph/1712.08596}
A The roots of $W(Z) = 0$

A.1 Factorization

In order to find the zeroes of the $W(Z)$, let us first write it in the ‘classical’ polynomial form

$$W(Z) = a_0 Z^4 + 4a_1 Z^3 + 6a_2 Z^2 + 4a_3 Z + a_4,$$

(122)

where, according to Eq. (17) the coefficients are

$$a_0 = -1, \quad a_1 = 0, \quad a_2 = -\frac{2E}{3}, \quad a_3 = -2a^2 b, \quad a_4 = -4E^2 + 8a^2 E - 4a^2 (a^2 + b^2 - 1).$$

(123)

Note the absence of the cubic term ($a_1 = 0$), meaning that equation $W(Z) = 0$ is already in the reduced form.

Solving equation $W(Z) = 0$, we essentially follow a simplified and slightly reformulated procedure of Neumark [1965]. In particular, the absence of the third power of $Z$ allows the factorization

$$W(Z) = a_0 W_+ W_- (Z) = a_0 \left(Z^2 + 2\sqrt{\xi} Z + h_+\right) \left(Z^2 - 2\sqrt{\xi} Z + h_-\right),$$

(124)

with only three parameters ($\xi, h_+, h_-$), and three conditions resulting from equating the coefficients of $W$ and $W_+ W_-:

$$h_+ h_- = -a_4, \quad h_- - h_+ = \frac{4a^2 b}{\sqrt{\xi}}, \quad h_+ + h_- = 4(E + \xi).$$

(125)

The last two equations are easily solved for $h_-$ and $h_+$

$$h_\pm = \frac{2\left(E + \xi \pm \frac{a^2 b}{\sqrt{\xi}}\right)}{3p}.$$

(126)

so the first of Eqs. (125), after the substitution of Eq. (126), is actually the resolvent cubic equation

$$\xi^3 + 2E \xi^2 + a^2 (3p + 2E) \xi - a^4 b^2 = 0,$$

(127)

where $3p = 1 - a^2 - b^2$, according to Eq. (30). The Descartes rule guarantees that (for nonzero $a$ and $b$) the resolvent has at least one positive real root to be used in factorization (124).
Before we proceed to solving the resolvent, let us make some important remarks. The four roots of the quartic equation $W(Z) = 0$ come in two pairs of the roots of $W_z(Z) = 0$ and $W_{-}(Z) = 0$, i.e.

$$W_z(Z) = 0:\quad Z_1 = -\sqrt{\xi} - \sqrt{\xi - h_z}, \quad Z_2 = -\sqrt{\xi} + \sqrt{\xi - h_z},$$

$$W_{-}(Z) = 0:\quad Z_3 = \sqrt{\xi} - \sqrt{\xi - h_{-}}, \quad Z_4 = \sqrt{\xi} + \sqrt{\xi - h_{-}},$$

(128)

Let $\xi_1$ be the only, or the greatest positive root of the resolvent (127). Then, from the Vieta’s formulas, we find for the remaining two roots $\xi_2, \xi_3 = -2E - \xi_1$, and $\xi_2, \xi_3 = a_3b_4^2 \xi_1^{-1}$, which allows to see that

$$\sqrt{\xi_2} \pm \sqrt{\xi_3} = \sqrt{\xi_2 + \xi_3} \pm 2\sqrt{\xi_2 \xi_3} = \sqrt{-\xi_1 - 2E \pm \frac{2a_3 b_4}{\sqrt{\xi_1}}} = \sqrt{\xi_1 - h_z}.$$ 

(129)

This leads to the Euler form of the solution

$$Z_1 = -\sqrt{\xi_1} - \sqrt{\xi_2 - \sqrt{\xi_1}}, \quad Z_2 = -\sqrt{\xi_1} + \sqrt{\xi_2 - \sqrt{\xi_1}},$$

$$Z_3 = \sqrt{\xi_1} - \sqrt{\xi_2 - \sqrt{\xi_1}}, \quad Z_4 = \sqrt{\xi_1} + \sqrt{\xi_2 - \sqrt{\xi_1}},$$

(130)

Assuming for the real roots $0 < \xi_1 \leq \xi_2 \leq \xi_3$, we guarantee a number of properties like the ordering $Z_1 \leq Z_2 \leq Z_3 \leq Z_4$, the fact that a given trajectory contains only a pair $(Z_1, Z_2)$ or $(Z_3, Z_4)$, and that if $\xi_2, \xi_3$ are complex conjugates, then $Z_2$ and $Z_4$ remain real, whereas $Z_1$ and $Z_3$ become complex.

