Mathematical Problems in Engineering

Theory, Methods, and Applications

Editor-in-Chief: Jose Manoel Balthazar

Special Issue
Space Dynamics

Guest Editors: Antonio F. Bertachini A. Prado, Maria Cecilia Zanardi, Tadashi Yokoyama, and Silvia Maria Giuliatti Winter

Hindawi Publishing Corporation
http://www.hindawi.com

Volume 2009

Space Dynamics

Mathematical Problems in Engineering

Space Dynamics

Guest Editors: Antonio F. Bertachini A. Prado, Maria Cecilia Zanardi, Tadashi Yokoyama, and Silvia Maria Giuliatti Winter

Contents

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 732758, 7 pages
doi:10.1155/2009/732758

Editorial

Space Dynamics

Antonio F. Bertachini A. Prado,[1] Maria Cecilia Zanardi,[2] Tadashi Yokoyama,[3] and Silvia Maria Giuliatti Winter[2]

[1] *INPE-DMC, Brazil*
[2] *FEG-UNESP, Guaratinguetá, Brazil*
[3] *UNESP, Campus de Rio Claro, Brazil*

Correspondence should be addressed to Antonio F. Bertachini A. Prado, prado@dem.inpe.br

Received 31 December 2009; Accepted 31 December 2009

The space activity in the world is one of the most important achievements of mankind. It makes possible live communications, exploration of Earth resources, weather forecast, accurate positioning and several other tasks that are part of our lives today.

The space dynamics plays a very important rule in these developments, since its study allows us to plan how to launch and control a space vehicle in order to obtain the results we need.

This field considers the study of Celestial Mechanics and Control applied to spacecraft and natural objects. The main tasks are to determine the orbit and the attitude of the spacecraft based in some observations, to obtain its position and attitude in space in a given time from some initial conditions, to find the best way to change their orbits and attitude, to analyze how to use the information of the satellites to find the position and the velocity of a given point (e.g., a personal receptor, a satellite or a car), etc.

This field of study comes from Astronomy. The main contributors from the past have important names like Johannes Kepler (1571–1630) and Isaac Newton (1642–1727). Based on the observations of the motion of the planets realized by Tycho Brahe (1546–1601), Kepler formulated the three basic laws, which govern the motion of the planets around the Sun. From these laws, Newton formulated the universal Law of Gravitation. According to this Law, mass attracts mass in a ratio that is proportional to the product of the two masses involved and inversely proportional to the square of the distance between them. Those laws are the scientific bases of the space exploration age that officially begin with the launch of the satellite Sputnik in 1957 by the former Soviet Union. Since then, a strong battle between the United States of America (USA) and the Soviet Union took place leading to many achievements in Space. One of the most important results was the landing of the man on the Moon, achieved

by the USA in 1969. From this point, several different applications of the space research were developed, changing for better the human life on Earth.

In that scope, this special issue of Mathematical Problems in Engineering is focused on the recent advances in space dynamics techniques. It has a total of 21 papers that are briefly described below.

Four of them are concerned with the attitude motion, control and determination. Optimal On-Off Attitude Control for the Brazilian Multimission Platform Satellite by G. Arantes Jr. et al. is the first one. This work deals with the analysis and design of the reaction thruster attitude control for the Brazilian Multi-Mission Platform satellite. The aim of this work is to provide smoother control for improved pointing requirements with less thruster activation or propellant consumption. The fuel is a deciding factor of the lifetime of the spacecraft and reduced propellant consumption is highly required, specially, regarding a multi-mission spacecraft wherein different payloads are being considering. The three-axis attitude control is considered and it is activated in pulse mode. Consequently a modulation of the torque command is compelling in order to avoid high non-linear control action. The paper considers the Pulse-Width Pulse-Frequency (PWPF) modulator, composed of a Schmidt trigger, a first order filter, and a feedback loop. This modulator holds several advantages over classical bang-bang controllers such as close to linear operations, high accuracy, and reduced propellant consumption. The Linear Gaussian Quadratic (LQG) technique is used to synthesize the control law during stabilization mode and the modulator is used to modulate the continuous control signal to discrete one. The results of the numerical simulations show that the obtained on-off thruster reaction attitude control system, based on the LQG/PWPF modulation, is optimal with respect to the minimization of the quadratic cost function of the states and control signals and propellant consumption. The paper presents a set of optimal parameter for the PWPF modulator by considering static and dynamics analysis. The obtained results demonstrate the feasibility of combining LQG/PWPF modulator in a unique controller for on-off thruster reaction attitude control system. Stability remains by adding the PWPF modulator and reasonable accuracy in attitude is achieved. Practical aspects are included in this study as filtering and presence of external impulsive perturbations. The advantages of less spent propellant shall contribute to the Brazilian Multi-Mission Platform project, specially, a satellite conceived to be used on a large number and different types of missions, in the context of an ever-advancing Brazilian space program.

The second paper on this subject is Highly Efficient Sigma Point Filter for Spacecraft Attitude and Rate Estimation by C. Fan and Z. You. In this paper, for spacecraft attitude determination problem, the multiplicative extended Kalman filter MEKF and other similar algorithms, have been good solutions for most nominal space missions. However, nowadays, due to their overload computational complexity, they are prohibitive for actual on board implementation. In this paper, the authors present a new and quite competitive algorithm, with significant lower computational complexity even when compared to the reduced sigma point algorithms. The precision is the same as the traditional unscented Kalman filters. In terms of efficiency, the proposed algorithm rivals MEKF, even in severe situations.

The next one is Spin-Stabilized Spacecraft: Analytical Attitude Propagation Using Magnetic Torques by R. V. Garcia et al.. This paper considers the problem in obtaining the attitude of a satellite in a given time based on information from a previous time. It analyzes the rotational motion of a spin stabilized Earth artificial satellite. It makes derivation of an analytical attitude prediction. Particular attention is given to torques, which come from residual magnetic and eddy currents perturbations, as well as their influences on the satellite

angular velocity and space orientation. A spherical coordinated system, fixed in the satellite, is used to locate the spin axis of the satellite in relation to the terrestrial equatorial system.

The last paper of this topic is Using of H-Infinity Control Method in Attitude Control System of Rigid-Flexible Satellite by X. C. M. Cubillos and L. C. G. Souza. This paper considers the attitude control systems of satellites with rigid and flexible components. In the current space missions, this problem is demanding a better performance, which implies in the development of several methods to approach this problem. For this reason, the methods available today need more investigation in order to know their capability and limitations. Therefore, in this paper, the H–Infinity method is studied in terms of the performance of the Attitude Control System of a Rigid-Flexible Satellite.

There were four papers studying the problem of finding space trajectories. The first one is Hill Problem Analytical Theory to the Order Four: Application to the Computation of Frozen Orbits around Planetary Satellites by M. Lara and J. F. Palacián. In this paper, applications to the computation of frozen orbits around planetary satellites are made. The Hill problem, a simplified model of the restricted three-body problem, also gives a very good approximation for the dynamics involving the motion of natural and artificial satellites, moons, asteroids and comets. Frozen orbits in the Hill problem are determined through the double averaged problem. The developed method provides the explicit equations of the transformation connecting averaged and non averaged models, making the computation of the frozen orbits straightforward.

The second one covering this topic is Collision and Stable Regions around Bodies with Simple Geometric Shape by A. A. Silva et al.. Collision and stable regions around bodies with simple geometric shape are studied. The gravitational potential of two simple geometric shapes, square and triangular plates, were obtained in order to study the orbital motion of a particle around them. Collision and stable regions were also derived from the well known Poincaré surface of section. These results can be applied to a particle in orbit around an irregular body, such as an asteroid or a comet.

The next paper is Dynamical Aspects of an Equilateral Restricted Four-Body Problem by M. Álvarez-Ramírez and C. Vidal. It is an immediate extension of the classical restricted three body problem (ERFBP): a particle is under the attraction of three nonzero masses (m_1, m_2, m_3) which move on circular orbits around their center of mass, fixed at the origin of the coordinate system in a such way that their configuration is always an equilateral triangle. In particular, it is assumed $m_2 = m_3$. In a synodical system, a first integral of the problem is obtained. Using Hamiltonian formalism the authors define Hill's regions. Equilibrium solutions are obtained for different cases and the number of them depends on the values of the masses. The Lyapunov stability of these solutions is studied in the symmetrical case assuming $m_1 = m_2 = m_3 = \mu$. Under certain conditions and for very small μ, circular and elliptic keplerian periodic solutions can be continued to ERFBP. For $\mu = 1/2$, Lyapunov Central theorem can provide a one-parameter family of periodic orbits. Some numerical applications are also shown.

The last one in this category is Nonsphericity of the Moon and Near Sun-Synchronous Polar Lunar Orbits by J. P. S. Carvalho et al.. Here, the dynamics of a lunar artificial satellite perturbed by the nonuniform distribution of mass of the Moon taking into account the oblateness (J_2) and the equatorial ellipticity (sectorial term C_{22}) is presented. A canonical perturbation method based on Lie-Hori algorithm is used to obtain the second order solutions. A study is performed for the critical inclination and the effect of the coupling terms J_2 and C_{22} are presented. A new second order formula is obtained for the critical inclination as a function of the argument of the pericenter and of the longitude of the ascending node. In the

same way, for Lunar Sun-synchronous and Near-Polar Orbits, a new formula is obtained to provide the value of the inclination. This formula depends on the semi-major axis, eccentricity and the longitude of the ascending node. For Lunar low altitude satellites, the authors call the attention for the importance of the additional harmonics J_3, J_5, and C_{31}, besides J_2 and C_{22}. In particular they mention that, for small inclinations, some contributions of the second order terms can become as large as the first order terms. Several numerical simulations are presented to illustrate the time variation of the eccentricity and inclination.

After that, there are five papers considering the problem of localization with information obtained from space, in particular using GPS and/or GLONASS constellations. The first paper of this topic is GPS Satellites Orbits: Resonance by L. D. D. Ferreira and R. V. Moraes. In this paper, the effects of the perturbations due to resonant geopotential harmonics on the semi major axis of GPS satellites are analyzed. The results show that it is possible to obtain secular perturbations of about 4m/day using numerical integration of the Lagrange planetary equations and considering, in the disturbing potential, the main secular resonant coefficients. The paper also shows the amplitudes for the long period terms due to the resonant coefficients for some hypothetical satellites orbiting in the neighborhood of the GPS satellites orbits. The results can be used to perform orbital maneuvers of the GPS satellites to keep them in their nominal orbits.

The second paper is Some Initial Conditions for Disposed Satellites of the Systems GPS and Galileo Constellations by D. M. Sanchez et al.. In this paper the stability of the disposed objects of the GPS and Galileo systems can be affected by the increasing in their eccentricities due to strong resonances. A search for initial conditions where the disposed objects remain at least 250 years, without crossing the orbits of the operational satellites, was performed. As a result, regions where the values of the eccentricity prevent possible risk of collisions have been identified in the phase space. The results also show that the initial inclination of the Moon plays an important role in searching these initial conditions.

Then, we have Quality of TEC Estimated with Mod Ion Using GPS and GLONASS Data by P. O. Camargo. The largest source of error in positioning and navigation with the Global Navigation Satellite System (GNSS) is the ionosphere, which depends on the Total Electron Content (TEC). The quality of the TEC was analyzed taking into account the ModIon model developed in UNESP-Brazil the more appropriate model to be used in the South America region.

After that, we have the paper The Impact on Geographic Location Accuracy due to Different Satellite Orbit Ephemeredes by C. C. Celestino et al.. Here, it is assumed that there are several satellites, hundreds of Data Collection Platforms (DCPs) deployed on ground (fixed or mobile) of a large country (e.g. Brazil), and also some ground reception stations. It considers the question of obtaining the geographic location of these DCPs. In this work, the impact on the geographic location accuracy, when using orbit ephemeris obtained through several sources, is assessed. First, by this evaluation is performed by computer simulation of the Doppler data, corresponding to real existing satellite passes. Then, real Doppler data are used to assess the performance of the location system. The results indicate that the use of precise ephemeris can improve the performance of the calculations involved in this process by reducing the location errors. This conclusion can then be extended to similar location systems.

There is also the paper Simulations under Ideal and Non ideal Conditions for Characterization of a Passive Doppler Geographical Location System Using Extension of Data Reception Network by C. T. Sousa et al.. It presents a Data Reception Network (DRN) software investigation to characterize the passive Doppler Geographical Location (GEOLOC)

software. The test scenario is composed by Brazilian Data Collection Satellite (SCD2) and the National Oceanic Atmospheric Administration satellite (NOAA-17) passes, a single Data Collecting Platform (DPC) and five ground received stations. The Doppler measurements data of a single satellite pass over a DCP, considering a network of ground reception stations, is the rule of the DNR. The DNR uses an ordering selection method that merges the collected Doppler shift measurements through the stations network in a single file. The pre-processed and analyzed measurement encompasses the DCP signal transmission time and the Doppler shifted signal frequency received on board of the satellite. Thus, the assembly to a single file of the measurements collected, considering a given satellite pass, will contain more information about the full Doppler effect behavior while decreasing the amount of measurement losses, as a consequence, an extended visibility between the relay satellite and the reception stations. The results and analyses were firstly obtained considering the ground stations separately, to characterize their effects in the geographical location result. Six conditions were investigated: ideal simulated conditions, random and bias errors in the Doppler measurements, errors in the satellites ephemeris and errors in the time stamp. To investigate the DNR importance to get more accurate locations, an analysis was performed considering the random errors of 1 HZ in the Doppler measurements. The results show that the developed GEOLOC is operating appropriately under the ideal conditions. The inclusion of biased errors degrades the location results more than the random errors. The random errors are filtered out by the least squares algorithm and they produce mean locations results that tend to zero error, mainly for high sampling rate. The simulations results, considering biased errors, yield errors that degrade the location for high and low sampling rates. The simulation results for ephemeredes error shows that it is fundamental to minimize them, because the location system cannot compensate these errors. The satellites ephemeredes errors are approximately similar in magnitude to their resulting transmitter location errors. The simulations results, using the DRN algorithm, show that to improve the locations results quality it would be necessary to have more Reception Stations spread over the Brazilian territory, to obtain additional amount of data. Then, on the other hand, it improves the geometrical coverage between satellite and DCPs, and better recovers the full Doppler curves, yielding, as a consequence, more valid and improved locations.

A similar problem, but concerned with the determination of an orbit of a satellite, is considered in A Discussion Related to Orbit Determination Using Nonlinear Sigma Point Kalman Filter by P. C. P. M. Pardal et al.. The goal of this work is to present a Kalman filter based on the sigma point unscented transformation, aiming at real-time satellite orbit determination using GPS measurements. Firstly, some underlying material is briefly presented before introducing SPKF (sigma point Kalman filter) and the basic idea of the unscented transformation in which this filter is based. Through the paper, the formulation about orbit determination via GPS, dynamic and observation models and unmodeled acceleration estimation are presented. The SPKF is investigated in many different applications and the results are discussed. The advantages indicate that SPKF can be used as an emerging estimation algorithm to nonlinear system.

Orbital maneuvers for space vehicles are also considered in three papers, as in Orbital Dynamics of a Simple Solar Photon Thruster by A. D. Guerman et al.. This paper studies the orbital dynamics and control for two systems of solar propulsion, a flat solar sail (FSS) and a simple solar photon thruster (SPT). The use of solar pressure to create propulsion can minimize the spacecraft on-board energy consumption during the mission. Modern materials and technologies made this propulsion scheme feasible, and many projects of solar sail are now under development, making the solar sail dynamics the subject of numerous studies.

To perform the analysis presented in this paper, the equations of the sailcraft's motion are deduced. Comparisons for the performance of two schemes of solar propulsion (Simple Solar Photon Thruster—SSPT and Dual Reflection Solar Photon Thruster—DRSPT) are shown for two test time-optimal control problems of trajectory transfer (Earth-Mars transfer and Earth-Venus transfer). The mathematical model for the force acting on SSPT due to the solar radiation pressure takes into account multiple reflections of the light flux on the sailcraft elements. In this analysis it is assumed that the solar radiation pressure follows inverse-square variation law, the only gravitational field is the one from the Sun (central Newtonian), and the sails are assumed to be ideal reflectors. For a planar motion of an almost flat sail with negligible attitude control errors, the SSPT equations of motion are similar to those for a DRSPT. The analysis showed a better performance of SPT in terms of response time and the results are more pronounced for Earth-Venus transfer. It can be explained by the greater values of the transversal component of the acceleration developed by SSPT compared to FSS.

Then, we have the paper Alternative Transfers to the NEOs 99942 Apophis, 1994 WR12, and 2007 UW1 via Derived Trajectories from Periodic Orbits of Family G by C. F. Melo et al.. This paper explores the existence of a natural and direct link between low Earth orbits and the lunar sphere of the influence to get low-energy transfer trajectory to the three Near Earth Objects through swing-bys with the Moon. The existence of this link is related to a family of retrograde periodic orbits around the Lagrangian equilibrium point L_1 predicted by the circular, planar, restricted three-body Earth-Moon-particle problem. Such orbits belong to the so-called Family G. The trajectories in this link are sensitive to small disturbances. This enables them to be conveniently diverted, reducing the cost of a swing-by maneuver. These maneuvers allow a gain in energy enough for the trajectories to escape from the Earth-Moon system and to be stabilized in heliocentric orbits between Earth and Venus or Earth and Mars. The result shows that the required increment of velocity by escape trajectories G is, in general, fewer than the ones required by conventional transfer (Patched-conic), between 2% up to 4%. Besides, the spacecraft velocities relative to the asteroids are also, in general, less than that value obtained by the conventional methods. In terms of the transfer time, the results show that in the Apophis and 1994WR12 it is possible to find Closest Point Approaches. The longest time always corresponds to the smallest relative velocity in Closest Point Approaches for trajectories G. Therefore, the trajectories G can intercept the Near Earth Objects orbits and, they can be a good alternative to design future missions destined to the Near Earth Objects.

After that, we have the paper Controlling the Eccentricity of Polar Lunar Orbits with Low-Thrust Propulsion by O. C. Winter et al.. This paper approaches the problem that lunar satellites in polar orbits suffer a high increase on the eccentricity, due to the gravitational perturbation of the Earth leading them to a collision with the Moon. Then, the control of the orbital eccentricity leads to the control of the satellite's lifetime. This paper introduces an approach in order to keep the orbital eccentricity of the satellite at low values. The method presented in the paper considers two systems: the 3-body problem, Moon-Earth-satellite and the 4-body problem, Moon-Earth-Sun-satellite. A system considering a satellite with initial eccentricity equals to 0.0001 and a range of initial altitudes, between 100 km and 5000 km, is considered. An empirical expression for the length of time needed to occur the collision with the Moon as a function of the initial altitude is derived. The results found for the 3-body model were not significantly different from those found for the 4-body model. After that, using low thrust propulsion, it is introduced a correction of the eccentricity every time it reaches the value 0.05.

Mechanical aspects of spacecrafts are considered in two papers. The first one is Internal Loading Distribution in Statically Loaded Ball Bearings Subjected to an Eccentric Thrust Load

by M. C. Ricci. In this paper an iterative method is introduced to calculate internal normal ball loads in statically loaded single-row, angular-contact ball bearings, subjected to a known thrust load which is applied to a variable distance from the geometric bearing center line. Numerical examples are shown and compared with the literature. Fifty figures are presented and the results are discussed.

The other paper is The Determination of the Velocities after Impact for the Constrained Bar Problem by A. Fenili et al.. In this paper, a mathematical model for a constrained manipulator is studied. Despite the fact that the model is simple, it has all the important features of the system. A fully plastic impact is considered. Analytical expressions for the velocities of the bodies involved after the collision are derived and used for the numerical integrations of the equations of motion. The theory presented in the paper can be used to problems where the robots have to follow some prescribed patterns or trajectories when in contact with the environment.

One paper deals with the astronomical side of the space dynamics: Gravitational Capture of Asteroids by Gas Drag by E. Vieira-Neto and O. C. Winter. The orbital configuration of the irregular satellites, present in the giant planets system, suggests that these bodies were asteroids in heliocentric orbits that have been captured by the planets. Since this capture is temporary, it has been necessary a dissipative effect in order to turn this temporary capture into a permanent one. This paper deals with this problem by analyzing the effects of the gas drag, from the Solar Nebula, in the orbital evolution of these asteroids after they have being captured by the planets. The results show that, although this dissipative effect is important, it is not the only mechanism responsible for keeping the asteroids in a permanent orbit about the planet.

Then, we also have one paper studying the motion of a spacecraft when traveling in the atmospheric region of the space: Atmospheric Reentry Dynamics of Conic Objects by J. P. Saldia et al.. In this paper, the accurate determination of the aerodynamics coefficients is an important issue in the calculation of the reentry trajectories of an object inside the terrestrial atmosphere. The methodology to calculate these coefficients and how to include them in a code, in order to compute the reentry trajectories, is considered. As a result, a sample of trajectories of conical objects for different initial flight conditions is presented.

Acknowledgments

The guest editors would like to thank all the authors, the reviews, the Editor of the journal, and all the staff involved in the preparation of this issue for the opportunity to publish the articles related to this important subject.

Antonio F. Bertachini A. Prado
Maria Cecilia Zanardi
Tadashi Yokoyama
Silvia Maria Giuliatti Winter

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 750945, 17 pages
doi:10.1155/2009/750945

Research Article

Optimal On-Off Attitude Control for the Brazilian Multimission Platform Satellite

Gilberto Arantes Jr.,[1] Luiz S. Martins-Filho,[2] and Adrielle C. Santana[3]

[1] *Center for Applied Space Technology and Microgravity, (ZARM) Am Fallturm, University of Bremen, 28359 Bremen, Germany*
[2] *Center of Engineering, Modelling and Applied Social Sciences, Federal University of ABC, Rua Catequese 242, 09090-400 Santo André, SP, Brazil*
[3] *Graduate Program on Information Engineering, Federal University of ABC, Rua Catequese 242, 09090-400 Santo André, SP, Brazil*

Correspondence should be addressed to Luiz S. Martins-Filho, luizsmf@gmail.com

Received 24 June 2009; Accepted 13 August 2009

Recommended by Maria Zanardi

This work deals with the analysis and design of a reaction thruster attitude control for the Brazilian Multimission platform satellite. The three-axis attitude control systems are activated in pulse mode. Consequently, a modulation of the torque command is compelling in order to avoid high nonlinear control action. This work considers the Pulse-Width Pulse-Frequency (PWPF) modulator which is composed of a Schmidt trigger, a first-order filter, and a feedback loop. PWPF modulator holds several advantages over classical bang-bang controllers such as close to linear operation, high accuracy, and reduced propellant consumption. The Linear Gaussian Quadratic (LQG) technique is used to synthesize the control law during stabilization mode and the modulator is used to modulate the continuous control signal to discrete one. Numerical simulations are used to analyze the performance of the attitude control. The LQG/PWPF approach achieves good stabilization-mode requirements as disturbances rejection and regulation performance.

1. Introduction

One of the intentions of this work is to support the ongoing Brazilian multimission platform (MMP) satellite project [1]. The project takes into consideration a special platform satellite which can supply multimissions capabilities supporting different payloads to lift up on the platform. Applications including Earth observation, communication, scientific experiments, and surveillance are few examples of suitable use of the MMP satellite. The MMP adopted pulse or on-off reaction thruster for attitude maneuvers, therefore, modulating continuous command signal to an on-off signal is a required task. Selecting the properly method to

modulate the control command signal is a key assignment. The aim of this work is to provide smoother control for improved pointing requirements with less thruster activation or propellant consumption. The fuel is a deciding factor of the lifetime of the spacecraft and reduced propellant consumption is highly required, specially, regarding a multimission spacecraft wherein different payloads are being considered.

In this paper a pulse-width pulse-frequency (PWPF) modulator is considered as a feasible option for the MMP reaction thruster modulation due to advantages over other types of pulse modulators as bang-bang controllers which has excessive thruster actuation [2, 3]. The PWPF modulator translates the continuous commanded control/torque signal to an on-off signal. Its behavior is a quasilinear mode which is possible by modulating the width of the activated reaction pulse proportionally to the level of the torque command input (pulse-width) and also the distance between the pulses (pulse-frequency). A PWPF modulator is composed of a Schmidt trigger, a lag network filter, and a feedback loop. The PWPF design requires iterative tuning of lag filter and Schmidt trigger. The optimal parameters achievement is based on the static (test signals) and dynamic (feedback signals) simulation results. The optimality is in respect to either the number of firings or spent fuel. The work in [3–5] provides good guidelines for the PWPF tuning task.

The PWPF is synthesized with a Linear Quadratic Gaussian (LQG) controller which is designed for the MIMO attitude system. The LQG controller, refered to as $\mathscr{H}_2$, allows a tradeoff between regulation performance and control effort. In order to reduce the control effort or fuel consumption, an iteratively searching of the trade-off can be carried out. Nevertheless the controller has to attempt all the involved requirements and specifications. A previous study of the LQG approach applied to the MMP satellite is presented in [6]. The reaction attitude control system is applied to the stabilization mode of the MMP. The paper is divided into 5 sections. Section 2 presents the nonlinear model of the satellite, assumed a rigid body, its linearization around the operation point, and the developed virtual reality model of the satellite for visualization purposes. Section 3 presents a brief description of the PWPF modulator and design of the LQG controller, which includes the description of the LQG controller and provides the tuning parameters range for the PWPF modulator. Section 4 presents the numerical simulation for the reaction thruster attitude control system during the stabilization mode. Regulation, filtering, and disturbance rejection are investigated and discussed. Conclusions are presented in Section 5 based on the obtained results.

2. Problem Formulation

In this section we describe the mathematical model of the attitude motion, including kinematics, dynamics, and the linerization of the satellite model around LHLV reference frame. Based on that linear model the LQG controller is designed for the stabilization mode.

2.1. Satellite Attitude Model

The attitude of the satellite will be defined in this work by the orientation of the body frame (x, y, z) (coincident with the three principal axes of inertia) with respect to the orbital reference frame (x_r, y_r, z_r), also known as Local-Vertical-Local-Horizontal (LVLH) [7]. The origin of the orbit reference frame moves with the center of mass of the satellite in orbit. The z_r axis points toward the center of mass of the Earth, x_r axis is in the plane of the orbit, perpendicular to z_r, in the direction of the velocity of the spacecraft. The y_r axis is normal to

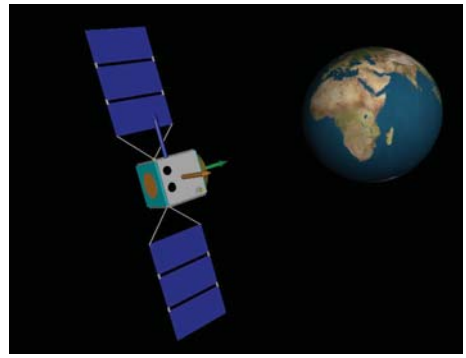


Figure 1: LVLH axis representation.

the local plane of the orbit, and completes a three-axis right-hand orthogonal system. Figure 1 illustrated the LVLH reference frame.

