Spring 1-1-2014

Long-term Dynamical Behavior of Highly Perturbed Natural and Artificial Celestial Bodies

Aaron Jay Rosengren
University of Colorado Boulder, aaron.rosengren@colorado.edu

Long-term Dynamical Behavior of Highly Perturbed Natural and Artificial Celestial Bodies

by

Aaron J. Rosengren

B.S., Mathematics, University of Missouri, 2010

B.S., Mechanical Engineering, University of Missouri, 2010

M.S., Aerospace Engineering Sciences, University of Colorado, 2012

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Aerospace Engineering Sciences

2014

This thesis entitled:
Long-term Dynamical Behavior of Highly Perturbed
Natural and Artificial Celestial Bodies
written by Aaron J. Rosengren
has been approved for the Department of Aerospace Engineering Sciences

Daniel J. Scheeres

Philip J. Armitage

Moriba K. Jah

Jay W. McMahon

Hanspeter Schaub

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Rosengren, Aaron J. (Ph.D., Aerospace Engineering Sciences)

Long-term Dynamical Behavior of Highly Perturbed
Natural and Artificial Celestial Bodies

Thesis directed by Prof. Daniel J. Scheeres

This thesis explores the dynamical evolution of celestial bodies, both natural and artificial, which are strongly perturbed by solar radiation pressure—a non-gravitational force that has played an increasingly important role in celestial mechanics since the early 1900s. The particular focus is on the high area-to-mass ratio (HAMR) space debris discovered in near geosynchronous Earth orbit (GEO) through optical observations in 2004, and on micron-sized circumplanetary dust particles in the outer Saturnian system. The formalism developed can also be applied to—and, indeed, was unquestionably influenced by—the orbital motion of spacecraft about small bodies (asteroids and comets). The chief difficulties which arise in getting an accurate understanding of the motion of such bodies in highly perturbed dynamical environments come, in part, from the nonlinearity of the dynamical system, but more so from the inadequacy of the classical approaches and methods. While modern formulations based on numerical integrations can give "precise" solutions for specific initial conditions, these afford little insight into the nature of the problem or the essential dependence of the perturbed motion on the system parameters.

The predominant perturbations acting on HAMR objects and circumplanetary dust grains are solar radiation pressure, planetary oblateness, and third-body gravitational interactions induced by the Sun and nearby natural satellites. We developed first-order averaged models, based on the Milankovitch formulation of perturbation theory, which govern the long-term evolution of orbits subject to these perturbing forces. The unexpectedly rich results obtained by the use of this vector formalism are due to certain important circumstances in celestial and quantum mechanics which gave rise to its origin and development. An attempt has been made to trace these historical developments and to put them into the perspective of the present.

The averaged equations of motion hold rigorously for all Keplerian orbits with nonzero angular momentum; they are free of the mathematical singularities associated with circular or equatorial orbits. These approximate equations are written in a concise analytical vector form, which allow our results and demonstrations to attain such extraordinary simplicity and clarity. As a first attempt to understand the disturbed motion, we consider separately the first-order effects of each principal perturbation in altering the orbital elements. We establish that each of these problems is integrable (under certain well-justified assumptions), and that they admit either an exact analytical solution or a complete qualitative description. We then explore the complex interplay between gravitational and non-gravitational perturbations, and examine stable "frozen" orbit configurations and resonances which can occur when these forces act in concert.

These results are applied to the study of the dynamics and stability of GEO orbits and to the identification of robust, long-term disposal orbits for geostationary satellites. We further apply these results towards an understanding of the initial albedo dichotomy of Iapetus, Saturn's enigmatic two-faced satellite. As dust is a ubiquitous component of the Solar System, we discuss the application of our averaged model to a number of other planetary systems. We highlight, in particular, the explanatory power of our approach and how it can modify or guide detailed numerical studies. Indeed, analytical and numerical techniques cannot be separated, and a complete, logically ordered picture is obtained only by the application of both methods jointly.

Dedication

I dedicate this thesis to two women: Glory, my mother and my first teacher, not because she did all the things that good mothers do, but because she does so much more; and Diliana, whose love, support, and encouragement, have kept me stable in the chaotic dynamics of graduate school.

Acknowledgements

During my long academic journey—nine years, as it turns out—I have incurred many debts and obligations, both personal and intellectual, which I now have great pleasure in acknowledging. To my advisor, Dan Scheeres, an encyclopedia of astronomical insight, I owe more than I can say for kindling and sustaining my active interest in research. This undertaking has been, in every aspect, a joint endeavor. On a more personal note, Dan has taught me that despite the great power and scope of modern computers, there is still a place for pen-and-paper calculations in the style of our forebears. During the slow ripening of this thesis, I have had the good fortune of innumerable discussion with my committee: Phil Armitage, Moriba Jah, Jay McMahon, and Hanspeter Schaub. I am grateful to Dr. Schaub, who taught me the importance of fundamentals, and that insight can often be obtained through simplicity; and to Dr. Jah, my mentor at the Air Force Research Laboratory, who introduced me to the HAMR debris problem and showed me that insight without accuracy is meaningless. I am especially indebted to Thomas Schildknecht, of the University of Bern, for permission to display his observational data and for many interesting and motivating conversations. I am also deeply indebted to the referees, who gave generously of their time for reviewing my journal articles, thereby enhancing substantially the readability of this work. I wish to especially acknowledge Luis Floría, of the University of Zaragoza, for the painstaking care with which he reviewed the 'Milankovitch' manuscript. A deep debt of gratitude is extended to the University of Colorado Interlibrary Loan and Document Delivery Request staff for their assistance and kindness in procuring many foreign manuscripts and articles out of circulation. I would also like to thank the bright and creative members, both past and present, of the Celestial and Spaceflight

Mechanics Laboratory. I am particularly grateful to Dr. Kohei Fujimoto, chief among them, for being an outstanding conference 'wingman' and mentor. I am also deeply and variously indebted to Steven Hart and Sarah Melssen, of the Colorado Center for Astrodynamics Research, and to my fellow graduate students, Fred Estante, Khashy Parsay, and Luis Zea notable among them. My sincerest thanks and gratitude to Ann Smead and Michael Byram for their invaluable support through the H. Joseph Smead Fellows Program. This research was also supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1144083, and from Grant FA9550-11-1-0188, administered by the Air Force Office of Scientific Research.

I also owe thanks to a host of mentors, professors, and colleagues at the University of Missouri, who provided me with the basic background on which to build: Ian Aberbach, Dario Cersosimo, Craig Kluever, Stephen Montgomery-Smith, Carlo Morpurgo, and P. Frank Pai. Professor Kluever and his former student, Dr. Cersosimo, not only introduced me to the field of aerospace engineering, but were instrumental in my decision to attend graduate school at CU under the guidance of Dan. My first foray into the vicinity of scientific research was through the McNair Scholars program and I should not fail to mention the splendid assistance I received from NaTashua Davis, Jeremy Bloss, and my mentor, Professor Pai. I am especially indebted to Professors Aberbach, Montgomery-Smith, and Morpurgo, who showed the most heroic reserves of patience in answering my often trivial mathematics questions. I cannot begin to thank adequately those friends in undergrad who helped me to survive a slew of engineering, physics, and mathematics courses. Among those to whom I am particularly grateful are Joseph Ayala, Chris Barr, Chris Brines, Daniel Coombs, Kathleen Frese, Jason Huffman, Steven McKee, Rob Nguyen, John Wagner, and Matt Wavada. Though I do not regard myself an even vaguely competent computer programmer, I owe much to John and Steve for helping me to ferret out insidious errors in my codes and mental logic. I must also acknowledge my friends, DeLon Hill, Jeremy Latty, Nate Massey, Matt Record, Ryan Sweeney, and Tim Wilson, and my family for their lasting encouragement. Above all, and as always, my profoundest thanks to Diliana Stoimenova, whose assiduous reading of various parts of this thesis was but the most tangible expression of her unflagging support.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1 Celestial Mechanics and Highly Perturbed Dynamical Systems

Celestial mechanics is essentially an application of the laws of mechanics to the study of the motions of celestial bodies, both translational and rotational, that result from and are controlled by gravitational and non-gravitational forces. In strictness the subject dates from antiquity when the early Greek philosophers observed and noted the simple and regularly recurring astronomical phenomena; in the modern scientific sense, however, the birth of celestial mechanics may be dated from the epoch when Newton deduced Kepler's laws of planetary motion from his theory of gravitation and his fundamental laws of motion. The long-term orbit evolution of the planets and natural satellites is one of the oldest subjects of investigation in celestial mechanics. The detailed motions of these bodies were more complicated than the simple picture portrayed by Kepler, but they could be explained as being due to gravitational perturbations from other bodies.

During the eighteenth century, Euler, Clairaut, d'Alembert, Lagrange, and Laplace brought increasing mathematical sophistication and analytical insight to bear on the problems of celestial mechanics. In seeking to solve what came to be known as the three-body problem, Euler, Clairaut, and d'Alembert undertook to develop general analytical methods for determining the perturbative effects caused by the mutual actions of the celestial bodies on one another. The crowning glory of mid-eighteenth-century astronomy was the demonstration, by Clairaut, that the inverse-square law of gravitation established by Newton could account, in a precise mathematical way, not only for the planets's apparent elliptical motion, but also for the complicated deviation of the Moon's

motion from this ideal orbit. The mathematical analysis marshaled and brought to bear in this development was applied by Clairaut, in the first large-scale numerical integration ever performed, to the perturbations of Halley's comet from its close encounters with the planets Jupiter and Saturn. Clairaut's accurate prediction of the perihelion date for the return of Halley's comet was without precedent and gave striking proof of the value of perturbation theory.

The theory of planetary motion was put on a firm foundation by Euler and Lagrange, who demonstrated that the apparent acceleration of Jupiter and deceleration of Saturn were not secular (increasing linearly with time), but periodic over timespans of centuries. Laplace, Lagrange, and Poisson, in considering the stability of the Solar System, showed that the semi-major axes of the planets are secularly invariant (to second order in the masses of the planets). The most proclaimed success of perturbation theory occurred in the middle of the nineteenth century with the dramatic mathematical prediction of the existence of Neptune by Adams and Le Verrier, on the basis of the unexplained irregularities in the motion of Uranus. The close of the nineteenth century marked an epoch distinguished by new perspectives and rigorous mathematical methods. It was then that Poincaré, whose contributions to celestial mechanics were of the most brilliant character, founded the qualitative theory of differential equations.[1]

During the Sputnik era there came a resurgence of interest in celestial mechanics, impelled forward by the prospect of the scientific exploration of interplanetary space, and greatly aided by the advent of sophisticated computers. Since many of the problems in the mechanics of space flight differ so markedly from the classical problems presented by nature, the classical methods often proved inadequate or had to be reconsidered in a new light, and new formulations had to be devised to handle novel situations. This extension and enrichment of celestial mechanics in relation to space age astronomy has tended to crystalize and develop into the subfield of astrodynamics, emphasizing the engineering and experimental aspects of dynamical astronomy.

[1] Dziobek (1892) and Moulton (1914) give a comprehensive account of the developments of perturbation theory up to the end of the nineteenth and beginning of the twentieth century. Understanding of how the theory evolved is greatly facilitated by the numerous historical sketches that accompany each main section. A more detailed historical account of the major endeavors through which celestial mechanics has evolved and of the observational discoveries through the eighteenth and nineteenth centuries can be found in Taton & Wilson (1995).

The application of a thoroughly grounded mathematical science, which is soundly embedded in hundreds of years of tradition, to the problem of highly perturbed natural and artificial celestial bodies seems at first to be superfluous. Indeed, celestial mechanics or astrodynamics is often considered to be a mature field. However, the unsolved problems in celestial mechanics are very numerous: in fact, the only part of celestial mechanics that has been solved with all the rigor, generality, and elegance that may be desired is the motion of two (point-mass) bodies. The main objects of study in this research are the long-term evolution of high area-to-mass ratio (HAMR) space debris in high-Earth orbits and the dynamical behavior of micron-sized dust particles in circumplanetary orbits. The key distinguishing feature of these bodies is their susceptibility to the effects of solar radiation pressure, and hence their dramatic variations in eccentricity and inclination over short timespans. The interplay between the gravitational and non-gravitational perturbations gives rise to interesting dynamical aspects; here the classical methods are not very effective. This naturally leads to the following thesis statement:

Thesis Statement

> The fundamental mechanics of motion of celestial bodies in highly perturbed dynamical environments are not fully understood and cannot be understood based solely on the application of classical perturbation theories and methods or from modern formulations based on numerical integrations alone. With the non-singular Milankovitch orbital elements, many problems are effectively treated and desirable abbreviation in the perturbation equations is obtained. Secular perturbation theory in terms of the Milankovitch elements provides a unified approach to the analysis and simulation of celestial bodies over long timespans and allows for the qualitative nature of their evolution to be understood.

1.2 A Note on Perturbation Theory

In general, celestial mechanics arrives at its results by making use of the most far-reaching simplification; namely, by treating all bodies as if they moved, in the first approximation, imperturbably on their Keplerian paths. Indeed, under the conditions prevailing in the Solar System, the motion of every celestial body is so decisively governed by the attraction of the central body that it does not depart appreciably from purely Keplerian orbits. The presence of the other members

of the system, or any irregularity in the shape and density of the bodies, is made manifest by the rather small and slowly changing perturbations they induce on this idealized pattern. Some very ingenious methods have been devised for computing and describing these perturbations. The principle upon which we shall study the effects of perturbations, and which became de rigueur with the works of Lagrange and Laplace, is known by the name of variation of parameters (§2.2). Although the curve described by the disturbed body is not strictly an ellipse, or for that matter, any regular formed curve, it may be represented by an ellipse provided we suppose the elements of the ellipse to change perpetually; hence why all discussions in celestial mechanics rest upon the solution of the two-body problem (q.v., Appendix B). We are thus led to the Lagrangian form of the equations of motion in celestial mechanics. A noteworthy feature of these equations is the use of a potential function in deriving the perturbative forces in the coordinate directions. The negative potential function, called the disturbing or perturbation function by Laplace, plays a fundamental role in celestial mechanics, giving rise to the notion of separation of perturbing effects into periodic and secular variations and the distinction between fast and slow time variables.

Classically, approximate analytical solutions of the Lagrange equations were sought in terms of developments in trigonometric series with respect to the small parameters of the problem (the masses, for instance); successive powers of these parameters figured as factors in the coefficients of the series, the arguments being functions of time (see, for instance, Kurth 1959, Chap. 3, §2). The procedure of integration brought to light a very remarkable class of perturbations; that is, the appearance of perturbations of seemingly secular character. These perturbations result from terms that contain the time variable explicitly, and not only in the arguments of sine and cosine, and seem to indicate that the solution increases without bound in the course of time. The fact that the general solution involves power series in the time suggests, however, that these secular terms are merely the first terms in the series representation and may actually be equivalent to terms of very long period. They must therefore be considered as only a first approximation to the long-term motion. "We can nevertheless," as Brouwer & Clemence (1961b, p. 48) point out, "learn something of a more general nature about planetary motions by limiting consideration to the secular perturbations alone[2]

alone." Poincaré (1892) demonstrated that, although we can theoretically obtain the real solution as a trigonometric series, such series are not uniformly convergent. Nevertheless, these expansions permit us to have solutions approximating as closely as we please to the real solution during a finite time interval, and therefore make possible a very accurate description of the perturbed motion.

In view of the great importance of osculating (or instantaneous) elements as variables in perturbation theory, we shall discuss a problem which is intimately connected and has been one of the main subjects in celestial mechanics; namely, the selection of quantities best suited to describe the disturbed motion. The application of the method of variation of parameters is naturally not restricted to the use of the classical orbital elements (q.v., §2.2; see Appendix B.2.3 for specific details on these elements). Although these elements possess the advantage of having a precise and simple geometrical significance, troublesome calculations would be necessary in order to derive the differential equations for their perturbations. The equations obtained would, moreover, not clearly disclose the characteristic features of the problem or give the best insight into the nature of the motion (Milankovitch 1941; Finlay-Freundlich 1958). The perturbed motion becomes much more intelligible if we describe it by the changes of the elements that arise naturally in the dynamical solution of the two-body problem; namely, the two vectorial first integrals of motion (i.e., the Milankovitch orbital elements; see §2.1). Establishing the system of differential equations they obey for the various perturbing forces is one of the chief subjects of this thesis.

1.3 The Principal Perturbations

The predominant perturbations acting on high area-to-mass ratio space debris and circum-planetary dust particles are solar radiation pressure, planetary oblateness, and third-body gravitational interactions induced by the Sun and nearby natural satellites (i.e., the Moon for HAMR objects in high-Earth orbits). Other perturbations, such as those arising from higher-order gravity field distributions, the planetary magnetic field, and the planet's atmosphere, inter alia, are usually quite small for distant orbits and will be neglected in this discussion.

[2] The theory of secular perturbation, or method of averaging, will be discussed in greater detail in §2.3.

1.3.1 Solar Radiation Pressure

Kepler, whose chief original contribution to astronomy was the discovery of the three laws of planetary motion which bear his name, also made a discerning observation and stimulating conjecture in 1619 about the mechanical nature of light. In his treatise, 'De Cometis', Kepler observed that the tail of a comet in the neighborhood of perihelion always streams out from the head in a direction which is almost exactly away from the Sun, as though blown by a strong wind emanating from the Sun (Jones 1953). Kepler attributed this repulsive force to the pressure exerted by the Sun's rays upon the matter vaporized in the comet's head, the experimental confirmation of which did not come for nearly three centuries. The view expressed by Kepler about light-pressure, or solar radiation pressure (SRP) as it is now known, was given a firm theoretical basis by Maxwell (1873, Art. 792), as a necessary consequence of the electromagnetic theory of light, and by Bartoli (1884), as essential to the validity of the second law of thermodynamics. The Maxwell-Bartoli theory was quantitatively confirmed by Lebedew (1902) and Nichols & Hull (1903) through direct laboratory experiments (see references therein if precise dates are wanted).

Although under terrestrial conditions the pressure of light is only of trivial importance, it has assumed great significance in astronomy and astrophysics (see, e.g., Burns, Lamy & Soter 1979; Armitage & Pringle 1997) and, more recently, in space flight mechanics with the development of solar sail propulsion systems for use in the robotic exploration of the Solar System (McInnes 1999). Radiation pressure is fundamental in the consideration of the structure of stars as a factor in their hydrostatic equilibrium, as was elucidated by Eddington. For stars of sufficient mass, Eddington (1916/17) shows that the radiation pressure outwards at the surface of a star may nearly neutralize the gravitational pull inwards, thereby setting an upper limit to their size. The dynamical effects of solar radiation pressure in the Solar System have been rigorously studied since the early 1900s (qq.v., Poynting 1904; Plummer 1905; Robertson 1937). This non-gravitational perturbation plays a vital role in the evolution of dust particles in interplanetary and circumplanetary orbits (Gustafson 1994; Hamilton 1996), as well as in the orbital motion about asteroids and comets (e.g., Richter & Keller

1995; Scheeres 2012a,b). It also acts on the orbits of dust grains through the Poynting-Robertson (P-R) drag effect (Wyatt & Whipple 1950), and on the orbits and rotation states of small bodies through the Yarkovsky and YORP effects (Rubincam 2000). Solar radiation pressure is, moreover, the largest non-gravitational perturbative force to affect the motion of artificial satellites and space debris in high-Earth orbits (Musen 1960; Allan 1962; Shapiro 1963; Friesen et al. 1992).

The physical origin of solar radiation pressure is the process of exchange of energy and momentum between fields and particles. Photons can be treated as quantum-mechanical particles that carry momentum and therefore impart a pressure when they interact with matter. The radiation pressure of the incident sunlight, assuming that the Sun acts as a point source, is proportional to the radiative flux and inversely proportional to the square of the distance from the Sun:

$$P(d_s) = \frac{P_\Phi}{d_s^2}, \tag{1.1}$$

where P_Φ is the solar radiation constant ($\sim 1 \times 10^8$ kg km^3/s^2/m^2), defined by the ratio of the radiant energy to the speed of light, and d_s is the distance between the central body (henceforth referred to, more simply, as the planet) and the Sun. As described by Scheeres (2012b, §2.6.2), the most relevant description of the solar photon force accounts for the momentum of the impinging photons on the body, the recoil momentum of those photons both specularly reflecting and diffusely scattering from the body, and the balance between the influx of photons, the temperature of the body, and the outflow of the re-emitted photons. Accounting for the total momentum transfer of the solar photons striking and recoiling off the surface element of a general body, the solar radiation force acting on a unit area A can be specified in dyadic notation[3] as (cf. Scheeres 2012b)

$$\boldsymbol{F}_i = -P(d_s)\left[\{\rho s(2\hat{\boldsymbol{n}}\hat{\boldsymbol{n}} - \boldsymbol{U}) + \boldsymbol{U}\} \cdot \hat{\boldsymbol{u}}\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{n}} + \{B(1-s)\rho + (1-\rho)B\}\,\hat{\boldsymbol{n}}\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{u}}\right] H(\hat{\boldsymbol{u}})A, \tag{1.2}$$

where ρ is the total reflectance or albedo of the body, s is the fraction of specularly reflected light, $\hat{\boldsymbol{n}}$ is the unit normal to the surface, $\hat{\boldsymbol{u}} = \hat{\boldsymbol{d}}_s - \hat{\boldsymbol{r}}$ is the unit vector from the surface to the Sun, $\hat{\boldsymbol{d}}_s$

[3] An expression $\boldsymbol{ab}$, formed by the juxtaposition of two vectors and written without a dot or cross between, constitutes a dyad, and is equivalent to the outer product: $\boldsymbol{ab} \equiv [\boldsymbol{a}][\boldsymbol{b}]^T$. The identity dyadic $\boldsymbol{U}$ has the general property $\boldsymbol{U} \cdot \boldsymbol{a} = \boldsymbol{a} \cdot \boldsymbol{U} = \boldsymbol{a}$. Appendix A gives a synopsis of the basic notation and properties of dyadics, which are used to advantage in nearly all parts of this research for the sake of brevity and clarity. A systematic development of dyadic algebra is given in Brand (1947, Chap. 4) and Gibbs & Wilson (1909, Chap. 5).

and $\hat{\boldsymbol{r}}$ are the position unit vectors of the Sun and object relative to the planet, respectively, $\boldsymbol{U}$ is the identity dyadic, B is a scattering coefficient that describes the fraction of light scattered normal to the surface (equal to 2/3 for an ideal Lambertian surface), and $H(\boldsymbol{u})$ is the visibility function for the surface and is equal to 1 when the Sun is in view and 0 otherwise. The net acceleration due to the solar photons on a body of mass m can be written in the general form

$$\boldsymbol{a}_{srp} = \frac{\beta'}{d_s^2}\hat{\boldsymbol{a}}, \tag{1.3}$$

in which $\hat{\boldsymbol{a}} = \sum_{i=1}^{N} \boldsymbol{F}_i / |\sum_{i=1}^{N} \boldsymbol{F}_i|$ is the net direction of the acceleration and

$$\beta' = \frac{P_\Phi |\sum_{i=1}^{N} \boldsymbol{F}_i|}{m}. \tag{1.4}$$

The simplest and most widely used model for computing solar radiation pressure acceleration is known in astrodynamical parlance as the cannonball model, which treats the body as a sphere with constant optical properties. The total momentum transfer from the incident solar photons is modeled as insolation plus reflection, and the force generated is independent of the body's attitude. Any force component normal to the object-Sun line that results from an aspherical shape or nonuniformly reflecting surface is thereby neglected. Then the net acceleration will act in the direction directly away from this line and have the general form (Scheeres 2012b)

$$\boldsymbol{a}_{srp} = -(1+\rho)(A/m)P_\Phi \frac{(\boldsymbol{d}_s - \boldsymbol{r})}{|\boldsymbol{d}_s - \boldsymbol{r}|^3} \tag{1.5}$$

$$= -\beta \frac{(\boldsymbol{d}_s - \boldsymbol{r})}{|\boldsymbol{d}_s - \boldsymbol{r}|^3}, \tag{1.6}$$

where $\beta = (1+\rho)(A/m)P_\Phi$, A/m is the appropriate cross-sectional area-to-mass ratio[4] in m^2/kg, and we ignore the possible effect of the object passing through the planet's shadow (q.v., Allan 1962). We also neglect effects depending on the wave properties of light that are known to be significant for very small particles (sizes on the order of the wavelength of the incident light; see Burns, Lamy & Soter 1979, and references therein).

[4] For example, $A/m = 0.75/\rho_s s$ for spherical particles of uniform density ρ_s and radius s.

This solar radiation pressure model can be rewritten as a disturbing function (the negative potential function of the disturbing acceleration)

$$\mathcal{R}_{srp} = -\beta \frac{1}{|\boldsymbol{d}_s - \boldsymbol{r}|}, \tag{1.7}$$

where $\boldsymbol{a}_{srp} = \partial \mathcal{R}_{srp} / \partial \boldsymbol{r}$. If the object is close to the planet, or $r \ll d_s$, the disturbing function can be simplified by expanding $1/|\boldsymbol{d}_s - \boldsymbol{r}|$ and keeping the first term that contains the position vector:

$$\mathcal{R}_{srp} = -\frac{\beta}{d_s^3} \boldsymbol{d}_s \cdot \boldsymbol{r}, \tag{1.8}$$

with the gradient giving an acceleration independent of the object's position relative to the planet:

$$\boldsymbol{a}_{srp} = -\frac{\beta}{d_s^2} \hat{\boldsymbol{d}}_s. \tag{1.9}$$

For a given semi-major axis a, reflectivity ρ, and A/m value, we define the SRP perturbation angle (first defined by Mignard & Hénon 1984) as

$$\tan \Lambda = \frac{3\beta}{2 V_{lc} H_s}, \tag{1.10}$$

in which $V_{lc} = \sqrt{\mu/a}$ is the local circular speed of the object about the planet, μ the being planet's gravitational parameter and a the object's semi-major axis, $H_s = \sqrt{\mu_s a_s (1 - e_s^2)}$ is the specific angular momentum of the planet about the Sun, and μ_s is the gravitational parameter of the Sun, assumed to be in an elliptic orbit with semi-major axis a_s and eccentricity e_s. We note that as the SRP perturbation becomes strong, $\Lambda \to \pi/2$; and as it becomes weak $\Lambda \to 0$. The angle Λ can be used to rigorously characterize the strength of the SRP perturbation acting on a body as a function of its orbit, its non-gravitational parameter, and the heliocentric orbit of the planet. As it combines these quantities into a single parameter, which completely defines the secular SRP-only solution (qq.v., Richter & Keller 1995; Scheeres 2012b), we find it efficacious to use as the fundamental defining characteristic of SRP dominated bodies.

1.3.2 Planetary Oblateness

No planetary body is precisely spherical; consequently, further terms appear in their gravitational potential, in addition to the basic term representing spherically symmetric (Keplerian)

attraction. The true shape of the planets, however, can be approximated as oblate (i.e., flattened) spheroids—the equilibrium shape of homogeneous uniformly rotating masses under the combined influence of gravity and centrifugal forces, as was inferred by Newton during his investigation on the figure of the Earth (Taton & Wilson 1995). Thus, the flattened planets, while no longer spherically symmetric, retain axial symmetry about their rotation axes.

The study of the motion of a satellite in the field of an oblate spheroid dates back to Laplace, as it seems with all fundamental problems in celestial mechanics. In the Fifth Book of his monumental 'Mécanique Céleste', Laplace treats the effects of the oblateness of Saturn on its rings and proves that "the action of Saturn, arising from the oblateness of its form, constantly retains the rings, so as to keep them nearly in the plane of its equator; and the different rings of Saturn are by this cause retained in the same plane (Laplace 1799, §21)." He further asserts that, as a ring may be considered as an aggregation of satellites, the same holds true for the plane of the orbits of the satellites that nearly coincide with the planet's equator (i.e., all known satellites at that time except Iapetus, which is treated separately in his Eight Book).

In the brief span of time after the launch of Sputnik, a whole succession of analyses was devoted to the problem poised by the drag-free motion of an artificial satellite about an oblate planet, employing almost every known perturbation method. Although in a sense, as we have seen, the problem is a classic one that also occurred among the natural satellites, it was necessary in the applications of artificial satellite motion to obtain a more general, detailed, and accurate solution. The most intricate and notable investigations were presented by Kozai (1959), Brouwer (1959), and Garfinkel (1959) in the renowned 1959 issue of the Astronomical Journal.

For a planet possessing axial symmetry, the multipole expansion for its gravitational force-function at a point external to it has the form (Roy 2005)

$$\mathcal{U} = \frac{\mu}{r}\left[1 - \sum_{i=2}^{\infty} J_i \left(\frac{R}{r}\right)^i P_{i,0}(\sin\delta)\right], \tag{1.11}$$

where r is the radius, δ is the latitude angle measured from the equatorial plane, R is the mean equatorial radius of the planet, the coefficients J_i are constants that depend on the distribution of

mass within the planet, and the functions $P_{i,0}$ are known as Legendre polynomials.[5] If the satellite is far enough from the planet, then a reasonable approximation is to truncate the series after the quadrupole moment J_2 (often referred to as the dominant oblateness gravity field coefficient). Evaluating this term explicitly, with the Legendre polynomial $P_{2,0}(x) = 1/2(3x^2 - 1)$, the quadrupole disturbing function arising from an oblate planet can be written as

$$\mathcal{R}_2 = \mathcal{U}_2 - \frac{\mu}{r} \tag{1.12}$$

$$= \frac{\mu J_2 R^2}{2r^3}\left(1 - 3\sin^2\delta\right). \tag{1.13}$$

We can state the disturbing function in a general vector expression

$$\mathcal{R}_2 = \frac{\mu J_2 R^2}{2r^3}\left[1 - 3(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{p}})^2\right], \tag{1.14}$$

where $\hat{\boldsymbol{p}}$ is a unit vector aligned with the maximum axis of inertia of the planet (i.e., the planet's rotation pole). The perturbing acceleration is then (cf. Ward 1962)

$$\boldsymbol{a}_2 = \frac{\partial \mathcal{R}_2}{\partial \boldsymbol{r}} = -\frac{3\mu J_2 R^2}{2r^4}\left\{\left[1 - 5(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{p}})^2\right]\hat{\boldsymbol{r}} + 2(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{p}})\hat{\boldsymbol{p}}\right\}. \tag{1.15}$$

For outer satellites at relatively large planetocentric distances, the gravitational potential due to the inner satellites and the rings can be accounted for by augmenting the planet's effective oblateness (Ward 1981; Tremaine, Touma & Namouni 2009). Assuming the inner satellites are on circular orbits in the equatorial plane of the planet with semi-major axes a_i and masses m_i, the quadrupole moment of the planet may be enhanced as (Tremaine, Touma & Namouni 2009)

$$J_2' R^2 \equiv J_2 R^2 + \frac{1}{2}\sum_{i=1}^{n} a_i^2 m_i / M_c, \tag{1.16}$$

where M_c is the planet's mass.

[5] The Legendre polynomials are defined by

$$P_{i,0} = \frac{1}{2^i i!}\frac{\mathrm{d}^i}{\mathrm{d}x^i}(x^2 - 1)^i.$$

1.3.3 Third-Body Gravitation

The most interesting and long-standing problem of celestial mechanics is that of determining the perturbing effects due to the gravitational force of bodies other than the central mass; more formally known as third-body perturbations. The Moon, its motion around the Earth disturbed by the gravitational attraction of the exceedingly large and relatively near Sun, presented one of the most complex problems within the Solar System (Goldreich 1966); "the development of [lunar] theory represents the most dramatic chapter in the history of Newton's theory of gravitation (Finlay-Freundlich 1958, p. 97)." The third-body perturbation problem has also occupied a prominent place in modern celestial mechanics (see, e.g., Musen 1961; Allan 1962; Cook 1962; Kaula 1962; Lidov 1962; Vashkov'yak 1974; Farago & Laskar 2010; Naoz et al. 2013).

Taking the planet as the center of our dynamical system, the perturbation acceleration from a body with gravitational parameter μ_p is (Kurth 1959; Scheeres 2012b)

$$\boldsymbol{a}_p = -\mu_p \left[\frac{\boldsymbol{r} - \boldsymbol{d}_p}{|\boldsymbol{r} - \boldsymbol{d}_p|^3} + \frac{\boldsymbol{d}_p}{|\boldsymbol{d}_p|^3} \right], \tag{1.17}$$

where $\boldsymbol{d}_p$ is the position vector of the perturbing body relative to the planet. It is important to note that the second term, which is independent of the dynamical variable $\boldsymbol{r}$, results from the fact that the coordinate system centered in the planet is not an inertial system.

For use in perturbation analysis it is convenient to recast this as a disturbing function

$$\mathcal{R}_p = \mu_p \left[\frac{1}{|\boldsymbol{r} - \boldsymbol{d}_p|} - \frac{\boldsymbol{d}_p \cdot \boldsymbol{r}}{|\boldsymbol{d}_p|^3} \right], \tag{1.18}$$

where $\boldsymbol{a}_p = \partial \mathcal{R}_p / \partial \boldsymbol{r}$. If the satellite's distance from the planet is small compared to the distance between the planet and the third body, or $r/d_p \ll 1$, the disturbing function can be represented as an infinite series using the Legendre expansion, resulting in (Scheeres 2012b)

$$\mathcal{R}_p = \frac{\mu_p}{d_p} \left[\sum_{i=0}^{\infty} \left(\frac{r}{d_p} \right)^i P_{i,0} \left(\frac{\boldsymbol{r} \cdot \boldsymbol{d}_p}{r d_p} \right) - \frac{\boldsymbol{d}_p \cdot \boldsymbol{r}}{d_p^2} \right]. \tag{1.19}$$

The first term is independent of $\boldsymbol{r}$ and can be dropped. Also, as the Legendre polynomial $P_{1,0} = x$, the $i = 1$ term is canceled by the term outside the summation that results from the fictitious

acceleration of the planet's center of mass by the perturbing body. Thus, the relevant perturbation terms begin at the quadrupole ($i = 2$) term. Retaining only this term yields the quadrupole disturbing function

$$\mathcal{R}_p = \frac{\mu_p}{2d_p^3}\left[3(\boldsymbol{r}\cdot\hat{\boldsymbol{d}}_p)^2 - r^2\right]. \tag{1.20}$$

Thus, to lowest order, the gravitational attraction of the perturbing body can be represented as a quadratic form, which is the fundamental approximation made in the Hill lunar problem (Brouwer & Clemence 1961a). Under this approximation, the perturbing acceleration simplifies to

$$\boldsymbol{a}_p = \frac{\mu_p}{d_p^3}\left[3(\boldsymbol{r}\cdot\hat{\boldsymbol{d}}_p)\hat{\boldsymbol{d}}_p - \boldsymbol{r}\right]. \tag{1.21}$$

The third-body gravitational attraction also acts on the equatorial bulge of the planet, causing a slow rotation of the planter's spin axis $\hat{\boldsymbol{p}}$. It is precisely this effect that produces the precession of the equinoxes, a phenomenon known since antiquity, but not understood until Newton's time (Taton & Wilson 1995). The gravitational action of the Sun and Moon on the Earth's equatorial bulge causes the Earth's rotation axis to precess about the ecliptic's pole with a period of nearly 26,000 years, so that the nodes of the equatorial plane, the equinoxes, move in the retrograde direction. This axial precession along with the cyclical changes in Earth's orbit—its length and shape, and the rhythmic shifts in its angle of orientation to the Sun—might have precipitated Ice Ages (Milankovitch 1941), as discussed in Section 2.1.1.

1.4 Original Contributions and Brief Summary of Chapters

> *In the several hundred year old subject of celestial mechanics, a reference to a seemingly new idea almost always can be located in the early literature. The point is, of course, what happens to the development of the idea.*[6]
>
> — Victor G. Szebehely, 1964

This thesis explores the dynamical evolution of celestial bodies, which are strongly perturbed by solar radiation pressure.[7] The particular focus is on the high area-to-mass ratio space debris

[6] Szebehely VG, ed. 1964. *Celestial Mechanics and Astrodynamics.* New York: Academic Press, pp. xi–xii

[7] The papers upon which this thesis is based are listed in the bibliography: Rosengren & Scheeres (2013, 2014a,b); Rosengren, Scheeres & McMahon (2014).

discovered in near geosynchronous Earth orbit (GEO) through optical observations by Schildknecht et al. (2004), and on micron-sized circumplanetary dust particles with an emphasis on the Saturnian-Sun system. The orbital motion of spacecraft about small bodies (asteroids and comets) and planetary satellites, and the long-term dynamics of these celestial bodies as well as extrasolar planetary systems, inter alia, can also be treated within our framework; from the point of view of mechanics, only the emphasis and nature of the forces involved may differ.

Most modern studies on the long-term dynamics of highly perturbed bodies, both natural and artificial, concentrate on numerical integration of the precise set of differential equations. This process leads to the construction of accurate ephemerides and the investigation of the empirical evolution of such objects, but not necessarily to general insight into the nature of the problem or the essential dependence of the perturbed motion on the system parameters. The efficient use of the computer, moreover, depends in a highly important way on an understanding of these dependencies, and on a careful preliminary selection of the force models to be integrated. Our purpose is thus to adopt the simplest possible expressions useful for studying the long-term evolution of these celestial bodies. These equations must reveal the qualitative regularities of motion, and they must provide a way of obtaining quantitative predictions of long-term changes. This thesis presents some of the analyses and research investigations being performed in support of these goals.

Chapter 2 opens with an explanation of the vector formulation of perturbation theory, which leads naturally to a description of the perturbed motion in terms of the angular momentum and eccentricity vectors—the Milankovitch orbital elements (Milankovitch 1939). These elements possess a geometric significance clearer than that which can be deduced from the classical orbital elements, or the various alternative sets employed in celestial mechanics; hence the effect of the perturbation on the orbit can be seen immediately. The use of these vectorial elements for perturbed Keplerian motion revives the physical insight which has often been lost in classical accounts under the burden and undesirable complexity of routine and laborious astronomical calculations. After a detailed discussion of the origin and development of such elements, drawing in particular on some historical sidelights on quantum mechanics, we present two new sets of orbital elements, which are based on

the Milankovitch elements, but have wider scope and generality. In §2.2, we derive the perturbation equations of our Milankovitch-like element sets in Lagrangian form, assuming the perturbing accelerations are expressible as gradients of a disturbing function, and in Gaussian form when the acceleration is arbitrary. The perturbation equations for the first element set hold rigorously for all Keplerian orbits with nonzero angular momentum; the second set is free of singularities associated with zero eccentricity and vanishing line of nodes. The main results are given by Propositions 1, 2, and 3. After averaging (§2.3), a compact, elegant, and symmetrical form of secular Milankovitch-like equations is obtained, which reminds of the structure of canonical systems of equations in Hamiltonian mechanics.

The vectorial elements are particularly useful in finding the first-order long-period and secular variations by averaging over the fast variables of the system, as elucidated in Chapter 3. We develop a first-order averaged model, based on the Milankovitch formulation of secular perturbation theory, for solar radiation pressure, the oblateness (equatorial bulge) of the host planet, and third-body gravitational interactions; each written in a concise analytical vector form. The secular equations hold rigorously for all Keplerian orbits with nonzero angular momentum; they are free of the singularities associated with zero eccentricity and vanishing line of nodes. We show in §3.1 how the secular Milankovitch equations that govern the motion of an object orbiting about a planet (idealized as a point mass moving in a heliocentric elliptical orbit) and subject to an arbitrary model of solar radiation pressure acceleration (obeying an inverse-square law) can be solved in closed form yielding a remarkably simple solution (Theorem 1). The averaged cannonball solution, previously derived for the secular motion of an orbiter about a small body in a solar radiation pressure dominated environment (qq.v., Mignard 1981; Richter & Keller 1995; Scheeres 2012a), can be easily recovered from this more general result.

There exist three mutually perpendicular planes of equilibrium for the gravitational motion of bodies under the perturbing action of the Sun and from the oblateness of the central body: the classical Laplace plane, characteristic of satellite systems, and the remaining two planes discovered by Sekiguchi (1961) and treated recently in great detail by Tremaine, Touma & Namouni (2009).

We rigorously show in Chapter 4 how the dynamical effects of solar radiation pressure lead to a modification of these equilibrium planes. The modified Laplace equilibria, determined by the competition between the gravitational and non-gravitational perturbations, turn out to be qualitatively different, in some cases, from those obtained without considering the radiation pressure. The warping of the surfaces swept out by the equilibria as the semi-major axis varies depends critically on the cross-sectional area of the body exposed and exhibit remarkable differences in behavior than in the classical case. We present an analytical solution for the modified equilibria, determine their stability, and calculate the detailed warp of the modified Laplace surfaces.

Chapter 5 constitutes one of the primary applications of this research; namely, the study of the high area-to-mass ratio space debris. In §5.3, we demonstrate the validity of the first-order averaged equations presented in Chapter 3 by comparing numerical integrations of them to integrations of the exact Newtonian equations. The analysis given here brings out the essential physics involved and gives quite sensible predictions on how a population of HAMR objects will evolve over time. We show how the dynamical configuration of the Earth-Moon-Sun system, and in particular the regression of the lunar node, causes a secular resonance effect with a particular class of HAMR objects, leading to complex behavior and possibly generating orbital chaos. Furthermore, we investigate the spatial distribution of the HAMR debris population and identify a systematic structure associated with their distribution in inclination and ascending node phase space, which can aid the space surveillance community in future search campaigns. We apply the classical and modified Laplace plane solutions to the study of the dynamics and stability of GEO orbits and in §5.4 to the identification of robust, long-term disposal orbits for geostationary satellites.

In Chapter 6, we study the dynamical effects of radiation and gravitational perturbations on circumplanetary dust particles, and apply the results of Chapter 4 towards an understanding of the the striking hemispheric brightness asymmetry of Iapetus, Saturn's two-faced satellite. With Poynting-Robertson drag implicated as the physical mechanism and the modified surfaces as the transfer path, we show how dust ejected from Phoebe (Saturn's largest irregular satellite) spirals inward toward Iapetus creating a heterogeneous deposition pattern. As dust is a ubiquitous compo-

nent of the Solar System, we discuss the application of the modified Laplace surfaces to a number of other planetary systems.

Auxiliary concepts and formulae, as well as a considerable part of reckoning work for intermediate calculations, can be found in a series of Appendices.

Chapter 2

Perturbation Formulation and Averaging

There has been, apparently, no systematic attempt to find the parameters most suited to the problem in either special or general perturbations, even when they are restricted to the constants of the two-body problem.[1]

— Samuel Herrick, 1948

2.1 The Milankovitch Orbital Elements

2.1.1 Milankovitch and the Astronomical Theory of Climate Change

Serbian astronomer and climatologist Milutin Milankovitch (1879-1958) is best known for his fundamental contribution to the scientific understanding of long-term climatic variations. According to the astronomical theory of paleoclimates, often referred to as the Milankovitch theory, the secular variations of the distribution of solar radiation on the Earth's surface is the result of orbital and orientation changes of the Earth under planetary perturbations (Milankovitch 1941; Berger 1988). Milankovitch provided a clear, scientifically valid, and mathematically precise explanation of the origin and recurrence of the glacial-interglacial cycles (Ice Ages) using only Newton's law of gravitation and the law of irradiation. What is often not known about Milankovitch is his contribution to celestial mechanics proper.[2] Using the two vectorial first integrals of the unperturbed two-body problem—the angular momentum (areal) vector and the Laplace vector[3] —Milankovitch

[1] Herrick S. op. cit., p. 321

[2] For an interesting and detailed account of his scientific achievements, see Dimitrijević (2002).

[3] Also referred to as the Runge-Lenz vector in theoretical physics; though, as Goldstein (1976) suggests, a more fitting eponym is the Hermann-Bernoulli-Laplace vector. This vector invariant was discovered by Jakob Hermann and Johann Bernoulli in 1710 (Volk 1976; Bartnik, Haberzettl & Sandhas 1988), expounded on at the end of the

(1939) reformulated the classical perturbation equations in vector form providing the rudiments of a coordinate-free perturbation theory. The vectorial elements describe the spatial orientation, geometrical shape, and size of the osculating Keplerian orbit, and, together with the sixth scalar integral that represents the motion in time, constitutes a complete set of orbital elements.

Assuming that the disturbing force is conservative and can thus be represented as the (negative) gradient of the perturbing potential, Milankovitch derived the Lagrange planetary equations for the angular momentum vector $\boldsymbol{H}$, the Laplace vector $\boldsymbol{b} = \mu \boldsymbol{e}$, where μ denotes the gravitational parameter of the central body and $\boldsymbol{e}$ is the eccentricity vector, and the time of periapsis passage. For the astronomical theory of paleoclimates, he was only concerned with the secular evolution of the Earth's orbit, which had to be traced over hundreds of thousands of years into the past. Eliminating the time-dependent part from the perturbation function in a manner similar to Lagrange and Laplace, Milankovitch developed secular equations for the evolution of $\boldsymbol{H}$ and $\boldsymbol{b}$, which are nonsingular and possess an unusual degree of symmetry and brevity (Milankovitch 1939, 1941).

The Milankovitch orbital elements exhibit with elegant simplicity the geometrical relationships that exist; with this vector formalism, a deeper insight into the physics of the problem can be achieved, and many fundamental results in celestial mechanics can be obtained quite directly and much more simply (e.g., Musen 1961; Allan & Cook 1964; Richter & Keller 1995; Tremaine, Touma & Namouni 2009; Correia et al. 2011; Katz, Dong & Malhotra 2011, inter alia). Despite all its inherent advantages, this vectorial approach is still viewed as unconventional and abstruse (Tamayo et al. 2013). This is in part due to the absence of a systematic and accurate presentation of the various ideas and equations, free from second-hand references to foreign journals and articles out of circulation, and in a clear and consistent mathematical language. This section is an attempt to restructure and clarify the piecemeal contributions brought to the Milankovitch formulation of

eighteenth century in Laplace's 'Mécanique Céleste', and independently rediscovered in 1845 by Hamilton during his application of quaternions to Keplerian motion. It was later presented by Gibbs in a simple vector analytical derivation of the solution of the Kepler problem (see Gibbs & Wilson 1909, Chap. 3, pp. 135–136), and subsequently quoted in the then widely used monograph on vector analysis of Runge (1919, Chap. 2, §5). Lenz (1924) used this vector in his treatment of the motion and spectrum of the hydrogen atom in the framework of the Bohr-Sommerfeld theory of quantization. Pauli (1926) made essential use of the Laplace-Runge-Lenz vector in his admirable solution of the hydrogen atom, performed on the basis of Heisenberg's matrix mechanics.

special and general perturbations.

2.1.2 The Kepler Problem and Quantum Mechanics

The problem of describing the motion of a particle under the influence of a central force, the intensity of which varies inversely as the square of the distance, has acquired considerable celebrity under the name of the Kepler problem.[4] The atomic analogue to planetary motion—namely, the orbital motion of an electron in the Coulomb field of a positively charged nucleus—underlies both the Rutherford-Bohr model of the hydrogen atom and the old quantum theory of Bohr and Sommerfeld (Mehra & Rechenberg 2001, Vol. 1). The characteristic feature of the Kepler problem (both classically and quantum mechanically) is the existence of the Laplace-Runge-Lenz vector constant of motion. Whereas the angular momentum vector is a constant of the motion for every spherically symmetrical field of force, the constancy of $\boldsymbol{b}$ is a special aspect of the $1/|\boldsymbol{r}|$ potential ($\boldsymbol{r}$ being the radius vector of the particle), and expresses the fact that the orbits of the Kepler problem do not precess.[5] The Laplace-Runge-Lenz vector has a clear geometrical significance: it defines the orientation of the major axis in the orbital plane, pointing from the center of the force to the periapsis of the orbit, and its magnitude is proportional to the eccentricity (Bartnik, Haberzettl & Sandhas 1988). At a more fundamental level, both the polar equation of the orbit (Kepler's first law) and the hodograph of Keplerian motion (i.e., the shape of trajectories in velocity or momentum space) follow from $\boldsymbol{b}$ in remarkably simple ways; these two equations embody the complete representation of the motion (qq.v., Runge 1919, Chap. 2, §5; Battin 1999, Chap. 3).

One of the earliest successes of quantum mechanics, in a field where ordinary methods had proved inadequate, was the calculation, by Epstein and Schwarzschild in 1916, of the Stark effect[6] of the hydrogen atom (Bohr 1918; Mehra & Rechenberg 2001, Vol. 1, Chap. 2). Bohr (1918,

[4] A systematic exposition of many of the classical and modern aspects of the Kepler problem, including its group-geometric features, can be found in Cordani (2003). Appendix B gives a brief review of some well known, and for that reason simple, classical results of the Kepler problem in the framework of the Newtonian theory of gravitation.

[5] A small deviation of the potential from this form causes the major axis of the orbit to precess slowly in the plane perpendicular to $\boldsymbol{H}$; closed orbits do not exist in this case (Weinberg 1972).

[6] The Stark effect is the separation of spectral lines emitted by atoms under the action of an external homogeneous electric field. A detailed theory of the analogous Zeeman effect, i.e., the separation of atomic spectral lines in the presence of an external homogeneous magnetic field, was given later that same year by Sommerfeld and Debye.

§4), adapting the classical perturbation techniques of celestial mechanics to atomic systems, gave a more direct and illuminating method of calculating the secular motion and energy levels of the hydrogen atom perturbed by an electric field. Using the components of the averaged position vector (which, incidentally, is intrinsically related to the Laplace-Runge-Lenz vector), referred to by Bohr as the coordinates of the 'electrical centre' of the orbit, together with the components of the angular momentum vector, Bohr (1918) developed secular (Gauss) equations for the evolution of these vectors and obtained from their solutions the characteristic effect of an electric field on the hydrogen spectrum. Having provided a satisfactory explanation of the Stark effect, the subsequent problem of atomic physics that occupied the minds of the architects of quantum mechanics was the simultaneous action of electric and magnetic fields (both static and uniform, and perpendicular to each other) on the one-electron atom (Mehra & Rechenberg 2001, Vol. 1, Chap. 4, §5; Vol. 3, Chap. 4). Bohr's method for treating the Stark effect was cast into a particularly elegant form by Klein (1924) through vector analysis, allowing for a simple solution to the crossed-field problem. Lenz (1924) replaced the equations of motion of the electron (in the hydrogen atom under the influence of crossed electric and magnetic fields) by equations describing the evolution of the Laplace-Runge-Lenz vector, which he called the 'axial vector' ('Achsenvektor'), and the angular momentum vector. After averaging, the integration of the secular equations immediately yielded the quantized energy of the perturbed motion in agreement with the results obtained by Klein (1924). Pauli (1926) studied the crossed-field problem in the framework of Heisenberg's matrix mechanics, which renounced the Bohr-Sommerfeld mechanical, space-time pictures of atomic structure (Mehra & Rechenberg 2001, Vol. 3). Replacing the dynamical variables in Lenz's perturbation formulation (i.e., the Laplace-Runge-Lenz vector and the angular momentum vector) with Hermitian matrices in accordance with the fundamental postulates of the "new" theory, Pauli (1926) completely solved the quantum-mechanical Kepler problem, obtaining the Balmer formula for the hydrogen spectrum, calculating the splitting of the spectral lines in an electric field (Stark

See Mehra & Rechenberg (2001) for a comprehensive account of the content and impact of these important scientific papers, and the others discussed in this section, as well as a detailed historical study of the central aspects of quantum mechanics.

effect) and in a magnetic field (Zeeman effect), and determining the electron's behavior in crossed fields in perfect agreement with experiment (Mehra & Rechenberg 2001, Vol. 3). Pauli's solution to the hydrogen atom was generally regarded as a most convincing proof of the quantum theory of matter; however, the procedure, which worked brilliantly in the case of the hydrogen atom, could not be extended to two- and more-electron atoms, so matrix mechanics was largely superseded by a mathematically equivalent—though far more convenient—formalism introduced by Schrödinger (Mehra & Rechenberg 2001, Vol. 3; Vol. 5). While matrix mechanics is now considered to be an historical relic, the notion of employing the two vectorial first integrals of the Kepler problem in perturbation theory is still of great contemporary interest in quantum mechanics (qq.v., Redmond 1964; Stahlhofen 1994; Bellomo et al. 1998; Schleif & Delos 2008, and references therein).[7] This concept, whose genesis belongs to Bohr (1918), Klein (1924), and Lenz (1924), was introduced into the field of celestial mechanics by Milankovitch (1939), who formulated the Lagrange planetary equations in the language of vector calculus.

2.1.3 Applications to Celestial Mechanics

Allan & Ward (1963) regard the Milankovitch orbital elements[8] as the natural constants of motion in the problem of two bodies since they do not involve any particular frame of reference in their definitions. As noted by Hestenes (1983), the classical orbital elements and various alternative sets employed in celestial mechanics can be regarded as different parameterizations of these fundamental vectors. The classical elements, taken from observational astronomy, were built into the theory of planetary motion, though they are actually ill-fitting for this physical problem. This is evident from the great mathematical difficulties which must be surmounted in the deduction of the Lagrange planetary equations, and is further demonstrated by the complicated and asymmetrical

[7] It even appears of advantage to use these quantities of indisputable basic importance from the general standpoint of modern physics: "By far the most convenient technique for calculating the precession of perihelia," notes Steven Weinberg, on the post-Newtonian formalism of gravitation, "is to compute the rate of change of the Runge-Lenz vector (Weinberg 1972, Chap. 9, §5, p. 230)."

[8] We refer to any set of orbital elements which emanate from $\boldsymbol{H}$ and $\boldsymbol{b}$ as the Milankovitch elements to give precedence to Milankovitch. Although it may be more historically accurate to refer to these types of orbit parameters as the Bohr-Klein-Lenz vectorial elements, Milankovitch presented them in a particularly suitable form for celestial mechanics and its applications.

structure of these equations, as emphasized by Milankovitch (1941, Chap. 3, §18). Actually, the classical perturbation equations of celestial mechanics can be obtained in an elementary way from the vectorial equations as demonstrated in Appendix E, thereby giving more physical intuition into the nature of these classical equations (qq.v., Milankovitch 1941; Kurth 1959, Chap. 3, §4; Ward 1962, p. 142; Pollard 1966, Chap. 1, §17; Burns 1976; Battin 1999, §10.5). One of the characteristic features of Milankovitch's perturbation formulation is the introduction of redundant variables. The angular momentum vector and the Laplace vector are equivalent to only five independent quantities because of the orthogonality constraint $\boldsymbol{H}\cdot\boldsymbol{b}=0$. While redundancy in the set of variables increases the order of the dynamical system, it often leads to equations which are remarkably simple and elegant (Broucke, Lass & Ananda 1971).[9]

Planetary equations in terms of the Milankovitch elements in Lagrangian form were quoted and used after their inception, most notably by Bilimovitch, Musen, Allan, and Ward. Bilimovitch (1943) obtained the Milankovitch equations using the Pfaffian equations and Cartan's integral invariants, which led to a much simpler derivation than the Lagrange or Poisson bracket operations. Musen (1964) extended the work of Bilimovitch, permitting supplementary variables to be present in the system and forming the constraints with Lagrange multipliers. The Pfaffian method was used to search for new sets of elements and in the formation of equations for their variations. Musen discovered a canonical set of elements, of pure vectorial form, defined by the vectors $\sqrt{\mu a}\hat{\boldsymbol{e}}$ and $\sqrt{1-e^2}\hat{\boldsymbol{e}}_\perp - M\hat{\boldsymbol{e}}$, where a is the semi-major axis, e is the eccentricity, and M is the mean anomaly.[10] Fleckenstein (1952) presented a concise derivation of the Milankovitch planetary equations, noted their connection with Lenz (1924), and recognized their importance for general perturbations. Allan & Ward (1963) introduced vector and dyadic Poisson brackets resulting in a very convenient and

[9] Apart from these advantages, there exists an intrinsic relationship between the introduction of redundant variables and regularization theory in celestial mechanics: in fact, "regularization of elements enforces ... the use of redundant sets of elements (Stiefel & Scheifele 1971)." There is a wealth of interesting applications in orbital mechanics and rotational dynamics in which regularized and stabilized equations can be obtained in terms of redundant variables (see, for instance, Stiefel & Scheifele 1971; Broucke, Lass & Ananda 1971; Deprit 1975; Cordani 2003), but this topic is beyond our current scope.

[10] The unit vectors $\hat{\boldsymbol{e}}=\boldsymbol{e}/e$ and $\hat{\boldsymbol{e}}_\perp=\hat{\boldsymbol{h}}\times\hat{\boldsymbol{e}}$ (where $\hat{\boldsymbol{h}}=\boldsymbol{H}/H$) are the orientation-defining integrals of the two-body problem, and can be specified using the classical orbital elements relative to an inertial frame (q.v., Appendix B.2.3). They are referred to as the Gibbsian vectorial elements by Musen (1961, 1964) and the vectorial orbital constants by Brouwer & Clemence (1961a). The geometrical significance of these vectors can be seen from Figure B.1.

elegant derivation of the perturbation equations. They also showed how the Gibbsian unit vectors of the two-body problem ($\hat{\boldsymbol{e}}$, $\hat{\boldsymbol{e}}_\perp$, and $\hat{\boldsymbol{h}}$) may be employed in perturbation theory leading to various sets of redundant elements. Actually, as Musen (1954) had observed, the vector $(e/\sqrt{p})\hat{\boldsymbol{e}}_\perp$, where $p = a(1 - e^2)$ is the orbit parameter, was used in Hansen's lunar theory, albeit in an implicit, and somewhat encrypted, form (see Tisserand 1896, Vol. 3, pp. 305–306, and Vol. 4, p. 328, for specific details on Hansen's method). Planetary equations in terms of the Milankovitch elements in Gaussian form, that is, with the disturbing acceleration given explicitly, were derived by Popović (1950), and subsequently expounded by a number of celestial mechanicians (Musen 1954; Allan 1962; Wen 1962; Ward 1962; Roy & Moran 1973). Herrick (1948) formulated the Gauss equations in terms of the vectors $e\hat{\boldsymbol{e}}$ and $e\sqrt{p}\hat{\boldsymbol{e}}_\perp$, and seemed to have been unaware of the developments of Milankovitch (see also Newton (1961) in this respect).