### A.2 Weierstrass resolvent and its roots

The cubic resolvent equation (127) can be brought to a reduced form without the square term in a number of ways. We choose the substitution based upon the seminvariant (Janson, 2011)

$$P_2 = a_1 a_2 - a_1^3 = \frac{2}{3} E,$$  

(131)

with

$$\xi = s - P_2 = s - \frac{2}{3} E.$$  

(132)

Applying it to Eq. (127), and multiplying both sides by 4, we obtain

$$S(s) = 4s^3 - g_2 s - g_3 = 0,$$  

(133)

The cubic polynomial $S(s)$ plays a special role in the theory of the Weierstrass elliptic functions, thus let us call it the Weierstrass resolvent. The symbols $g_2$ and $g_3$ that appear in Eq. (133) are the two, algebraically independent, basic invariants of the quartic $W(Z)$:

- the apolar invariant of degree 2

$$g_2 = a_0 a_4 + 3a_1^2 - 4a_1 a_3 = 12 \left( \frac{2E}{3} \right)^2 - a_1^2 \left( \frac{2E}{3} \right) - a_2^2 \rho,$$  

(134)

- the Hankel determinant of degree 3

$$g_3 = \begin{vmatrix} a_0 & a_1 & a_2 \\ a_1 & a_2 & a_3 \\ a_2 & a_3 & a_4 \end{vmatrix} = -8 \left( \frac{2E}{3} \right)^3 + 12a_2^3 \left( \frac{2E}{3} \right)^2 + 12a_3^2 \rho \left( \frac{2E}{3} \right) + 4a_4 b_2^2,$$  

(135)

known also as a catalecticant (Janson, 2011).

Notably, both the discriminants: $\Delta_z$ of the quartic $W(Z)$, and $\Delta_1$ of the cubic $S(s)$ are not only expressible in terms of $g_2$ and $g_3$, but they are equal up to a constant factor. If

$$\Delta = \left( \frac{g_2}{g_3} \right)^3 - g_3,$$  

(136)

then $\Delta_z = 16\Delta_1 = 4^3 3^3 \Delta$. 
Let the roots of the Weierstrass resolvent equation $S(s) = 0$ be $s = e_1$, $s = e_2$, and $s = e_3$. By the Vieta’s formulae, they satisfy

$$e_1 + e_2 + e_3 = 0, \quad e_1 e_2 + e_1 e_3 + e_2 e_3 = -\frac{g_2}{3}, \quad e_1 e_2 e_3 = \frac{g_3}{4}$$

(137)

and by the definition of the scaled discriminant (136)

$$(e_1 - e_2)^2 (e_1 - e_3)^2 (e_2 - e_3)^2 = \frac{27 \Delta}{16}$$

(138)

Introducing auxiliary quantities $\beta$ and $\phi$, such that

$$g_2 = 3 \beta^2, \quad g_3 = \beta^3 \cos \phi, \quad \Delta = \beta^6 \sin^2 \phi$$

(139)

hence

$$\beta = \sqrt[3]{\frac{g_2}{3}}, \quad \cos \phi = g_3 \left(\frac{3}{g_2}\right)^2$$

(140)

we can establish the universal formula for the roots [Brizard 2015]

$$e_1 = \beta \cos \frac{\phi}{3},$$

$$e_2 = \beta \cos \frac{\phi - 2\pi}{3} = -\frac{e_1}{2} + \frac{\sqrt{3}}{2} \beta \sin \frac{\phi}{3} = -\frac{e_1}{2} + \frac{\sqrt{g_2 - 3e_1^2}}{2},$$

$$e_3 = \beta \cos \frac{\phi + 2\pi}{3} = -\frac{e_1}{2} - \frac{\sqrt{3}}{2} \beta \sin \frac{\phi}{3} = -\frac{e_1}{2} - \frac{\sqrt{g_2 - 3e_1^2}}{2}.$$  