The attitude is represented by the direction cosine matrix R between body frame and reference frame. During the stabilization mode only small angular variations are considered, in this case the Euler angles parametrization is an appropriate choice due to the guarantee of nonsingularity. Thus, by using Euler angles (ϕ, θ, ψ) in an asymmetric sequence 3-2-1 (z-y-x) for describing a rotation matrix, one finds [7, 8]

$$R_{zyx} = \begin{bmatrix} c\psi c\theta & s\psi c\theta & -s\theta \\ -c\phi s\psi + s\phi s\theta c\psi & c\phi c\psi + s\phi s\theta s\psi & s\phi c\theta \\ s\phi s\psi + c\phi s\theta c\psi & -s\phi c\psi + c\phi s\theta s\psi & c\phi c\theta \end{bmatrix}. \tag{2.1}$$

For a rotating body the elements of the direction cosine matrix change with time, this change relative to any reference frame fixed in inertial space can be written as follows [9]:

$$\dot{R}(t) = S\left(\boldsymbol{\omega}_{ib}^{b}\right) R(t), \tag{2.2}$$

where $\boldsymbol{\omega}_{ib}^{b} = (\omega_x, \omega_y, \omega_z)^T$ is the angular velocity of the body frame relative to the inertial frame, expressed in the body frame, S is the skew-symmetric operator given by

$$S\left(\boldsymbol{\omega}_{ib}^{b}\right) = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{2.3}$$

According to [10] the angular velocity can be expressed as function of the mean orbital motion (ω_0) and the derivatives $(\dot{\phi}, \dot{\theta}, \dot{\phi})$, thereafter the kinematics of the rigid body is

expressed by

$$\boldsymbol{\omega}_{ib}^{b} = \begin{bmatrix} 1 & 0 & -s\theta \\ 0 & c\phi & s\phi c\theta \\ 0 & -s\phi & c\phi c\theta \end{bmatrix} \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} - \omega_0 \begin{bmatrix} c\theta s\psi \\ s\phi s\theta s\psi + c\phi c\psi \\ c\phi s\theta s\psi - s\phi c\psi \end{bmatrix} \tag{2.4}$$

since large slewing maneuvers of the satellite are not considered, it is save to approximate $c\theta \approx 1$, $s\theta \approx \theta$, $\phi\psi \approx 0$. According to (2.4) for small Euler angles, the kinematics can be approximated as

$$\boldsymbol{\omega}_{ib}^{b} = \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} + \omega_0 \begin{bmatrix} -\psi \\ -1 \\ \phi \end{bmatrix}. \tag{2.5}$$

The dynamics of a satellite attitude, equipped with six one-sides thrusters is modelled by using the Euler equations. Furthermore, the attitude dynamic is written in the body frame, it yields

$$\tau_{ext} = \left[\frac{d\mathbf{h}}{dt}\right]_b + \boldsymbol{\omega}_{ib}^{b} \times \mathbf{h}_b, \tag{2.6}$$

where $\mathbf{h}_b = J\boldsymbol{\omega}_{ib}^{b}$ is the momentum of the rigid body, J is the satellite inertia matrix, and $\boldsymbol{\tau}_{ext}$ are the external torques acting in the system including perturbation and thruster actuation. Using $[d\mathbf{h}/dt]_b = J\dot{\boldsymbol{\omega}}_{ib}^{b}$, (2.6) becomes

$$J\dot{\boldsymbol{\omega}}_{ib}^{b} + S\left(\boldsymbol{\omega}_{ib}^{b}\right) J\boldsymbol{\omega}_{ib}^{b} = \boldsymbol{\tau}_{d}^{b} + \boldsymbol{\tau}_{c}^{b}, \tag{2.7}$$

where $\boldsymbol{\tau}_{d}^{b}$ represents all the disturbance torques, for example, atmosphere drag, gravity gradient, and so on, and $\boldsymbol{\tau}_{c}^{b}$ represents the control torques used for controlling the attitude motion. The control torques about the body axes, x, y, and z provide by the thrusters are $\boldsymbol{\tau}_x, \boldsymbol{\tau}_y \boldsymbol{\tau}_z$, respectively. The thruster reaction system is discussed in detail in the following section. The torque effect caused by the gravity gradient is taken into account and it is included in the linearization process. An asymmetric body subject to a gravitational field experience a torque tending to align the axis of the least moment of inertia with the field direction [8]. For small angle maneuvers, the model of the gravity gradient torque is approximated as [8, 9]

$$\boldsymbol{\tau}_{g}^{b} = 3\omega_0^2 \begin{bmatrix} (J_z - J_y)\phi \\ (J_x - J_z)\theta \\ 0 \end{bmatrix}. \tag{2.8}$$

Substituting (2.4) into (2.6) and adding the control and gradient gravity torque, we linearize the satellite attitude model. Moreover, the linearization is performed around the LHLV orbital

frame, it is thus adopted for the stabilization mode. Afterwards the attitude model can be represented in the state space form [6, 10]

$$\begin{aligned} \dot{\mathbf{x}} &= A\mathbf{x} + B\mathbf{u}, \\ \mathbf{y} &= C\mathbf{x} + D\mathbf{u}, \end{aligned} \tag{2.9}$$

with states $\mathbf{x} = [\phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi}]^T$, and inputs $\mathbf{u} = [\tau_x, \tau_y, \tau_z]^T$.$A$ is the state matrix, B is the input matrix, C is the output matrix, and D is the direct transmission matrix. In the particular problem they are given by

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \dfrac{4\omega_0^2(J_z - J_y)}{J_x} & 0 & 0 & 0 & 0 & \dfrac{\omega_0(J_x - J_y + J_z)}{J_x} \\ 0 & \dfrac{3\omega_0^2(J_x - J_z)}{J_y} & 0 & 0 & 0 & 0 \\ 0 & 0 & \dfrac{\omega_0^2(J_x - J_y)}{J_z} & \dfrac{\omega_0(J_y - J_x - J_z)}{J_z} & 0 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \dfrac{l}{J_x} & 0 & 0 \\ 0 & \dfrac{l}{J_y} & 0 \\ 0 & 0 & \dfrac{l}{J_z} \end{bmatrix}, \qquad C = I_{6\times 6}, \qquad D = 0_{6\times 3}. \tag{2.10}$$

It is worth to note that x row and z yaw axes belong to a multi-input and multi-output (MIMO) system 4×2 and the y pitch axis could be dealt as a single input and single output system (SISO) by assuming a tachometry feedback control. Although the controller is project over the linear model, the nonlinear model is used in the simulations.

2.2. Virtual Reality Model of the Spacecraft

In this work a Virtual Reality (VR) model are developed as a visualization tool. The purpose is to visualize the simulations giving a fast and a visual feedback of the simulation models over time. The model is produced by using the virtual reality model language (VRML) format which includes a description of 3-dimensional scenes, sounds, internal actions, and WWW anchors. It enables us to view moving three-dimensional scenes driven by signals from the

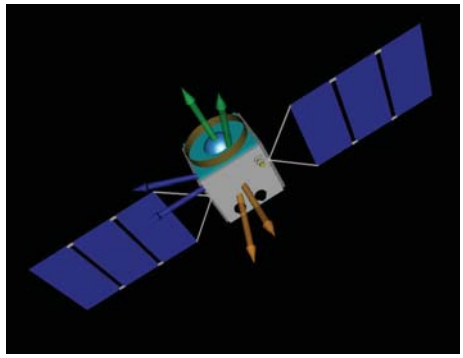


Figure 2: A graphical interface in VRML for visualization.

dynamic model, that is, attitude dynamics. The VR model was created with the use of V-Realm builder tool, a more detailed description can be found in [11]. Figure 2 shows the basic structure representation of the spacecraft's bus. The payload is not illustrated.

3. Thruster Attitude Control System

In this section the controller design based on the Linear Quadratic Gaussian (LQG) technique is briefly described, afterwards the PWPF modulator is presented in details.

3.1. LQG Controller Design

The Linear Quadratic Gaussian (LQG) or $\mathscr{H}_2$ control consist of a technique for designing optimal controllers. The approach is based on the search of the tradeoff between regulation performance of the states and control effort [12]. The referred optimality is expressed by a quadratic cost function and allows the designer to shape the principal gains of the return ration, at either the input or the output of the plant, to achieve required performance or robustness specifications. Moreover the method is easily designed for Multi-Input Multi-Output (MIMO) systems. The controller design takes into account disturbances in the plant and measurement noise from the sensors. Formally, the LQG approach addresses the problem where we consider a linear system model perturbed by disturbances $\mathbf{w}$, and measurements of the sensor corrupted by noise $\boldsymbol{\nu}$ which includes also the effects of the disturbances by measurement environment. The state-space model representation of the linear or linearized system with the addition of the disturbance effects can be mathematically expressed by

$$\begin{aligned} \dot{\mathbf{x}} &= A\mathbf{x} + B\mathbf{u} + G\mathbf{w}, \\ \mathbf{y}_v &= C\mathbf{x} + D\mathbf{u} + \boldsymbol{\nu}, \end{aligned} \tag{3.1}$$

in our problem A, B, C, and D are given by (2.10). The matrix G is the disturbance balance matrix. The disturbance and measurements noises are assumed both white noises. The principle of the LQG is combine the linear quadratic regulator (LQR) and the linear-quadratic estimator (LQE), that is, a steady-state Kalman filter. The separation principle guarantees that those can be design and computed independently [13].

3.1.1. LQR Problem

The solution for the optimal state feedback controller is obtained by solving the LQR problem. Namely the LQR optimal controller automatically ensures a stable close-loop system, and achieves guarantee levels of stability and robustness for minimal phase systems, for example, multivariable margins of phase and gain. The LQR approach gives the optimal controller gain, denoted by K, with linear control law:

$$\mathbf{u} = -K\mathbf{x}, \tag{3.2}$$

which minimizes the quadratic cost function, given by

$$J_{LQR} = \int_0^\infty \left(\mathbf{x}^T Q\mathbf{x} + \mathbf{u}^T R\mathbf{u}\right) dt, \tag{3.3}$$

where Q is positive definite, and R is semipositive definite, these are weighting or tuning matrices that define the trade-off between regulation performance and control efforts. The first term in (3.3) corresponds to the energy of the controlled output ($\mathbf{y} = \mathbf{x}$) and the second term corresponds to the energy of the control signal. The gain matrix K for the optimization problem is obtained by solving the algebraic matrix Riccati equation:

$$A^T P + PA - PBR^{-1}B^T P + Q = 0. \tag{3.4}$$

The optimal control gain is then obtained by

$$K = R^{-1}B^T P. \tag{3.5}$$

The close-loop dynamics model is obtained by substituting (3.5) into (3.1), and taking $\mathbf{w} = \mathbf{v} = 0$, as follows

$$\dot{\mathbf{x}} = (A - BK)\mathbf{x}, \tag{3.6}$$

which corresponds to an asymptotically stable system.

In order to adopt the LQR formulation the whole state $\mathbf{x}$ of the process has to be measurable. In this case it is necessary to estimate the absent states, so the estimated states are denoted by $\hat{\mathbf{x}}$. Notice that the output matrix in our case is $C = I_{6\times 6}$, it means that the whole state is measurable. Physically, the angular rates are obtained from the gyros and the attitude/orientation from the solar sensor. Nevertheless, because of the presence of noise, an estimation is advice in order to produce better and reliable information about the real states. The estimation is performed by employing the steady-state Kalman filter.

3.1.2. Kalman Filter Design

The Kalman filter is used to obtain the estimated state $\hat{\mathbf{x}}$. The filter equation in view of the attitude model is given by

$$\dot{\hat{\mathbf{x}}} = A\hat{\mathbf{x}} + B\mathbf{u} + L(\mathbf{y}_v - C\hat{\mathbf{x}}), \tag{3.7}$$

where L is the Kalman filter gain. The optimal gain L minimizes the covariance of the error E between real $\mathbf{x}$ and estimated $\hat{\mathbf{x}}$ states, by defining the state estimation error as $\mathbf{e} := \hat{\mathbf{x}} - \mathbf{x}$, the cost function is given by [13]

$$J_{LQE} = \lim_{t \to \infty} E\left\{ \mathbf{e} \cdot \mathbf{e}^T \right\}. \tag{3.8}$$

We assume that the disturbances affecting the process $\mathbf{w}$ and $\mathbf{v}$ are zero-mean Gaussian white-noise process with covariances $Q_e = E(\mathbf{w}\mathbf{w}^T)$ and $R_e = E(\mathbf{v}\mathbf{v}^T)$, respectively. The process and measurement noises are uncorrelated from each other. The gain L is obtained solving the algebraic matrix Riccati equation:

$$A^T S + SA - SCR_e^{-1}C^T S + Q_e = 0. \tag{3.9}$$

The optimal estimator gain is then obtained by

$$L = R_e^{-1}C^T S, \tag{3.10}$$

and the error dynamics is given by

$$\dot{\mathbf{e}} = (A - LC)\mathbf{e}, \tag{3.11}$$

where $A - LC$ is asymptotically stable. From (3.6) and (3.11) the open-loop transfer function for the LQG controller is found as follows:

$$K_{lqg}(s)G(s) = K(sI - A + BK + LC)^{-1}L\Phi(s), \tag{3.12}$$

where $G(s) = \Phi(s) = C(sI - A)^{-1}B$ is the transfer function of the attitude model, in this case G_s is a matrix of transfer functions.

3.2. Pulse-Width Pulse-Frequency Modulator

The control signals from the LQG controller are of continuous type. However, pulse thruster devices can provide only on-off signals generating nonlinear control action. Nonetheless, those can be used in a quasilinear mode by modulating the width of the activate reaction pulse proportionally to the level of the torque command input. This is known as pulse-width modulation (PW). In the pulse-width pulse-frequency (PWPF) modulation the distance between the pulses is also modulated. Its basic structure is shown in Figure 3.

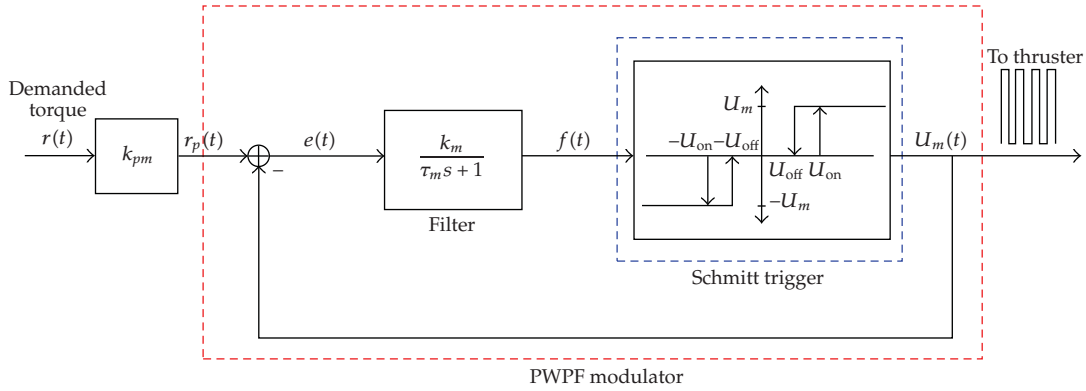


Figure 3: Pulse-width pulse-frequency (PWPF) modulator.

The modulator includes a Schmitt trigger which is a relay with dead zone and hysteresis, it includes also a first-order-filter, lag network type, and a negative feedback loop. When a positive input to the Schmitt trigger is greater than U_{on}, the trigger input is U_m. If the input falls below U_{off} the trigger output is 0. This response is also reflected for negative inputs in case of two side-thrusters or those thruster that produce negative torques (clockwise direction). The error signal $e(t)$ is the difference between the Schmitt trigger output U_{on} and the system input $r(t)$. The error is fed into the filter whose output signal $f(t)$ and it feeds the Schmitt trigger. The parameters of interest for designing the PWPF are: the filter coefficients k_m and τ_m, the Schmitt trigger parameters U_{on}, U_{off}, it defines the hysteresis as $h = U_{\text{on}} - U_{\text{off}}$, and the maximal/minimal $\pm U_m$. The PWPF modulator can incorporate an additional gain k_{pm} which will be considered separately from the control gain.

In the case of a constant input, the PWPF modulator drives the thruster valve with on-off pulse sequence having a nearly linear duty cycle with input amplitude. It is worth to note that the modulator has a behavior independent of the system in which it is used [3]. The static characteristics of the continuous time modulator for a constant input C are presented as follows:

(i) on-time

$$T_{\text{on}} = PW = -\tau_m \ln\left(1 + \frac{h}{k_m(C - U_m) - U_{\text{on}}}\right), \tag{3.13}$$

(ii) off-time

$$T_{\text{off}} = -\tau_m \ln\left(1 - \frac{h}{k_m C - (U_{\text{on}} - h)}\right), \tag{3.14}$$

(iii) modulator frequency

$$f = \frac{1}{T_{\text{on}} + T_{\text{off}}}, \tag{3.15}$$

Table 1: Recommended range for the PWPF parameters.

	Static analysis	Dynamic analysis	Recommended
k_m	$2 < k_m < 7$	N/A	$2 < k_m < 7$
τ_m	$0.1 < \tau_m < 1$	$0.1 < \tau_m < 0.5$	$0.1 < \tau_m < .5$
U_{on}	$U_{\text{on}} > 0.3$	N/A	$U_{\text{on}} > 0.3$
h	$h > 0.2U_{\text{on}}$	N/A	$h > 0.2U_{\text{on}}$
k_{pm}	N/A	$k_{pm} \geq 20$	$k_{pm} \geq 20$

(iv) duty cycle

$$DC = \left[\frac{\ln[1 + a/(1-x)]}{1 + \ln[1 + a/x]}\right]^{-1}, \tag{3.16}$$

(v) minimum pulse-width (PW)

$$\Delta = -\tau_m \ln\left[1 - \frac{h}{k_m U_m}\right], \tag{3.17}$$

where the following internal parameters are also defined: dead zone $C_d = U_{\text{on}}/k_m$, saturation level $C_s = U_m + (U_{\text{on}} - h)/k_m$, normalized hysteresis width $a = h/k_m(C_s - C_d)$, and normalized input $x = (C - C_d)/(C_s - C_d)$.

In order to determine the range of parameters for the PWPF modulator, static and dynamic analyses are carried out. The static analysis involves test input signals, for example, step, ramp, and sinusoidal signals. The dynamic analysis uses plant and controller. Afterwards the choice is based upon the number of firings and level of fuel consumption results. The number of firings gives an indication of the life-time of the thrusters. Table 1 presents the obtained results for the particular problem.

3.3. Specifications and Tuning Schemes

The specification of the requirements for the attitude control system are determined by the capabilities of the MMP satellite to attempt some desired nominal performance for the linked payload. Considering the stabilization mode the following specifications are given in terms of time and frequency domain:

(i) steady state error less than 0.5° degrees for each axis;

(ii) overshoot less than 40%;

(iii) short rise time or fast response against disturbances;

(iv) stability margins gain GM ≥ 6 db and PM ≥ 60° for each channel.

For the control design, it is necessary to check the limitations and constraints imposed by the plant. In this sense the optimality of the LQG only holds for the following assumptions: the matrix $[A \quad B]$ must be stabilizable and $[A \quad C]$ must be detectable. In the case of the attitude model, both conditions are satisfied. The next step is to design a controller which achieves the required system performance. During the stabilization mode, it is desired

attenuation of the effects of disturbances acting on the satellite and accomplishment of regulation to maintain the satellite in the required attitude. Moreover the output has to be insensitive to measurements errors. Unfortunately there is an unavoidable tradeoff between attenuation disturbances and filtering out measurement error. This tradeoff has to be kept in mind during the design of the controller. In the case of attitude model, the disturbances acting in the system belong to the spectrum at low frequencies, note that the regulation signals belongs also to spectrum at low frequencies. On the other hand, the measurement noises and unmodeled system terms are concentrated at high frequencies. In order to fulfill the specifications, tuning of LQG gains and PWPF gains have to be careful performed. The nature of the tuning is an iterative process which turns out less arduous with the use of a computational tool, in this work the Matlab package is used. In the following, the obtained weights for LQG controller and PWPF modulator are presented.

3.3.1. LQR Tuning

The first choice for the tuning matrices Q and R is taken from the Bryson's rule, selecting Q and R diagonal matrices with the form

$$\begin{aligned} Q_{ii} &= \frac{1}{\text{maximum acceptable value of } x_i^2} \quad i \in \{1,2,\ldots,n\}, \\ R_{ii} &= \frac{1}{\text{maximum acceptable value of } u_j^2} \quad j \in \{1,2,\ldots,m\}, \end{aligned} \tag{3.18}$$

where x_i and u_j are the states input signals boundaries, respectively. The rule is used to keep the states and inputs below some boundaries. It is advised to avoid large control signals which from the engineering point of view are unacceptable. On the other hand, the controller has to fulfill all the system specifications and the LQR formulation does not directly allow one to achieve standard control system specifications. Nevertheless those can be achieved by iteration over the values of the weights of Q and R in the cost function till it arrives at satisfactory controller. For the proposed reaction attitude control system the boundaries for the states are kept $\pm 5^\circ$ in attitude (ϕ, θ, ψ), and ± 1 degree per second for the rates. The boundary for the input signals are 1 Newton meter. The result weighting matrices for the controller which achieved satisfactory controller are

$$Q = Q_{ii}, \qquad R = 1 \times 10^{-1} \cdot R_{ii}. \tag{3.19}$$

The control tuning matrices R and Q were obtained through iterative process following expectable requirements, for example, allowed (non-saturation) control effort and reasonable stabilization time.

3.3.2. Filter Tuning

The tuning weight matrices R_e and Q_e for the Kalman filter are obtained considering R_e large compared to Q_e. It corresponds to weighting the measurements less than the dynamics model. This also leads to a reduction of the poles values for $A - LC$. The relative magnitude

Table 2: PWPF parameters used to compose the ACS.

k_m	τ_m	U_{on}	h	K_p
1	0.1	0.45	0.3	20

Table 3: Simulation parameters.

Parameters	Values
Principal momentum of inertia (without payload) (kgm^2)	$J_x = 305.89126$ $J_y = 314.06488$ $J_z = 167.33919$
Torque arm (m)	$l = 1.0$
Mean orbital motion (rad/s)	$\omega_0 = 0.001$
Mass (kg)	578.05239
Orbit altitude (km)	750
Maximum force (N)	5
Eccentricity	$\cong 0$
Initial attitude (degrees) slew maneuver	$(\phi,\theta,\psi)=(10,10,10)$
Initial Angular Rate (degrees/s)	$\boldsymbol{\omega}_{ib}^{b} = [1,1,1]^T$

of R_e and Q_e is determined iteratively till achieves satisfactory gain L in terms of filtering and smoothing of the measurement vector signal $\mathbf{y}_v$. The matrices values are given by

$$\begin{aligned} Q_e &= \operatorname{diag}(0,0,0,q_e,q_e,q_e), \\ R_e &= \operatorname{diag}(r_e,r_e,r_e,v_e,v_e,v_e), \end{aligned} \tag{3.20}$$

where $q_e = 5 \times 10^{-3}$, $r_e = 1 \times 10^{-1}$, and $v_e = 1 \times 10^{-2}$. Note that the precision for the rate measurements is bigger than for the attitude measurements, and the tuning values for the dynamic noise in the attitude are selected as zeros.

3.3.3. Selected PWPF Parameters

In order to compose the entire reaction thruster attitude control system and to achieve the desire performance the parameters for the PWPF are selected from the optimal range. Table 2 presents those PWPF parameters.

Next section presents the performance of the reaction thruster attitude controller during the stabilization mode. Filtering noise, rejection of impulse disturbances, and regulation performance are analyzed.

4. Numerical Simulation and Results

The reaction thruster attitude control is tested through numerical simulations. The tuning matrices schemes presented in Section 3 are used to obtain the controller and observer gains. They are able to attempt pointing requirements (<0.5°) and reasonable margins of stability for the attitude control system during the stabilization mode. Table 3 includes the values of

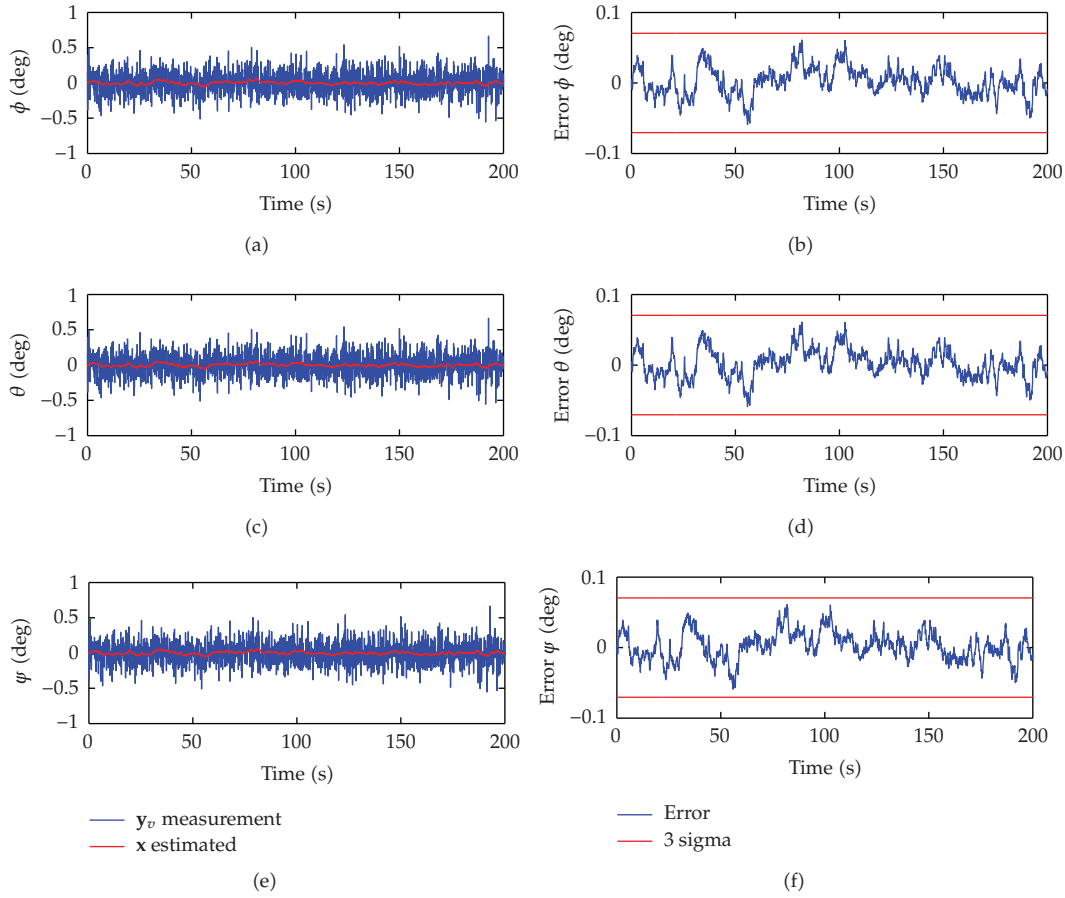


Figure 4: Filtering of the measurement data.

principal momentum of inertia without payload [1]. Although several simulations over a wide range of initial condition for attitude were performed, just one case is shown.

4.1. Noise Filtering

In order to filter the noise in the measurement $\mathbf{y}_v$ the steady-state Kalman filter is applied. The estimated and measurement attitude is shown in Figure 4, on the left side. The errors and the respective 3 sigma boundary results are shown on the right side of Figure 4. The steady-state Kalman achieves good estimation of the real attitude with a standard deviation $\sigma \cong 0.027$. The estimation results present smoother profile compare to the measurements which is favorable wish for the control system.

4.2. Short Slew Maneuver during Stabilization Mode

Although a set of different initial conditions are simulated and analyzed, we present only the case when the satellite has a displacement of 10 degrees for each axis in attitude with respect to LHLV orbital frame. To regulate or stabilize the satellite a short maneuver is needed. The

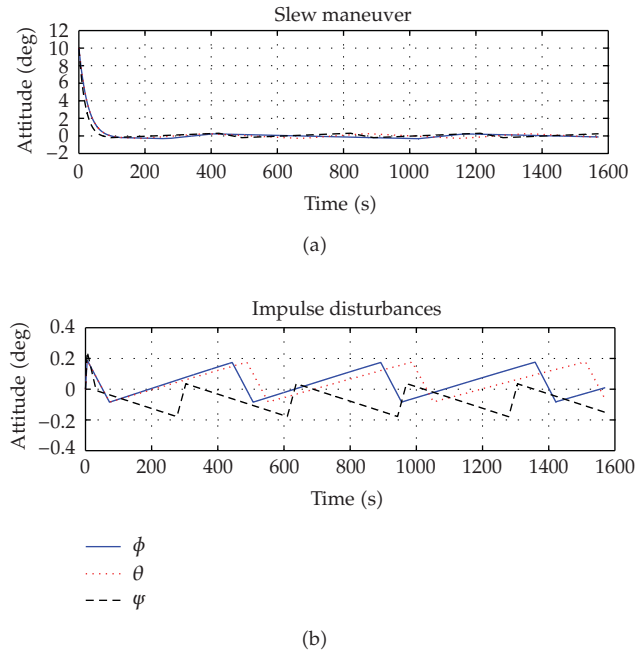


Figure 5: Attitude profile during slew maneuver and disturbance effect.

attitude profile is shown in Figure 5. The simulation time corresponds to quarter of the orbital period ($\cong$ 104) minutes. The satellite executes the maneuver in approximately 100 seconds.