Representing the Euler angles[11] (longitude of the ascending node, inclination, and argument of periapsis) as vectorial elements amounts to referring perturbed Keplerian motion to a slowly moving frame (Deprit 1975). This viewpoint is also the basis of Strömgren's method, which incorporates the unit vectors $\hat{\boldsymbol{e}}$ and $\hat{\boldsymbol{h}}$ into the Gibbs rotation vector[12] to study the instantaneous angular velocity of rotation of the osculating ellipse (Strömgren 1929). Strömgren's equations possess a mathematical simplicity that rivals Milankovitch's formulation, and represent the first successful attempt at applying vector analysis to the theory of special perturbations. Musen (1954) showed how the perturbation formalisms of Strömgren and Milankovitch could be combined yielding differential equations for the angular momentum vector, the Gibbs rotation vector, and the mean anomaly. Following Musen's work, Herget (1962) defined an arbitrary 'departure point' within the osculating orbit plane to overcome the problem of small eccentricity in the mean anomaly equation. In eliminating the singularity due to small eccentricities, Herget introduced a spurious singularity resulting from rotations that are half turns. Deprit (1975) used Euler parameters (quaternions) to describe the rotation giving a highly redundant set of elements, which does not generate this singu-

[11] Euler angles are coordinates or parameters used to describe the orientation of a rigid body, and are customary in the analytical specification of torque-free rigid body rotation (see, e.g., Schaub & Junkins 2009, Chaps. 3 and 4)

[12] See Gibbs and Wilson, op. cit., pp. 343–345.

larity. A new spinor formulation of Milankovitch's perturbation theory was constructed by Hestenes (1983, 1999) based on geometric algebra. Hestenes eliminated all redundancies, and showed how the unification of vector and quaternion algebra can lead to much insight and clarity.

The Milankovitch elements are particularly useful in finding the first-order long-period and secular variations by averaging over the fast, orbital timescales of the system. When the disturbing function is limited to its secular part, as in the case of averaging, the semi-major axis does not undergo any secular changes (Brouwer & Clemence 1961a). Since the semi-major axis is secularly invariable, the angular momentum vector can be scaled by $\sqrt{\mu a}$. For this vector, denoted here as $\boldsymbol{h}$, together with the eccentricity vector $\boldsymbol{e}$, the secular Milankovitch equations take a compact and symmetrical form (qq.v., Equations 2.36 and 2.37; see also Allan & Cook (1964), Equation 13). Musen (1960) used these elements to study the effects of solar radiation pressure acceleration on the long-term orbit evolution of the Vanguard I satellite. Richter & Keller (1995) studied the averaged SRP dynamics acting on dust particles orbiting cometary nuclei, and showed that Musen's equations could be solved in closed form. Scheeres (2012a), using vector and dyadic analysis, demonstrated a simpler solution technique, and explored several aspects of the solution that are relevant for the orbital motion of spacecraft about small bodies. This solution is generalized to an arbitrary model of solar radiation pressure in Chapter 3. Musen (1961) also developed secular vectorial equations for third-body perturbations, which are valid for all eccentricity and all inclination. The third-body disturbing function was truncated at the lowest-order term in the Legendre expansion (i.e., the Hill, or quadrupolar, approximation), and averaged over the periods of both the orbiter and the disturbing body. Allan (1962) extended Musen's work by incorporating the parallactic term (third harmonic) into the disturbing function expansion, and by including the effects of the Earth's shadow for SRP acceleration. Because these equations are not widely known, and in order to draw attention to their applicability and elegance, we give an outline of their derivations in Chapter 3. Allan & Cook (1964) studied the secular equations resulting from Earth's oblateness (second zonal harmonic in Earth's gravity field) and the gravitational attractions of the Sun and the Moon. They investigated a region of space, called the Laplace plane, where the long-term

evolution of circular orbits driven by the combined effects of these forces is zero, so that the orbits are "frozen." This is an extension of Laplace's treatment in 1805 for the motion of Iapetus about Saturn (Laplace 1805, Vol. 4, Eight Book, Chap. 17; Tisserand 1896, Vol. 4, Chap. 6). Tremaine, Touma & Namouni (2009) studied the properties of the Laplace plane, including the stability of its generating orbits, and generalized it to eccentric orbits. Allan & Cook (1967), in considering the possibility of a geocentric contribution to the zodiacal light, showed that solar radiation pressure modified the classical Laplace equilibrium plane[13] (see also Tamayo et al. 2013, for an interesting discussion of the modified Laplace plane in the context of circumplanetary dust particles in unstable orbital dynamics environments). Breiter & Ratajczak (2005) applied Allan and Cook's vectorial equations to the averaged Galactic disk tide perturbations in cometary motion (see also Breiter, Fouchard & Ratajczak 2008). Namouni (2005) applied the secular Milankovitch equations towards an understanding of the origin of the large eccentricities of extrasolar planets, which may have resulted from precessing stellar jets during star formation. Touma, Tremaine & Kazandjian (2009) developed formulas for the numerical integration of the first-order secular effects of softened gravitational interactions, using a modified Gauss method and similar vectorial elements, which can be used to study the averaged dynamical evolution of nearly-Keplerian, many-body systems. McMahon & Scheeres (2010a,b) used the Gaussian form of the secular Milankovitch equations for long term predictions of the system evolution of binary asteroids under the Binary YORP effect. Correia et al. (2011), using the Milankovitch elements, investigated the secular evolution of hierarchical three-body systems (such as "hot Jupiters" around binary stars), where the innermost bodies undergo tidal interactions, using the Hill approximation for gravitational interactions with general relativity corrections (see also Farago & Laskar 2010, and references therein for more details). Katz, Dong & Malhotra (2011) applied these vectorial elements to the hierarchical restricted three-body problem, finding that interesting dynamical behavior occurs for the Kozai-Lidov cycles when the parallactic, or octupole, term is included in the gravitational

[13] A concise review of the classical Laplace plane, along with a more rigorous treatment of the modified Laplace plane, is given in Chapter 4

potential. Tremaine & Yavetz (2013) used this geometric and vector-based formalism to study the stability of satellites in low Earth orbits under gravitational perturbations, and note emphatically that the Milankovitch elements "deserve to be more widely known."

2.2 Variation of Parameters

The fundamental method of perturbation theory in celestial mechanics is variation of parameters, which evolved over half a century beginning with Euler's cursory initial steps and reaching its zenith with Lagrange, Laplace, and Gauss (Taton & Wilson 1995). The basic concept is that the inclusion of perturbation forces into the two-body problem generates small changes to the constants of motion, causing them to become time-varying parameters (i.e., osculating elements). The original transformation in the two-body problem between the constants and the solution still stands, allowing the time-varying parameters to describe the solution.[14] For the perturbed Kepler problem to be well posed, the chosen set of osculating orbital elements must comprise a complete set of functionally-independent first integrals of the unperturbed system; however, there may be redundant variables in the set.

The importance of selecting a set of osculating elements most suited to describe the perturbed motion has already been discussed in Chapter 1. The six components of the vectorial first integrals $\boldsymbol{H}$ and $\boldsymbol{b}$ furnish a set of only five functionally independent (scalar) constants of motion which characterize the form, orientation, and size of the osculating conic-section path; the sixth (and last) element, functionally independent of the preceding ones, and necessary to completely describe the motion, is related to the specification of the instantaneous position of the moving particle along its orbit at any given instant of time, and therefore accounts for the relationship between position in orbit and time.

In our perturbation formulation, we use the angular momentum vector $\boldsymbol{H}$, the eccentricity

[14] For the method of variation of parameters, the parameters may be varied either in solution functions of the unperturbed system (Lagrange's method) or in their first integrals (Poisson's method). Poisson's method is particularly useful in that it gives the variation of each element with time directly, without the necessity of computing and inverting the Lagrangian bracket matrix (Poisson 1809). For this section, we are in part following the exposition of R. H. Battin, op. cit., §10.5, to which we refer for the omitted derivations.

vector $\boldsymbol{e}$, and either the true longitude L or the mean longitude l. The 'angles' L and l locate where the solution is within the trajectory at a given time, and allow for a nonsingular description of the motion. The elements $\boldsymbol{H}$, $\boldsymbol{b} = \mu\boldsymbol{e}$, and L were used by Roy & Moran (1973) for special perturbation methods, in which they derived the planetary equations in Gaussian form only. We present here a separate and independent derivation of both the Lagrange and Gauss equations for these two element sets, and examine the choice and merits of the corresponding scalar element.

The magnitude of $\boldsymbol{H}$ is $\sqrt{\mu p}$, where p is the focal parameter ($p = a(1 - e^2)$ for elliptical ($a > 0$) and hyperbolic ($a < 0$) orbits), and the angular momentum and eccentricity vectors satisfy the identity $\boldsymbol{H} \cdot \boldsymbol{e} = 0$. We can write these vectors in terms of the position $\boldsymbol{r}$ and velocity $\boldsymbol{v}$ in dyadic notation[15] as

$$\boldsymbol{H} = \widetilde{\boldsymbol{r}} \cdot \boldsymbol{v}, \tag{2.1}$$

$$\boldsymbol{e} = \frac{1}{\mu}\widetilde{\boldsymbol{v}} \cdot \widetilde{\boldsymbol{r}} \cdot \boldsymbol{v} - \frac{\boldsymbol{r}}{|\boldsymbol{r}|}. \tag{2.2}$$

The true longitude of a body in orbit is the angle measured from the planetary equinox to the ascending node, and then in the direction of the orbit to the body. Namely,

$$L = \Omega + \theta = \varpi + f, \tag{2.3}$$

where Ω is the longitude of the ascending node, $\theta = \omega + f$ is the argument of latitude, ω is the argument of periapsis, f is the true anomaly, and $\varpi = \Omega + \omega$ is the longitude of periapsis (q.v., Figure B.1a in Appendix B). Correspondingly, the mean longitude is defined by

$$l = \varpi + M, \tag{2.4}$$

in which M is the mean anomaly (Battin 1999).

2.2.1 Lagrange Planetary Equations

For generality, let $\boldsymbol{\alpha} \in \boldsymbol{R}^7$ denote our set of osculating orbital elements. The variation equation of α_i in Lagrangian form can be stated, using the Einstein summation convention for the

[15] The operator $\widetilde{\ }$ turns a vector into a skew-symmetric dyadic and is defined such that $\boldsymbol{a} \times \boldsymbol{b} = \widetilde{\boldsymbol{a}} \cdot \boldsymbol{b} = \boldsymbol{a} \cdot \widetilde{\boldsymbol{b}} = -\widetilde{\boldsymbol{b}} \cdot \boldsymbol{a} = -\boldsymbol{b} \cdot \widetilde{\boldsymbol{a}}$. Note that the associative law is valid for the continued direct product in Equation 2.2; we can therefore write without ambiguity $(\widetilde{\boldsymbol{v}} \cdot \widetilde{\boldsymbol{r}}) \cdot \boldsymbol{v} = \widetilde{\boldsymbol{v}} \cdot (\widetilde{\boldsymbol{r}} \cdot \boldsymbol{v}) = \widetilde{\boldsymbol{v}} \cdot \widetilde{\boldsymbol{r}} \cdot \boldsymbol{v}$.

repeated indices, as (Milankovitch 1941, §19; Battin 1999, §10.5)

$$\dot{\alpha}_i = -(\alpha_i, \alpha_j)\frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial \alpha_j} + \frac{\partial \alpha_i}{\partial t}, \tag{2.5}$$

where $\mathcal{R}(\boldsymbol{\alpha})$ is the disturbing function (the negative potential function of the disturbing force), and $\partial\alpha_i/\partial t$ represents the change in α_i due to the unperturbed two-body motion. The operator $(-,-)$ is called the Poisson bracket, and for a general term i, j yields

$$(\alpha_i, \alpha_j) = \frac{\partial \alpha_i}{\partial \boldsymbol{r}} \cdot \frac{\partial \alpha_j}{\partial \boldsymbol{v}} - \frac{\partial \alpha_i}{\partial \boldsymbol{v}} \cdot \frac{\partial \alpha_j}{\partial \boldsymbol{r}}, \tag{2.6}$$

where the gradient of a scalar with respect to a vector defines a vector in $\boldsymbol{R}^3$. The Poisson bracket operator can be extended to include vectors and dyadics.

Definition 1. *For vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\boldsymbol{R}^3$, define (Allan & Ward 1963)*

$$(\boldsymbol{a}, \alpha_j) = -(\alpha_j, \boldsymbol{a}) = \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{r}} \cdot \frac{\partial \alpha_j}{\partial \boldsymbol{v}} - \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{v}} \cdot \frac{\partial \alpha_j}{\partial \boldsymbol{r}}, \tag{2.7}$$

$$(\boldsymbol{a}, \boldsymbol{b}) = -(\boldsymbol{b}, \boldsymbol{a})_c = \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{r}} \cdot \left(\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{v}}\right)_c - \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{v}} \cdot \left(\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{r}}\right)_c, \tag{2.8}$$

where the suffix c denotes the conjugate dyadic (q.v., Appendix A).

As noted by Goldstein (1950, §8.7) and Allan & Ward (1963), forming the Poisson bracket with $\boldsymbol{H}$ can be interpreted as a rotation operation, and thus all the brackets with scalar elements that do not contain fixed vectors independent of $\boldsymbol{r}$ and $\boldsymbol{v}$ vanish identically, and all the brackets with the vector elements can be written as a skew-symmetric dyadic. However, we will recourse to direct evaluation from the definitions, as it is instructive to do so.

The partial derivatives of the Milankovitch elements with respect to $\boldsymbol{r}$ and $\boldsymbol{v}$ are given in Appendix D. It follows from the definition of the dyadic Poisson bracket, and the identities in Equations A.15, A.17, and A.18 (from Appendix A), that the Poisson bracket of the angular momentum vector with itself is

$$(\boldsymbol{H}, \boldsymbol{H}) = -\widetilde{\boldsymbol{H}}. \tag{2.9}$$

Similarly, from Equations D.1-D.4 and the dyadic cross product identities given in Appendix A, the dyadic Poisson bracket of $\boldsymbol{H}$ and $\boldsymbol{e}$ can be written as

$$(\boldsymbol{H}, \boldsymbol{e}) = -\widetilde{\boldsymbol{e}}. \tag{2.10}$$

By direct substitution of Equations D.3 and D.4 into Equation 2.8, we obtain after simplifying

$$(\boldsymbol{e}, \boldsymbol{e}) = \frac{2}{\mu^2}\left[\frac{\boldsymbol{v}\cdot\boldsymbol{v}}{2} - \frac{\mu}{|\boldsymbol{r}|}\right](\boldsymbol{v}\boldsymbol{r} - \boldsymbol{r}\boldsymbol{v}). \tag{2.11}$$

Using Equation A.18, the dyadic term, $\boldsymbol{v}\boldsymbol{r} - \boldsymbol{r}\boldsymbol{v}$, simplifies to $\widetilde{\boldsymbol{H}}$ (see also Battin 1999, §3.6, p. 130, Problem 3-26; for a definition of the dyadic product, see §2.3, p. 92). The bracketed term is the energy integral of the unperturbed two-body problem (see Equation B.11). Taking the dot product of the eccentricity vector, Equation 2.2, with itself shows that the energy can be expressed in terms of H and e as $E = -\mu^2(1-e^2)/2H^2$. Therefore, the final dyadic Poisson bracket can be written in the form

$$(\boldsymbol{e}, \boldsymbol{e}) = -\frac{1-e^2}{H^2}\widetilde{\boldsymbol{H}}. \tag{2.12}$$

An important property of the Poisson brackets is the Jacobi-Poisson theorem which states that for any two constants of motion of a given holonomic dynamical system, their Poisson bracket is also a constant of motion (Whittaker 1937, §145). It is therefore permitted to evaluate the Poisson bracket of any two integrals of the system at any selected convenient point in the orbit. This fact is exploited to great advantage in determining the Lagrange equations for the classical orbital elements (Battin 1999). This theorem also has profound significance for averaging, as it allows us to pull the averaging operator across the Poisson bracket term to the disturbing function, greatly simplifying the derivation of the mean equations (Scheeres 2012b).

In the classical form of the Lagrange planetary equations (i.e., the time rate of change of the elliptic elements: a, e, i, Ω, ϖ, l_0, where l_0 is the mean longitude at epoch), the expression involving l_0 is often replaced with the mean longitude to avoid an undesirable complication caused by the presence of the linear function of time in the expression for $\partial\mathcal{R}/\partial a$ (this procedure is often referred to in the literature as "Eliminating the Secular Term"; see, for instance Brouwer & Clemence 1961a,

Chapter 11, §6). It is therefore customary to start with an anomaly at epoch as the remaining constant of motion, thereby taking advantage of the time-invariance property of the brackets in the derivation of the variational equations, but to later replace that element with its time-varying counterpart. However, the true longitude at epoch is implicitly defined though the solution of Kepler's equation, and thus cannot be used as an initial alternate element. On the other hand, the mean longitude at epoch is explicitly defined and has been used in other 'nonsingular' element sets, such as the equinoctial variables (Broucke & Cefola 1972; Battin 1999, p. 492). Nevertheless, we will not recourse to this conventional approach, but instead derive the Lagrange equations for our two element sets directly, at the expense of not being able to apply the Jacobi-Poisson theorem.[16] As will be evident, each set has advantages that suit themselves for different applications.

Proposition 1. *Let the set of osculating orbital elements, $\boldsymbol{\alpha} \in \boldsymbol{R}^7$, be given by $\boldsymbol{H}$, $\boldsymbol{e}$, and L. Then their perturbation equations in Lagrangian form can be stated as*

$$\dot{\boldsymbol{H}} = \widetilde{\boldsymbol{H}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{H}} + \widetilde{\boldsymbol{e}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{e}} + \frac{\boldsymbol{H} + H\hat{\boldsymbol{z}}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}} \frac{\partial \mathcal{R}}{\partial L}, \tag{2.13a}$$

$$\dot{\boldsymbol{e}} = \widetilde{\boldsymbol{e}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{H}} + \frac{1 - e^2}{H^2} \widetilde{\boldsymbol{H}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{e}} + \frac{1}{H} \left[(2 + \hat{\boldsymbol{r}} \cdot \boldsymbol{e}) \hat{\boldsymbol{r}} + \boldsymbol{e} - \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{e}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}} \boldsymbol{H} \right] \frac{\partial \mathcal{R}}{\partial L}, \tag{2.13b}$$

$$\dot{L} = -\frac{\boldsymbol{H} + H\hat{\boldsymbol{z}}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{H}} - \frac{1}{H} \left[(2 + \hat{\boldsymbol{r}} \cdot \boldsymbol{e}) \hat{\boldsymbol{r}} + \boldsymbol{e} - \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{e}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}} \boldsymbol{H} \right] \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{e}} + \frac{H}{r^2}, \tag{2.13c}$$

where $\mathcal{R} = \mathcal{R}(\boldsymbol{\alpha})$.

Proof. Having computed the dyadic Poisson brackets of the vectorial elements (qq.v., Equations 2.9, 2.10, and 2.12), the only remaining computations are the vector Poisson brackets of the true longitude with $\boldsymbol{H}$ and $\boldsymbol{e}$. The partial derivatives of these elements with respect to $\boldsymbol{v}$ and $\boldsymbol{r}$ are given in Appendix D. From Equation 2.7, and using the identities given in Equations A.15 and A.16, we find

$$(L, \boldsymbol{H}) = \frac{\partial L}{\partial \boldsymbol{r}} \cdot \left(\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{v}} \right)_c - \frac{\partial L}{\partial \boldsymbol{v}} \cdot \left(\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{r}} \right)_c \tag{2.14}$$

[16] Since both the true longitude and the mean longitude vary with time (implicitly through f and M, respectively), they are not constants of motion, and thus their Poisson brackets with $\boldsymbol{H}$ and $\boldsymbol{e}$ may not, in general, be independent of time. This can result in time-varying coefficients in front of the partials of $\mathcal{R}$ in the perturbation equations, which is undesirable when these equations must be averaged.

$$= \frac{1}{H}\boldsymbol{H} + \frac{1}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})}\widetilde{\boldsymbol{H}} \cdot [\boldsymbol{r}\boldsymbol{v} - \boldsymbol{v}\boldsymbol{r}] \cdot \hat{\boldsymbol{z}}. \tag{2.15}$$

The dyadic $\boldsymbol{r}\boldsymbol{v} - \boldsymbol{v}\boldsymbol{r}$ is equivalent to $-\widetilde{\boldsymbol{H}}$. From Equation A.17 and the properties of dyadic algebra, $(L, \boldsymbol{H})$ simplifies to

$$(L, \boldsymbol{H}) = \frac{\boldsymbol{H} + H\hat{\boldsymbol{z}}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}}. \tag{2.16}$$

Similarly, the Poisson bracket of L with $\boldsymbol{e}$ can be written, after some reduction, as

$$(L, \boldsymbol{e}) = \frac{2H}{\mu r^2}\boldsymbol{r} - \frac{\boldsymbol{r} \cdot \boldsymbol{v}}{\mu r^2 H}\widetilde{\boldsymbol{H}} \cdot \boldsymbol{r} - \frac{1}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})}\hat{\boldsymbol{z}} \cdot \left[\left(\frac{\boldsymbol{v} \cdot \boldsymbol{v}}{\mu} - \frac{1}{|\boldsymbol{r}|}\right)\boldsymbol{r} - \frac{\boldsymbol{r} \cdot \boldsymbol{v}}{\mu}\boldsymbol{v}\right]\boldsymbol{H}. \tag{2.17}$$

The term in brackets can be identified with the eccentricity vector, as seen by Equation 2.2. Applying several identities from the two-body problem, $(L, \boldsymbol{e})$ reduces to the form

$$(L, \boldsymbol{e}) = \frac{1}{H}\left[(2 + \hat{\boldsymbol{r}} \cdot \boldsymbol{e})\hat{\boldsymbol{r}} + \boldsymbol{e} - \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{e}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}}\boldsymbol{H}\right]. \tag{2.18}$$

From Equation 2.5, the perturbation equations in Lagrangian form can be stated as

$$\dot{\boldsymbol{H}} = -(\boldsymbol{H}, \boldsymbol{H}) \cdot \frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial \boldsymbol{H}} - (\boldsymbol{H}, \boldsymbol{e}) \cdot \frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial \boldsymbol{e}} - (\boldsymbol{H}, L)\frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial L}, \tag{2.19a}$$

$$\dot{\boldsymbol{e}} = -(\boldsymbol{e}, \boldsymbol{H}) \cdot \frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial \boldsymbol{H}} - (\boldsymbol{e}, \boldsymbol{e}) \cdot \frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial \boldsymbol{e}} - (\boldsymbol{e}, L)\frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial L}, \tag{2.19b}$$

$$\dot{L} = -(L, \boldsymbol{H}) \cdot \frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial \boldsymbol{H}} - (L, \boldsymbol{e}) \cdot \frac{\partial \mathcal{R}(\boldsymbol{\alpha})}{\partial \boldsymbol{e}} + \frac{\partial L}{\partial t}, \tag{2.19c}$$

where $\partial L/\partial t$ represents the change in L due to the unperturbed two-body motion, as specified by Kepler's second law of planetary motion:

$$\frac{\partial L}{\partial t} = \frac{H}{r^2}. \tag{2.20}$$

With the Poisson brackets determined, Equation 2.13 follows directly, thereby completing the proof. □

Remark 1. *Although Equation 2.13 is slightly more complicated than the original Milankovitch equations (cf. Milankovitch 1939, Equations 204-206), they become singular only for $H = 0$; that is, when the resulting motion is rectilinear and hence only has a single degree of freedom.*

Remark 2. *Due to the presence of the time-dependent term $\hat{\boldsymbol{r}}$, which varies with true anomaly, in the Poisson bracket $(L, \boldsymbol{e})$, the element set $\boldsymbol{H}$, $\boldsymbol{e}$, and L is not convenient for averaging applications.*

Proposition 2. *Let the set of osculating orbital elements, $\boldsymbol{\alpha} \in \boldsymbol{R}^7$, be given by $\boldsymbol{H}$, $\boldsymbol{e}$, and l. Then their perturbation equations in Lagrangian form are given by*

$$\dot{\boldsymbol{H}} = \widetilde{\boldsymbol{H}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{H}} + \widetilde{\boldsymbol{e}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{e}} + \frac{\boldsymbol{H} + H\hat{\boldsymbol{z}}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}} \frac{\partial \mathcal{R}}{\partial l}, \tag{2.21a}$$

$$\dot{\boldsymbol{e}} = \widetilde{\boldsymbol{e}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{H}} + \frac{1 - e^2}{H^2} \widetilde{\boldsymbol{H}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{e}} - \left(\frac{\sqrt{1 - e^2}}{\sqrt{\mu a} + H} \boldsymbol{e} + \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{e}}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})} \boldsymbol{H} \right) \frac{\partial \mathcal{R}}{\partial l}, \tag{2.21b}$$

$$\dot{l} = -\frac{\boldsymbol{H} + H\hat{\boldsymbol{z}}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}} \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{H}} + \left(\frac{\sqrt{1 - e^2}}{\sqrt{\mu a} + H} \boldsymbol{e} + \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{e}}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})} \boldsymbol{H} \right) \cdot \frac{\partial \mathcal{R}}{\partial \boldsymbol{e}} + n, \tag{2.21c}$$

where $\mathcal{R} = \mathcal{R}(\boldsymbol{\alpha})$.

Proof. It follows from Equation 2.4, the definition of the Poisson bracket, and the properties of differentiation, that

$$(l, \boldsymbol{H}) = (\varpi + M, \boldsymbol{H}) = (\varpi, \boldsymbol{H}) + (M, \boldsymbol{H}). \tag{2.22}$$

The partial derivatives of ϖ with respect to $\boldsymbol{v}$ and $\boldsymbol{r}$ are given by Equations D.20 and D.21. By direct substitution into Definition 1, we find

$$(\varpi, \boldsymbol{H}) = \frac{\boldsymbol{H} + H\hat{\boldsymbol{z}}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}}. \tag{2.23}$$

Note that this is the same result as that found for $(L, \boldsymbol{H})$, Equation 2.16. Indeed, $(L, \boldsymbol{H}) = (\varpi + f, \boldsymbol{H}) = (\varpi, \boldsymbol{H}) + (f, \boldsymbol{H})$, and since $(f, \boldsymbol{H}) = 0$ from the condition given in Allan & Ward (1963), we have $(\varpi, \boldsymbol{H}) = (L, \boldsymbol{H})$. Also from Allan & Ward (1963), $(M, \boldsymbol{H})$ vanishes identically, yielding the simple form for the Poisson bracket of the mean longitude and the angular momentum vector:

$$(l, \boldsymbol{H}) = (\varpi, \boldsymbol{H}) = \frac{\boldsymbol{H} + H\hat{\boldsymbol{z}}}{H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H}}. \tag{2.24}$$

Since the longitude of periapsis and the eccentricity vector are both constants of motion, by the Jacobi-Poisson theorem, their Poisson bracket is constant throughout the motion. At periapsis,

when $\boldsymbol{r} \cdot \boldsymbol{v} = 0$, $\boldsymbol{r} = a(1-e)\hat{\boldsymbol{e}}$, and $\boldsymbol{v} = (\mu/H)(1+e)\hat{\boldsymbol{e}}_\perp$, we find

$$(\varpi, \boldsymbol{e}) = -\frac{H}{\mu a e^2}\boldsymbol{e} - \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{e}}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})}\boldsymbol{H}. \tag{2.25}$$

From Allan & Ward (1963), the Poisson bracket of the mean anomaly with the eccentricity vector has the form

$$(M, \boldsymbol{e}) = \frac{nH^2}{\mu^2 e^2}\boldsymbol{e}, \tag{2.26}$$

in which $n = \sqrt{\mu/a^3}$ is the orbiter's mean motion. Note that both $(\varpi, \boldsymbol{e})$ and $(M, \boldsymbol{e})$ become singular when $e = 0$. Adding these expressions, we obtain the Poisson bracket of the mean longitude with $\boldsymbol{e}$:

$$(l, \boldsymbol{e}) = (\varpi, \boldsymbol{e}) + (M, \boldsymbol{e}) \tag{2.27}$$

$$= -\frac{\sqrt{1-e^2}}{\sqrt{\mu a} + H}\boldsymbol{e} - \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{e}}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})}\boldsymbol{H}, \tag{2.28}$$

which is independent of time and devoid of singularities related to orbits of zero inclination angle and zero eccentricity. The Lagrange equations for $\boldsymbol{H}$, $\boldsymbol{e}$, and l follow from Equation 2.5, with $\partial l/\partial t = n$, thereby completing the proof of Proposition 2. □

Remark 3. *Equation 2.21 is free of the singularities associated with zero eccentricity and vanishing line of nodes. It should be noted that the use of the mean longitude necessarily implies that the motion must be elliptic; thus, the equations lose their meaning for values of eccentricity $e > 1$. In this case, a different formulation is needed.*

Remark 4. *The Poisson brackets for the element set $\boldsymbol{H}$, $\boldsymbol{e}$, and l are independent of time, and thus the averaging can be carried across these terms to the disturbing function, making this set well-suited for averaging applications.*

2.2.2 Gauss Equations

The variation equation of $\boldsymbol{\alpha}$ in Gaussian form can be stated as (Battin 1999, §10.5)

$$\dot{\boldsymbol{\alpha}} = \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{v}} \cdot \boldsymbol{a}_d + \frac{\partial \boldsymbol{\alpha}}{\partial t}, \tag{2.29}$$

where $\boldsymbol{a}_d$ is an arbitrary disturbing acceleration and $\partial\boldsymbol{\alpha}/\partial t$ represents the change in $\boldsymbol{\alpha}$ due to the unperturbed two-body motion. Thus, to compute the Gauss equation for an integral of motion, we must only take its partial with respect to the velocity, and dot that expression with the perturbing acceleration vector.

Proposition 3. *The Gauss perturbation equations of the element set $\boldsymbol{H}$, $\boldsymbol{e}$, and L can be stated as (Roy & Moran 1973)*

$$\dot{\boldsymbol{H}} = \widetilde{\boldsymbol{r}} \cdot \boldsymbol{a}_d, \tag{2.30a}$$

$$\dot{\boldsymbol{e}} = \frac{1}{\mu}\left(\widetilde{\boldsymbol{v}} \cdot \widetilde{\boldsymbol{r}} - \widetilde{\boldsymbol{H}}\right) \cdot \boldsymbol{a}_d, \tag{2.30b}$$

$$\dot{L} = \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{r}}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})} \boldsymbol{H} \cdot \boldsymbol{a}_d + \frac{H}{r^2}. \tag{2.30c}$$

Proof. The partial derivatives of $\boldsymbol{H}$, $\boldsymbol{e}$, and L, with respect to $\boldsymbol{v}$, are given in Equations D.1, D.3, and D.12, respectively. Substituting each of these expressions into Equation 2.29, and noting Equation 2.20 for the unperturbed motion, the Gauss variational equations follow trivially. □

The Gauss equation for l, parameterized using the classical orbital elements, is given by Battin (1999, §10.4). For completeness, we restate it here in terms of the Milankovitch vectorial elements:

$$\dot{l} = \left(-\frac{e}{\mu(1+\sqrt{1-e^2})}\left[H(\hat{\boldsymbol{e}} \cdot \hat{\boldsymbol{r}})\hat{\boldsymbol{r}} + (r+p)(\hat{\boldsymbol{e}} \cdot \boldsymbol{v})\hat{\boldsymbol{\theta}}\right] - \frac{2}{na^2}\boldsymbol{r} + \frac{\hat{\boldsymbol{z}} \cdot \boldsymbol{r}}{H(H + \hat{\boldsymbol{z}} \cdot \boldsymbol{H})}\boldsymbol{H}\right) \cdot \boldsymbol{a}_d + n, \tag{2.31}$$

where $\hat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}}/H$. Note that Equation 2.31 consists of terms that collect the contributions due to the three components (in the radial, transverse, and normal directions) of the disturbing acceleration vector $\boldsymbol{a}_d$, while in Equation 2.30c only the contribution stemming from the normal direction is relevant, yielding a much simpler variational equation. Also, the Gauss equation for L, Equation 2.30c, is valid for all Keplerian orbits with nonzero angular momentum, whereas Equation 2.31 for $\dot{l}$ is only valid for elliptical orbits in which $e < 1$.

2.3 The Method of Averaging

The method of averaging, or theory of secular perturbation, originated fundamentally in the circumstance that the usual procedure of integration by series expansion brought to light terms proportional to the time (Dziobek 1892). This method, being one of the oldest and most developed topics of investigation in celestial mechanics,[17] provides a particularly effective method for the approximate evaluation of the dynamics of a system (Arnold, Kozlov & Neishtadt 2006). In the classical Laplace-Lagrange secular theory and the Gaussian ring-averaging method (Hill 1882; Murray & Dermott 1999, §7.6), as in that of Delaunay, Halphen, Poincaré-Lindstedt, Brouwer-von Zeipel, and the modern Hamiltonian methods based on the Lie transform (Hori 1966; Deprit 1969), the object is to eliminate the short-period perturbation terms and derive the equations of motion that capture the secular evolution of the system (Boccaletti & Pucacco 1999; Ferraz-Mello 2007; Sanders, Verhulst & Murdock 2007). When the equations of motion do not have a canonical form, short-period terms can be eliminated systematically by means of the Krylov-Bogoliubov-Mitropolsky method of averaging, developed for the asymptotic analysis of nonlinear dynamical systems (Bogoliubov & Mitropolsky 1961). This method taken at the first order alone provides a direct means of obtaining what can be considered the averaged or secular form of the equations of motion. The averaged equations can be numerically integrated, with significantly reduced computational requirements, or, in some cases, solved in closed form (e.g., Bohr 1918; Klein 1924; Struble 1961; Mignard & Hénon 1984; Richter & Keller 1995). This approach allows one to easily capture both the qualitative and quantitative effects of perturbations over very long (but, in general, finite) timespans, and reveals the essential dependence of the evolution on the system parameters in a more satisfactory way than does a numerical solution of the non-averaged equations.

[17] Sanders, Verhulst & Murdock (2007) give a detailed historical survey of the development and justification of the averaging principle up to the middle of the twentieth century. They also provide a theoretical discussion on the mathematical bases of averaging.

2.3.1 First-Order Averaging

The basic idea in the averaging method is to obtain approximate equations for the evolution that contain only slowly changing variables by exploiting the presence of a small dimensionless parameter ϵ that characterizes the size of the perturbation. The tacit assumption is that the perturbing forces are sufficiently weak enough that these approximate secular equations of motion can be used to describe the orbital evolution. The perturbation equations in celestial mechanics, relating the time variation of the orbit parameters to the perturbing accelerations, are nonlinear first-order differential equations of the general form

$$\dot{\boldsymbol{x}} = \epsilon \boldsymbol{g}(\boldsymbol{x}, t), \tag{2.32}$$

where $\boldsymbol{g}(\boldsymbol{x}, t)$ is assumed to be T-periodic in t. Equation 2.32 is trivially solved when $\epsilon = 0$, yielding the integrals in the unperturbed problem. The method of averaging consists in replacing Equation 2.32 by the averaged autonomous system

$$\dot{\bar{\boldsymbol{x}}} = \epsilon \bar{\boldsymbol{g}}(\bar{\boldsymbol{x}}), \tag{2.33}$$

$$\bar{\boldsymbol{g}}(\bar{\boldsymbol{x}}) = \frac{1}{T} \int_0^T \boldsymbol{g}(\boldsymbol{x}, t)\, \mathrm{d}t, \tag{2.34}$$

in which the average is performed over time, and it is understood that $\boldsymbol{x}$ in the integrand is to be regarded as a constant during the averaging process. The basis for this approximation is the averaging principle which states that in the general (non-resonant) case, the short period terms removed by averaging cause only small oscillations, which are superimposed on the long-term drift described by the averaged system (Arnold, Kozlov & Neishtadt 2006); such conjectures seemed to be substantiated to a certain extent.

2.3.2 Averaging the Perturbation Equations

Secular Lagrange Equations : A characteristic feature of the classical Lagrange planetary equations is that the disturbing function can be averaged prior to application to the system (Scheeres 2012b, §6.6.4). This remarkable property, which greatly facilitates the derivation of the

averaged equations, remains valid only for the element set $\boldsymbol{H}$, $\boldsymbol{e}$, and l, because of the (incidental) time independence of their Poisson brackets. The averaged disturbing function is defined as

$$\bar{\mathcal{R}}(\boldsymbol{\alpha}) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{R}(\boldsymbol{\alpha}, M)\,\mathrm{d}M, \tag{2.35}$$

where the averaging is (generally) performed over the mean anomaly through the transformation $t = M/n$ (the orbit period $T = 2\pi/n$), while keeping the remaining orbital elements constant. Note that $\bar{\mathcal{R}}(\boldsymbol{\alpha})$ is independent of the mean anomaly, in which case the fast variable gets averaged out. By limiting the disturbing function to its secular part, the semi-major axis becomes a constant of the motion, as can be seen from the classical Lagrange equation for $\dot{a}$ (Brouwer & Clemence 1961a, Chap. 16, §2). This fact is a reflection of the classical Laplace theorem on the immunity of the semi-major axis to secular change. Since the semi-major axis is secularly invariable, $\boldsymbol{H}$ can be scaled by $\sqrt{\mu a}$. The first-order averaged equations for the scaled angular momentum vector, $\boldsymbol{h}$, and the eccentricity vector take the symmetric and elegant form (Allan & Cook 1964)

$$\dot{\boldsymbol{h}} = \tilde{\bar{\boldsymbol{h}}} \cdot \frac{\partial \bar{\mathcal{R}}^*}{\partial \bar{\boldsymbol{h}}} + \tilde{\bar{\boldsymbol{e}}} \cdot \frac{\partial \bar{\mathcal{R}}^*}{\partial \bar{\boldsymbol{e}}}, \tag{2.36}$$

$$\dot{\boldsymbol{e}} = \tilde{\bar{\boldsymbol{e}}} \cdot \frac{\partial \bar{\mathcal{R}}^*}{\partial \bar{\boldsymbol{h}}} + \tilde{\bar{\boldsymbol{h}}} \cdot \frac{\partial \bar{\mathcal{R}}^*}{\partial \bar{\boldsymbol{e}}}, \tag{2.37}$$

where the over bar indicates a mean value of the quantity under it, and $\bar{\mathcal{R}}^* = \bar{\mathcal{R}}(\boldsymbol{\alpha})/\sqrt{\mu a}$ represents the scaled averaged disturbing function of the perturbation; which, by virtue of the orbit averaging, is independent of time and therefore a first integral of the secular system. These equations admit two additional integrals: $\bar{\boldsymbol{h}} \cdot \bar{\boldsymbol{e}}$ and $\bar{\boldsymbol{h}} \cdot \bar{\boldsymbol{h}} + \bar{\boldsymbol{e}} \cdot \bar{\boldsymbol{e}}$; physically meaningful solutions are restricted to the four-dimensional manifold on which $\bar{\boldsymbol{h}} \cdot \bar{\boldsymbol{e}} = 0$ and $\bar{\boldsymbol{h}} \cdot \bar{\boldsymbol{h}} + \bar{\boldsymbol{e}} \cdot \bar{\boldsymbol{e}} = 1$ (Tremaine, Touma & Namouni 2009). Topologically, the structure of bounded Keplerian orbits of fixed energy (semi-major axis) is equivalent to the product of two two-dimensional spheres. This consequence, established by Moser (1970) during his studies on the equivalence of Keplerian motion with geodesic flows, is the deeper reason for the singularities introduced by the use of the classical elements (q.v., Stiefel & Scheifele 1971). The averaged equation for the mean longitude is given by

$$\dot{\bar{l}} = -\frac{\bar{\boldsymbol{h}} + \bar{h}\hat{\boldsymbol{z}}}{\bar{h} + \hat{\boldsymbol{z}} \cdot \bar{\boldsymbol{h}}} \cdot \frac{\partial \bar{\mathcal{R}}^*}{\partial \bar{\boldsymbol{h}}} + \left(\frac{\bar{h}}{1 + \bar{h}} \bar{\boldsymbol{e}} + \frac{\hat{\boldsymbol{z}} \cdot \bar{\boldsymbol{e}}}{\bar{h}(\bar{h} + \hat{\boldsymbol{z}} \cdot \bar{\boldsymbol{h}})} \bar{\boldsymbol{h}} \right) \cdot \frac{\partial \bar{\mathcal{R}}^*}{\partial \bar{\boldsymbol{e}}} + \bar{n}. \tag{2.38}$$

Secular Gauss Equations : The Gauss equations have time-varying terms multiplying the accelerations involving the true anomaly, and thus they must each be averaged separately. The secular evolution of these elements can be computed as

$$\dot{\bar{\boldsymbol{H}}} = \frac{1}{2\pi}\int_0^{2\pi} \dot{\boldsymbol{H}}\,\mathrm{d}M, \tag{2.39}$$

$$\dot{\bar{\boldsymbol{e}}} = \frac{1}{2\pi}\int_0^{2\pi} \dot{\boldsymbol{e}}\,\mathrm{d}M, \tag{2.40}$$

$$\dot{\bar{L}} = \frac{1}{2\pi}\int_0^{2\pi} \dot{L}\,\mathrm{d}M, \tag{2.41}$$

$$\dot{\bar{l}} = \frac{1}{2\pi}\int_0^{2\pi} \dot{l}\,\mathrm{d}M. \tag{2.42}$$

Averaging in Multiple Timescale Systems : Up to this point, we have assumed that the perturbing acceleration will be periodic in time, and, therefore, that our equations contain only one fast variable. The case of multiple timescales or frequencies over which the relevant dynamical motion occurs can be treated in an analogous way, provided there is no commensurability of frequencies of the fast variables (i.e., resonances). If a resonance is present, then a special study is needed and the resulting secular equations take a qualitatively different form. Let the two timescales be defined by an orbital rate n and an angular rate N, generally associated with the motion of an external body in the system. In the non-resonance case, if sufficient distance between the two timescales exists (i.e., $N/n \ll 1$), it is acceptable to hold N constant while averaging over n; in which case, the remaining system has a time-varying term associated with N. If the singly-averaged disturbing function is a function of a second periodic term of period T, or $\mathcal{R}(\boldsymbol{x}, t, M) = \mathcal{R}(\boldsymbol{x}, t+T, M)$, then another independent averaging may be performed:

$$\bar{\bar{\mathcal{R}}}(\bar{\bar{\boldsymbol{x}}}) = \frac{1}{2\pi T}\int_0^{T}\int_0^{2\pi} \mathcal{R}(\boldsymbol{x}, s, M)\,\mathrm{d}M\mathrm{d}s. \tag{2.43}$$

The doubly-averaged disturbing function can then be substituted into Equations 2.36, 2.37, and 2.38, yielding a greatly simplified set of equations. A similar procedure can be carried out for the secular Gauss equations.

Chapter 3

Averaged Model of Gravitational and Non-Gravitational Perturbations

We shall now consider separately the first-order effects of each principal perturbation in altering the Milankovitch elements; in Chapters 4 and 5, respectively, we examine stable "frozen" orbit configurations and resonances which can occur when these perturbations act in concert. Appendix C contains a systematic account of the development of averaging in elliptic motion and gathers the averaged formulae that are used in the computation of the secular equations of motion.

3.1 Solar Radiation Pressure

The remarkable fact that the secular motion of an orbiter subject to solar radiation pressure, and orbiting about a point mass, can be solved in closed-form has been known for some time. Mignard & Hénon (1984) showed that a particle disturbed by solar radiation pressure, modeled in its simplest form and averaged over the particle's unperturbed orbit, is analogous to the Stark effect once the secularized SRP problem is cast into a rotating frame (q.v. Bohr 1918).[1] Solar radiation pressure was treated as a constant force in magnitude, and directed along the Sun-planet line, and the planet's heliocentric orbit was assumed to be circular. Richter & Keller (1995), using the secular Milankovitch equations and accounting for the elliptic motion of the planet about the Sun, generalized Mignard and Hénon's solution to an inverse-square law SRP force (see also Scheeres 2012b). These earlier analyses are limited to the cannonball model of solar radiation pressure, which was discussed in §1.3.1. Although this captures the main aspects of SRP, higher-fidelity

[1] See also Lantoine & Russell (2011) for a complete account on the integrability of the Stark problem (i.e., the Kepler problem perturbed by a uniform force of constant magnitude and direction).

models are needed for accurate long-term predictions. In this section, we consider the general SRP model (Equation 1.3) that describes the interaction of the solar photons on an arbitrary body. We establish that the secular problem is integrable (under certain assumptions) and that the averaged cannonball solution can be easily recovered from this more general result.

3.1.1 Averaged SRP Dynamics

The disturbing function associated with the SRP acceleration, modeled by Equation 1.3, can be written as

$$\mathcal{R}_{srp} = \frac{\beta'}{d_s^2}\hat{\boldsymbol{a}} \cdot \boldsymbol{r}, \tag{3.1}$$

provided the object is sufficiently close to the planet (i.e., $r/d_s \ll 1$). Using the Lagrange planetary equations for the element set $\boldsymbol{H}$, $\boldsymbol{e}$, and l, the disturbing function can be averaged prior to application in this system. From Equations 2.35 and 3.1, we can compute the averaged disturbing function as

$$\bar{\mathcal{R}}_{srp} = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{R}_{srp}\,\mathrm{d}M \tag{3.2}$$

$$= \frac{\beta'}{d_s^2}\hat{\boldsymbol{a}} \cdot \bar{\boldsymbol{r}}, \tag{3.3}$$

where we initially hold the direction $\hat{\boldsymbol{a}}$ fixed in inertial space.[2] Thus, we only need to compute the average of the position vector, a classically known result[3]

$$\bar{\boldsymbol{r}} = -\frac{3}{2}a\boldsymbol{e}. \tag{3.4}$$

[2] The same form of the equations holds if $\hat{\boldsymbol{a}}$ is assumed to be fixed or assumed to be rotating about the Sun. Note that the bar ($\bar{-}$) operator is omitted from the Milankovitch elements in what follows because there is no ambiguity; i.e., all variables are averaged variables.

[3] From a historical point of view, for many problems in celestial mechanics it was necessary to express the fundamental equations of elliptic motion by means of periodic functions in the form of series (see, for example, Brouwer & Clemence 1961a, Chap. 2). The development of the position vector by Bessel functions in terms of the mean anomaly is given by Tisserand (1896, Vol. 1, Chap. 13, p. 226); the constant term in the series is simply the average value over the orbit. Bohr (1918), in the framework of the old quantum theory, called $\bar{\boldsymbol{r}}$ the position of the 'electrical centre' of the orbit, and established the relation in Equation 3.4 based on this classical elliptic expansion. Lenz (1924) found the relation $\bar{\boldsymbol{r}} = -\frac{3}{2}a_n\boldsymbol{b}$, where $\boldsymbol{b}$ is the Laplace-Runge-Lenz vector for the Coulomb problem and a_n is the semi-major axis of the Kepler ellipse having the quantum number n. In Heisenberg's matrix mechanics, Pauli (1926) demonstrated that the same relation holds between the corresponding Hermitian matrices of the dynamical variables as in the classical theory by calculating the quantum-mechanical equivalent of averaging over orbit cycles.

This leads to

$$\bar{\mathcal{R}}^*_{srp} = -\frac{3a_{srp}}{2}\sqrt{\frac{a}{\mu}}\hat{\boldsymbol{a}}\cdot\boldsymbol{e}, \tag{3.5}$$

where $a_{srp} = \beta'/d_s^2$. Stated in this form, the scaled averaged disturbing function $\bar{\mathcal{R}}^*_{srp}$ can be substituted into the secular Milankovitch equations, Equations 2.36 and 2.37, giving

$$\dot{\boldsymbol{h}}_{srp} = \frac{3a_{srp}}{2}\sqrt{\frac{a}{\mu}}\widetilde{\hat{\boldsymbol{a}}}\cdot\boldsymbol{e}, \tag{3.6}$$

$$\dot{\boldsymbol{e}}_{srp} = \frac{3a_{srp}}{2}\sqrt{\frac{a}{\mu}}\widetilde{\hat{\boldsymbol{a}}}\cdot\boldsymbol{h}. \tag{3.7}$$

Note that Equations 3.6 and 3.7 are autonomous linear equations with constant coefficients if the motion of $\hat{\boldsymbol{a}}$ is neglected. As such, they can be solved in closed form, with the details given by Scheeres (2012a).

The secular evolution of the mean longitude can be found from Equations 2.38 and 3.5 as

$$\dot{l} = -\frac{3a_{srp}}{2}\sqrt{\frac{a}{\mu}}\left(\frac{h}{1+h}\boldsymbol{e} + \frac{\hat{\boldsymbol{z}}\cdot\boldsymbol{e}}{h(h+\hat{\boldsymbol{z}}\cdot\boldsymbol{h})}\boldsymbol{h}\right)\cdot\hat{\boldsymbol{a}} + n. \tag{3.8}$$

Once the solutions for $\boldsymbol{h}$ and $\boldsymbol{e}$ are found, Equation 3.8 reduces to quadratures.

For the sake of completeness, we also derive Equations 3.6 and 3.7 using the Gaussian form of the perturbation equations. Substituting the perturbation acceleration, Equation 1.3, into Equations 2.30a and 2.30b, and scaling the angular momentum by $\sqrt{\mu a}$, the differential equations for $\boldsymbol{h}$ and $\boldsymbol{e}$, resulting from SRP acceleration, can be written as

$$\dot{\boldsymbol{h}}_{srp} = -\frac{\beta'}{\sqrt{\mu a}d_s^2}\widetilde{\hat{\boldsymbol{a}}}\cdot\boldsymbol{r}, \tag{3.9}$$

$$\dot{\boldsymbol{e}}_{srp} = \frac{\beta'}{\mu d_s^2}\left(-\sqrt{\mu a}\widetilde{\boldsymbol{h}}\cdot\hat{\boldsymbol{a}} + \widetilde{\boldsymbol{v}}\cdot\widetilde{\boldsymbol{r}}\cdot\hat{\boldsymbol{a}}\right). \tag{3.10}$$

Averaging these equations amounts to averaging the position vector and the term $\widetilde{\boldsymbol{v}}\cdot\widetilde{\boldsymbol{r}}$; the latter is equivalent to $\boldsymbol{r}\boldsymbol{v} - (\boldsymbol{r}\cdot\boldsymbol{v})\boldsymbol{U}$ and its average can be computed from the results given in Appendix C as

$$\overline{\widetilde{\boldsymbol{v}}\cdot\widetilde{\boldsymbol{r}}} = -\frac{1}{2}\sqrt{\mu a}\widetilde{\boldsymbol{h}}. \tag{3.11}$$

Substituting Equations 3.4 and 3.11 into the above differential equations, we obtain, as expected, exactly the same expressions for the secular evolution of $\boldsymbol{h}$ and $\boldsymbol{e}$ as from the Lagrange approach.

Substituting Equation 1.3 into Equation 2.30c, and averaging, we find

$$\dot{L} = \frac{\beta'}{d_s^2} \frac{\hat{\boldsymbol{z}} \cdot \bar{\boldsymbol{r}}}{\sqrt{\mu a} h (h + \hat{\boldsymbol{z}} \cdot \boldsymbol{h})} \boldsymbol{h} \cdot \hat{\boldsymbol{a}} + \sqrt{\mu a} h \overline{\frac{1}{r^2}}. \tag{3.12}$$

From Hestenes (1999, Chap. 8, §2), we have

$$\overline{\frac{1}{r^2}} = \frac{1}{a^2 \sqrt{1 - e^2}} = \frac{1}{a^2 h}. \tag{3.13}$$

Therefore, the average equation for the true longitude takes the form

$$\dot{L} = -\frac{3 a_{srp}}{2} \sqrt{\frac{a}{\mu}} \frac{\boldsymbol{e} \cdot \hat{\boldsymbol{z}} \hat{\boldsymbol{a}} \cdot \boldsymbol{h}}{h (h + \hat{\boldsymbol{z}} \cdot \boldsymbol{h})} + n, \tag{3.14}$$

which can be reduced to quadratures. It is worth noting from Equations 3.8 and 3.14 that the same object will have different periods in average mean longitude and average true longitude.

3.1.2 Closed-Form Solution to the SRP Perturbed Orbiter Problem

Theorem 1. *Assume that the direction of SRP acceleration is fixed relative to the planet-Sun line. Then, when transformed into a frame rotating with the Sun, and taking the planet's heliocentric true anomaly as the independent variable, the averaged equations for* $\boldsymbol{e}$ *and* $\boldsymbol{h}$ *are reduced to an autonomous linear system, and thus can be solved in closed form. The "state transition" matrix (STM, also called "fundamental" matrix) or dyadic of the system reduces to elementary circular functions, yielding the explicit solution*

$$\begin{bmatrix} \boldsymbol{e}(f) \\ \boldsymbol{h}(f) \end{bmatrix} = \boldsymbol{\Phi}(f - f_o) \cdot \begin{bmatrix} \boldsymbol{e}_o \\ \boldsymbol{h}_o \end{bmatrix}, \tag{3.15}$$

$$\boldsymbol{\Phi}(f) = \frac{1}{2} \begin{bmatrix} \left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos \Lambda'} \right) + \cos\left(\frac{f\sqrt{1+\nu}}{\cos \Lambda'} \right) \right) \boldsymbol{U} & \left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos \Lambda'} \right) - \cos\left(\frac{f\sqrt{1+\nu}}{\cos \Lambda'} \right) \right) \boldsymbol{U} \\ \left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos \Lambda'} \right) - \cos\left(\frac{f\sqrt{1+\nu}}{\cos \Lambda'} \right) \right) \boldsymbol{U} & \left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos \Lambda'} \right) + \cos\left(\frac{f\sqrt{1+\nu}}{\cos \Lambda'} \right) \right) \boldsymbol{U} \end{bmatrix}$$

$$+ \begin{bmatrix} F_1 \left(\cos^2 \Lambda' \hat{\boldsymbol{H}}_s \hat{\boldsymbol{H}}_s + \sin^2 \Lambda' \hat{\boldsymbol{a}} \hat{\boldsymbol{a}} \right) - F_2 \sin \Lambda' \cos \Lambda' \left(\hat{\boldsymbol{H}}_s \hat{\boldsymbol{a}} + \hat{\boldsymbol{a}} \hat{\boldsymbol{H}}_s \right) \\ F_2 \left(\cos^2 \Lambda' \hat{\boldsymbol{H}}_s \hat{\boldsymbol{H}}_s + \sin^2 \Lambda' \hat{\boldsymbol{a}} \hat{\boldsymbol{a}} \right) - F_1 \sin \Lambda' \cos \Lambda' \left(\hat{\boldsymbol{H}}_s \hat{\boldsymbol{a}} + \hat{\boldsymbol{a}} \hat{\boldsymbol{H}}_s \right) \end{bmatrix}$$

$$
\left.
\begin{array}{r}
F_2\left(\cos^2\Lambda'\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s+\sin^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}\right)-F_1\sin\Lambda'\cos\Lambda'\left(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}}+\hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s\right)\\
F_1\left(\cos^2\Lambda'\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s+\sin^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}\right)-F_2\sin\Lambda'\cos\Lambda'\left(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}}+\hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s\right)
\end{array}
\right]
$$

$$
+\left[
\begin{array}{cc}
-G_1\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_2\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}} & -G_2\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_1\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}}\\
-G_2\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_1\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}} & -G_1\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_2\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}}
\end{array}
\right], \tag{3.16}
$$

in which $\boldsymbol{U}$ *denotes the 3-dimensional identity dyadic, and the coefficients* F_1, F_2, G_1, *and* G_2 *are scalar functions of* ν, f, *and* Λ', *given by*

$$
F_1=\frac{1}{2}\left(\frac{2}{1-\nu^2}-\frac{1}{1-\nu}\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)-\frac{1}{1+\nu}\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right), \tag{3.17a}
$$

$$
F_2=\frac{1}{2}\left(\frac{2\nu}{1-\nu^2}-\frac{1}{1-\nu}\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)+\frac{1}{1+\nu}\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right), \tag{3.17b}
$$

$$
G_1=\frac{1}{2}\left(\frac{1}{\sqrt{1-\nu}}\sin\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)+\frac{1}{\sqrt{1+\nu}}\sin\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right), \tag{3.17c}
$$

$$
G_2=\frac{1}{2}\left(\frac{1}{\sqrt{1-\nu}}\sin\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)-\frac{1}{\sqrt{1+\nu}}\sin\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right), \tag{3.17d}
$$

where f *is the planet's true anomaly,* Λ' *is the SRP perturbation angle defined by Equation 3.22, and* $\nu=\sin(2\Lambda')(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})$, *with* $|\nu|\leq 1$. *Equation 3.15 is the particular solution to the initial-value problem posed by the differential system obtained in Equation 3.24, with the initial conditions* $\boldsymbol{e}(f_o)=\boldsymbol{e}_o$, $\boldsymbol{h}(f_o)=\boldsymbol{h}_o$.

Proof. The secular evolution of the scaled angular momentum vector and the eccentricity vector is governed by Equations 3.6 and 3.7, when the direction in which the SRP acceleration acts is fixed. Consider a frame rotating with the planet-Sun line. The rotation rate equals the planet's true anomaly rate of $\dot{f}=H_p/d_s^2$, where H_p is the specific angular momentum of the planet about the Sun. The rotation direction is perpendicular to the planet-Sun line and defined as the direction $\hat{\boldsymbol{H}}_s$; that is, $\hat{\boldsymbol{H}}_s$ is the planet's angular momentum unit vector. Under the assumption that the orbit rate about the planet is fast as compared to the orbit rate of the planet about the Sun, the same equations for $\dot{\boldsymbol{h}}$ and $\dot{\boldsymbol{e}}$ can be used. However, these equations must be shifted into the rotating frame using the transport theorem (i.e., formulae for differentiation in moving reference frames, relating variations of a vector in a moving, non-inertial, frame to its variations in an inertial

reference system; see, e.g., Battin 1999, §2.5, pp. 101–106.) Given a frame rotational velocity vector of $\dot{f}\hat{\boldsymbol{H}}_s$, we have

$$\dot{\boldsymbol{e}}_r + \dot{f}\widetilde{\hat{\boldsymbol{H}}_s} \cdot \boldsymbol{e} = \frac{3a_{srp}}{2}\sqrt{\frac{a}{\mu}}\widetilde{\hat{\boldsymbol{a}}} \cdot \boldsymbol{h}, \tag{3.18}$$

$$\dot{\boldsymbol{h}}_r + \dot{f}\widetilde{\hat{\boldsymbol{H}}_s} \cdot \boldsymbol{h} = \frac{3a_{srp}}{2}\sqrt{\frac{a}{\mu}}\widetilde{\hat{\boldsymbol{a}}} \cdot \boldsymbol{e}, \tag{3.19}$$

where the r subscript indicates time derivative with respect to a rotating frame and will be dropped from this point on. We can then write the secular equations as a linear system

$$\begin{bmatrix} \dot{\boldsymbol{e}} \\ \dot{\boldsymbol{h}} \end{bmatrix} = \frac{3a_{srp}}{2}\sqrt{\frac{a}{\mu}} \begin{bmatrix} -\dfrac{2\dot{f}}{3a_{srp}}\sqrt{\dfrac{\mu}{a}}\widetilde{\hat{\boldsymbol{H}}_s} & \widetilde{\hat{\boldsymbol{a}}} \\ \widetilde{\hat{\boldsymbol{a}}} & -\dfrac{2\dot{f}}{3a_{srp}}\sqrt{\dfrac{\mu}{a}}\widetilde{\hat{\boldsymbol{H}}_s} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{h} \end{bmatrix}. \tag{3.20}$$

These linear differential equations are not autonomous, as both $\dot{f}$ and a_{srp} vary in time. However, we note that the ratio $\dot{f}/a_{srp}$ is time-independent, as both vary inversely with d_s^2 and hence

$$\frac{2\dot{f}}{3a_{srp}}\sqrt{\frac{\mu}{a}} = \frac{2H_p}{3\beta'}\sqrt{\frac{\mu}{a}}. \tag{3.21}$$

This quantity is a constant for a given body and its orbit. For convenience, and to remain consistent with Mignard & Hénon (1984), we define an equivalent quantity

$$\tan\Lambda' = \frac{3\beta'}{2H_p}\sqrt{\frac{a}{\mu}}. \tag{3.22}$$

As the SRP perturbation becomes strong, $\Lambda' \to \pi/2$; and as it becomes weak $\Lambda' \to 0$.