(141)

If $\Delta > 0$ (hence $g_2 > 0$) there are three simple real roots $e_1 < e_2 < e_3$ given directly by Eqs. (141). When $\Delta < 0$, there is one real root $e_1$ and two complex ones, with $e_2 = \overline{e_3}$. Equations (139) and (141) remain valid in principle, but they involve complex quantities and require distinguishing the sign of $g_2$. In these circumstances it is more convenient to use the Cardano form for the real root

$$e_1 = \frac{1}{2} \left(\sqrt[3]{g_3 + \sqrt{-\Delta}} + \sqrt[3]{g_3 - \sqrt{-\Delta}}\right).$$

(142)

and

$$e_2 = -\frac{e_1 - ie_3}{2}, \quad e_3 = \frac{e_1 + ie_3}{2},$$

(143)

with

$$e_3 = \frac{\sqrt[3]{-g_2}}{2} \left(\sqrt[3]{g_3 + \sqrt{-\Delta}} - \sqrt[3]{g_3 - \sqrt{-\Delta}}\right).$$

(144)

for the complex roots.

Finally, the degeneracy $\Delta = 0$ implies real roots: one simple and one double, or one triple root. The former case requires $g_3 \neq 0$ and $g_2 > 0$; then, according to the sign of $g_3$, either

$$e_1 = \sqrt[3]{g_3} = \frac{g_1^3}{3}, \quad e_2 = e_3 = e_{12} = -\frac{e_1}{2}, \quad \text{for } g_3 > 0$$

(145)

or

$$e_1 = \frac{\sqrt[3]{27}}{3}, \quad e_2 = e_3 = e_{12} = -\frac{e_1}{2}, \quad \text{for } g_3 < 0.$$  

(146)

The ordering of roots in (146) is exceptional ($e_3$ is the greatest), but helps to maintain a coherent notation in further applications.

The triple root $e_{123} = 0$ may appear only for $g_2 = g_3 = 0$, which is possible only when $\sigma^2 + \beta^2 = 1$.

\[2\] In this approach, $\sqrt{x}$ of a real argument $x$ is used as a real-valued function for $x < 0$, i.e. $\sqrt{-1} = -1$.  

A.3 The roots $Z_j$

Although the roots of $W(Z)$ are to be expressed in terms of the roots of $S_4(s)$, we need to include in the discussion not only the invariants $\Delta$, $g_2$, and $g_3$, but also seminvariants $P_2$ and

$$Q_2 = 2g_3^2 a_3 - 6a_0 a_1 a_2 + 4a_1^3 = -4a^2 b.$$  

(147)

This is due to the fact that although formally it is enough to substitute

$$\xi_j = e_j - \frac{2}{3} \rho = e_j - P_2,$$  

(148)

into Eq. (130), the signs of $\xi_j$ play a significant role in determining which of the roots are real and which are complex, and there are various ways to create multiple roots.

In the following discussion we will refer to the Theorem 9.3 of Janson (2011), adjusted to the different scaling of our invariants and seminvariant (namely, his $P = 48P_2$, $Q = 16Q_2$, $J = 432g_3$, and $I = 12g_2$).

A.3.1 Four simple real roots ($\Delta > 0$ and all $\xi_j \geq 0$)

If the four real roots exist, they take the form (139) with $\xi_j$ defined in Eq. (145) and $e_j$ as in (143). This requires not only $\Delta > 0$ to have three simple real roots $e_j$, but also that $\xi_j \geq 0$ for each $j \in \{1, 2, 3\}$. The latter is secured by $P_2 < 0$ and $12P_2^2 - a_0^2 g_2 \geq 0$ (Janson 2011). Substituting (151), (153) and (124), we obtain

$$\Delta > 0, \quad \text{and} \quad -\rho \leq \frac{2E}{3} < 0,$$  

(149)

as the condition for the real quadruple $Z_1 < Z_2 < Z_3 < Z_4$. If $\Delta$ is positive, but the second condition in (149) is not fulfilled, there are no real roots, and $Z_j$ form two distinct pairs of complex conjugate numbers.