The duty cycle generated by the PWPF modulation is shown in Figure 6. The duty cycle for row, pitch, and yaw angles are the same order of magnitude. The maximal spent time to complete a close path is quasi 800 seconds and it occurs in row direction. The specification of pointing accuracy is achieved, less than 0.5 degrees. In fact reading out Figure 6 the maximal errors in row, pitch, and yaw are ±0.3, ±0.3, and ±0.25 degrees, respectively. It shows a high accurate performance of the reaction thruster which is possible by modulating the control signal using the PWPF modulator.

Figure 7 shows the control command, executed by LQG controller, and the modulation during the slew maneuver. The thrusters' profile present small pulse-width modulation (t_{on}) which leads small impulses and hence less fuel consumption. Positive torques are executed by 3 of thrusters and negative ones by another 3 thrusters.

4.3. Disturbance Rejection During Stabilization Mode

In order to test the action of the controller against disturbance effects (e.g., atmospheric drag) or uncertainties in the system (e.g., sloshing), simulations considering impulse disturbance signals acting in the system are carried out. Figure 5 presents the results obtained for the attitude. It shows a maximal error in attitude of ±0.2 degrees which fulfill the desired precision specification. The attitude control system is capable of respond fast to the disturbance effects, less than 60 seconds for an exogenous pulse of 1 Newton. The control signal command and the PWPF modulation results are presented in Figure 8. The results are satisfactory in terms of accuracy and fuel since the modulation of the pulse-width is kept small.

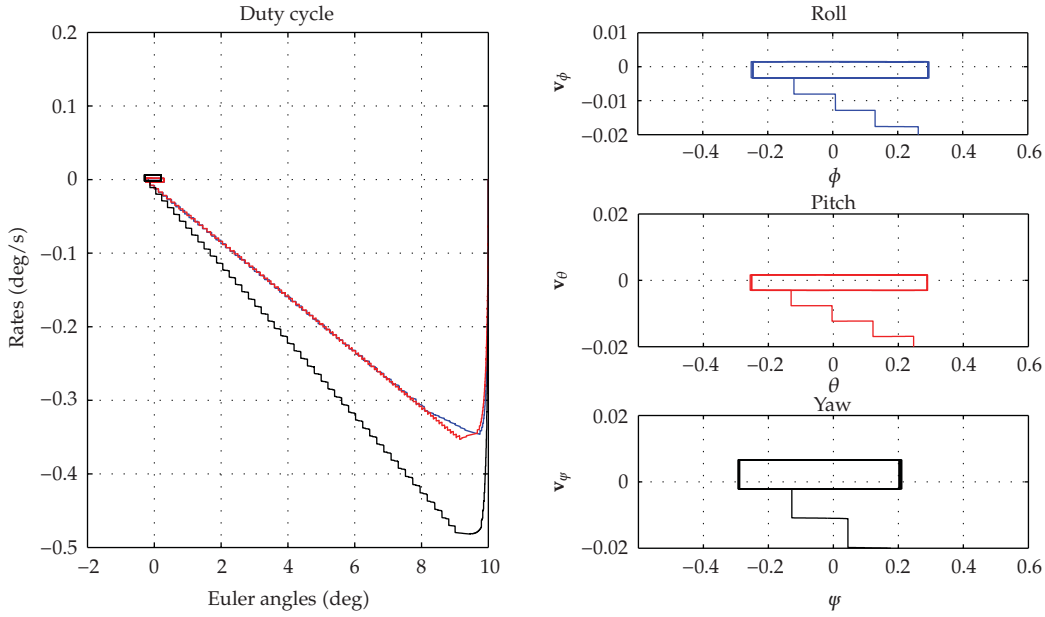


Figure 6: Duty cycle response for stabilization during a short slew maneuver.

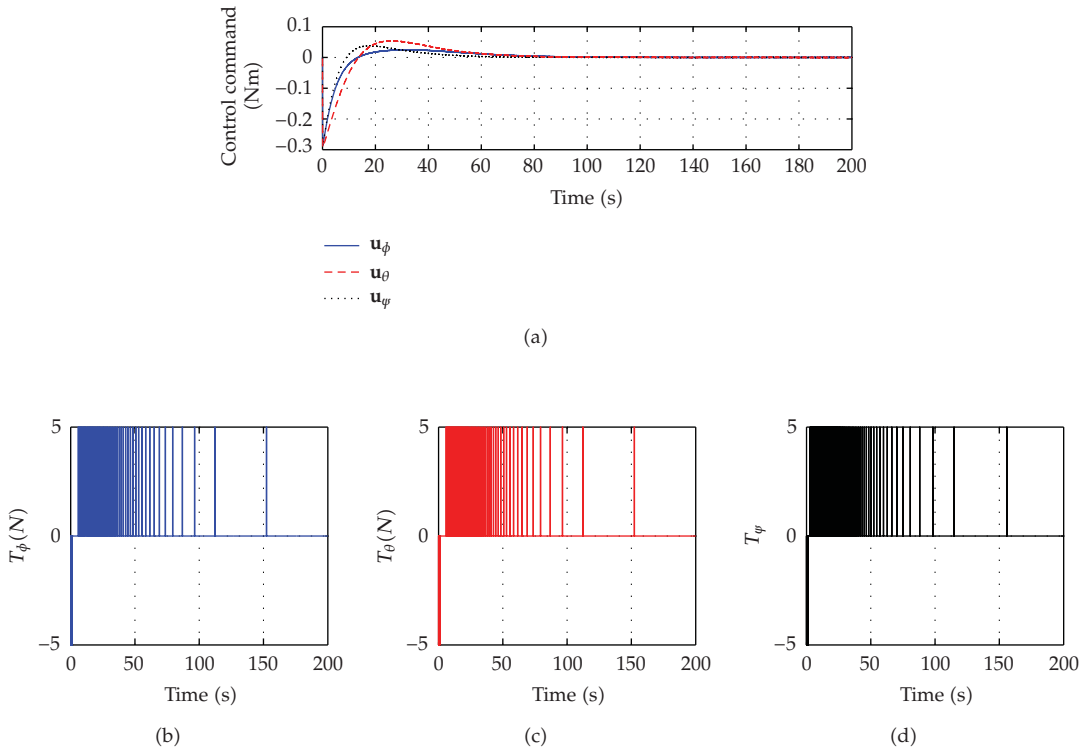


Figure 7: Control command and PWPF modulation during the slew maneuver.

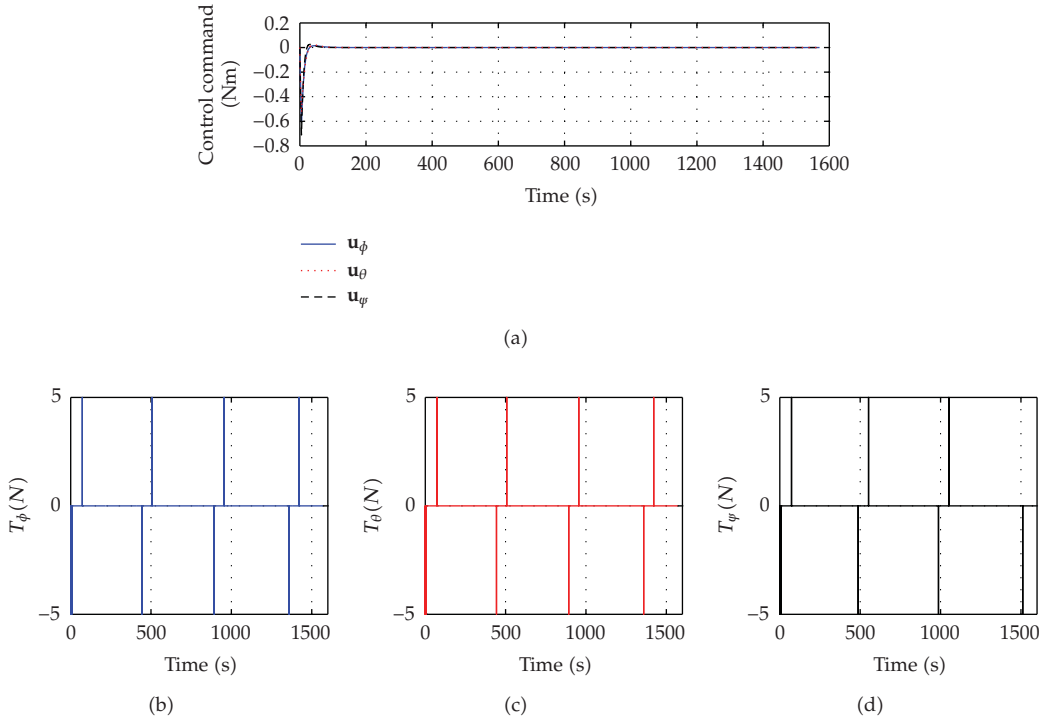


Figure 8: Control command and PWPF modulation for disturbance rejection.

It is worth to note that the use of magnetic coils actuators or reaction wheel devices can zero the error residue, for example, duty cycle, by a damped actuation. This actuation will be very small because of the level of accuracy in attitude achieved by the thruster actuation.

5. Conclusions

The obtained on-off thruster reaction attitude control system based on the LQG/PWPF modulation is optimal with respect to regulation, (i.e., minimizing the quadratic cost function of states and control signals), and propellant consumption. The optimality for fuel is obtained through off-line simulations varying the parameters of the PWPF modulator till less fuel consumption is achieved. This work presents the set of optimal parameters for the PWPF modulator by considering static and dynamic analysis.

The LQG design is an efficient way to achieve exponentially stability, moreover it allows to weight the magnitude of input signal $\mathbf{u}$, restricting the torque commands till acceptable performance is achieved. The weighting matrices for tuning the optimal LQG controller are presented and discussed in this work. The previous work, see [6], using the LQG design, demonstrated successfully, the applicability and suitability of the controller for the stabilization mode. However, in the foregoing work the required on-off modulation was not taken into consideration. It is worth to note that the LQG controller is able to stabilize the system even for large initial attitude displacements within nonlinear dynamics. It shows how resistent the controller is in face of internal changes in behavior.

The obtained results demonstrate the feasibility of combining LQG/PWPF modulator in an unique controller for on-off thruster reaction attitude control system. Stability remains

by adding the PWPF modulator and reasonable accuracy in attitude is achieved, that is, magnitude of the duty cycle. Practical aspects are included in this study as filtering and presence of external impulsive perturbations. The advantages of less spent propellant shall contribute to the MMP project, specially, a satellite conceived to be used on a large number and different types of missions, in the context of an ever-advancing Brazilian space program.

Acknowledgments

The authors acknowledge the support of *Conselho Nacional de Desenvolvimento Cientifico e Tecnológico*-CNPq, *Coordenação de Aperfeiçoamento* de Pessoal *de Nivel Superior* CAPES (Brazil), and *Fundação de Amparo à Pesquisa do* Estado de S. *Paulo*-FAPESP (São Paulo, Brazil).

References

[1] AEB—Agência Espacial Brasileira, "National space activities program PNAE 1998–2007," Report Brazilian Space Agency, AEB, Brasilia, Brazil, 1998.
[2] Q. Hu and Y. Liu, "A hybrid scheme of feed-forward/feedback control for vibration suppression of flexible spacecraft with on-off actuators during attitude maneuver," *International Journal of Information Technology*, vol. 11, no. 12, pp. 95–107, 2005.
[3] T. D. Krovel, *Optimal tunning of PWPF modulator for attitude control*, M.S. thesis, Department of Engineering Cybernetics, Norwegian University of Science and Technology, 2005.
[4] T. D. Krovel, "PWPF modulation of thrusters on the micro-satellite SSETI/ESMO," Tech. Rep., Department of Engineering Cybernetics, Norwegian University of Science and Technology, 2005.
[5] R. S. McClelland, *Spacecraft attitude control system performance using pulse-width pulse-frequency modulated thrusters*, M.S. thesis, Naval Postgraduate School, Monterey, Calif, USA, 1994.
[6] A. C. Santana, L. S. Martins-Filho, and G. Arantes Jr., "Attitude stabilization of the PMM satellite using a LQG-based control strategy," *Tendências em Matemática Aplicada e Computacional*, vol. 9, no. 2, pp. 321–330, 2008.
[7] J. R. Wertz, Ed., *Spacecraft Attitude Determination and Control*, Reidel, Dordrecht, The Netherlands, 1978.
[8] M. H. Kaplan, *Modern Spacecraft Dynamics and Control*, John Wiley & Sons, New York, NY, USA, 1976.
[9] B. Wie, H. Weiss, and A. Arapostathis, "Quaternion feedback regulator for spacecraft eigenaxis rotations," *Journal of Guidance, Control, and Dynamics*, vol. 12, no. 3, pp. 375–380, 1989.
[10] G. Arantes Jr., *Estudo comparativo de técnicas de controle de atitude em três eixos para satelites artificias*, M.S. thesis, INPE, S. José dos Campos, Brazil, 2005.
[11] LIGOS, "V-Relam Builder: User's Guide and Reference," Ligos Corporation, 1997.
[12] J. B. Burl, *Linear Optimal Control*, Prentice Hall, Upper Saddle River, NJ, USA, 1998.
[13] B. D. O. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*, Dover, Mineola, Tex, USA, 2007.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 507370, 23 pages
doi:10.1155/2009/507370

Research Article

Highly Efficient Sigma Point Filter for Spacecraft Attitude and Rate Estimation

Chunshi Fan[1] and Zheng You[2]

[1] *Department of Precision Instruments and Mechanology, Tsinghua University, Beijing 100084, China*
[2] *State Key Laboratory of Precision Measurement Technology and Instruments, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Chunshi Fan, fancspim@gmail.com

Received 9 July 2009; Accepted 30 September 2009

Recommended by Tadashi Yokoyama

Nonlinearities in spacecraft attitude determination problem has been studied intensively during the past decades. Traditionally, multiplicative extended Kalman filter_MEKF_algorithm has been a good solution for most nominal space missions. But in recent years, advances in space missions deserve a revisit of the issue. Though there exist a variety of advanced nonlinear filtering algorithms, most of them are prohibited for actual onboard implementation because of their overload computational complexity. In this paper, we address this difficulty by developing a new algorithm framework based on the marginal filtering principle, which requires only 4 sigma points to give a complete 6-state attitude and angular rate estimation. Moreover, a new strategy for sigma point construction is also developed to further increase the efficiency and numerical accuracy. Incorporating the presented framework and novel sigma points, we proposed a new, nonlinear attitude and rate estimator, namely, the Marginal Geometric Sigma Point Filter. The new algorithm is of the same precision as traditional unscented Kalman filters, while keeping a significantly lower computational complexity, even when compared to the reduced sigma point algorithms. In fact, it has truly rivaled the efficiency of MEKF, even when simple closed-form solutions are involved in the latter.

1. Introduction

Nonlinearities in spacecraft attitude determination problem have been studied intensively during the past decades. Since the early 1980's, multiplicative extended Kalman filtering (MEKF) algorithm [1] has proven to be a successful solution for engineering application. The MEKF algorithm has a very low computing cost and performs quite well in most nominal space missions where the spacecraft's angular rate is slow and the nonlinearities are not so impactive.

In recent years, advances in space missions, such as the greater agility demand and the application of lower cost sensors, deserve a revisit of the nonlinearity issue. Although a variety of advanced nonlinear filtering algorithms exist, only few of them are close to the restrict numerical expense requirements of actual onboard implementations. In the existing methods, the well-known sigma point Kalman filters (SPKFs) [2, 3] have approven to be among the most efficient ones. SPKF bases on a Gaussian distribution approximation technique, namely, the unscented transformation (UT), where a deterministic set of weighted points (known as the sigma points) are used to make probabilistic inference [4]. Eliminating the complex Jacobian matrix derivations, SPKF algorithms are much easier to implement and have better performance than traditional widely used EKF algorithms; so they have found widespread application in a variety of fields. In recent years, spacecraft attitude estimation problems have also been addressed by SPKF approaches in literature and engineering practice [5–7].

In spite of being efficient among nonlinear filters, baseline SPKF still seems computational costly for engineering implementation. If we denote m as the number of sigma points required, then for an n-dimension random state vector $\mathbf{x}$, standard unscented transformation needs a set of $m = 2n + 1$ points to capture the state's statistical distribution properties. More seriously, when we develop a complete SPKF estimator for the n-dimensional state model, the actual m needed is no longer $2n + 1$—it easily becomes $4n$ or even larger, because standard SPKF algorithm requires a state augmentation to include all the propagation and measurement noise terms, hence leading to an unacceptable increase in computational burden. For avoidance of state augment, iterated and trapezoidal approximation approaches [5, 8] have been developed. Both approaches are suitable for additive noise case only, and they are able to reduce m back to $2n + 1$. To further reduce the complexity, strategies for introducing fewer sigma points are exploited, known as the reduced sigma point filters. Several simplex points selection strategy have been developed, including the $n + 2$ point minimal-skew simplex points [9], the spherical simplex points [10, 11], and some enhanced higher-order extensions [12]. Each of the above sigma point sets involves a zero-valued "central point," which is usually endowed with a negative weight so as to minimize higher order effects, known as the scaled UT technique [13]. In recent years, new sigma point selection strategy involving no central-point is introduced, such as the Schmidt orthogonalization-based simplex set [14], which includes $n + 1$ equally weighted points. It is also proved that [15] such negative weight-free, equally weighted sigma point sets are numerically more stable and accurate as well as having a better efficiency. In fact, both central-point elimination and equal weight assignment improve the symmetry property of the sigma point set, and a better symmetry property provides a better numerical behavior, as it has a better capability to suppress the impact of the round off errors. In this article, we address the construction strategies to make a best symmetric structure in simplex sigma point set.

Anyhow, applications of simplex sets have reduced the required sigma points to 50% of the traditional nonaugmented algorithms and have made a significant improvement in numerical efficiency. In fact, the numerical efficiency of simplex SPKF algorithm is able to rival or even exceed its EKF counterpart if general formed Riccati equations and numerical integration process are involved in the latter. However, for typical quaternion-based spacecraft attitude problems, since simple analytical solutions and sparse matrices exist for MEKF′ covariance propagation and measurement updates, general simplex SPKF algorithm still compares unfavorably with MEKF in efficiency.

Clearly, to develop a competitive algorithm alternative to the MEKF for practical applications, we need a still further reduction of m. Recently, a new sigma point selection

strategy for a class of "partially linear" transformation problems has been proposed, namely, the marginal unscented transformation [16]. By exploiting the linear substructures in system model and margining out the corresponding variables, the marginal UT algorithms only needs a set of sigma points that adequately describes the statistics property of the nonlinear part of the states. It provides a possible approach to further shrink the size of the sigma point set. As attitude dynamics also has linear substructures in gyro bias drift model and the observation equations, it is imaginable that we can also address the attitude and rate estimator design with similar ideas.

The main contributions of this article include two parts. First, we have derived in detail a marginal version of SPKF algorithm for typical 6-state attitude determination system. The new algorithm uses merely 4 sigma points to give a complete attitude and angular rate estimation; hence it is able to achieve a high numerical efficiency that truly rivals the MEKF. Second, we have proposed a new set of simplex sigma points for Euclidean Geometric space, named the Geometric Simplex sigma point set. The new set has a symmetric structure, a lower computational expense and is numerically more accurate. It would be of use in a variety of 3-dimensional modeled dynamic problems.

The organization of this paper proceeded as follows. First, we established a general 6-state stellar-inertial spacecraft attitude kinemics and measurement model and analyzed the partially linear structure in the system. Then a marginal SPKF estimator is derived in detail. Next, we looked into the asymmetrical properties of existing sigma point construction algorithms and proposed the Geometric simplex set. Finally, we incorporated the proposed sigma point set into the marginal filtering framework to configure a complete attitude estimator, named the Marginal Geometric Sigma Point Filter, and inspected its performance in simulation with comparisons to the traditional MEKF and Spherical Simplex SPKF.

2. Models and Traditional Filtering Frameworks

2.1. Attitude Dynamics and the MEKF Framework

For spacecraft attitude estimation, quaternion has been the most widely used attitude parameterization. The quaternion is given by a 4-dimension vector defined as $\overline{\mathbf{q}} = [\mathbf{q}^T, q_4]^T$, with $\mathbf{q} \equiv [q_1, q_2, q_3]^T = \hat{\mathbf{n}} \sin(\phi/2)$ and $q_4 = \cos(\phi/2)$, where $\hat{\mathbf{n}}$ is Euler axis and ϕ is the rotation angle. Quaternion parameter satisfies a single constraint given by $\overline{\mathbf{q}}^T\overline{\mathbf{q}} = 1$. The kinematics equation is given by

$$\dot{\overline{\mathbf{q}}} = \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega})\mathbf{q}, \tag{2.1}$$

where $\boldsymbol{\omega}$ is the angular rate vector given from the gyro's measurement $\boldsymbol{\omega}^{\text{meas}}$ by compensating the gyro bias $\mathbf{b}$:

$$\boldsymbol{\omega} = \boldsymbol{\omega}^{\text{meas}} - \mathbf{b} - \boldsymbol{\eta}_{\text{ARW}}, \tag{2.2}$$

where $\boldsymbol{\eta}_{\text{ARW}}$ is a zero-mean Gaussian angular random walk noise with a covariance of $\sigma^2_{\text{ARW}}\mathbf{I}_3$. $\mathbf{b}$ is often modeled as a rate random walk process with white noise $\boldsymbol{\eta}_{\text{RRW}}$ and a covariance of $\sigma^2_{\text{RRW}}\mathbf{I}_3$:

$$\dot{\mathbf{b}} = \boldsymbol{\eta}_{\text{RRW}}. \tag{2.3}$$

From (2.1)~(2.3), we can derive the discrete-time version of the above models with numerical integration. In fact, numerically simpler closed-form solution exists for (2.1) if we approximately consider the direction of $\boldsymbol{\omega}$ as a constant vector during the propagation interval T for each filtering circle from t_{k-1} to t_k:

$$\widehat{\overline{\mathbf{q}}}_{k/k-1} = \widehat{\overline{\mathbf{q}}}_{k-1} \otimes \widehat{\overline{\mathbf{q}}}^{\boldsymbol{\omega}}_{k-1}, \tag{2.4}$$

where

$$\widehat{\overline{\mathbf{q}}}^{\boldsymbol{\omega}}_{k-1} = \left[\frac{\widehat{\boldsymbol{\phi}}^T_{k-1}}{2} \frac{\sin\left(\widehat{\phi}_{k-1}/2\right)}{\cos\left(\widehat{\phi}_{k-1}/2\right)}, \cos\frac{\widehat{\phi}_{k-1}}{2} \right]^T \tag{2.5}$$

with

$$\begin{aligned} \widehat{\boldsymbol{\omega}}_{k-1} &= \boldsymbol{\omega}^{\text{meas}}_{k-1} - \widehat{\mathbf{b}}_{k-1}, \\ \widehat{\boldsymbol{\phi}}_{k-1} &= \widehat{\boldsymbol{\omega}}_{k-1} T, \qquad \widehat{\phi}_{k-1} = \left| \widehat{\boldsymbol{\phi}}_{k-1} \right|. \end{aligned} \tag{2.6}$$

The approximation is tenable as far as the period T is small enough, which is usually well satisfied in practice, and a second-order accuracy is guaranteed.

In actual calculation process, (2.5) seems too complex, and a 2nd-order approximation is enough. To associate the quaternion $\overline{\mathbf{q}}(\boldsymbol{\phi}) = [\hat{\mathbf{n}}^T \sin(\phi/2), \cos(\phi/2)]^T$ to its pertinent rotation vector $\boldsymbol{\phi} = \hat{\mathbf{n}}\phi$ in a straightforward way, we can choose an arbitrarily 3-dimensional attitude parameter as the media. In this paper, we choose the Modified Rodrigues Parameters (MRPs) recommended in [17] to give a 2nd-order, trigonometric function, and square-root function-free approximation of (2.5):

$$\begin{aligned} \boldsymbol{\phi}(\delta\mathbf{q}) &= \frac{4\delta\mathbf{q}}{(1 + \delta q_4)}, \\ \delta\overline{\mathbf{q}}(\boldsymbol{\phi}) &= \frac{\left[8\boldsymbol{\phi}^T, (16 - \phi^2)\right]^T}{(16 + \phi^2)} \end{aligned} \tag{2.7}$$

with $\phi^2 = \boldsymbol{\phi}^T\boldsymbol{\phi}$. Clearly, such MRP-based expression is quite economic for computation. Then the discrete-time propagation equations can be written as follows:

$$\begin{aligned} \widehat{\overline{\mathbf{q}}}^{\boldsymbol{\omega}}_{k-1} &= \frac{\left[8\widehat{\boldsymbol{\phi}}^T_{k-1}, \left(16 - \widehat{\phi}^2_{k-1}\right)\right]^T}{\left(16 + \widehat{\phi}^2_{k-1}\right)}, \\ \widehat{\mathbf{b}}_{k/k-1} &= \widehat{\mathbf{b}}_{k-1}. \end{aligned} \tag{2.8}$$

To cope with the unit-length constraint of quaternion, local error states (also known as the local disturbance states) are introduced into filter design. We describe the local

attitude error as a 3-dimensional rotation vector $\mathbf{a}$ and the local gyro bias error as another 3-dimensional vector $\Delta\mathbf{b}$. Then the actual state vector processed by the filter is the 6-dimension disturbance state:

$$\mathbf{x} = \left[\mathbf{a}^T, \Delta\mathbf{b}^T\right]^T. \tag{2.9}$$

For clarity, we hence address the original states $\widehat{\overline{\mathbf{q}}}$ and $\mathbf{b}$ as the "global states," which would mainly serve as singular-point-free reference and storage for the filter. The global states and the local error state are affiliated as

$$\begin{aligned} \overline{\mathbf{q}} &= \widehat{\overline{\mathbf{q}}} \otimes \delta\overline{\mathbf{q}}(\hat{\mathbf{a}}), \\ \mathbf{b} &= \widehat{\mathbf{b}} + \Delta\mathbf{b}, \end{aligned} \tag{2.10}$$

where $\delta\overline{\mathbf{q}}$ is the local error quaternion corresponding to $\hat{\mathbf{a}}$. Again we choose to use the MRP approximation, as

$$\begin{aligned} \mathbf{a}(\delta\mathbf{q}) &= \frac{4\delta\mathbf{q}}{(1+\delta q_4)}, \\ \delta\overline{\mathbf{q}}(\mathbf{a}) &= \frac{\left[8\mathbf{a}^T, (16-a^2)\right]^T}{(16+a^2)} \end{aligned} \tag{2.11}$$

with $a^2 = \mathbf{a}^T\mathbf{a}$. Clearly, such an MRPs-based expression is free from any square-root or trigonometric functions, economic in computation.