Despite the time invariance of the ratio, the multiplying factor of the matrix is still time varying. This can be eliminated, however, by changing the independent parameter from time to the true anomaly of the planet about the Sun. To make this transformation we let

$$\dot{\boldsymbol{e}} = \frac{\mathrm{d}\boldsymbol{e}}{\mathrm{d}f}\dot{f} = \boldsymbol{e}'\dot{f}, \quad \dot{\boldsymbol{h}} = \frac{\mathrm{d}\boldsymbol{h}}{\mathrm{d}f}\dot{f} = \boldsymbol{h}'\dot{f}. \tag{3.23}$$

Then, the factor in front of the matrix becomes our newly defined SRP strength parameter, leading to the autonomous linear differential equations

$$\begin{bmatrix} \boldsymbol{e}' \\ \boldsymbol{h}' \end{bmatrix} = \begin{bmatrix} -\widetilde{\hat{\boldsymbol{H}}_s} & \tan\Lambda'\widetilde{\hat{\boldsymbol{a}}} \\ \tan\Lambda'\widetilde{\hat{\boldsymbol{a}}} & -\widetilde{\hat{\boldsymbol{H}}_s} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{h} \end{bmatrix}. \tag{3.24}$$

The state transition (or fundamental) matrix or dyadic for this system can be obtained using the power series definition of the matrix exponential, as demonstrated in Appendix G, giving the explicit solution to the SRP perturbed orbiter problem. □

Remark 5. *The details of this solution rely only on the assumption that SRP acceleration is an inverse-square law, and not on the specific acceleration magnitude or optical properties of the object. It is important to remember that in this analysis we assumed that the direction* $\hat{\boldsymbol{a}}$ *is fixed in the rotating coordinate frame. Note that if the object's orientation changes with respect to the planet-Sun line (i.e., tumbling motion occurs), the object can evolve through a variety of solutions characterized by the SRP perturbation angle* Λ'*.*

As an aside, let us assume that $\hat{\boldsymbol{a}}$ is constrained to lie within the heliocentric orbit plane; that is, $\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{a}} = 0$ and hence, $\nu = 0$. This leads to

$$F_1 = 1 - \cos\left(\frac{f}{\cos\Lambda'}\right), \quad F_2 = 0, \quad G_1 = \sin\left(\frac{f}{\cos\Lambda'}\right), \quad G_2 = 0. \tag{3.25}$$

Substituting these expressions into Equation 3.16, we find

$$\begin{bmatrix} \boldsymbol{e}(\psi) \\ \boldsymbol{h}(\psi) \end{bmatrix} = \boldsymbol{\Phi}(\psi - \psi_o) \cdot \begin{bmatrix} \boldsymbol{e}_o \\ \boldsymbol{h}_o \end{bmatrix}, \tag{3.26}$$

$$\begin{aligned} \boldsymbol{\Phi}(\psi) = {} & \cos\psi \boldsymbol{U} \\ & + (1 - \cos\psi) \begin{bmatrix} \cos^2\Lambda' \hat{\boldsymbol{H}}_s \hat{\boldsymbol{H}}_s + \sin^2\Lambda' \hat{\boldsymbol{a}}\hat{\boldsymbol{a}} & -\sin\Lambda'\cos\Lambda' \left(\hat{\boldsymbol{H}}_s \hat{\boldsymbol{a}} + \hat{\boldsymbol{a}} \hat{\boldsymbol{H}}_s\right) \\ -\sin\Lambda'\cos\Lambda' \left(\hat{\boldsymbol{H}}_s \hat{\boldsymbol{a}} + \hat{\boldsymbol{a}} \hat{\boldsymbol{H}}_s\right) & \cos^2\Lambda' \hat{\boldsymbol{H}}_s \hat{\boldsymbol{H}}_s + \sin^2\Lambda' \hat{\boldsymbol{a}}\hat{\boldsymbol{a}} \end{bmatrix} \\ & + \sin\psi \begin{bmatrix} -\cos\Lambda' \widetilde{\hat{\boldsymbol{H}}_s} & \sin\Lambda' \widetilde{\hat{\boldsymbol{a}}} \\ \sin\Lambda' \widetilde{\hat{\boldsymbol{a}}} & -\cos\Lambda' \widetilde{\hat{\boldsymbol{H}}_s} \end{bmatrix}, \end{aligned} \tag{3.27}$$

where $\psi = f/\cos\Lambda'$. The solutions are 2π-periodic in ψ, repeating their values every time that the true anomaly f is increased by the amount of an integral multiple of $2\pi\cos\Lambda'$. Thus, over one heliocentric orbit, the solution will repeat $1/\cos\Lambda'$ times. As the perturbation grows large, and Λ' approaches $\pi/2$, the solution will repeat many times. On the contrary, as the perturbation grows small, the solution will repeat only once every heliocentric orbit. The STM is also an orthogonal

matrix (i.e., $\mathbf{\Phi}^{-1} = \mathbf{\Phi}^T$), and defines a rotation matrix ($\det \mathbf{\Phi} = 1$) in 6-dimensional space. With this re-orientation of the SRP force, the form of the averaged solution is exactly the same as that for a cannonball model (cf. Richter & Keller 1995, Equation 20; see also Scheeres 2012a, Equations 48 and 49). Thus, if the net direction in which the SRP acceleration acts lies within the planet's heliocentric orbit plane, the object will have similar dynamics to a cannonball.

3.2 Planetary Oblateness

3.2.1 Averaged Oblateness Dynamics

The quadrupole disturbing function arising from planetary oblateness is given by Equation 1.14. Averaging over the orbital motion gives

$$\bar{\mathcal{R}}_2 = \frac{\mu J_2 R^2}{2}\left[\overline{\frac{1}{r^3}} - 3\hat{\boldsymbol{p}}\cdot\overline{\left(\frac{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}{r^3}\right)}\cdot\hat{\boldsymbol{p}}\right], \tag{3.28}$$

where the planet's spin axis $\hat{\boldsymbol{p}}$ is assumed to be fixed in inertial space; that is, we neglect any tidal torques on the planet's equatorial bulge, as this precession period is generally much longer than the precession period of the satellite orbit (Milankovitch 1941; Goldreich 1965; Boué & Laskar 2006). From Musen (1961), we have

$$\overline{\frac{1}{r^3}} = \frac{1}{a^3h^3}, \tag{3.29}$$

$$\overline{\left(\frac{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}{r^3}\right)} = \frac{1}{2a^3h^3}\left[\boldsymbol{U} - \hat{\boldsymbol{h}}\hat{\boldsymbol{h}}\right]. \tag{3.30}$$

Substituting these expressions into the Equation 3.28, and scaling the resulting function by $\sqrt{\mu a}$, yields

$$\bar{\mathcal{R}}_2^* = -\frac{nJ_2R^2}{4a^2h^3}\left[1 - 3(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}})^2\right], \tag{3.31}$$

in which $n = \sqrt{\mu/a^3}$ is the satellite's mean motion. Substituting $\bar{\mathcal{R}}_{20}^*$ into Equations 2.36 and 2.37, the secular equations for the oblateness gravity field perturbation can be written as

$$\dot{\boldsymbol{h}}_2 = -\frac{3nJ_2R^2}{2a^2h^5}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\widetilde{\hat{\boldsymbol{p}}}\cdot\boldsymbol{h}, \tag{3.32}$$

$$\dot{\boldsymbol{e}}_2 = -\frac{3nJ_2R^2}{4a^2h^5}\left\{\left[1 - \frac{5}{h^2}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})^2\right]\widetilde{\boldsymbol{h}} + 2(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\widetilde{\hat{\boldsymbol{p}}}\right\}\cdot\boldsymbol{e}. \tag{3.33}$$

3.2.2 Secular Solution

Taking the scalar product of each side of Equation 3.32 with $\boldsymbol{h}$ and integrating shows that h is constant. Dotting Equation 3.32 with $\hat{\boldsymbol{p}}$ yields zero, implying that the product $\hat{\boldsymbol{p}} \cdot \boldsymbol{h}$ is constant, which means that the orbit plane inclination relative to the rotation pole $\hat{\boldsymbol{p}}$ is a fixed quantity, which follows necessarily because the averaged disturbing function, Equation 3.31, is axisymmetric (i.e., the shape of the orbit is unchanged if $\hat{\boldsymbol{h}} \to -\hat{\boldsymbol{h}}$). Note that $\widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{h}$ is constant in magnitude but not constant in direction. We will now show that the solution to Equation 3.32 is of the form of a rotation matrix, rotating about $\hat{\boldsymbol{p}}$, times $\boldsymbol{h}_0$.

Assume we have a rotation angle ϕ_1 about the unit vector $\hat{\boldsymbol{p}}$. We can specify the rotation matrix corresponding to this rotation in dyadic notation as (Gibbs & Wilson 1909, p. 338)

$$\boldsymbol{C}_1 = \cos\phi_1 \boldsymbol{U} + (1 - \cos\phi_1)\hat{\boldsymbol{p}}\hat{\boldsymbol{p}} - \sin\phi_1 \widetilde{\hat{\boldsymbol{p}}}. \tag{3.34}$$

We then let $\boldsymbol{h} = \boldsymbol{C}_1 \cdot \boldsymbol{h}_0$ and insert it into the differential equation (Equation 3.32) to obtain

$$\dot{\boldsymbol{C}}_1 \cdot \boldsymbol{h}_0 = \dot{\Omega}\left(\widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{C}_1 \cdot \boldsymbol{h}_0\right), \tag{3.35}$$

where the constant $\dot{\Omega}$ is given by

$$\dot{\Omega} = -\frac{3nJ_2R^2}{2a^2h^5}(\hat{\boldsymbol{p}} \cdot \boldsymbol{h}). \tag{3.36}$$

Since the orientation of $\hat{\boldsymbol{p}}$ is fixed, ϕ_1 is the only thing that changes in the time derivative of $\boldsymbol{C}_1$. Also note that

$$\widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{C}_1 \cdot \boldsymbol{h}_0 = \cos\phi_1 \widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{h}_0 - \sin\phi_1(\hat{\boldsymbol{p}} \cdot \boldsymbol{h}_0)\hat{\boldsymbol{p}} + \sin\phi_1 \boldsymbol{h}_0.$$

Therefore, Equation 3.35 reduces to

$$-\dot{\phi}_1(\sin\phi_1 \boldsymbol{h}_0 - \sin\phi_1 \hat{\boldsymbol{p}}\hat{\boldsymbol{p}} \cdot \boldsymbol{h}_0 + \cos\phi_1 \widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{h}_0) = \dot{\Omega}(\cos\phi_1 \widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{h}_0 - \sin\phi_1(\hat{\boldsymbol{p}} \cdot \boldsymbol{h}_0)\hat{\boldsymbol{p}} + \sin\phi_1 \boldsymbol{h}_0).$$

Grouping like terms yields

$$-(\dot{\phi}_1 + \dot{\Omega})\sin\phi_1 \boldsymbol{h}_0 - (\dot{\phi}_1 + \dot{\Omega})\cos\phi_1 \widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{h}_0 + \dot{\phi}_1 \sin\phi_1 \hat{\boldsymbol{p}}\hat{\boldsymbol{p}} \cdot \boldsymbol{h}_0 + \dot{\Omega}\sin\phi_1(\hat{\boldsymbol{p}} \cdot \boldsymbol{h}_0)\hat{\boldsymbol{p}} = \boldsymbol{0}.$$

Analysis of this equation will show that it yields a single differential equation for $\dot{\phi}_1$ that is constant. Namely,

$$\dot{\phi}_1 + \dot{\Omega} = 0. \tag{3.37}$$

Hence, if we let $\phi_1 = -\Omega$, then the explicit solution to Equation 3.32 is given by

$$\boldsymbol{h} = \boldsymbol{C}_1 \cdot \boldsymbol{h}_0, \tag{3.38}$$

$$\boldsymbol{C}_1 = \cos\Omega \boldsymbol{U} + (1 - \cos\Omega)\hat{\boldsymbol{p}}\hat{\boldsymbol{p}} + \sin\Omega\widetilde{\hat{\boldsymbol{p}}}, \tag{3.39}$$

where $\Omega(t) = \Omega_0 + \dot{\Omega}t$, and $\dot{\Omega}$ is given in Equation 3.36. Therefore, Equation 3.32 describes retrograde precession of the satellite's angular momentum vector around the planet's rotation axis, with angular frequency $\dot{\Omega}$.

The differential equation for $\boldsymbol{e}$, Equation 3.33, can be rewritten as

$$\dot{\boldsymbol{e}}_2 = \dot{\Omega}\widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{e} + \dot{\omega}\widetilde{\hat{\boldsymbol{h}}} \cdot \boldsymbol{e}, \tag{3.40}$$

where $\dot{\omega}$ is also a constant by virtue of the previous arguments:

$$\dot{\omega} = -\frac{3nJ_2R^2}{4a^2h^5}\left[1 - \frac{5}{h^2}(\hat{\boldsymbol{p}} \cdot \boldsymbol{h})^2\right]. \tag{3.41}$$

Dotting Equation 3.40 with $\boldsymbol{e}$ yields zero implying that the eccentricity is constant. We will show that the solution to Equation 3.40 has the form of two successive rotations, one about the rotation pole $\hat{\boldsymbol{p}}$ and the other about the orbit normal $\hat{\boldsymbol{h}}$.

We rewrite Equation 3.40 in a rotating frame, rotating with the angular momentum vector (accounting for its J_2 dynamics given in Equation 3.38). The frame rotation rate equals $\dot{\Omega}$ and the rotation is about the symmetry axis of the mass distribution. Given a frame rotational velocity of $\dot{\Omega}\hat{\boldsymbol{p}}$, we find

$$\dot{\boldsymbol{e}}_r + \dot{\Omega}\widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{e} = \dot{\Omega}\widetilde{\hat{\boldsymbol{p}}} \cdot \boldsymbol{e} + \dot{\omega}\widetilde{\hat{\boldsymbol{h}}} \cdot \boldsymbol{e},$$

which reduces to

$$\dot{\boldsymbol{e}}_r = \dot{\omega}\widetilde{\hat{\boldsymbol{h}}} \cdot \boldsymbol{e}. \tag{3.42}$$

We now show that the solution to Equation 3.42 is of the form of a rotation matrix, rotating about $\hat{\boldsymbol{h}}$, times $\boldsymbol{e}_0$. The rotation matrix corresponding to a rotation angle ϕ_2 about $\hat{\boldsymbol{h}}$ can be specified as

$$\boldsymbol{C}_2 = \cos\phi_2 \boldsymbol{U} + (1 - \cos\phi_2)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}} - \sin\phi_2 \widetilde{\hat{\boldsymbol{h}}}. \tag{3.43}$$

Letting $\boldsymbol{e} = \boldsymbol{C}_2 \cdot \boldsymbol{e}_0$ and inserting it into the differential equation, we find

$$\dot{\boldsymbol{C}}_2 \cdot \boldsymbol{e}_0 = \dot{\omega}\left(\widetilde{\hat{\boldsymbol{h}}} \cdot \boldsymbol{C}_2 \cdot \boldsymbol{e}_0\right). \tag{3.44}$$

We note that the angular momentum vector is fixed in this rotating frame; therefore, ϕ_2 is the only thing that changes in the time derivative of $\boldsymbol{C}_2$. Performing the differentiation and the dyadic dot product, we find that Equation 3.44 simplifies to

$$-(\dot{\phi}_2 + \dot{\omega})\sin\phi_2 \boldsymbol{e}_0 - (\dot{\phi}_2 + \dot{\omega})\cos\phi_2 \widetilde{\hat{\boldsymbol{h}}} \cdot \boldsymbol{e}_0 + \dot{\phi}_2 \sin\phi_2 \hat{\boldsymbol{h}}\hat{\boldsymbol{h}} \cdot \boldsymbol{e}_0 + \dot{\omega}\sin\phi_2(\hat{\boldsymbol{h}} \cdot \boldsymbol{e}_0)\hat{\boldsymbol{h}} = \boldsymbol{0}.$$

Analysis of this equation shows that $\dot{\phi}_2 + \dot{\omega} = 0$, which implies that $\phi_2 = -\omega$. Thus, the explicit solution to the rotating frame differential equation, Equation 3.42, is given by

$$\boldsymbol{e} = \boldsymbol{C}_2 \cdot \boldsymbol{e}_0, \tag{3.45}$$

$$\boldsymbol{C}_2 = \cos\omega \boldsymbol{U} + (1 - \cos\omega)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}} + \sin\omega \widetilde{\hat{\boldsymbol{h}}}, \tag{3.46}$$

where $\omega(t) = \omega_0 + \dot{\omega}t$, $\dot{\omega}$ is given in Equation 3.41, and $\hat{\boldsymbol{h}}$ follows its dynamics given by Equation 3.38.

Finally, given the solution for eccentricity in the rotating frame, we can map it into the inertial frame by premultiplying by $\boldsymbol{C}_1$, which rotates a vector in the angular momentum-fixed frame into inertial space. Combining successive rotations, the final solution to Equation 3.33 is

$$\boldsymbol{e} = \boldsymbol{C}_1 \cdot \boldsymbol{C}_2 \cdot \boldsymbol{e}_0, \tag{3.47}$$

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are given in Equations 3.39 and 3.46, respectively.

In the classical problem of the motion of a satellite in the field of an oblate planet, the rates $\dot{\Omega}$ and $\dot{\omega}$ can be identified as the regression of the nodes and apsidal precession rate, respectively. Note that if $\hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{h}} = \pm 1/\sqrt{5}$, then $\dot{\omega} = 0$; that is, if the orbit plane inclination relative to the planet's rotation pole is approximately 63.43° of 116.57°, the apsidal rate vanishes. This is known in the

astrodynamics literature as the critical inclination, and has been the subject of much research and a matter of considerable controversy ever since (Roy 2005, §11.4).

3.3 Third-Body Gravitation

For third-body perturbations, there are two fundamental frequencies, defined by the satellite's orbital rate n and the motion of the perturbing body with its angular rate n_p. Averaging in two-frequency systems was discussed at the end of §2.3. We shall restrict our attention to the general case when the timescales between the two periods are sufficiently far apart (and not commensurate). Accordingly, we can average over the fast variables, independently with respect to each of them.

3.3.1 Averaged Third-Body Dynamics

3.3.1.1 Singly-Averaged Equations

The Hill-approximated third-body disturbing function, Equation 1.20, can be averaged over the satellite's unperturbed two-body motion about the planet as

$$\bar{\mathcal{R}}_p = \frac{\mu_p}{2d_p^3}\left[3\hat{\boldsymbol{d}}_p \cdot \overline{\boldsymbol{rr}} \cdot \hat{\boldsymbol{d}}_p - \overline{r^2}\right], \tag{3.48}$$

in which we note (q.v., Appendix C)

$$\overline{r^2} = a^2\left(1 + \frac{3}{2}e^2\right), \tag{3.49}$$

$$\overline{\boldsymbol{rr}} = \frac{1}{2}a^2\left[5\boldsymbol{ee} - \boldsymbol{hh} + (1-e^2)\boldsymbol{U}\right]. \tag{3.50}$$

Substituting Equations 3.49 and 3.50 into Equation 3.48, and disregarding the irrelevant constant term, we have

$$\bar{\mathcal{R}}_p^* = \frac{3\mu_p}{4nd_p^3}\left[5(\hat{\boldsymbol{d}}_p \cdot \boldsymbol{e})^2 - (\hat{\boldsymbol{d}}_p \cdot \boldsymbol{h})^2 - 2e^2\right]. \tag{3.51}$$

From the secular Milankovitch equations, Equations 2.36 and 2.37, the singly-averaged third-body equations can be written in the form (cf. Allan 1962)

$$\dot{\bar{\boldsymbol{h}}}_p = \frac{3\mu_p}{2nd_p^3}\hat{\boldsymbol{d}}_p \cdot (5\boldsymbol{ee} - \boldsymbol{hh}) \cdot \widetilde{\hat{\boldsymbol{d}}}_p, \tag{3.52}$$

$$\dot{\tilde{\boldsymbol{e}}}_p = \frac{3\mu_p}{2nd_p^3}\left[\hat{\boldsymbol{d}}_p \cdot (5\boldsymbol{e}\boldsymbol{h} - \boldsymbol{h}\boldsymbol{e}) \cdot \tilde{\hat{\boldsymbol{d}}}_p - 2\tilde{\boldsymbol{h}} \cdot \boldsymbol{e}\right]. \tag{3.53}$$

3.3.1.2 Doubly-Averaged Equations

If the perturbing body is assumed to be in an elliptic orbit with semi-major axis a_p and eccentricity e_p, and sufficient distance between the two timescales exists, then another averaging may be performed. Averaging Equation 3.51 over the mean anomaly of the third body M_p gives

$$\bar{\bar{\mathcal{R}}}_p^* = -\frac{3\mu_p}{8na_p^3h_p^3}\left[5(\hat{\boldsymbol{H}}_p \cdot \boldsymbol{e})^2 - (\hat{\boldsymbol{H}}_p \cdot \boldsymbol{h})^2 - 2e^2\right], \tag{3.54}$$

where $h_p = \sqrt{1-e_p^2}$ and $\hat{\boldsymbol{H}}_p$ is the angular momentum unit vector of the perturbing body. Stated in this form, we can substitute $\bar{\bar{\mathcal{R}}}_p^*$ into Equations 2.36 and 2.37 to obtain the doubly-averaged equations of motion. However, since averaging is a linear process, it can also be performed over the secular Milankovitch equations directly to yield

$$\dot{\bar{\bar{\boldsymbol{h}}}}_p = \frac{1}{2\pi}\int_0^{2\pi} \dot{\boldsymbol{h}}_p \,\mathrm{d}M_p \tag{3.55}$$

$$= -\frac{3\mu_p}{2n}\left[5\boldsymbol{e} \cdot \overline{\left(\frac{\hat{\boldsymbol{d}}_p\hat{\boldsymbol{d}}_p}{d_p^3}\right)} \cdot \tilde{\boldsymbol{e}} - \boldsymbol{h} \cdot \overline{\left(\frac{\hat{\boldsymbol{d}}_p\hat{\boldsymbol{d}}_p}{d_p^3}\right)} \cdot \tilde{\boldsymbol{h}}\right], \tag{3.56}$$

$$\dot{\bar{\bar{\boldsymbol{e}}}}_p = \frac{1}{2\pi}\int_0^{2\pi} \dot{\boldsymbol{e}}_p \,\mathrm{d}M_p \tag{3.57}$$

$$= -\frac{3\mu_p}{2n}\left[5\boldsymbol{e} \cdot \overline{\left(\frac{\hat{\boldsymbol{d}}_p\hat{\boldsymbol{d}}_p}{d_p^3}\right)} \cdot \tilde{\boldsymbol{h}} - \boldsymbol{h} \cdot \overline{\left(\frac{\hat{\boldsymbol{d}}_p\hat{\boldsymbol{d}}_p}{d_p^3}\right)} \cdot \tilde{\boldsymbol{e}} + \overline{\left(\frac{1}{d_p^3}\right)} 2\tilde{\boldsymbol{h}} \cdot \boldsymbol{e}\right], \tag{3.58}$$

where $(\bar{\bar{\ }})$ denotes the double averaged value. The average of these quantities are given by Equations 3.29 and 3.30. Consequently, the doubly-averaged third-body dynamics for an elliptically orbiting disturbing body become (cf. Musen 1961)

$$\dot{\bar{\bar{\boldsymbol{h}}}}_p = -\frac{3\mu_p}{4na_p^3h_p^3}\hat{\boldsymbol{H}}_p \cdot (5\boldsymbol{e}\boldsymbol{e} - \boldsymbol{h}\boldsymbol{h}) \cdot \widetilde{\hat{\boldsymbol{H}}}_p, \tag{3.59}$$

$$\dot{\bar{\bar{\boldsymbol{e}}}}_p = -\frac{3\mu_p}{4na_p^3h_p^3}\left[\hat{\boldsymbol{H}}_p \cdot (5\boldsymbol{e}\boldsymbol{h} - \boldsymbol{h}\boldsymbol{e}) \cdot \widetilde{\hat{\boldsymbol{H}}}_p - 2\tilde{\boldsymbol{h}} \cdot \boldsymbol{e}\right]. \tag{3.60}$$

A noteworthy feature of these equations, as Musen (1961) has pointed out, is that the long-period effects depend only upon the orientation of the perturbing body's orbital plane and are not influenced by its periapsis.

Initially Circular Orbits : For a satellite on an initially circular orbit ($\boldsymbol{e} = \boldsymbol{0}$), the orbit remains circular throughout and the differential equation for the orbit pole $\hat{\boldsymbol{h}}$ can be written as

$$\dot{\hat{\boldsymbol{h}}} = -\frac{3\mu_p}{4na_p^3h_p^3}(\hat{\boldsymbol{H}}_p \cdot \hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{H}}}_p \cdot \hat{\boldsymbol{h}}. \tag{3.61}$$

Dotting with $\hat{\boldsymbol{H}}_p$ yields zero, implying that the product $\hat{\boldsymbol{H}}_p \cdot \hat{\boldsymbol{h}}$ is constant; hence, the satellite's inclination relative to the perturber's orbit pole $\hat{\boldsymbol{H}}_p$ is a fixed quantity. We therefore have precisely the same situation as in the oblateness perturbation case; it follows that the solution is of the form:

$$\boldsymbol{h} = \boldsymbol{C} \cdot \boldsymbol{h}_0, \tag{3.62}$$

$$\boldsymbol{C} = \cos\Omega\boldsymbol{U} + (1 - \cos\Omega)\hat{\boldsymbol{H}}_p\hat{\boldsymbol{H}}_p + \sin\Omega\widetilde{\hat{\boldsymbol{H}}}_p, \tag{3.63}$$

in which $\Omega(t) = \Omega_0 + \dot{\Omega}t$ and

$$\dot{\Omega} = -\frac{3\mu_p}{4na_p^3h_p^3}(\hat{\boldsymbol{H}}_p \cdot \hat{\boldsymbol{h}}). \tag{3.64}$$

Thus, for $\boldsymbol{e} = \boldsymbol{0}$, Equation 3.59 describes uniform precession of $\boldsymbol{h}$ around $\hat{\boldsymbol{H}}_p$, the rate of rotation being $\dot{\Omega}$. In fact, the same results holds true to first order in e (q.v., Tremaine & Yavetz 2013).

3.3.2 Lidov-Kozai Oscillations

The dynamical system given by Equations 3.59 and 3.60, known in astrophysical parlance as the secular hierarchical restricted three-body problem (Correia et al. 2011; Katz, Dong & Malhotra 2011; Naoz et al. 2013), is mathematically equivalent to the system originally treated by Lidov (1962) and Kozai (1962b), who discovered a phenomenon of the most far-reaching significance. Since the disturbing function is axisymmetric, the projection of the scaled angular momentum vector on the orbit axis of the perturbing body is conserved; namely, $\hat{\boldsymbol{H}}_p \cdot \boldsymbol{h}$ is an integral of the motion. The scaled doubly-averaged disturbing function $\overset{=*}{\mathcal{R}}_p$ is also conserved because it is autonomous, and the semi-major axis is a constant of motion under averaging. Since the disturbing function, Equation 3.54, depends on the satellite's orbital elements through only four variables (a, e, $\hat{\boldsymbol{H}}_p \cdot \boldsymbol{h}$, and $\hat{\boldsymbol{H}}_p \cdot \boldsymbol{e}$), and we have three conserved quantities, the orbit-averaged evolutions have only one

degree of freedom and are therefore integrable (Lidov 1962; Kozai 1962b; Tremaine & Yavetz 2013). Kozai (1962b) expressed the analytical solutions in terms of Weierstrass elliptic functions; however, the integrals of motion provide a qualitative picture of the phase space, and better disclose the characteristic features of the motion.

Since $\hat{\boldsymbol{H}}_p \cdot \boldsymbol{h}$ and a are constant, Equation 3.54 permits us to rewrite the disturbing function ("energy-like") integral as

$$5(\hat{\boldsymbol{H}}_p \cdot \boldsymbol{e})^2 - 2e^2 = \text{constant}, \quad \text{or} \tag{3.65}$$

$$e^2(2/5 - \sin^2 i \sin^2 \omega) = \text{constant}, \tag{3.66}$$

where the satellite's angular elements are referenced to the orbital plane of the perturbing body and to the direction toward the perturber's equinox (Vashkov'yak 1999). From $\hat{\boldsymbol{H}}_p \cdot \boldsymbol{h} = \text{constant}$, we can write the other classical integral as (Lidov 1962; Vashkov'yak 1999)

$$(1 - e^2)\cos^2 i = \text{constant}. \tag{3.67}$$

These two independent integrals enable one to determine the various phase portraits of the evolution and to place constraints on the possible motion. Of particular interest is when the satellite is initially on a circular orbit. Then the time-averaged third-body gravitational perturbation can induce a variation in its eccentricity, causing it to oscillate through a maximum value given by $e^2 \approx 1 - 5/3\cos^2 i_0$, depending on its initial inclination, provided that $|\cos i_0| \leq \sqrt{3/5}$ (Vashkov'yak 1999). Accordingly, from Equation 3.67, both the eccentricity and inclination execute periodic oscillations known as Lidov-Kozai cycles. These cycles have been implicated as the essential physical mechanism in the formation and evolution of a variety of astrophysical systems, from the distribution of irregular satellites, extrasolar planets, and binary stars to the galaxy mergers of supermassive black holes (Blaes, Lee & Socrates 2002; Fabrycky & Tremaine 2007; Tremaine & Yavetz 2013). Katz, Dong & Malhotra (2011) and Naoz et al. (2013) have recently found that qualitatively different behaviors can occur when the octupole moment of the Legendre expansion is accounted for in the secular equations. Orbits can reach extremely high eccentricities and undergo chaotic flips from prograde to retrograde orientation.

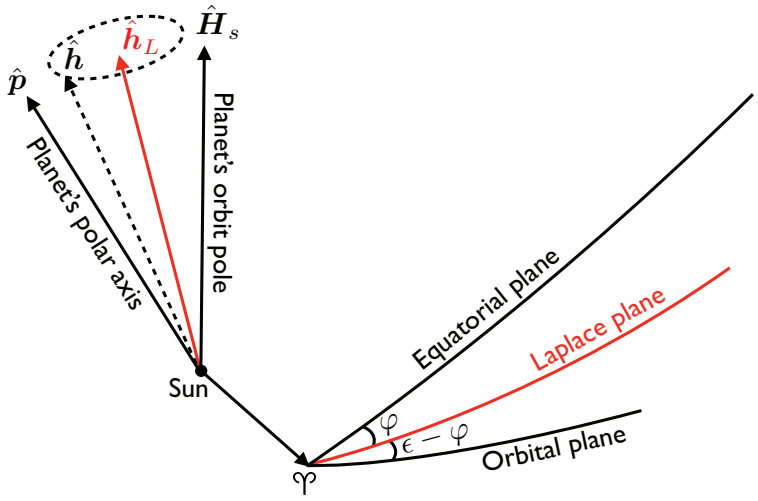
Chapter 4

Laplace Plane Modifications Arising from Solar Radiation Pressure

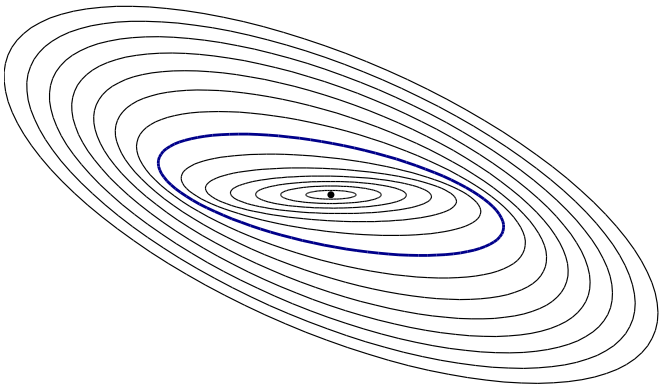
4.1 Background

Laplace in his theory of the Galilean satellites of Jupiter and the rings and satellites of Saturn showed that their orbits, under the perturbing gravitational effects of planetary oblateness and solar tides, maintain a nearly constant inclination relative to a local invariable plane—the Laplace plane (Laplace 1805). By itself, oblateness causes the pole of the satellite's orbital plane to precess around the planet's rotation pole. Solar third-body perturbations, if acting alone, will have a similar effect, but the precession will now take place about the pole of the planet's heliocentric orbital plane. Under both perturbations, a compromise is reached and the precession takes place around a mean pole that lies between the planet's rotational and orbital poles; this intermediate pole defines the classical Laplace plane (Tisserand 1896; Allan & Cook 1964; Burns 1986; Tremaine, Touma & Namouni 2009). For a satellite near its planet, this plane lies approximately in the planet's equatorial plane, while for distant satellites, it coincides with the planet's orbital plane; all three planes sharing a common node (Figure 4.1). Between these two degenerate states, the Laplace plane at a given semi-major axis lies at some intermediate orientation generating the warped Laplace surface (Ward 1981; Dobrovolskis 1993; Tremaine, Touma & Namouni 2009). A circular orbit in the Laplace plane experiences no secular precession and is therefore a frozen orbit, or equilibrium solution for the averaged system (Dobrovolskis 1993; Kudielka 1997).

The Laplace equilibrium plane provides a qualitative explanation for the overall shape and structure of planetary rings (Burns et al. 1979; Dobrovolskis, Borderies & Steiman-Cameron 1989),



(a) The Laplace equilibrium plane lies between the planet's equatorial and orbital planes, which differ by the planet's obliquity ϵ, and shares a common node with them (i.e., the equinox of the planet). The pole $\hat{\boldsymbol{h}}$ of any orbit at small inclination to the pole of the Laplace plane $\hat{\boldsymbol{h}}_L$ will regress about it at constant inclination and rate.



(b) The warped Laplace surface, shown as a collection of tilted rings, for a planet with modest obliquity. At the Laplace radius, the Laplace plane lies halfway between the planet's equatorial and orbital planes (thick blue ring).

Figure 4.1: Geometry of the Laplace equilibrium plane and warped Laplace surface. The symbols shown in the figure are described in more detail in the following sections.

and is of profound importance in understanding the origin of natural satellites (Goldreich 1965, 1966; Ward 1981). The Laplace surface defines the mean symmetry plane for particles on gravitationally dominated orbits. In the formation of planetary rings, mutual collisions will damp the motions out of the local Laplacian plane, which is essentially equatorial, and eventually all of the

ring particles will lie in a thin, flattened disk in that plane (Jeffreys 1947). The same dynamical processes also govern the evolution of protoplanetary disks from which the regular satellites were presumably formed. The Laplace surface is the steady-state shape of a dissipative, low-viscosity circumplanetary disk of gas and dust (Dobrovolskis, Borderies & Steiman-Cameron 1989; Tremaine, Touma & Namouni 2009); thus, satellites formed from such a disk (with the possible exception of Saturn's moon, Iapetus) generally orbit in or near the Laplace surface (Ward 1981). Understanding the physical mechanisms that can result in warping of an initially flat disk is a fundamental problem in the theory of astrophysical accretion disks, not just in the circumplanetary context, but also for protostellar disks (Armitage & Pringle 1997) and for disks around black holes and other compact objects (Petterson 1977; Pringle 1996; Tremaine & Davis 2013, and references therein). Tremaine & Davis (2013) have determined the analog of the Laplace surface in accretion disks around spinning black holes in X-ray binary systems, where the general-relativistic Lense-Thirring precession rather than the quadruple precession from oblateness is the dominant non-Keplerian perturbation from the central body.

Tremaine, Touma & Namouni (2009) defines the Laplace equilibria to be orbits in which the secular evolution due to the quadrupole potential of the gravitational perturbations is zero. Laplace equilibria exist for both circular and eccentric orbits in three planes: the classical, orthogonal, and polar Laplace planes (Kudielka 1997; Tremaine, Touma & Namouni 2009). The orbit poles of the circular Laplace equilibria lie along three orthogonal directions, two of them in the principal plane defined by the planet's rotation pole and heliocentric orbit pole, and the remaining in the direction orthogonal to this plane (Sekiguchi 1961). Their stability was investigated by Vashkov'yak (1974) and Kudielka (1997), and subsequently studied independently by Tremaine, Touma & Namouni (2009). Circular orbits in the classical Laplace plane are stable if the planetary obliquity is less than 68.875° or greater than 111.125°; within this range, however, orbits with semi-major axes near the Laplace radius (see Figure 4.1b), in which the two perturbations are of equal size, are unstable and undergo large and possibly chaotic oscillations in eccentricity and inclination (see Tremaine, Touma & Namouni 2009, for detailed stability maps). Circular orbits in the orthogonal Laplace

plane are always unstable, while those in the polar Laplace plane, which cross over the planet's pole, are stable at small planetocentric distances where the oblateness perturbation dominates and are unstable otherwise. Kudielka (1997) and Tremaine, Touma & Namouni (2009) found that a bifurcation into two equilibria of elliptic orbits occurs at the boundary lines of stability for circular orbits in the classical Laplace plane. Similarly, stable, eccentric polar orbits bifurcate from the circular polar Laplace equilibria at the semi-major axis where the latter become unstable. The eccentricity vectors of these stable elliptic orbits lie either in or orthogonal to the principal plane.

Allan & Cook (1967) showed that solar radiation pressure—which plays an important role in the dynamical behavior of dust grains in circumplanetary orbits—modifies the classical Laplace equilibrium plane. They found an approximate solution to the modified Laplace plane and noted that the orbital plane of a given dust particle, for a given semi-major axis, will regress around this plane. Tamayo et al. (2013), in considering circumplanetary dust particles in high-obliquity systems, found that solar radiation pressure shifts the location where the equilibrium plane transitions between the planet's orbital and equatorial planes, thereby shifting the point where dynamical instability sets in. These analyses are limited to particles that are only weakly perturbed by solar radiation pressure; that is, particles which have low values of effective area-to-mass ratio.

The purpose of this chapter is to rigorously show how solar radiation pressure modifies the classical Laplace plane, leading to a generalization of the treatments of Allan & Cook (1967) and Tamayo et al. (2013). We first give a concise exposition of the general results of Kudielka (1997) and Tremaine, Touma & Namouni (2009) regarding the stability of the classical equilibria. We then derive an analytical solution, valid for all particle sizes, for the modified equilibria, calculate the detailed warp of the modified Laplace surfaces, and determine their stability. These results are applied in Chapter 5 to the study of the dynamics and stability of GEO orbits and to the identification of robust, long-term disposal orbits for geostationary satellites. A further application of these results is given in Chapter 6 towards an understanding of the initial albedo dichotomy of Iapetus, Saturn's enigmatic satellite.

4.2 The Classical Laplace Plane and Equilibria

We shall consider first the dynamical effects of gravitational perturbations. The secular equations governing the motion of a satellite around an oblate planet perturbed by the Sun, in the quadrupolar approximation, are (see §3.2 and §3.3)

$$\begin{aligned}
\dot{\boldsymbol{h}} &= -\frac{\omega_2}{h^5}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\widetilde{\hat{\boldsymbol{p}}}\cdot\boldsymbol{h} - \omega_s\hat{\boldsymbol{H}}_s\cdot(5\boldsymbol{e}\boldsymbol{e}-\boldsymbol{h}\boldsymbol{h})\cdot\widetilde{\hat{\boldsymbol{H}}}_s, \\
\dot{\boldsymbol{e}} &= -\frac{\omega_2}{2h^5}\left\{\left[1-\frac{5}{h^2}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})^2\right]\widetilde{\boldsymbol{h}} + 2(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\widetilde{\hat{\boldsymbol{p}}}\right\}\cdot\boldsymbol{e} \\
&\quad -\omega_s\left[\hat{\boldsymbol{H}}_s\cdot(5\boldsymbol{e}\boldsymbol{h}-\boldsymbol{h}\boldsymbol{e})\cdot\widetilde{\hat{\boldsymbol{H}}}_s - 2\widetilde{\boldsymbol{h}}\cdot\boldsymbol{e}\right],
\end{aligned} \tag{4.1}$$

where $\omega_2 = 3nJ_2R^2/2a^2$ and $\omega_s = 3\mu_s/4na_s^3h_s^3$ are the factors of oblateness solar third-body perturbation, respectively. It is important to note the symmetries present in Equations 4.1: namely, these equations are invariant under the transformations (Tremaine, Touma & Namouni 2009)

$$\hat{\boldsymbol{p}}\to-\hat{\boldsymbol{p}} \quad \text{or} \quad \hat{\boldsymbol{H}}_s\to-\hat{\boldsymbol{H}}_s \quad \text{or} \quad \boldsymbol{e}\to-\boldsymbol{e} \quad \text{or} \quad (\boldsymbol{h}\to-\boldsymbol{h}, t\to-t). \tag{4.2}$$

Invariance under $\hat{\boldsymbol{p}}\to-\hat{\boldsymbol{p}}$ implies that the range of obliquity ϵ (i.e., the angle between the planet's equatorial and orbital planes) can be restricted from $(0,\pi)$ to $(0,\pi/2)$.

In the general case, for arbitrary values of the parameters of the problem, Equations 4.1 admit neither an exact analytical solution nor a complete qualitative description. Lidov & Yarskaya (1974) indicate the known integrable cases and investigate the geometrical behavior of the resulting motion. In the limiting case for which $\omega_s\to 0$, the evolution reduces to monotonic variations in Ω and ω (angles reckoned from from the planet's equatorial plane) according to Equations 3.36 and 3.41. For $\omega_2\to 0$, when the oblateness perturbation is absent, Equations 4.1 provide a vector formulation of the Lidov-Kozai oscillations (described in §3.3.2), in which the orbit eccentricity, inclination, and argument of periapsis (angles reckoned from the orbital plane of the perturbing body) undergo a coupled evolution according to Equations 3.65 and 3.67. For the special cases in which the planet's equatorial plane either coincides with or is orthogonal to the orbital plane of the perturbing body, phase portraits can be obtained through similar integral conditions (q.v., Lidov

& Yarskaya 1974, and references therein). It is particularly noteworthy that for an initially circular orbit ($\boldsymbol{e} = \boldsymbol{0}$), a complete general solution to Equations 4.1 exists (Allan & Cook 1964; Tremaine, Touma & Namouni 2009). The dynamics of circular orbits are governed by the three-dimensional system of equations

$$\dot{\hat{\boldsymbol{h}}} = -\omega_2(\hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{p}}} \cdot \hat{\boldsymbol{h}} - \omega_s(\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{H}}}_s \cdot \hat{\boldsymbol{h}}, \tag{4.3}$$

which admit, besides the trivial integral $|\hat{\boldsymbol{h}}| = 1$, an energy-like integral governed by the associated averaged disturbing function (i.e., through the addition of Equations 3.31 and 3.54). Equation 4.3 simplifies considerably if we recast the problem in the form of Euler's equations for the motion of a rigid body under no external torques.[1] The nature of the motion can be expressed analogously to that of an asymmetrical top, with an elliptical cylinder representing the energy surface standing in for the ellipsoid in Euler's formation, with the details given by Allan & Cook (1964) and Tremaine, Touma & Namouni (2009). We can therefore determine the possible trajectories of the orbit pole $\hat{\boldsymbol{h}}$, period of oscillation, equilibria and the global phase-space topology around them; however, we limit our discussion here to the equilibrium solutions, $\boldsymbol{e} = \boldsymbol{0}$ and $\boldsymbol{h} =$ constant, since these become especially important when we consider the joint effects of gravitational and radiation perturbations.

Tremaine, Touma & Namouni (2009) defines the Laplace equilibria to be frozen orbits of Equations 4.1, or orbits where the average angular momentum and eccentricity vectors remain constant. There are five types of equilibria for the system, each classified by the orientation of the vectors $\boldsymbol{h}$ and $\boldsymbol{e}$. The properties of these equilibria, especially their stability, were studied extensively by Kudielka (1997) and Tremaine, Touma & Namouni (2009), to which we refer for the omitted details. We will focus only on the circular Laplace equilibria, in which case $\boldsymbol{e} = \boldsymbol{0}$ and $\boldsymbol{h}$ lies either in the principal plane specified by the vectors $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{H}}_s$, or is orthogonal to this plane (qq.v., Sekiguchi 1961; Vashkov'yak 1974; Tremaine, Touma & Namouni 2009). There are six circular Laplace equilibria, of both prograde and retrograde orbits, lying in three distinct planes.

[1] See Schaub & Junkins (2009) for a discussion of the solutions to the rotational dynamics of a torque-free body.

4.2.1 Circular Laplace Equilibria

For an initially circular orbit, $\dot{\boldsymbol{e}}$ is identically zero and the condition for equilibrium becomes

$$\omega_2(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{p}}}\cdot\hat{\boldsymbol{h}}+\omega_s(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{H}}}_s\cdot\hat{\boldsymbol{h}}=\boldsymbol{0}. \tag{4.4}$$

In the coplanar case, $\hat{\boldsymbol{h}}$ lies in the principal plane and may be specified by an azimuthal angle φ, measured from the planet's spin axis $\hat{\boldsymbol{p}}$ (see Figure 4.1). Using oblique axes (contravariant coordinates), we may write

$$\hat{\boldsymbol{h}}=\frac{1}{\sin\epsilon}\left[\sin(\epsilon-\varphi)\hat{\boldsymbol{p}}+\sin\varphi\hat{\boldsymbol{H}}_s\right]. \tag{4.5}$$

Consequently, the frozen orbit condition becomes (cf. Allan & Cook 1964; Dobrovolskis 1993; Tremaine, Touma & Namouni 2009)

$$\begin{aligned}&\omega_2\sin 2\varphi+\omega_s\sin 2(\varphi-\epsilon)=0\quad\text{or}\\ &\tan 2\varphi=\frac{\sin 2\epsilon}{\cos 2\epsilon+(r_L/a)^5},\end{aligned} \tag{4.6}$$

where r_L is the Laplace radius defined by (Goldreich 1966; Dobrovolskis 1993)

$$r_L^5=a^5\frac{\omega_2}{\omega_s}=2J_2R^2a_s^3h_s^3\frac{\mu}{\mu_s}. \tag{4.7}$$

Equation 4.6 has four solutions for φ in a 2π interval, giving two planes of equilibrium (Kudielka 1997). The classical Laplace plane has the properties $\varphi\to 0$ (or π) as $a\to 0$ and $\varphi\to\epsilon$ (or $\epsilon+\pi$) as $a\to\infty$, so that the corresponding Laplace surface coincides with the planet's equator at small planetocentric distances and with its orbital plane at large distances (Ward 1981; Tremaine, Touma & Namouni 2009). The Laplace radius r_L is the critical distance at which the Laplace plane lies halfway between the equatorial and orbital planes: it is thus the planetocentric distance where the effects of oblateness and solar tides are equal.[2] We call the prograde and retrograde circular orbits lying in the classical Laplace plane the classical Laplace equilibria. The "orthogonal"

[2] Note that the Laplace radius defined by Equation 4.7 differs from that originally defined by Goldreich (1966) and subsequently quoted by Tremaine, Touma & Namouni (2009) by a factor of $2^{1/5}$. This factor, being of order unity, is usually omitted, but provides a better indicator of when the Laplace plane transitions between the two limiting cases (Dobrovolskis 1993; Tamayo et al. 2013).

Laplace equilibria lie in the orthogonal Laplace plane, which has the property $\pi/2 < \varphi < \pi$ or $3\pi/2 < \varphi < 2\pi$.

The remaining two circular Laplace equilibria lie in the plane orthogonal to the classical Laplace plane and to the line of nodes of the planetary equator on the planetary orbit, and can be specified as $\hat{\boldsymbol{h}} = \pm\widetilde{\hat{\boldsymbol{p}}}\cdot\hat{\boldsymbol{H}}_s/\sin\epsilon$ (Vashkov'yak 1974; Dobrovolskis 1980; Tremaine, Touma & Namouni 2009). As these orbits cross over the planet's pole, we call them the polar Laplace equilibria.

4.2.2 Stability

The characteristic polynomials and the structure of the eigenvalues for the classical and orthogonal Laplace equilibria are the same. The corresponding eigenvalue equations are given by (Tremaine, Touma & Namouni 2009)

$$\begin{aligned}
\lambda_h^2 &= -\omega_2^2\cos^2\varphi - \omega_s^2\cos^2(\epsilon-\varphi) - \frac{\omega_2\omega_s}{2}\left[\cos 2\varphi + \cos 2(\epsilon-\varphi) + 2\cos 2\epsilon\right], \\
\lambda_e^2 &= -\frac{\omega_2^2}{4}\left(5\cos^4\varphi - 2\cos^2\varphi + 1\right) - \frac{\omega_s^2}{2}\left[7\cos 2(\epsilon-\varphi) - 5\right] \\
&\quad -\frac{\omega_2\omega_s}{16}\left[2 + 3\cos 2\epsilon + 6\cos 2\varphi + 6\cos 2(\epsilon-\varphi) + 15\cos 2(\epsilon-2\varphi)\right].
\end{aligned} \tag{4.8}$$

From these equations and the equilibrium condition (4.6), the space of possible parameter values can be partitioned into stability and instability regions, as shown in Figures 4.2 and 4.3. The classical Laplace equilibria ($\varphi < \pi/2$) are always linear stable to variations in the angular momentum vector; however, instability in the eccentricity vector appears at obliquities between 68.875° and 111.125° and is restricted to semi-major axes near the Laplace radius ($0.8r_L \le a \le 1.02r_L$). The orthogonal Laplace equilibria are always unstable to spatial variations, however, there exists a small, fairly complex domain in which they are stable to changes in $\boldsymbol{e}$ (see also Kudielka 1997, Figure 2).

The eigenvalue equations for the polar Laplace equilibria have the form

$$\begin{aligned}
\lambda_h^2 &= -\omega_2\omega_s\sin^2\epsilon, \\
\lambda_e^2 &= \frac{1}{4}(6\omega_s + \omega_2)(4\omega_s - \omega_2).
\end{aligned} \tag{4.9}$$

Since $\lambda_h^2 < 0$, the polar Laplace equilibria are linearly stable to variations in $\boldsymbol{h}$. They are stable to

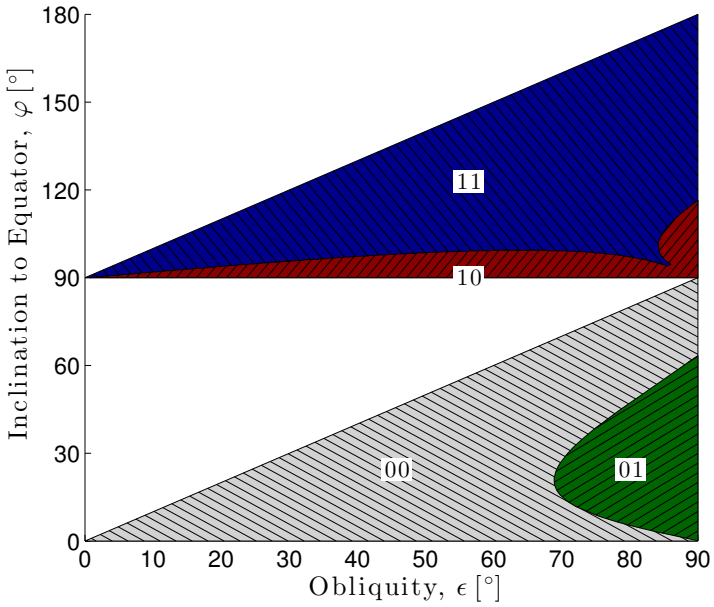


Figure 4.2: Stability and instability domains for the classical and orthogonal Laplace equilibria, adapted from Tremaine, Touma & Namouni (2009). The labels in each region indicate wether the equilibria are stable ("0") or unstable ("1") to changes in $\boldsymbol{h}$ and $\boldsymbol{e}$, respectively. Because of the symmetries in the secular equations of motion and equilibrium condition, the results are unchanged if $\varphi \to \pi + \varphi$ or $\epsilon \to \pi - \epsilon$ or $(\varphi, \epsilon) \to (\pi + \varphi, \pi - \epsilon)$.

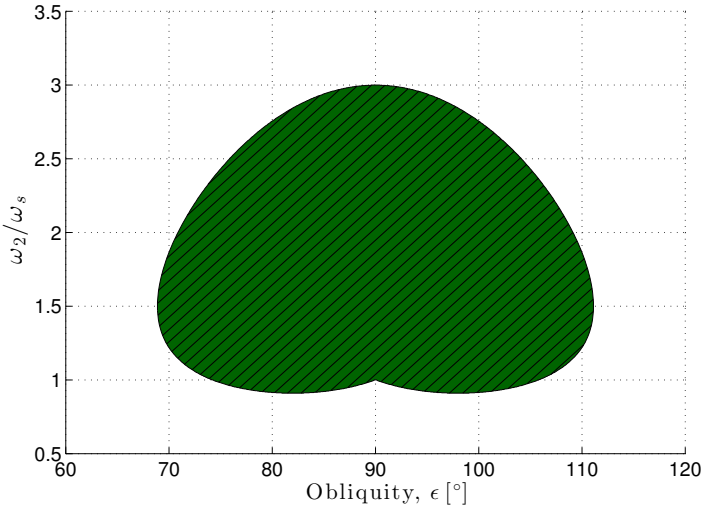


Figure 4.3: Instability region for the classical Laplace equilibria (Kudielka 1997). The vertical coordinate is the ratio of planetary oblateness and solar third-body perturbations (see Equation 4.7).

variations in $\boldsymbol{e}$ if and only if (Kudielka 1997; Tremaine, Touma & Namouni 2009)

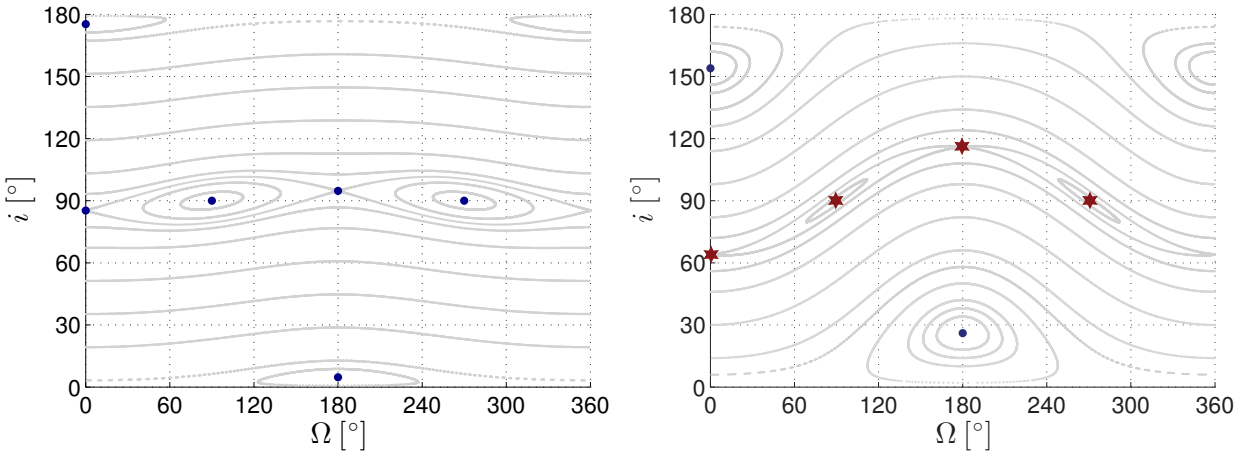
$$4\omega_s < \omega_2 \quad \text{or} \quad a < 4^{-1/5} r_L. \tag{4.10}$$

Note that the stability is independent of ϵ.

Figure 4.4 shows the trajectories of the system, Equations 4.1, in the inclination and ascending node phase space for initially circular orbits ($\boldsymbol{e} = \boldsymbol{0}$) of radius $a = 3r_L/4$ and $a = 2r_L$, and for $\epsilon = 26.7°$, corresponding to the Saturnian-Sun system with Saturn's oblateness suitably modified to include the attraction of the rings and inner moons. The six Laplace equilibria are indicated by blue points (stable) or red stars (unstable) according to their stability to variations in eccentricity vector. The trajectories through the orthogonal Laplace equilibria (saddle points) define separatrices, separating the domains of circulation and libration about the polar Laplace equilibria (Dobrovolskis, Borderies & Steiman-Cameron 1989). The prograde and retrograde Laplace equilibria are surrounded by closed precessional trajectories and are always stable for the Saturnian-Sun system. Kudielka (1997) provides similar (i, Ω)-phase space diagrams for a simplified Earth-Moon-Sun system and the Uranian-Sun system. Dobrovolskis, Borderies & Steiman-Cameron (1989) studied the Neptunian-Triton system and suggest that the polar Laplace equilibria could be sites for stable polar rings around Neptune (Dobrovolskis 1980). Kudielka (1994) and Ulivieri et al. (2013) point out the applicability of these frozen, high-altitude and highly-inclined, orbits for Earth satellites, as one with such an orientation would be nearly fixed in space, requiring minimal orbit maintenance.