A.3.2 Two simple real roots ($\Delta < 0$)

When $e_2$ and $e_3$ are complex, the pair $(Z_1, Z_3)$ is complex, whereas $(Z_2, Z_4)$ in the formula (139) formally remain real-valued, yet only by canceling the imaginary parts. Using $\xi_j = e_1 - P_2$, with $e_1$ given by Eq. (142), we can obtain $Z_1$ and $Z_2$ directly from Eqs. (129) and (130). Alternatively, we can find the expressions for

$$Z_1 = -\sqrt{\xi_1 - \sqrt{2|\xi_2|^3 - \xi_1 - 3P_2}}, \quad Z_2 = -\sqrt{\xi_1 + \sqrt{2|\xi_2|^3 - \xi_1 - 3P_2}},$$  

(150)

and

$$Z_3 = \sqrt{\xi_1 + i \sqrt{2|\xi_2|^3 + \xi_1 + 3P_2}}, \quad Z_4 = \overline{Z_3},$$  

(151)

where

$$\xi_1 = e_1 - P_2, \quad 2|\xi_2| = 2|\xi_3| = |e_1 + 2P_2 + i\rho| = \sqrt{(e_1 + 2P_2)^2 + e_2^2},$$  

(152)

with $e_1$ given by Eq. (142).

A.3.3 Multiple roots ($\Delta = 0$)

The statement $\Delta = 0$ means only that at least one of the roots is at least a double root. Further distinction is based upon the signs and values of $g_2$, $g_3$ and $P_2$. Let us inspect five possibilities involving multiple real roots from the Theorem 9.3 of Janson (2011).

A quadruple real root is not possible, because it requires $P_2 = g_2 = g_3 = 0$, whereas substituting $E = 0$ we obtain $g_3 = 4a^2 b^2 \neq 0$. Two real double roots are also impossible, because they require (among other conditions) that $Q_2 = 0$, which is not the case. The remaining three cases are the following.
1. A triple real root and one single real root appear when \( g_2 = g_3 = 0 \), and \( P_2 < 0 \). Taking the resultant of \( g_2 \) and \( g_3 \) considered as the polynomials in \( E \), one finds that both the invariants admit the common root if
\[
a^2 + b^2 = 1,
\]
the relation well known from [Henrard and Marignand 1987]. With this constraint, \( g_2 = 0 \) can be solved to give a unique negative root
\[
E = -\frac{3}{2} (a^2 b)^{\frac{2}{3}}.
\]
According to the statement below Eq. (146), \( g_2 = g_3 = 0 \) refers to the triple root \( e_{123} = 0 \), hence, with \( \xi_2 = \xi_3 = \xi_1 = -\frac{4}{3} E \), we obtain
\[
Z_1 = -3 \sqrt{a^2 b}, \quad Z_{34} = \sqrt{a^2 b},
\]
where \( Z_1 \) is the single, and \( Z_{34} \) is the triple root.

2. Two simple real roots \( Z_1, Z_2 \) and double real root \( Z_{34} \) appear when \( g_2 > 0, P_2 < 0, \) and \( 12 P_2^2 - g_2 > 0 \). So, if \( E \) is a real root of \( \Delta = 0 \) in the interval
\[
-\rho < \frac{2E}{3} < 0,
\]
then either
\[
Z_1 = -\sqrt{\xi_1} - 2 \sqrt{\xi_{23}}, \quad Z_2 = -\sqrt{\xi_1} + 2 \sqrt{\xi_{23}}, \quad Z_{34} = \sqrt{\xi_1}, \quad \text{for } g_3 > 0,
\]
or
\[
Z_1 = -\sqrt{\xi_1} - 2 \sqrt{\xi_{12}}, \quad Z_{34} = \sqrt{\xi_1}, \quad Z_4 = -\sqrt{\xi_1} + 2 \sqrt{\xi_{12}}, \quad \text{for } g_3 < 0,
\]
where \( \xi_1 = e_1 - P_2, \) and \( \xi_{ij} = e_{ij} - P_2, \) with the Weierstrass resolvent roots given by Eq. (145) or (146), according to the sign of \( g_3 \).