After all generality, the observation model in this article is established as an automatic star sensor with quaternion measurements $\overline{\mathbf{q}}_k^{\text{meas}}$. But in actual practice this information is presented to the Kalman filter in a more convenient way as in terms of a 3-dimensional parameter. Again we choose to use the MRP parameter, and then the star sensor's observation model is simply defined as the local error between the predicted and observed attitudes:

$$\mathbf{z}_k^{\text{meas}} = \mathbf{h}(\hat{\mathbf{x}}_k) = \mathbf{a}\left(\delta\overline{\mathbf{q}}_k^{\text{meas}}\right) + \mathbf{v}_k, \tag{2.12}$$

where

$$\delta\overline{\mathbf{q}}_k^{\text{meas}} = \widehat{\overline{\mathbf{q}}}_{k/k-1}^{-1} \otimes \overline{\mathbf{q}}_k^{\text{meas}}, \tag{2.13}$$

and $\mathbf{v}_k$ is the measurement noise covariance modeled as

$$\mathbf{R}_k = \sigma_r^2\mathbf{I}_3. \tag{2.14}$$

With the above models, we can derive an MEKF estimator as in [1] and briefly reviewed in blocked matrix form in Table 4. Note that the most computational cost parts of the

algorithm involve the covariance propagation, measurement update, and the Kalman filter gain computing, namely,

$$\mathbf{P}_{k/k-1}^{X} = \boldsymbol{\Theta}_{k-1}\mathbf{P}_{k-1}^{X}\boldsymbol{\Theta}_{k-1}^{T} + \mathbf{Q}_{k}, \tag{2.15}$$

$$\mathbf{K}_{k} = \mathbf{P}_{k/k-1}^{X}\mathbf{H}_{k}^{T}\left(\mathbf{H}_{k}\mathbf{P}_{k/k-1}^{X}\mathbf{H}_{k}^{T} + \mathbf{R}_{k}\right)^{-1}, \tag{2.16}$$

$$\mathbf{P}_{k}^{X} = (\mathbf{I}_{6} - \mathbf{K}_{k}\mathbf{H}_{k})\mathbf{P}_{k/k-1}^{X}, \tag{2.17}$$

where $\mathbf{P}_{k}^{X}$ is the estimation of $\boldsymbol{x}$'s covariance, $\mathbf{K}_{k}$ is the Kalman filter gain, $\mathbf{H}_{k}$ is the observation matrix, and $\boldsymbol{\Theta}_{k}$ and $\mathbf{Q}_{k}$ are, respectively, the transition matrix and the propagation noise matrix, as

$$\boldsymbol{\Theta}_{k-1} = \begin{bmatrix} \boldsymbol{\Phi}_{k-1} & \boldsymbol{\Psi}_{k-1} \\ \mathbf{0}_{3\times 3} & \mathbf{I}_{3} \end{bmatrix}, \qquad \mathbf{Q}_{k} = \begin{bmatrix} \mathbf{Q}_{k}^{A} & \left(\mathbf{Q}_{k}^{BA}\right)^{T} \\ \mathbf{Q}_{k}^{BA} & \mathbf{Q}_{k}^{B} \end{bmatrix}. \tag{2.18}$$

General problem would involve numerical integration of Riccati equations to evaluate the $\boldsymbol{\Theta}_{k}$ matrix in (2.15), and complex derivations to evaluate the $\mathbf{H}_{k}$ matrix in (2.16) and (2.17). Fortunately, it is also found that simple analytical solutions exist for $\boldsymbol{\Phi}_{k-1}$, $\boldsymbol{\Psi}_{k-1}$ (see Table 4) [18], and $\mathbf{Q}_{k}$ [19] as

$$\begin{aligned} \mathbf{Q}_{k}^{A} &= T\left(\sigma_{\mathrm{ARW}}^{2} + \left(\frac{1}{3}\right)\sigma_{\mathrm{RRW}}^{2}T^{2}\right)\mathbf{I}_{3}, \\ \mathbf{Q}_{k}^{BA} &= -\left(\frac{1}{2}\right)\sigma_{\mathrm{RRW}}^{2}T^{2}\mathbf{I}_{3}, \\ \mathbf{Q}_{k}^{B} &= \sigma_{\mathrm{RRW}}^{2}T\mathbf{I}_{3}. \end{aligned} \tag{2.19}$$

Moreover, $\mathbf{H}_{k}$ involves sparse and unit matrices as

$$\mathbf{H}_{k} = [\mathbf{I}_{3}, \mathbf{0}_{3\times 3}]. \tag{2.20}$$

Hence the MEKF algorithm is numerically very efficient. The computational complex of a typical stellar-inertial MEKF attitude estimator is evaluated in Table 4.

2.2. Traditional Sigma Point Filtering Framework

We consider now the application of SPKF to the system discussed above. A set of m sigma points are chosen to approximate the statistic distribution of the 6-dimension disturbance state $\mathbf{x} = [\mathbf{a}^{T}, \Delta\mathbf{b}^{T}]^{T}$:

$$\boldsymbol{\chi}(i) = \left[\boldsymbol{\alpha}^{(i)T}, \boldsymbol{\beta}^{(i)T}\right]^{T}, \quad i = 1, 2, \ldots, m, \tag{2.21}$$

where (i) is the index, $\boldsymbol{\alpha}^{(i)}$, $\boldsymbol{\beta}^{(i)}$ represent the attitude and bias error, respectively, each point is assigned with a weight $W^{(i)}$, and all the weights satisfy the normalization constraint:

$$\sum_{i=1}^{m} W^{(i)} = 1. \tag{2.22}$$

For clarity, write the sigma points and their associated weight in matrix form, as

$$\begin{aligned} \mathcal{X} &= \left[\boldsymbol{\chi}^{(1)}, \boldsymbol{\chi}^{(2)}, \ldots, \boldsymbol{\chi}^{(m)}\right] = \begin{bmatrix} a \\ \mathcal{B} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \ldots, \boldsymbol{\alpha}^{(m)} \\ \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(m)} \end{bmatrix}, \\ \mathcal{W} &= \left[W^{(1)}, W^{(2)}, \ldots, W^{(m)}\right], \\ \boldsymbol{\Lambda}^{\mathcal{W}} &= \operatorname{diag}(\mathcal{W}). \end{aligned} \tag{2.23}$$

For unbiased distribution of $\mathbf{x} \in \mathbb{R}^n$, $\mathcal{X}$ is constructed as

$$\mathcal{X} = \mathbf{S}^X \mathcal{U}, \tag{2.24}$$

where $\mathbf{S}^X$ is an arbitrary square-root matrix of $\mathbf{P}^X$ with $\mathbf{S}^X(\mathbf{S}^X)^T = \mathbf{P}^X = \widehat{\mathbf{cov}(\mathbf{x}, \mathbf{x})}$, also denoted as $\mathbf{S}^X = \sqrt{\mathbf{P}^X}$. $\mathcal{U} = [\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}] \in \mathbb{R}^{n \times m}$ is a base set of sigma points, and it can be defined in several different rules, depending on the sigma point construction strategy we use. Anyhow, $\mathcal{U}$ has an unbiased mean and a unit covariance:

$$\begin{aligned} \mathcal{U} \cdot \mathcal{W}^T &= \mathbf{0}, \\ \mathcal{U} \cdot \boldsymbol{\Lambda}^{\mathbf{W}} \cdot \mathcal{U}^T &= \mathbf{I}_m. \end{aligned} \tag{2.25}$$

The construction strategy of $\mathcal{U}$ is also the dominating differentiation between different SPKF algorithms. With the definitions above, the set $\mathbf{X}$ is able to capture the statistics of $\mathbf{x}$'s distribution precisely up to 2nd-order.

The SPKF attitude estimator is as follows. At the beginning time of each filtering step t_{k-1}, the local error state is reset to zero:

$$\hat{\mathbf{x}}_{k-1} = \left[\hat{\mathbf{a}}_{k-1}^T, \Delta\hat{\mathbf{b}}_{k-1}^T\right]^T = [\mathbf{0}_{3\times1}, \mathbf{0}_{3\times1}]^T. \tag{2.26}$$

Then we construct the sigma points of $\mathbf{X}_{k-1}$. As in [5], we would like to use the trapezoidal approximation to avoid state augmentation; so actually $\mathbf{S}_{k-1}^X$ is computed as

$$\mathbf{S}_{k-1}^X = \sqrt{\mathbf{P}_{k-1}^X + \overline{\mathbf{Q}}_{k-1}^X}, \tag{2.27}$$

where

$$\overline{\mathbf{Q}}_{k-1}^{X} = \begin{bmatrix} \overline{\mathbf{Q}}_{k-1}^{A}, \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3}, \overline{\mathbf{Q}}_{k-1}^{B} \end{bmatrix} = \begin{bmatrix} \dfrac{T}{2}\left(\sigma_{\mathrm{ARW}}^2 - \dfrac{T^2}{6}\sigma_{\mathrm{RRW}}^2\right)\mathbf{I}_3, \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3}, \ \dfrac{T}{2}\sigma_{\mathrm{RRW}}^2\mathbf{I}_3 \end{bmatrix}, \tag{2.28}$$

and it is equivalent to have the $\mathbf{Q}_k$ in (2.18) implicitly propagated together with the sigma points [5].

Next, propagate the m sigma points as follows:

$$\begin{aligned} \boldsymbol{\phi}_{k-1}^{\beta(i)} &= \left(\widehat{\boldsymbol{\omega}}_{k-1} - \boldsymbol{\beta}_{k-1}^{(i)}\right)T, \\ \overline{\mathbf{q}}_{k-1}^{\beta(i)} &= \frac{\left[8\boldsymbol{\phi}_{k-1}^{\beta(i)T}, \left(16 - \phi_{k-1}^{\beta(i)2}\right)\right]^T}{\left(16 + \phi_{k-1}^{\beta(i)2}\right)}, \end{aligned} \tag{2.29}$$

where $\phi_{k-1}^{\beta(i)} = |\boldsymbol{\phi}_{k-1}^{\beta(i)}|$, and

$$\begin{aligned} \overline{\mathbf{q}}_{k-1}^{\boldsymbol{\alpha}(i)} &= \frac{\left[8\boldsymbol{\alpha}_{k-1}^{(i)T}, \left(16 - \alpha_{k-1}^{(i)2}\right)\right]^T}{\left(16 + \alpha_{k-1}^{(i)2}\right)}, \\ \overline{\mathbf{q}}_{k/k-1}^{\boldsymbol{\alpha}(i)} &= \left(\widehat{\overline{\mathbf{q}}}_{k-1}^{\boldsymbol{\omega}}\right)^{-1} \otimes \overline{\mathbf{q}}_{k-1}^{\boldsymbol{\alpha}(i)} \otimes \overline{\mathbf{q}}_{k-1}^{\beta(i)}. \end{aligned} \tag{2.30}$$

Thereby we have the propagated sigma points $a_{k/k-1}$ and $\mathcal{B}_{k/k-1}$, respectively, as

$$\boldsymbol{\alpha}_{k/k-1}^{(i)} = \boldsymbol{\alpha}\left(\overline{\mathbf{q}}_{k/k-1}^{\boldsymbol{\alpha}(i)}\right) = \frac{4\mathbf{q}_{k/k-1}^{\boldsymbol{\alpha}(i)}}{\left(1 + q_{4k/k-1}^{\boldsymbol{\alpha}(i)}\right)}, \tag{2.31}$$

$$\boldsymbol{\beta}_{k/k-1}^{(i)} = \boldsymbol{\beta}_{k-1}^{(i)}. \tag{2.32}$$

Note we have propagated the attitude-related sigma points $\boldsymbol{\alpha}^{(i)}$ directly without adhering it to the global state $\widehat{\overline{\mathbf{q}}}_{k-1}$ as in [5]; so a decrease of computational complexity is achievable. This approach stands as far as the time interval of T is guaranteed to be small enough in order that both $\boldsymbol{\alpha}^{(i)}$ and $\boldsymbol{\phi}^{(i)}$ can be taken as small rotation vectors, and the MRP approximation is valid. Then the predicted mean of $\mathbf{x}$ is

$$\widehat{\mathbf{x}}_{k/k-1} = \mathcal{X}_{k/k-1}\mathcal{W}^T. \tag{2.33}$$

Now we compute the predicted measurements. Noting that the sigma points have already implicitly changed their reference from $\widehat{\overline{\mathbf{q}}}_{k-1}$ to $\widehat{\overline{\mathbf{q}}}_{k/k-1}$ during propagation, so the predicted star sensor quaternion measurements are

$$\widehat{\overline{\mathbf{q}}}_{k/k-1}^{\mathcal{X}(i)} = \widehat{\overline{\mathbf{q}}}_{k/k-1} \otimes \delta\overline{\mathbf{q}}_{k/k-1}^{\boldsymbol{\alpha}(i)}. \tag{2.34}$$

Following (2.12) and (2.13), it is straight forward to derive the observation model for the SPKF as

$$\boldsymbol{\zeta}_{k/k-1}^{(i)} = \boldsymbol{\alpha}_{k/k-1}^{(i)}. \tag{2.35}$$

Denoted in matrix form as $\mathcal{Z} = [\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \ldots, \boldsymbol{\zeta}^{(m)}]$, then we can get the predicted measurement $\widehat{z}_{k/k-1}$ as

$$\widehat{z}_{k/k-1} = \boldsymbol{\mathcal{Z}}_{k/k-1} \cdot \boldsymbol{\mathcal{W}}^T. \tag{2.36}$$

Next we compute the covariance predictions. If we take $\widehat{\mathbf{x}}_{k/k-1}$ as a bias of the sigma points, then to appropriately compute the covariance, we must remove this bias at first. Hence the covariance prediction process would involve two steps:

$$\mathcal{X}_{k/k-1} = \mathcal{X}_{k/k-1} - \mathbf{1}_{1\times m} \circ \widehat{\mathbf{x}}_{k/k-1}, \tag{2.37}$$

$$\mathbf{P}_{k/k-1}^{X} = \boldsymbol{\mathcal{X}}_{k/k-1} \cdot \boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}} \cdot \boldsymbol{\mathcal{X}}_{k/k-1}{}^T, \tag{2.38}$$

where "$\circ$" denotes the Kronecker product, and $\mathbf{1}_{1\times m} = [1, 1, \ldots, 1]_{1\times m}$. Similarly, we give the unbiased measurement points set as

$$\boldsymbol{\mathcal{Z}}_{k/k-1} = \boldsymbol{\mathcal{Z}}_{k/k-1} - \mathbf{1}_{1\times m} \circ \widehat{z}_{k/k-1} \tag{2.39}$$

the innovation covariance and cross covariance matrix are then, respectively,

$$\begin{aligned} \mathbf{P}_{k/k-1}^{Z} &= \boldsymbol{\mathcal{Z}}_{k/k-1} \cdot \boldsymbol{\Lambda}^{\mathcal{W}} \cdot \boldsymbol{\mathcal{Z}}_{k/k-1}{}^T + \mathbf{R}_k, \\ \mathbf{P}_{k/k-1}^{XZ} &= \boldsymbol{\mathcal{X}}_{k/k-1} \cdot \boldsymbol{\Lambda}^{\mathcal{W}} \cdot \boldsymbol{\mathcal{Z}}_{k/k-1}{}^T \end{aligned} \tag{2.40}$$

and the measurement update procedure is

$$\mathbf{K}_k = \mathbf{P}_{k/k-1}^{XZ}\left(\mathbf{P}_{k/k-1}^{Z}\right)^{-1}, \tag{2.41}$$

$$\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_{k/k-1} + \mathbf{K}_k\left(\widehat{z}_k^{\text{meas}} - \widehat{z}_{k/k-1}\right), \tag{2.42}$$

$$\mathbf{P}_k^{X} = \mathbf{P}_{k/k-1}^{X} - \mathbf{K}_k\mathbf{P}_{k/k-1}^{Z}(\mathbf{K}_k)^T. \tag{2.43}$$

Finally, update the global states $\widehat{\overline{\mathbf{q}}}$ and $\widehat{\mathbf{b}}$ as

$$\begin{aligned}\widehat{\overline{\mathbf{q}}}_k &= \widehat{\overline{\mathbf{q}}}_{k/k-1} \otimes \delta\overline{\mathbf{q}}(\widehat{\mathbf{a}}_k),\\ \widehat{\mathbf{b}}_k &= \widehat{\mathbf{b}}_{k/k-1} + \Delta\widehat{\mathbf{b}}_k.\end{aligned} \tag{2.44}$$

Above is the framework of an SPKF version attitude estimator. A blocked-form procedure summary is listed in Table 5. It is not difficult to evaluate the computational complexity of the SPKF estimator. Even when reduced-point algorithms, such as the spherical simplex unscented Kalman filter, are adopted, the computing effort is still a double of the MEKFs, as listed in Table 3.

3. Marginal Geometric Sigma Point Filters

3.1. Marginal Sigma Point Filtering Framework for Spacecraft Attitude and Rate Estimations

Now we look into some special structures of the above estimator. First, noting (2.32), we find that the bias-related sigma points $\mathcal{B} = [\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(m)}]$ remain unchanged during the whole process of the propagation, indicating that their mean would also remain unchanged as $\Delta\widehat{\mathbf{b}}_{k/k-1} = \Delta\widehat{\mathbf{b}}_{k-1} = \mathbf{0}_{3\times1}$. In other words, no information has been introduced into the state $\Delta\mathbf{b}$'s mean and covariance $\mathbf{cov}(\Delta\mathbf{b}, \Delta\mathbf{b})$ during the propagation. Therefore, once we are able to capture the information of $\widehat{\mathbf{a}}$, $\mathbf{cov}(\mathbf{a}, \mathbf{a})$ and $\mathbf{cov}(\mathbf{b}, \mathbf{a})$, we have already obtained all the information available during time propagation.

Then noting (2.35), we find only the attitude-related sigma points $a_{k/k-1}$ are explicitly used to construct the measurement predictions $\mathcal{Z}_{k/k-1}$. We can write a formal expression of this transform as

$$z = \mathbf{h}(\mathbf{x}) = \boldsymbol{\gamma}(\mathbf{a}). \tag{3.1}$$

In fact, (3.1) belongs to a special class of nonlinear transformation, namely, the partially linear transformation [16]. Clearly, for the measurement update process, the random variable z's mean $\overline{z}$, $\mathbf{cov}(z, z)$, and cross covariance $\mathbf{cov}(\mathbf{a}, z)$ are all independent of $\Delta\mathbf{b}$, and it is proved that the cross covariance $\mathbf{cov}(\Delta\mathbf{b}, z)$ is also independent of $\mathbf{cov}(\Delta\mathbf{b}, \Delta\mathbf{b})$ up to the 2nd-order.

The above discussions leads to the following conclusion: as long as we can construct a set of sigma points that matches the given mean estimations of $\widehat{\mathbf{a}}$, $\Delta\widehat{\mathbf{b}}$, and the covariances of $\mathbf{P}^A = \widehat{\mathbf{cov}(\mathbf{a}, \mathbf{a})}$ and $\mathbf{P}^{BA} = \widehat{\mathbf{cov}(\Delta\mathbf{b}, \mathbf{a})}$, it is enough to capture the first two moments' statistics properties of the random state $\mathbf{x}$ with a precision up to 2nd-order. Before moving on to construct such a sigma point set, we should also note from (2.26), (2.37) and (2.39) that with the reset and de-bias steps embeding the filter, actually we are on the assumption that we have an unbiased distribution as $\mathbf{x} = [\mathbf{a}^T, \Delta\mathbf{b}^T]^T = [\mathbf{0}_{3\times1}^T, \mathbf{0}_{3\times1}^T]^T$.

So now our goal becomes to construct a minimum set of sigma points, which is able to fully capture all the available information given as (a) unbiased mean estimation, that is, $\overline{\mathbf{a}} = \mathbf{0}_{3\times1}$, $\Delta\overline{\mathbf{b}} = \mathbf{0}_{3\times1}$, and (b) the covariance estimation as $\mathbf{P}^A = \mathbf{cov}(\mathbf{a}, \mathbf{a})$ $\mathbf{P}^{BA} = \mathbf{cov}(\Delta\mathbf{b}, \mathbf{a})$.

To match the unbiased mean, we have

$$a \cdot \boldsymbol{\mathcal{W}}^T = W_m\left(\boldsymbol{\alpha}^{(1)} + \boldsymbol{\alpha}^{(2)} + \cdots + \boldsymbol{\alpha}^{(m)}\right) = \mathbf{0}, \tag{3.2}$$

$$\mathcal{B} \cdot \boldsymbol{\mathcal{W}}^T = W_m\left(\boldsymbol{\beta}^{(1)} + \boldsymbol{\beta}^{(2)} + \cdots + \boldsymbol{\beta}^{(m)}\right) = \mathbf{0}. \tag{3.3}$$

As stated in [9], to fully capture the mean of an n-dimensional state vector, at least $m = n + 1$ points are needed. Noting both $\mathbf{a}$ and $\Delta\mathbf{b} \in \mathbb{R}^3$, the minimum m which satisfies both (3.2) and (3.3) is $m = 4$.

To further reduce the computational expense and make a better symmetry property, assign equal weights W for all the sigma points. Then we have

$$\sum_{i=1}^{m} W_i = mW = 1. \tag{3.4}$$

Clearly,

$$W = \frac{1}{m} = \frac{1}{4}, \tag{3.5}$$

$$\boldsymbol{\mathcal{W}} = \left(\frac{1}{m}\right)\mathbf{1}_{1\times m}, \qquad \boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}} = \left(\frac{1}{m}\right)\mathbf{I}_m. \tag{3.6}$$

To match $\mathbf{a}$'s covariance estimation $\mathbf{P}^A = \widehat{\mathbf{cov}(\mathbf{a}, \mathbf{a})}$ with outer products approximation, we have

$$a\boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}}(a)^T = \left(\frac{1}{m}\right)a(a)^T = \mathbf{P}^A. \tag{3.7}$$

Again we make use of a base set $\boldsymbol{\mathcal{U}}^a \in \mathbb{R}^{3\times m}$ to help matching $\mathbf{P}^A$. Denoting $\boldsymbol{\mathcal{U}}^a = [\mathbf{u}^{a(1)}, \mathbf{u}^{a(2)}, \mathbf{u}^{a(3)}, \mathbf{u}^{a(4)}]$, the following relationship should be satisfied:

$$\boldsymbol{\mathcal{U}}^a \cdot \boldsymbol{\mathcal{W}}^T = \mathbf{0}, \tag{3.8}$$

$$\boldsymbol{\mathcal{U}}^a \cdot \boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}} \cdot (\boldsymbol{\mathcal{U}}^a)^T = \mathbf{I}_3. \tag{3.9}$$

Substituting (3.5) and (3.6), it is straightforward to get

$$\sum_{i=1}^{m} \mathbf{u}^{a(i)} = \mathbf{0}, \tag{3.10}$$

$$\boldsymbol{\mathcal{U}}^a \cdot (\boldsymbol{\mathcal{U}}^a)^T = m\mathbf{I}_3. \tag{3.11}$$

As $m = 4$, $\boldsymbol{\mathcal{U}}^a$ must be a simplex set. In spite of the constraints in (3.8) and (3.9), we still have the freedom to choose an arbitrary realization of $\boldsymbol{\mathcal{U}}^a$. Later we will propose a novel set

derived from a heuristic geometry approach. Now supposing that $\boldsymbol{\mathcal{U}}^a$ is given, then we may construct a simply as

$$a = \mathbf{S}^A \cdot \boldsymbol{\mathcal{U}}^a, \tag{3.12}$$

where $\mathbf{S}^A$ is an arbitrary matrix square-root of $\mathbf{P}^A$ fulfilling $\mathbf{S}^A(\mathbf{S}^A)^T = \mathbf{P}^A$. It is straightforward to validate (3.12) as substituting it to (3.7):

$$\begin{aligned} a\boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}}(a)^T &= \left(\frac{1}{m}\right)\mathbf{S}^A \cdot \boldsymbol{\mathcal{U}}^a\left(\mathbf{S}^A\boldsymbol{\mathcal{U}}^a\right)^T \\ &= \left(\frac{1}{m}\right)\mathbf{S}^A \cdot \left(\boldsymbol{\mathcal{U}}^a(\boldsymbol{\mathcal{U}}^a)^T\right) \cdot \left(\mathbf{S}^A\right)^T \\ &= \left(\frac{1}{m}\right)\mathbf{S}^A \cdot (m\mathbf{I}_3) \cdot \left(\mathbf{S}^A\right)^T = \mathbf{P}^A. \end{aligned} \tag{3.13}$$

To match the cross covariance $\mathbf{P}^{BA}$ is

$$\boldsymbol{\mathcal{B}}\boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}}(a)^T = \left(\frac{1}{m}\right)\boldsymbol{\mathcal{B}}(a)^T = \mathbf{P}^{BA}. \tag{3.14}$$

Here we propose a simple and convenient algorithm to compute $\boldsymbol{\mathcal{B}}$. Define $\mathbf{S}_{k-1}^{BA} = \mathbf{P}^{BA}(\mathbf{S}^A)^{-T}$, which we could get with the low computational cost Gaussian elimination:

$$\mathbf{S}_{k-1}^{BA} = \mathbf{P}^{BA}\left(\mathbf{S}^A\right)^{-T} = \frac{\mathbf{P}_{k-1}^{BA}}{\left(\mathbf{S}_{k-1}^A\right)^T}. \tag{3.15}$$

Then we have

$$\boldsymbol{\mathcal{B}} = \mathbf{S}_{k-1}^{BA}\boldsymbol{\mathcal{U}}^a. \tag{3.16}$$

Proof of (3.15) is straightforward as substituting it to (3.14):

$$\begin{aligned} \boldsymbol{\mathcal{B}}\boldsymbol{\Lambda}^{\mathcal{W}}(a)^T &= \left(\frac{1}{m}\right)\left(\mathbf{S}_{k-1}^{BA}\boldsymbol{\mathcal{U}}^a\right)\left(\mathbf{S}^A\boldsymbol{\mathcal{U}}^a\right)^T \\ &= \left(\frac{1}{m}\right)\mathbf{S}_{k-1}^{BA} \cdot \left(\boldsymbol{\mathcal{U}}^a(\boldsymbol{\mathcal{U}}^a)^T\right) \cdot \left(\mathbf{S}^A\right)^T \\ &= \left(\frac{1}{m}\right)\mathbf{P}^{BA}\left(\mathbf{S}^A\right)^{-T} \cdot (m\mathbf{I}_3) \cdot \left(\mathbf{S}^A\right)^T \\ &= \mathbf{P}^{BA}\left(\mathbf{S}^A\right)^{-T} \cdot \left(\mathbf{S}^A\right)^T = \mathbf{P}^{BA}. \end{aligned} \tag{3.17}$$

Thereupon, we have constructed a desired set of marginal sigma points as in (3.12) and (3.16). Note that it is enough to use merely one base set $\boldsymbol{\mathcal{U}}^a$ to construct the full-length Sigma point set $\mathcal{X} = [a^T, \boldsymbol{\mathcal{B}}^T]^T$. So hereafter we would suppress the superscript of $\boldsymbol{\mathcal{U}}^a$ as $\boldsymbol{\mathcal{U}}$.