4.3 The Modified Laplace Plane and Equilibria

We now explore the interplay between gravitational and non-gravitational perturbations and the equilibria that result when these forces act in concert. We shall focus on the simple, yet fundamental, cannonball model of solar radiation pressure (Equation 1.9), as it provides useful insight into the nature of the problem, and reveals the degree of complexity, which arises, in part, from the nonlinearity of the dynamical system.



(a) $a = 3r_L/4$. (b) $a = 2r_L$.

Figure 4.4: The circular Laplace equilibria and arbitrary trajectories of the orbit poles of circular orbits of radius $a = 3r_L/4$ and $a = 2r_L$ in the inclination and ascending node phase space in the Saturn-equatorial frame ($r_L = 55.58$ Saturn radii). The inclination i is measured with respect to Saturn's equatorial plane and the ascending node Ω is measured in the equatorial plane, referred to the vernal equinox of Saturn. Red stars indicate that the equilibria are unstable to changes in the eccentricity vector.

4.3.1 Modified Laplace Equilibria

Consider first the effects of solar radiation pressure acting alone. For solar radiation pressure perturbations, there are two fundamental timescales associated with the orbital motion of the perturbed body and the Sun. As sufficient distance between the two frequencies exists, we can average over the fast variables, independently with respect to each of them. The (singly) averaged equations governing the SRP perturbed orbiter problem can be written as

$$\begin{aligned} \dot{\boldsymbol{h}}_{srp} &= -\frac{H_s \tan\Lambda}{d_s^2} \tilde{\hat{\boldsymbol{d}}}_s \cdot \boldsymbol{e}, \\ \dot{\boldsymbol{e}}_{srp} &= -\frac{H_s \tan\Lambda}{d_s^2} \tilde{\hat{\boldsymbol{d}}}_s \cdot \boldsymbol{h}, \end{aligned} \tag{4.11}$$

where $H_s = \sqrt{\mu_s a_s (1 - e_s^2)}$ is the specific angular momentum of the planet about the Sun, and the SRP perturbation angle Λ is defined by Equation 1.10.

The secular equations of motion governing the gravitational perturbations, Equations 4.1, are autonomous if the slow variations of $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{H}}_s$ due to the torques on the planet's equatorial bulge

and orbit, respectively, are ignored. The relevant timespans involved for planetary spin and orbit precession are many orders of magnitude longer than what interests us here. The singly-averaged equations (4.11) governing the SRP perturbations, on the other hand, still have a time-varying term associated with the orbital motion of the Sun. This can be removed through another averaging, however, the application of the averaging method here is less obvious than it may appear at first sight.[3] The tacit assumption in carrying out these averaging results was that the orbital elements $\boldsymbol{h}$ and $\boldsymbol{e}$ do not change much over the characteristic timescales; that is, they are regarded as constants during the averaging process. In the case of solar radiation pressure perturbations, however, these elements change significantly over the heliocentric orbit period of the planet; in fact, as Allan & Cook (1967) point out, the integrated first-order effects are of order unity when radiation pressure is important. Accordingly, the second averaging must be carried out in a different manner; here, we exploit the known analytical solution to the SRP perturbed orbiter problem to determine the main secular dynamics involved.

When transformed into a frame rotating with the Sun, and taking the planet's heliocentric true anomaly as the independent variable, the averaged equations for $\boldsymbol{e}$ and $\boldsymbol{h}$, Equations 4.11, reduce to an autonomous linear system, yielding the explicit solution (qq.v., Richter & Keller 1995; Scheeres 2012a):

$$\begin{bmatrix} \boldsymbol{e}(\psi) \\ \boldsymbol{h}(\psi) \end{bmatrix} = \boldsymbol{\Phi}(\psi - \psi_o) \cdot \begin{bmatrix} \boldsymbol{e}_o \\ \boldsymbol{h}_o \end{bmatrix}, \tag{4.12}$$

$$\begin{aligned} \boldsymbol{\Phi}(\psi) = {} & \cos\psi \boldsymbol{U} \\ & + (1 - \cos\psi) \begin{bmatrix} \cos^2\Lambda \hat{\boldsymbol{H}}_s \hat{\boldsymbol{H}}_s + \sin^2\Lambda \hat{\boldsymbol{d}}_s \hat{\boldsymbol{d}}_s & \sin\Lambda\cos\Lambda \left(\hat{\boldsymbol{H}}_s \hat{\boldsymbol{d}}_s + \hat{\boldsymbol{d}}_s \hat{\boldsymbol{H}}_s \right) \\ \sin\Lambda\cos\Lambda \left(\hat{\boldsymbol{H}}_s \hat{\boldsymbol{d}}_s + \hat{\boldsymbol{d}}_s \hat{\boldsymbol{H}}_s \right) & \cos^2\Lambda \hat{\boldsymbol{H}}_s \hat{\boldsymbol{H}}_s + \sin^2\Lambda \hat{\boldsymbol{d}}_s \hat{\boldsymbol{d}}_s \end{bmatrix} \\ & - \sin\psi \begin{bmatrix} \cos\Lambda \widetilde{\hat{\boldsymbol{H}}_s} & \sin\Lambda \widetilde{\hat{\boldsymbol{d}}_s} \\ \sin\Lambda \widetilde{\hat{\boldsymbol{d}}_s} & \cos\Lambda \widetilde{\hat{\boldsymbol{H}}_s} \end{bmatrix}, \end{aligned} \tag{4.13}$$

where $\psi = f/\cos\Lambda$ and f is the planet's true anomaly (as demonstrated in §3.1.2 for the more

[3] Indeed, if a second averaging is performed following the usual intuitive process, we would find that the SRP perturbations vanish—a very misleading result (q.v., Chapter 5).

general model of solar radiation pressure). Notice that $\boldsymbol{\Phi}(\psi + 2k\pi) = \boldsymbol{\Phi}(\psi)$ for all $k \in \boldsymbol{Z}$, where $\boldsymbol{Z}$ is the set of integers. This means that the state transition matrix $\boldsymbol{\Phi}$, and hence the solutions, are 2π-periodic in ψ, or periodic in f with period $2\pi \cos\Lambda$.

Equation 4.13 is an orthogonal matrix (i.e., $\boldsymbol{\Phi}^{-1} = \boldsymbol{\Phi}^T$), and defines a rotation ($\det \boldsymbol{\Phi} = 1$) in $\boldsymbol{R}^6$; its eigenvalues are therefore of modulus 1. The eigenvalues of $\boldsymbol{\Phi}$ are 1 and $\cos\psi \pm i \sin\psi$, each of multiplicity 2. Its eigenvectors arranged in column form are

$$\begin{bmatrix} \cos\Lambda\hat{\boldsymbol{H}}_s & \sin\Lambda\hat{\boldsymbol{d}}_s & i\sin\Lambda\hat{\boldsymbol{H}}_s & -i\cos\Lambda\hat{\boldsymbol{d}}_s + \hat{\boldsymbol{d}}_{s_\perp} & -i\sin\Lambda\hat{\boldsymbol{H}}_s & i\cos\Lambda\hat{\boldsymbol{d}}_s + \hat{\boldsymbol{d}}_{s_\perp} \\ \sin\Lambda\hat{\boldsymbol{d}}_s & \cos\Lambda\hat{\boldsymbol{H}}_s & -i\cos\Lambda\hat{\boldsymbol{d}}_s + \hat{\boldsymbol{d}}_{s_\perp} & i\sin\Lambda\hat{\boldsymbol{H}}_s & i\cos\Lambda\hat{\boldsymbol{d}}_s + \hat{\boldsymbol{d}}_{s_\perp} & -i\sin\Lambda\hat{\boldsymbol{H}}_s \end{bmatrix}, \tag{4.14}$$

where $\hat{\boldsymbol{d}}_{s_\perp} = \widetilde{\hat{\boldsymbol{H}}}_s \cdot \hat{\boldsymbol{d}}_s$.

The dynamics of $\boldsymbol{e}$ and $\boldsymbol{h}$ are characterized by the motion in the center manifolds and pure rotations about two different directions: $\hat{\boldsymbol{d}}_s$ and $\hat{\boldsymbol{H}}_s$. In inertial space, $\hat{\boldsymbol{d}}_s$ and $\hat{\boldsymbol{d}}_{s_\perp}$ both rotate with the planet-Sun line and their average over one solar year will produce secular effects of nearly third order in the perturbing force (on account of the small eccentricities of the planetary orbits); consequently, these effects can be safely ignored in this analysis. The orbit pole $\hat{\boldsymbol{H}}_s$, on the other hand, will remain nominally constant. Thus, the averaged eigenvectors only retain the net rotation of the angular momentum and eccentricity vectors about the $\hat{\boldsymbol{H}}_s$ direction. Accordingly, the secular equations for the solar radiation pressure perturbation become

$$\begin{aligned} \dot{\boldsymbol{h}}_{srp} &= -\omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s \cdot \boldsymbol{h}, \\ \dot{\boldsymbol{e}}_{srp} &= -\omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s \cdot \boldsymbol{e}, \end{aligned} \tag{4.15}$$

where the exact rate of rotation ω_{srp} can be determined by mapping the periodic solution into the inertial frame and is given by

$$\omega_{srp} = \frac{2\pi(1 - \cos\Lambda)}{T_s \cos\Lambda}, \tag{4.16}$$

in which T_s is the planet's orbital period. It should be emphasized that Equations 4.15 represent the doubly-averaged (i.e., long-term) orbital behavior of spherical objects perturbed by SRP. Thus, although the second of Equations 4.15 implies that initially circular orbits will remain circular

throughout, these orbits undergo periodic changes in their eccentricities over one solar year with the amplitude of these oscillations increasing with increasing Λ (q.v., §5.3). This subtlety will be of significance in Chapter 6.

Including the secular effect of SRP, Equations 4.15, into Equations 4.1, we can find the condition for the modified Laplace equilibrium:

$$\omega_2(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{p}}}\cdot\hat{\boldsymbol{h}}+\omega_s(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{H}}}_s\cdot\hat{\boldsymbol{h}}+\omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s\cdot\hat{\boldsymbol{h}}=\boldsymbol{0}\quad\text{or}\tag{4.17}$$

$$\omega_2\sin 2\varphi+\omega_s\sin 2(\varphi-\epsilon)+2\omega_{srp}\sin(\varphi-\epsilon)=0.\tag{4.18}$$

We define the modified Laplace surfaces to be the locus of all orbits that are modified Laplace equilibria. The resulting equilibrium planes turn out to be qualitatively different from those obtained without considering the radiation pressure. In particular, Equation 4.18 does not necessarily have four solutions for φ in a 2π interval, as in the classical case, but the number of roots depends on the relative strength of the perturbations. More importantly, solar radiation pressure destroys the symmetries present in the classical and orthogonal equilibria; namely, the prograde and retrograde equilibria will no longer lie in the same plane (the equilibrium condition is no longer invariant under the transformation $\boldsymbol{h}\to-\boldsymbol{h}$), and the polar Laplace equilibria vanish completely (i.e., the condition $\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}}=\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{h}}=0$ no longer satisfies the equilibrium equation).

4.3.2 Stability

To study the stability of the equilibria, we form the linearized equations of motion and compute the characteristic polynomial. Substituting $\boldsymbol{h}=\boldsymbol{h}^\star+\delta\boldsymbol{h}$ and $\boldsymbol{e}=\delta\boldsymbol{e}$ into Equations 4.1 and 4.15, and expanding to first order in $\delta\boldsymbol{h}$ and $\delta\boldsymbol{e}$ gives

$$\begin{aligned}
\delta\dot{\boldsymbol{h}}&=\left\{\omega_2\left[\widetilde{\hat{\boldsymbol{h}}^\star}\cdot\hat{\boldsymbol{p}}\hat{\boldsymbol{p}}-(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}}^\star)\widetilde{\hat{\boldsymbol{p}}}\right]+\omega_s\left[\widetilde{\hat{\boldsymbol{h}}^\star}\cdot\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s-(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{h}}^\star)\widetilde{\hat{\boldsymbol{H}}}_s\right]-\omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s\right\}\cdot\delta\boldsymbol{h},\\
\delta\dot{\boldsymbol{e}}&=\left\{-\frac{\omega_2}{2}\left[\left(1-5(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}}^\star)^2\right)\widetilde{\hat{\boldsymbol{h}}^\star}+2(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}}^\star)\widetilde{\hat{\boldsymbol{p}}}\right]\right.\\
&\quad\left.-\omega_s\left[5\widetilde{\hat{\boldsymbol{h}}^\star}\cdot\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s+(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{h}}^\star)\widetilde{\hat{\boldsymbol{H}}}_s-2\widetilde{\hat{\boldsymbol{h}}^\star}\right]-\omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s\right\}\cdot\delta\boldsymbol{e}.
\end{aligned}\tag{4.19}$$

As in the classical case (q.v., Tremaine, Touma & Namouni 2009), the two equations are decoupled: the linearized evolution of $\boldsymbol{h}$ is independent of the linearized evolution of $\boldsymbol{e}$. Substituting Equation 4.5 into Equations 4.19 and evaluating the eigenvalue equations gives

$$\begin{aligned}
\lambda_h^2 = &-\omega_2^2 \cos^2\varphi - \omega_s^2 \cos^2(\epsilon-\varphi) - \frac{\omega_2\omega_s}{2}\left[\cos 2\varphi + \cos 2(\epsilon-\varphi) + 2\cos 2\epsilon\right] \\
&-\omega_{srp}^2 - \frac{\omega_2\omega_{srp}}{2}\left[\cos(\epsilon-\varphi) + 3\cos(\epsilon+\varphi)\right] - 2\omega_s\omega_{srp}\cos(\epsilon-\varphi), \\
\lambda_e^2 = &-\frac{\omega_2^2}{4}\left(5\cos^4\varphi - 2\cos^2\varphi + 1\right) - \frac{\omega_s^2}{2}\left[7\cos 2(\epsilon-\varphi) - 5\right] \\
&-\frac{\omega_2\omega_s}{16}\left[2 + 3\cos 2\epsilon + 6\cos 2\varphi + 6\cos 2(\epsilon-\varphi) + 15\cos 2(\epsilon-2\varphi)\right] \\
&-\omega_{srp}^2 + \frac{\omega_2\omega_{srp}}{4}\left[2\cos(\epsilon-\varphi) + \cos(\epsilon+\varphi) + 5\cos(\epsilon-3\varphi)\right] + 2\omega_s\omega_{srp}\cos(\epsilon-\varphi).
\end{aligned} \tag{4.20}$$

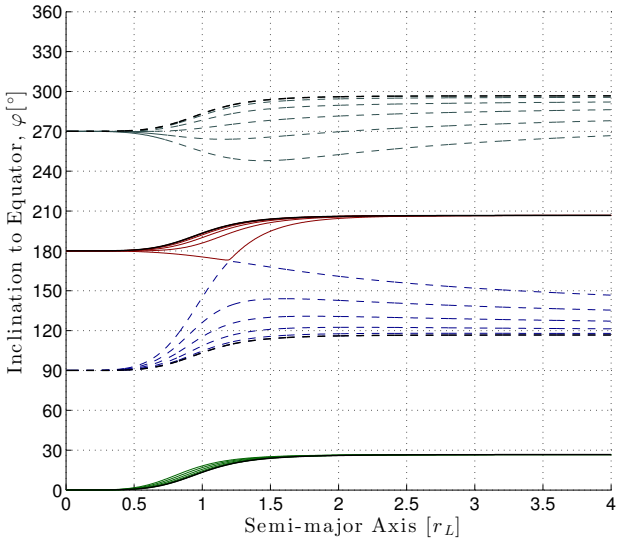
The modified Laplace equilibria are linearly stable if both $\lambda_h^2 < 0$ and $\lambda_e^2 < 0$.

The solutions to Equation 4.18 for spherical circum-Saturnian dust particles with effective area-to-mass ratios from 0 up to 150 m^2/kg are shown in Figure 4.5, which has been broken into three pieces for the sake of clarity. Solid lines denote stable equilibria, while dashed lines denote equilibria that have at least one unstable mode. Note that if the SRP perturbation strength vanishes ($A/m \to 0$), the modified surfaces coincide with the classical and orthogonal Laplace surfaces (solutions to Equation 4.6), shown by the black lines (Figure 4.5a). The warped surface shown by the solid green lines, which is composed of stable prograde orbits, is qualitatively similar to the classical case. Solar radiation pressure simply rotates the classical Laplace plane, increasing its inclination relative to Saturn's equator with increasing area-to-mass ratio (or decreasing particle size, as $s = 0.75/\rho_s/(A/m)$); each particle has its own modified Laplace plane for a given semi-major axis. Consequently, solar radiation pressure moves the warp inward with increasing A/m and shifts the location where the equilibrium plane transitions between the two degenerate states (equatorial and orbital). This behavior was first noted by Allan & Cook (1967), during their investigation of the zodiacal light, but the validity of their arguments is limited to particles with area-to-mass ratios less than 20 m^2/kg (see also Tamayo et al. 2013).

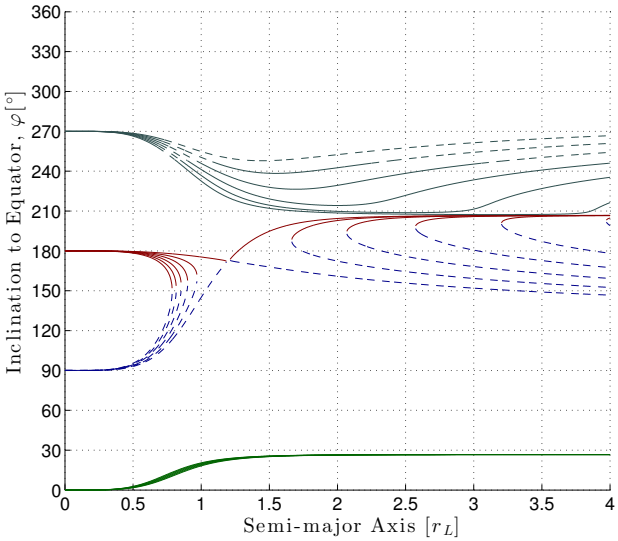
Only in the prograde solution to Equation 4.18 ($\varphi < \pi/2$) does the azimuthal angle of the

equilibrium orbit pole increase smoothly and continuously. The other modified Laplace surfaces in Figure 4.5 exhibit drastically different behavior as the area-to-mass ratio of the particle increases. In contrast to the classical case, the surfaces shown by the solid red lines, consisting of stable retrograde orbits, divides into inner and outer parts if $(1+\rho)A/m > 31.2$ m^2/kg; and is mirrored by an unstable solution, shown as dashed blue lines. The surfaces emanating from $\varphi = 3\pi/2$ become significantly distorted with increasing area-to-mass ratio and approach $\varphi = \epsilon + \pi$ for very small particles (i.e., the particles angular momentum vector becomes anti-aligned with that of Saturn's orbit). These surfaces are stable at small Saturno-centric distances for nearly all particles and becomes stable at larger distances for smaller dust grains. The solutions to Equation 4.18 for $(1+\rho)A/m > 44.2$ m^2/kg (Figure 4.5c) look remarkably similar to the warped surfaces for inviscid accretion disks orbiting spinning black holes in binary systems (Tremaine & Davis 2013); although, there are fundamental differences in the physical mechanisms and interpretation.

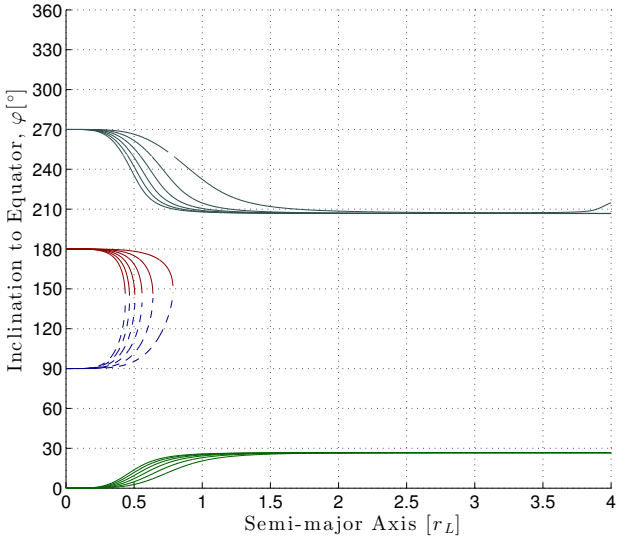
Figures 4.6-4.8 show inclination and ascending node phase space diagrams for initially circular orbits of radius $a = 3r_L/4$ and $a = 2r_L$ for a range of dust particles in the Saturnian-Sun system, reported in the Saturn-equatorial frame described in Figure 4.4. The corresponding SRP perturbation angles Λ (Equation 1.10) are listed for reference. Note that the structure of the orbit pole evolution shows distinctly different behavior than Figure 4.4 and across the various particle sizes. In particular, the symmetries that existed in the classical case are completely destroyed, as each of the modified Laplace equilibria at a given semi-major axis lie on distinct planes. Moreover, one of the orthogonal Laplace equilibria, which corresponded to a saddle point for gravitationally dominated orbits, is now surrounded by closed precessional trajectories and is stable for large particles.



(a) $0 \leq (1+\rho)A/m \leq 31.2$ m^2/kg.

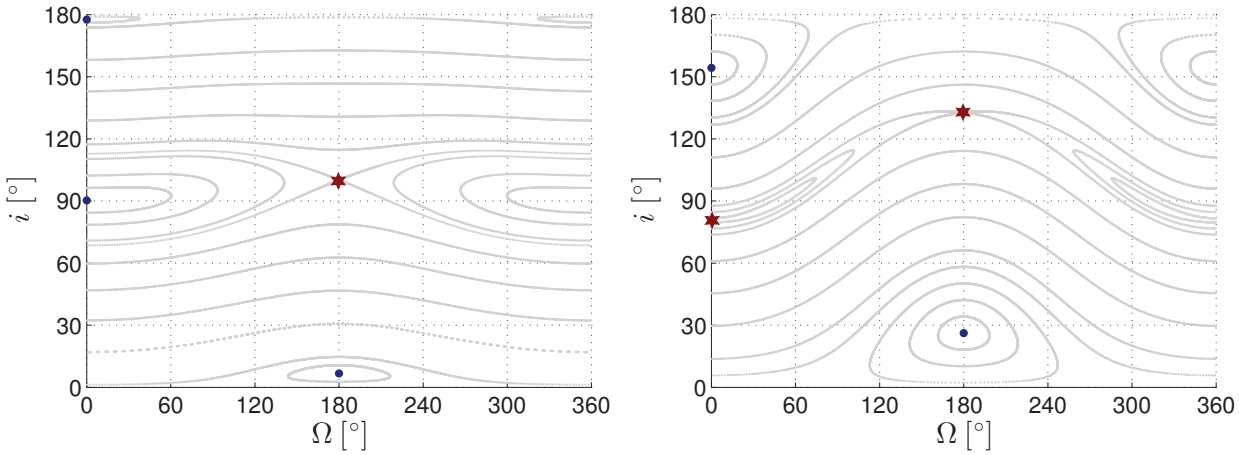


(b) $31.2 < (1+\rho)A/m < 44.2$ m^2/kg.



(c) $44.2 \leq (1+\rho)A/m \leq 150$ m^2/kg.

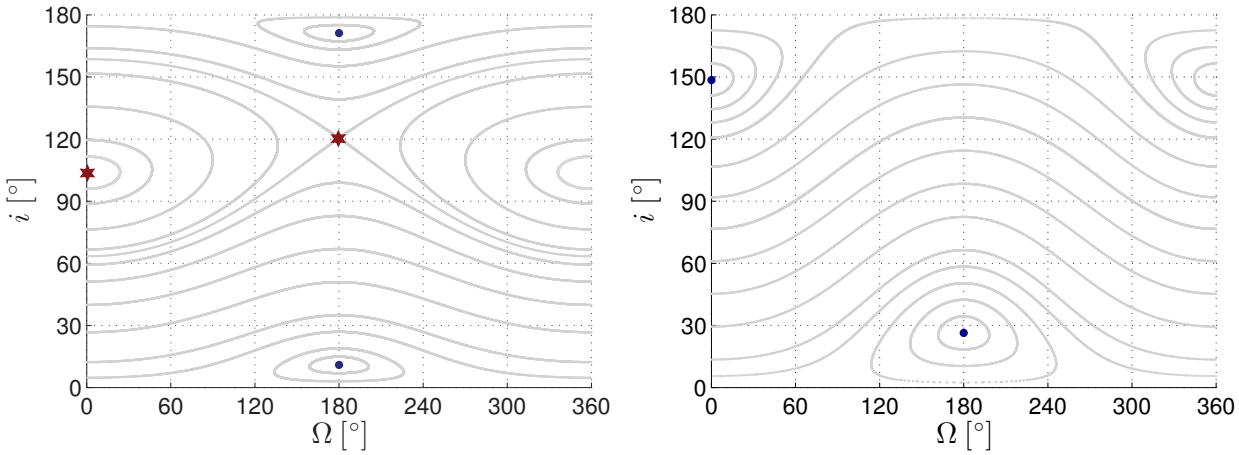
Figure 4.5: Modified Laplace equilibria and their stability for the Saturnian-Sun system. The vertical axis is the azimuthal angle φ in the principal plane, measured from Saturn's spin axis $\hat{\boldsymbol{p}}$, and the horizontal axis is the semi-major axis in Laplace radii (Equation 4.7). Solutions are shown for eighteen values of effective area-to-mass ratio, split into three evenly spaced intervals. Unstable equilibria are shown by dashed lines.



(a) $a = 3r_L/4$.

(b) $a = 2r_L$.

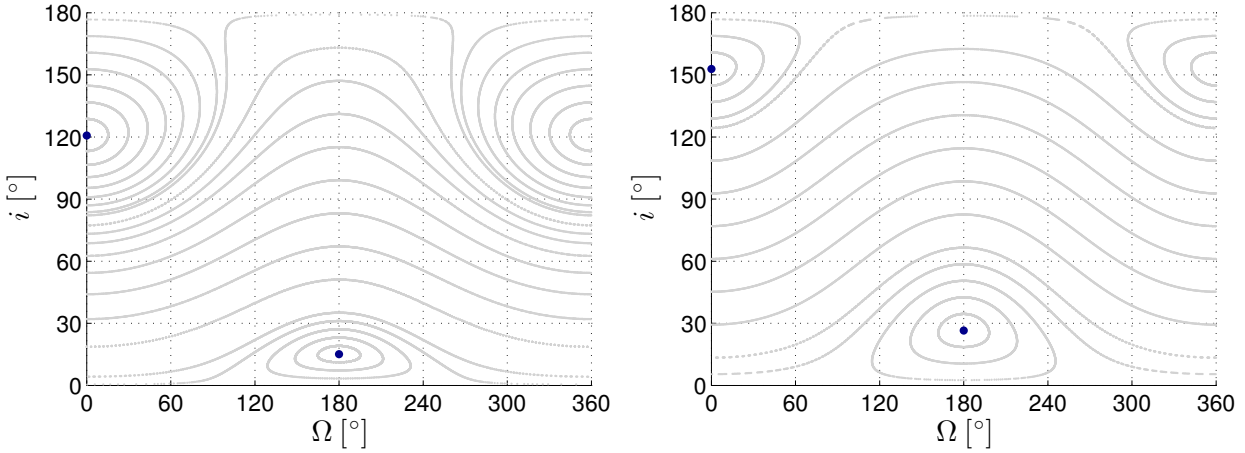
Figure 4.6: Modified Laplace equilibria and i-Ω diagrams for a circum-Saturnian particle with $(1+\rho)A/m = 20$ m^2/kg for $a = 3r_L/4$, $\Lambda = 3.22°$ (left) and $a = 2r_L$, $\Lambda = 5.25°$ (right). Red stars indicate that the equilibria are unstable to changes in the eccentricity vector.



(a) $a = 3r_L/4$.

(b) $a = 2r_L$.

Figure 4.7: Modified Laplace equilibria and i-Ω diagrams for a circum-Saturnian particle with $(1+\rho)A/m = 40$ m^2/kg for $a = 3r_L/4$, $\Lambda = 6.42°$ (left) and $a = 2r_L$, $\Lambda = 10.41°$ (right). Red stars indicate that the equilibria are unstable to changes in the eccentricity vector.



(a) $a = 3r_L/4$.

(b) $a = 2r_L$.

Figure 4.8: Modified Laplace equilibria and i-Ω diagrams for a circum-Saturnian particle with $(1+\rho)A/m = 60$ m^2/kg for $a = 3r_L/4$, $\Lambda = 9.58°$ (left) and $a = 2r_L$, $\Lambda = 15.40°$ (right).

Chapter 5

High Area-to-Mass Ratio Space Debris

Nearly four centuries of experience since Galileo's time has shown that it is frequently useful to depart from the real and to construct a "model" of the system being studied; some of the complications are stripped away, so a simple and generalized mathematical structure can be built up out of what is left. Once that is done, the complicating factors can be restored one by one, and the relationship suitably modified. To try to achieve the complexities of reality at one bound, without working through a simplified model first, is so difficult that it is virtually never attempted and, we can feel certain, would not succeed if it were attempted.[1]

— Isaac Asimov, 1966

5.1 A New Class of Geosynchronous Debris

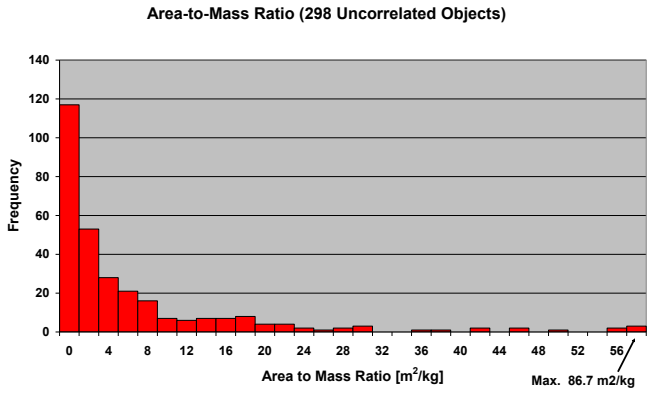
Nearly a half century has elapsed since satellites were first launched into the geostationary (equatorial, circular-synchronous) orbit—the altitude of 35,786 km where satellites appear to remain fixed over a single point on the equator throughout the day, providing a unique vantage point for communication, meteorology, science, and military applications (Zee 1989; Johnson 2012). The geostationary ring is the least forgiving region to space debris because there is no natural cleansing mechanism to limit the lifetimes of the debris at this altitude. Only objects in low-altitude orbits will return to Earth without human intervention through the influence of atmospheric drag, which steadily reduces their orbital energy until they re-enter or burn up within the atmosphere. In some preferential low Earth orbit (LEO) regions, where the population is above a critical spatial density, random collisions are predicted to produce new debris at a rate that is greater than the removal

[1] Asimov I. 1966. *Understanding Physics.* New York: New American Library, p. 13

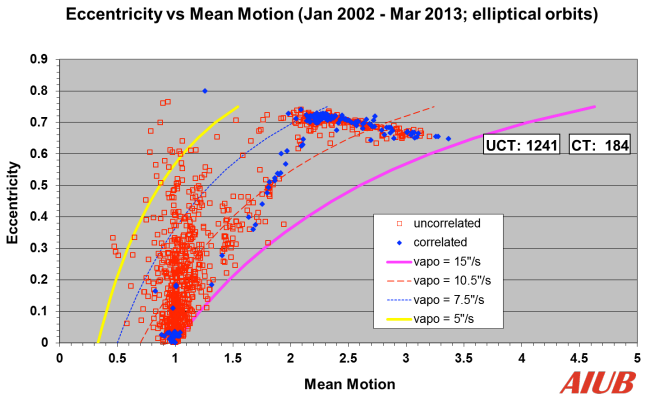
rate due to orbital decay (i.e., the Kessler syndrome; see, for instance, Kessler & Cour-Palais 1978; Liou & Johnson 2006; Rossi & Valsecchi 2006). In GEO, the relative velocities are much lower (less than 1 km/s), meaning that the the damage done by impact and the amount of detritus generated in a collision is not as severe. However, because debris would contaminate this unique and valuable resource practically forever, placing satellites in super-synchronous disposal orbits at the end of their operational lifetimes has been recommended and practiced as one possible means of protecting this orbital environment (IADC 2002; ITU 2010). These satellites are also passivated to reduce the probability of future explosions by removing any on-board stored energy, such as residual fuel or pressurants and charged batteries.

The high area-to-mass ratio (HAMR) debris population in GEO space (Figures 5.1), discovered through optical observations by Schildknecht and colleagues (ca. 2004), demonstrates that energetic breakups and collisions are not the only source of concern (Schildknecht 2007). This hitherto unknown class of body in high-Earth orbits—having area-to-mass ratios hundreds or thousands of times greater than that of a typical satellite and thus strongly perturbed by solar radiation pressure—has been linked to aging satellites in the storage orbits (Liou & Weaver 2005). Indeed, it is a plausible hypothesis, and generally believed within the debris community, that HAMR objects are sheets of multilayer insulation detaching from payloads or buses of objects in the GEO disposal orbit due to material deterioration and surface degradation in the space environment (Liou & Weaver 2005). Though such objects can be generated in a variety of other ways; i.e., YORP spin-up, collisions, and explosions, to name just a few. The orbits of HAMR debris are difficult to track and predict due to the strength of SRP and our lack of knowledge of the body characteristics: object geometry, attitude behavior, surface properties, thermal characteristics, etc. (Kelecy & Jah 2011). A better understanding of their nature, orbital evolution, and possible origin is critical.

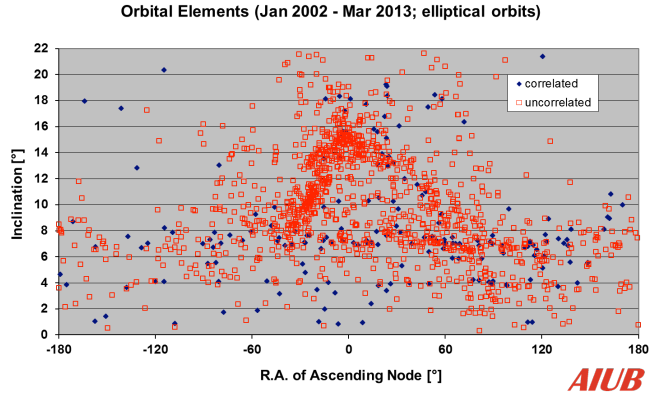
The motion of high area-to-mass ratio objects in high-Earth orbits has been studied extensively since the discovery of this debris. The orbits of HAMR objects are highly perturbed from the combined effect of solar radiation pressure, anomalies of the Earth gravitational field, and third-body gravitational interactions induced by the Sun and the Moon (Chao 2006; Valk, Lemaître &



(a) Distribution of the area-to-mass ratio of 298 uncorrelated objects in the AIUB/ESA catalog.



(b) Eccentricity as a function of mean motion (rev/day) for all objects listed in the AIUB/ESA catalog. The HAMR population is the vertically dispersed cloud concentrated at 1 rev/day.



(c) Inclination and ascending node of the cataloged objects.

Figure 5.1: Observational data from 2002 to 2013 of 298 HAMR objects, for which 6-parameter orbits were determined, from the internal catalog maintained at the Astronomical Institute of the University of Bern (Credit: T. Schildknecht, personal communication; updated figures from Schildknecht 2007). The uncorrelated objects are those that could not be identified with known objects listed in the US Strategic Command catalog; see Schildknecht (2007) for more details.

Anselmo 2008; Anselmo & Pardini 2010; Früh, Kelecy & Jah 2013). The evolution of individual orbits of HAMR debris over a considerable time interval, taking into account both short-period and long-period terms, can be calculated by numerical integration of the precise set of differential equations. Numerical computations of this type have revealed that the amplitudes of the short-term fluctuations in the osculating orbital elements are, in general, quite small when compared with the long-term variations. For this reason, and the necessity of considering the orbital behavior of a large population of HAMR debris over many thousands of revolutions, it seems reasonable to investigate the equations that govern the long-term evolution of orbits. Such equations, as discussed in §2.3, can be derived by the theory of secular perturbation.

Liou & Weaver (2005) used PROP3D, a fast orbit propagator based on the averaging principle that was developed for NASA's debris evolutionary models, to investigate the HAMR debris problem. PROP3D accounts for the perturbations from Earth gravity up to the fourth zonal harmonic, low-order lunisolar gravitational interactions, and SRP with consideration of the Earth shadow effect. Through comparisons with a high-fidelity orbit integrator, they showed that HAMR objects in GEO are dominated by major perturbations (e.g., those discussed in §1.3), not those of higher order. Chao (2006) performed long-term studies of the orbital evolution of GEO objects with high area-to-mass ratios through analytically averaged equations. Chao investigated the secular effects of SRP on the eccentricity and argument of perigee by averaging over the object's orbital period. The long-term motion of the orbit inclination and longitude of the ascending node were studied though doubly-averaged equations for the lunisolar attraction, ignoring higher-order terms in eccentricity, and singly-averaged equations for the SRP perturbation. The secular equations of motion were written in terms of the classical orbital elements, which leads to the presence of small numerical divisors—the eccentricity and the sine of the inclination.

We present here a complete non-singular model of first-order averaging, explicitly given in terms of the Milankovitch elements, which accounts for solar radiation pressure, the oblateness of the Earth's gravitational figure, and lunisolar perturbations. This secular model combines the first-order averaged equations, derived in Chapter 3, for each perturbing force, and naturally accounts

for the complexities of the Earth-Moon-Sun system. We first review in §5.2 the environment and force models for each perturbation, and discuss any underlying assumptions and approximations. In §5.3, we demonstrate the validity of our averaged equations by comparing numerical integrations of them to integrations of the exact Newtonian equations, along with comparisons with explicit high-fidelity simulations of HAMR objects in GEO (e.g., Anselmo & Pardini 2010). The extent to which the qualitative properties of the orbit persist with increasing area-to-mass is investigated. We then apply the non-singular theory of first-order averaging toward an understanding of the general nature and spatial distribution of the HAMR debris population, and discuss the implications of these results for future observing campaigns. In §5.4, using our understanding of the classical and modified Laplace planes, we propose new and more robust disposal orbits for end-of-life retirement of geostationary satellites, which take advantage of the dynamics and stability of GEO orbits.

5.2 Environment and Force Models

5.2.1 Earth-Moon-Sun System

Any account of motion in the Earth-Moon-Sun systems has to start with a description of the dynamical configuration of this three-body problem. Perozzi et al. (1991), using eclipse records, the JPL ephemeris, and results from numerical integration of the three-body problem, showed that the mean geometry of the Earth-Moon-Sun system repeats itself closely after a period of time equal in length to the classical eclipse prediction cycle known as the Saros.[2] That is, this dynamical system is moving in a nearly periodic orbit. Saros means repetition, and indicates a period of 223 synodic months ($\sim$ 6585.3213 days), after which the Sun has returned to the same place it occupied with respect to the nodes of the Moon's orbit when the cycle began. While the motion of the Earth around the Sun can, over timespans of interest, be considered Keplerian, the Moon is incessantly subject to solar perturbations resulting in periodic and secular variations of its orbital elements (Roy 2005). The semi-major axis, eccentricity, and inclination are subject to periodic variations

[2] The Saros has been the basis for which predicting eclipses rests since the very dawn of Chaldean history; after the lapse of the Saros period of 6585 days, solar and lunar eclipses recur under almost identical circumstances except that they are displaced about 120° westward on the Earth (Moulton 1914, p. 131).

about their mean values of 384 400 km, 0.0549, and $5°09'$, respectively. The line of apses advances, making one revolution in about 8.85 years. The node of the Moon's orbital plane (its intersection with the ecliptic) regresses in the ecliptic plane with a sidereal period of 6798.3 days (about 18.61 years); the node moving westward on the ecliptic at a rate of roughly $1°$ in 18.9 days. The regression of the lunar nodes is intrinsically related with the Saros cycle. This complexity of lunar motion must be taken into account for long-term studies of HAMR object dynamics.

To describe the motion of the Earth about the Sun, we define a heliocentric orbit frame, $(\hat{\boldsymbol{E}}_e, \hat{\boldsymbol{E}}_{e_\perp}, \hat{\boldsymbol{H}}_e)$, in which $\hat{\boldsymbol{E}}_e$ is the unit vector pointing to the orbit perihelion, $\hat{\boldsymbol{E}}_{e_\perp}$ is the unit vector in the heliocentric plane of motion and normal to $\hat{\boldsymbol{E}}_e$, and the cross product of these two vectors defines the orbit normal, specified as $\hat{\boldsymbol{H}}_e$, about which the Earth revolves (q.v., Appendix B.2.3). With this formulation, the varying position vector between the Earth and the Sun is specified as $\boldsymbol{d}_e = d_e \hat{\boldsymbol{d}}_e$ and split into a magnitude d_e and a direction $\hat{\boldsymbol{d}}_e$, both of which are functions of the Earth true anomaly f_e

$$d_e = \frac{a_e(1-e_e^2)}{1+e_e\cos f_e}, \tag{5.1}$$

$$\hat{\boldsymbol{d}}_e = \cos f_e \hat{\boldsymbol{E}}_e + \sin f_e \hat{\boldsymbol{E}}_{e_\perp}, \tag{5.2}$$

where a_e and e_e are the semi-major axis and eccentricity of the Earth's orbit, respectively.

The Moon's actual motion is very complex: high-precision lunar ephemerides are available through the Jet Propulsion Laboratory, which account for the relativistic n-body equations of motion for the point-mass Sun, Moon, planets, and major asteroids, perturbations on the orbit of the Earth-Moon barycenter from the interaction of the point-mass Sun with the figure, solid-body tides of both the Earth and Moon, and observations of lunar laser ranging (Folkner, Williams & Boggs 2009). This precise description of the Moon's motion, while useful for the construction of accurate ephemerides, conceals the analytical character that is essential for a clear understanding of the processes involved. To that end, we assume that the Moon is on an osculating elliptical orbit in which the lunar node precesses clockwise in the ecliptic plane with a period of about 18.61 years. Note that we neglect the rotation of the lunar perigee as well as the periodic variations

in the Moon's semi-major axis, eccentricity, and inclination. We define a geocentric orbit frame, $(\hat{\boldsymbol{E}}_m, \hat{\boldsymbol{E}}_{m_\perp}, \hat{\boldsymbol{H}}_m)$, where $\hat{\boldsymbol{E}}_m$ is the unit vector pointing to the orbit perigee, $\hat{\boldsymbol{E}}_{m_\perp} = \hat{\boldsymbol{H}}_m \times \hat{\boldsymbol{E}}_m$, and $\hat{\boldsymbol{H}}_m$ is the Moon's angular momentum unit vector, about which the Moon revolves. These vectors are resolved using the Moon's mean ecliptic orbital elements in which $\Omega_m(t) = \Omega_{m_0} + \dot{\Omega}_m(t - t_0)$, where $\dot{\Omega}_m = -2\pi/P_{\text{saros}}$ and P_{saros} is the sidereal period of nodal regression in seconds. The position vector from the Earth to the Moon is then specified as $\boldsymbol{d}_m = d_m \hat{\boldsymbol{d}}_m$, where d_m and $\hat{\boldsymbol{d}}_m$ are given by Equations 5.1 and 5.2, respectively, using the Moon's orbit parameters.

5.2.2 Newtonian Equations of Motion

Numerical integration of the precise non-averaged equations of motion represents the most accurate means of calculating the exact trajectory of an orbiting body in a given time interval. Anselmo and Pardini have made several numerical investigations of the HAMR debris problem, mapping out the dynamics of these objects over long timespans with all relevant perturbations included. Their most recent work presents a detailed analysis concerning the long-term evolution of HAMR debris in high-Earth orbit subject to SRP with Earth's shadow effects, geopotential harmonics up to degree and order eight, and third-body gravitational interactions induced by the Sun and the Moon (Anselmo & Pardini 2010). Comparison of these solutions with the results obtained from the averaged formulas is a significantly reliable estimate of the accuracy of the approximated equations (Lidov 1962). Such comparison permits us to conclude about the applicability of the averaged equations for considering the evolution of HAMR debris orbits.

In order to have a firm basis for comparison and validation, we perform direct numerical simulations of the long-term dynamics of HAMR objects in GEO using a simplified dynamical model, which includes solar radiation pressure in the cannonball approximation (Equation 1.6), the Earth's gravity potential limited at degree and order 2, and the gravitational tidal action of the point-mass Sun and Moon (governed by Equation 1.17). The longitude-dependent spherical harmonics of the gravitational potential of the Earth, in particular the (2,2) term, leads to resonance effects for geosynchronous orbits, acting as long-term perturbations of the semi-major axis (see,

for instance, Lemaître, Delsate & Valk 2009). For this reason, we consider the effects of both the C_{20} and C_{22} terms of the harmonic expansion of Earth's gravitational potential in our Newtonian equations, which account for the polar and equatorial flattening of the Earth's figure. Inclusion of these predominant harmonics is sufficient to capture the main effects of nonsphericity in the Earth's mass distribution at high-altitude orbits. The force model for planetary oblateness was already discussed in §1.3.2. We review this model here and combine it with the tesseral term.

The customary way to represent the disturbing function of the second degree and order gravity field perturbation is using a body-fixed frame with latitude angle δ measured from the equatorial plane and longitude angle λ measured in the equator from the axis of minimum moment of inertia

$$\mathcal{R}_2 = -\frac{\mu C_{20}}{2r^3}\left(1 - 3\sin^2\delta\right) + \frac{3\mu C_{22}}{r^3}\cos^2\delta\cos 2\lambda, \tag{5.3}$$

where $C_{20} = -J_2 R^2$ is the dimensional oblateness gravity field coefficient, R is the mean equatorial radius of the Earth, and C_{22} is the dimensional ellipticity gravity field coefficient (Scheeres 2012b). We can state the disturbing function in a general vector expression

$$\mathcal{R}_2 = -\frac{\mu C_{20}}{2r^3}\left[1 - 3(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{p}})^2\right] + \frac{3\mu C_{22}}{r^3}\left[(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{s}})^2 - (\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{q}})^2\right], \tag{5.4}$$

where we assume that the unit vectors $\hat{\boldsymbol{p}}$, $\hat{\boldsymbol{q}}$, and $\hat{\boldsymbol{s}}$ are aligned with the Earth's maximum, intermediate, and minimum axes of inertia. The perturbing acceleration is then

$$\begin{aligned}\boldsymbol{a}_2 = {} & \frac{3\mu C_{20}}{2r^4}\left\{\left[1 - 5(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{p}})^2\right]\hat{\boldsymbol{r}} + 2(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{p}})\hat{\boldsymbol{p}}\right\} \\ & - \frac{3\mu C_{22}}{r^4}\left\{5\left[(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{s}})^2 - (\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{q}})^2\right]\hat{\boldsymbol{r}} - 2\left[(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{s}})\hat{\boldsymbol{s}} - (\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{q}})\hat{\boldsymbol{q}}\right]\right\}.\end{aligned} \tag{5.5}$$

A few words must also be said about solar radiation pressure. The fundamental approximation adopted when using the cannonball model is that the non-radial component of the radiation pressure is negligible or will average out over time periods that are small compared to the orbital period. Although the cannonball model captures the general nature of SRP, it does not provide a precise prediction of how an individual object will evolve. However, this simple model is commonly used in the propagation of HAMR debris orbits since there is no method to incorporate

a physically realistic SRP model with a lack of a priori information (i.e., object shape and size, rotational motion, surface and thermal characteristics, etc.). Even though the cannonball model may not realistically represent the SRP force acting on these objects, the fact that this model gives rise to many complex dynamical behaviors necessitates a deeper understanding of this basic model before exploring these more complex models. Thus, in this analysis, we will mainly focus on the cannonball model as that will allow a direct comparison with earlier analyses of HAMR debris orbit dynamics (cf. Liou & Weaver 2005; Chao 2006; Valk, Lemaître & Anselmo 2008; Anselmo & Pardini 2010).

Combining the above force models with those expounded on in §1.3.2, we can define the equations of motion for an object in orbit about the oblate Earth disturbed by SRP and lunisolar gravity. In an inertially fixed frame centered at the Earth, they can be stated in relative form

$$\ddot{\boldsymbol{r}} = \frac{\partial U}{\partial \boldsymbol{r}}, \tag{5.6}$$

$$U(\boldsymbol{r}) = \frac{\mu}{r} + \mathcal{R}_{srp}(\boldsymbol{r}) + \mathcal{R}_2(\boldsymbol{r}) + \mathcal{R}_s(\boldsymbol{r}) + \mathcal{R}_m(\boldsymbol{r}), \tag{5.7}$$

where $\mathcal{R}_s$ and $\mathcal{R}_m$ are tidal disturbing functions for the Sun and the Moon, respectively. Performing the partial derivatives, we can state the problem in terms of the perturbation accelerations:

$$\ddot{\boldsymbol{r}} = -\frac{\mu}{r^3}\boldsymbol{r} + \boldsymbol{a}_{srp} + \boldsymbol{a}_2 + \boldsymbol{a}_s + \boldsymbol{a}_m. \tag{5.8}$$

5.2.3 Secular Equations of Motion

The secular evolution of the Milankovitch orbital elements in the presence of SRP, J_2 and lunisolar perturbations can be stated as

$$\dot{\boldsymbol{h}} = \dot{\boldsymbol{h}}_{srp} + \dot{\boldsymbol{h}}_2 + \dot{\boldsymbol{h}}_s + \dot{\boldsymbol{h}}_m, \tag{5.9}$$

$$\dot{\boldsymbol{e}} = \dot{\boldsymbol{e}}_{srp} + \dot{\boldsymbol{e}}_2 + \dot{\boldsymbol{e}}_s + \dot{\boldsymbol{e}}_m, \tag{5.10}$$

where the first-order averaged models for these perturbations, developed in the last two chapters, are shown in Tables 5.1 and 5.2 for convenience. We will conduct explicit long-term integrations of the single-averaged equations (Table 5.1) and compare the predicted evolution with the results

obtained from the Newtonian non-averaged formulation. The doubly-averaged, autonomous, equations (Table 5.2) will be used mainly for descriptive purposes and to gain an accurate dynamical understanding of the perturbed motion.

Combining all of these perturbations leads to a highly-nonlinear system, which does not appear integrable, but such simplifications help to understand some qualitative features of the system. Although the exact averaged solution is presumably inaccessible, the expressions given in Equations 5.9 and 5.10 are several hundred times faster to numerically integrate than their non-averaged Newtonian counterparts. With these results it is possible to predict accurately the long term orbital behavior of HAMR objects, given the initial values of the orbital elements and the initial geometry of the Earth-Moon-Sun system; the latter being important for calculation of Moon-induced perturbations.

Table 5.1: Singly-averaged equations of motion governing solar radiation pressure, planetary oblateness, and third-body gravitational perturbations.

	SRP	Oblateness	Third-Body
$\dot{\boldsymbol{h}}$	$-\frac{H_s \tan\Lambda}{d_s^2}\tilde{\hat{\boldsymbol{d}}}_s \cdot \boldsymbol{e}$	$-\frac{3nJ_2R^2}{2a^2h^5}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\tilde{\hat{\boldsymbol{p}}}\cdot\boldsymbol{h}$	$\frac{3\mu_p}{2nd_p^3}\hat{\boldsymbol{d}}_p\cdot(5\boldsymbol{ee}-\boldsymbol{hh})\cdot\tilde{\hat{\boldsymbol{d}}}_p$
$\dot{\boldsymbol{e}}$	$-\frac{H_s \tan\Lambda}{d_s^2}\tilde{\hat{\boldsymbol{d}}}_s \cdot \boldsymbol{h}$	$-\frac{3nJ_2R^2}{4a^2h^5}\left\{\left[1-\frac{5}{h^2}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})^2\right]\tilde{\boldsymbol{h}}+2(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\tilde{\hat{\boldsymbol{p}}}\right\}\cdot\boldsymbol{e}$	$\frac{3\mu_p}{2nd_p^3}\left[\hat{\boldsymbol{d}}_p\cdot(5\boldsymbol{eh}-\boldsymbol{he})\cdot\tilde{\hat{\boldsymbol{d}}}_p-2\tilde{\boldsymbol{h}}\cdot\boldsymbol{e}\right]$

Table 5.2: Doubly-averaged equations of motion governing solar radiation pressure, planetary oblateness, and third-body gravitational perturbations. The factors are the perturbation strength frequencies defined in Chapter 4: $\omega_{srp} = 2\pi(1-\cos\Lambda)/T_s\cos\Lambda$, $\omega_2 = 3nJ_2R^2/2a^2$, and $\omega_p = 3\mu_p/4na_p^3h_p^3$. These nonlinear equations are autonomous if the slow variations of $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{H}}_p$ due to the torques on the planet's equatorial bulge and orbit, respectively, are ignored.

	SRP	Oblateness	Third-Body
$\dot{\boldsymbol{h}}$	$-\omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s\cdot\boldsymbol{h}$	$-\frac{\omega_2}{h^5}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\tilde{\hat{\boldsymbol{p}}}\cdot\boldsymbol{h}$	$-\omega_p\hat{\boldsymbol{H}}_p\cdot(5\boldsymbol{ee}-\boldsymbol{hh})\cdot\widetilde{\hat{\boldsymbol{H}}}_p$
$\dot{\boldsymbol{e}}$	$-\omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s\cdot\boldsymbol{e}$	$-\frac{\omega_2}{2h^5}\left\{\left[1-\frac{5}{h^2}(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})^2\right]\tilde{\boldsymbol{h}}+2(\hat{\boldsymbol{p}}\cdot\boldsymbol{h})\tilde{\hat{\boldsymbol{p}}}\right\}\cdot\boldsymbol{e}$	$-\omega_p\left[\hat{\boldsymbol{H}}_p\cdot(5\boldsymbol{eh}-\boldsymbol{he})\cdot\widetilde{\hat{\boldsymbol{H}}}_p-2\tilde{\boldsymbol{h}}\cdot\boldsymbol{e}\right]$

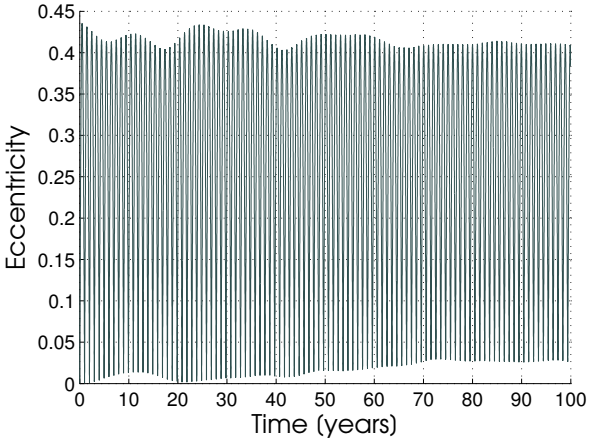
5.3 Long-Term Orbit Evolution

We restrict our attention to objects released in geostationary orbit ($a \sim 42164.2$ km) on 1950 January 01 12:00:00 UTC; therefore, $\boldsymbol{e}_0 \equiv \mathbf{0}$ and $\boldsymbol{h}_0 \equiv \hat{\boldsymbol{p}}$ (Earth's rotation pole). To understand the main characteristics of motion and to determine the extent to which the qualitative properties of the orbit persist with increasing area-to-mass, the dynamics were simulated using a reflectance value of 0.36, and area-to-mass ratios between 0 and 40 m^2/kg. For a given semi-major axis, reflectivity, and A/m value, we compute the corresponding Λ angle (Equation 1.10), which we use to characterize the evolutionary behavior of HAMR debris orbits. The SRP perturbation angles and corresponding effective area-to-mass ratios are shown in Table 5.3.

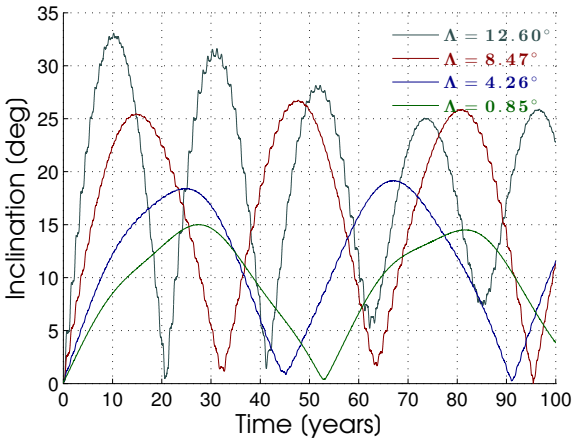
5.3.1 Newtonian Non-Averaged Dynamics

For the non-averaged dynamics, we use the high-accuracy Solar System ephemeris (DE421), provided by JPL, to calculate the position vectors of the Sun and the Moon (Folkner, Williams & Boggs 2009). The long-term orbit evolution of several HAMR objects, obtained using numerical integrations of the Newtonian equations of motion (Equation 5.8), are shown in Figure 5.2. The eccentricity and inclination evolution, over 100 years, shown in Figures 5.2a and 5.2b, closely match the results obtained by Anselmo & Pardini (2010).[3] For the object with $\Lambda = 12.60^\circ$ (corresponding to $(1+\rho)A/m = 20.4$ m^2/kg), the eccentricity undergoes an approximately yearly oscillation with amplitude of about 0.4 and long-period modulations of ~ 0.05. The inclination undergoes approximately yearly oscillations that are superimposed on the long-term drift, which has a varying maximum amplitude between 25° and 35° and a long-term oscillation period of about 22 years. As shown in Figure 5.2c, the evolution of the two-dimensional eccentricity vector, $e\,[\cos\omega\cos\Omega - \cos i\sin\omega\sin\Omega; \cos\omega\sin\Omega + \cos i\sin\omega\cos\Omega]$, is characterized by both a yearly and long-term regression; the latter exhibiting complex evolutionary behavior. The orbit pole, $\hat{\boldsymbol{h}}$, precesses clockwise, having the same characteristics as the inclination oscillation (see Figure 5.2d).

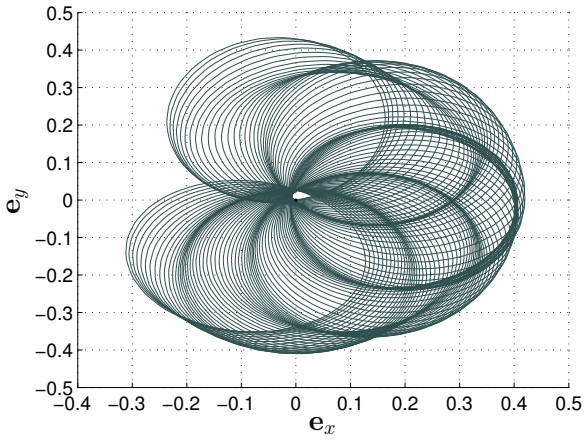
[3] Note that we used slightly different initial conditions and a different release epoch for our simulations; the latter will change the initial dynamical configuration of the Earth-Moon-Sun system.