3. If the energy \( E \) is a real root of \( \Delta = 0 \) outside the interval \((155), i.e.
\[
\frac{2E}{3} < -\rho, \quad \text{or} \quad \left( E > 0, \quad \text{and} \quad \frac{2E}{3} \neq -\rho \right),
\]
then a double real root \( Z_{12} \) is accompanied by two simple complex roots \( Z_3 \) and \( Z_4 \). This case appears when \( e_1 = 0 \) in Eq. (145). Accordingly,
\[
Z_{12} = -\sqrt{\xi_1} = -\sqrt{e_1 - P_2},
\]
where \( e_1 \) is given by Eq. (145).

B Cassini states coordinates \( z_j \)

The left-hand side of the quartic equation \( w(z - b) = 0 \) is the polynomial
\[
w(z - b) = a'_0 z^4 + 4 a'_1 z^3 + 6 a'_2 z^2 + 4 a'_3 z + a'_4,
\]
where
\[
a'_0 = -1, \quad a'_1 = \frac{b}{2}, \quad a'_2 = \frac{1 - a^2 - b^2}{6} = \frac{\rho}{2}, \quad a'_3 = -\frac{b}{2}, \quad a'_4 = b^2.\]

Evaluating the invariants and seminvariants from the primed coefficients, we find
\[
g'_2 = \frac{3 p^2}{4} \geq 0, \quad g'_3 = -\left( \frac{\rho}{2} \right)^3 + a^2 b^2, \quad P'_2 = -\frac{\rho}{2} - \frac{b^2}{4}, \quad Q'_2 = -\frac{(1 + a^2) b}{2} < 0,
\]
ant the scaled discriminant is
\[
\Delta' = \left( \frac{g'_2}{3} \right)^3 - (g'_3)^2 - \frac{a^2 b^2}{16} (p^3 - a^2 b^3),
\]
where \( p = \sqrt{1 - a^2} \).
The transformation
\[ z = \frac{z_r + b}{2} \]
(165)
converts the equation \( w(z - b) = 0 \) into \( w(z_r - b/2) = 0 \), which is free of the \( e^j_i \) term, and so is ready for the factorization from Sect. A.1. Fortunately, we do not need to know the coefficients of the equation in \( z_r \), because we require only the invariants (163) which are conserved under the simple transformation (165). Thus, tracing backward the procedure from Appendix A we start with solving the Weierstrass resolvent
\[ 4x^3 - g_2x^2 - g_3 = 0, \]
(166)
finding the roots \( e'_1, e'_2, \) and \( e'_3 \). These define
\[ e'_j = e'_j - P'_2 = \frac{e'_j + \rho}{2} + \frac{b^2}{4}, \]
(167)
as in Eqs. (132) and (138). Finally, four roots \( z_r \) are given by Eq. (130) with \( z_i \to z_i, \xi_j \to \xi'_j \), and then \( z'_i = z_i + (b/2) \). Each real solution \( z'_i \) is the \( z \) coordinate of some Cassini state.

The number of real roots of \( w(z - b) = 0 \) depends on the sign of \( \Delta' \).