Looking into (2.38), we find

$$\mathbf{P}^X = \begin{bmatrix} \mathbf{P}^A, & \mathbf{P}^{AB} \\ \mathbf{P}^{BA}, & \mathbf{P}^B \end{bmatrix} = \begin{bmatrix} \mathbf{P}^A, & \left(\mathbf{P}^{BA}\right)^T \\ \mathbf{P}^{BA}, & \mathbf{P}^B \end{bmatrix}. \tag{3.18}$$

With the proposed sigma points, the propagation and innovation steps of $\Delta\mathbf{b}$'s covariance estimation $\mathbf{P}^B$ are no longer necessary. Eliminating the $\mathbf{P}^B$-related term and making use of the symmetric structure of the matrix, it is enough to have $\mathbf{P}^A$ and $\mathbf{P}^{BA}$ propagated. Accordingly, it is only necessary to have the same matrices updated. There by, we will replace (2.38) and (2.43) with

$$\begin{aligned}
\mathbf{P}^A_{k/k-1} &= a_{k/k-1} \cdot \boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}} \cdot (a_{k/k-1})^T, \\
\mathbf{P}^{BA}_{k/k-1} &= \boldsymbol{\mathcal{B}}_{k/k-1} \cdot \boldsymbol{\Lambda}^{\boldsymbol{\mathcal{W}}} \cdot (a_{k/k-1})^T, \\
\mathbf{P}^A_k &= \mathbf{P}^A_{k/k-1} - \mathbf{P}^{AZ}_{k/k-1}\left(\mathbf{P}^Z_{k/k-1}\right)^{-1}\left(\mathbf{P}^{AZ}_{k/k-1}\right)^T \\
&= \mathbf{P}^{AA}_{k/k-1} - \mathbf{K}^{AZ}_k\left(\mathbf{P}^{AZ}_{k/k-1}\right)^T, \\
\mathbf{P}^{BA}_k &= \mathbf{P}^{BA}_{k/k-1} - \mathbf{P}^{BZ}_{k/k-1}\left(\mathbf{P}^Z_{k/k-1}\right)^{-1}\left(\mathbf{P}^{AZ}_{k/k-1}\right)^T \\
&= \mathbf{P}^{BA}_{k/k-1} - \mathbf{K}^{BZ}_k\left(\mathbf{P}^{AZ}_{k/k-1}\right)^T.
\end{aligned} \tag{3.19}$$

To avoid state augmentation, we would like to have the propagation noise terms incorporated into the filter with trapezoidal approximation. However, as $\mathbf{P}^B$ is no longer used, we have to seek for alternate approach. Denote

$$\begin{aligned}
\overline{\mathbf{S}}^{QA}_{k-1} &= \sqrt{\overline{\mathbf{Q}}^A_{k-1}} = \sqrt{\left(\frac{T^2}{2}\right) \cdot \left(\sigma^2_{\mathrm{ARW}} - \left(\frac{T^2}{6}\right)\sigma^2_{\mathrm{RRW}}\right)} \cdot \mathbf{I}_3, \\
\overline{\mathbf{S}}^{QB}_{k-1} &= \sqrt{\overline{\mathbf{Q}}^B_{k-1}} = \left(\sigma_{\mathrm{RRW}}\sqrt{\frac{T}{2}}\right) \cdot \mathbf{I}_3.
\end{aligned} \tag{3.20}$$

Then similar to [11], we can add the noise terms directly to the sigma points with the help of $\boldsymbol{\mathcal{U}}$:

$$a_{k-1} = \mathbf{S}^A_{k-1}\boldsymbol{\mathcal{U}} + \mathbf{S}^{QA}_{k-1}\boldsymbol{\mathcal{U}} = \left(\mathbf{S}^A_{k-1} + \mathbf{S}^{QA}_{k-1}\right)\boldsymbol{\mathcal{U}}, \tag{3.21}$$

$$\boldsymbol{\mathcal{B}}_{k-1} = \mathbf{S}^{BA}_{k-1}\boldsymbol{\mathcal{U}} - \mathbf{S}^B_Q\boldsymbol{\mathcal{U}} = \left(\mathbf{S}^{BA}_{k-1} - \mathbf{S}^B_Q\right)\boldsymbol{\mathcal{U}}. \tag{3.22}$$

Table 1: Simplex base sets for 3-dimensional space.

Sigma set	$\mathcal{U} \in \mathbb{R}^{n\times m}$ for $n = 3$
Spherical simplex $m = 5$	$\mathcal{U} = (0.5/\sqrt{W_m})\begin{bmatrix} 0, & -1/\sqrt{2}, & 1/\sqrt{2}, & 0, & 0 \\ 0, & -1/\sqrt{6}, & -1/\sqrt{6}, & 2/\sqrt{6}, & 0 \\ 0, & -1/\sqrt{12}, & -1/\sqrt{12}, & -1/\sqrt{12}, & 3/\sqrt{12} \end{bmatrix}$ $0 < W_0 < 1,\ W_m = (1 - W_0)/(n+1)$
Schmidt orthogonal $m = 4$	$\mathcal{U} = (1/\sqrt{W})\begin{bmatrix} 1/\sqrt{2}, & -1/\sqrt{2}, & 0, & 0 \\ 1/\sqrt{6}, & 1/\sqrt{6}, & -1/\sqrt{3/2}, & 0 \\ 1/\sqrt{12}, & 1/\sqrt{12}, & 1/\sqrt{12}, & -1/\sqrt{4/3} \end{bmatrix}$ $W = 1/(n+1) = 1/4$
Geometric simplex $m = 4$	$\mathcal{U} = \begin{bmatrix} +1, & +1 & -1 & -1 \\ +1, & -1 & -1 & +1 \\ +1, & -1 & +1 & -1 \end{bmatrix}$ $W = 1/(n+1) = 1/4$

Table 2: Residues of numerical mean estimation.

	$\mathbf{S}$	$\begin{bmatrix} 10^{-1}, & 0, & 0; \\ 10^{-2}, & 10^{-1}, & 0; \\ 10^{-3}, & 10^{-4}, & 10^{-1} \end{bmatrix}$	$\begin{bmatrix} 10^{-2}, & 0, & 0; \\ 10^{-3}, & 10^{-2}, & 0; \\ 10^{-4}, & 10^{-5}, & 10^{-2} \end{bmatrix}$	$\begin{bmatrix} 10^{-4}, & 0, & 0; \\ 10^{-5}, & 10^{-4}, & 0; \\ 10^{-6}, & 10^{-7}, & 10^{-4} \end{bmatrix}$
Mean	SS	2.151464×10^{-18}	1.319254×10^{-19}	5.591741×10^{-22}
	SO	1.734723×10^{-18}	6.505213×10^{-19}	1.694066×10^{-21}
	GS	0 (precise)	0 (precise)	0 (precise)
Covariance	SS	3.878960×10^{-18}	4.065758×10^{-19}	8.271806×10^{-24}
	SO	4.249187×10^{-18}	3.030437×10^{-19}	7.018860×10^{-24}
	GS	1.734723×10^{-18}	0 (precise)	2.339620×10^{-24}

SS: Spherical simplex set. SO: Schmidt orthogonal set. GS: Geometric simplex set.

Note in (3.22) that the sign before $\mathbf{S}_Q^B$ is negative. The reason is that with a given covariance estimation $\mathbf{P}$ and its square-root $\mathbf{S}$, it is equivalent for us to choose either $\mathbf{S}$ or $-\mathbf{S}$ when constructing the sigma points. Both "directions" work because the given mean is unbiased. But when we add noise terms directly to the sigma points as in (3.22), we must guarantee that $\mathbf{S}_{k-1}^{BA}$ and $\mathbf{S}_Q^B$ have a common sign. Recalling (25) as

$$\mathbf{Q}_{k-1}^{BA} = -\left(\frac{1}{2}\right)\sigma_{\mathrm{RRW}}^2 T^2 \mathbf{I}_3, \tag{3.23}$$

here the negative sign in the right side of (3.23) clearly indicates opposite signs between $\mathbf{S}_{k-1}^{BA}$ and $\mathbf{S}_Q^B$.

Table 3: Comparisons of total arithmetic operations.

Algorithm	Phase	Multiplies	Adds	Square roots
MEKF	Propagation	350	270	2
	Measurement update	260	185	1
	One full filter step*	*610*	*455*	*3*
	One observation circle†	**1680**	**1265**	**9**
MGSPF	Propagation	350	285	5
	Measurement update	380	280	1
	One full filter step	*750*	*565*	*6*
	One observation circle	**1860**	**1480**	**21**
SSUKF	Propagation	810	565	8
	Measurement update	455	370	1
	One full filter step	*1265*	*935*	*9*
	One observation circle	**3695**	**2630**	**33**

*Including 1 propagation step and 1 measurement update step.
†Including several propagation steps and 1 measurement update step. In this case we take a typical value as 4 propagations for each 1 update.

3.2. The Geometric Simplex Sigma Points

Construction of the base set $\boldsymbol{\mathcal{U}}$ plays an important role in simplex sigma point algorithms. Existing strategies had mainly focused on the design of general operation flow for getting a base set for any arbitrary dimension n. In [9, 10], direction-extending technique is developed and used to build the minimal-skew and spherical simplex set. While in [14], Schmidt orthogonalization is employed to develop a new set. Table 2 demonstrates both the spherical simplex set [10] and the Schmidt orthogonal simplex set [14] for $n = 3$. Close comparison could reveal that both sets are equivalent after a sign shift except for the existence of an additional central point in the spherical simplex set. Both sets are easy to be extended to higher dimension space, and because all the points are equally weighted and equidistantly placed on a hyper sphere, they are numerically stable over the increase of n.

However, both sets lack numerical accuracy, and they are complex to compute, as irrational numbers $\sqrt{2}$, $\sqrt{6}$, and so on exist. Further, they do not have symmetric structures. A fully symmetric set needs that for every point $\mathbf{u}^{(i)} \in \boldsymbol{\mathcal{U}}$, we can get another point $\mathbf{u}^{(j)} \in \boldsymbol{\mathcal{U}}$, $i \neq j$ simply by element permutation or sign-changing point of the generator point $\mathbf{u}^{(i)}$ [15]. Clearly, except for the first dimension (or describing in matrix language, the first row of $\boldsymbol{\mathcal{U}}$), not even element level symmetry is guaranteed in the spherical or Schmidt orthogonal simplex sets.

In order to make a better symmetry property, we propose a new base set of sigma points here as in the 3rd row of Table 1 and Figure 1. It is straight forward to find that both (3.10) and (3.11) are completely satisfied. For clarity, we name this new set as "the Geometric Simplex Set," for it has a nice symmetrical structure as a tetrahedron in the 3-dimensional Euclidean space (Figure 1). The proposed new sigma point set has several benefits.

(i) The new set is more intuitive to comprehend and apply, especially for the 3-dimensional Euclidean space, the true space where we are, and the true space in which a variety of dynamical problems as guidance, navigation, and so on take place.

Table 4: Pseudocode and computational expense evaluation of the MEKF with closed-form solutions.

	Algorithm	×	+	$\sqrt{\ }$
Initialize	$\widehat{\mathbf{x}}_0 = [\widehat{\mathbf{a}}_0^T, \Delta\widehat{\mathbf{b}}_0^T]^T = \mathbf{0}_{6\times1}$, $\widehat{\overline{\mathbf{q}}}_0 = [\mathbf{0}_{3\times1}^T, 1]^T$, $\widehat{\mathbf{b}}_0 = \mathbf{0}_{3\times1}$			
	$\mathbf{P}_0^A = \varepsilon_a^2\mathbf{I}_3$, $\mathbf{P}_0^B = \varepsilon_b^2\mathbf{I}_3$,			
	$\mathbf{Q}_k^A = T(\sigma_{\mathrm{ARW}}^2 + (1/3)\sigma_{\mathrm{RRW}}^2T^2)\mathbf{I}_3$			
	$\mathbf{Q}_k^{BA} = -(1/2)\sigma_{\mathrm{RRW}}^2T^2\mathbf{I}_3$, $\mathbf{Q}_k^B = \sigma_{\mathrm{RRW}}^2T\mathbf{I}_3$			
Propagation	$\widehat{\boldsymbol{\omega}}_{k-1} = \boldsymbol{\omega}_{k-1}^{\mathrm{meas}} - \widehat{\mathbf{b}}_{k-1}$, $\widehat{\omega}_{k-1} = \lvert\widehat{\boldsymbol{\omega}}_{k-1}\rvert$, $\widehat{\phi}_{k-1} = \widehat{\omega}_{k-1}T$	6	5	1
	$\widehat{\overline{\mathbf{q}}}_{k-1}^{\boldsymbol{\omega}} = [8\widehat{\boldsymbol{\phi}}_{k-1}^T, (16 - \widehat{\phi}_{k-1}^2)]^T/(16 + \widehat{\phi}_{k-1}^2)$	6	2	
	$\widehat{\mathbf{n}} = \widehat{\boldsymbol{\omega}}_{k-1}/\widehat{\omega}_{k-1}$	3		
	$\boldsymbol{\Phi}_{k-1} = \mathbf{I}_3 - \sin\widehat{\phi}_{k-1}[\widehat{\mathbf{n}}\times] + (1 - \cos\widehat{\phi}_{k-1})[\widehat{\mathbf{n}}\times]^2$	49	30	
	$\approx \mathbf{I}_3 - (\widehat{\phi}_{k-1} - (\widehat{\phi}_{k-1}^3/6))[\widehat{\mathbf{n}}\times] + (\widehat{\phi}_{k-1}^2/2)[\widehat{\mathbf{n}}\times]^2$			
	$\boldsymbol{\Psi}_{k-1} = (T/2)\mathbf{I}_3 - ((1 - \cos\widehat{\phi}_{k-1})/\widehat{\omega}_{k-1})[\widehat{\mathbf{n}}_{k-1}\times]$			
	$+((\widehat{\phi}_{k-1} - \sin\widehat{\phi}_{k-1})/\widehat{\omega}_{k-1})[\widehat{\mathbf{n}}_{k-1}\times]^2$	51	30	
	$\approx (T/2)\mathbf{I}_3 - (\widehat{\phi}_{k-1}^2/2\widehat{\omega}_{k-1})[\widehat{\mathbf{n}}_{k-1}\times] + (\widehat{\phi}_{k-1}^3/6\widehat{\omega}_{k-1})[\widehat{\mathbf{n}}_{k-1}\times]^2$			
	$\mathbf{P}_{k/k-1}^A = \boldsymbol{\Phi}_{k-1}\mathbf{P}_{k-1}^A\boldsymbol{\Phi}_{k-1}^T + \boldsymbol{\Psi}_{k-1}\mathbf{P}_{k-1}^B\boldsymbol{\Psi}_{k-1}^T$	108	81	
	$+\boldsymbol{\Psi}_{k-1}\mathbf{P}_{k-1}^{BA}\boldsymbol{\Phi}_{k-1}^T + (\boldsymbol{\Psi}_{k-1}\mathbf{P}_{k-1}^{BA}\boldsymbol{\Phi}_{k-1}^T)^T + \mathbf{Q}_k^A$	54	57	
	$\mathbf{P}_{k/k-1}^{BA} = \mathbf{P}_{k-1}^{BA}\Phi_{k-1}^T + \mathbf{P}_{k-1}^B\boldsymbol{\Psi}_{k-1}^T + \mathbf{Q}_k^{BA}$	54	48	
	$\mathbf{P}_{k/k-1}^B = \mathbf{P}_{k-1}^B + \mathbf{Q}_k^B$		3	
	$\widehat{\overline{\mathbf{q}}}_{k/k-1} = \widehat{\overline{\mathbf{q}}}_{k-1} \otimes \widehat{\overline{\mathbf{q}}}_{k-1}^{\boldsymbol{\omega}}$, $\widehat{\overline{\mathbf{q}}}_{k/k-1} = \widehat{\overline{\mathbf{q}}}_{k/k-1}/\lVert\widehat{\overline{\mathbf{q}}}_{k/k-1}\rVert$	22	15	1
	$\widehat{\mathbf{b}}_{k/k-1} = \widehat{\mathbf{b}}_{k-1}$			
		353	**271**	**2**
Measurement update	$\mathbf{P}_{k/k-1}^Z = \mathbf{H}_k\mathbf{P}_{k/k-1}^X\mathbf{H}_k^T + \mathbf{R}_k = \mathbf{P}_{k/k-1}^A + \mathbf{R}_k$		3	
	$\mathbf{P}_{k/k-1}^{AZ} = \mathbf{P}_{k/k-1}^A\mathbf{H}_k^T = \mathbf{P}_{k/k-1}^A$			
	$\mathbf{P}_{k/k-1}^{BZ} = \mathbf{P}_{k/k-1}^{BA}\mathbf{H}_k^T = \mathbf{P}_{k/k-1}^{BA}$			
	$\mathbf{K}_k^{AZ} = \mathbf{P}_{k/k-1}^A(\mathbf{P}_{k/k-1}^Z)^{-1}$, $\mathbf{K}_k^{BZ} = \mathbf{P}_{k/k-1}^{BA}(\mathbf{P}_{k/k-1}^Z)^{-1}$	114	58	
	$\delta\overline{\mathbf{q}}_k^{\mathrm{meas}} = \widehat{\overline{\mathbf{q}}}_{k/k-1}^{-1} \otimes \overline{\mathbf{q}}_k^{\mathrm{meas}}$	16	12	
	$z_k^{\mathrm{meas}} = 4\mathbf{q}_k^{\mathrm{meas}}/(1 + q_{4_k}^{\mathrm{meas}})$, $\boldsymbol{\gamma}_k^{\mathrm{meas}} = \hat{z}_k^{\mathrm{meas}}$	5	4	
	$\widehat{\mathbf{a}}_k = \mathbf{K}_k^{AZ}\boldsymbol{\gamma}_k^{\mathrm{meas}}$, $\Delta\widehat{\mathbf{b}}_k = \mathbf{K}_k^{BZ}\boldsymbol{\gamma}_k^{\mathrm{meas}}$	18	9	
	$\mathbf{P}_k^A = \mathbf{P}_{k/k-1}^A - \mathbf{K}_k^{AZ}(\mathbf{P}_{k/k-1}^{AZ})^T$	27	27	
	$\mathbf{P}_k^{BA} = \mathbf{P}_{k/k-1}^{BA} - \mathbf{K}_k^{BZ}(\mathbf{P}_{k/k-1}^{AZ})^T$	27	27	
	$\mathbf{P}_k^B = \mathbf{P}_{k/k-1}^B - \mathbf{K}_{k/k-1}^{BZ}(\mathbf{P}_{k/k-1}^{BZ})^T$	27	27	
	$\widehat{\overline{\mathbf{q}}}_k = \widehat{\overline{\mathbf{q}}}_{k/k-1} \otimes \delta\widehat{\overline{\mathbf{q}}}(\widehat{\mathbf{a}}_k)$, $\widehat{\overline{\mathbf{q}}}_k = \widehat{\overline{\mathbf{q}}}_k/\lVert\widehat{\overline{\mathbf{q}}}_k\rVert$	22	15	1
	$\widehat{\mathbf{b}}_k = \widehat{\mathbf{b}}_{k/k-1} + \Delta\widehat{\mathbf{b}}_k$		3	
		256	**185**	**1**

Table 5: Pseudocode and computational expense evaluation of the MGSPF algorithm.

	Algorithm	×	+	$\sqrt{}$
Initialize	$\widehat{\mathbf{x}}_0 = [\widehat{\mathbf{a}}_0^T, \Delta\widehat{\mathbf{b}}_0^T]^T = \mathbf{0}_{6\times1}$, $\widehat{\overline{\mathbf{q}}}_0 = [\mathbf{0}_{3\times1}^T, 1]^T$, $\widehat{\mathbf{b}}_0 = \mathbf{0}_{3\times1}$			
	$\mathbf{P}_0^A = \varepsilon_a^2\mathbf{I}_3$, $\mathbf{P}_0^{BA} = \varepsilon_a\varepsilon_b\mathbf{I}_3$ $\quad m = \mathbf{4}$			
	$\overline{\mathbf{S}}_{k-1}^{QA} = T\sqrt{((6\sigma_{\mathrm{ARW}}^2 - T^2\sigma_{\mathrm{RRW}}^2)/12)}\mathbf{I}_3$ $\overline{\mathbf{S}}_{k-1}^{QB} = (\sigma_{\mathrm{RRW}}\sqrt{T/2})\mathbf{I}_3$			
Propagation	$\mathbf{S}_{k-1}^A = \sqrt{\mathbf{P}_{k-1}^A}$, $\mathbf{S}_{k-1}^{BA} = \mathbf{P}_{k-1}^{BA}/(\mathbf{S}_{k-1}^A)^T$	25	13	3
	$a_{k-1} = (\mathbf{S}_{k-1}^A + \mathbf{S}_{k-1}^{QA})\boldsymbol{\mathcal{U}}^{GS}$, $\boldsymbol{\mathcal{B}}_{k-1} = (\mathbf{S}_{k-1}^{BA} - \mathbf{S}_{k-1}^{QB})\boldsymbol{\mathcal{U}}^{GS}$	0	54	
	$\widehat{\boldsymbol{\omega}}_{k-1} = \boldsymbol{\omega}_{k-1}^{\mathrm{meas}} - \widehat{\mathbf{b}}_{k-1}$, $\widehat{\omega}_{k-1} = \lvert\widehat{\boldsymbol{\omega}}_{k-1}\rvert$, $\widehat{\boldsymbol{\phi}}_{k-1} = \widehat{\omega}_{k-1}T$	6	5	1
	$\widehat{\overline{\mathbf{q}}}_{k-1}^{\boldsymbol{\omega}} = [8\widehat{\boldsymbol{\phi}}_{k-1}^T, (16 - \widehat{\phi}_{k-1}^2)]^T/(16 + \widehat{\phi}_{k-1}^2)$	6	2	
	For i = 1 : m, $\quad \boldsymbol{\phi}_{k-1}^{\boldsymbol{\beta}(i)} = (\widehat{\boldsymbol{\omega}}_{k-1} - \boldsymbol{\beta}_{k-1}^{(i)})T$	$3m$	$3m$	
	$\overline{\mathbf{q}}_{k-1}^{\boldsymbol{\beta}(i)} = [8\boldsymbol{\phi}_{k-1}^{\boldsymbol{\beta}(i)T}, (16 - \phi_{k-1}^{\boldsymbol{\beta}(i)2})]^T/(16 + \phi_{k-1}^{\boldsymbol{\beta}(i)2})$	$6m$	$2m$	
	$\overline{\mathbf{q}}_{k-1}^{\boldsymbol{\alpha}(i)} = [8\boldsymbol{\alpha}_{k-1}^{(i)T}, (16 - \alpha_{k-1}^{(i)2})]^T/(16 + \alpha_{k-1}^{(i)2})$	$6m$	$2m$	
	$\overline{\mathbf{q}}_{k/k-1}^{\boldsymbol{\alpha}(i)} = (\widehat{\overline{\mathbf{q}}}_{k-1}^{\boldsymbol{\omega}})^{-1} \otimes \overline{\mathbf{q}}_{k-1}^{\boldsymbol{\alpha}(i)} \otimes \overline{\mathbf{q}}_{k-1}^{\boldsymbol{\beta}(i)}$	$32m$	$24m$	
	$\boldsymbol{\alpha}_{k/k-1}^{(i)} = 4\mathbf{q}_{k/k-1}^{\boldsymbol{\alpha}(i)}/(1 + q_{4\,k/k-1}^{\boldsymbol{\alpha}(i)})$ $\quad$ **end**	$5m$	$1m$	
	$\boldsymbol{\mathcal{B}}_{k/k-1} = \boldsymbol{\mathcal{B}}_{k-1}$, $\boldsymbol{\mathcal{Z}}_{k/k-1} = a_{k/k-1}$			
	$\widehat{\mathbf{a}}_{k/k-1} = (1/m)\, a_{k/k-1} \cdot \mathbf{1}_{1\times m\ 1\times m}$, $\Delta\widehat{\mathbf{b}}_{k/k-1} = \Delta\widehat{\mathbf{b}}_{k-1} = \mathbf{0}_{3\times1}$	3	$3(m-1)$	
	$a_{k/k-1} = a_{k/k-1} - \mathbf{1}_{1\times m} \circ \widehat{\mathbf{a}}_{k/k-1}$		$3m$	
	$\mathbf{P}_{k/k-1}^A = (1/m)\, a_{k/k-1}(a_{k/k-1})^T$	$6 + 6m$	$6(m-1)$	
	$\mathbf{P}_{k/k-1}^{BA} = (1/m)\boldsymbol{\mathcal{B}}_{k/k-1}(a_{k/k-1})^T$	$9 + 9m$	$9(m-1)$	
	$\widehat{\overline{\mathbf{q}}}_{k/k-1} = \widehat{\overline{\mathbf{q}}}_{k-1} \otimes \widehat{\overline{\mathbf{q}}}_{k-1}^{\boldsymbol{\omega}}$, $\widehat{\overline{\mathbf{q}}}_{k/k-1} = \widehat{\overline{\mathbf{q}}}_{k/k-1}/\lVert\widehat{\overline{\mathbf{q}}}_{k/k-1}\rVert$	22	15	1
	$\widehat{\mathbf{b}}_{k/k-1} = \widehat{\mathbf{b}}_{k-1}$			
		For $m = 4$ **351**	**285**	**5**
Measurement update	$\widehat{z}_{k/k-1} = (1/m)\boldsymbol{\mathcal{Z}}_{k/k-1} \bullet 1_{m\times1}$	3	$3(m-1)$	
	$\boldsymbol{\mathcal{Z}}_{k/k-1} = \boldsymbol{\mathcal{Z}}_{k/k-1} - \mathbf{1}_{1\times m} \circ \widehat{z}_{k/k-1}$		$3m$	
	$\mathbf{P}_{k/k-1}^Z = (1/m)\boldsymbol{\mathcal{Z}}_{k/k-1}\boldsymbol{\mathcal{Z}}_{k/k-1}{}^T + \mathbf{R}_k$	$6 + 6m$	$9(m-1) + 3$	
	$\mathbf{P}_{k/k-1}^{AZ} = (1/m)\, a_{k/k-1}\boldsymbol{\mathcal{Z}}_{k/k-1}{}^T$	$9 + 9m$	$9(m-1)$	
	$\mathbf{P}_{k/k-1}^{BZ} = (1/m)\boldsymbol{\mathcal{B}}_{k/k-1}\boldsymbol{\mathcal{Z}}_{k/k-1}{}^T$	$9 + 9m$	$9(m-1)$	
	$\mathbf{K}_k^{AZ} = \mathbf{P}_{k/k-1}^A(\mathbf{P}_{k/k-1}^Z)^{-1}$, $\mathbf{K}_k^{BZ} = \mathbf{P}_{k/k-1}^{BA}(\mathbf{P}_{k/k-1}^Z)^{-1}$	114	58	
	$\delta\overline{\mathbf{q}}_k^{\mathrm{meas}} = \overline{\mathbf{q}}_{k/k-1}^{-1} \otimes \overline{\mathbf{q}}_k^{\mathrm{meas}}$	16	12	
	$z_k^{\mathrm{meas}} = 4\mathbf{q}_k^{\mathrm{meas}}/(1 + q_{4\,k}^{\mathrm{meas}})$, $\boldsymbol{\gamma}_k^{\mathrm{meas}} = \widehat{z}_k^{\mathrm{meas}} - \widehat{z}_{k/k-1}$	5	4	
	$\widehat{\mathbf{a}}_k = \mathbf{K}_k^{AZ}\boldsymbol{\gamma}_k^{\mathrm{meas}}$, $\Delta\widehat{\mathbf{b}}_k = \mathbf{K}_k^{BZ}\boldsymbol{\gamma}_k^{\mathrm{meas}}$	18	12	
	$\mathbf{P}_k^A = \mathbf{P}_{k/k-1}^A - \mathbf{K}_k^{AZ}(\mathbf{P}_{k/k-1}^{AZ})^T$	27	27	
	$\mathbf{P}_k^{BA} = \mathbf{P}_{k/k-1}^{BA} - \mathbf{K}_k^{BZ}(\mathbf{P}_{k/k-1}^{AZ})^T$	27	27	
	$\mathbf{P}_k^B = \mathbf{P}_{k/k-1}^B - \mathbf{K}_{k/k-1}^{BZ}(\mathbf{P}_{k/k-1}^{BZ})^T$	27	27	
	$\widehat{\overline{\mathbf{q}}}_k = \widehat{\overline{\mathbf{q}}}_{k/k-1} \otimes \delta\overline{\mathbf{q}}(\widehat{\mathbf{a}}_k)$, $\widehat{\overline{\mathbf{q}}}_k = \widehat{\overline{\mathbf{q}}}_k/\lVert\widehat{\overline{\mathbf{q}}}_k\rVert$	22	15	1
	$\widehat{\mathbf{b}}_k = \widehat{\mathbf{b}}_{k/k-1} + \Delta\widehat{\mathbf{b}}_k$		3	
		For $m = 4$ **379**	**278**	**1**

(ii) Lower computation expense and better round-off error behavior. The new set is free from calculating any irrational numbers. Furthermore, as it is only constituted of ±1, we can replace the multiplication operations in (3.12), (3.16) with simple sign changes. By elimination of both irrational number and multiplication, we made (3.12), (3.16) precise, and free from round-off errors.