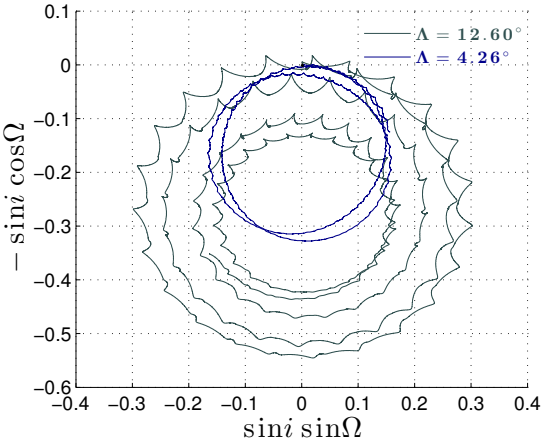
(a) Eccentricity evolution ($\Lambda = 12.60°$).



(b) Inclination evolution.



(c) Two-dimensional eccentricity vector evolution ($\Lambda = 12.60°$).



(d) Two-dimensional angular momentum unit vector evolution.

Figure 5.2: Long-term orbit evolution (100 years) in the Earth-equatorial frame for different values of the SRP perturbation angle, as predicted by the full non-averaged equations of motion, Equation 5.8. The position vectors of both the Sun and the Moon were computed using the JPL ephemeris (DE421).

5.3.2 Singly-Averaged Dynamics

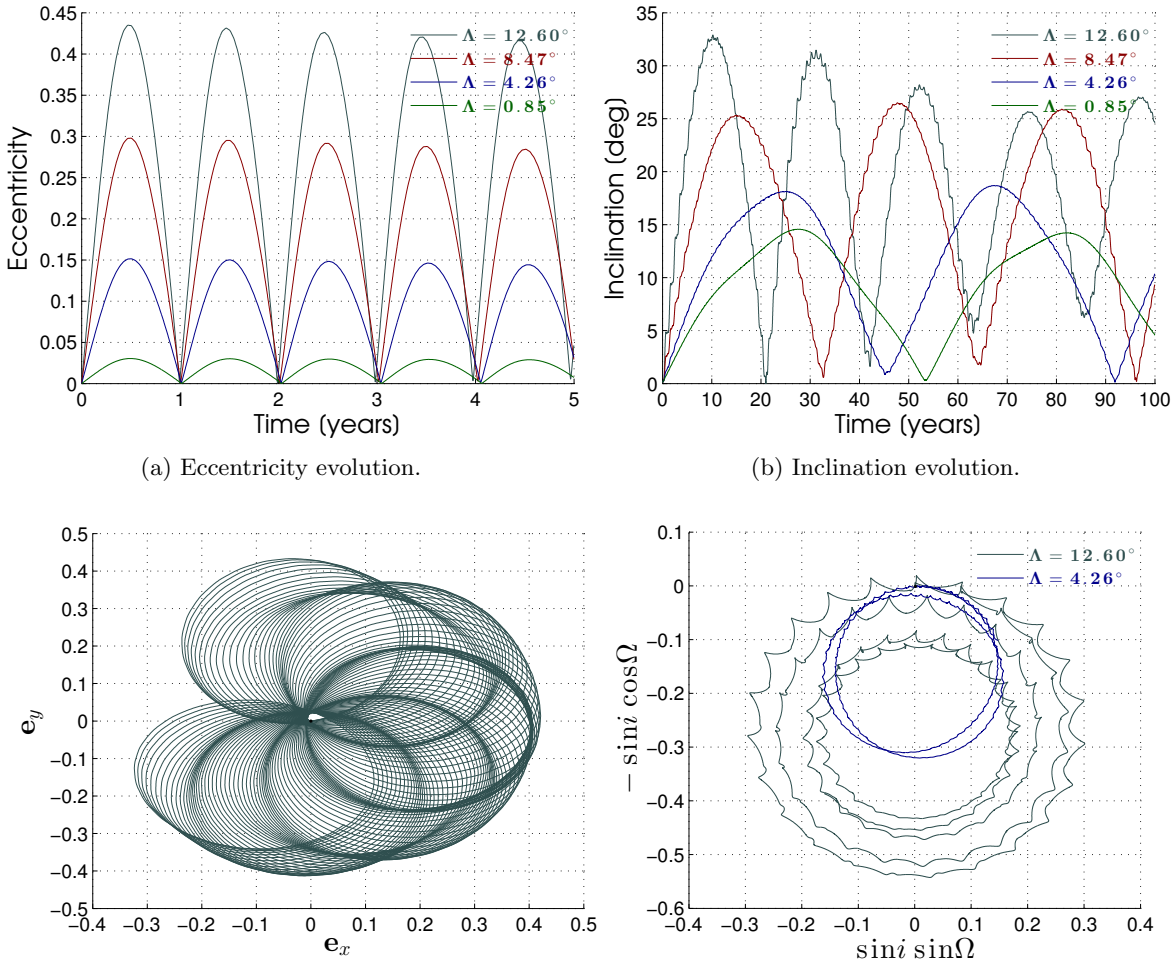
Since our Newtonian non-averaged results compare well, both quantitatively and qualitatively, with those of Anselmo & Pardini (2010),[4] they can be used as a logical basis for assessing the validity of our averaged model. We are particularly interested in distinguishing between cause and effect and in identifying the precise origin of any perturbation experienced by the HAMR object. To that end, we avoid using the precise JPL ephemeris, and instead assume two-body dynamics for the Sun and the Moon, for which the lunar node regresses in the ecliptic plane with a sidereal period of ~ 18.61 years (see §5.2.1).

The evolution of several HAMR objects, obtained using numerical integrations of the singly-averaged equations of motion, are shown in Figure 5.3. The eccentricity evolution in Figure 5.3a is shown over a shorter timescale to emphasize the respective amplitudes of the yearly oscillations for each object. The approximated averaged equations, using two-body dynamics and accounting for the regression of the Moon's node, gives us nearly identical plots at this level of resolution as the full Newtonian non-averaged simulations. We do not use any special formalism in our integrations to preserve the constraints on $\boldsymbol{h}$ and $\boldsymbol{e}$,[5] yet after 100 years, they are satisfied to over one part per billion.

With the Newtonian non-averaged formulation, the various perturbations are all lumped together and we obtain no indication as to the form and nature of any of them. However, our averaged formulation gives significant qualitative indications and allows us to understand many aspect of HAMR debris motion. Concerning the eccentricity evolution, solar radiation pressure acting alone induces sub-yearly oscillations with period $2\pi \cos \Lambda$; the amplitude increasing with increasing Λ. Inclusion of the J_2 dynamics causes only slight changes in the short-term oscillations, but induces long-period small fluctuations in the maximum amplitudes. The dynamical coupling between SRP and oblateness becomes more pronounced with increasing Λ. The addition of third-body perturbations, primarily the attraction of the Moon, causes a slight increase in the short-term amplitudes,

[4] In addition to our force model, they account for higher-order gravity field perturbations and Earth shadow effects in their numerical integrations.

[5] Recall that $\boldsymbol{e} \cdot \boldsymbol{h} = 0$ and $\boldsymbol{e} \cdot \boldsymbol{e} + \boldsymbol{h} \cdot \boldsymbol{h} = 1$.



(a) Eccentricity evolution.

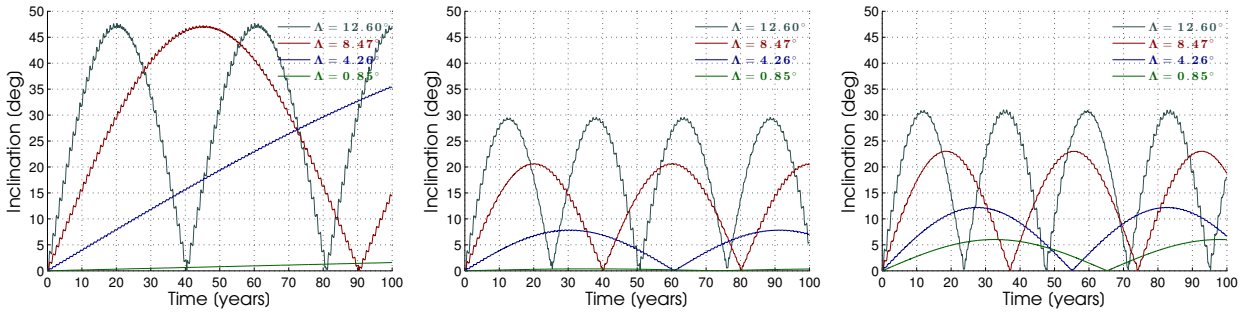
(b) Inclination evolution.

(c) Two-dimensional eccentricity vector evolution ($\Lambda = 12.60°$).

(d) Two-dimensional angular momentum unit vector evolution.

Figure 5.3: Long-term orbit evolution in the Earth-equatorial frame for different values of the SRP perturbation angle, as predicted by the singly-averaged equations of motion. The position vectors of both the Sun and Moon were computed using two-body dynamics, accounting for lunar nodal regression.

and gives rise to long-term aperiodic oscillations in the maximum amplitudes. Regarding the inclination evolution, solar radiation pressure accounts for the sub-yearly oscillations that ride on top of the longer-term secular drift, and the reduction in the long-term oscillation periods with increasing Λ (see Figure 5.4). The addition of Earth oblateness brings about a reduction in both the amplitude and period of the long-term oscillations. Inclusion of lunisolar perturbations causes a slight increase in the long-term amplitudes and a decrease in the long-term oscillation periods, and for certain values of Λ—most notably $\Lambda = 12.60°$ (corresponding to $(1+\rho)A/m = 20.4\ \mathrm{m}^2/\mathrm{kg}$)—causes large fluctuations (peak-to-peak changes) in the maximum amplitudes. These fluctuations manifest themselves as a beating phenomenon in the evolution of the two-dimensional angular momentum unit vector, $[\sin i \sin\Omega,\ -\sin i \cos\Omega]$. Note that for lower values of Λ, this complex behavior is not observed, as shown in Figure 5.3d. The origin of this phenomenon, which also appears in the numerical results of Anselmo & Pardini (2010), will be discussed in the following section.



(a) SRP-only. (b) SRP and oblateness. (c) SRP, J_2, and solar gravity.

Figure 5.4: Accumulation of the effects on the long-term inclination evolution, as each perturbing force is added to the system.

5.3.3 Secular Saros Resonance

The epoch date determines the initial geometry of the Earth-Moon-Sun system, and thus the initial location of the lunar ascending node. We found that when the nodal rate of the perturbed system is commensurate with the nodal rate of the Moon (i.e., the Saros), the perturbations build

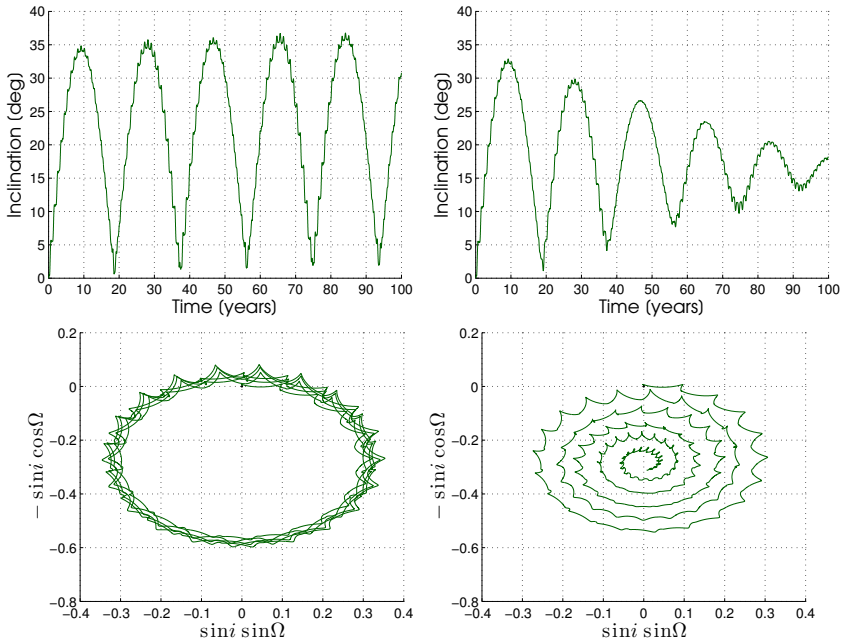
up more effectively over long periods to produce significant resonant effects on the orbit. Such resonances, which occur for a class of HAMR objects that are not cleared out of orbit through their eccentricity growth, gives rise to strongly changing dynamics over longer time periods. This resonant behavior explains the long-term beating phenomenon that occurs for $\Lambda = 12.60°$ (see Figure 5.3d). Its nodal period in the equatorial frame is close enough to the Saros (sidereal period of lunar nodal regression) that there is a strong interaction between the lunar effects and the overall precession rate.

Figures 5.5-5.7 show the inclination and two-dimensional angular momentum vector evolution in the Earth equatorial frame, for several HAMR objects propagated using the same initial conditions, but varying the initial lunar node. Varying the initial location of the Moon's ascending node over 2π is, in a sense, equivalent to varying the release epoch within a Saros cycle. For $\Lambda = 13.81°$, the nodal period in the equatorial frame is approximately equal to 18.61 years, thereby inducing a $1:1$ resonance with the Saros. The qualitative picture of the evolution changes drastically based on this angle, which is indicative of resonance. Figures 5.6 and 5.7 show the evolution of objects with nodal rates either too slow or too fast to resonantly interact with the Saros. Note that for these objects, their orbits will change quantitatively based on the initial node angle, but the qualitative evolution remains the same. The approximate range of SRP perturbation angles for which resonance can be important at GEO is between $\Lambda = 10.5°$ and $\Lambda = 15.5°$ (corresponding to $16.93\ \mathrm{m^2/kg} \leq (1+\rho)A/m \leq 25.33\ \mathrm{m^2/kg}$).

5.3.4 Global Behavior: The Primacy of the Laplace Plane

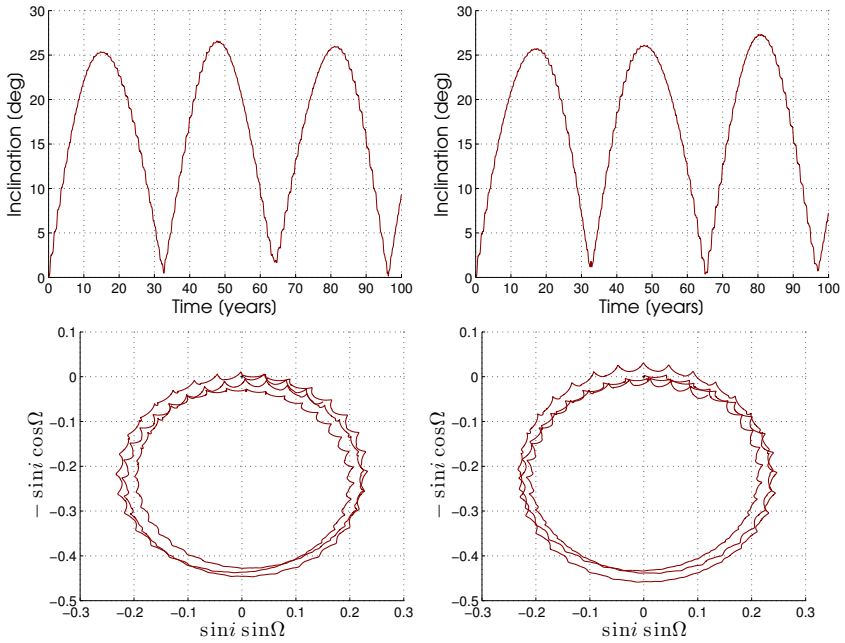
5.3.4.1 The Classical Laplace Plane at GEO

In the geostationary orbit, station-keeping maneuvers are required to constantly maintain the orderly arrangements of operational satellites (Zee 1989). The orbital dynamics of uncontrolled geostationary satellites is governed by the oblateness of the Earth and third-body gravitational interactions induced by the Sun and the Moon. By itself, Earth's oblateness causes the pole of the



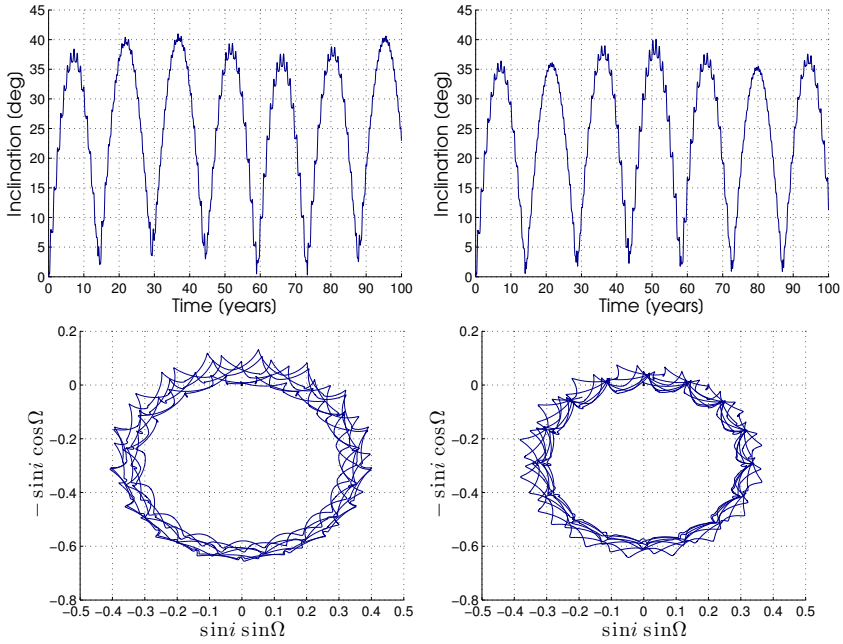
(a) $\Omega_m(0) = \Omega_{m_0}$. (b) $\Omega_m(0) = \Omega_{m_0} + 3\pi/2$.

Figure 5.5: Long-term orbit evolution in the Earth-equatorial frame of an object with $\Lambda = 13.81°$, as a function of the initial lunar node. Note the significant resonant effect caused by the Saros.



(a) $\Omega_m(0) = \Omega_{m_0}$. (b) $\Omega_m(0) = \Omega_{m_0} + 3\pi/2$.

Figure 5.6: Long-term orbit evolution in the Earth-equatorial frame of an object with $\Lambda = 8.47°$, as a function of the initial lunar node. Note that the Saros resonance is not observable.



(a) $\Omega_m(0) = \Omega_{m_0}$. (b) $\Omega_m(0) = \Omega_{m_0} + 3\pi/2$.

Figure 5.7: Long-term orbit evolution in the Earth-equatorial frame of an object with $\Lambda = 16.59°$, as a function of the initial lunar node. Note that the Saros resonance is not observable.

orbital plane to precess around Earth's rotation pole, the rate of rotation being proportional to $J_2(R/a)^2 n$. Lunisolar perturbations, if acting alone, will have a similar effect, but the precession will now take place about the poles of the orbital planes of the Moon and the Sun, respectively, at a rate proportional to n_p^2/n. The motion of the orbit pole of the satellite is a combination of simultaneous precession about these three different axes, one of which, the pole of the Moon's orbit, regresses around the pole of the ecliptic with a period of 18.61 years. As expounded on in Chapter 4, the classical Laplace plane is the mean reference plane about whose axis the satellite's orbit precesses. Under the assumption that the lunar orbit lies in the ecliptic, Allan & Cook (1964) found the condition for the approximate Laplace plane at GEO, which lies between the plane of the Earth's equator and that of the ecliptic and passes through their intersection (i.e., the vernal equinox), and which has an inclination φ with respect to the equator given by (cf. Equation 4.6)

$$\tan 2\varphi = \frac{\sin 2\epsilon}{\cos 2\epsilon + (r_L/a)^5}, \tag{5.11}$$

where the Laplace radius is

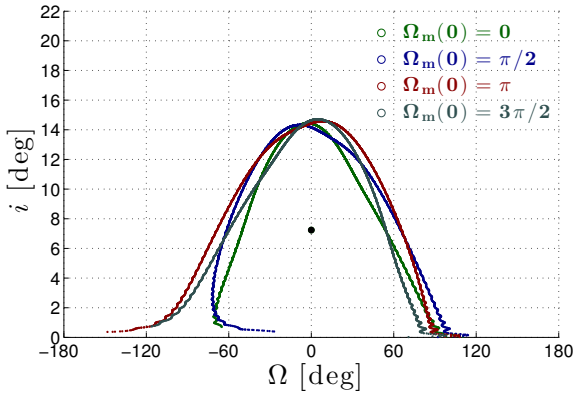
$$r_L^5 = a^5 \frac{\omega_2}{\omega_m + \omega_s}, \tag{5.12}$$

and $\omega_2 = 3nJ_2R^2/2a^2$, $\omega_m = 3\mu_m/4na_m^3h_m^3$, and $\omega_s = 3\mu_s/4na_s^3h_s^3$. Recall that the Laplace radius is the geocentric distance where the effects of oblateness and lunisolar forces are equal. Note that for Earth-orbital dynamics, this equilibrium is stable to both changes in the orbit plane orientation and eccentricity (q.v., Figure 4.2). The geometrical relations between the orbital plane of Earth, its equatorial plane, and the Laplace plane are shown in Figure 4.1a in Chapter 4.

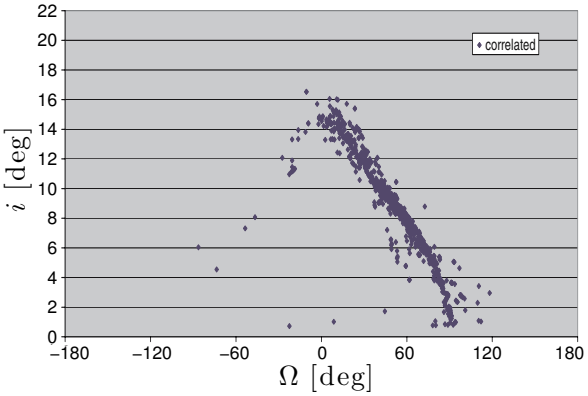
Shown in Figure 5.8 is the long-term evolution of the inclination and ascending node, in the Earth equatorial frame, of initially geostationary satellites. These satellites, following cessation of active station-keeping, precess at a nearly constant inclination about the pole of the Laplace plane with a period of about 53 years (cf. Friesen et al. 1992; Friesen, Kessler & Zook 1993; Schildknecht 2007). Sufficient time has now passed for the orbits of the earliest uncontrolled satellites to complete their predicted long-period motion, as indicated by Schildknecht (2007). The orbital planes of these objects evolve in a predictable way; that is, their inclinations and ascending nodes are strongly correlated (see Figures 5.8a and 5.8b). In fact, the former Soviet Union designed their geostationary satellite constellation to take advantage of this systematic structure: by selecting the initial inclination and ascending node such that the perturbations will naturally reduce the inclination to zero before increasing again, the satellite's inclination is kept below a few degrees over its lifetime without the need of expensive north-south station-keeping (Johnson 1982, 2012).

5.3.4.2 The Modified Laplace Plane at GEO

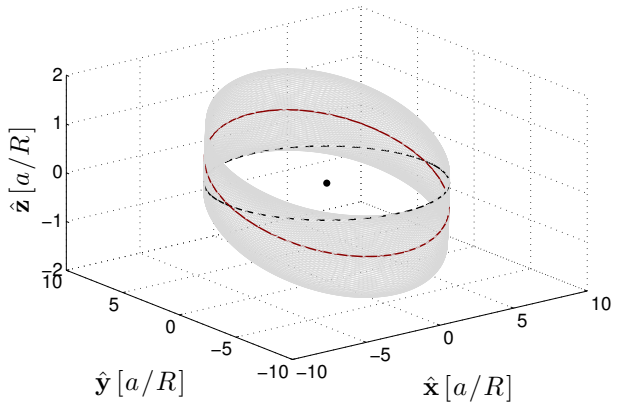
To understand the spatial distribution of the HAMR debris population, and to determine whether their (i, Ω) pattern is also systematic, we investigated a range of HAMR objects, released in geostationary orbit with area-to-mass ratios from 0 up to 40 m^2/kg. We propagated over 80 different Λ values for 100 years, uniformly distributed in initial lunar node (360 different node



(a) Scatter plot of the time-series, over 53 years, of inclination and ascending node as predicted by the singly-averaged model. The classical Laplace plane has $i = 7.2°$ and $\Omega = 0°$ as indicated by the black dot.



(b) Inclination and ascending node of the correlated uncontrolled satellites from January 2004 to August 2005 (figure adapted with permission from Schildknecht 2007).



(c) Three-dimensional picture of the evolution of one of the satellites. The Earth is at the center, the initial geostationary orbit is indicated by the black orbit, and the classical Laplace plane is shown in red. The Laplace plane is the plane of symmetry for the evolution.

Figure 5.8: Long-term motion, in the Earth-equatorial frame, of the orbital planes of initially geostationary satellites. A numerical integration was performed over 53 years of the singly-averaged model for four different initial positions of the lunar orbit, i.e., four different launch dates as follows: July 1964 ($\Omega_m(0) = \pi/2$), March 1969 ($\Omega_m(0) = 0$), November 1973 ($\Omega_m(0) = 3\pi/2$), July 1978 ($\Omega_m(0) = \pi$). Note that the maximum values of inclination occur at $\Omega = 0°$ (i.e., the ascending node of the Sun's orbit) and the minimum values occur at $\Omega = 180°$ (i.e., the descending node).

values from 0 to 2π), giving nearly 30,000 simulations. As evident from Figures 5.5-5.7, the initial location of the lunar node is an important parameter as different behavior occurs depending on where the object is relative to the Moon. Moreover, the initial lunar node can be correlated with any epoch within a Saros cycle, and thus may shed insight into the source of this debris.

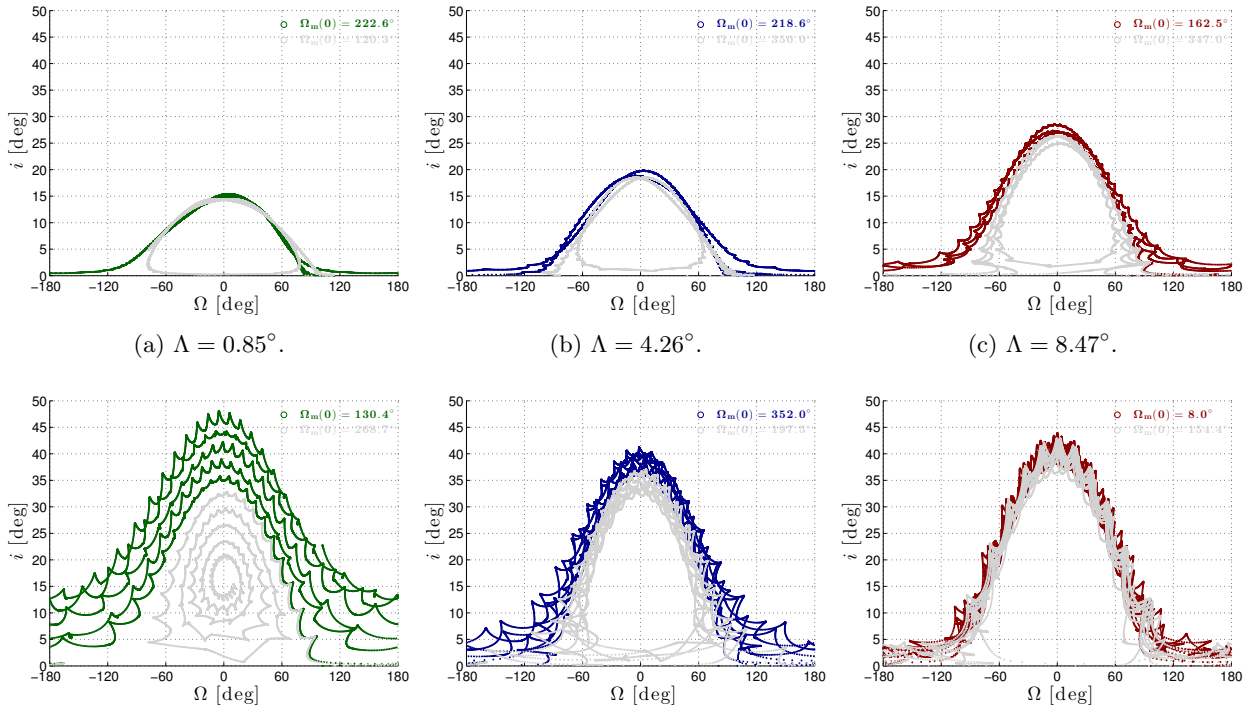
Figure 5.9 shows the time-series, over 100 years, of inclination and right ascension of the ascending node in the Earth-equatorial frame, for a range of SRP perturbation angles, for two trajectories selected out of the 360 different trajectories that are a function of the initial lunar node. That is, we track the statistics over all 360 initial lunar nodes, and show the evolution that gives the largest value in maximum long-term inclination reached, and the evolution that gives the smallest value in maximum long-term inclination reached. Note that the initial node angles that produces these extreme cases will differ depending on the Λ angle. The maximum long-term inclination reached for each HAMR object is given in Table 5.3. For the non-Saros-resonant objects, the qualitative behavior is the same with the maximum long-term inclination increasing with increasing Λ. The $1:1$ Saros resonance for $\Lambda = 13.81^\circ$ can be observed in Figure 5.9d, and leads to this object having the largest maximum inclination across all Λ angles. In all cases, the pattern associated with their distribution in inclination and ascending node phase space is systematic (i.e., strongly correlated), which means that HAMR objects evolve in predictable ways (cf. Figure 5.1c).

The systematic structure of the orbital planes for HAMR objects is a direct reflection of the fact that the classical Laplace plane can be generalized to accommodate solar radiation pressure, as demonstrated analytically in Chapter 4. The approximate mean pole of the modified Laplace plane is given by (cf. Equation 4.18)

$$\omega_2(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{p}}}\cdot\hat{\boldsymbol{h}} + (\omega_m+\omega_s)(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{h}})\widetilde{\hat{\boldsymbol{H}}}_s\cdot\hat{\boldsymbol{h}} + \omega_{srp}\widetilde{\hat{\boldsymbol{H}}}_s\cdot\hat{\boldsymbol{h}} = \boldsymbol{0} \quad \text{or} \tag{5.13}$$

$$\omega_2\sin 2\varphi + (\omega_m+\omega_s)\sin 2(\varphi-\epsilon) + 2\omega_{srp}\sin(\varphi-\epsilon) = 0, \tag{5.14}$$

where the perturbation strength frequencies are defined in Table 5.2. The orientation of the modified Laplace surface between the degenerate states (equatorial and ecliptic) is given as a function



(a) $\Lambda = 0.85°$. (b) $\Lambda = 4.26°$. (c) $\Lambda = 8.47°$.

(d) $\Lambda = 13.81°$. (e) $\Lambda = 16.59°$. (f) $\Lambda = 20.43°$.

Figure 5.9: Scatter plot of the time-series, over 100 years, of inclination and right ascension of the ascending node, for two selected trajectories for each value of Λ. The long-term motion of the orbital plane is shown for two different initial positions of the lunar node for each HAMR object, corresponding to the respective extreme cases. These results agree well with the observed HAMR objects, shown in Figure 5.1c, for which the bulk of the uncorrelated objects lie on the predicted evolution tracks, but with a much larger spread than the correlated objects (cf. Figure 5.8b; for a discussion on the observations, see Schildknecht 2007). The fact that the poles of HAMR objects have a wider precession, as well as the Saros resonance, may account for the observed distribution.

of a/R in Figure 5.10. Solar radiation pressure modifies the classical Laplace plane, increasing its inclination relative to the equator with increasing Λ; each HAMR objects has its own modified Laplace plane for a given semi-major axis and effective area-to-mass ratio (or corresponding modified Laplace surface).

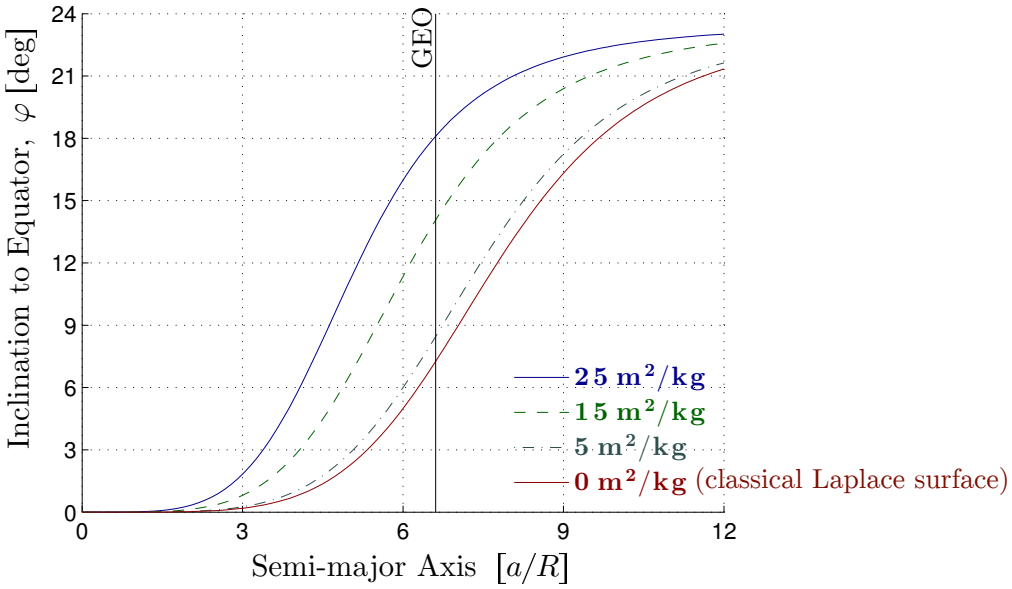


Figure 5.10: Inclination of the modified Laplace equilibrium plane relative to the Earth's equator as a function of semi-major axis in Earth radii for a range of HAMR values. For an object near the Earth, the Laplace plane (both classical and modified) lies approximately in the Earth's equatorial plane, while for distant objects, it coincides with the ecliptic plane; all three planes sharing a common node (the vernal equinox). Between these two limiting states, the Laplace plane at a given semi-major axis lies at some intermediate orientation, generating the warped Laplace surface. Note that an object with $A/m = 0$ m^2/kg corresponds to the classical Laplace surface.

5.3.4.3 Orbital Lifetime of HAMR Debris

The orbital lifetime of HAMR objects is determined by their eccentricity growth and hence the decrease of their orbit perigee radius. An eccentricity above 0.849 for a semi-major axis corresponding to GEO would result in an impact with the Earth. Note that an object occupying this critical eccentricity orbit may not impact the Earth, if the eccentricity changes rapidly enough.

Table 5.3 lists the lowest minimum periapsis reached across all 360 trajectories for each SRP perturbation angle. For a value of $\Lambda = 28.21°$ ($(1+\rho)A/m = 49.0$ m^2/kg), all 360 trajectories result in an eccentricity growth above 0.849 in less than a year, which should lead to shorter lifetimes for this debris. However, if the object does not go through perigee when the critical eccentricity is reached, it could persist in orbit for several decades. The results obtained are consistent with those of Anselmo & Pardini (2010). These limits may actually serve as a test for where such HAMR debris originated, but a more complete understanding of how the maximum eccentricity of a debris varies over time would need to be made.

Table 5.3: Maximum long-term inclination and minimum periapsis radius reached over all 360 trajectories for each SRP perturbation angle. The corresponding effective area-to-mass ratios are also listed. Recall that the Saros resonance becomes important between $\Lambda = 10.5°$ and $\Lambda = 15.5°$.

Λ [°]	$(1+\rho)A/m$ [m^2/kg]	max i [°]	min r_p [R]
0.85	1.36	15.40	6.4
4.26	6.8	19.79	5.6
8.47	13.6	28.56	4.6
12.60	20.4	39.64	3.7
13.81	22.44	48.04	3.3
16.59	27.2	41.21	2.9
20.43	34.0	43.88	2.2
24.08	40.8	44.28	1.5
27.54	47.6	48.03	1.0

5.4 Disposal Orbits for Geostationary Satellites

The importance of managing space debris is acknowledged by all space-faring nations, as the long-term financial, legal, and environmental implications of collisions between high-value satellites are manifest. The discovery of the high area-to-mass ratio debris reveals that the situation in the geostationary orbit region is more critical than previously thought, and is becoming as compelling a problem as in LEO. As we begin to discover the full scope of the debris problem in GEO, we are finding that the current graveyard and mitigation practices are now obsolescent. The low

energy release of HAMR objects from aging satellites abandoned in disposal orbits is not directly addressed in the international policies that established the graveyard (IADC 2002; Johnson 2012). The current disposal region, which is several hundred kilometers above GEO depending on the spacecraft characteristics, is not well suited as a graveyard for two important reasons: it does not mitigate the possibility of collisions between the uncontrolled objects residing in this region,[6] and, more importantly, it cannot contain the HAMR population.

In the early 1990s, when the problems of overcrowding of geostationary orbit with operational and defunct satellites began to emerge, Friesen et al. (1992) suggested the use of the classical Laplace plane for satellite applications and as an orbital debris management strategy for GEO orbit (Friesen et al. 1992; Friesen, Kessler & Zook 1993). The orbit plane of a geosynchronous satellite with such an orientation, notes Otis Graf, would be fixed in space (Graf 1975). The significance of the Laplace plane for use as a GEO disposal orbit is that the orbits of satellites placed in this stable equilibrium will be fixed on average, and that any orbit at small inclination to it regresses around this plane at nearly constant inclination and rate (Allan & Cook 1964). This stable graveyard can be specified for a range of semi-major axes above (or below) GEO, and satellites located in this region will have drastically reduced relative encounter velocities, compared to the conventional graveyard orbits (Friesen, Kessler & Zook 1993). Thus, if collisions were to occur between satellites in the stable graveyard, they would occur at very low velocities, thereby damping out the relative motion of these objects and keeping them in this stable plane (akin to the nature of planetary rings and satellites; see, for instance, Dobrovolskis, Borderies & Steiman-Cameron 1989; Tremaine, Touma & Namouni 2009).

The purpose of this section is to investigate the robustness of the Laplace plane graveyard orbit for the containment of highly perturbed objects such as HAMR debris. Since placing a satellite in this inclined orbit can be expensive, we discuss some alternative disposal schemes that

[6] There is circumstantial evidence that collisions have occurred in the super-synchronous disposal regime. For instance, the decommissioned GOES-10 NOAA spacecraft (International Designator 1997-019A, U.S. Satellite Number 24786) experienced a distinct change in orbital period on 5 September 2011, abruptly falling to a slightly lower orbit in the graveyard region above GEO. As the derelict GOES-10 spacecraft had been completely passivated, the cause of this anomalous orbital perturbation might have been from an impact with an unknown object (NASA Orbital Debris Program Office 2012).

have acceptable cost-to-benefit ratios.

5.4.1 Super-synchronous Disposal Orbit

The current disposal orbit scheme, established by the Inter-Agency Space Debris Coordination Committee and supported by the International Telecommunication Union, is to boost retired satellites into super-synchronous orbits several months before station-keeping fuel is expected to be exhausted (IADC 2002; ITU 2010). The minimum altitude threshold for re-orbiting incorporates the geostationary protected region (i.e., the operational station-keeping zone and maneuver corridor) as well as an allowance for gravitational and non-gravitational perturbations, as shown in Figure 5.11. The stability of the super-synchronous disposal orbits and their potential to reduce collision hazards have been investigated extensively in the literature (Hechler & Van der Ha 1981; Johnson 2012); however, these studies have focused on long-term simulations of intact satellites, which have very low area-to-mass ratios. We have found that this current disposal scheme for end-of-life retirement of GEO payloads is not well suited as a long-term graveyard because it does not mitigate the possibility of collisions between the uncontrolled objects residing in this region, nor does it reduce the severity of such collisions, and it cannot contain the high area-to-mass ratio debris population (Schildknecht 2007).

The long-term orbit evolution of HAMR objects, released from super-synchronous disposal orbits, is shown in Figure 5.12. An increase in Λ results in a faster and wider clockwise precession of the orbit pole (Friesen et al. 1992; Anselmo & Pardini 2010), with the secular precession period as a function of Λ given in Figure 5.13. When HAMR objects return to zero inclination, which occurs much faster than the uncontrolled satellites, they may very rapidly cross the geostationary protected region due to their significant eccentricity oscillations (illustrated in Figure 5.3a). More specifically, the probability of a particular debris object in the disposal orbit striking an active satellite in the GEO belt is largely dependent on the geometry of its orbital plane. Evidently, for the objects to collide, their orbital paths must cross and they must arrive at the intersection at the same time (Greenberg 1982). HAMR objects in eccentric orbits that are inclined to the equatorial

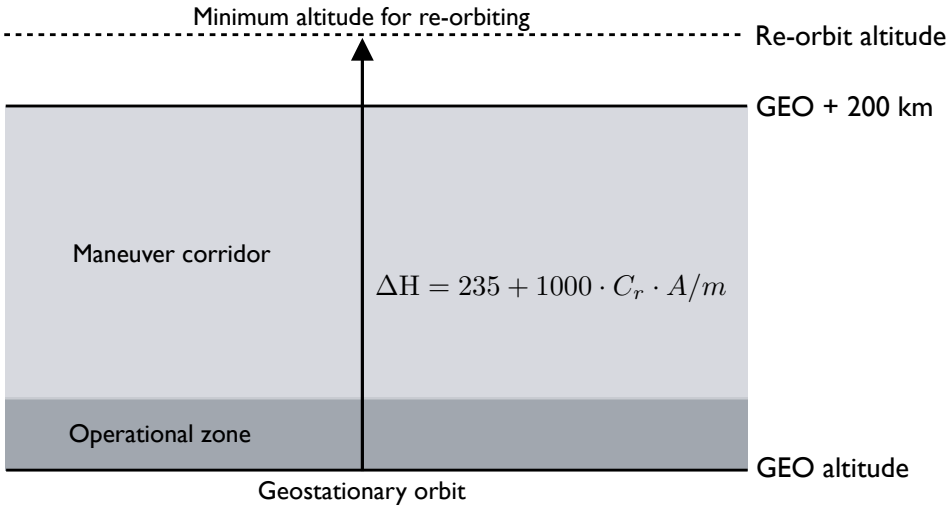
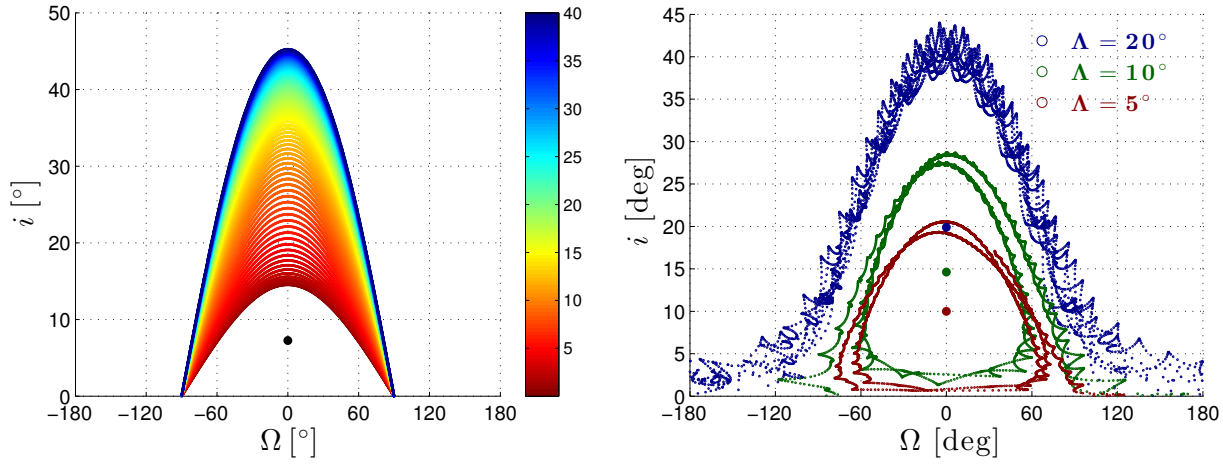


Figure 5.11: The current, internationally established, disposal scheme for end-of-life retirement of GEO payloads. The accepted re-orbiting altitude is specified by ΔH, which accounts for the geostationary protected region and an allowance for perigee oscillation due to gravitational and non-gravitational perturbations (Adapted from ITU 2010).

plane will have a very small but non-zero probability of crossing the geostationary ring; however, HAMR objects that reside in the equatorial plane will have a high probability of impact because they will intersect this ring twice per orbit.

5.4.2 The Laplace Plane Graveyard Orbit

We now reconsider the possibility of using the stable Laplace plane equilibrium as a robust GEO graveyard, a notion originally put forward by Friesen, Kessler & Zook (1993), but which has not been fully appreciated in the scholarly world. Not only will satellites orbiting in this region have drastically reduced relative encounter velocities, thereby reducing the likelihood of a collisional cascade, but we have found that this region is robust to large SRP perturbations. In particular, if satellites located in the classical Laplace plane graveyard orbit shed HAMR objects, they will be trapped in inclination and ascending node phase space, and will not likely cross the GEO protected region, as demonstrated in Figure 5.14.



(a) Qualitative evolution as predicted by the doubly-averaged model (Table 5.2). The colorbar indicates the value of Λ in degrees and the position of the classical Laplace plane is indicated by the black dot.

(b) Long-term orbit evolution based on the singly-averaged model (Table 5.1). The corresponding predicted equilibriums for each object are also shown.

Figure 5.12: Evolution of the orbital planes of HAMR debris, released from the super-synchronous disposal orbit, illustrating the nature of the problem.

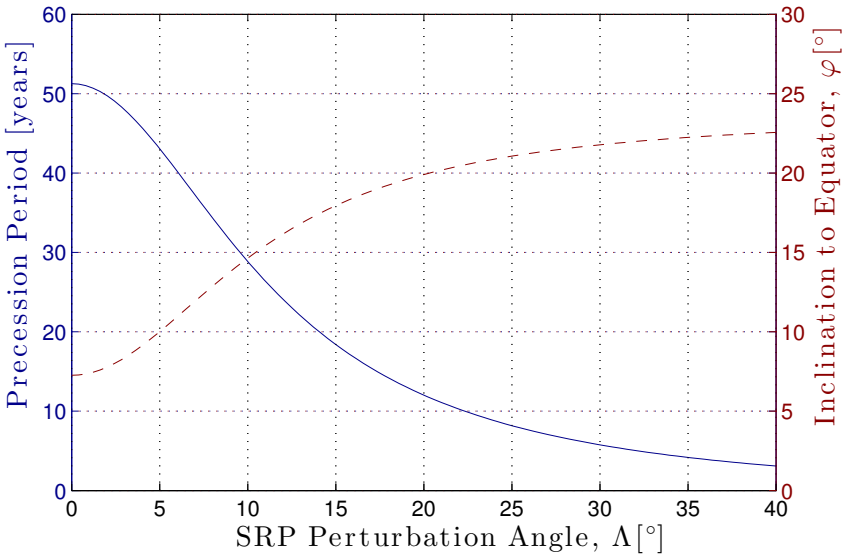
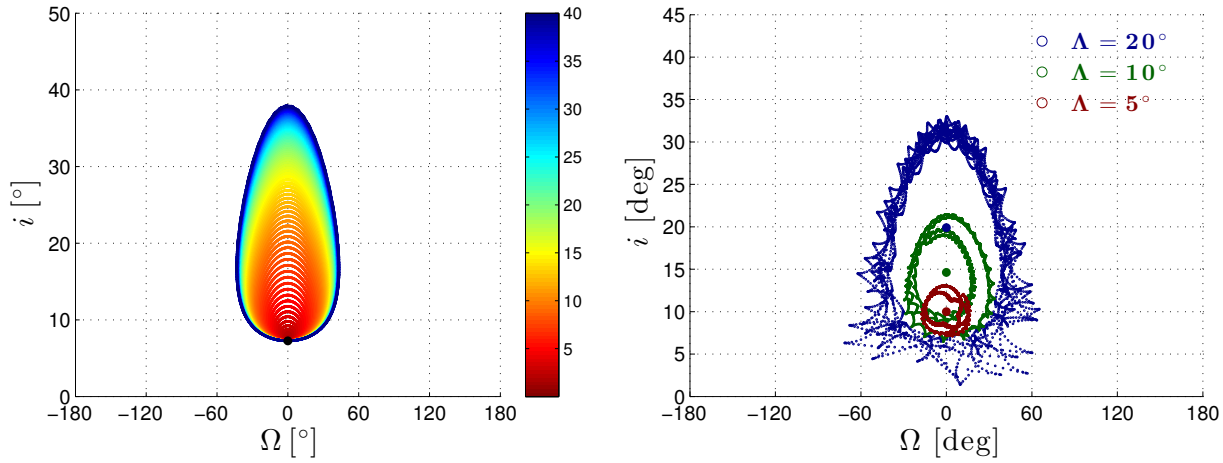


Figure 5.13: Approximate secular precession period of the orbital pole as a function of Λ, and inclination of the modified Laplace equilibrium plane (from Equation 5.14). The pole of the orbit precesses at constant rate and inclination around the approximate mean pole $\hat{\boldsymbol{h}}_L$ with a period $2\pi/\omega$, where $\omega\hat{\boldsymbol{h}}_L = \omega_2\hat{\boldsymbol{p}} + (\omega_m + \omega_s + \omega_{srp})\hat{\boldsymbol{H}}_s$. The modified Laplace plane is the plane of symmetry for the HAMR objects of Figure 5.12.



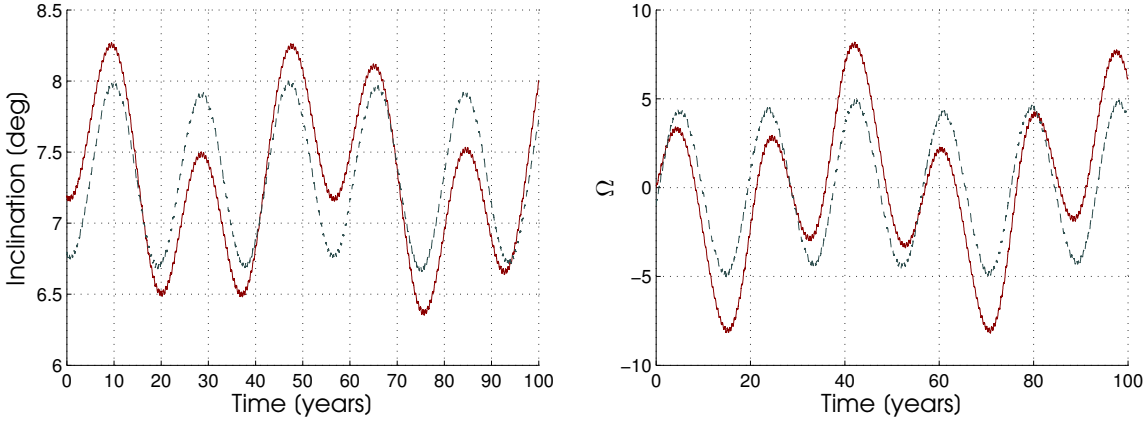
(a) Qualitative evolution as predicted by the doubly-averaged model (Table 5.2). The colorbar indicates the value of Λ in degrees and the position of the classical Laplace plane is indicated by the black dot.

(b) Long-term orbit evolution based on the singly-averaged model (Table 5.1). The corresponding predicted equilibriums for each object are also shown.

Figure 5.14: Evolution of the orbital planes of HAMR debris, released from the classical Laplace plane disposal orbit, illustrating the containment of these peculiar objects.

5.4.2.1 Nearness to reality

We now concern ourselves with the issue of realism. We called the Laplace plane a frozen orbit because it is an equilibrium for the doubly-averaged, autonomous, equations of motion. The method of averaging is based on the notion that the short-period terms discarded in averaging cause only small oscillations, which are superimposed on the long-term, slowly varying, evolution described by the averaged system. The natural question arises as to whether the secular equilibrium will "unfreeze" when these short-term variations are included or when the averaging is pushed to higher order. Figure 5.15 shows the evolution of the Laplace plane, as predicted by our singly-averaged model (which agrees well with the non-averaged equations). The orbit planes of GEO satellites placed in such an orbit experience very little precession, remaining always within 1° of their initial orientation. As indicated by Friesen, Kessler & Zook (1993) and shown by the dashed curve, these excursions, which are caused by the regression of the lunar nodes, can be further reduced by choosing the initial orbit plane orientation to be in phase with the Moon's nodal precession.



(a) Inclination evolution.

(b) Ascending node evolution.

Figure 5.15: Long-term evolution of the inclination and right ascension of the ascending node for a geosynchronous satellite released from the classical Laplace plane equilibrium, showing a comparison between our predicted result (red, solid curve) and the empirical result of Friesen, Kessler & Zook (1993) (gray, dashed curve).

5.4.2.2 Robustness of Laplace graveyard to high-fidelity SRP models

The solution to the modified Laplace equilibrium plane and its implications for the high area-to-mass ratio debris are based on the cannonball model of solar radiation pressure, which assumes that the force acts along the Earth-Sun line. This attitude-independent model neglects any force component normal to the Earth-Sun line resulting from asymmetries in the body's shape or optical properties, which may be important in determining the behavior of these objects.

The general model presented in §1.3.1, however, depends on a large number of complicating factors, and a systematic study of the entire parameter space represents a difficult and laborious task.[7] We have shown in §3.1.2 that if the net direction in which the SRP acceleration acts lies within the Earth's heliocentric orbit plane, the object will have similar dynamics to a cannonball. Therefore, the main "non-cannonball" effects are associated with the out-of-plane component of solar radiation pressure. For simplicity, consider a flat plate object which maintains a fixed orientation with respect to the Earth-Sun line. The disturbing acceleration—in the case of perfect

[7] McMahon & Scheeres (2010a,b) have developed a general model for solar radiation pressure acting on natural and artificial celestial bodies that enables the forces and torques to be analyzed independent of the objects's specific characteristics. Such an analytical model is well-suited for such applications, but will not be pursued in this work.

specular reflection ($\rho s = 1$)—can be represented as

$$\boldsymbol{a}_{srp} = -\frac{2(A/m)P_{\Phi}\left(\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{d}}_s\right)^2}{d_s^2}\hat{\boldsymbol{n}}, \tag{5.15}$$

where we assume that the object is close to the Earth, or $r \ll d_s$, and we ignore the possible effects of the Earth's shadow. The surface normal direction can be specified as

$$\hat{\mathbf{n}} = \cos\theta\cos\phi\hat{\boldsymbol{d}}_s - \cos\theta\sin\phi\hat{\boldsymbol{d}}_{s_\perp} + \sin\theta\hat{\boldsymbol{H}}_s, \tag{5.16}$$

where $\hat{\boldsymbol{d}}_{s_\perp} = \widetilde{\hat{\boldsymbol{H}}}_s \cdot \hat{\boldsymbol{d}}_s$.

Simulations have been carried out for a range of HAMR objects, released from the Laplace plane graveyard orbit with effective area-to-mass ratios from 0 up to 40 m^2/kg. The SRP acceleration, Equation 5.15, was averaged over the object's unperturbed orbit and included in the singly-averaged model . The angle θ in Equation 5.16 was varied between 0° and 90° and ϕ assumed a value between −45° and 45°. For the object with $\theta = 0°$, the evolution in the (i, Ω) phase space is similar to what was predicted using the cannonball model. When the SRP acceleration direction is tilted out of the ecliptic plane, the symmetry of the motion about the modified Laplace plane is eradicated for large values of ϕ. Nevertheless, in all cases considered, these HAMR objects if started from this inclined graveyard orbit never crossed through zero inclination over a hundred year evolution. Moreover, only the objects that are in near resonance with the Saros ($17\ \mathrm{m}^2/\mathrm{kg} \leq (1+\rho)A/m \leq 25\ \mathrm{m}^2/\mathrm{kg}$) came within 5° of the geostationary orbit.

5.4.3 Economic Viability and Alternative Disposal Option

The current disposal orbits (Figure 5.11) free desirable longitudinal positions for replacement satellites and reduce immediate collision hazards in GEO with an acceptable cost-to-benefit ratio, an important criterion. The cost in terms of incremental velocity is not more than 3.65 m/s per 100 km increase in altitude, which amounts to the fuel needed for one month operational station-keeping (Hechler & Van der Ha 1981). To place a satellite into the Laplace plane graveyard, not only must the current practice be implemented to remove the satellite from GEO altitude, but the

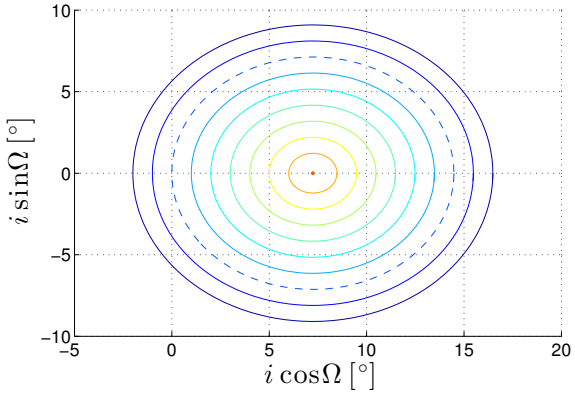
satellite's orbit must be inclined by about 7.2°. This expensive plane change maneuver requires an incremental velocity of roughly 388 m/s, a third the cost of placing the satellite on an escape trajectory (for comparison, a de-orbit maneuver at GEO requires ∼1.5 km/s and an Earth-escape maneuver costs ∼1.2 km/s; Petro 1992). The IADC (2002) guidelines for storage orbits are based on a one-dimensional problem: define a safe minimum re-orbiting distance above GEO needed to isolate the retired satellites from GEO. It was noted, however, that these guidelines should be updated as new information becomes available regarding the space environment, as to define what constitutes an effective graveyard.

An alternative disposal option to the Laplace plane graveyard is to simply incline the current disposal orbits by a minimum inclination, as shown in Figure 5.16, such that the HAMR objects released here will not cross through the geostationary orbit. For instance, a modest plane change maneuver of half a degree would only require an incremental velocity of less than 27 m/s. The basic physical principle behind this two-dimensional disposal scheme is easy to grasp. The orbit's angular momentum vector sweeps out a circular cone around the fixed pole of the classical Laplace plane, with the radius of the circular path being a function of the orbit's inclination to this fixed plane (Figure 5.16a). Accordingly, tilting the orbit normal towards the axis of the Laplace plane shrinks this circle, and as the modified Laplace plane is further inclined, the HAMR objects released here will not cross below the initial inclination of the disposal orbit.