1. \( \Delta' > 0 \) is equivalent to \( a^2 + b^2 < 1 \). Three real roots \( e'_1 > e'_2 > e'_3 \) of the resolvent (166) are, by analogy with (141),
\[ e'_1 = \frac{\rho}{2} \cos \frac{\phi'}{3}, \]
\[ e'_2 = \frac{\rho}{2} \cos \frac{\phi' - 2\pi}{3} = \frac{e'_1}{2} + \frac{\sqrt{g'_3 - 3(e'_1)^2}}{2}, \]
\[ e'_2 = \frac{\rho}{2} \cos \frac{\phi' - 2\pi}{3} = \frac{e'_1}{2} + \frac{\sqrt{g'_3 - 3(e'_1)^2}}{2}, \]
(168)
where
\[ \phi' = \arccos \left( \frac{2\rho b^2}{\rho^3} - 1 \right) = 2\pi - \arcsin \left( \frac{ab}{\rho^3} \right) \]
(169)
They always define four real roots \( z'_i \), because all \( \xi'_j \) are positive. Indeed
\[ \xi'_1 = \frac{b^2}{4} + \frac{\rho}{2} \left( 1 + \cos \frac{\phi'}{3} \right) = \frac{b^2}{4} + \frac{\rho \cos \frac{\phi'}{3}}{6} > 0, \]
\[ \xi'_2 = \frac{b^2}{4} + \frac{\rho}{2} \left( 1 + \cos \frac{\phi' - 2\pi}{3} \right) = \frac{b^2}{4} + \frac{\rho \cos \frac{\phi' - 2\pi}{6}}{6} > 0, \]
\[ \xi'_3 = \frac{b^2}{4} + \frac{\rho}{2} \left( 1 + \cos \frac{\phi' + 2\pi}{3} \right) = \frac{b^2}{4} + \frac{\rho \cos \frac{\phi' + 2\pi}{6}}{6} > 0. \]
(170)
Unlike in eq. (140), we label the roots \( z'_i \) not according to their ordering in magnitude, but so that the subscript \( i \) matches the the Cassini state label \( C_j \) according to Colombo (1966), it is
\[ z'_1 = \frac{b}{2} + \sqrt{\xi'_1 + \sqrt{\xi'_2 + \sqrt{\xi'_3}}}, \]
\[ z'_2 = \frac{b}{2} - \sqrt{\xi'_1 - \sqrt{\xi'_2 + \sqrt{\xi'_3}}}, \]
\[ z'_3 = \frac{b}{2} + \sqrt{\xi'_2 - \sqrt{\xi'_1 + \sqrt{\xi'_3}}}, \]
\[ z'_4 = \frac{b}{2} - \sqrt{\xi'_2 - \sqrt{\xi'_1 + \sqrt{\xi'_3}}}, \]
(171)
with \( z'_1 < z'_3 < z'_2 < z'_1 \), as expected.

2. \( \Delta' < 0 \) means \( a^2 + b^2 > 1 \). The Weierstrass resolvent has one real root \( e'_1 \) and two complex roots \( e'_2, e'_3 \) given by Eqs. (132), (133), and (144) with \( e'_1, e'_2, e'_3, \) and \( \Delta' \). Accordingly, we obtain two real roots \( z'_i \) – the ones involving \( \sqrt{\xi'_1 + \sqrt{\xi'_2}} \), where the imaginary part cancels out. Adapting the expressions (150) and (152), and adjusting the subscripts of \( z'_i \) to the Cassini states \( C_3 \) and \( C_2 \), we obtain two real roots of \( w(z - b) = 0 \) as
\[ z'_1 = \frac{b}{2} - \sqrt{\xi'_1 - \sqrt{2[\xi'_2] - \xi'_1 - 3\xi'_3}}, \]
\[ z'_2 = \frac{b}{2} - \sqrt{\xi'_1 + \sqrt{2[\xi'_2] - \xi'_1 - 3\xi'_3}}, \]
(172)
with \( z'_3 < z'_2 \). The final substitution is made in Sect. 2.4.
\[ \Delta' = 0, \text{ hence } a^2 + b^2 = 1, \text{ implies one single and one double real root of the Weierstrass resolvent. A triple root is excluded, because } g_2' = \frac{1}{4} (ab)^{\frac{3}{2}} \neq 0. \text{ Thus, observing that } g_3' = a^2 b^2 / 8 > 0, \text{ and } \rho = (ab)^{\frac{3}{4}}, \]

\[
\begin{align*}
\epsilon_1' &= \frac{(ab)^{\frac{3}{2}}}{2}, \\
\epsilon_23' &= \frac{\epsilon_1'}{2}, \\
\xi_1' &= \frac{4a^2 b^2 + b^2}{4}, \\
\xi_{23}' &= \frac{a^2 b^2 + b^2}{4},
\end{align*}
\]

(173)

provide three Cassini states: two usual \( C_3, C_2 \), and one degenerate \( C_{14} \), with

\[
\begin{align*}
\xi_1' &= \frac{b}{2} - \sqrt{\xi_1' - 2 \sqrt{\xi_{23}'},} \\
\xi_2' &= \frac{b}{2} - \sqrt{\xi_1' + 2 \sqrt{\xi_{23}'},} \\
\xi_{14}' &= \frac{b}{2} + \sqrt{\xi_1'}. \end{align*}
\]

(174)

listed in the ascending order.