(iii) The Geometric simplex set has a better symmetrical structure, which would help to further increase the numerical accuracy, including (a) single dimension symmetry completely fulfilled (or in matrix language, each row of $\boldsymbol{\mathcal{U}}$ is constituted with symmetrically distributed elements). (b) interdimensional symmetry (or per mutational symmetry) partly fulfilled. Define the generator point as $\mathbf{u} = [1, 1, -1]^T$, and construct new points from $\mathbf{u}$ by permutation and sign-changing, altogether we can make 8 points occupying the 8 vertices of the unit cube in Figure 1. Note $\boldsymbol{\mathcal{U}}$ has included 4 of them with a symmetric structure, which is enough to capture the random state's first two order moments (mean and variance). Actually, the other 4 points can be found in $-\boldsymbol{\mathcal{U}}$, and clearly, for an unbiased problem, choosing either $\boldsymbol{\mathcal{U}}$ or $-\boldsymbol{\mathcal{U}}$ is equivalent.

Numerical Demonstration

Suppose that we have already obtained a 3-dimensional unbiased state $\mathbf{a}$, covariance $\mathbf{P}$, and its coresponding square-root matrix $\mathbf{S}$. Then we are going to generate a set of sigma points with a base set $\boldsymbol{\mathcal{U}}$ as $a = \mathbf{S} \cdot \boldsymbol{\mathcal{U}}$. As had been claimed, theoretically we should have

$$\begin{aligned} \widehat{\mathbf{a}} &= W_m\left(\boldsymbol{\alpha}^{(1)} + \cdots + \boldsymbol{\alpha}^{(m)}\right) = \mathbf{0}, \\ \widehat{\mathbf{P}} &= a \cdot \boldsymbol{\Lambda}^{\boldsymbol{\mathcal{U}}} \cdot (a)^T = \mathbf{S} \cdot \mathbf{S}^T, \end{aligned} \tag{3.24}$$

where $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{P}}$ represent the result by numerical computation. Then we can take the norm of the residues $|\widehat{\mathbf{a}} - \mathbf{0}|$ and $|\operatorname{diag}(\widehat{\mathbf{P}} - \mathbf{P})|$ as a criterion of a sigma set's numerical accuracy. Three typical base sets, namely, the Spherical Simplex set, the Schmidt Orthogonal set and the new proposed Geometry Simplex set are compared over a series of different S matrices with their diagonal elements ranked form $10^2 \sim 10^{-6}$.

The numerical experiment is programmed with double precision float numbers in MATLAB, and some typical results are listed in Table 2. As can be seen, numerical behaviors of the Geometric simplex set are quite encouraging. On mean computation, both spherical and Schmidt sets introduced a residue error at the scale of about 10^{-16} of the diagonal elements of $\mathbf{S}$, while the geometric set's computed residue had always been precisely 0. On covariance computation, the new set's accuracy is also significantly superior. Biased sets are also studied and the result is similar.

3.3. *The Highly Efficient New Filter*

Incorporating the Geometric Simplex sigma point set into the Marginal SPKF framework, we would have a new nonlinear SPKF estimator for attitude estimation, namely, the Marginal Geometric Sigma Point Filter (MGSPF) algorithm, summarized in Table 5. As a sigma point filtering algorithm, the MGSPF has a significant increase in numerical efficiency, while it still guarantees a same order precision as the traditional SPKF algorithms. A general comparison

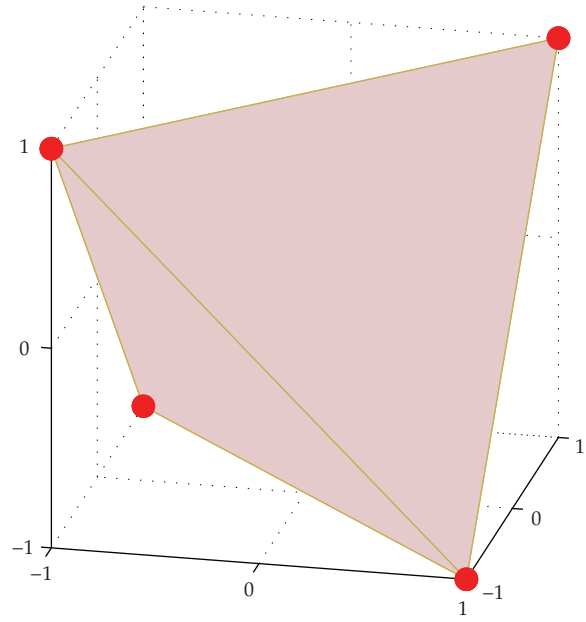


Figure 1: Geometry simplex sigma points in 3-dimensional Euclidean space.

of computation expense is taken between the MEKF and MGSPF algorithms as listed in Tables 4 and 5, and the result is summarized in Table 3.

For the computing effort of the propagation phase, there is little difference between MGSPF and MEKF, both are about half of the SSUKFs. For measurement update phase, the MGSPF takes some more arithmetic operations, but still only 80% of the SSUFK. In fact, if we take into account that in most actual implementations, there exist more propagation steps than observation steps, the total computational expense of MEKF and MGSPF would be very close. It is clear that the MGSPF has achieved a truly rivalizing efficiency as the MEKF, even when simple analytical closed-form solutions are included in the MEKF, and they are almost 50% of the SSUKF.

4. Simulations

In this section we apply the proposed Marginal Geometric Sigma Point Filter (MGSPF) algorithm to the typical stellar-inertial spacecraft attitude determination system with numerical simulations. To give a comparison, the multiple extended Kalman filter (MEKF) and a nonaugmented spherical simplex unscented Kalman filter (SSUKF) with trapezoidal approximation of the propagation noise are also simulated.

Parameters of the simulated model are set as follows. The spacecraft's initiation attitude is $\overline{\mathbf{q}}_0 = [0.1, 0.15, 0.2, 1]^T$, or expressed in 3-1-2 Euler Angles as [14°,15°,21°]. The initial angular velocity is $\boldsymbol{\omega}_0 = [0.05^\circ/\mathrm{s}, 0.1^\circ/\mathrm{s}, 0.15^\circ/s]^T$, and it runs a sinusoidal maneuver at an Amplitude of 0.5°/s and periods of 100 s, 120 s, and 125 s for each axis. The gyroscope is modeled as initial bias 3.4°/s, drift instability (also known as the flicker noise) 0.001°/s; angular random walk (ARW) $1 \times 10^{-3\circ}/\mathrm{s}^{1/2}$, rate random walk $1.4 \times 10^{-3\circ}/\mathrm{s}^{3/2}$, and sampling

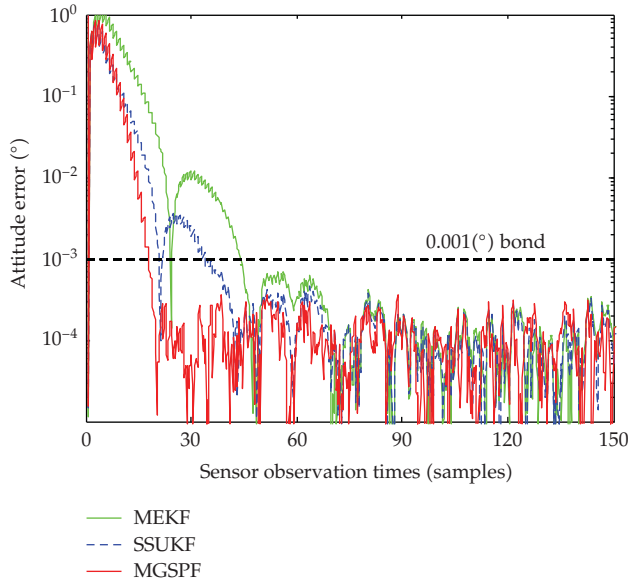


Figure 2: Estimation error history of the attitude.

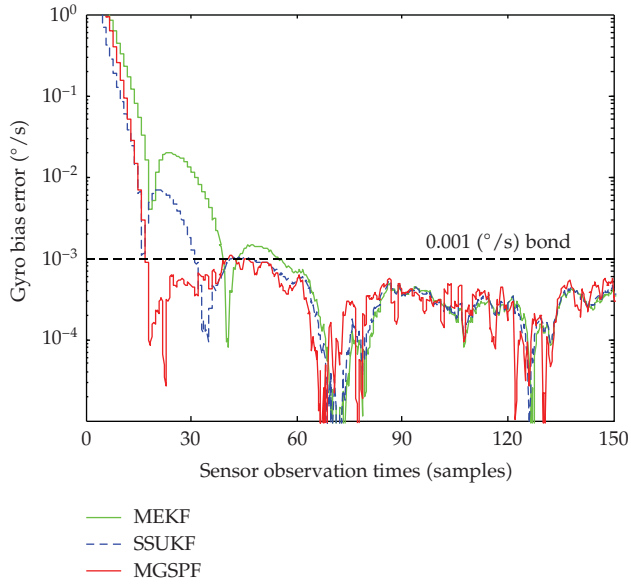


Figure 3: Estimation error history of the gyroscope bias.

frequency 20 Hz. The star sensor is simulated with 1σ accuracy as cross boresight 10 arc-seconds and around boresight 30 arc-seconds, and the sensor's update-rate is 5 Hz.

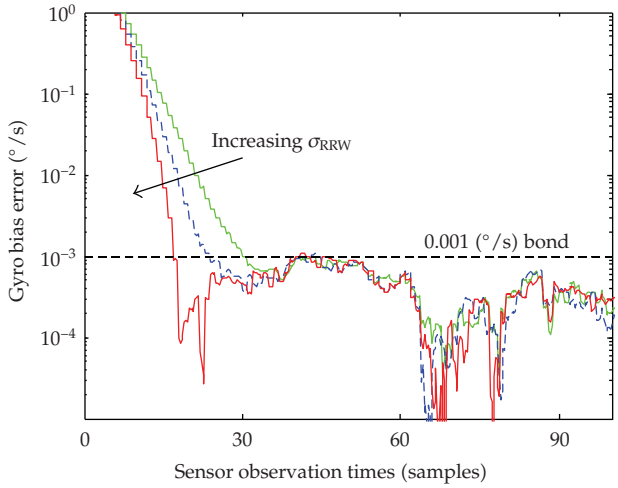


Figure 4: Influence of the bias noise parameter in MGSPF.

The initial states of all filters are set equivalently as $\hat{\mathbf{x}}_0 = [\hat{\mathbf{a}}_0^T, \Delta\hat{\mathbf{b}}_0^T]^T = \mathbf{0}_{6\times1}$, $\hat{\bar{\mathbf{q}}}_0 = [\mathbf{0}_{3\times1}^T, 1]^T$, and $\hat{\mathbf{b}}_0 = \mathbf{0}_{3\times1}$. The initial covariance $\mathbf{P}_0^X$ is set with the attitude-related elements $\mathbf{P}_0^A = (10^\circ)^2\mathbf{I}_3$, and bias-related elements $\mathbf{P}_0^B = (0.1^\circ/s)^2\,\mathbf{I}_3$. For MGSPF, as $\mathbf{P}_0^B$ is no longer used, we equivalently set a $\mathbf{P}_0^{BA} = \sqrt{\mathbf{P}_0^A\mathbf{P}_0^B}$. Specific elements in $\mathbf{R}_k$ and $\mathbf{Q}_k$ are chosen through tuning, set as $\sigma_r = 2\times10^{-7}$, $\sigma_{\mathrm{ARW}} = 4.5\times10^{-4}$, and $\sigma_{\mathrm{RRW}} = 4\times10^{-4}$, respectively, for all three filters.

The simulation results of attitude estimation error and and gyro bias estimation error are, respectively, illustrated in Figures 2 and 3. As the star sensor has a high precision and a 5 Hz Data Update Rate, the three filters' steady-state accuracies are close to each other; so we mainly focus on the initial stage of the estimation process. As shown in Figure 2, of the attitude estimation error, to converge to a value below 0.001°, MEKF takes more than 60 star observations, SSUKF takes about 40, while the MGSPF takes only about 20. Meanwhile, as in Figure 3 of the gyro bias estimation, to achieve an estimation precision of 0.001°/s, MEKF takes more than 50 star samples, SSUKF takes 30, and the MGSPF takes about 20. This indicates that the MGSPF algorithm, once properly implemented, provides a better performance than MEKF at a similar numerical expense, while it is able to achieve, if not better, at least a comparable performance to traditional sigma point filters, at a significant lower expense.

We now address the issue of tuning. The main difference between MGSPF and traditional sigma point filters in parameter selection is mainly reflected in the usage of σ_{RRW}; so we focus on the effect of different σ_{RRW} on the performance of MGSPF. As illustrated in Figure 4, a larger σ_{RRW} has the advantage of enabling a faster convergence or can also be said as an enhancement of the filter's tracking ability. On the other hand, a large σ_{RRW} also has a drawback of instable parameter estimation in steady state; that is, the estimation of $\Delta\mathbf{b}$ will "jump". An optimized value of σ_{RRW} would be a tradeoff between the two. In addition, as the sigma points are created from the covariance matrices which are added with the noise terms arisen from σ_{RRW}, the scale of σ_{RRW} should be kept in a reasonable range that would not obscure the information contained in the covariances.

5. Conclusion

A new, minimum sigma points algorithm for spacecraft attitude and angular rate estimation has been developed. By marginalizing out the linear substructures within the random walk gyro bias model and the attitude involving, only observation model, the new algorithm needs only 4 sigma points to give a complete 6-state attitude and angular rate estimation. The algorithm's computational expense is only 50% of the traditional SSUKF algorithm. It has truly rivaled the MEKF algorithm's computing speed even when simple analytical closed-form solutions are included. Yet it is still able to achieve the same accuracy as traditional unscented Kalman filters.

A new, symmetrical, and numerically more efficient simplex sigma set has been presented. The new set is completely free from irrational numbers and is free from any multiplication operations during sigma point construction. The new set introduces almost none of round-off error for mean reference and smaller error for covariance reference. It would be of use for the implementation in a variety of 3-dimensional Euclidean space involving dynamical problems such as positioning and attitude estimation problems.

With the remarkable reduction in computational expense, the sigma point Kalman filter would gain a significant upgrading in its competitiveness as a candidate algorithm for actual onboard implementation.

References

[1] E. J. Lefferts, F. L. Markley, and M. D. Shuster, "Kalman filtering for spacecraft attitude estimation," *Journal of Guidance, Control, and Dynamics*, vol. 5, no. 5, pp. 417–429, 1982.

[2] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[3] R. D. van Merwe, E. A. Wan, and S. I. Julier, "Sigma-point Kalman filters for nonlinear estimation and sensor-fusion—applications to integrated navigation," in *Proceedings of the Guidance, Navigation, and Control Conference (AIAA '04)*, vol. 3, pp. 1735–1764, Providence, RI, USA, August 2004.

[4] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "New approach for filtering nonlinear systems," in *Proceedings of the American Control Conference (ACC '95)*, vol. 3, pp. 1628–1632, Seattle, Wash, USA, 1995.

[5] J. L. Crassidis and F. L. Markley, "Unscented filtering for spacecraft attitude estimation," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 4, pp. 536–542, 2003.

[6] K. Lai, J. Crassidis, and R. Harman, "In-space spacecraft alignment calibration using the unscented filter," in *Proceedings of the Guidance, Navigation, and Control Conference and Exhibit (AIAA '03)*, Honolulu, Hawaii, USA, August 2003.

[7] J. Côté and J. de Lafontaine, "Magnetic-only orbit and attitude estimation using the square-root unscented Kalman filter: application to the PROBA-2 spacecraft," in *Proceedings of the Guidance, Navigation, and Control Conference and Exhibit (AIAA '08)*, Honolulu, Hawaii, USA, August 2008.

[8] R. van der Merwe and E. Wan, "The square-root unscented Kalman filter for state and parameter-estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '01)*, vol. 6, pp. 3461–3464, Salt Lake City, Utah, USA, May 2001.

[9] S. J. Julier and J. K. Uhlmann, "Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations," in *Proceedings of the American Control Conference (ACC '02)*, vol. 2, pp. 887–892, Anchorage, Alaska, USA, May 2002.

[10] S. J. Julier, "The spherical simplex unscented transformation," in *Proceedings of the American Control Conference (ACC '03)*, vol. 3, pp. 2430–2434, Denver, Colo, USA, June 2003.

[11] J. G. Castrejôn-Lozano, L. R. Garca Carrillo, A. Dzul, and R. Lozano, "Spherical simplex sigma-point Kalman filters: a comparison in the inertial navigation of a terrestrial vehicle," in *Proceedings of the American Control Conference (ACC '08)*, pp. 3536–3541, 2008.

[12] J.-F. Lévesque, "Second-order simplex sigma points for nonlinear estimation," in *Proceedings of the Guidance, Navigation, and Control Conference (AIAA '06)*, vol. 2, pp. 819–830, Keystone, Colo, USA, August 2006.

[13] S. J. Julier, "The scaled unscented transformation," in *Proceedings of the American Control Conference (ACC '02)*, vol. 6, pp. 4555–4559, Anchorage, Alaska, USA, May 2002.

[14] W.-C. Li, P. Wei, and X.-C. Xiao, "A novel simplex unscented transform and filter," in *Proceedings of the International Symposium on Communications and Information Technologies (ISCIT '07)*, pp. 926–931, Sydney, Australia, 2007.

[15] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, 2009.

[16] M. R. Morelande and B. Ristic, "Smoothed state estimation for nonlinear Markovian switching systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 4, pp. 1309–1325, 2008.

[17] F. L. Markley, "Attitude error representations for Kalman filtering," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 2, pp. 311–317, 2003.

[18] P. Singla, *A new attitude determination approach using split field of view star camera*, M.S. thesis, Texas A&M University, College Station, Tex, USA, 2002.

[19] R. L. Farrenkopf, "Analytic steady-state accuracy solutions for two common spacecraft attitude estimators," *Jounal of Guidance and Control*, vol. 1, no. 4, pp. 282–284, 1978.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 242396, 18 pages
doi:10.1155/2009/242396

Research Article

Spin-Stabilized Spacecrafts: Analytical Attitude Propagation Using Magnetic Torques

Roberta Veloso Garcia,[1] Maria Cecília F. P. S. Zanardi,[2] and Hélio Koiti Kuga[1]

[1] *Space Mechanic and Control Division, National Institute for Space Research (INPE), Sao Paulo, São José dos Campos 12227-010, Brazil*
[2] *Department of Mathematics, Group of Orbital Dynamics and Planetology, Sao Paulo State University (UNESP), Sao Paulo, Guaratinguetá 12516-410, Brazil*

Correspondence should be addressed to Roberta Veloso Garcia, rvelosogarcia@yahoo.com.br

Received 28 July 2009; Accepted 4 November 2009

Recommended by Antonio Prado

An analytical approach for spin-stabilized satellites attitude propagation is presented, considering the influence of the residual magnetic torque and eddy currents torque. It is assumed two approaches to examine the influence of external torques acting during the motion of the satellite, with the Earth's magnetic field described by the quadripole model. In the first approach is included only the residual magnetic torque in the motion equations, with the satellites in circular or elliptical orbit. In the second approach only the eddy currents torque is analyzed, with the satellite in circular orbit. The inclusion of these torques on the dynamic equations of spin stabilized satellites yields the conditions to derive an analytical solution. The solutions show that residual torque does not affect the spin velocity magnitude, contributing only for the precession and the drift of the spacecraft's spin axis and the eddy currents torque causes an exponential decay of the angular velocity magnitude. Numerical simulations performed with data of the Brazilian Satellites (SCD1 and SCD2) show the period that analytical solution can be used to the attitude propagation, within the dispersion range of the attitude determination system performance of Satellite Control Center of Brazil National Research Institute.

1. Introduction

This paper aims at analyzing the rotational motion dynamics of spin-stabilized Earth's artificial satellites, through derivation of an analytical attitude prediction. Emphasis is placed on modeling the torques steaming from residual magnetic and eddy currents perturbations, as well as their influences on the satellite angular velocity and space orientation. A spherical coordinated system fixed in the satellite is used to locate the spin axis of the satellite in relation

to the terrestrial equatorial system. The directions of the spin axis are specified by the right ascension (α) and the declination (δ) as represented in Figure 1. The magnetic residual torque occurs due to the interaction between the Earth magnetic field and the residual magnetic moment along the spin axis of the satellite. The eddy currents torque appears due to the interaction of such currents circulating along the satellite structure chassis and the Earth's magnetic field.

The torque analysis is performed through the quadripole model for the Earth's magnetic field and the satellite in circular and elliptical orbits. Essentially an analytical averaging method is applied to determine the mean torque over an orbital period.

To compute the average components of both the residual magnetic and eddy current torques in the satellite body frame reference system (satellite system), an average time in the fast varying orbit element, the mean anomaly, is utilized. This approach involves several rotation matrices, which are dependent on the orbit elements, right ascension and declination of the satellite spin axis, the magnetic colatitudes, and the longitude of ascending node of the magnetic plane.

Unlike the eddy currents torques, it is observed that the residual magnetic torque does not have component along the spin axis; however, it has nonzero components in satellite body x-axis and y-axis. Afterwards, the inclusion of such torques on the rotational motion differential equations of spin-stabilized satellites yields the conditions to derive an analytical solution [1]. The theory is developed accounting also for orbit elements time variation, not restricted to circular orbits, giving rise to some hundreds of curvature integrals solved analytically.

In order to validate the analytical approach, the theory developed has been applied for the spin-stabilized Brazilian Satellites (SCD1 and SCD2), which are quite appropriated for verification and comparison of the theory with the data generated and processed by the Satellite Control Center (SCC) of Brazil National Research Institute (INPE). The oblateness of the orbital elements is taken into account.

The behaviors of right ascension, declination, and spin velocity of the spin axis with the time are presented and the results show the agreement between the analytical solution and the actual satellite behavior.

2. Geomagnetic Field

It is well known that the Earth's magnetic field can be obtained by the gradient of a scalar potential V [2]; it means that

$$\overrightarrow{B} = -\nabla V, \tag{2.1}$$

with the magnetic potential V given by

$$V(r', \phi, \theta) = r_T \sum_{n=1}^{k} \left(\frac{r_T}{r}\right)^{n+1} \sum_{m=0}^{n} \left(g_n^m \cos m\theta + h_n^m \operatorname{sen} m\theta\right) P_n^m(\phi), \tag{2.2}$$

where r_T is the Earth's equatorial radius, g_n^m, h_n^m are the Gaussian coefficients, $P_n^m(\phi)$ are the Legendre associated polynomial and r, ϕ, θ mean the geocentric distance, the local colatitudes, and local longitude, respectively.

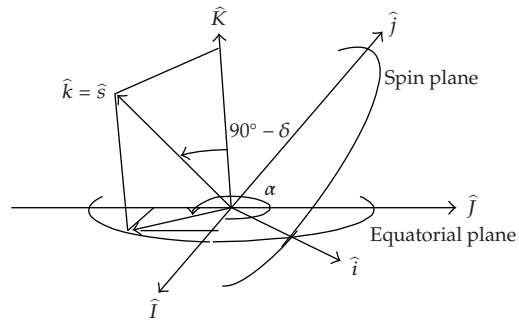


Figure 1: Orientation of the spin axis ($\hat{s}$): equatorial system ($\hat{I}, \hat{J}, \hat{K}$), satellite body frame reference system ($\hat{i}, \hat{j}, \hat{k}$), right ascension ($\alpha$), and declination ($\delta$) of the spin axis.

In terms of spherical coordinates, the geomagnetic field can be expressed by [2],

$$\overrightarrow{B} = B_r \hat{r} + B_\phi \hat{\phi} + B_\theta \hat{\theta}, \tag{2.3}$$

with

$$B_r = -\frac{\partial V}{\partial r}, \qquad B_\phi = -\frac{1}{r}\frac{\partial V}{\partial \phi}, \qquad B_\theta = -\frac{1}{r \operatorname{sen} \phi}\frac{\partial V}{\partial \theta}. \tag{2.4}$$

For the quadripole model, it is assumed that n equals 1 and 2 and m equals 0, 1 and 2 in (2.2). After straightforward computations, the geomagnetic field can be expressed by [3, 4]

$$B_r = 2\left(\frac{r_T}{r}\right)^3 f_1(\theta, \phi) + 3\left(\frac{r_T}{r}\right)^4 f_2(\theta, \phi), \tag{2.5}$$

$$B_\phi = -\left(\frac{r_T}{r}\right)^3 f_3(\theta, \phi) - \left(\frac{r_T}{r}\right)^4 f_4(\theta, \phi), \tag{2.6}$$

$$B_\theta = -\frac{1}{\operatorname{sen} \phi}\left\{\left(\frac{r_T}{r}\right)^3 f_5\ (\theta, \phi) + \left(\frac{r_T}{r}\right)^4 f_6(\theta, \phi) + 2\left(\frac{r_T}{r}\right)^4 f_7(\theta, \phi)\right\}, \tag{2.7}$$

where the functions f_i, $i = 1, 2, \ldots, 7$, are shown in [3] and depend on the Gaussian coefficients g_2^2, h_1^1, h_2^1, h_2^2 .

In the Equator reference system, the geomagnetic field is expressed by [2]

$$B_X = \left(B_r \cos \overline{\delta} + B_\phi \operatorname{sen} \overline{\delta}\right) \cos \overline{\alpha} - B_\theta \operatorname{sen} \overline{\alpha}, \tag{2.8}$$

$$B_Y = \left(B_r\ \cos \overline{\delta}\ + B_\phi \operatorname{sen} \overline{\delta}\right) \operatorname{sen} \overline{\alpha} - B_\theta \cos \overline{\alpha}, \tag{2.9}$$

$$B_Z = B_r \operatorname{sen} \overline{\delta} + B_\phi \cos \overline{\delta}, \tag{2.10}$$

where $\overline{\alpha}$ and $\overline{\delta}$ are the right ascension and declination of the satellite position vector, respectively, which can be obtained in terms of the orbital elements; B_r, B_ϕ, and B_θ are given by (2.5), (2.6), and (2.7), respectively.

In a satellite reference system, in which the axis z is along the spin axis, the geomagnetic field is given by [4, 5]

$$\overrightarrow{B} = B_x\hat{i} + B_y\hat{j} + B_z\hat{k}, \tag{2.11}$$

where

$$\begin{aligned} B_x &= -\ B_X \operatorname{sen}\alpha + B_Y \cos\alpha, \\ B_y &= -B_X \operatorname{sen}\delta\ \cos\alpha - B_Y \operatorname{sen}\delta \operatorname{sen}\alpha + B_Z \cos\delta, \\ B_z &= -\ B_X\ \cos\delta \cos\alpha - B_Y \cos\delta \operatorname{sen}\alpha + B_Z \operatorname{sen}\delta, \end{aligned} \tag{2.12}$$

with B_X, B_Y, and B_Z given by (2.8)–(2.10).

3. Residual and Eddy Currents Torques

Magnetic residual torques result from the interaction between the spacecraft's residual magnetic field and the Earth's magnetic fields. If $\overrightarrow{m}$ is the magnetic moment of the spacecraft and $\overrightarrow{B}$ is the geomagnetic field, then the residual magnetic torques are given by [2]

$$\overrightarrow{N}_r = \overrightarrow{m} \times \overrightarrow{B}. \tag{3.1}$$

For the spin-stabilized satellite, with appropriate nutation dampers, the magnetic moment is mostly aligned along the spin axis and the residual torque can be expressed by [5]

$$\overrightarrow{N}_r = M_s\hat{k} \times \overrightarrow{B}, \tag{3.2}$$

where M_s is the satellite magnetic moment along its spin axis and $\hat{k}$ is the unit vector along the spin axis of the satellite.