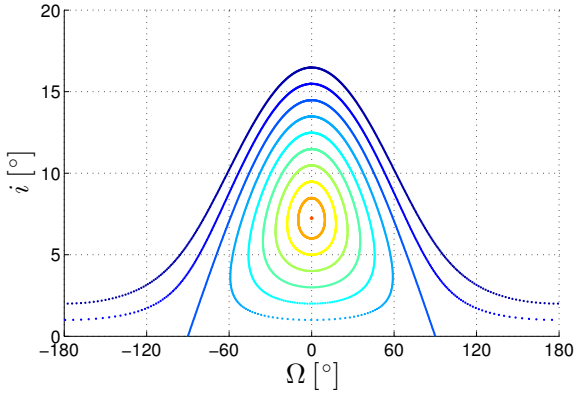
5.5 Discussion

5.5.1 Accuracy of Averaged Equations

For our averaged model, the disturbing function was limited to the cannonball model of SRP without Earth shadow effects, the dominant zonal harmonic in the harmonic expansion of Earth's gravitational potential, and the lowest-order term in the Legendre expansion of the lunar and solar disturbing functions (i.e., Hill's approximation). Under these approximations, the semi-major axis does not undergo any secular changes and the problem reduces to understanding the remaining



(a) Evolution as viewed from Earth's polar axis.



(b) Structure of the evolution in inclination and ascending node phase space.

Figure 5.16: Qualitative evolution, over 55 years, of the orbital planes of satellites released from inclined super-synchronous orbits, showing how the alternative graveyard orbit scheme would keep objects out of the equatorial plane. The dashed curve in Figure 5.16a and the separatrix-like curve in Figure 5.16b are the plane precession of a satellite released from zero inclination (i.e., geostationary orbit). The classical Laplace equilibrium plane is the dot in the center of these plots.

four orbital elements, e, i, Ω, and ω, at a given semi-major axis. For geosynchronous orbits, the tesseral harmonics in the geopotential coupled with solar radiation pressure can introduce long-term changes and even chaotic behavior in the semi-major axis (Lemaître, Delsate & Valk 2009). Hubaux et al. (2013) have demonstrated that the Earth shadow effects, and even the precise model of shadow, can drastically modify the chaotic zone. However, these effects are localized to a narrow range of semi-major axis (hundreds of meters) and will not significantly affect the long-term orbit evolution. Indeed, even a 100 km change in semi-major axis will cause only a 0.1 percent change in the SRP perturbation angle.

The dynamical behavior underlined by our averaged model is in good agreement with earlier researchers (cf. Liou & Weaver 2005; Chao 2006; Valk, Lemaître & Anselmo 2008; Anselmo & Pardini 2010). We attribute any quantitative differences between the singly-averaged and non-averaged simulations to our use of the Hill approximation for the lunar perturbing potential; namely, in the assumption that $r/d_m \ll 1$. Since geosynchronous orbits are relatively small in comparison with that of the Earth and the Moon, only the first non-constant term needs to be retained in the

Legendre expansion of the third-body disturbing function to obtain a good representation of the object's motion. However, since HAMR objects are in highly-eccentric GEO orbits, higher-order terms in the lunar disturbing function expansion may become important. Lidov (1962) showed that the parallactic term (third harmonic or octupole moment) can cause the eccentricity to build up more rapidly if the orbit is sufficiently large. In particular, the parallactic term becomes effective when the object is at a distance of about 10 Earth radii. Thus, we find that for an initial orbit at GEO semi-major axis, if the object has an eccentricity greater than 0.5, which occurs for $\Lambda > 15°$, the parallactic term may need to be considered. The perturbation cause by this term, as well as other secondary perturbations such as Earth's shadow effects, can easily be included into our general framework (see Appendix F).

5.5.2 Saros Resonance Phenomenon and (i, Ω) Phase Space

The recognition of the Saros resonance raises many questions of interest concerning the nature and evolution of the HAMR debris population. This phenomenon actually appears in the numerical results of Anselmo & Pardini (2010), and is relevant for many of the observed HAMR debris in near GEO orbits. Since the singly-averaged results capture this subtle behavior, our averaged model accounts for the full dynamics precisely, and can be used for accurate long-term predictions. This resonance leads to complex evolutionary behavior and may play a role in generating orbital chaos.

Observers are interested in knowing what the best search strategy is to maximize the detection efficiency of HAMR objects for future surveys. From this point of view, the distribution in (i, Ω) phase space indicates exactly where their surveys should concentrate (Figure 5.17). The systematic orientation of the orbital planes indicates that an anti-solar direction survey should concentrate near the equator during the spring and fall. During the summer and winter, the observers should look at high latitudes for highly inclined objects.

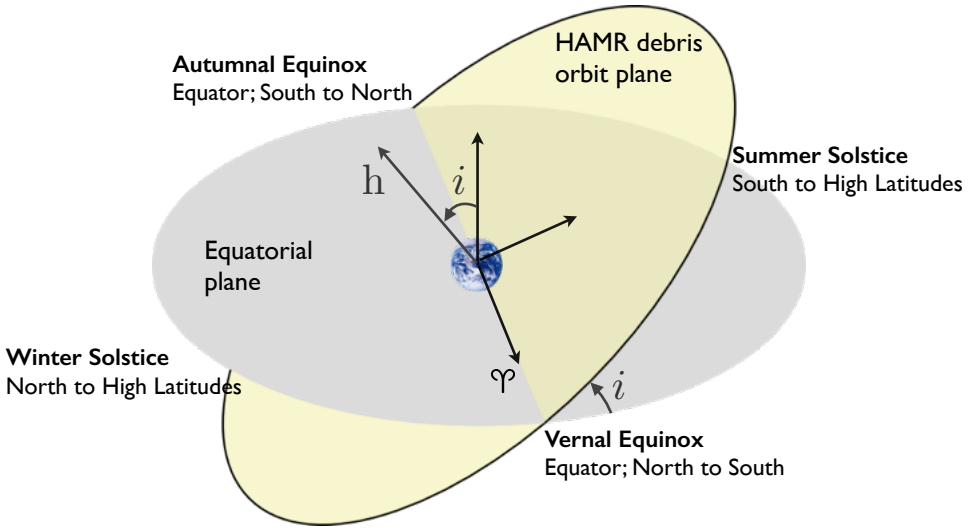
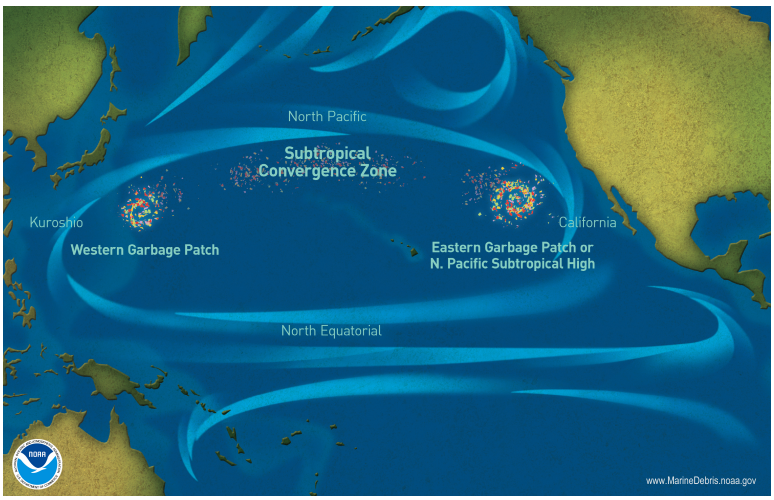


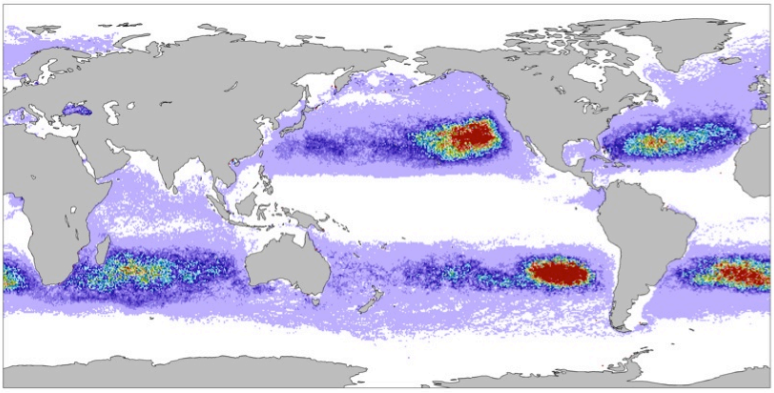
Figure 5.17: A geosynchronous orbit debris search strategy based on the distribution of the orbital planes of HAMR objects.

5.5.3 Graveyard Orbits

In accordance with Friesen, Kessler & Zook (1993), we propose the use of the stable Laplace equilibrium plane as a long-term graveyard for GEO. The Laplace plane graveyard would trap HAMR debris in (i, Ω) phase space and would not allow these objects to rain down through geostationary orbit. This new graveyard region, being surrounded by closed precessional trajectories (Figure 5.16b), may also be robust for the containment of objects released from an explosion or fragmentation event; although future numerical studies are needed to confirm this proposition. The Laplace plane graveyard orbit is, in some sense, analogous to the ocean gyres (Figure 5.18), which are characterized by exceptionally high concentrations of marine debris that have been trapped by the currents (Law et al. 2010). This analogy is, however, misleading as to the nature of these stable environments: the currents naturally bring floating debris toward the ocean gyres (Figure 5.18b), whereas it is up to the spacecraft operators to place their retired satellites in the Laplace plane. For this reason, we proposed an alternative, cost-effective graveyard orbit based on the structure of the orbit plane evolution in the (i, Ω) phase space.



(a) The North Pacific subtropical gyre—one of several major ocean gyres, where large, wind-driven currents circle and push water toward the center (Credit: NOAA).



(b) A ten year numerical simulation of the distribution of marine debris, initially seeded uniformly on the sea surface, showing how the floating bits will accumulate in the five subtropical gyres due to converging surface currents (Credit: Dohan & Maximenko 2010).

Figure 5.18: The general ocean surface circulation, governed by prevailing global winds, Earth's rotation, and the restriction of flow by continental boundaries, produce the five subtropical convergence zones or gyres: the Indian Ocean gyre, the North and South Pacific gyres, and the North and South Atlantic gyres (qq.v., Law et al. 2010; Dohan & Maximenko 2010).

Chapter 6

Circumplanetary Dust Particles

But it seems, that one part of [Iapetus's] surface is not so capable of reflecting to us the light of the Sun which maketh it visible, as the other part is.[1]

— Giovanni Domenico Cassini, 1677

6.1 Saturn's Enigmatic Satellite Iapetus

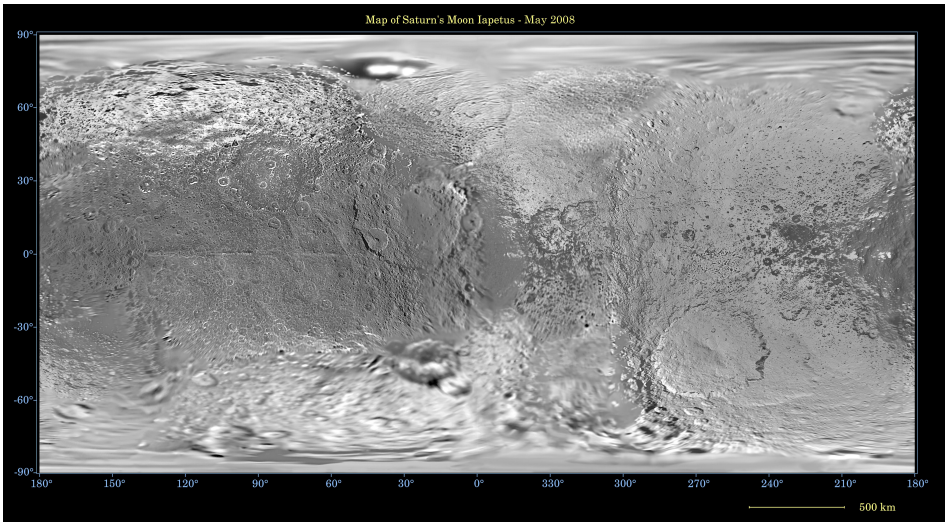
Iapetus, Saturn's third largest satellite, is one of the most peculiar bodies in the Solar System (Wilson & Sagan 1996). Ever since its discovery by Cassini in 1671, Iapetus has confounded astronomers with its distant, near-circular orbit of high inclination to Saturn's equatorial plane, its striking hemispheric brightness asymmetry, its heavily cratered surface, its synchronous rotation period, its odd shape, and its peerless equatorial ridge (Mosqueira, Estrada & Charnoz 2010). While the origin of Iapetus is still unknown, with many theories of formation challenged by Iapetus's unique orbital configuration and large size (Ward 1981; Burns 1986; Tremaine, Touma & Namouni 2009), a clearer physical picture for the origin of Iapetus's extreme albedo dichotomy is beginning to emerge (Soter 1974; Verbiscer, Skrutskie & Hamilton 2009; Spencer & Denk 2010; Tamayo et al. 2011). Iapetus's surface is darkest at the apex of motion ($\sim 90°$W) with a high degree of symmetry about this point and extending in a roughly elliptical region, covering nearly 40% of the leading hemisphere (Figure 6.1a). The bright trailing-side terrain extends over the poles onto the leading side. Spencer & Denk (2010) have suggested that an exogenic deposition of dark material,

[1] Cassini GD. 1677. *Philosophical Transactions of The Royal Society of London* 12: 831–833

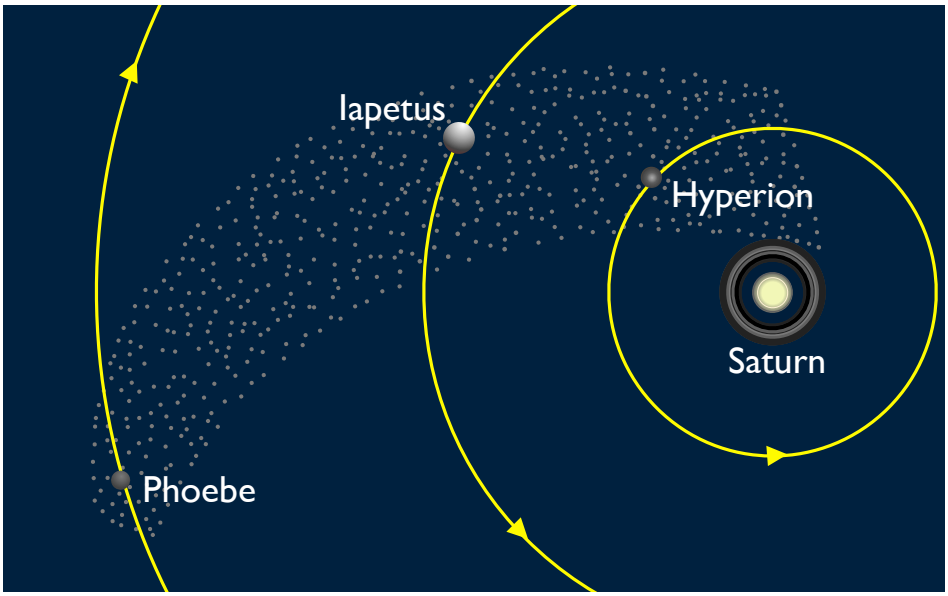
from either a heliocentric source or from Saturn's outer retrograde satellites, results in an initial darkening of the leading hemisphere (Figure 6.1b). This initial albedo dichotomy then triggered ice sublimation and redeposition, which can make equatorial ice migrate towards Iapetus's poles. The accretion of exogenous particles coupled with thermal migration of water ice proves to be the most plausible explanation for the main geometrical features of Cassini Regio, the dark side of Iapetus. However, the mechanism and source for the dark material needed to initiate thermal segregation remains unknown. Generally, dust from Phoebe that spirals in due to Poynting-Robertson (P-R) drag is implicated as the initial source (Soter 1974; Verbiscer, Skrutskie & Hamilton 2009; Tamayo et al. 2011), although a qualitative foundation for this hypothesis is hitherto missing (Leiva & Briozzo 2013). We show here how certain features of the modified Laplace surfaces provide strong support for this supposition, but predict the existence of grain-size-dependent heterogeneity in the surface deposition pattern.

It was shown by Laplace (1799) that the plane of the planet's equator, in its very slow motion caused by the action of the Sun, carries with it those satellites whose orbits initially lie in that plane (Vol. 2, Fifth Book, Chap. 3). More generally, if the precession rate of the satellite's orbit is fast compared with the rate of change of the planet's spin axis, then the satellite's inclination relative to the Laplace surface is an adiabatic invariant (Oudemans 1888; Goldreich 1965). The same is true in the case of tidal dissipation; accordingly, Cazenave, Dobrovolskis & Lago (1980) and Mignard (1981), during their investigation of the origin of the Martian satellites, showed that a gradually evolving satellite that lies in its planet's orbital plane and tidally migrates inward would end up in the planet's equatorial plane. Thus, for slow enough evolutions the orbital inclination remains fixed relative to the Laplace surface. Analogously, a dust particle ejected from Phoebe would maintain a fixed inclination relative to its modified Laplace surface (Figure 4.5) as it slowly migrated inward under P-R drag.

The orbit of Iapetus and its evolution has been studied for over two centuries (Laplace 1805; Hall 1885; Tisserand 1896; Brouwer & Clemence 1961b; Harper 1987). At the distance of Iapetus from Saturn ($\sim$ 59 Saturn radii), the Laplace plane nearly bisects the angle formed by Saturn's



(a) Global map of Iapetus from the Cassini and Voyager spacecraft flybys, depicting the geometrical features of Cassini Regio (Credit: NASA/JPL/Space Science Institute).



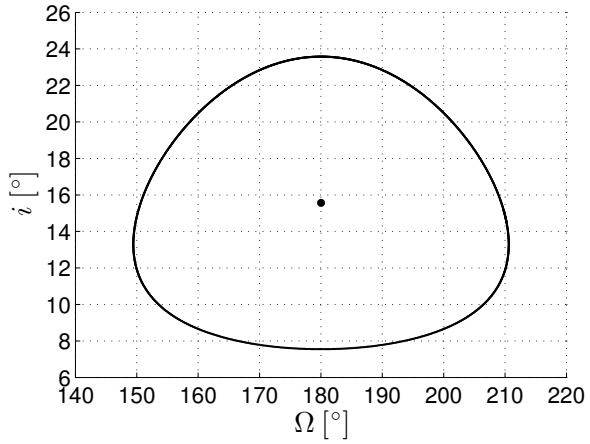
(b) Soter's (1974) Phoebe dust model postulating that dark material impact-ejected from Phoebe, slowly spirals in toward Saturn, and is then collected preferentially on the leading side of the tidally locked Iapetus (adapted from McKinnon 1999).

Figure 6.1: Global surface map of Iapetus showing its striking albedo dichotomy, and an illustration of Soter's (1974) model of dust infall from Phoebe, which presumably triggered the process of temperature-driven water-ice sublimation invoked to explain Iapetus's bright poles (Spencer & Denk 2010).

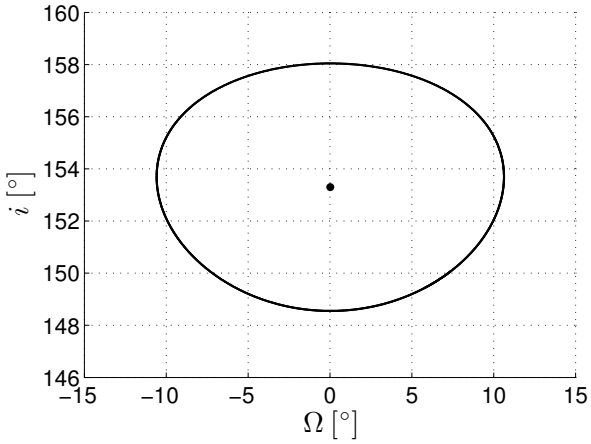
equatorial plane and its orbit. Iapetus has an eccentricity of 0.028 and its orbit precesses with a period of about 3300 (Earth) years while maintaining a nearly constant inclination of 7.9° relative to the Laplace surface, as shown in Figure 6.2. In general, the probability of a particular dust grain striking Iapetus depends critically on the mutual geometry of their respective orbital planes. For the dust and Iapetus to collide, their orbital paths must cross and they must arrive at the intersection at the same time (Greenberg 1982). The dust particles generated in an impact with Phoebe will approximately share that moon's eccentricity of 0.156, and as SRP causes periodic changes in the orbit eccentricity over one Saturn year, the dust grains will remain in eccentric orbits throughout their inward migration (Tamayo et al. 2011). Nevertheless, since Iapetus's orbital period is several orders of magnitude shorter than the migration timescale,[2] even dust grains in eccentric orbits that are highly inclined to Iapetus's orbital plane can strike the moon. Indeed, Tamayo et al. (2011) have shown through detailed numerical simulations that nearly all dust particles of size $\gtrsim 10\,\mu$m will eventually impact Iapetus. Phoebe dust, spiraling inward on retrograde orbits, would strike Iapetus's leading hemisphere with some wrapping onto the edge of the trailing hemisphere by particles with non-zero eccentricities, with smaller particles extending farther (Tamayo et al. 2011). Just as dust eccentricities act to extend coverage longitudinally, dust-orbit inclinations and Iapetus's orbital precession (Figure 6.2a) should extend coverage latitudinally (Wilson & Sagan 1996; Tamayo et al. 2011), with particles reaching beyond the poles by approximately 24° (the maximum relative tilt of the orbits of Iapetus and Phoebe). We now investigate whether the sparse sampling of particle sizes (5, 10, 25, 50, 100, and 500 μm) considered by Tamayo et al. (2011) was sufficient to capture all of the possible behaviors; that is, if it can account for the entire spatial distribution observed on Cassini Regio and necessary to seed the thermal segregation process.

Saturn's diffuse outer dust ring associated with Phoebe (Figure 6.3) forms a symmetric structure about Saturn's orbital plane extending out of this plane by about 5°, and its dust grains, like Phoebe (Figure 6.2b), follow retrograde orbits (Verbiscer, Skrutskie & Hamilton 2009). The dust

[2] The P-R decay timescale grows linearly with particle size (Wyatt & Whipple 1950; Burns et al. 1979); for reference, 10 μm particles liberated from Phoebe will reach Iapetus in ~ 1 Myr (Tamayo et al. 2011).



(a) Orbital plane evolution of Iapetus over 115 Saturn years.



(b) Orbital plane evolutionPhoebe over 27 Saturn years.

Figure 6.2: Long-term evolution of the orbital planes of Iapetus and Phoebe in the Saturn-equatorial frame, where the node is reckoned from the vernal equinox of Saturn. The local Laplace plane at Iapetus's semi-major axis ($1.06r_L$) has $i = 15.6°$ and $\Omega = 180°$, and that at Phoebe's distance ($3.85r_L$) coincides with Saturn's orbital plane, as indicated by the points. See Figure 4.4 for the global (i, Ω) phase space structure and how these trajectories fit within those diagrams.

grains in the Phoebe ring would maintain a fixed inclination relative to their respective modified Laplace surfaces as they spiral in towards Iapetus. We expect from Figure 4.5 that decaying particle orbits lying initially in Saturn's orbital plane with $0 \leq (1+\rho)A/m \leq 31.2$ m^2/kg (Figure 4.5a) will follow the surfaces shown by the solid red lines, while those with $44.2 \leq (1+\rho)A/m \leq 150$ m^2/kg (Figure 4.5c) will follow the gray surfaces. The orbital planes of dust grains with $31.2 < (1+\rho)A/m < 44.2$ m^2/kg (Figure 4.5b) spiraling in from Phoebe along the red surface should begin precessing around the nearest stable surface at the point where the former surface terminates, before proceeding inwards to Iapetus; the amplitude of this large scale oscillation is given by the angular distance to the new stable surface. Note that these retrograde orbits can be mapped into corresponding prograde orbits ($\hat{\boldsymbol{h}} \to -\hat{\boldsymbol{h}}$) to see which dust particles will lie in the same plane as Iapetus,[3] as shown in Figure 6.4. For the particles that evolve along the red surface, only those with $18 \lesssim (1+\rho)A/m \lesssim 28$ m^2/kg (or $26 \gtrsim s \gtrsim 16.7\,\mu$m, assuming particles share Phoebe's density

[3] It is evident from simple geometrical consideration that dust particles lying in Iapetus's orbital plane will preferentially impact near the equator, while those in inclined orbits will have a greater probability of impacting Iapetus at higher latitudes, nearer the poles.

of 1.6 g/cm^3) will, during the course of their long-term migration, lie in Iapetus's orbital plane and thus coat this moon at low latitudes (near the apex of motion). Of the very small particles ($(1+\rho)A/m \gtrsim 44.2$ m^2/kg or $s \lesssim 10.6\,\mu$m), those with radius approximately less than $4.3\,\mu$m will have a high probability of impacting around the equator. However, dust grains smaller than $3\,\mu$m will likely impact Saturn or escape the system due to their eccentricities approaching unity.[4] Particles in the range $31.2 < (1+\rho)A/m < 44.2$ m^2/kg ($15 > s > 10.6\,\mu$m) will not reside in the same plane as Iapetus due to their large amplitude oscillations. Particles outside these two narrow ranges will, based on simple geometry, preferentially strike the surface of Iapetus at higher latitudes. We therefore conclude that the sample of dust particles studied by Tamayo et al. (2011), although too sparse at small sizes for precise calculations, is representative of all possible global interactions of the dust with the Iapetian surface.

Given the higher impact flux at Iapetus's apex from dust particles of sizes $26 \gtrsim s \gtrsim 16.7\,\mu$m and $4.3 \gtrsim s \gtrsim 3\,\mu$m, we hypothesize that the spatial distribution of dark material on Cassini Regio is grain-size dependent: there exists a substantial latitudinal heterogeneity in the surface deposition pattern. We note that Tamayo et al. (2011) have predicted a similar longitudinal heterogeneity due to dust eccentricities induced by radiation pressure. These two predictions, taken together, may help constrain models of thermal ice migration invoked by Spencer & Denk (2010) to explain the bright polar regions beyond $\pm 60°$ latitude (see Figure 6.1a). Unfortunately, as Tamayo et al. (2011) point out, it is difficult to observationally determine the true nature and depth of the dark layer. Nevertheless, these results have direct implications for the dynamics of dust grains in the ring associated with Phoebe, revealing the complexity of the perpetual process of mass transfer occurring in the Solar System (Tamayo et al. 2011; Tamayo, Burns & Hamilton 2013).

[4] An eccentricity above 0.995 for a semi-major axis corresponding to Phoebe's orbit would result in an impact with Saturn (or its rings). We note that this corresponds to $\Lambda \sim 44.9°$, and thus we do not expect dust grains with Λ angles larger than this value to persist in orbit if ejected from Phoebe.

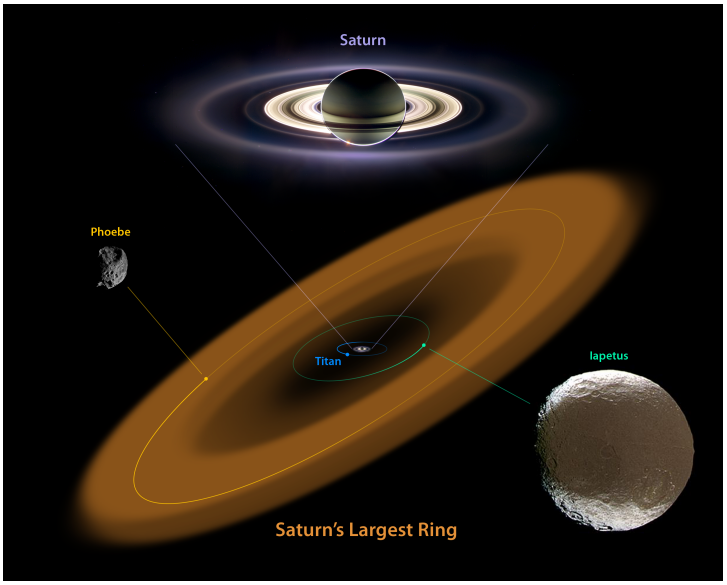


Figure 6.3: Diagram of the tenuous Phoebe ring, made up of a sparse collection of ice and dust particles, discovered by NASA's Spitzer Space Telescope (Credit: NASA/JPL-Caltech). (see Verbiscer, Skrutskie & Hamilton 2009, for more details).

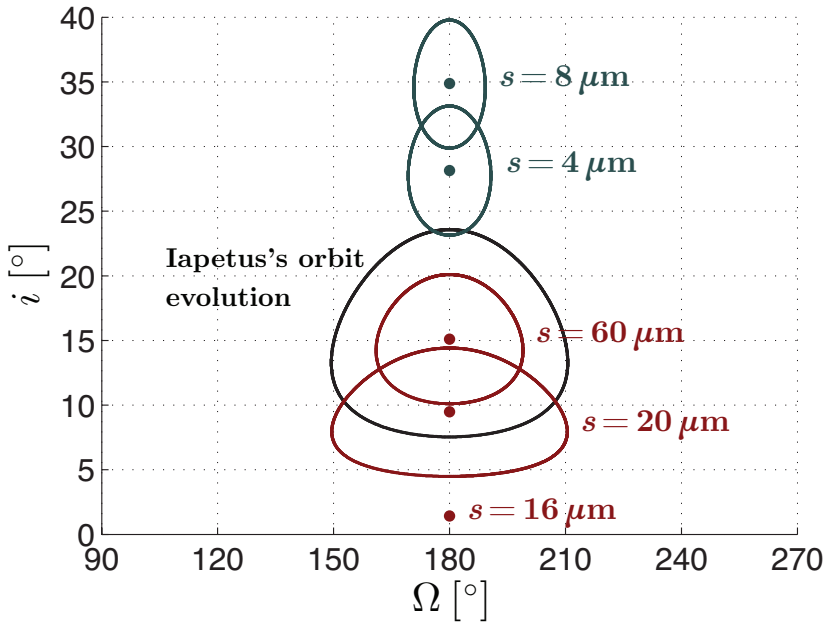


Figure 6.4: Evolution of the orbital planes of Iapetus and of particles at 5° inclinations to their modified Laplace surfaces (mapped into prograde orbits). The precessional trajectories around the modified Laplace planes of particles $\gtrsim 26$ will lie within the black curve (corresponding to Iapetus's orbital plane evolution); while those for particles with $4.3 \lesssim s \lesssim 10.6\,\mu$m will lie outside the black curve. The figure shows that only particle sizes of a certain range will reside in the same plane as Iapetus during their inward migration.

6.2 Other Applications

Dust is a ubiquitous feature of the Solar System, being largely created by micrometeoroid bombardment and mutual collisions among celestial bodies. Bottke et al. (2010) have shown that an enormous amount of dust—one-thousandth of the mass of the Earth's moon, by their estimates—has been produced in the outer regions of each giant planet through collisional cascading among large primordial irregular satellites. The modified Laplace surfaces, therefore, have direct application to many planetary systems.

Dobrovolskis (1980) concludes from a variety of considerations that the polar Laplace plane might host planetary rings. In particular, Dobrovolskis, Borderies & Steiman-Cameron (1989) have proven, in the context of hypothetical polar rings around Neptune, that such rings are stable in the presence of energy-dissipating collisions between the ring particles, and have demonstrated the settling of these rings into the polar equilibrium orientations. However, such a ring system cannot be composed micron-sized dust particles since the absence of polar equilibria in the modified case precludes this possibility. If these faint, highly inclined dust rings exist, then radiation pressure will produce a diffuse concentration of dust particles in the modified orthogonal planes (see Figures 4.6-4.8 for illustrations).

It has long been suggested that the continuous hypervelocity micrometeoroid impacts into the Martian moons should release clouds of dust, forming tenuous tori around Mars (see Hamilton 1996, and references therein). These putative dust belts have not been observed, with both attempts producing negative results (Showalter, Hamilton & Nicholson 2006), and several uncertainties exist about the numerical density and size distribution of the dust grains. The modified Laplace plane implies that these dust rings would be vertically asymmetric as each dust grain would have its own distinct plane of symmetry, in accordance with the prediction of Hamilton (1996). However, each of these planes intersect at the Martian equinoxes: this would be the point where the density of dust grains would be greatest.

There has been considerable interest in determining the radial and vertical extent of Saturn's

diffuse outer Phoebe ring ever since its discovery in 2009 (Verbiscer, Skrutskie & Hamilton 2009; Tamayo, Hedman & Burns 2014, and references therein). While dust material has been observed between 130 and 270 Saturn radii, dynamical arguments suggest that this dust ring extends radially outward up to 300 Saturn radii and inward to the orbit of Iapetus. Future observations, at shorter optical wavelengths, are being planned with the Cassini spacecraft to place constraints on ring-particle size and make measurements of the ring closer to Saturn (Tamayo, Hedman & Burns 2014). Such measurements mostly observe the smallest grains ($\lesssim 12\,\mu$m) that make up the bulk of the population's surface area. We expect that particles with $4.3 \lesssim s \lesssim 10.6\,\mu$m, which have the fastest migration timescales, will actually drift past Iapetus due to their large eccentricities and inclinations. Our approach can thus be used to determine the expected radial extent of the Phoebe ring. The modified Laplace surfaces, moreover, suggest that the ring is going to have dramatic asymmetries in its vertical structure.

This formalism is ideal for studying dust infall in the Uranian system, and may explain the similar—though far less drastic—hemispherical brightness asymmetries observed on the outermost four regular satellites of Uranus (q.v., Tamayo, Burns & Hamilton 2013, and references therein). In this system, since Uranus has an obliquity of 97.9°, Tremaine, Touma & Namouni (2009) noted that gravitationally dominated orbits following the classical Laplace surface would become unstable, executing large oscillations in eccentricity and inclination near the Laplace radius (see §4.2.2, and also Tamayo et al. 2013, for more details). Tamayo, Burns & Hamilton (2013) investigated how these instabilities manifest when solar radiation pressure was considered, and observed a fundamental asymmetry between prograde and retrograde particles in regard to the distribution of collision probabilities among the inner moons. The dynamical cause of these asymmetries appears to be related to the warping and stability of the modified Laplace surfaces, expounded on in §4.3.2, although a more detailed analysis of this system would need to be made.

Chapter 7

Conclusions and Future Work

In the mid-twentieth century, Samuel Herrick, an astronomer of towering reputation, posed the vital question of determining the most suitable system of coordinates for solving problems in celestial mechanics (Herrick 1948). As Finlay-Freundlich remarks:

> In every theory of natural science we have to distinguish between two essential parts: the one concerns the fundamental concepts derived by abstraction and generalization from the most primitive observable quantities; the second, the fundamental laws formulated in these quantities designated to describe all phenomena belonging to the realm of experiences to which the theory refers. There will always exist various possibilities of selecting suitable sets of quantities which may be understood as being of basic importance.[1]

We have stressed the considerations brought forward by Milankovitch (1939, 1941) during his long and laborious investigations of the astronomical theory of climate change; with regard to physical insight and from a perturbations point of view, there seems to be little debate over the advantages of the use of the Milankovitch vectorial elements in formulating the equations of motion.[2] They have a deeper connection with the dynamical problem than the classical elements, which characterize only the simple geometry of conic sections. The angular momentum and eccentricity vectors, the invariants which emerge fundamentally from the integrals of the Kepler problem, have an essential dynamical importance, not just in celestial mechanics, but from the general standpoint of modern physics.

[1] Finlay-Freundlich E. op. cit., p. 53.

[2] The fact that these elements have not been widely used in celestial mechanics can perhaps best be explained by the argument that tradition is stronger than the desire for simplicity.

We presented two new sets of Milankovitch-like vectorial elements, examined the choice and merits of the corresponding scalar element, and derived the perturbation equations for these sets in both Lagrangian and Gaussian form. The application of this vectorial formulation in secular perturbation theory to the determination of gravitational and non-gravitational effects leads to equations for perturbed elements in which the small numerical divisors are not present. These secular equations are several hundred times faster to numerically integrate than the non-averaged counterpart, and provides a very accurate description of the long-term orbit behavior. Comparison with observations is the ultimate test of any theory or model; in this work, the comparison has been supplemented with numerical integration. It should be noted that the comparison between numerical integrations and the averaged equations will, in general, show a divergence between the true and secular solutions. In some cases, this deviation is due to higher-order effects,[3] but often it results from an inconsistent choice of initial conditions (Scheeres 2012b). Future work will focus on obtaining the short-period correction terms for each perturbing force; while the mathematical developments here are straightforward, they can be algebraically quite complicated.

Understanding the long period effects in the motions of celestial bodies is the central problem in the determination of orbital stability over long timespans. In treating the problem of stability, two types of long period perturbations are of primary importance: secular perturbations and mean motion resonances. The tesseral harmonics in the Earth's gravitational potential can introduce long-term effects, especially if the mean motion of the object is commensurable with the angular velocity of Earth's rotation (Musen & Bailie 1962; Allan 1967). The most interesting case is the influence of the ellipticity of Earth's equator on the motion of a geostationary object. The averaged effect of this perturbation must be treated using a resonance theory and will be studied in future work.

For gravitationally dominated orbits, solar radiation pressure is negligible and the resulting motion is largely governed by the oblateness of the primary and the attraction of the Sun. The interplay between these gravitational perturbations gives rise to three mutually perpendicular planes

[3] The extension to second-order perturbations is easy in principle, but the algebra becomes much heavier.

of equilibrium for circular satellite orbits. We have investigated the effect of solar radiation pressure on these equilibrium planes and have studied the properties of the modified Laplace equilibria. The modified Laplace surfaces, swept out by these equilibria as the semi-major axis varies, exhibit qualitatively different behavior than those obtained without considering the radiation pressure. These results provide useful insight into the Earth-orbiting debris problem, in particular the dynamical behavior of the high area-to-mass ratio space debris, as well as into the fundamental aspects of the evolution of dust particles in circumplanetary orbits.

A long-standing enigma of Saturn's satellite Iapetus is its striking albedo dichotomy, the origin and nature of which represents one of the outstanding problems of planetary science. First observed by Cassini in 1677, Iapetus is much darker on its leading hemisphere (which faces forward in its orbit about Saturn) than its opposite trailing hemisphere; the reflectance varies across its surface by a factor of 10 to 20—the largest brightness asymmetry known in the Solar System (Wilson & Sagan 1996). Numerous Earth-based investigations and several flybys by the Voyager and Cassini spacecrafts have been conducted to dispel some of the controversy surrounding the origin of the dark side of Iapetus (Spencer & Denk 2010). The modified Laplace surfaces, together with an adiabatic invariance argument on Poynting-Robertson drag, provide a natural qualitative explanation for the initial albedo dichotomy of this prodigious and bewildering moon.

It would now remain for us to apply the preceding theory to the different planets; but given the already excessive length of this memoir, we shall follow in the footsteps of Laplace[4] and put off these applications to another time.

[4] Having found a new method for integrating differential equations by approximation—a method which allowed for the determination of all perturbations, both secular and periodic—Laplace concluded: "Il nous resterait présentement à appliquer la théorie précédente aux différentes planètes; mais la longueur déjà trop grande de ce Mémoire m'oblige de renvoyer ces applications à un autre temps ..." (Laplace PS. 1772/76. Recherches sur le calcul intégral et sur le système du monde, Mémoires de l'Académie Royale des Sciences de Paris, p. 464).

Bibliography

Allan RR. 1962. Satellite orbit perturbations due to radiation pressure and luni-solar forces, *Quarterly Journal of Mechanics and Applied Mathematics* 15:283–301

Allan RR. 1967. Resonance effects due to the longitude dependence of the gravitational field of a rotating primary, *Planetary and Space Science* 15:53–76

Allan RR, Cook GE. 1964. The long-period motion of the plane of a distant circular orbit, *Proceedings of the Royal Society A* 280:97–109

Allan RR, Cook GE. 1967. Discussion of paper by S. J. Peale, 'dust belt of the Earth', *Journal of Geophysical Research* 72:1124–1127

Allan RR, Ward GN. 1963. Planetary equations in terms of vectorial elements, *Mathematical Proceedings of the Cambridge Philosophical Society* 59:669–677

Anselmo L, Pardini C. 2010. Long-term dynamical evolution of high area-to-mass ratio debris released into high Earth orbits, *Acta Astronautica* 67:204–216

Armitage PJ, Pringle JP. 1997. Radiation-induced warping of protostellar accretion disks, *The Astrophysical Journal* 488:L47–L50

Arnold VI, Kozlov VV, Neishtadt AI. 2006. *Mathematical Aspects of Classical and Celestial Mechanics*. Berlin: Springer-Verlag, 3rd Edn.

Bartnik EA, Haberzettl H, Sandhas W. 1988. Equations of motion using the dynamical evolution of the Runge-Lenz vector, *The Astrophysical Journal* 334:517–526

Bartoli A. 1884. Il calorico raggiante e il secondo principio di termodinamica, *Il Nuovo Cimento* 15:193–202

Battin RH. 1999. *An Introduction to the Mathematics and Methods of Astrodynamics*. Reston: American Institute of Aeronautics and Astronautics, Rev. Edn.

Bellomo P, Stroud Jr. CR, Farrelly D, Uzer T. 1998. Quantum-classical correspondence in the hydrogen atom in weak external fields, *Physical Review A* 58:3896–3913

Berger A. 1988. Milankovitch theory and climate, *Reviews of Geophysics* 26:624–657

Bilimovitch A. 1943. Über die anwendungen der Pfaffschen methode in der störungstheorie, *Astronomische Nachrichten* 273:161–178

Blaes O, Lee MH, Socrates A. 2002. The Kozai mechanism and the evolution of binary supermassive black holes, *The Astrophysical Journal* 578:775–786

Boccaletti D, Pucacco G. 1999. *Theory of Orbits*, Vol. 2: Perturbative and Geometrical Methods. Berlin: Springer-Verlag. Corrected Second Printing 2002

Bogoliubov NN, Mitropolsky YA. 1961. *Asymptotic Methods in the Theory of Non-Linear Oscillations*. New York: Gordon and Breach

Bohr N. 1918. On the quantum theory of line-spectra. Part II. On the hydrogen spectrum, *Kongelige Danske Videnskabernes Selskabs Skrifter* 8:37–100

Born M. 1927. *The Mechanics of the Atom*. London: G. Bell and Sons

Bottke WF, Nesvorný D, Vokrouhlický D, Morbidelli A. 2010. The irregular satellites: The most collisionally evolved populations in the solar system, *The Astronomical Journal* 139:994–1014

Boué G, Laskar J. 2006. Precession of a planet with a satellite, *Icarus* 185:312–330

Boué G, Laskar J. 2009. Spin axis evolution of two interacting bodies, *Icarus* 201:750–767

Brand L. 1947. *Vector and Tensor Analysis*. New York: John Wiley & Sons

Breiter S, Fouchard M, Ratajczak R. 2008. Stationary orbits of comets perturbed by galactic tides, *Monthly Notices of the Royal Astronomical Society* 383:200–208

Breiter S, Ratajczak R. 2005. Vectorial elements for the galactic disc tide effects in cometary motion, *Monthly Notices of the Royal Astronomical Society* 364:1222–1228

Broucke R, Lass H, Ananda M. 1971. Redundant variables in celestial mechanics, *Astronomy & Astrophysics* 13:390–398

Broucke RA, Cefola PJ. 1972. On the equinoctial orbit elements, *Celestial Mechanics* 5:303–310

Brouwer D. 1959. Solution of the problem of artificial satellite theory without drag, *The Astronomical Journal* 64:378–397

Brouwer D, Clemence GM. 1961a. *Methods of Celestial Mechanics*. London: Academic Press

Brouwer D, Clemence GM. 1961b. Orbits and masses of planets and satellites, In *Planets and Satellites*, eds. GP Kuiper, BM Middlehurst. Chicago: The University of Chicago Press, 31–94

Burns JA. 1976. Elementary derivation of the perturbation equations of celestial mechanics, *American Journal of Physics* 44:944–949

Burns JA. 1986. The evolution of satellite orbits, In *Satellites*, eds. JA Burns, MS Matthews. Tucson: The University of Arizona Press, 117–158

Burns JA, Hamill P, Cuzzi JN, Durisen RH. 1979. On the "thickness" of Saturn's rings caused by satellite and solar perturbations and by planetary precession, *The Astronomical Journal* 84:1783–1801

Burns JA, Lamy PL, Soter S. 1979. Radiation forces on small particles in the solar system, *Icarus* 40:1–48

Cazenave A, Dobrovolskis A, Lago B. 1980. Orbital history of the Martian satellites with inferences on their origin, *Icarus* 44:730–744

Chao CC. 2006. In *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference*, Keystone, Colorado, Paper AIAA-2006-6514

Cook GE. 1962. Luni-solar perturbations of the orbit of an Earth satellite, *The Geophysical Journal of the Royal Astronomical Society* 6:271–291

Cordani B. 2003. *The Kepler Problem: Group Theoretical Aspects, Regularization and Quantization, with Application to the Study of Perturbations.* Basel: Birkhäuser

Correia ACM, Laskar J, Farago F, Boué G. 2011. Tidal evolution of hierarchical and inclined systems, *Celestial Mechanics and Dynamical Astronomy* 111:105–130

Deprit A. 1969. Canonical transformations depending on a small parameter, *Celestial Mechanics* 1:12–30

Deprit A. 1975. Ideal elements for perturbed keplerian motions, *Journal of Research of the National Bureau of Standards B* 79:1–15

Dimitrijević MS. 2002. Milutin Milanković (1879 - 1958) and his contribution to European astronomy, *Astronomische Nachrichten* 323:570–573

Dobrovolskis AR. 1980. Where are the rings of Neptune?, *Icarus* 43:222–226

Dobrovolskis AR. 1993. The Laplace planes of Uranus and Pluto, *Icarus* 105:400–407

Dobrovolskis AR, Borderies NJ, Steiman-Cameron TY. 1989. Stability of polar rings around Neptune, *Icarus* 81:132–144

Dohan K, Maximenko N. 2010. Monitoring ocean currents with satellite sensors, *Oceanography* 23:94–103

Dziobek O. 1892. *Mathematical Theories of Planetary Motions.* Ann Arbor: Register Publishing Company. Reprint 1962. New York: Dover Publications

Eddington AS. 1916/17. On the radiative equilibrium of the stars, *Monthly Notices of the Royal Astronomical Society* 77:16–35, 596–612

Fabrycky D, Tremaine S. 2007. Shrinking binary and planetary orbits by Kozai cycles with tidal friction, *The Astrophysical Journal* 669:1298–1315

Farago F, Laskar J. 2010. High-inclination orbits in the secular quadrupolar three-body problem, *Monthly Notices of the Royal Astronomical Society* 401:1189–1198

Ferraz-Mello S. 2007. *Canonical Perturbation Theories: Degenerate Systems and Resonance.* New York: Springer

Finlay-Freundlich E. 1958. *Celestial Mechanics.* New York: Pergamon Press

Fleckenstein JO. 1952. Les théorèmes de Laplace sur les perturbations séculaires dans les éléments vectoriels des orbites planétaires, *Experientia* 8:136–137

Folkner WM, Williams JG, Boggs DH. 2009. The planetary and lunar ephemeris DE 421. JPL Interplanetary Network Progress Report 42-178

Friesen LJ, IV AAJ, Zook HA, Kessler DJ. 1992. Analysis of orbital perturbations acting on objects in orbits near geosynchronous Earth orbit, *Journal of Geophysical Research* 97:3845–3863

Friesen LJ, Kessler DJ, Zook HA. 1993. Reduced debris hazard resulting from a stable inclined geosynchronous orbit, *Advances in Space Research* 13:231–241

Früh C, Kelecy TM, Jah MK. 2013. Coupled orbit-attitude dynamics of high area-to-mass ratio (HAMR) objects: influence of solar radiation pressure, Earth's shadow and the visibility in light curves, *Celestial Mechanics and Dynamical Astronomy* 117:385–404

Früh C, Schildknecht T. 2012. Variation of the area-to-mass ratio of high area-to-mass ratio space debris objects, *Monthly Notices of the Royal Astronomical Society* 419:3521–3528

Garfinkel B. 1959. The orbit of a satellite of an oblate planet, *The Astronomical Journal* 64:353–367

Gibbs JW, Wilson EB. 1909. *Vector Analysis*. New York: Charles Scribner's Sons. Reprint 1960. New York: Dover Publications

Goldreich P. 1965. Inclination of satellite orbits about an oblate precessing planet, *The Astronomical Journal* 70:5–9

Goldreich P. 1966. History of the lunar orbit, *Reviews of Geophysics* 4:411–439

Goldstein H. 1950. *Classical Mechanics*. Cambridge: Addison-Wesley Publishing Co.

Goldstein H. 1976. More on the prehistory of the Laplace or Runge-Lenz vector, *American Journal of Physics* 44:1123–1124

Graf Jr. OF. 1975. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference*, Nassau, Bahamas, Paper AAS 75-023

Greenberg R. 1982. Orbital interactions: A new geometrical formalism, *The Astronomical Journal* 87:184–195

Gustafson BÅS. 1994. Physics of zodiacal dust, *Annual Review of Earth and Planetary Sciences* 22:553–595

Hall A. 1885. The orbit of Iapetus, the outer satellite of Saturn, In *Washington Observations for 1882*. Washington: Government Printing Office, 5–82

Hamilton DP. 1996. The asymmetric time-variable rings of Mars, *Icarus* 119:153–172

Harper D. 1987. *The Dynamics of the Outer Satellites of Saturn*. Ph.D. thesis, University of Liverpool

Hechler M, Van der Ha JC. 1981. Probability of collisions in the geostationary ring, *Journal of Spacecraft and Rockets* 18:361–366

Herget P. 1962. On the variation of arbitrary vectorial constants, *The Astronomical Journal* 67:16–18

Herrick S. 1948. A modification of the "variation-of-constants" method for special perturbations, *Publications of the Astronomical Society of the Pacific* 60:321–323

Hestenes D. 1983. Celestial mechanics with geometric algebra, *Celestial Mechanics* 30:151–170

Hestenes D. 1999. *New Foundations for Classical Mechanics*. New York: Kluwer Academic Publishers, 2nd Edn.

Hill GW. 1882. On Gauss's method of computing secular perturbations, with an application to the action of Venus on Mercury, *Astronomical Papers of the American Ephemeris* 1:315–361

Hori G. 1966. Theory of general perturbations with unspecified canonical variables, *Publications of the Astronomical Society of Japan* 18:287–296

Hubaux C, Libert AS, Delsate N, Carletti T. 2013. Influence of Earth's shadowing effects on space debris stability, *Advances in Space Research* 51:25–38

Hughes S. 1981. The computation of tables of Hansen coefficients, *Celestial Mechanics* 29:101–107

Inter-Agency Space Debris Coordination Committee. 2002. IADC space debris mitigation guidelines. IACD-02-01; revised 2007

International Telecommunication Union. 2010. Environmental protection of the geostationary-satellite orbit. ITU-R S.1003-2

Jeffreys H. 1947. The effects of collisions on Saturn's rings, *Monthly Notices of the Royal Astronomical Society* 107:263–267

Johnson NL. 1982. The development and deployment of Soviet geosynchronous satellites, *Journal of the British Interplanetary Society* 35:450–458

Johnson NL. 2012. A new look at the GEO and near-GEO regimes: Operations, disposals, and debris, *Acta Astronautica* 80:82–88

Jones RV. 1953. Pressure of radiation, *Nature* 171:1089–1093

Katz B, Dong S, Malhotra R. 2011. Long-term cycling of Kozai-Lidov cycles: Extreme eccentricities and inclinations excited by a distant eccentric perturber, *Physical Review Letters* 107:181101

Kaula WM. 1962. Development of the lunar and solar disturbing functions for a close satellite, *The Astronomical Journal* 67:300–303

Kelecy T, Jah M. 2011. Analysis of high area-to-mass ratio (HAMR) GEO space object orbit determination and prediction performance: Iinitial strategies to recover and predict HAMR GEO trajectories with no a priori information, *Acta Astronautica* 69:551–558

Kelly TJ. 1989. A note on first-order normalizations of perturbed Keplerian systems, *Celestial Mechanics and Dynamical Astronomy* 46:19–25

Kessler DJ, Cour-Palais BG. 1978. Collision frequency of artificial satellites: The creation of a debris belt, *Journal of Geophysical Research* 83:2637–2646

Klein O. 1924. Über die gleichzeitige wirkung von gekreuzten homogenen elektrischen und magnetischen feldern auf das wasserstoffatom. i., *Zeitschrift für Physik* 22:109–118

Kovalevsky J. 1963. *Introduction à la Mécanique Céleste*. Paris: Librairie Armand Colin. Reprint 1967. Introduction to Celestial Mechanics. Dordrecht: D. Reidel Publishing Co.

Kozai Y. 1959. The motion of a close Earth satellite, *The Astronomical Journal* 64:367–377

Kozai Y. 1962a. Mean values of cosine functions in elliptic motion, *The Astronomical Journal* 67:311–312

Kozai Y. 1962b. Secular perturbations of asteroids with high inclination and eccentricity, *The Astronomical Journal* 67:591–598

Kudielka V. 1994. Balanced Earth satellite orbits, *Celestial Mechanics and Dynamical Astronomy* 60:455–470

Kudielka VW. 1997. Equilibria bifurcations of satellite orbits, In *The Dynamical Behaviour of our Planetary System*, eds. R Dvorak, J Henrard. Dordrecht: Kluwer Academic Publishers, 243–255

Kurth R. 1959. *Introduction to the Mechanics of the Solar System*. New York: Pergamon Press

Lantoine G, Russell RR. 2011. Complete closed-form solutions of the Stark problem, *Celestial Mechanics and Dynamical Astronomy* 109:333–366

Laplace PS. 1825. *Traité de Mécanique Céleste*. 5 Vols. Paris: Courcier. (Vol. 1: 1799; Vol. 2: 1799; Vol. 3: 1802; Vol. 4: 1805; Vol. 5: 1825)

Law KL, Morét-Ferguson S, Maximenko NA, Proskurowski G, Peacock EE, et al. 2010. Plastic accumulation in the north atlantic subtropical gyre, *Science* 329:1185–1188

Lebedew P. 1902. The physical causes of the deviations from Newton's law of gravitation, *The Astrophysical Journal* 16:155–161

Leiva AM, Briozzo CB. 2013. Low-energy impact distribution and the albedo dichotomy of Iapetus, *Monthly Notices of the Royal Astronomical Society* 430:858–868

Lemaître A, Delsate N, Valk S. 2009. A web of secondary resonances for large A/m geostationary debris, *Celestial Mechanics and Dynamical Astronomy* 104:383–402

Lenz W. 1924. Über den bewegungsverlauf und die quantenzustände der gestörten Keplerbewegung, *Zeitschrift für Physik* 24:197–207

Lidov ML. 1962. The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies, *Planetary and Space Science* 9:719–759

Lidov ML, Yarskaya MV. 1974. Integrable cases in the problem of the evolution of a satellite orbit under the joint effect of an outside body and of the noncentrality of the planetary field, *Cosmic Research* 12:139–152

Liou JC, Johnson NL. 2006. Risks in space from orbiting debris, *Science* 311:340–341

Liou JC, Weaver JK. 2005. In *Proceedings of the Fourth European Conference on Space Debris*, Darmstadt, Germany, Paper ESA SP-587

Maxwell JC. 1873. *A Treatise on Electricity and Magnetism*, Vol. 2. Oxford: Clarendon Press

McInnes CR. 1999. *Solar Sailing. Technology, Dynamics and Mission Applications.* London: Springer

McKinnon WB. 1999. Midsize icy satellites, In *The New Solar System*, eds. JK Beatty, CC Peterson, A Chaikin. Cambridge: Sky Publishing Corporation, 4th Edn., 297–310

McMahon J, Scheeres D. 2010a. Detailed prediction for the BYORP effect on binary near-Earth asteroid (66391) 1999 KW4 and implications for the binary population, *Icarus* 209:494–509

McMahon J, Scheeres D. 2010b. Secular orbit variation due to solar radiation effects: a detailed model for byorp, *Celestial Mechanics and Dynamical Astronomy* 106:261–300

McMahon JW. 2011. *An Analytical Theory for the Perturbative Effect of Solar Radiation Pressure on Natural and Artificial Satellites.* Ph.D. thesis, University of Colorado

Mehra J, Rechenberg H. 2001. *The Historical Development of Quantum Theory.* 6 Vols. New York: Springer

Mignard F. 1981. Evolution of the Martian satellites, *Monthly Notices of the Royal Astronomical Society* 194:365–379

Mignard F, Hénon M. 1984. About an unsuspected integrable problem, *Celestial Mechanics* 33:239–250

Milankovitch M. 1939. Über die verwendung vektorieller bahnelemente in der störungsrechnung, *Bulletin de l'Académie des Sciences Mathématiques et Naturelles A* 6:1–70

Milankovitch M. 1941. *Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem.* Belgrade: Königlich Serbische Akademie. Reprint 1969. Canon of Insolation and the Ice-Age Problem. Jerusalem: Israel Program for Scientific Translations

Moser J. 1970. Regularization of Kepler's problem and the averaging method on a manifold, *Communications on Pure and Applied Mathematics* 23:609–636

Mosqueira I, Estrada PR, Charnoz S. 2010. Deciphering the origin of the regular satellites of gaseous giants — Iapetus: The Rosetta ice-moon, *Icarus* 207:448–460

Moulton FR. 1914. *An Introduction to Celestial Mechanics.* New York: Macmillan, 2nd Edn. Reprint 1970. New York: Dover Publications

Murray CD, Dermott SF. 1999. *Solar System Dynamics.* Cambridge: Cambridge University Press

Musen P. 1947. Über die vektoriell-skalaren gleichungen der astronomischen störungstheorie, *Zeitschrift für Naturforschung A* 2:365–369

Musen P. 1954. Special perturbations of the vectorial elements, *The Astronomical Journal* 59:262–267

Musen P. 1960. The influence of the solar radiation pressure on the motion of an artificial satellite, *Journal of Geophysical Research* 65:1391–1396

Musen P. 1961. On the long-period lunar and solar effects on the motion of an artificial satellite, 2, *Journal of Geophysical Research* 66:2797–2805

Musen P. 1964. On the application of Pfaff's method in the theory of variation of astronomical constants. NASA Technical Note D-2301

Musen P, Bailie AE. 1962. On the motion of a 24-hour satellite, *Journal of Geophysical Research* 67:1123–1132

Namouni F. 2005. On the origin of the eccentricities of extrasolar planets, *The Astronomical Journal* 130:280–294

Naoz S, Farr WM, Lithwick Y, Rassio FA, Teyssandier J. 2013. Secular dynamics in hierarchical three-body systems, *Monthly Notices of the Royal Astronomical Society* 431:2155–2171

NASA Orbital Debris Program Office. 2012. Two derelict NOAA satellites experience anomalous events, In *Orbital Debris Quarterly News*, eds. JC Liou, D Shoots. Houston: NASA Johnson Space Center, 16(1):1–2

Newton RR. 1961. Variables that are determinate for any orbit, *American Rocket Society Journal* 31:364–366

Nichols EF, Hull GF. 1903. The pressure due to radiation, *The Astrophysical Journal* 17:315–351

Oudemans JAC. 1888. On the retrogradation of the plane of Saturn's ring and of those of his satellites whose orbits coincide with that plane, *Monthly Notices of the Royal Astronomical Society* 49:54–64

Pauli Jr. W. 1926. Über das wasserstoffspektrum vom standpunkt der neuen quantenmechanik, *Zeitschrift für Physik* 36:336–363. English translation 1967. In *Sources of Quantum Mechanics*, ed. B. L. van der Waerden, Amsterdam: North-Holland, 387–416

Perozzi E, Roy AE, Steves BA, Valsecchi GB. 1991. Significant high number of commensurabilities in the main lunar problem. I: The Saros as a near-periodicity of the Moon's orbit, *Celestial Mechanics and Dynamical Astronomy* 52:241–261

Petro AJ. 1992. Techniques for orbital debris control, *Journal of Spacecraft and Rockets* 29:260–263

Petterson JA. 1977. Twisted accretion disks. II. Applications to x-ray binary systems, *The Astrophysical Journal* 216:827–837

Plummer HC. 1905. On the possible effects of radiation on the motion of comets, with special reference to Encke's comet, *Monthly Notices of the Royal Astronomical Society* 65:229–237

Plummer HC. 1918. *An Introductory Treatise on Dynamical Astronomy.* Cambridge: Cambridge University Press. Reprint 1960. New York: Dover Publications

Poincaré H. 1899. *Les Méthodes Nouvelles de la Mécanique Céleste.* 3 Vols. Paris: Gauthier-Villars. (Vol. 1: 1892; Vol. 2: 1893; Vol. 3: 1899) Reprint 1993. New Methods of Celestial Mechanics, ed. DL Goroff. New York: American Institute of Physics

Poisson SD. 1809. Mémoire sur la variation des constantes arbitraires dans les questions de mécanique, *Journal de l'école polytechnique* 8:266–344

Pollard H. 1966. *Mathematical Introduction to Celestial Mechanics.* Englewood Cliffs: Prentice-Hall

Popović B. 1950. New forms of the equations of perturbation in planetary motion (in Serbian), *Glas Srpske Akademije Nauka i Umetnosti* 198:129–139

Poynting JH. 1904. Radiation in the solar system: Its effect on temperature and its pressure on small bodies, *Philosophical Transactions of The Royal Society of London A* 202:525–552

Pringle JE. 1996. Self-induced warping of accretion discs, *Monthly Notices of the Royal Astronomical Society* 281:357–361

Redmond PJ. 1964. Generalization of the Runge-Lenz vector in the presence of an electric field, *Physical Review B* 133:1352–1353

Richter K, Keller HU. 1995. On the stability of dust particle orbits around cometary nuclei, *Icarus* 114:355–371

Robertson HP. 1937. Dynamical effects of radiation in the solar system, *Monthly Notices of the Royal Astronomical Society* 97:423–438

Rosengren AJ, Scheeres DJ. 2013. Long-term dynamics of high area-to-mass ratio objects in high-Earth orbit, *Advances in Space Research* 52:1545–1560

Rosengren AJ, Scheeres DJ. 2014a. Laplace plane modifications arising from solar radiation pressure, *The Astrophysical Journal* In Press

Rosengren AJ, Scheeres DJ. 2014b. On the Milankovitch orbital elements for perturbed Keplerian motion, *Celestial Mechanics and Dynamical Astronomy* 118:197–220

Rosengren AJ, Scheeres DJ, McMahon JW. 2014. The classical Laplace plane as a stable disposal orbit for geostationary satellites, *Advances in Space Research* 53:1219–1228

Rossi A, Valsecchi GB. 2006. Collision risk against space debris in Earth orbits, *Celestial Mechanics and Dynamical Astronomy* 95:345–356

Roy AE. 2005. *Orbital Motion*. New York: Taylor & Francis, 4th Edn.