By substituting the geomagnetic field (2.11) in (3.1), the instantaneous residual torque is expressed by

$$\overrightarrow{N}_r = M_s\left(-\ B_y\hat{i} + B_x\hat{j}\right). \tag{3.3}$$

On the other hand, the eddy currents torque is caused by the spacecraft spinning motion. If $\overrightarrow{W}$ is the spacecraft's angular velocity vector and p is the Foucault parameter representing the geometry and material of the satellite chassis [2], then this torque may be modeled by [2]

$$\overrightarrow{N}_i = p\overrightarrow{B} \times \left(\overrightarrow{\mathbf{B}}\ \times \overrightarrow{W}\right). \tag{3.4}$$

For a spin-stabilized satellite, the spacecraft's angular velocity vector and the satellite magnetic moment, along the z-axis and induced eddy currents torque, can be expressed by [5, 6]

$$\overrightarrow{N}_i = pW\left(-B_x\, B_z\hat{i} - B_yB_z\ \hat{j} + \left(B_y^2 + B_x^2\right)\hat{k}\right). \tag{3.5}$$

4. Mean Residual and Eddy Currents Torques

In order to obtain the mean residual and eddy currents torques, it is necessary to integrate the instantaneous torques $\overrightarrow{N}_r$ and $\overrightarrow{N}_i$, given in (3.3) and (3.5), over one orbital period T as

$$\overrightarrow{N}_{r_m} = \frac{1}{T}\int_{t_i}^{t_i+T}\overrightarrow{N}_r dt, \qquad \overrightarrow{N}_{i_m} = \frac{1}{T}\int_{t_i}^{t_i+T}\overrightarrow{N}_i dt, \tag{4.1}$$

where t is the time t_i the initial time, and T the orbital period. Changing the independent variable to the fast varying true anomaly, the mean residual and eddy currents torque can be obtained by [4]

$$\overrightarrow{N}_{r_m} = \frac{1}{T}\int_{\upsilon_i}^{\upsilon_i+2\pi}\overrightarrow{N}_r\frac{r^2}{h}d\upsilon, \qquad \overrightarrow{N}_{i_m} = \frac{1}{T}\int_{\upsilon_i}^{\upsilon_i+2\pi}\overrightarrow{N}_i\frac{r^2}{h}d\upsilon, \tag{4.2}$$

where υ_i is the true anomaly at instant t_i, r is the geocentric distance, and h is the specific angular moment of orbit.

To evaluate the integrals of (4.2), we can use spherical trigonometry properties, rotation matrix associated with the references systems, and the elliptic expansions of the true anomaly in terms of the mean anomaly [7], including terms up to first order in the eccentricity (e). Without losing generality, for the sake of simplification of the integrals, we consider the initial time for integration equal to the instant that the satellite passes through perigee. After extensive but simple algebraic developments, the mean residual and eddy currents torques can be expressed by [3, 6]

$$\overrightarrow{N}_{r_m} = N_{rxm}\hat{i} + N_{rym}\hat{j}, \qquad \overrightarrow{N}_{i_m} = \frac{pW}{2\pi}\left(N_{ixm}\hat{i} + N_{iym}\hat{j} + N_{izm}\hat{k}\right), \tag{4.3}$$

with

$$\begin{aligned} N_{rxm} &= \frac{M_s}{2\pi}(A\,\text{sen}\,\delta\cos\alpha + B\,\text{sen}\,\delta\,\text{sen}\,\alpha - C\,\text{sen}\,\delta), \\ N_{rym} &= \frac{M_s}{2\pi}(-\,D\,\text{sen}\,\alpha + E\,\cos\,\delta) \end{aligned} \tag{4.4}$$

and N_{ixm}, N_{iym}, N_{izm} as well as the coefficients A, B, C, D, and E are presented in the appendix. It is important to observe that the mean components of these torques depend on the attitude angles (δ, α) and the orbital elements (orbital major semi-axis: a, orbital eccentricity: e, longitude of ascending node: Ω, argument of perigee: ω, and orbital inclination: i).

5. The Rotational Motion Equations

The variations of the angular velocity, the declination, and the ascension right of the spin axis for spin-stabilized artificial satellites are given by Euler equations in spherical coordinates [5] as

$$\begin{aligned} \dot{W} &= \frac{1}{I_z}\, N_z, \\ \dot{\delta} &= \frac{1}{I_z W} N_y, \\ \dot{\alpha} &= \frac{1}{I_z W \mathrm{Cos}\ \delta} N_x, \end{aligned} \tag{5.1}$$

where I_z is the moment of inertia along the spin axis and N_x, N_y, N_z are the components of the external torques in the satellite body frame reference system. By substituting N_{rm}, given in (4.3), in (5.1), the equations of motion are

$$\frac{dW}{dt} = 0, \tag{5.2}$$

$$\frac{d\delta}{dt} = \frac{N_{rym}}{I_z W}, \tag{5.3}$$

$$\frac{d\alpha}{dt} = \frac{N_{rxm}}{I_z W \cos\delta}, \tag{5.4}$$

where it is possible to observe that the residual torque does not affect the satellite angular velocity (because its z-axis component is zero).

By substituting N_{im}, given in (4.3), in (5.1), the equations of motion are

$$\frac{dW}{dt} = \frac{pW}{2\pi I_z} N_{izm}, \tag{5.5}$$

$$\frac{d\delta}{dt} = \frac{pW}{2\pi I_z} N_{iym}, \tag{5.6}$$

$$\frac{d\alpha}{dt} = \frac{p}{2\pi I_z \cos\delta}\, N_{ixm}. \tag{5.7}$$

The differential equations of (5.2)–(5.4) and (5.5)–(5.7) can be integrated assuming that the orbital elements (I, Ω, w) are held constant over one orbital period and that all other terms on right-hand side of equations are equal to initial values.

6. Analysis of the Angular Velocity Magnitude

The variation of the angular velocity magnitude, given by (5.5), can be expressed as:

$$\frac{dW}{W} = k\,dt, \quad \text{with } k = \frac{N_{izm}p}{2\,\pi I_z}. \tag{6.1}$$

If the parameter k is considered constant for one orbital period, then the analytical solution of (6.1) is

$$W = W_0 e^{kt}, \tag{6.2}$$

where W_0 is the initial angular velocity. If the coefficient $k < 0$ in (6.2), then the angular velocity magnitude decays with an exponential profile.

7. Analysis of the Declination and Right Ascension of Spin Axis

For one orbit period, the analytical solutions of (5.3)-(5.4) and (5.6)-(5.7) for declination and right ascension of spin axis, respectively, can simply be expressed as,

$$\delta = k_1 t + \delta_0, \tag{7.1}$$

$$\alpha = k_2 t + \alpha_0, \tag{7.2}$$

with:

(i) for the case where the residual magnetic torque is considered in the motion equations,

$$\begin{aligned} k_1 &= \frac{N_{rym}}{I_z W_o}, \\ k_2 &= \frac{N_{rym}}{I_z W_o \cos\delta_o}, \end{aligned} \tag{7.3}$$

(ii) for the case where the eddy currents torque is considered in the motion equations,

$$\begin{aligned} k_1 &= \frac{pN_{iym}}{2\pi I_z}, \\ k_2 &= \frac{pN_{iym}}{2\pi I_z \cos\delta_o}, \end{aligned} \tag{7.4}$$

where W_0, δ_0, and α_0 are the initial values for spin velocity, declination, and right ascension of spin axis.

Table 1: INPE's Satellite Control Center Data (index SCC) and computed results with the satellite in elliptical orbit and under the influence of the residual magnetic torque (index QER) for declination and right ascension and SCD1 (in degrees).

Day	α_{SCC}	α_{QER}	$\alpha_{SCC} - \alpha_{QER}$	δ_{SCC}	δ_{QER}	$\delta_{SCC} - \delta_{QER}$
22/08/93	282.70	282.7000	0.0000	79.64	79.6400	0.0000
23/08/93	282.67	282.7002	−0.0302	79.35	79.6399	−0.2899
24/08/93	283.50	282.6999	0.8001	79.22	79.6394	−0.4194
25/08/93	283.01	282.7004	0.3096	78.95	79.6395	−0.6895
26/08/93	282.43	282.7015	−0.2715	78.70	79.6399	−0.9399
27/08/93	281.76	282.7019	−0.9419	78.48	79.6398	−1.1598
28/08/93	281.01	282.7019	−1.6919	78.27	79.6393	−1.3693
29/08/93	280.18	282.7024	−2.5224	78.08	79.6392	−1.5592
30/08/93	279.29	282.7036	−3.4136	77.91	79.6396	−1.7296
31/08/93	278.34	282.7043	−4.3643	77.78	79.6397	−1.8597
01/09/93	277.36	282.7044	−5.3444	77.67	79.6391	−1.9691

The solutions presented in (7.1) and (7.2), for the spin velocity magnitude, declination and right ascension of the spin axis, respectively, are valid for one orbital period. Thus, for every orbital period, the orbital data must be updated, taking into account at least the main influences of the Earth's oblateness. With this approach, the analytical theory will be close to the real attitude behavior of the satellite.

8. Applications

The theory developed has been applied to the spin-stabilized Brazilian Satellites (SCD1 and SCD2) for verification and comparison of the theory against data generated by the Satellite Control Center (SCC) of INPE. Operationally, SCC attitude determination comprises [8, 9] sensors data preprocessing, preliminary attitude determination, and fine attitude determination. The preprocessing is applied to each set of data of the attitude sensors that collected every satellite that passes over the ground station. Afterwards, from the whole preprocessed data, the preliminary attitude determination produces estimates to the spin velocity vector from every satellite that passes over a given ground station. The fine attitude determination takes (one week) a set of angular velocity vector and estimates dynamical parameters (angular velocity vector, residual magnetic moment, and Foucault parameter). Those parameters are further used in the attitude propagation to predict the need of attitude corrections. Over the test period, there are not attitude corrections. The numerical comparison is shown considering the quadripole model for the geomagnetic field and the results of the circular and elliptical orbits. It is important to observe that, by analytical theory that included the residual torque, the spin velocity is considered constant during 24 hours. In all numerical simulations, the orbital elements are updated, taking into account the main influences of the Earth's oblateness.

9. Results for SCD1 Satellite

The initial conditions of attitude had been taken on 22 of August of 1993 to the 00:00:00 GMT, supplied by the INPE's Satellite Control Center (SCC). Tables 1, 2, and 3 show the results

Table 2: INPE's Satellite Control Center Data (index SCC) and computed results with the satellite in circular orbit and under the influence of the residual magnetic torque (index QCR) for declination and right ascension and SCD1 (in degrees).

Day	α_{SCC}	α_{QCR}	$\alpha_{SCC} - \alpha_{QCR}$	δ_{SCC}	δ_{QCR}	$\delta_{SCC} - \delta_{QCR}$
22/08/93	282.70	282.7000	0	79.64	79.6400	0
23/08/93	282.67	282.7216	−0.0516	79.35	79.6251	−0.2751
24/08/93	283.50	282.7151	0.7849	79.22	79.6172	−0.3972
25/08/93	283.01	282.6737	0.3363	78.95	79.6182	−0.6682
26/08/93	282.43	282.5877	−0.1577	78.70	79.6303	−0.9303
27/08/93	281.76	282.4526	−0.6926	78.48	79.6552	−1.1752
28/08/93	281.01	282.2631	−1.2531	78.27	79.6940	−1.4240
29/08/93	280.18	282.0158	−1.83588	78.08	79.7473	−1.6673
30/08/93	279.29	281.7091	−2.4191	77.91	79.8140	−1.9040
31/08/93	278.34	281.3439	−3.0039	77.78	79.8962	−2.1162
01/09/93	277.36	280.9240	−3.5640	77.67	79.9892	−2.3191

with the data from SCC and computed values by the present analytical theory, considering the quadripole model for the geomagnetic field and the satellite in circular and elliptical orbit, under influence of the residual and eddy currents torques.

The mean deviation errors for the right ascension and declination are shown in Table 4 for different time simulations. The behavior of the SCD1 attitude over 11 days is shown in Figure 2. It is possible to note that mean error increases with the time simulation. For more than 3 days, the mean error is bigger than the required dispersion range of SCC.

Over the 3 days of test period, better results are obtained for the satellite in circular orbit with the residual torque. In this case, the difference between theory and SCC data has mean deviation error in right ascension of 0.2444° and −0.2241° for the declination. Both are within the dispersion range of the attitude determination system performance of INPE's Control Center.

In Table 5 is shown the computed results to spin velocity when the satellite is under influence of the eddy currents torque, and its behavior over 11 days is shown in Figure 3. The mean error deviation for the spin velocity is shown in Table 6 for different time simulation. For the test period of 3 days, the mean deviation error in spin velocity was of −0.0312 rpm and is within the dispersion range of the attitude determination system performance of INPE's Control Center.

10. Results for SCD2 Satellite

The initial conditions of attitude had been taken on 12 February 2002 at 00:00:00 GMT, supplied by the SCC. In the same way for SCD1, Tables 7, 8, and 9 presented the results with the data from SCC and computed values by circular and elliptical orbits with the satellite under the influence of the residual magnetic torque and eddy currents torque.

The mean deviation errors are shown in Table 10 for different time simulations. For this satellite, there is no significant difference between the circular and elliptical orbits when considering the residual magnetic torque. The behavior of the SCD2 attitude over 12 days is shown in Figure 4.

Table 3: INPE's Satellite Control Center Data (index SCC) and computed results with the satellite in circular orbit and under the influence of the eddy current torque (index QCI) for declination and right ascension and SCD1 (in degrees).

Day	α_{SCC}	α_{QCI}	$\alpha_{SCC} - \alpha_{QCI}$	δ_{SCC}	δ_{QCI}	$\delta_{SCC} - \delta_{QCI}$
22/08/93	282.70	282.7000	0.0000	79.64	79.6400	0.0000
23/08/93	282.67	282.6848	−0.0148	79.35	79.6490	−0.2990
24/08/93	283.50	282.6723	0.8277	79.22	79.6457	−0.4257
25/08/93	283.01	282.6621	0.3479	78.95	79.6352	−0.6852
26/08/93	282.43	282.6538	−0.2238	78.70	79.6220	−0.9220
27/08/93	281.76	282.6468	−0.8868	78.48	79.6092	−1.1292
28/08/93	281.01	282.6412	−1.6312	78.27	79.6001	−1.3301
29/08/93	280.18	282.6366	−2.4566	78.08	79.5942	−1.5142
30/08/93	279.29	282.6331	−3.3431	77.91	79.5913	−1.6813
31/08/93	278.34	282.6305	−4.2905	77.78	79.5904	−1.8104
01/09/93	277.36	282.6287	−5.2687	77.67	79.5909	−1.9209

Table 4: Mean deviations for different time simulations for declination and right ascension and SCD1 (in degrees).

Time Simulation (days)	11	8	3	2
$\alpha_{SCC} - \alpha_{QER}$	−1.5882	−0.5435	0.2566	−0.0151
$\alpha_{SCC} - \alpha_{QCR}$	−1.0779	−0.3587	0.2444	−0.0258
$\alpha_{SCC} - \alpha_{QCI}$	−1.5400	−0.5047	0.2710	0.0074
$\delta\alpha_{SCC} - \delta\alpha_{QER}$	−1.0896	−0.8034	−0.2364	−0.1449
$\delta_{SCC} - \delta_{QCR}$	−1.1707	−0.8172	−0.2241	−0.1376
$\delta_{SCC} - \delta_{QCI}$	−1.0653	−0.7882	−0.2416	−0.1495

Table 5: INPE's Satellite Control Center Data (index SCC) and computed results for spin velocity, with the satellite in circular orbit and under the influence of the eddy currents torque (index QCI) (in rpm).

Day	W_{SCC}	W_{QCI}	$W_{SCC} - W_{QCI}$
22/08/93	86.2100	86.2100	0.0000
23/08/93	86.0400	86.3156	−0.2756
24/08/93	85.8800	86.4985	−0.6185
25/08/93	85.8000	86.7144	−0.9144
26/08/93	85.7300	86.9439	−1.2139
27/08/93	85.6600	87.1719	−1.5119
28/08/93	85.5800	87.3631	−1.7831
29/08/93	85.5100	87.5296	−2.0196
30/08/93	85.4400	87.6657	−2.2257
31/08/93	85.3700	87.7658	−2.3958
01/09/ 93	85.3100	87.8426	−2.5326

Table 6: Mean deviations for different time simulations for spin velocity and SCD1 (in degrees).

Time Simulation (days)	11	8	3	2
$W_{SCC} - W_{QCI}$ (rpm)	−0.1475	−0.1091	−0.0312	−0.0144

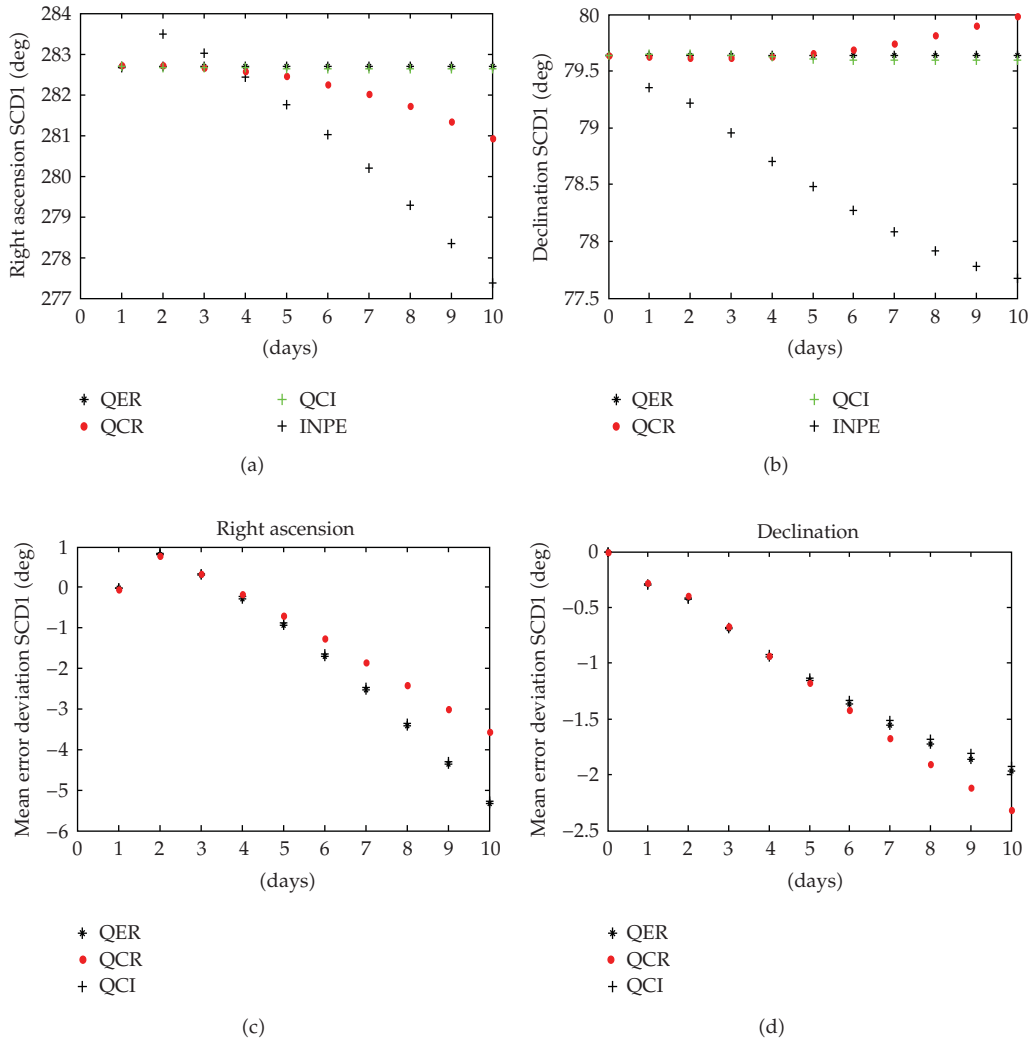


Figure 2: Evolution of the declination (δ) and right ascension (α) of satellite spin axis for SCD1 and its mean deviation error.

Over the test period of the 12 days with the satellite in elliptical orbit and considering the residual magnetic torque, the difference between theory and SCC data has mean deviation error in right ascension of −0.1266 and −0.1358 in the declination. Both torques are within the dispersion range of the attitude determination system performance of INPE's Control Center, and the solution can be used for more than 12 days.

In Table 11 the computed results to spin velocity are shown when the satellite is under the influence of the eddy currents torque. The mean deviation error for the spin velocity is shown in Table 12 for different time simulation. For the test period, the mean deviation error in spin velocity was of 0.0253 rpm and it is within the dispersion range of the attitude determination system performance of INPE's Control Center. The behavior of the spin velocity is shown in Figure 5.

Table 7: INPE's Satellite Control Center Data (index SCC) and computed results with the satellite in elliptical orbit and under the influence of the residual magnetic torque (index QER) for declination and right ascension and SCD2 (in degrees).

Day	α_{SCC}	α_{QER}	$\alpha_{SCC} - \alpha_{QER}$	δ_{SCC}	δ_{QER}	$\delta_{SCC} - \delta_{QER}$
12/02/02	278.71	278.710000	0.0000	63.47	63.470000	0.0000
13/02/02	278.73	278.709999	0.0200	63.45	63.469998	−0.0200
14/02/02	278.74	278.710000	0.0300	63.42	63.470002	−0.0500
15/02/02	278.74	278.710000	0.0300	63.39	63.470005	−0.0800
16/02/02	278.72	278.709999	0.0100	63.36	63.470002	−0.1100
17/02/02	278.68	278.709999	−0.0300	63.33	63.470000	−0.1400
18/02/02	278.63	278.710000	−0.0800	63.31	63.470003	−0.1600
19/02/02	278.57	278.710001	−0.1400	63.29	63.470006	−0.1800
20/02/02	278.50	278.710000	−0.2100	63.27	63.470004	−0.2000
21/02/02	278.42	278.709999	−0.2900	63.25	63.470000	−0.2200
22/02/02	278.33	278.710000	−0.3800	63.24	63.470002	−0.2300
23/02/02	278.23	278.710002	−0.4800	63.23	63.470006	−0.2400

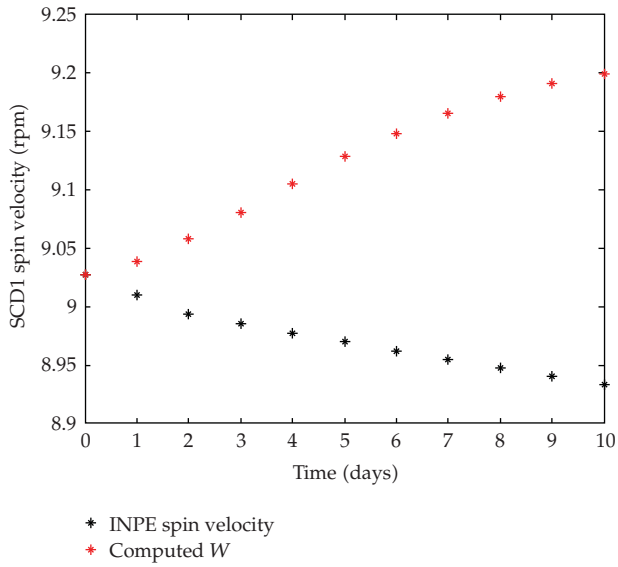


Figure 3: Evolution of the spin velocity (W) for SCD1.

11. Mean Pointing Deviation

For the tests, it is important to observe the deviation between the actual SCC supplied and the analytically computed attitude, for each satellite. It can be computed by

$$\theta = \cos^{-1}\left(\hat{i}\,\hat{i}_c + \hat{j}\,\hat{j}_c + \hat{k}\,\hat{k}_c\right), \tag{11.1}$$

where $(\hat{i}, \hat{j}, \hat{k})$ indicates the unity vectors computed by SCC and $(\hat{i}_c, \hat{j}_c, \hat{k}_c)$ indicates the unity vector computed by the presented theory.

Table 8: INPE's Satellite Control Center Data (index SCC) and computed results with the satellite in circular orbit and under the influence of the residual magnetic torque (index QCR) for declination and right ascension and SCD2 (in degrees).

Day	α_{SCC}	α_{QCR}	$\alpha_{SCC} - \alpha_{QCR}$	δ_{SCC}	δ_{QCR}	$\delta_{SCC} - \delta_{QCR}$
12/02/02	278.71	278.710000	0	63.47	63.470000	0
13/02/02	278.73	278.7113	0.01870	63.45	63.4692	−0.0192
14/02/02	278.74	278.7127	0.02733	63.42	63.4683	−0.0482
15/02/02	278.74	278.7141	0.0259	63.39	63.4673	−0.0773
16/02/02	278.72	278.7155	0.0045	63.36	63.4664	−0.1064
17/02/02	278.68	278.7168	−0.0368	63.33	63.4654	−0.1354
18/02/02	278.63	278.7180	−0.0880	63.31	63.4646	−0.1546
19/02/02	278.57	278.7191	−0.1491	63.29	63.4638	−0.1738
20/02/02	278.50	278.7200	−0.2200	63.27	63.4631	−0.1931
21/02/02	278.42	278.7207	−0.3007	63.25	63.4625	−0.2125
22/02/02	278.33	278.7212	−0.3913	63.24	63.4621	−0.2221
23/02/02	278.23	278.7215	−0.4916	63.23	63.4618	−0.2318

Table 9: INPE's Satellite Control Center Data (index SCC) and computed results with the satellite in circular orbit and under the influence of the eddy currents torque (index QCI) for declination and right ascension and SCD2 (in degrees).

Day	α_{SCC}	α_{QCI}	$\alpha_{SCC} - \alpha_{QCI}$	δ_{SCC}	δ_{QCI}	$\delta_{SCC} - \delta_{QCI}$
12/02/02	278.71	278.7100	0.0000	63.47	63.4700	0.0000
13/02/02	278.73	278.7170	0.0130	63.45	63.4921	−0.0421
14/02/02	278.74	278.7261	0.0139	63.42	63.5119	−0.0919
15/02/02	278.74	278.7371	0.0029	63.39	63.5268	−0.1368
16/02/02	278.72	278.7497	−0.0296	63.36	63.5352	−0.1752
17/02/02	278.68	278.7635	−0.0835	63.33	63.5370	−0.2070
18/02/02	278.63	278.7772	−0.1472	63.31	63.5345	−0.2245
19/02/02	278.57	278.7912	−0.2212	63.29	63.5302	−0.2402
20/02/02	278.50	278.8044	−0.3043	63.27	63.5285	−0.2585
21/02/02	278.42	278.8159	−0.3959	63.25	63.5334	−0.2834
22/02/02	278.33	278.8253	−0.4953	63.24	63.5477	−0.3077
23/02/02	278.23	278.8321	−0.6021	63.23	63.5724	−0.3423

Table 10: Mean deviations for different time simulation for declination and right ascension and SCD2 (in degrees).

Time Simulation (days)	12	8	5	2
$\alpha_{SCC} - \alpha_{QER}$	−0.1266	−0.0200	0.0180	0.0100
$\alpha_{SCC} - \alpha_{QCR}$	−0.1334	−0.0247	−0.0153	−0.0093
$\alpha_{SCC} - \alpha_{QCI}$	−0.1875	−0.0565	−0.0139	0.0065
$\delta\alpha_{SCC} - \delta\alpha_{QER}$	−0.1358	−0.0925	−0.0520	−0.0099
$\delta_{SCC} - \delta_{QCR}$	−0.1312	−0.0894	−0.0502	−0.0096
$\delta_{SCC} - \delta_{QCI}$	−0.1925	−0.1397	−0.1088	−0.0210

Table 11: INPE's Satellite Control Center Data (index SCC) and computed results of spin velocity, with the satellite in circular orbit and under the influence of the eddy currents torque (index QCI) (in rpm).

Day	W_{SCC}	W_{QCI}	$W_{SCC} - W_{QCI}$
12/02/02	34.4800	34.4800	0.0000
13/02/02	34.4200	34.4942	−0.0742
14/02/02	34.3700	34.4572	−0.0872
15/02/02	34.3100	34.3561	−0.04617
16/02/02	34.2600	34.1831	0.0769
17/02/02	34.2000	33.9323	0.2678
18/02/02	34.1400	34.6059	−0.4659
19/02/02	34.0800	34.2108	−0.1308
20/02/02	34.0200	33.7703	0.2497
21/02/02	33.9600	33.3067	0.6533
22/02/02	33.9000	32.8493	1.0508
23/02/02	33.8300	32.4199	1.4101

Table 12: Mean deviations for different time simulations for spin velocity and SCD1 (in degrees).