Roy AE, Moran PE. 1973. Studies in the application of recurrence relations to special perturbation methods III. Non-singular differential equations for special perturbations, *Celestial Mechanics* 7:236–255

Rubincam DP. 2000. Radiative spin-up and spin-down of small asteroids, *Icarus* 148:2–11

Runge C. 1919. *Vektoranalysis*, Vol. 1. Leipzig: Hirzel

Sanders JA, Verhulst F, Murdock J. 2007. *Averaging Methods in Nonlinear Dynamical Systems*. New York: Springer, 2nd Edn.

Schaub H, Junkins JL. 2009. *Analytical Mechanics of Space Systems*. Reston: American Institute of Aeronautics and Astronautics, 2nd Edn.

Scheeres DJ. 2012a. Orbit mechanics about asteroids and comets, *Journal of Guidance, Control, and Dynamics* 35:987–997

Scheeres DJ. 2012b. *Orbital Motion in Strongly Perturbed Environments: Applications to Asteroid, Comet and Planetary Satellite Orbiters*. Berlin: Springer-Praxis

Schildknecht T. 2007. Optical surveys for space debris, *Annual Review of Astronomy and Astrophysics* 14:41–111

Schildknecht T, Musci R, Ploner M, Beutler G, Flury W, et al. 2004. Optical observations of space debris in GEO and in highly-eccentric orbits, *Advances in Space Research* 34:901–911

Schleif CR, Delos JB. 2008. Semiclassical theory of the structure of the hydrogen spectrum in near-perpendicular electric and magnetic fields: Derivations and formulas for Einstein-Brillouin-Keller-Maslov quantization and description of monodromy, *Physical Review A* 77:043422

Sekiguchi N. 1961. On the libration of the orbital plane of a stationary artificial satellite with a circular orbit, *Publications of the Astronomical Society of Japan* 13:207–211

Shapiro II. 1963. The prediction of satellite orbits, In *Dynamics of Satellites*, ed. M Roy. New York: Academic Press, 257–312

Showalter MR, Hamilton DP, Nicholson PD. 2006. A deep search for Martian dust rings and inner moons using the Hubble space telescope, *Planetary and Space Science* 54:844–854

Soter S. 1974. The brightness asymmetry of Iapetus, Presented at the *International Astronomical Union Colloquium 28*, Ithaca, NY

Spencer JR, Denk T. 2010. Formation of Iapetus' extreme albedo dichotomy by exogenically triggered thermal ice migration, *Science* 327:432–435

Stahlhofen AA. 1994. Once more the perturbed Kepler problem, *American Journal of Physics* 62:1145–1147

Stiefel EL, Scheifele G. 1971. *Linear and Regular Celestial Mechanics.* Berlin: Springer-Verlag

Strömgren B. 1929. Formeln zur genäherten störungsrechnung in bahnelementen, *Publikationer og mindre Meddelelser fra Københavns Observatorium* 65:1–27

Struble RA. 1961. An application of the method of averaging in the theory of satellite motion, *Journal of Mathematics and Mechanics* 10:691–704

Tamayo D, Burns JA, Hamilton DP. 2013. Chaotic dust dynamics and implications for the hemispherical color asymmetries of the Uradian satellites, *Icarus* 226:655–662

Tamayo D, Burns JA, Hamilton DP, Hedman MM. 2011. Finding the trigger to Iapetus' odd global albedo pattern: Dynamics of dust from Saturn's irregular satellites, *Icarus* 215:260–278

Tamayo D, Burns JA, Hamilton DP, Nicholson PD. 2013. Dynamical instabilities in high-obliquity systems, *The Astronomical Journal* 145:54

Tamayo D, Hedman MM, Burns JA. 2014. First observations of the Pphoebe ring in optical light, *Icarus* http://dx.doi.org/10.1016/j.icarus.2014.01.021

Taton R, Wilson C, eds. 1995. *The General History of Astronomy*, Vol. 2: Planetary astronomy from the Renaissance to the rise of Astrophysics. Part B: The eighteenth and nineteenth centuries. Cambridge: Cambridge University Press

Tisserand F. 1896. *Traité de Mécanique Céleste*. 4 Vols. Paris: Gauthier-Villars. (Vol. 1: 1889; Vol. 2: 1891; Vol. 3: 1894; Vol. 4: 1896)

Touma JR, Tremaine S, Kazandjian MV. 2009. Gauss's method for secular dynamics, softened, *Monthly Notices of the Royal Astronomical Society* 394:1085–1108

Tremaine S, Davis SW. 2013. Dynamics of warped accretion discs. ArXiv:1308.1964

Tremaine S, Touma J, Namouni F. 2009. Satellite dynamics on the Laplace surface, *The Astronomical Journal* 137:3706–3717

Tremaine S, Yavetz T. 2013. Why do Earth satellites stay up? ArXiv:1309.5244

Ulivieri C, Circi C, Ortore E, Bunkheila F, Todino F. 2013. Frozen orbital plane solutions for satellites in nearly circular orbit, *Journal of Guidance, Control, and Dynamics* 36:935–945

Valk S, Lemaître A, Anselmo L. 2008. Analytical and semi-analytical investigations of geosynchronous space debris with high area-to-mass ratios, *Advances in Space Research* 41:1077–1090

Vashkov'yak MA. 1974. Stability of circular satellite orbits for combined action of perturbations from an external body and from the noncentrality of the planetary gravitational field, *Cosmic Research* 12:757–769

Vashkov'yak MA. 1999. Evolution of the orbits of distant satellites of uranus, *Astronomy Letters* 25:476–481

Verbiscer AJ, Skrutskie MF, Hamilton DP. 2009. Saturn's largest ring, *Nature* 461:1098–1100

Volk O. 1976. Miscellanea from the history of celestial mechanics, *Celestial Mechanics* 14:365–382

Ward GN. 1962. On the secular variations of the elements of satellite orbits, *Proceedings of the Royal Society of London A* 266:130–142

Ward WR. 1981. Orbital inclination of Iapetus and the rotation of the Laplacian plane, *Icarus* 46:97–107

Weinberg S. 1972. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. New York: John Wiley & Sons

Wen LS. 1962. A unified treatment of "variation of parameters" and "differential expressions" methods in trajectory prediction and error analysis, *Journal of the Aerospace Sciences* 29:61–66

Whittaker ET. 1937. *Analytical Dynamics*. Cambridge: Cambridge University Press, 4th Edn.

Wilson PD, Sagan C. 1996. Spectrophotometry and organic matter on Iapetus: 2. Models of interhemispheric asymmetry, *Icarus* 122:92–106

Wyatt Jr. SP, Whipple FL. 1950. The Poynting-Robertson effect on meteor orbits, *The Astrophysical Journal* 111:134–141

Zee CH. 1989. *Theory of Geostationary Orbits*. Dordrecht: Kluwer Academic Publishers

Appendix A

Dyadic Notation and Operations

> *A branch of mathematics of especial value to theoretical physics is vector analysis, which has been developed largely by physicists. Its concepts have immediate perceptual meaning, so that many physical laws show their full significance only when stated in the language of vectors. It is therefore quite a mistaken view to regard vector analysis merely as a computational shorthand.*[1]
>
> — Georg Joos, 1958

In this treatise, we adopt a more general perspective that is independent of any particular system of reference by adhering to vectors and, their higher-dimensional counterparts, dyadics (or tensors of rank or valence two). Since the now-quaint topic of dyadics is likely to be unfamiliar to most readers, we present here a synopsis of basic concepts, properties, and notations,[2] following the exposition of Gibbs & Wilson (1909) and Brand (1947), to which we refer for the omitted details.

An expression $\boldsymbol{ab}$, formed by the juxtaposition of two vectors in $\boldsymbol{R}^3$ and written without a dot or cross between, constitutes what is called a dyad, or *indeterminate product*. The first vector of a dyad is called its *antecedent* and the second its *consequent*. Note that while $\boldsymbol{a}\cdot\boldsymbol{b}$ is a scalar and $\boldsymbol{a}\times\boldsymbol{b}$ is a vector, the dyad $\boldsymbol{ab}$ represents a new mathematical entity. A dyadic $\boldsymbol{\Phi}$ is defined as a linear polynomial of dyads, such as $\boldsymbol{\Phi} = \boldsymbol{a}_i\boldsymbol{b}_i = \boldsymbol{a}_1\boldsymbol{b}_1 + \boldsymbol{a}_2\boldsymbol{b}_2 + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n$, and is used to treat linear vector functions. The dyadic $\boldsymbol{\Phi}_c = \boldsymbol{b}_i\boldsymbol{a}_i = \boldsymbol{b}_1\boldsymbol{a}_1 + \boldsymbol{b}_2\boldsymbol{a}_2 + \cdots + \boldsymbol{b}_n\boldsymbol{a}_n$, obtained by interchanging the

[1] Joos G. *Theoretical Physics*. London: Blackie & Son, 3rd Edn., p. 6

[2] See Brand (1947, Chap. 4, Linear Vector Functions) for a systematic development of dyadic algebra. A detailed, and more encyclopedic, exposition is given by Gibbs & Wilson (1909, Chap. 5), and a brief summary with applications to celestial mechanics can be found in Scheeres (2012b, §1.5) and to rigid body dynamics in Goldstein (1950, §5.3). The elements of the subject will be sufficient for our purpose.

antecedent and consequent in each term of $\boldsymbol{\Phi}$, is called the conjugate of $\boldsymbol{\Phi}$. If a dyadic is equal to its conjugate as $\boldsymbol{\Phi} = \boldsymbol{\Phi}_c$, it is called self-conjugate, or symmetric. A dyadic equal to the negative of its conjugate ($\boldsymbol{\Phi} = -\boldsymbol{\Phi}_c$) is called anti-self-conjugate, or antisymmetric (skew-symmetric).

The direct, or scalar, products of a dyadic and a vector $\boldsymbol{r}$ are denoted by $\boldsymbol{\Phi} \cdot \boldsymbol{r}$ and $\boldsymbol{r} \cdot \boldsymbol{\Phi}$, and are defined as follows:

$$\boldsymbol{\Phi} \cdot \boldsymbol{r} = \boldsymbol{a}_1\boldsymbol{b}_1 \cdot \boldsymbol{r} + \boldsymbol{a}_2\boldsymbol{b}_2 \cdot \boldsymbol{r} + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n \cdot \boldsymbol{r}, \tag{A.1}$$

$$\boldsymbol{r} \cdot \boldsymbol{\Phi} = \boldsymbol{r} \cdot \boldsymbol{a}_1\boldsymbol{b}_1 + \boldsymbol{r} \cdot \boldsymbol{a}_2\boldsymbol{b}_2 + \cdots + \boldsymbol{r} \cdot \boldsymbol{a}_n\boldsymbol{b}_n. \tag{A.2}$$

In the first of these equations, the consequents of the dyadic $\boldsymbol{\Phi}$ are multiplied into $\boldsymbol{r}$ by direct or scalar multiplication, and $\boldsymbol{\Phi}$ is said to act as a *prefactor*; in the second, $\boldsymbol{r}$ is multiplied into the antecedents of $\boldsymbol{\Phi}$, and $\boldsymbol{\Phi}$ is said to act as a *postfactor*. It should be noted that a dyadic $\boldsymbol{\Phi}$ acting as a prefactor or postfactor upon a vector $\boldsymbol{r}$ determines a linear vector function of $\boldsymbol{r}$; evidently the vectors $\boldsymbol{\Phi} \cdot \boldsymbol{r}$ and $\boldsymbol{r} \cdot \boldsymbol{\Phi}$ are in general different. Furthermore, a dyadic used as a prefactor gives the same result as its conjugate acting as postfactor; that is,

$$\boldsymbol{\Phi} \cdot \boldsymbol{r} = \boldsymbol{r} \cdot \boldsymbol{\Phi}_c \tag{A.3}$$

define the same linear vector function (i.e., produce the same vector). Any two dyadics $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are said to be equal when

$$\boldsymbol{s} \cdot \boldsymbol{\Phi} \cdot \boldsymbol{r} = \boldsymbol{s} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{r} \tag{A.4}$$

for all values of $\boldsymbol{s}$ and $\boldsymbol{r}$, where the products in both members of this expression are associative. We write $\boldsymbol{\Phi} = \boldsymbol{0}$, the zero dyadic, when $\boldsymbol{\Phi} \cdot \boldsymbol{r} = \boldsymbol{0}$ for every vector $\boldsymbol{r}$.

The distributive law of ordinary algebra is valid in the expansion of the combination of vectors in a dyad, provided that the order of the vectors be maintained. That is,

$$(\boldsymbol{a} + \boldsymbol{b})(\boldsymbol{c} + \boldsymbol{d}) = \boldsymbol{a}\boldsymbol{c} + \boldsymbol{a}\boldsymbol{d} + \boldsymbol{b}\boldsymbol{c} + \boldsymbol{b}\boldsymbol{d}. \tag{A.5}$$

The order of the vectors in each dyad must not be altered, for, in general, $\boldsymbol{a}\boldsymbol{b} \neq \boldsymbol{b}\boldsymbol{a}$; namely, the commutative law does not hold.

The direct product of two dyads $\boldsymbol{ab}$ and $\boldsymbol{cd}$ is defined as the dyad

$$\boldsymbol{ab} \cdot \boldsymbol{cd} = (\boldsymbol{b} \cdot \boldsymbol{c})\boldsymbol{ad}, \tag{A.6}$$

whose antecedent and consequent are the two outer vectors, taken in order, with a scalar coefficient consisting of the scalar product of the adjacent vectors. Note that $\boldsymbol{cd} \cdot \boldsymbol{ab} = (\boldsymbol{d} \cdot \boldsymbol{a})\boldsymbol{cb}$, which in general differs from Equation A.6. The direct product of a dyadic $\boldsymbol{\Phi}$ into a dyadic $\boldsymbol{\Psi}$ is defined as the dyadic obtained by forming the sum of the direct products of each dyad of $\boldsymbol{\Phi}$ into each dyad of $\boldsymbol{\Psi}$. The defining equations for the product $\boldsymbol{\Phi} \cdot \boldsymbol{\Psi}$, when regarded as an operator to be used as either a prefactor or postfactor, are therefore

$$(\boldsymbol{\Phi} \cdot \boldsymbol{\Psi}) \cdot \boldsymbol{r} = \boldsymbol{\Phi} \cdot (\boldsymbol{\Psi} \cdot \boldsymbol{r}) = \boldsymbol{\Phi} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{r}, \tag{A.7}$$

$$\boldsymbol{r} \cdot (\boldsymbol{\Phi} \cdot \boldsymbol{\Psi}) = (\boldsymbol{r} \cdot \boldsymbol{\Phi}) \cdot \boldsymbol{\Psi} = \boldsymbol{r} \cdot \boldsymbol{\Phi} \cdot \boldsymbol{\Psi}. \tag{A.8}$$

Hence the direct product of two dyadics and a vector is associative, provided the vector follows or precedes both dyadics. However, the associative law is not valid in general for the product of a dyadic, a vector, and another dyadic: $(\boldsymbol{\Phi} \cdot \boldsymbol{r}) \cdot \boldsymbol{\Psi} \neq \boldsymbol{\Phi} \cdot (\boldsymbol{r} \cdot \boldsymbol{\Psi})$.

A dyadic $\boldsymbol{U}$ which, when applied as prefactor or postfactor to any vector, leaves the vector unchanged is called the identity dyadic, or *idemfactor*. Hence, for any vector $\boldsymbol{r}$,

$$\boldsymbol{U} \cdot \boldsymbol{r} = \boldsymbol{r} \cdot \boldsymbol{U} = \boldsymbol{r}. \tag{A.9}$$

Moreover, the direct product of a dyadic $\boldsymbol{\Phi}$ and the identity dyadic $\boldsymbol{U}$ is the dyadic itself:

$$\boldsymbol{\Phi} \cdot \boldsymbol{U} = \boldsymbol{U} \cdot \boldsymbol{\Phi} = \boldsymbol{\Phi}. \tag{A.10}$$

The skew, or vector, product of a dyadic $\boldsymbol{\Phi} = \boldsymbol{a}_i\boldsymbol{b}_i$ and a vector $\boldsymbol{r}$ are denoted by $\boldsymbol{\Phi} \times \boldsymbol{r}$ and $\boldsymbol{r} \times \boldsymbol{\Phi}$, and are defined analogous to the definitions of the direct products of a dyadic and a vector. Accordingly, we have

$$\boldsymbol{\Phi} \times \boldsymbol{r} = \boldsymbol{a}_1\boldsymbol{b}_1 \times \boldsymbol{r} + \boldsymbol{a}_2\boldsymbol{b}_2 \times \boldsymbol{r} + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n \times \boldsymbol{r}, \tag{A.11}$$

$$\boldsymbol{r} \times \boldsymbol{\Phi} = \boldsymbol{r} \times \boldsymbol{a}_1\boldsymbol{b}_1 + \boldsymbol{r} \times \boldsymbol{a}_2\boldsymbol{b}_2 + \cdots + \boldsymbol{r} \times \boldsymbol{a}_n\boldsymbol{b}_n, \tag{A.12}$$

the skew product being itself a dyadic. The associative law holds for the continued product formed in the dot or cross multiplication by a vector of the direct product of any number of dyadics, provided the vector is an extreme factor in the product. For any vectors $\boldsymbol{r}$ and $\boldsymbol{s}$, we can therefore write without ambiguity:

$$\boldsymbol{r} \times \boldsymbol{\Phi} \cdot \boldsymbol{\Psi}, \quad \boldsymbol{\Phi} \cdot \boldsymbol{\Psi} \times \boldsymbol{r}, \quad \boldsymbol{r} \times \boldsymbol{\Phi} \cdot \boldsymbol{s}, \quad \boldsymbol{r} \cdot \boldsymbol{\Phi} \times \boldsymbol{s}, \quad \boldsymbol{r} \times \boldsymbol{\Phi} \times \boldsymbol{s}. \tag{A.13}$$

Any vector $\boldsymbol{a}$ used in cross multiplication with $\boldsymbol{b}$ defines a linear vector function; consequently, the operator $\boldsymbol{a}\times$ may be represented as a dyadic. Namely,

$$\boldsymbol{a} \times \boldsymbol{b} = (\boldsymbol{U} \times \boldsymbol{a}) \cdot \boldsymbol{b} = (\boldsymbol{a} \times \boldsymbol{U}) \cdot \boldsymbol{b}. \tag{A.14}$$

The dyadics $\boldsymbol{U} \times \boldsymbol{a}$ and $\boldsymbol{a} \times \boldsymbol{U}$ are skew-symmetric (anti-self-conjugate); that is,

$$(\boldsymbol{U} \times \boldsymbol{a})_c = -\boldsymbol{U} \times \boldsymbol{a}. \tag{A.15}$$

For notational convenience, we will denote the cross-product dyadic $\boldsymbol{U} \times \boldsymbol{a} = \boldsymbol{a} \times \boldsymbol{U}$ as $\widetilde{\boldsymbol{a}}$; thus, the operator $\widetilde{-}$ turns a vector into a skew-symmetric dyadic, allowing us to write the cross product as a linear vector function. Note that the equivalent statements of the cross product follow:

$$\boldsymbol{a} \times \boldsymbol{b} = \widetilde{\boldsymbol{a}} \cdot \boldsymbol{b} = \boldsymbol{a} \cdot \widetilde{\boldsymbol{b}} = -\widetilde{\boldsymbol{b}} \cdot \boldsymbol{a} = -\boldsymbol{b} \cdot \widetilde{\boldsymbol{a}}. \tag{A.16}$$

The cross-product dyadic satisfies the following important identities:

$$\widetilde{\boldsymbol{a}} \cdot \widetilde{\boldsymbol{b}} = \boldsymbol{b}\boldsymbol{a} - (\boldsymbol{a} \cdot \boldsymbol{b})\, \boldsymbol{U}, \tag{A.17}$$

$$\widetilde{\boldsymbol{a} \cdot \boldsymbol{b}} = \boldsymbol{b}\boldsymbol{a} - \boldsymbol{a}\boldsymbol{b}. \tag{A.18}$$

A dyadic $\boldsymbol{\Phi} = \boldsymbol{a}_i\boldsymbol{b}_i$ represented as a polynomial consisting of an arbitrary number of dyads may be reduced to a form consisting of but three dyads whose antecedents or consequents may be chosen as any three non-coplanar vectors. A dyadic $\boldsymbol{\Phi}$ can also be expressed with respect to an orthonormal basis $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$, as $\boldsymbol{\Phi} = \varphi_{xx}\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \varphi_{xy}\hat{\boldsymbol{x}}\hat{\boldsymbol{y}} + \varphi_{xz}\hat{\boldsymbol{x}}\hat{\boldsymbol{z}} + \varphi_{yx}\hat{\boldsymbol{y}}\hat{\boldsymbol{x}} + \varphi_{yy}\hat{\boldsymbol{y}}\hat{\boldsymbol{y}} + \varphi_{yz}\hat{\boldsymbol{y}}\hat{\boldsymbol{z}} + \varphi_{zx}\hat{\boldsymbol{z}}\hat{\boldsymbol{x}} + \varphi_{zy}\hat{\boldsymbol{z}}\hat{\boldsymbol{y}} + \varphi_{zz}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}$. This is called the nonion form of $\boldsymbol{\Phi}$; in it are represented all the nine dyads formable

with the basis vectors. The identity dyadic $\boldsymbol{U}$ is a very special case of a symmetric dyadic, and may be expressed in the general form:

$$\boldsymbol{U} = \boldsymbol{a}\boldsymbol{a}' + \boldsymbol{b}\boldsymbol{b}' + \boldsymbol{c}\boldsymbol{c}', \tag{A.19}$$

in which the antecedents $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are arbitrary non-coplanar vectors, and $\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{c}'$ are the vectors of their reciprocal system. In particular, since the reciprocal system to the $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$-system is this system itself, the nonion form of the identity dyadic is

$$\boldsymbol{U} = \hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\hat{\boldsymbol{z}}. \tag{A.20}$$

The cross-product dyadic expressed relative to the basis $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$ is

$$\widetilde{\boldsymbol{a}} = a_x(\hat{\boldsymbol{z}}\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}\hat{\boldsymbol{z}}) + a_y(\hat{\boldsymbol{x}}\hat{\boldsymbol{z}} - \hat{\boldsymbol{z}}\hat{\boldsymbol{x}}) + a_z(\hat{\boldsymbol{y}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}). \tag{A.21}$$

Matrix Algebra : In the language of coordinates and matrices, a dyad is equivalent to the outer product; viz.,

$$\boldsymbol{a}\boldsymbol{b} \equiv [\boldsymbol{a}][\boldsymbol{b}]^T = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \begin{bmatrix} b_x & b_y & b_z \end{bmatrix} = \begin{bmatrix} a_x b_x & a_x b_y & a_x b_z \\ a_y b_x & a_y b_y & a_y b_z \\ a_z b_x & a_z b_y & a_z b_z \end{bmatrix}. \tag{A.22}$$

That is, $(\boldsymbol{a}\boldsymbol{b})_{ij} = a_i b_j$, where the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ have been replaced with the column matrices corresponding to their coordinates with respect to the chosen basis. Furthermore, the cross-product dyadic can be represented by the skew-symmetric matrix

$$\widetilde{\boldsymbol{a}} \equiv \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}. \tag{A.23}$$

Finally, the identity dyadic is equivalent to the identity matrix:

$$\boldsymbol{U} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{A.24}$$

Appendix B

Keplerian Motion: The Two-Body Problem

B.1 Importance and Problem Statement

The problem of two bodies is that of determining the relative motion of two spherical masses, whose densities are homogeneous in concentric layers, under the Newtonian law of gravitation. This simple dynamical problem admits a general solution in terms of well-known algebraic and trigonometric functions, and has lost none of its attraction over the centuries. Because of the conditions prevailing in the Solar System, the two-body problem has been the representative system in celestial mechanics and is the basis from which to begin studying more complex problems. Indeed, as the eminent astronomer Jean Kovalevsky explains in his concise and masterfully written book:

> A solution of the two-body problem often represents physical reality in an acceptable way, but this is not the main reason for the importance of this problem. ... all the most complete theories of celestial motion use functions appearing in solutions of two-body problems (elliptic case) as elementary functions. Solutions of the two-body problem constitutes the basic algebra of the dynamics of the Solar System—hence its importance in celestial mechanics.[1]

An exhaustive treatment of the two-body problem may be found in the extensive literature on this subject. The account given here is meant to emphasize the characteristic features of the problem and follows that of Finlay-Freundlich (1958) and Pollard (1966), with the diagrams adapted from Murray & Dermott (1999).

The center of mass of the system of two bodies of mass m_1 and m_2, respectively, is either at rest or moving in a straight line with uniform velocity. Thus, the motion of two bodies about

[1] Kovalevsky J. op. cit., Chap. 2, §8.

their center of mass can always be reduced to an equivalent one-body problem. A still simpler formulation can be attained by shifting the origin of the system from the center of mass to one of the point masses itself, thereby reducing the problem to its most usually observable form: the motion of one mass relative to the other. This transformation does not alter the form of the equations and so no change in the character of the motion is consequently produced. "Each body moves as if, in the place of the second body, a mass equal to the sum of the two masses attracted it according to Newton's law [of gravitation] (Finlay-Freundlich 1958, p. 13)." Accordingly, the relative motion is governed by the equation

$$\ddot{\boldsymbol{r}} = -\frac{\mu}{r^3}\boldsymbol{r}, \tag{B.1}$$

where $\boldsymbol{r}$ is the position vector of mass m_2 relative to m_1, $\mu = \mathcal{G}(m_1 + m_2)$, and $\mathcal{G}$ is the Newtonian gravitational constant. As the equation of relative motion is invariant under $\boldsymbol{r} \to -\boldsymbol{r}$, the relative orbits are geometrically identical. Note that Equation B.1 is equivalent to three simultaneous second-order, nonlinear, differential equations requiring six constants of integration for their complete solution; these constants, called integrals of the motion, are of remarkable significance in conveying the geometrical and dynamical properties of the solution.

B.2 The Vectorial and Classical Integrals of Motion

B.2.1 Angular Momentum Vector and Kepler's Second Law

Forming the vector product of Equation B.1 with the position vector $\boldsymbol{r}$, we obtain $\boldsymbol{r} \times \ddot{\boldsymbol{r}} \equiv \boldsymbol{0}$, since the acceleration $\ddot{\boldsymbol{r}}$ is collinear with the position vector. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r} \times \dot{\boldsymbol{r}}) = \boldsymbol{r} \times \ddot{\boldsymbol{r}} = \boldsymbol{0}.$$

Hence,

$$\boldsymbol{r} \times \boldsymbol{v} = \boldsymbol{H}, \tag{B.2}$$

where $\boldsymbol{v}$ denotes the velocity vector $\dot{\boldsymbol{r}}$ and $\boldsymbol{H}$ is a constant vector. The vectorial (first) integral $\boldsymbol{H}$ represents the angular momentum and indicates that the motion takes place either in a fixed plane

or along a fixed straight line through the origin. If we take the scalar product of Equation B.2 with $\boldsymbol{r}$, we have

$$\boldsymbol{H} \cdot \boldsymbol{r} = 0. \tag{B.3}$$

Thus, if $H \neq 0$, since $\boldsymbol{r}$ must always be orthogonal to the fixed vector $\boldsymbol{H}$, the motion occurs in a plane orthogonal to $\boldsymbol{H}$. Considering the motion in this plane (the orbital plane), the position and velocity vectors can be written in polar coordinates as

$$\boldsymbol{r} = r\hat{\boldsymbol{r}}, \quad \boldsymbol{v} = \dot{r}\hat{\boldsymbol{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}}, \tag{B.4}$$

where $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{h}} \times \hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{h}} = \boldsymbol{H}/H$. Substituting Equations B.4 into Equation B.2, yields

$$H = r^2\dot{\theta}. \tag{B.5}$$

Since $r^2\dot{\theta}$ equals twice the rate at which the radius vector sweeps out area, Equation B.5 expresses Kepler's second law of planetary motion in mathematical form (see Moulton 1914, Chap. 3, Art. 55, for a statement of Kepler's three laws). Consequently, the angular momentum vector is sometimes referred to as the integral of area or vector of areal velocity (Finlay-Freundlich 1958, p. 13; Kurth 1959, p. 57).

If $H = 0$, since

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\boldsymbol{r}}{r}\right) = \frac{\boldsymbol{H} \times \boldsymbol{r}}{r^3}, \tag{B.6}$$

the vector $\boldsymbol{r}/r$ is a constant, and the motion occurs on a straight line passing through the center of attraction. In this case, the position and velocity vectors are parallel and the resulting motion is said to be rectilinear.

B.2.2 Eccentricity Vector and Kepler's First Law

It turns out that the two-body problem admits another vectorial constant of motion, in addition to the angular momentum vector $\boldsymbol{H}$. Multiplying Equation B.6 by μ and substituting

Equation B.1, we have

$$\mu \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\boldsymbol{r}}{r}\right) = \dot{\boldsymbol{v}} \times \boldsymbol{H}.$$

Integrating both sides of this relation, yields

$$\boldsymbol{b} = \boldsymbol{v} \times \boldsymbol{H} - \mu \frac{\boldsymbol{r}}{r}, \tag{B.7}$$

where the constant of integration $\boldsymbol{b}$ is known as the Laplace-Runge-Lenz vector; the history of its rediscoveries being etched in its name. Scaling this vector by μ gives the eccentricity vector

$$\boldsymbol{e} = \frac{1}{\mu}\boldsymbol{v} \times \boldsymbol{H} - \frac{\boldsymbol{r}}{r}. \tag{B.8}$$

Note that if $H = 0$, $\boldsymbol{e} = -\boldsymbol{r}/r$, so that $\boldsymbol{e}$ lies along the line of motion with unity magnitude ($e \equiv 1$). From Equations B.3 and B.8, it follows that $\boldsymbol{H} \cdot \boldsymbol{e} = 0$. Thus, if $H \neq 0$, the vectors $\boldsymbol{H}$ and $\boldsymbol{e}$ are orthogonal and $\boldsymbol{e}$ lies in the plane of motion. Forming the scalar product of Equation B.8 with $\boldsymbol{r}$ gives

$$\boldsymbol{e} \cdot \boldsymbol{r} = \frac{H^2}{\mu} - r. \tag{B.9}$$

Let f, called the true anomaly, denote the angle between $\boldsymbol{r}$ and $\boldsymbol{e}$ so that we have

$$r = \frac{H^2/\mu}{1 + e\cos f}; \tag{B.10}$$

this is known as the polar equation of the orbit. Consequently, the relative orbits of two bodies attracting each other according to Newton's law of gravitation are conic sections of eccentricity e and parameter $p = H^2/\mu$; we generally refer to such motion as Keplerian. Each body lies in the focus of the conic section described by the other. The nature of this conic is an ellipse, a parabola, or a branch of the hyperbola, according to whether the eccentricity is less than, equal to, or greater than unity. When the eccentricity is zero, we have circular motion, the limiting case of an ellipse. It is assumed here that $H \neq 0$; otherwise all three forms of motion degenerate to rectilinear motion. Note that the case of elliptic motion in which the orbit is bounded ($e < 1$) represents Kepler's first law of planetary motion.

A further integral is obtained from Equation B.1 by scalar multiplying by $\boldsymbol{v}$, as we get an exact differential:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}v^2\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mu}{r}\right).$$

Therefore,

$$\frac{1}{2}v^2 - \frac{\mu}{r} = E. \tag{B.11}$$

Equation B.11 is known as the law of conservation of energy and E represents the specific (mechanical) energy of the system; $v^2/2$ is related to the kinetic energy and $-\mu/r$ to the potential energy of the system. The constancy of E and H are sufficient to provide all the information required for a complete description of the relative motion in the plane defined by $\hat{\boldsymbol{h}}$; they give explicit constraints between the velocity and position of the relative solution (see, for instance, Scheeres 2012b, §3.1).

The angular momentum vector $\boldsymbol{H}$, eccentricity vector $\boldsymbol{e}$, and energy integral E furnish a set of only five functionally independent (scalar) constants of motion since there exist two relationships between them. We have already noted one relation among $\boldsymbol{H}$ and $\boldsymbol{e}$; namely, $\boldsymbol{H} \cdot \boldsymbol{e} = 0$. To obtain the second one, take the scalar product of the eccentricity vector, Equation B.8, with itself:

$$\mu^2(e^2 - 1) = 2EH^2. \tag{B.12}$$

If $H \neq 0$, then we can immediately deduce that the orbit will be elliptical ($e < 1$) if $E < 0$, parabolic ($e = 1$) if $E = 0$, and hyperbolic ($e > 1$) if $E > 0$. From the theory of conic sections, the parameter p of an ellipse and hyperbola is $p = a(1 - e^2)$; consequently, Equation B.12 yields the semi-major axis a of the orbit

$$a = -\frac{\mu}{2E}. \tag{B.13}$$

The semi-major axis depends only on the energy integral E. In the parabolic case, $E = 0$ and the semi-major axis tends to infinity.

B.2.3 Classical Orbital Elements

The two vector constants of motion, $\boldsymbol{H}$ and $\boldsymbol{e}$, give us in effect the necessary integrals to complete the problem, and, formally, the system given by Equations B.1 can be considered reduced to quadratures. The representation of the motion by the conserved quantities of the orbit is a characteristic feature of celestial mechanics; however, the integrals discussed thus far are not those traditionally used to describe the orbit.

The orbital plane is usually defined geometrically relative to an inertial coordinate frame by the inclination i, which is the angle between $\hat{\boldsymbol{h}}$ and the $\hat{\boldsymbol{z}}$-axis of the frame, and the longitude of the ascending node Ω, which is the angle from the $\hat{\boldsymbol{x}}$-axis to the node vector $\hat{\boldsymbol{n}}$ defined by $\hat{\boldsymbol{z}} \times \boldsymbol{H}/|\hat{\boldsymbol{z}} \times \boldsymbol{H}|$.[2] We have thus defined two angles, called orbital elements, required to fix the position of the orbit in space.

Similarly, the fact that the major axis in the orbital plane is fixed in space must also be able to be expressed as an angle. The manner in which this is usually done is by defining another orbital element, the argument of periapsis ω, as the angle between node vector and $\boldsymbol{e}$; the periapsis being the nearest point to the focus in the orbital plane (in the case of elliptic orbits, the opposite point being the farthest one from the focus is called the apoapsis).

The orbital elements Ω, i, and ω are known more generally, in the study of rigid body dynamics, as the Euler angles (Battin 1999, p. 84). According to known formula in spherical trigonometry, the two vectorial integrals, $\boldsymbol{H}$ and $\boldsymbol{e}$, can be specified using the Euler angles relative to an inertial frame as

$$\boldsymbol{H} = H\hat{\boldsymbol{h}}, \tag{B.14}$$

$$\boldsymbol{e} = e\hat{\boldsymbol{e}}, \tag{B.15}$$

in which $\hat{\boldsymbol{h}}$ and $\hat{\boldsymbol{e}}$ are the orientation-defining integrals of the two-body problem (Strömgren 1929;

[2] The line of intersection between the orbital plane and the $\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}$ plane is called the line of nodes, and the point at which the orbit crosses this reference plane moving from below to above the plane is called the ascending node.

Brouwer & Clemence 1961a; Scheeres 2012b):

$$\hat{\boldsymbol{h}} = \sin i \sin\Omega \hat{\boldsymbol{x}} - \sin i \cos\Omega \hat{\boldsymbol{y}} + \cos i \hat{\boldsymbol{z}}, \tag{B.16}$$

$$\begin{aligned}\hat{\boldsymbol{e}} = &(\cos\omega\cos\Omega - \cos i \sin\omega \sin\Omega)\hat{\boldsymbol{x}} \\ &+ (\cos\omega \sin\Omega + \cos i \sin\omega \cos\Omega)\hat{\boldsymbol{y}} \\ &+ \sin i \sin\omega \hat{\boldsymbol{z}}.\end{aligned} \tag{B.17}$$

Another important unit vector, necessary to complete the triad, is parallel to the minor axis of the orbit and may be specified by

$$\hat{\boldsymbol{e}}_\perp = \hat{\boldsymbol{h}} \times \hat{\boldsymbol{e}} \tag{B.18}$$

$$\begin{aligned}= &-(\sin\omega\cos\Omega + \cos i \cos\omega \sin\Omega)\hat{\boldsymbol{x}} \\ &- (\sin\omega \sin\Omega - \cos i \cos\omega \cos\Omega)\hat{\boldsymbol{y}} \\ &+ \sin i \cos\omega \hat{\boldsymbol{z}}.\end{aligned} \tag{B.19}$$

The geometrical relationships between these vectors and the Euler angles are shown in Figure B.1a.

The shape the orbit is characterized by the eccentricity e and the size by either the semi-major axis a or the parameter p (see Figure B.1b for the elliptic case). These scalar integrals, together with the three Euler angles, constitute the set of classical orbital elements excepting only for the integral that links the motion in the orbit to the time.

From Equations B.4, B.5, and B.10, the position and velocity vectors describing the relative orbit may be expressed in terms of the orbital parameters defined thus far as

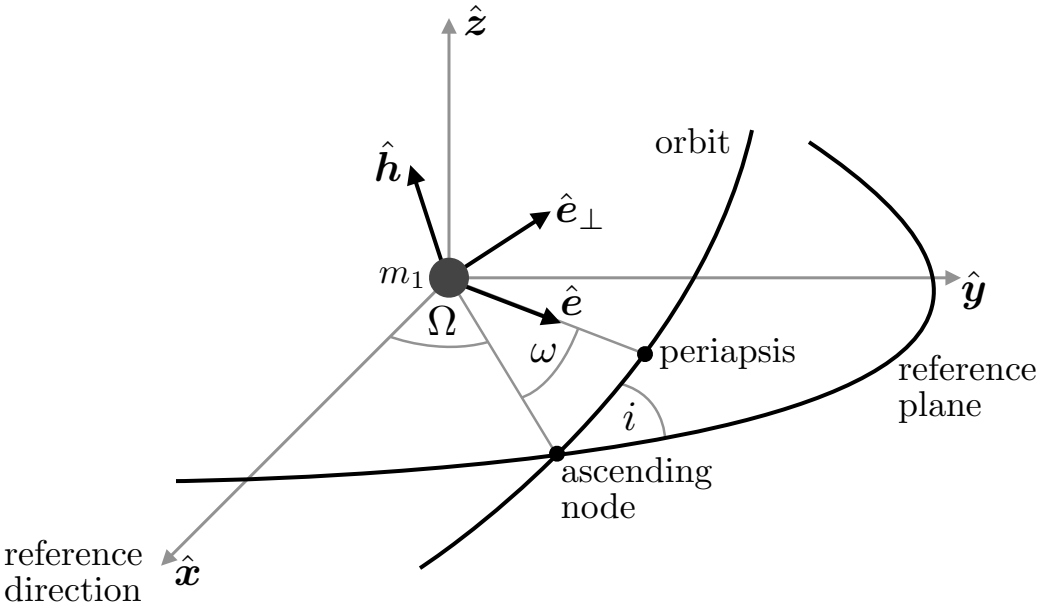
$$\boldsymbol{r} = r\left(\cos f \hat{\boldsymbol{e}} + \sin f \hat{\boldsymbol{e}}_\perp\right), \tag{B.20}$$

$$\boldsymbol{v} = \frac{\mu}{H}\left[-\sin f \hat{\boldsymbol{e}} + (e + \cos f)\,\hat{\boldsymbol{e}}_\perp\right], \tag{B.21}$$

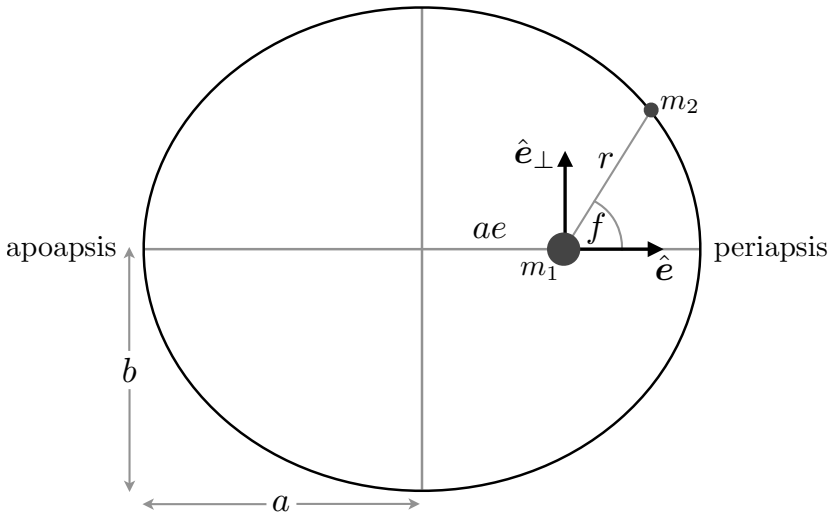
where we note that the components of the velocity vector along and perpendicular to the radius vector are given by

$$\dot{r} = \frac{\mu}{H} e \sin f, \tag{B.22}$$

$$r\dot{f} = \frac{\mu}{H}(1 + e\cos f). \tag{B.23}$$



(a) Orientation of the orbital plane in space.



(b) Geometry of the ellipse of eccentricity e, semi-major axis a, and semi-minor axis $b = a\sqrt{1-e^2}$.

Figure B.1: Geometrical description of the classical orbital elements

B.3 The Orbit in Time

The motion of the body in orbit will be described by an equation connecting the true anomaly f to the time. Such an equation can be derived by integrating Equation B.5, which expresses the constancy of the angular momentum magnitude (Kepler's second law). By combining Equations B.5 and B.10 and noting that the argument of latitude $\theta = \omega + f$, where ω is a constant, we may write

$$\sqrt{\frac{\mu}{p^3}}\,\mathrm{d}t = \frac{\mathrm{d}f}{(1 + e\cos f)^2}. \tag{B.24}$$

Stated in this form, the quadrature can be carried out, yielding the remaining constant of integration (classically chosen to be the time of periapsis passage τ); however, the relations become simpler and more lucid if we introduce other angular variables. Note that it is necessary to consider separately the three kinds of conic section. For a complete treatment of this subject, the literature is rich in detailed explanations (see, for example, Finlay-Freundlich 1958, Chap. 2, §4). For the case $E < 0$ and $0 < e < 1$ (elliptic orbits), the integrated form of Equation B.24 is Kepler's equation:

$$M - M_0 = E - e\sin E, \tag{B.25}$$

where $M = nt$ is the mean anomaly, M_0 is the mean anomaly at epoch (an integral of motion related to τ), E is the eccentric anomaly, and n represents the mean angular motion of the body in its orbit and can be expressed by the elements of the orbit:

$$n = \sqrt{\frac{\mu}{a^3}} = \frac{2\pi}{T}, \tag{B.26}$$

where T is the period of the orbit, which is in accordance with Kepler's third law. The seemingly simple Kepler's equation, being transcendental in nature, has intrigued astronomers and mathematicians for centuries and has exerted significant influence on the development of many branches of mathematics (q.v., Moulton 1914).

Kepler's equation (or the equivalent formulations for parabolic and hyperbolic orbits) and the last integral τ make possible the resolution of the problem of two bodies. For elliptic motion, the solution may be written directly in terms of the eccentric anomaly as

$$\boldsymbol{r} = a\left[(\cos E - e)\hat{\boldsymbol{e}} + \sqrt{1 - e^2}\sin E\hat{\boldsymbol{e}}_\perp\right], \tag{B.27}$$

$$\boldsymbol{v} = \frac{\sqrt{\mu a}}{r}\left[\sin E\hat{\boldsymbol{e}} + \sqrt{1-e^2}\cos E\hat{\boldsymbol{e}}_\perp\right], \tag{B.28}$$

in which $r = a(1 - e\cos E)$. It is important to note here that the solution is given in the form of the six independent first integrals, called the elements of the orbit, and cannot be written explicitly as a function of time. The fact that the coordinates used in the equations of motion do not appear in the solution indicates that they are not well-suited to the problem; the introduction of more suitable parameters will both facilitate the solution, present it in an especially simple form, and provide a conceptual basis for our future mathematical formulations (Finlay-Freundlich 1958), as demonstrated in Chapter 2.

Appendix C

Mean Values in Elliptic Motion

In the perturbed Kepler problem, the greater part of the work in obtaining the secular equations of motion by the method of averaging (§2.3) consists in calculating certain mean values. The time-average is performed over a periodic motion having a period by far shorter than the time which characterizes the evolution of the dynamical system; this periodicity necessarily implies that averaging is taken for elliptical orbits. The removal of time or (what amounts to the same thing) mean anomaly requires computing the quadrature of functions depending implicitly on this variable through the true anomaly (q.v., Appendix B). The traditional approach has resorted to more or less elaborate expansions of various functions into sums of periodic terms, which depend chiefly on Taylor's series and Fourier's theorem, mainly because the integrals of these functions cannot be obtained conveniently in any other way. These expansions in elliptic motion are also fundamentally connected with the analytical development of the disturbing function—a subject which has and continues to engage the attention of many astronomers and mathematicians (see, e.g., Tisserand 1889; Brouwer & Clemence 1961a, Chap. 2 and Chap. 15; Murray & Dermott 1999, Chap. 6). The several-centuries-old literature on this subject is too extensive to give in detail, but there are some particular cases of special simplicity and importance. We therefore consider briefly a few formal manipulations required for our purposes, as well as present a new general procedure which copes with a variety of functions encountered in the averaging methods of celestial mechanics.

Given a quantity $\mathcal{F}(\boldsymbol{\alpha}, M)$, defined as a function of the dimensionless time variable M in

addition to the other orbital elements represented as $\boldsymbol{\alpha}$, the average is defined by

$$\overline{\mathcal{F}}(\boldsymbol{\alpha}) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{F}(\boldsymbol{\alpha}, M)\,\mathrm{d}M, \tag{C.1}$$

where the orbital elements $\boldsymbol{\alpha}$ are held constant in the integration. The fundamental equations of elliptic motion presented in Appendix B give at once the means of calculating any desired orbital average. Although the average is defined with respect to mean anomaly, it is often more easily calculated by means of the true or eccentric anomaly, using the differential relationships

$$\frac{\mathrm{d}M}{2\pi} = \frac{r\mathrm{d}E}{2\pi a} = \frac{r^2\mathrm{d}f}{2\pi ab}, \tag{C.2}$$

where $b = a\sqrt{1-e^2}$ is the semi-minor axis; yielding the equivalent forms for averaging (Hestenes 1999; Scheeres 2012b):

$$\overline{\mathcal{F}}(\boldsymbol{\alpha}) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{F}(\boldsymbol{\alpha}, M)\,\mathrm{d}M = \frac{1}{2\pi a}\int_0^{2\pi} \mathcal{F}(\boldsymbol{\alpha}, E) r\,\mathrm{d}E = \frac{1}{2\pi ab}\int_0^{2\pi} \mathcal{F}(\boldsymbol{\alpha}, f) r^2\,\mathrm{d}f. \tag{C.3}$$

Note that r can be expressed in terms of f and E as

$$r = \begin{cases} \dfrac{a(1-e^2)}{1+e\cos f}, \\ \\ a(1-e\cos E). \end{cases} \tag{C.4}$$

We shall often require the mean values of certain powers such as r^n, $r^n\cos mf$, $r^n\cos^m f$,..., where n and m are integers. While detailed explicit tabulations for these quantities may be found in Cayley's and Newcomb's tables (q.v., Hughes 1981, and references therein), these results may be deduced from the Hansen coefficients. The outstanding importance of Hansen's coefficients was recognized by Tisserand, who expresses them as series invoking Bessel and hypergeometric functions (Tisserand 1889, Chap. 15); from which, several formulae of recurrence can be derived that greatly facilitate their calculation (Hughes 1981).

The general Hansen coefficient $X_k^{n,m}(e)$ is a function of the orbit eccentricity and is defined by the integral

$$X_k^{n,m}(e) = \frac{1}{2\pi}\int_0^{2\pi} \left(\frac{r}{a}\right)^n \cos(mf - kM)\,\mathrm{d}M. \tag{C.5}$$

For $k = 0$, exact analytical expressions exist for the zero-order Hansen coefficients $X_0^{n,m}$ for all values of n and m; for $k \neq 0$, the analytical expressions for them do not terminate and, consequently, the series have to be truncated at some particular order in the eccentricity (Hughes 1981).

The integrals for $X_0^{n,m}$ and $X_0^{-(n+1),m}$ can be written as

$$X_0^{n,m} = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{r}{a}\right)^n \cos mf \, \mathrm{d}M, \tag{C.6}$$

and

$$X_0^{-(n+1),m} = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{a}{r}\right)^{(n+1)} \cos mf \, \mathrm{d}M. \tag{C.7}$$

Note that since cosine is an even function, it is only necessary to obtain expressions for $m > 0$ as $X_0^{n,m} = X_0^{n,-m}$ and $X_0^{-(n+1),m} = X_0^{-(n+1),-m}$. For $n \geq 1$ and $0 \leq m \leq n$, the integrals can be evaluated as (Tisserand 1889, p. 258; Plummer 1918, pp. 44–46)

$$X_0^{n,m} = \left(-\frac{e}{2}\right)^m \binom{n+m+1}{m} F\left(\frac{m-n-1}{2}, \frac{m-n}{2}, m+1; e^2\right), \tag{C.8}$$

and

$$X_0^{-(n+1),m} = \left(\frac{e}{2}\right)^m \frac{1}{(1-e^2)^{(2n-1)/2}} \sum_{j=0}^{\mathrm{int}[(n-1-m)/2]} \frac{1}{2^j} \binom{n-1}{2j+m} \binom{2j+m}{j} e^{2j}, \tag{C.9}$$

where $F(\alpha, \beta, \gamma; x)$ is a hypergeometric function in e^2 (q.v., Battin 1999, Chap. 1); $\binom{m}{n}$ is the binomial coefficient:

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}; \tag{C.10}$$

and $\mathrm{int}[a]$ denotes the integer part of a and is related to the floor $\lfloor a \rfloor$ and ceiling $\lceil a \rceil$ functions by

$$\mathrm{int}[a] = \begin{cases} \lfloor a \rfloor & \text{if } a \geq 0, \\ \lceil a \rceil & \text{if } a < 0. \end{cases} \tag{C.11}$$

The recursive relations for $X_0^{n,m}$ are given by (Hughes 1981, and references therein)

$$X_0^{n+1,m} = \frac{2n+3}{n+2} X_0^{n,m} - \frac{(n+1-m)(n+1+m)}{(n+1)(n+2)} (1-e^2) X_0^{n-1,m}, \tag{C.12}$$

and

$$eX_0^{n,m+1} = \frac{1}{(n-m+1)}\left[e(n+m+1)X_0^{n,m-1} + 2mX_0^{n,m}\right]; \tag{C.13}$$

the corresponding recursive formulae for $X_0^{-(n+1),m}$ are given by

$$(n-m+1)(n-m-1)X_0^{-(n+3),m} = \frac{n+1}{1-e^2}\left[(2n+1)X_0^{-(n+2),m} - nX_0^{-(n+1),m}\right], \tag{C.14}$$

and

$$\begin{aligned} X_0^{-(n+1),m} = \frac{1}{n-m-1}\Big[&2(m+1)(1-e^2)^{1/2}X_0^{-(n+1),(m+1)} \\ &+(n+m+1)e^2(1-e^2)^{3/2}X_0^{-(n+1),(m+2)}\Big]. \end{aligned} \tag{C.15}$$

From Equations C.8 and C.9, the first few coefficients (needed for the recursion) are

$$X_0^{1,0} = 1+\frac{1}{2}e^2, \quad X_0^{1,1} = -\frac{3}{2}e, \quad X_0^{2,0} = 1+\frac{3}{2}e^2, \quad X_0^{2,1} = -2e-\frac{1}{2}e^3, \tag{C.16}$$

and

$$X_0^{-2,0} = (1-e^2)^{-1/2}, \quad X_0^{-2,1} = 0, \quad X_0^{-3,0} = (1-e^2)^{-3/2}, \quad X_0^{-3,1} = \frac{1}{2}e(1-e^2)^{-3/2}. \tag{C.17}$$

For $n = 0$, Kozai (1962a) proved the general result

$$X_0^{0,m} = \frac{1}{2\pi}\int_0^{2\pi} \cos mf \,\mathrm{d}M = \frac{(-e)^m(1+m\sqrt{1-e^2})}{(1+\sqrt{1-e^2})^m}, \tag{C.18}$$

so that $X_0^{0,0} = 1$, $X_0^{0,1} = -e$. We also have (Kelly 1989)

$$\frac{1}{2\pi}\int_0^{2\pi} \sin mf \,\mathrm{d}M = 0. \tag{C.19}$$

Mean values of the form $\overline{r^n \cos^m f \sin^l f}$ vanish for odd values of l (Born 1927, §22). For even values of l, both $\cos^m f$ and $\sin^m f$ may be replaced by their power-reduction formulae and the mean value reduced to the forms just considered.

The average of certain powers of radius are easily obtained from the Hansen coefficients as

$$\overline{r^n} = a^n X_0^{n,0}; \tag{C.20}$$

in particular

$$\overline{r} = a\left(1+\frac{1}{2}e^2\right), \quad \overline{r^2} = a^2\left(1+\frac{3}{2}e^2\right), \quad \overline{r^{-2}} = \frac{1}{ab}, \quad \overline{r^{-3}} = \frac{1}{b^3}, \quad \overline{r^{-4}} = \frac{a\left(1+\frac{1}{2}e^2\right)}{b^5}, \tag{C.21}$$

while the special case of $1/r$ is (Born 1927, p. 144)

$$\overline{r^{-1}} = \frac{1}{a}. \tag{C.22}$$

These results can also be derived from the useful formulae of (Boué & Laskar 2009) and (Scheeres 2012b), based on Fourier series expansions of quantities of the form $(1+e\cos x)^n$. Specifically, for $n \geq 0$,

$$\overline{r^n} = \frac{a^n}{2\pi}\int_0^{2\pi}(1-e\cos E)^{n+1}\,\mathrm{d}E \tag{C.23}$$

$$= a^n \sum_{l=0}^{\mathrm{int}[\frac{n+1}{2}]} \frac{(n+1)!}{(l!)^2(n+1-2l)!}\left(\frac{e}{2}\right)^{2l}. \tag{C.24}$$

Similarly, for r^{-n}, $n \geq 2$, we can write

$$\overline{\frac{1}{r^n}} = \frac{1}{2\pi a^n(1-e^2)^{n-3/2}}\int_0^{2\pi}(1+e\cos f)^{n-2}\,\mathrm{d}f \tag{C.25}$$

$$= \frac{1}{a^n(1-e^2)^{n-3/2}} \sum_{l=0}^{\mathrm{int}[\frac{n-2}{2}]} \frac{(n-2)!}{(l!)^2(n-2-2l)!}\left(\frac{e}{2}\right)^{2l}. \tag{C.26}$$

For $n \geq 3$, these two general results are related by the general formula (Hestenes 1999)

$$\overline{\frac{1}{r^n}} = \frac{\overline{r^{n-3}}}{b^{2n-3}}. \tag{C.27}$$

We can now determine in a straightforward manner the averages of the position and velocity vectors; with the help of Equations B.20 and B.21, we obtain (recognizing that the odd sine functions vanish)

$$\overline{\boldsymbol{r}} = \overline{r\cos f}\,\hat{\boldsymbol{e}} = aX_0^{1,1}\hat{\boldsymbol{e}}$$

$$= -\frac{3}{2}a\boldsymbol{e}, \tag{C.28}$$

$$\overline{\boldsymbol{v}} = \frac{\mu}{H}\left(e+\overline{\cos f}\right)\hat{\boldsymbol{e}}_\perp = \frac{\mu}{H}\left(e+X_0^{0,1}\right)\hat{\boldsymbol{e}}_\perp$$

$$= \boldsymbol{0}. \tag{C.29}$$

The time-average of the velocity vector makes physically intuitive sense as (Hestenes 1999)

$$\overline{\boldsymbol{v}} = \frac{1}{T}\int_0^T \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\,\mathrm{d}t = \frac{1}{T}\oint \mathrm{d}\boldsymbol{r} = \boldsymbol{0}. \tag{C.30}$$

While the above procedure is straightforward, it is certainly laborious (especially when considering dyadics and higher rank tensors of the dynamical variables); we can circumvent some of the tedium by employing an alternative elegant approach akin to that of Hestenes (1999, §8.2). The expression (B.20) for the position vector $\boldsymbol{r}$ may be written

$$\boldsymbol{r} = \frac{1}{e^2}\left[\left(\frac{b^2}{a} - r\right)\boldsymbol{e} + \frac{1}{\mu}(\boldsymbol{r}\cdot\boldsymbol{v})\widetilde{\boldsymbol{H}}\cdot\boldsymbol{e}\right], \tag{C.31}$$

by virtue of Equation B.9, and upon forming the vector product of Equation B.8 with $\boldsymbol{r}$. Since $2(\boldsymbol{r}\cdot\boldsymbol{v}) = \mathrm{d}(\boldsymbol{r}\cdot\boldsymbol{r})/\mathrm{d}t$, its mean value vanishes identically; consequently,

$$\overline{\boldsymbol{r}} = \frac{1}{e^2}\left(\frac{b^2}{a} - \overline{r}\right)\boldsymbol{e} = -\frac{3}{2}a\boldsymbol{e}. \tag{C.32}$$

More generally,

$$r^n\boldsymbol{r} = \frac{1}{e^2}\left[\left(\frac{b^2}{a}r^n - r^{n+1}\right)\boldsymbol{e} + \frac{1}{\mu}(r^{n+1}\dot{r})\widetilde{\boldsymbol{H}}\cdot\boldsymbol{e}\right]. \tag{C.33}$$

In the same way, the second term on the right side of this equation can be expressed as a time derivative and its average must therefore vanish; from which it follows that (Hestenes 1999)

$$\overline{r^n\boldsymbol{r}} = \frac{1}{e^2}\left[\frac{b^2}{a}\overline{r^n} - \overline{r^{n+1}}\right]\boldsymbol{e}. \tag{C.34}$$

A few specific examples from this general formula are

$$\overline{r\boldsymbol{r}} = -\frac{1}{2}a^2(4+e^2)\boldsymbol{e}, \quad \overline{r^2\boldsymbol{r}} = -\frac{5}{8}a^3(4+3e^2)\boldsymbol{e}. \tag{C.35}$$

As Equation C.33 holds for all $n \in \boldsymbol{Z}$, we have (Hestenes 1999)

$$\overline{\left(\frac{\boldsymbol{r}}{r^n}\right)} = \frac{1}{e^2}\left[\frac{b^2}{a}\overline{\left(\frac{1}{r^n}\right)} - \overline{\left(\frac{1}{r^{n-1}}\right)}\right]\boldsymbol{e}; \tag{C.36}$$

which, for $n \geq 4$, can be expressed in the slightly simpler form

$$\overline{\left(\frac{\boldsymbol{r}}{r^n}\right)} = \frac{1}{e^2b^{2n-5}}\left[\frac{1}{a}\overline{r^{n-3}} - \overline{r^{n-4}}\right]\boldsymbol{e}, \tag{C.37}$$

with the help of Equation C.27. In particular, we have

$$\overline{\left(\frac{\boldsymbol{r}}{r}\right)} = -\boldsymbol{e}, \quad \overline{\left(\frac{\boldsymbol{r}}{r^2}\right)} = -\frac{1}{a+b}\boldsymbol{e}, \quad \overline{\left(\frac{\boldsymbol{r}}{r^3}\right)} = \boldsymbol{0}, \quad \overline{\left(\frac{\boldsymbol{r}}{r^4}\right)} = \frac{1}{2b^3}\boldsymbol{e}, \quad \overline{\left(\frac{\boldsymbol{r}}{r^5}\right)} = \frac{a}{b^5}\boldsymbol{e}. \tag{C.38}$$

We shall now allude to some averaging results involving dyadics and triadics which have always appeared to be of the greatest difficulty, but which, when considered in light of this analytical approach, will be found to be as simple and natural as those already discussed. Proceeding as before, it is not difficult to prove for $n \in \boldsymbol{Z}$ that (Hestenes 1999)

$$\overline{\left(\frac{\boldsymbol{rr}}{r^n}\right)} = \frac{1}{e^2}\left[\mathcal{A}_1\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} - (\mathcal{A}_1 - e^2\mathcal{B}_1)\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\right], \tag{C.39}$$

in which

$$\mathcal{A}_1 = \frac{b^4}{a^2}\overline{\frac{1}{r^n}} - \frac{2b^2}{a}\overline{\frac{1}{r^{n-1}}} + \overline{\frac{1}{r^{n-2}}}, \quad \mathcal{B}_1 = \overline{\frac{1}{r^{n-2}}}; \tag{C.40}$$

or, for $n \geq 5$,

$$\overline{\left(\frac{\boldsymbol{rr}}{r^n}\right)} = \frac{1}{e^2b^{2n-7}}\left[\mathcal{A}_2\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} - (\mathcal{A}_2 - e^2\mathcal{B}_2)\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\right],$$

where

$$\mathcal{A}_2 = \frac{1}{a^2}\overline{r^{n-3}} - \frac{2}{a}\overline{r^{n-4}} + \overline{r^{n-5}}, \quad \mathcal{B}_2 = \overline{r^{n-5}}. \tag{C.41}$$

In the same way, we have for $n \in \boldsymbol{Z}$

$$\overline{\left(\frac{\boldsymbol{rrr}}{r^n}\right)} = \frac{1}{e^3}\left\{\mathcal{A}_3\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} - (\mathcal{A}_3 - e^2\mathcal{B}_3)\left[\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\right]\right\}, \tag{C.42}$$

where

$$\begin{aligned}\mathcal{A}_3 &= \frac{b^6}{a^3}\overline{\frac{1}{r^n}} - \frac{3b^4}{a^2}\overline{\frac{1}{r^{n-1}}} + \frac{3b^2}{a}\overline{\frac{1}{r^{n-2}}} - \overline{\frac{1}{r^{n-3}}}, \\ \mathcal{B}_3 &= \frac{b^2}{a}\overline{\frac{1}{r^{n-2}}} - \overline{\frac{1}{r^{n-3}}}.\end{aligned} \tag{C.43}$$

From the general formula (C.27), it follows that

$$\overline{\left(\frac{\boldsymbol{rrr}}{r^n}\right)} = \frac{1}{e^3b^{2n-9}}\left\{\mathcal{A}_4\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} - (\mathcal{A}_4 - e^2\mathcal{B}_4)\left[\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\right]\right\}, \tag{C.44}$$

where

$$\mathcal{A}_4 = \frac{1}{a^3}\overline{r^{n-3}} - \frac{3}{a^2}\overline{r^{n-4}} + \frac{3}{a}\overline{r^{n-5}} - \overline{r^{n-6}},$$
$$\mathcal{B}_4 = \frac{1}{a}\overline{r^{n-5}} - \overline{r^{n-6}}, \quad \text{for } n \geq 6. \tag{C.45}$$

Several useful formulae from the above results are

$$\overline{\boldsymbol{rr}} = \frac{1}{2}a^2\left[(1+4e^2)\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + (1-e^2)\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\right], \tag{C.46}$$

$$\overline{\left(\frac{\boldsymbol{rr}}{r}\right)} = \frac{1}{2}a\left[(1+2e^2)\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + (1-e^2)\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\right], \tag{C.47}$$

$$\overline{\left(\frac{\boldsymbol{rr}}{r^5}\right)} = \frac{1}{2b^3}\left[\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\right], \tag{C.48}$$

$$\overline{\boldsymbol{rrr}} = -\frac{5}{8}a^3e\left\{(3+4e^2)\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + (1-e^2)\left[\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\right]\right\}, \tag{C.49}$$

$$\overline{\left(\frac{\boldsymbol{rrr}}{r}\right)} = -\frac{1}{2}a^2e\left\{(3+2e^2)\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + (1-e^2)\left[\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\right]\right\}, \tag{C.50}$$

$$\overline{\left(\frac{\boldsymbol{rrr}}{r^7}\right)} = \frac{1}{4}\frac{ae}{b^5}\left[3\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\right]. \tag{C.51}$$

To understand how these triadics manifest in astronomical applications, consider an arbitrary constant unit vector $\hat{\boldsymbol{p}}$. For gravitational field perturbations, it becomes necessary to average inverse powers of radius times the scalar product of $\hat{\boldsymbol{r}}$ with $\hat{\boldsymbol{p}}$. In particular, Scheeres (2012b, Appendix C) and Boué & Laskar (2006, p. 329) require the mean values of the quantities

$$\frac{(\hat{\boldsymbol{p}}\cdot\boldsymbol{r})^3}{r^7} \quad \text{and} \quad \frac{(\hat{\boldsymbol{p}}\cdot\boldsymbol{r})^2\boldsymbol{r}}{r^7}, \tag{C.52}$$

respectively. Both of these averages can be obtained simply from Equation C.51. In the first case, we dot the constant vector $\hat{\boldsymbol{p}}$ with each of the orientation-defining integrals, yielding

$$\overline{\left(\frac{(\hat{\boldsymbol{p}}\cdot\boldsymbol{r})^3}{r^7}\right)} = \frac{3}{4}\frac{a}{b^5}(\hat{\boldsymbol{p}}\cdot\boldsymbol{e})\left[1-(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{h}})^2\right], \tag{C.53}$$

by virtue of the dyadic identity $\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp = \boldsymbol{U} - \hat{\boldsymbol{h}}\hat{\boldsymbol{h}}$. The average of the second quantity can be obtained by dotting $\hat{\boldsymbol{p}}$ with any two of the vectorial orbital constants in each triad (the order is not important here) in Equation C.51 (cf. Boué & Laskar 2006):

$$\overline{\left(\frac{(\hat{\boldsymbol{p}}\cdot\boldsymbol{r})^2\boldsymbol{r}}{r^7}\right)} = \frac{1}{4}\frac{ae}{b^5}\left\{\left[3(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{e}})^2 + (\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{e}}_\perp)^2\right]\hat{\boldsymbol{e}} + 2(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{e}})(\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{e}}_\perp)\hat{\boldsymbol{e}}_\perp\right\}. \tag{C.54}$$

Note that this same procedure may be carried out to tensors of any rank, as needed.

Hitherto, we have mainly considered the average of quantities dependent on the position of the satellite. Such is the case when treating gravitational perturbations from the Lagrange formulation (qq.v., §3.2 and §3.3). However, in the Gaussian perturbation equations, it becomes necessary to average functions that depend on both dynamical variables.

By taking the vector product of the angular momentum vector $\boldsymbol{H}$ and the expression (B.8) for the eccentricity vector $\boldsymbol{e}$, we can write the velocity vector $\boldsymbol{v}$ in the special form[1]

$$\begin{aligned}\boldsymbol{v} &= \frac{\mu}{H^2}\widetilde{\boldsymbol{H}} \cdot \left(\boldsymbol{e} + \frac{\boldsymbol{r}}{r}\right) \\ &= -\frac{\mu}{H^2}\left(\boldsymbol{e} + \frac{\boldsymbol{r}}{r}\right) \cdot \widetilde{\boldsymbol{H}}.\end{aligned} \tag{C.55}$$

The mean values of any quantity involving the radius and position vector being known, those depending on the velocity vector can, of course, be determined by merely substituting for $\boldsymbol{v}$ the hodograph expression (C.55), expanding and reducing; and a convenient mode of arranging the calculations presents itself so obviously, that little explanation seems required.

We shall sometimes require the mean values of the scalar product of $\boldsymbol{r}$ and $\boldsymbol{v}$ times the vector $\boldsymbol{r}$ (see, for instance, Correia et al. 2011, p. 128). Equation C.55 implies

$$\boldsymbol{r} \cdot \boldsymbol{v} = \frac{\mu e}{H}\boldsymbol{r} \cdot \hat{\boldsymbol{e}}_\perp, \tag{C.56}$$

so that

$$\overline{\frac{(\boldsymbol{r} \cdot \boldsymbol{v})\boldsymbol{r}}{r^n}} = \frac{\mu e}{H}\overline{\left(\frac{\boldsymbol{r}\boldsymbol{r}}{r^n}\right)} \cdot \hat{\boldsymbol{e}}_\perp \tag{C.57}$$

$$= \frac{\mu}{eH}(\mathcal{A}_1 - e^2\mathcal{B}_1)\hat{\boldsymbol{e}}_\perp, \tag{C.58}$$

by virtue of Equation C.39; which, of course, can be put into the simpler form for $n \geq 5$:

$$\overline{\frac{(\boldsymbol{r} \cdot \boldsymbol{v})\boldsymbol{r}}{r^n}} = \frac{aH}{eb^{2n-5}}(\mathcal{A}_2 - e^2\mathcal{B}_2)\hat{\boldsymbol{e}}_\perp. \tag{C.59}$$

As an example, we have

$$\overline{(\boldsymbol{r} \cdot \boldsymbol{v})\boldsymbol{r}} = \frac{1}{2}aeH\hat{\boldsymbol{e}}_\perp. \tag{C.60}$$

[1] Equation C.55 provides a vector representation of the hodograph of Keplerian motion (q.v., Battin 1999, §3.5).

Consider next the dyad of the position and velocity vectors:

$$\boldsymbol{rv} = -\frac{\mu}{H^2}\left(\boldsymbol{re} + \frac{\boldsymbol{rr}}{r}\right)\cdot\widetilde{\boldsymbol{H}}. \tag{C.61}$$

Thus, averaging this dyad amounts to averaging the position vector and the dyad $\boldsymbol{rr}/r$. From Equations C.32 and C.47, we find (cf. Scheeres 2012b, Appendix C)

$$\begin{aligned}\overline{\boldsymbol{rv}} &= -\frac{\mu}{H^2}\left[\overline{\boldsymbol{r}}\boldsymbol{e} + \overline{\left(\frac{\boldsymbol{rr}}{r}\right)}\right]\cdot\widetilde{\boldsymbol{H}} \\ &= -\frac{1}{2}\widetilde{\boldsymbol{H}}. \end{aligned}\tag{C.62}$$

We can proceed in an identical manner for triads; namely,

$$\overline{\boldsymbol{rrv}} = -\frac{\mu}{H^2}\left[\overline{\boldsymbol{rr}}\boldsymbol{e} + \overline{\left(\frac{\boldsymbol{rrr}}{r}\right)}\right]\cdot\widetilde{\boldsymbol{H}} \tag{C.63}$$

$$= \frac{1}{2}aeH\left(-2\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp + \hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\right). \tag{C.64}$$

For $n \in \boldsymbol{Z}$, we can state the general formula

$$\overline{\left(\frac{\boldsymbol{rrv}}{r^n}\right)} = -\frac{\mu}{H^2}\left[\overline{\left(\frac{\boldsymbol{rr}}{r^n}\right)}\boldsymbol{e} + \overline{\left(\frac{\boldsymbol{rrr}}{r^{n+1}}\right)}\right]\cdot\widetilde{\boldsymbol{H}}. \tag{C.65}$$

We have thus presented a systematic method for computing the mean values of any particular function encountered in elliptic motion. All orbital averages occurring in perturbed Keplerian motion reduce to averages of certain powers of r; hence, the fundamental significance and profundity of Hansen's coefficients is now clearly seen.

Appendix D

Partial Derivatives of the Milankovitch Orbital Elements with Respect to the Position and Velocity Vectors

The partial derivatives of the orbital elements with respect to the position and velocity vectors are needed for the Poisson bracket operations and the Gauss equations (q.v., Chapter 2). Taking the partial derivative of Equation B.2 with respect to $\boldsymbol{v}$ and $\boldsymbol{r}$, respectively, and using the cross product identity in Equation A.16, gives

$$\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{v}} = \widetilde{\boldsymbol{r}}, \tag{D.1}$$

$$\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{r}} = -\widetilde{\boldsymbol{v}}. \tag{D.2}$$

Similarly, the partial derivatives of $\boldsymbol{e}$ with respect to these vectors follow from Equation B.8 as

$$\frac{\partial \boldsymbol{e}}{\partial \boldsymbol{v}} = \frac{1}{\mu}\left(\widetilde{\boldsymbol{v}}\cdot\widetilde{\boldsymbol{r}} - \widetilde{\boldsymbol{H}}\right) = \frac{1}{\mu}\left[2\boldsymbol{r}\boldsymbol{v} - \boldsymbol{v}\boldsymbol{r} - (\boldsymbol{r}\cdot\boldsymbol{v})\boldsymbol{U}\right], \tag{D.3}$$

$$\frac{\partial \boldsymbol{e}}{\partial \boldsymbol{r}} = \frac{1}{r^3}\widetilde{\boldsymbol{r}}\cdot\widetilde{\boldsymbol{r}} - \frac{1}{\mu}\widetilde{\boldsymbol{v}}\cdot\widetilde{\boldsymbol{v}} = \frac{\boldsymbol{r}\boldsymbol{r}}{|\boldsymbol{r}|^3} - \frac{1}{\mu}\boldsymbol{v}\boldsymbol{v} + \left(\frac{\boldsymbol{v}\cdot\boldsymbol{v}}{\mu} - \frac{1}{|\boldsymbol{r}|}\right)\boldsymbol{U}. \tag{D.4}$$

The trigonometric functions of the longitude of the ascending node Ω are given by

$$\cos\Omega = -\frac{\hat{\boldsymbol{y}}\cdot\boldsymbol{H}}{|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|}, \quad \sin\Omega = \frac{\hat{\boldsymbol{x}}\cdot\boldsymbol{H}}{|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|}, \quad \tan\Omega = \frac{\hat{\boldsymbol{x}}\cdot\boldsymbol{H}}{-\hat{\boldsymbol{y}}\cdot\boldsymbol{H}}. \tag{D.5}$$

Taking the partial derivative of $\tan\Omega$ with respect to the velocity, and letting $\boldsymbol{H}$ vary according to Equation D.1, yields

$$\frac{\partial \Omega}{\partial \boldsymbol{v}} = \frac{1}{|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|^2}\left[-(\hat{\boldsymbol{y}}\cdot\boldsymbol{H})\hat{\boldsymbol{x}}\cdot\widetilde{\boldsymbol{r}} + (\hat{\boldsymbol{x}}\cdot\boldsymbol{H})\hat{\boldsymbol{y}}\cdot\widetilde{\boldsymbol{r}}\right]. \tag{D.6}$$

From Equation A.21, the term inside the brackets can be rewritten as $-\boldsymbol{H}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\widetilde{\boldsymbol{r}}$. Using the identity $\boldsymbol{H}\cdot\widetilde{\hat{\boldsymbol{z}}}=-\hat{\boldsymbol{z}}\cdot\widetilde{\boldsymbol{H}}$ and noting from Equation A.17 and the orthogonality of $\boldsymbol{H}$ and $\boldsymbol{r}$ that $\widetilde{\boldsymbol{H}}\cdot\widetilde{\boldsymbol{r}}=\boldsymbol{r}\boldsymbol{H}$, we find

$$\frac{\partial\Omega}{\partial\boldsymbol{v}}=\frac{\hat{\boldsymbol{z}}\cdot\boldsymbol{r}}{|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|^2}\boldsymbol{H}. \tag{D.7}$$

The argument of latitude θ is defined as the angle between $\boldsymbol{r}$ and $\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}$:

$$\cos\theta=\frac{\boldsymbol{r}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}}{r|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|},\quad \sin\theta=\frac{(\hat{\boldsymbol{z}}\cdot\boldsymbol{r})H}{r|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|},\quad \tan\theta=\frac{(\hat{\boldsymbol{z}}\cdot\boldsymbol{r})H}{\boldsymbol{r}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}}. \tag{D.8}$$

Differentiating $\tan\theta$ with respect to $\boldsymbol{v}$, and noting that $\partial H/\partial\boldsymbol{v}=\hat{\boldsymbol{h}}\cdot\partial\boldsymbol{H}/\partial\boldsymbol{v}$, we find

$$\frac{\partial\theta}{\partial\boldsymbol{v}}=\frac{(\hat{\boldsymbol{z}}\cdot\boldsymbol{r})H}{r^2|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|^2}\boldsymbol{r}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\left(\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}-\boldsymbol{U}\right)\cdot\widetilde{\boldsymbol{r}}. \tag{D.9}$$

Using the identity $\boldsymbol{U}=\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}+\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}$, where $\hat{\boldsymbol{\theta}}=\widetilde{\hat{\boldsymbol{h}}}\cdot\hat{\boldsymbol{r}}$, it follows that

$$\frac{\partial\theta}{\partial\boldsymbol{v}}=-\frac{(\hat{\boldsymbol{z}}\cdot\boldsymbol{r})H}{|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|^2}\hat{\boldsymbol{r}}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}\cdot\widetilde{\hat{\boldsymbol{r}}}. \tag{D.10}$$

Finally, from Equations A.17 and D.7, we obtain

$$\frac{\partial\theta}{\partial\boldsymbol{v}}=-\frac{(\hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{h}})(\hat{\boldsymbol{z}}\cdot\boldsymbol{r})}{|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|^2}\boldsymbol{H}=-\hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{h}}\frac{\partial\Omega}{\partial\boldsymbol{v}}. \tag{D.11}$$

Consequently, the partial derivative of the true longitude with respect to $\boldsymbol{v}$ can be written

$$\frac{\partial L}{\partial\boldsymbol{v}}=\frac{\partial\Omega}{\partial\boldsymbol{v}}+\frac{\partial\theta}{\partial\boldsymbol{v}}=\frac{\hat{\boldsymbol{z}}\cdot\boldsymbol{r}}{H(H+\hat{\boldsymbol{z}}\cdot\boldsymbol{H})}\boldsymbol{H}. \tag{D.12}$$

Following a similar procedure, the partial derivative of L with respect to $\boldsymbol{r}$ reads

$$\frac{\partial L}{\partial\boldsymbol{r}}=\frac{1}{Hr^2}\widetilde{\boldsymbol{H}}\cdot\boldsymbol{r}-\frac{\hat{\boldsymbol{z}}\cdot\boldsymbol{v}}{H(H+\hat{\boldsymbol{z}}\cdot\boldsymbol{H})}\boldsymbol{H}. \tag{D.13}$$

The argument of periapsis ω is defined as the angle between $\boldsymbol{e}$ and $\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}$:

$$\cos\omega=\frac{\boldsymbol{e}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}}{e|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|},\quad \sin\omega=\frac{(\hat{\boldsymbol{z}}\cdot\boldsymbol{e})H}{e|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|},\quad \tan\omega=\frac{(\hat{\boldsymbol{z}}\cdot\boldsymbol{e})H}{\boldsymbol{e}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}}. \tag{D.14}$$

Differentiating $\tan\omega$ with respect to $\boldsymbol{v}$, we obtain

$$\frac{\partial\omega}{\partial\boldsymbol{v}}=\frac{1}{e^2|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|^2}\left\{\boldsymbol{e}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}\frac{\partial}{\partial\boldsymbol{v}}\left[(\hat{\boldsymbol{z}}\cdot\boldsymbol{e})H\right]-H(\hat{\boldsymbol{z}}\cdot\boldsymbol{e})\frac{\partial}{\partial\boldsymbol{v}}\left[\boldsymbol{e}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}\right]\right\}, \tag{D.15}$$

where $\boldsymbol{H}$ and $\boldsymbol{e}$ vary according to Equations D.1 and D.3, respectively, giving

$$\frac{\partial}{\partial \boldsymbol{v}}\left[(\hat{\boldsymbol{z}}\cdot\boldsymbol{e})H\right] = \frac{H}{\mu}\hat{\boldsymbol{z}}\cdot\left(\widetilde{\boldsymbol{v}}\cdot\widetilde{\boldsymbol{r}} - \widetilde{\boldsymbol{H}}\right) + (\hat{\boldsymbol{z}}\cdot\boldsymbol{e})\hat{\boldsymbol{h}}\cdot\widetilde{\boldsymbol{r}}, \tag{D.16}$$

$$\frac{\partial}{\partial \boldsymbol{v}}\left[\boldsymbol{e}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}\right] = \frac{1}{\mu}\hat{\boldsymbol{z}}\cdot\widetilde{\boldsymbol{H}}\cdot\left(\widetilde{\boldsymbol{v}}\cdot\widetilde{\boldsymbol{r}} - \widetilde{\boldsymbol{H}}\right) + \boldsymbol{e}\cdot\widetilde{\hat{\boldsymbol{z}}}\cdot\widetilde{\boldsymbol{r}}. \tag{D.17}$$

Performing considerable algebra, Equation D.15 can be written in radial-transversal-normal coordinates as

$$\frac{\partial\omega}{\partial\boldsymbol{v}} = -\frac{1}{\mu e}\left[H(\hat{\boldsymbol{e}}\cdot\hat{\boldsymbol{r}})\hat{\boldsymbol{r}} + (r+p)(\hat{\boldsymbol{e}}\cdot\boldsymbol{v})\hat{\boldsymbol{\theta}}\right] - \frac{(\hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{h}})(\hat{\boldsymbol{z}}\cdot\boldsymbol{r})}{|\widetilde{\hat{\boldsymbol{z}}}\cdot\boldsymbol{H}|^2}\boldsymbol{H}. \tag{D.18}$$

Consequently, the partial derivative of the longitude of periapsis, ϖ, with respect to $\boldsymbol{v}$ becomes

$$\frac{\partial\varpi}{\partial\boldsymbol{v}} = \frac{\partial\Omega}{\partial\boldsymbol{v}} + \frac{\partial\omega}{\partial\boldsymbol{v}} \tag{D.19}$$

$$= -\frac{1}{\mu e}\left[H(\hat{\boldsymbol{e}}\cdot\hat{\boldsymbol{r}})\hat{\boldsymbol{r}} + (r+p)(\hat{\boldsymbol{e}}\cdot\boldsymbol{v})\hat{\boldsymbol{\theta}}\right] + \frac{\hat{\boldsymbol{z}}\cdot\boldsymbol{r}}{H(H+\hat{\boldsymbol{z}}\cdot\boldsymbol{H})}\boldsymbol{H}. \tag{D.20}$$

The partial derivative of ϖ with respect to $\boldsymbol{r}$ can be computed in a similar manner, yielding

$$\frac{\partial\varpi}{\partial\boldsymbol{r}} = \frac{1}{eH}\left[\frac{1}{r}(\hat{\boldsymbol{e}}\cdot\hat{\boldsymbol{r}})\hat{\boldsymbol{r}} - \frac{1}{\mu}(\hat{\boldsymbol{e}}\cdot\boldsymbol{v})\boldsymbol{v}\right]\cdot\widetilde{\boldsymbol{H}} - \frac{\hat{\boldsymbol{z}}\cdot\boldsymbol{v}}{H(H+\hat{\boldsymbol{z}}\cdot\boldsymbol{H})}\boldsymbol{H}. \tag{D.21}$$

Appendix E

The Perturbation Equations for the Classical Orbital Elements

The perturbation equations of celestial mechanics—the differential equations which describe how an orbit evolves under the action of small perturbating forces—are traditionally written in terms of the classical orbital elements (q.v., Appendix B.2.3). The foundations of celestial mechanics therefore rest on these elements, and it is necessary to establish the deeper connections between them and the vectorial elements used here. The equations that govern the time rates of change of the classical orbital elements can be expressed in terms of the angular momentum and eccentricity vectors, and their variations (qq.v., Milankovitch 1939; Kurth 1959; Ward 1962; Pollard 1966; Burns 1976; Battin 1999).

Recall the precise and clear geometrical significance of the vectorial elements: $\boldsymbol{H}$ points perpendicular to the instantaneous orbit plane and has magnitude (for elliptical and hyperbolic orbits) $H = \sqrt{\mu a(1-e^2)}$; $\boldsymbol{e}$ defines the orientation of the major axis in the orbital plane, pointing towards the instantaneous periapsis of the orbit, and its magnitude is the eccentricity e. From Appendix B.2.3, we can state the basic definition of these vectors in terms of the classical orbital elements relative to an inertial frame:

$$\boldsymbol{H} = H\hat{\boldsymbol{h}} = H(\sin i \sin\Omega\hat{\boldsymbol{x}} - \sin i \cos\Omega\hat{\boldsymbol{y}} + \cos i\hat{\boldsymbol{z}}), \tag{E.1}$$

$$\begin{aligned}\boldsymbol{e} = e\hat{\boldsymbol{e}} \quad &= e[(\cos\omega\cos\Omega - \cos i\sin\omega\sin\Omega)\hat{\boldsymbol{x}} \\ &\quad + (\cos\omega\sin\Omega + \cos i\sin\omega\cos\Omega)\hat{\boldsymbol{y}} \\ &\quad + \sin i\sin\omega\hat{\boldsymbol{z}}].\end{aligned} \tag{E.2}$$

Applying the formal chain rule results in

$$\frac{\mathrm{d}\boldsymbol{H}}{\mathrm{d}t} = H\sin i\frac{\mathrm{d}\Omega}{\mathrm{d}t}\hat{\boldsymbol{n}} - H\frac{\mathrm{d}i}{\mathrm{d}t}\hat{\boldsymbol{n}}_\perp + \frac{\mathrm{d}H}{\mathrm{d}t}\hat{\boldsymbol{h}}, \tag{E.3}$$

$$\frac{\mathrm{d}\boldsymbol{e}}{\mathrm{d}t} = \frac{\mathrm{d}e}{\mathrm{d}t}\hat{\boldsymbol{e}} + e\frac{\mathrm{d}\omega}{\mathrm{d}t}\hat{\boldsymbol{e}}_\perp + e\sin\omega\frac{\mathrm{d}i}{\mathrm{d}t}\hat{\boldsymbol{h}} - e\frac{\mathrm{d}\Omega}{\mathrm{d}t}\widetilde{\hat{\boldsymbol{e}}}\cdot\hat{\boldsymbol{z}}, \tag{E.4}$$

where $\hat{\boldsymbol{e}}_\perp$ is defined by Equation B.19, the node vector is given by $\hat{\boldsymbol{n}} = \cos\Omega\hat{\boldsymbol{x}} + \sin\Omega\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{n}}_\perp = \widetilde{\hat{\boldsymbol{h}}}\cdot\hat{\boldsymbol{n}}$. If we take the scalar product of Equation E.3 with $\hat{\boldsymbol{n}}$, $\hat{\boldsymbol{n}}_\perp$, and $\hat{\boldsymbol{h}}$, respectively, we have

$$\frac{\mathrm{d}\Omega}{\mathrm{d}t} = \frac{1}{H\sin i}\hat{\boldsymbol{n}}\cdot\dot{\boldsymbol{H}}, \tag{E.5}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = -\frac{1}{H}\hat{\boldsymbol{n}}_\perp\cdot\dot{\boldsymbol{H}}, \tag{E.6}$$

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \hat{\boldsymbol{h}}\cdot\dot{\boldsymbol{H}}. \tag{E.7}$$

Similarly, dotting Equation E.4 with $\hat{\boldsymbol{e}}$ and $\hat{\boldsymbol{e}}_\perp$, respectively, yields

$$\frac{\mathrm{d}e}{\mathrm{d}t} = \hat{\boldsymbol{e}}\cdot\dot{\boldsymbol{e}}, \tag{E.8}$$

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{1}{e}\hat{\boldsymbol{e}}_\perp\cdot\dot{\boldsymbol{e}} - \cos i\frac{\mathrm{d}\Omega}{\mathrm{d}t}. \tag{E.9}$$

From Equations B.12, B.13, E.7, and E.8, the perturbation equation for the semi-major axis can be stated as

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{2}{1-e^2}\left(\mu a\boldsymbol{e}\cdot\dot{\boldsymbol{e}} + \frac{1}{\mu}\boldsymbol{H}\cdot\dot{\boldsymbol{H}}\right). \tag{E.10}$$

Appendix F

Averaged Model for Secondary Perturbations

F.1 Parallactic Term in Third-Body Perturbations

The octupole disturbing function arising from the third harmonic in the Legendre expansion of the third-body perturbations is given by

$$\mathcal{R}_{p,3} = \frac{\mu_p}{2d_p^4}\left[5(\boldsymbol{r}\cdot\hat{\boldsymbol{d}}_p)^3 - 3r^2(\boldsymbol{r}\cdot\hat{\boldsymbol{d}}_p)\right]. \tag{F.1}$$

Averaging Equation F.1 requires the average of the triad $\boldsymbol{rrr}$ and vector $r^2\boldsymbol{r}$; both can be computed as (q.v., Appendix C)

$$\overline{\boldsymbol{rrr}} = -\frac{5}{8}a^3 e\left\{(3+4e^2)\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}\hat{\boldsymbol{e}} + (1-e^2)\left[\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}} + \hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}_\perp + \hat{\boldsymbol{e}}_\perp\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}_\perp\right]\right\}, \tag{F.2}$$

$$\overline{r^2\boldsymbol{r}} = -\frac{5}{8}a^3(4+3e^2)\boldsymbol{e}. \tag{F.3}$$

Substituting these into Equation F.1, we find

$$\bar{\mathcal{R}}^*_{p,3} = -\frac{5a\mu_p}{16nd_p^4}\left\{5\left[7(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})^3 - 3(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{h})^2(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})\right] + 3(1-8e^2)(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})\right\}. \tag{F.4}$$

Computing the partial derivatives and substituting them into Equations 2.36 and 2.37, the secular equations resulting from the parallactic term can be stated as (cf. Allan 1962)

$$\dot{\boldsymbol{h}}_{p,3} = -\frac{15a\mu_p}{16nd_p^4}\left\{5\left[7(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})^2 - (\hat{\boldsymbol{d}}_p\cdot\boldsymbol{h})^2\right]\boldsymbol{e} - 10(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{h})\boldsymbol{h} + (1-8e^2)\boldsymbol{e}\right\}\cdot\widetilde{\hat{\boldsymbol{d}}}_p, \tag{F.5}$$

$$\dot{\boldsymbol{e}}_{p,3} = -\frac{15a\mu_p}{16nd_p^4}\Bigg[\left\{5\left[7(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})^2 - (\hat{\boldsymbol{d}}_p\cdot\boldsymbol{h})^2\right]\boldsymbol{h} - 10(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{h})\boldsymbol{e} + (1-8e^2)\boldsymbol{h}\right\}\cdot\widetilde{\hat{\boldsymbol{d}}}_p$$
$$- 16(\hat{\boldsymbol{d}}_p\cdot\boldsymbol{e})\widetilde{\boldsymbol{h}}\cdot\boldsymbol{e}\Bigg]. \tag{F.6}$$

These equations can subsequently be averaged over the perturber's motion (cf. Katz, Dong & Malhotra 2011; Naoz et al. 2013); however, this will not be pursued here.

F.2 Solar Radiation Pressure with Planetary Shadow Effects

The determination of what part of the orbit, if any, is in the planet's shadow is a well-determined geometrical problem, which can be specified by the eccentric anomaly E.[1] The relevant two-body orbit relations, expressed in terms of the eccentric anomaly, are given in Appendix B.

Suppose the object emerges into sunlight at $E = E_1$ and that E increases continuously over the sunlit part of the orbit to $E = E_2$ when it then re-enters the planet's shadow. Then, under the cannonball model approximation, the averaged equation for the angular momentum vector can be stated as

$$\dot{\boldsymbol{H}} = \frac{1}{2\pi}\int_{E_1}^{E_2} \frac{a\beta}{d_s^2}(e\cos E - 1)\left[(\cos E - e)\widetilde{\hat{\boldsymbol{e}}}\cdot\hat{\boldsymbol{d}}_s + \sqrt{1-e^2}\sin E\widetilde{\hat{\boldsymbol{e}}}_\perp\cdot\hat{\boldsymbol{d}}_s\right]\mathrm{d}E \tag{F.7}$$

$$= \frac{a}{2\pi}\frac{\beta}{d_s^2}\left\{\left[\frac{3}{2}eE - (1+e^2)\sin E + \frac{1}{4}e\sin 2E\right]\Big|_{E_1}^{E_2}\widetilde{\hat{\boldsymbol{e}}}\cdot\hat{\boldsymbol{d}}_s + \sqrt{1-e^2}\left(\cos E - \frac{1}{4}e\cos 2E\right)\Big|_{E_1}^{E_2}\widetilde{\hat{\boldsymbol{e}}}_\perp\cdot\hat{\boldsymbol{d}}_s\right\}. \tag{F.8}$$

Similarly, for the eccentricity vector, we have

$$\dot{\boldsymbol{e}} = \frac{n}{2\pi}\int_{E_1}^{E_2}\frac{a^2\beta}{\mu d_s^2}\Big[\sin E(\cos E - e)\widetilde{\hat{\boldsymbol{e}}}\cdot\widetilde{\hat{\boldsymbol{e}}} + \sqrt{1-e^2}(1 - e\cos E + \sin^2 E)\widetilde{\hat{\boldsymbol{e}}}\cdot\widetilde{\hat{\boldsymbol{e}}}_\perp - \sqrt{1-e^2}(1 - 2e\cos E + \cos^2 E)\widetilde{\hat{\boldsymbol{e}}}_\perp\cdot\widetilde{\hat{\boldsymbol{e}}} - (1-e^2)\cos E\sin E\widetilde{\hat{\boldsymbol{e}}}_\perp\cdot\widetilde{\hat{\boldsymbol{e}}}_\perp\Big]\cdot\hat{\boldsymbol{d}}_s\,\mathrm{d}E \tag{F.9}$$

$$= \frac{1}{2\pi}\sqrt{\frac{a}{\mu}}\frac{\beta}{d_s^2}\left[\left(e\cos E - \frac{1}{4}\cos 2E\right)\Big|_{E_1}^{E_2}\widetilde{\hat{\boldsymbol{e}}}\cdot\widetilde{\hat{\boldsymbol{e}}} + \sqrt{1-e^2}\left(\frac{3}{2}E - e\sin E - \frac{1}{4}\sin 2E\right)\Big|_{E_1}^{E_2}\widetilde{\hat{\boldsymbol{e}}}\cdot\widetilde{\hat{\boldsymbol{e}}}_\perp - \sqrt{1-e^2}\left(\frac{3}{2}E - 2e\sin E + \frac{1}{4}\sin 2E\right)\Big|_{E_1}^{E_2}\widetilde{\hat{\boldsymbol{e}}}_\perp\cdot\widetilde{\hat{\boldsymbol{e}}} + (1-e^2)\left(\frac{1}{4}\cos 2E\right)\Big|_{E_1}^{E_2}\widetilde{\hat{\boldsymbol{e}}}_\perp\cdot\widetilde{\hat{\boldsymbol{e}}}_\perp\right]\cdot\hat{\boldsymbol{d}}_s. \tag{F.10}$$

Neglecting the shadow region will give an approximation to the general case. Setting $E_2 = E_1 + 2\pi$ in Equations F.8 and F.10, and scaling $\boldsymbol{H}$ by $\sqrt{\mu a}$, the secular evolution of $\boldsymbol{h}$ and $\boldsymbol{e}$, in

[1] In this derivation, we follow the exposition of Allan (1962), but not quite in the form presented.

the case of no shadow, are governed by

$$\dot{\boldsymbol{h}} = -\frac{3}{2}\sqrt{\frac{a}{\mu}}\frac{\beta}{d_s^2}\widetilde{\hat{\boldsymbol{d}}}_s \cdot \boldsymbol{e}, \tag{F.11}$$

$$\dot{\boldsymbol{e}} = -\frac{3}{2}\sqrt{\frac{a}{\mu}}\frac{\beta}{d_s^2}\widetilde{\hat{\boldsymbol{d}}}_s \cdot \boldsymbol{h}; \tag{F.12}$$

precisely the same formulas as Equations 4.11, once the SRP perturbation angle (Equation 1.10) is introduced.

Appendix G

Derivation of State Transition Dyadic for the SRP Perturbed Orbiter Problem

The "state transition" matrix (STM) or dyadic for a linear autonomous system can be obtained using the power series definition of the (matrix or dyadic) exponential function

$$\boldsymbol{\Phi}(f) = e^{\boldsymbol{A}f} = \sum_{k=1}^{\infty} \frac{1}{k!} (\boldsymbol{A}f)^k = \boldsymbol{U} + f\boldsymbol{A} + \frac{f^2}{2!}\boldsymbol{A}^2 + \frac{f^3}{3!}\boldsymbol{A}^3 + \cdots, \tag{G.1}$$

$$\boldsymbol{A} = \begin{bmatrix} -\widetilde{\hat{\boldsymbol{H}}_s} & \tan\Lambda'\widetilde{\hat{\boldsymbol{a}}} \\ \\ \tan\Lambda'\widetilde{\hat{\boldsymbol{a}}} & -\widetilde{\hat{\boldsymbol{H}}_s} \end{bmatrix}, \tag{G.2}$$

where $\boldsymbol{U}$ is the 6-dimensional identity dyadic, and the block-partitioned dyadic $\boldsymbol{A}$ raised to a positive integer k is defined as the continued direct product $\boldsymbol{A} \cdot \boldsymbol{A} \cdot \ldots \cdot \boldsymbol{A}$, performed k times. Accordingly, we have

$$\begin{aligned}
\boldsymbol{A}^2 &= \begin{bmatrix} \widetilde{\hat{\boldsymbol{H}}_s} \cdot \widetilde{\hat{\boldsymbol{H}}_s} + \tan^2\Lambda'\widetilde{\hat{\boldsymbol{a}}} \cdot \widetilde{\hat{\boldsymbol{a}}} & -\tan\Lambda'(\widetilde{\hat{\boldsymbol{H}}_s} \cdot \widetilde{\hat{\boldsymbol{a}}} + \widetilde{\hat{\boldsymbol{a}}} \cdot \widetilde{\hat{\boldsymbol{H}}_s}) \\ \\ -\tan\Lambda'(\widetilde{\hat{\boldsymbol{H}}_s} \cdot \widetilde{\hat{\boldsymbol{a}}} + \widetilde{\hat{\boldsymbol{a}}} \cdot \widetilde{\hat{\boldsymbol{H}}_s}) & \widetilde{\hat{\boldsymbol{H}}_s} \cdot \widetilde{\hat{\boldsymbol{H}}_s} + \tan^2\Lambda'\widetilde{\hat{\boldsymbol{a}}} \cdot \widetilde{\hat{\boldsymbol{a}}} \end{bmatrix} \\
&= \begin{bmatrix} (\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s - \boldsymbol{U}) + \tan^2\Lambda'(\hat{\boldsymbol{a}}\hat{\boldsymbol{a}} - \boldsymbol{U}) & -\tan\Lambda'(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}} + \hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s - 2(\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{a}})\boldsymbol{U}) \\ \\ -\tan\Lambda'(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}} + \hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s - 2(\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{a}})\boldsymbol{U}) & (\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s - \boldsymbol{U}) + \tan^2\Lambda'(\hat{\boldsymbol{a}}\hat{\boldsymbol{a}} - \boldsymbol{U}) \end{bmatrix} \\
&= \begin{bmatrix} \hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s + \tan^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}} & -\tan\Lambda'(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}} + \hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s) \\ \\ -\tan\Lambda'(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}} + \hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s) & \hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s + \tan^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}} \end{bmatrix} \\
&+ \begin{bmatrix} -(1+\tan^2\Lambda')\boldsymbol{U} & 2\tan\Lambda'(\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{a}})\boldsymbol{U} \\ \\ 2\tan\Lambda'(\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{a}})\boldsymbol{U} & -\left(1+\tan^2\Lambda'\right)\boldsymbol{U} \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} \hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s + \tan^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}} & -\tan\Lambda'(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}} + \hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s) \\ -\tan\Lambda'(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}} + \hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s) & \hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s + \tan^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}} \end{bmatrix} - \left(1+\tan^2\Lambda'\right)\boldsymbol{U} + 2\tan\Lambda'(\hat{\mathbf{z}}\cdot\hat{\boldsymbol{a}})\boldsymbol{J}, \tag{G.3}$$

where

$$\boldsymbol{J} = \begin{bmatrix} \mathbf{0} & \boldsymbol{U} \\ \boldsymbol{U} & \mathbf{0} \end{bmatrix}, \tag{G.4}$$

in which $\boldsymbol{U}$ and $\mathbf{0}$ are the 3×3 identity and zero dyadics, respectively. It should be noted that we use $\boldsymbol{U}$ to represent both the 3- and 6-dimensional identity dyadic, as the distinction should be evident. Note that $\boldsymbol{J}$ raised to an even power is $\boldsymbol{U}$ and to an odd power is $\boldsymbol{J}$. Also, the dyadics $\boldsymbol{A}$ and $\boldsymbol{A}^2$ satisfy the conditions $\boldsymbol{A}\cdot\boldsymbol{J} = \boldsymbol{J}\cdot\boldsymbol{A}$ and $\boldsymbol{A}^2\cdot\boldsymbol{J} = \boldsymbol{J}\cdot\boldsymbol{A}^2$

The general powers of $\boldsymbol{A}$ can be represented as

$$\boldsymbol{A}^n = \begin{cases} (-1)^{k-1}\left[(1+\tan^2\Lambda')\boldsymbol{U} - 2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right]^{k-1}\cdot\boldsymbol{A} & \text{if } n = 2k-1 \\ (-1)^{k-1}\left[(1+\tan^2\Lambda')\boldsymbol{U} - 2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right]^{k-1}\cdot\boldsymbol{A}^2 & \text{if } n = 2k \end{cases}, \tag{G.5}$$

where k is a positive integer. Substituting Equation G.5 into the STM power series definition, we find that the state transition dyadic reduces to a linear combination of the first powers of $\boldsymbol{A}$, namely the zeroth order power of $\boldsymbol{A}$ (since $\boldsymbol{A}^0 = \boldsymbol{U}$), $\boldsymbol{A}$, and $\boldsymbol{A}^2$ as

$$\begin{aligned} \boldsymbol{\Phi}(f) = \boldsymbol{U} + \boldsymbol{A}\cdot\Bigg(f\boldsymbol{U} - \frac{f^3}{3!}\left[(1+\tan^2\Lambda')\boldsymbol{U} - 2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right] + \frac{f^5}{5!}\left[(1+\tan^2\Lambda')\boldsymbol{U}\right. \\ \left. -2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right]^2 - \frac{f^7}{7!}\left[(1+\tan^2\Lambda')\boldsymbol{U} - 2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right]^3 + \cdots\Bigg) \\ + \boldsymbol{A}^2\cdot\Bigg(\frac{f^2}{2!}\boldsymbol{U} - \frac{f^4}{4!}\left[(1+\tan^2\Lambda')\boldsymbol{U} - 2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right] + \frac{f^6}{6!}\left[(1+\tan^2\Lambda')\boldsymbol{U}\right. \\ \left. -2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right]^2 - \frac{f^8}{8!}\left[\left(1+\tan^2\Lambda'\right)\boldsymbol{U} - 2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right]^3 + \cdots\Bigg). \end{aligned} \tag{G.6}$$

The general term that describes the powers of $\boldsymbol{A}$ is

$$\left[(1+\tan^2\Lambda')\boldsymbol{U} - 2\tan\Lambda'(\hat{\boldsymbol{H}}_s\cdot\hat{\boldsymbol{a}})\boldsymbol{J}\right]^{k-1}. \tag{G.7}$$

Using $1+\tan^2\Lambda' = 1/\cos^2\Lambda'$ and $2\sin\Lambda'\cos\Lambda' = \sin(2\Lambda')$, we can extract and rewrite this general term as

$$\frac{1}{\cos^{2(k-1)}\Lambda'}\left[\boldsymbol{U} - \sin(2\Lambda')(\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{a}})\boldsymbol{J}\right]^{k-1}. \tag{G.8}$$

We will denote $\nu = \sin(2\Lambda')(\hat{\boldsymbol{H}}_s \cdot \hat{\boldsymbol{a}})$, where $|\nu| \leq 1$, and find a simpler form for the auxiliary matrix

$$\boldsymbol{U} - \nu\boldsymbol{J}. \tag{G.9}$$

This 6-dimensional matrix is real and symmetric and thus diagonalizable over the field $\boldsymbol{R}$ of real numbers. This means that all eigenvalues and associated eigenvectors are real, and its eigenvectors make up a basis of $\boldsymbol{R}^6$. Moreover, the eigenvectors corresponding to distinct eigenvalues are mutually orthogonal with respect to the usual inner product in Euclidean vector space $\boldsymbol{R}^6$. In fact, the eigenvectors can be chosen to be orthonormal.

The eigenvalues of $\boldsymbol{U} - \nu\boldsymbol{J}$ have the general form $1 \mp \nu$ and have a multiplicity of 3 each, and its orthonormalized eigenvectors arranged in column form (as a block-partitioned dyadic) are

$$\boldsymbol{T} = \frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{U} & \boldsymbol{U} \\ \\ \boldsymbol{U} & -\boldsymbol{U} \end{bmatrix}. \tag{G.10}$$

Accordingly, the matrix $\boldsymbol{T}$ is orthogonal, since the set of its columns constitutes an orthonormal basis of Euclidean vector space $\boldsymbol{R}^6$; consequently, the inverse of $\boldsymbol{T}$ is the transpose, or conjugate, of $\boldsymbol{T}$: $\boldsymbol{T}^{-1} = \boldsymbol{T}^T$. It is easily seen that the value of its determinant is $\det \boldsymbol{T} = -1$, and thus $\boldsymbol{T}$ is a "reflection". This allows for a significant simplification in the original equation statement.

Using eigendecomposition,

$$\boldsymbol{T}^{-1} \cdot (\boldsymbol{U} - \nu\boldsymbol{J}) \cdot \boldsymbol{T} = \boldsymbol{D} = \begin{bmatrix} (1-\nu)\boldsymbol{U} & \boldsymbol{0} \\ \\ \boldsymbol{0} & (1+\nu)\boldsymbol{U} \end{bmatrix}, \tag{G.11}$$

where $0 \in \boldsymbol{R}^{3\times 3}$, and we can rewrite the auxiliary matrix in Equation G.9 as

$$\boldsymbol{U} - \nu\boldsymbol{J} = \boldsymbol{T} \cdot \boldsymbol{D} \cdot \boldsymbol{T}^{-1} = \boldsymbol{T} \cdot \begin{bmatrix} (1-\nu)\boldsymbol{U} & \boldsymbol{0} \\ \\ \boldsymbol{0} & (1+\nu)\boldsymbol{U} \end{bmatrix} \cdot \boldsymbol{T}, \tag{G.12}$$

having used the fact that $\boldsymbol{T}$ is orthogonal and symmetric.

Equation G.8 can then be written as

$$\frac{1}{\cos^{2(k-1)}\Lambda'}\left[\boldsymbol{U}-\nu\boldsymbol{J}\right]^{k-1}=\frac{1}{\cos^{2(k-1)}\Lambda'}\left(\boldsymbol{T}\cdot\begin{bmatrix}(1-\nu)\boldsymbol{U} & \boldsymbol{0}\\ \boldsymbol{0} & (1+\nu)\boldsymbol{U}\end{bmatrix}\cdot\boldsymbol{T}\right)^{k-1}$$
$$=\boldsymbol{T}\cdot\begin{bmatrix}\left(\frac{\sqrt{1-\nu}}{\cos\Lambda'}\right)^{2(k-1)}\boldsymbol{U} & \boldsymbol{0}\\ \boldsymbol{0} & \left(\frac{\sqrt{1+\nu}}{\cos\Lambda'}\right)^{2(k-1)}\boldsymbol{U}\end{bmatrix}\cdot\boldsymbol{T}. \qquad \text{(G.13)}$$

Combining Equations G.13 and G.5, and substituting the result into the STM power series definition, as in Equation G.6, gives

$$\boldsymbol{\Phi}(f)=\boldsymbol{U}+\boldsymbol{A}\cdot\boldsymbol{T}\cdot\begin{bmatrix}\frac{\cos\Lambda'}{\sqrt{1-\nu}}\left(\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)^{2k+1}\right)\boldsymbol{U} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{\cos\Lambda'}{\sqrt{1+\nu}}\left(\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)^{2k+1}\right)\boldsymbol{U}\end{bmatrix}\cdot\boldsymbol{T}$$
$$+\boldsymbol{A}^2\cdot\boldsymbol{T}\cdot\begin{bmatrix}\frac{\cos^2\Lambda'}{1-\nu}\left(1-\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!}\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)^{2k}\right)\boldsymbol{U} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{\cos^2\Lambda'}{1+\nu}\left(1-\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!}\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)^{2k}\right)\boldsymbol{U}\end{bmatrix}\cdot\boldsymbol{T}. \qquad \text{(G.14)}$$

Recalling the infinite series formulas for sine and cosine, the STM can be further reduced:

$$\boldsymbol{\Phi}(f)=\boldsymbol{U}+\boldsymbol{A}\cdot\boldsymbol{T}\cdot\begin{bmatrix}\frac{\cos\Lambda'}{\sqrt{1-\nu}}\sin\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)\boldsymbol{U} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{\cos\Lambda'}{\sqrt{1+\nu}}\sin\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\boldsymbol{U}\end{bmatrix}\cdot\boldsymbol{T}$$
$$+\boldsymbol{A}^2\cdot\boldsymbol{T}\cdot\begin{bmatrix}\frac{\cos^2\Lambda'}{1-\nu}\left(1-\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)\right)\boldsymbol{U} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{\cos^2\Lambda'}{1+\nu}\left(1-\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right)\boldsymbol{U}\end{bmatrix}\cdot\boldsymbol{T}. \qquad \text{(G.15)}$$

Finally, substituting Equations G.2 and G.3 for $\boldsymbol{A}$ and $\boldsymbol{A}^2$, respectively, and Equation G.10 for T, the STM reduces to elementary circular functions:

$$\boldsymbol{\Phi}(f)=\frac{1}{2}\begin{bmatrix}\left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)+\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right)\boldsymbol{U} & \left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)-\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right)\boldsymbol{U}\\ \left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)-\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right)\boldsymbol{U} & \left(\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)+\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right)\boldsymbol{U}\end{bmatrix}$$

$$
+\begin{bmatrix}
F_1\left(\cos^2\Lambda'\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s+\sin^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}\right)-F_2\sin\Lambda'\cos\Lambda'\left(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}}+\hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s\right) & \\
F_2\left(\cos^2\Lambda'\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s+\sin^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}\right)-F_1\sin\Lambda'\cos\Lambda'\left(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}}+\hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s\right) & \\
& F_2\left(\cos^2\Lambda'\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s+\sin^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}\right)-F_1\sin\Lambda'\cos\Lambda'\left(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}}+\hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s\right) \\
& F_1\left(\cos^2\Lambda'\hat{\boldsymbol{H}}_s\hat{\boldsymbol{H}}_s+\sin^2\Lambda'\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}\right)-F_2\sin\Lambda'\cos\Lambda'\left(\hat{\boldsymbol{H}}_s\hat{\boldsymbol{a}}+\hat{\boldsymbol{a}}\hat{\boldsymbol{H}}_s\right)
\end{bmatrix}
$$

$$
+\begin{bmatrix}
-G_1\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_2\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}} & -G_2\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_1\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}} \\
-G_2\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_1\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}} & -G_1\cos\Lambda'\widetilde{\hat{\boldsymbol{H}}_s}+G_2\sin\Lambda'\widetilde{\hat{\boldsymbol{a}}}
\end{bmatrix}, \tag{G.16}
$$

where the coefficients F_1, F_2, G_1, and G_2 are scalar functions of ν, f, and Λ', given by

$$
F_1=\frac{1}{2}\left(\frac{2}{1-\nu^2}-\frac{1}{1-\nu}\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)-\frac{1}{1+\nu}\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right), \tag{G.17}
$$

$$
F_2=\frac{1}{2}\left(\frac{2\nu}{1-\nu^2}-\frac{1}{1-\nu}\cos\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)+\frac{1}{1+\nu}\cos\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right), \tag{G.18}
$$

$$
G_1=\frac{1}{2}\left(\frac{1}{\sqrt{1-\nu}}\sin\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)+\frac{1}{\sqrt{1+\nu}}\sin\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right), \tag{G.19}
$$

$$
G_2=\frac{1}{2}\left(\frac{1}{\sqrt{1-\nu}}\sin\left(\frac{f\sqrt{1-\nu}}{\cos\Lambda'}\right)-\frac{1}{\sqrt{1+\nu}}\sin\left(\frac{f\sqrt{1+\nu}}{\cos\Lambda'}\right)\right). \tag{G.20}
$$