Time simulation (days)	12	8	5	2
$W_{SCC} - W_{QCI}$	0.0253	−0.0060	−0.0027	−0.0039

Figures 6 and 7 present the pointing deviations for the test period. The mean pointing deviation for the SCD1 for different time simulations are presented in Table 13. Over the test period of 11 days, the mean pointing deviation with the residual magnetic torque and elliptical orbit was 1.1553°, circular orbit was 1.2003°, and eddy currents torque with circular orbit was 1.1306°. The test period of SCD1 shows that the pointing deviation is higher than the precision required for SCC. Therefore for SCD1, this analytical approach should be evaluated by a time less than 11 days.

For SCD2, the mean pointing deviation considering the residual magnetic torque and elliptical orbit was 0.1538, residual magnetic torque and circular orbit was 0.1507, and eddy current torque was 0.2160. All the results for SCD2 are within the dispersion range of the attitude determination system performance of INPE's Control Center of 0.5°.

12. Summary

In this paper an analytical approach was presented to the spin-stabilized satellite attitude propagation taking into account the residual and eddy currents torque. The mean components of these torques in the satellite body reference system have been obtained and the theory shows that, unlike the eddy currents torque, there is no residual torque component along the spin axis (z-axis). Therefore this torque does not affect the spin velocity magnitude, but it can cause a drift in the satellite spin axis.

The theory was applied to the spin-stabilized Brazilian satellites SCD1 and SCD2 in order to validate the analytical approach, using quadripole model for geomagnetic field and the satellite in circular and elliptical orbits.

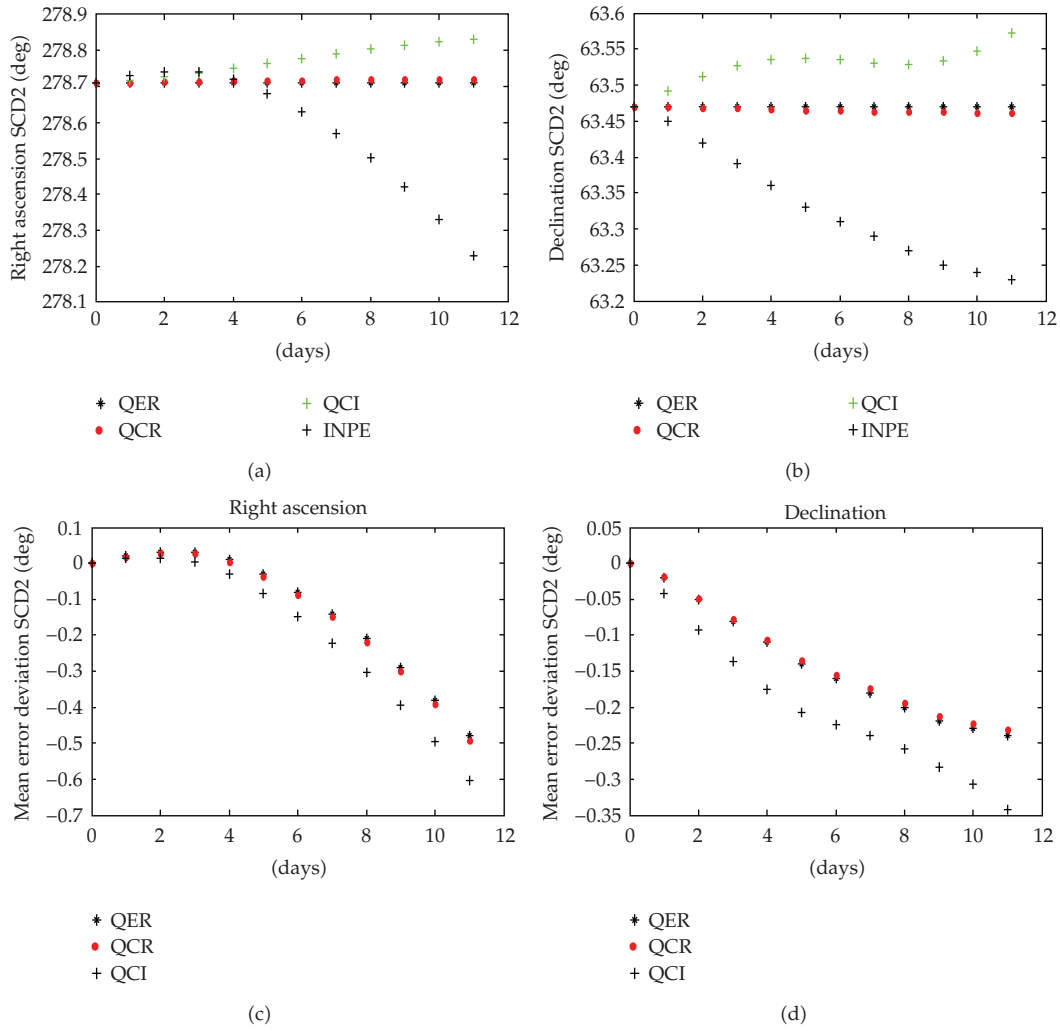


Figure 4: Evolution of the declination (δ) and right ascension (α) of satellite spin axis for SCD2 and its mean deviation error.

Table 13: Mean pointing deviation for SCD1.

Time simulation (days)	11	8	3	2
θ_{QER}	1.1553	0.8226	0.2448	0.1450
θ_{QCR}	1.2003	0.8288	0.2326	0.1376
θ_{QCI}	1.1306	0.8071	0.2503	0.1495

The result of the 3 days of simulations of SCD1, considering the residual magnetic torque, shows a good agreement between the analytical solution and the actual satellite behavior. For more than 3 days, the pointing deviation is higher than the precision required for SCC (0.5°).

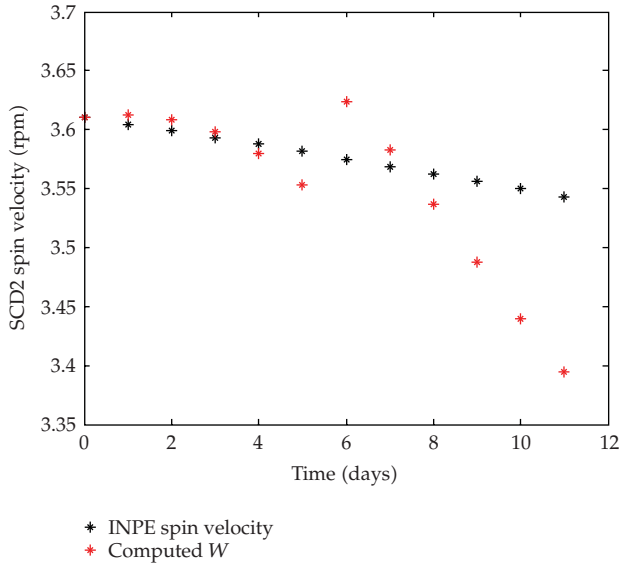


Figure 5: Evolution of the spin velocity magnitude (W) for SCD2.

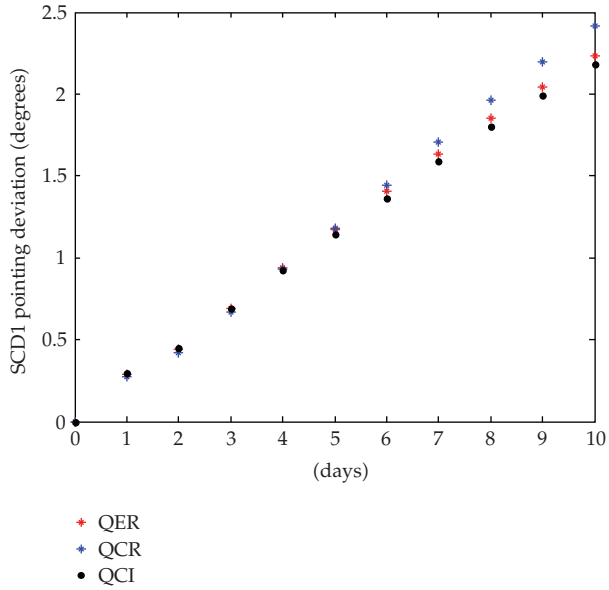


Figure 6: Pointing deviation evolution (in degrees) for SCD1.

For the satellite SCD2, over the test period of the 12 days, the difference between theory (when considering the residual or eddy currents torque) and SCC data is within the dispersion range of the attitude determination system performance of INPE's Control Center.

Thus the procedure is useful for modeling the dynamics of spin-stabilized satellite attitude perturbed by residual or eddy currents torques but the time simulation depends on the precision required for satellite mission.

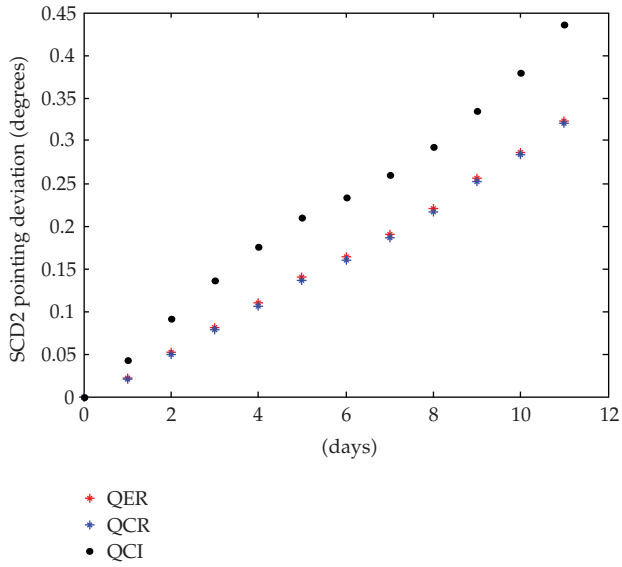


Figure 7: Pointing deviation evolution (in degrees) for SCD2.

Appendix

The coefficients of the mean components of the residual magnetic torques, given by (2.9), are expressed by

$$A = \sum_{i=1}^{7} a_{ia} + \sum_{i=1}^{7} a_{ib}, \qquad B = \sum_{i=1}^{7} b_{ia} + \sum_{i=1}^{7} b_{ib}, \qquad C = \sum_{i=1}^{7} c_{ia} + \sum_{i=1}^{7} c_{ib},$$
$$D = \sum_{i=1}^{7} a_{ia} + \sum_{i=1}^{7} a_{ib}, \qquad E = \sum_{i=1}^{7} b_{ia} + \sum_{i=1}^{7} b_{ib}, \tag{A.1}$$

where a_{ib}, b_{ib}, c_{jb}, $i = 1, 2, \ldots, 7$; $j = 1, \ldots, 4$, can be got by Garcia in [3]. It is important to note that the parcel b_{ib} is associated with the quadripole model and the satellite in an elliptical orbit. For circular, orbit, b_{ib} is zero.

The mean components N_{ixm}, N_{iym}, N_{izm} of the eddy currents torque are expressed by

$$N_{ixm} = \sum_{i=1}^{14\,712} \operatorname{tr} x(i) + \sum_{i=1}^{18\,426} Nx(i),$$
$$N_{iym} = \sum_{i=1}^{14\,712} \operatorname{tr} y(i) + \sum_{i=1}^{53\,765} Ny(i), \tag{A.3}$$
$$N_{izm} = \sum_{i=1}^{7\,350} \operatorname{tr} z(i) + \sum_{i=1}^{21\,435} Nz(i),$$

where $\operatorname{tr} x(i)$, $\operatorname{tr} y(i)$, $\operatorname{tr} z(i)$, $Nx(i)$, $Ny(i)$, and $Nz(i)$ are presented by Pereira [6].

The terms a_{ib}, b_{ib}, c_{jb}, $\operatorname{tr} x(i)$, $\operatorname{tr} y(i)$, $\operatorname{tr} z(i)$, $Nx(i)$, $Ny(i)$, and $Nz(i)$ depend on orbital elements (a, e, I, Ω, w) and attitude angles (δ, α).

Acknowledgment

This present work was supported by CNPq (National Counsel of Technological and Scientific Development).

References

[1] L. C. Thomas and J. O. Cappelari, "Attitude determination and prediction of spin-stabilized satellites," *The Bell System Technical Journal*, pp. 1656–1726, 1964.
[2] J. R. Wertz, *Spacecraft Attitude Determination and Control*, D. Reidel, Dordrecht, The Netherlands, 1978.
[3] R. V. Garcia, *Satélites estabilizados por rotação e torque magnético residual*, Dissertação de Mestrado, Faculdade de Engenharia, UNESP, Guaratinguetá, Brazil, 2007.
[4] M. C. Zanardi, "Dinâmica da atitude de satélites artificiais," *Tese de Livre Docência*, Universidade Estadual Paulista, Guaratinguetá, Brazil, 2005.
[5] H. K. Kuga, W. C. C. Silva, and U. T. V. Guedes, "Dynamics of attitude for spin stabilized satellites," Tech. Rep. INPE-4403, INPE, São José dos Campos, Brazil, 1987.
[6] A. J. Pereira, "Propagação da atitude de satélites estabilizados por totação com torque induzido," Trabalho de Conclusão de Curso, Faculdade de Engenharia, UNESP, Guaratinguetá, Brazil, 2006.
[7] D. Brouwer and G. M. Clemence, *Methods of Celestial Mechanics*, Academic Press, New York, NY, USA, 1961.
[8] V. Orlando, H. K. Kuga, and U. T. V. Guedes, "Flight dynamics leop and routine operations for SCD2, the INPE's second environmental data collecting satellite," *Advances in the Astronautical Sciences*, vol. 100, no. 2, pp. 1003–1013, 1998.
[9] H. K. Kuga, V. Orlando, and R. V. F. Lopes, "Flight dynamics operations during LEOP for the INPE's second environmental data collection satellite SCD2," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 21, pp. 339–344, 1999.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 173145, 9 pages
doi:10.1155/2009/173145

Research Article

Using of H-Infinity Control Method in Attitude Control System of Rigid-Flexible Satellite

Ximena Celia Méndez Cubillos and Luiz Carlos Gadelha de Souza

Space Mechanics and Control Division (DMC), National Institute for Space Research (INPE), Avenida dos Astronautas 1758, P.O. Box 515, São Paulo, 12201-940 São José dos Campos, Brazil

Correspondence should be addressed to Luiz Carlos Gadelha de Souza, gadelha@dem.inpe.br

Received 1 July 2009; Revised 7 October 2009; Accepted 18 November 2009

Recommended by Tadashi Yokoyama

The attitude control systems of satellites with rigid and flexible components are demanding more and more better performance resulting in the development of several methods control. For that reason, control design methods presently available, including parameters and states estimation, robust and adaptive control, as well as linear and nonlinear theory, need more investigation to know their capability and limitations. In this paper the investigated technique is H-Infinity method in the performance of the Attitude Control System of a Rigid-Flexible Satellite.

1. Introduction

The rapidly complexity increase of systems and processes to be controlled has stimulated the development of sophisticated analysis and design methods called advanced techniques. The H-Infinity (H∞) control theory, introduced by Zames [1], is one of the advanced techniques and its application in several problems of control has been growing rapidly.

The employment of flexible structures in the spatial area is another problem of control system which has been growing up too. Flexible systems offer several advantages compared with the rigid system. Some advantages are relatively smaller actuators, lower overall mass, faster response, lower energy consumption, in general, and lower cost. With the study of the Attitude Control System (ACS) of space structures with flexible antennas and/or panel and robotic manipulators, one becomes more complex when the dimensions of such structures increase due to necessity to consider a bigger number of vibration modes in its model in order to improve the model fidelity [2]. Examples of projects that involve flexible space structures are the International Space Station (ISS), the Lunar Reconnaissance Orbiter (LRO), the Lunar Crater Observation and Sensing Satellite (LCROSS), the Hubble Space Telescope, and so forth.

In Rigid-Flexible Satellite (RFS) the function of the ACS is to stabilize and orient the satellite during its mission, counteracting external disturbances torques and forces. In this paper is investigated multivariable control method H∞ for attitude control of an RFS consisting of a rigid body and two flexible panels. The satellite modeling was built following the Lagrangian approach and the discretization was done using the assumed-modes method. The equations of motion obtained were written in its modal state space form.

2. The Rigid-Flexible Satellite Model

Figure 1 shows the picture of the satellite used in this work, which is composed of a rigid body of cubic shape and two flexible panels. The center of mass of the satellite is in the point 0 origin of the system of coordinates (X, Y, Z) that coincides with its main axis of inertia. The elastic appendixes with the beam format are connected in the central body, being treated as a punctual mass in its free extremity. The length of the panel is represented by L, the mass is represented by m, and $v(x,t)$ is the elastic displacement in relation to the axis Z. The moment of inertia of the rigid body of the satellite in relation to the mass center is J_0. The moment of inertia of the panel in relation to its own mass center is given by J_p.

3. Equations of Motion

In the Lagrang approach are considered the equation of motion of the satellite around in Y and the elastic displacement of the panels. The Lagrange equations [3] for the problem can be written in the following form:

$$\frac{d}{dt}\left(\frac{\partial L^*}{\partial \dot{\theta}}\right) - \frac{\partial L^*}{\partial \theta} = \tau, \tag{3.1}$$

$$\frac{d}{dt}\left(\frac{\partial L^*}{\partial \dot{q}_i}\right) - \frac{\partial L^*}{\partial q_i} + \frac{\partial M}{\partial q_i} = 0. \tag{3.2}$$

In (3.1) τ is the torque of the reaction wheel, $L^* = T - V$ is the Lagrangian, and θ is the rotation angle of the satellite around the axis Y. In (3.2) M is the dissipation energy associated to the deformation of the panel q_i it represents each one of the generalized coordinates of the problem.

The beam deflection variable $v(x,t)$ is discretized using the expansion

$$v(x,t) = \sum_{i=1}^{n} \phi_i(x) q_i(t) \qquad 0 \le x \le L, \tag{3.3}$$

where n represents the number of manners to be adopted in the discretization and $\phi_i(x)$ represents each one of the own modes of the system. The admissible functions $\phi_i(x)$ are given by [4]

$$\phi_i(x) = \cosh(a_i x) - \cos(a_i x) - \alpha_i(\sinh(a_i x) - \sin(a_i x)), \tag{3.4}$$

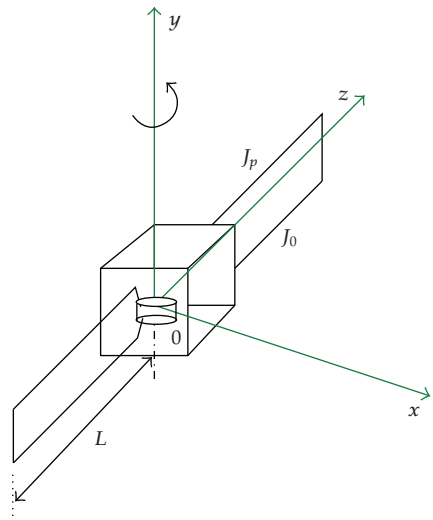


Figure 1: Satellite Model.

where

$$\alpha_i = \frac{\cosh(a_i L) + \cos(a_i L)}{\sinh(a_i L) + \sin(a_i L)}, \tag{3.5}$$

and $a_i L$ are the eigenvalues of the free system and undamped.

For the complete system, the total kinetic energy T is given by $T = T_{\text{Satellite}} + T_{\text{Panel}}$; therefore,

$$T = \frac{1}{2} J_0 \dot{\theta}^2 + \left[\rho A' \int_0^L \left[\dot{v}(x,t)^2 + 2(\dot{v}(x,t)) x\dot{\theta} + \left(x\dot{\theta}\right)^2 + \left(\dot{\theta} \cdot \dot{v}(x,t)\right)^2 \right] dx \right], \tag{3.6}$$

where ρ is the density of the panels and A' is the area. The dissipation energy function is

$$M = \dot{v}(x,t)^2 K_d, \tag{3.7}$$

where K_d is the dissipation constant. So $L^* = T - V$ is given by

$$L^* = \frac{1}{2} J_0 \dot{\theta}^2 + \left[\rho A' \int_0^L \left[\dot{v}(x,t)^2 + 2(\dot{v}(x,t)) x\dot{\theta} + \left(x\dot{\theta}\right)^2 + \left(\dot{\theta} \cdot \dot{v}(x,t)\right)^2 \right] dx \right] - v(x,t)^2 \cdot K. \tag{3.8}$$

In (3.8) K is constant elastic of the panels. After some manipulations of (3.8) [5] and using the orthogonalization property of vibration modes of the beam [6], one has

$$\int_0^L \phi_i \phi_j dx = 1 \quad \text{if } i = j, \qquad \int_0^L \phi_i \phi_j dx = 0 \quad \text{if } i \neq j. \tag{3.9}$$

Finally, two equations are obtained. These equations represent the dynamics of rotation motion of the satellite and the elastic displacement of the panels, respectively,

$$\ddot{\theta}\left(1 + a\sum_{i=1}^{n} q_i^2\right) + \alpha_i \cdot a\sum_{i=1}^{n} \ddot{q}_i = \frac{1}{J_1}\tau. \tag{3.10}$$

$$\ddot{q}_i + \alpha_i\ddot{\theta} - \dot{\theta}^2 q_i + d \cdot \dot{q}_i + c \cdot q_i = \tau_q. \tag{3.11}$$

where the term nonlinear α_i in (3.10) is defined as centripetal rigidity, and in (3.11) τ_q is a Piezoelectric actuator adapted for the following simulations where will be considered $i = 1$ (one mode), and the constants are given by

$$a = \frac{2\rho A'}{J_1}, \qquad J_1 = J_0 + 2J_p, \qquad c = \frac{K}{\rho A'}, \qquad d = \frac{K_d}{\rho A'}. \tag{3.12}$$

4. H-Infinity Control Method

4.1. Introduction

Throughout the decades of 1980 and 1990, H-Infinity control method had a significant impact in the development of control systems; nowadays the technique has become fully grown and it is applied on industrial problems [5]. In the control theory in order to achieve robust performance or stabilization, the H-Infinity control method is used. The control designer expresses the control problem as a mathematical optimization problem finding the controller solution. H∞ techniques have the advantage over classical control techniques in which the techniques are readily applicable to problems involving multivariable systems with cross-coupling between channels; disadvantages of H∞ techniques include the high level of mathematical understanding needed to apply them successfully and the need for a reasonably good model of the system to be controlled. The problem formulation is important, since any synthesized controller will be "optimal" in the formulated sense.

The H∞ name derives from the fact that mathematically the problem may be set in the space H∞, which consists of all bounded functions that are analytic in the right-half complex plane. We do not go to this length. The H∞ norm is the maximum singular value of the function; let us say that it can be interpreted as a maximum gain in any direction and at any frequency; for SISO (Single In, Single Out) systems, this is effectively the maximum magnitude of the frequency response. H∞ method is also used to minimize the closed-loop impact of a perturbation: depending on the problem formulation, the impact will be measured in terms of either stabilization or performance. Thus, one concludes that the procedures to project control systems are a difficult task due to the cited terms which are conflicting properties [7].

4.2. Modeling

This problem is defined by the configuration of Figure 2. The "plant" is a given system with two inputs and two outputs. It is often referred to as the generalized plant. The signal w is an external input and represents driving signals that generate disturbances, measurement noise, and reference inputs. The signal u is the control input. The output z has the meaning

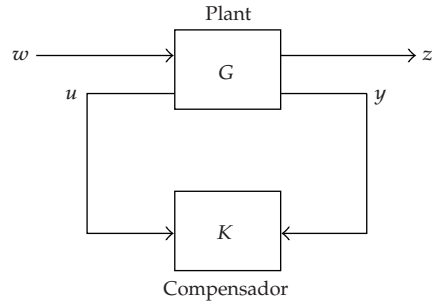


Figure 2: Generalized Plant.

of control error and ideally should be zero. The output y, finally, is the observed output and is available for feedback.

The project of control system is based given by

$$\begin{aligned} \dot{x}(t) &= Ax + Bu, \\ y(t) &= Cx + Du. \end{aligned} \tag{4.1}$$

A more general state space representation of the standard plant is

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1 w(t) + B_2 W(t), \\ z(t) &= C_1 x(t) + D_{11} w(t) + D_{12} u(t), \\ y(t) &= C_2 x(t) + D_{21} w(t) + D_{22} u(t), \\ \begin{bmatrix} \dot{x}(t) \\ z(t) \\ y(t) \end{bmatrix} &= \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix} . \begin{bmatrix} x(t) \\ w(t) \\ u(t) \end{bmatrix} = P. \end{aligned} \tag{4.2}$$

The solution of the corresponding H∞ problem based on Riccati equations is implemented requires the following conditions to be satisfied [8]:

(1) (A, B_2) is stabilizable and (C_2, A) is detectable,

(2) D_{12} and D_{21} have full rank,

(3) $[A\text{-}jwI\ B_2; C_1 D_{12}]$ has full column rank for all $w\ \epsilon\ \Re$ (hence, D_{12} is tall),

(4) $[A\text{-}jwI\ B_1; C_2 D_{21}]$ has full column rank for all $w\ \epsilon\ \Re$ (hence, D_{21} is wide).

The augmented plant is formed by accounting for the weighting functions W_1, W_2, and W_3, as shown in Figure 3. In order to reach the acting objectives, the outputs were chosen to be transfer weight functions: $z_1 = W_1;\ z_2 = W_2 y;\ z_3 = W_3 u$.

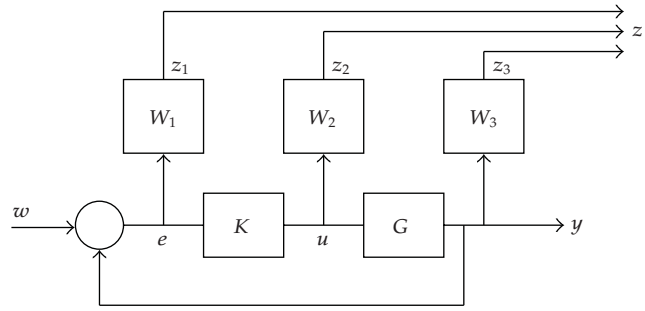


Figure 3: Plant with weighting functions for H∞ design.

Table 1: Parameters.

Parameter	Description	Value
J_0	Moments of inertia of the rigid body of the satellite	720 Kg·m^2
J_p	Moment of inertia of the panel	40 Kg·m^2
K	Constant elastic of the panels	320 Kg·rad^2/s^2
K_d	Dissipation constant	0,48 Kg·rad^2/s
L	Length of the panel	2 m
m	Mass of the satellite	200 kg

The function cost of mixed sensibility is given for

$$T_{y1u1} = \begin{bmatrix} W_1 S \\ W_2 R \\ W_3 T \end{bmatrix}, \quad \begin{aligned} S &= (I + GK)^{-1}, \\ R &= K(I + GK)^{-1}, \\ T &= GK(I + GK)^{-1}, \end{aligned} \tag{4.3}$$

where S is called sensibility, T is complementary sensitivity function, and R does not have any name. The cost function of mixed sensibility is named alike, because it punishes S, R, and T at the same time; it can also be said project requirement. The transfer function from w to z_1 is the weighted sensitivity function W_1S, which characterizes the performance objective of good tracking; the transfer function from w to z_2 is the complementary sensitivity function T, whose minimization ensures low control gains at high frequencies, and the transfer function from w to z_3 is KS, which measures the control effort. It is also used to impose the constraints on the control input for example, the saturation limits.

5. Simulations

The simulations were carried out by computational implementation of the software MatLab. The initials conditions used here are $\theta = 0.001$ rad. and $\dot{\theta} = 0$ rad./s. The values considered for the physical parameters in the numerical simulation are presented in Table 1.

The procedure of the project of H∞ is different from other control projects knowledge such as LQR (Linear Quadratic Regulator) and LQG (Linear Quadratic Gaussian) [5]; the

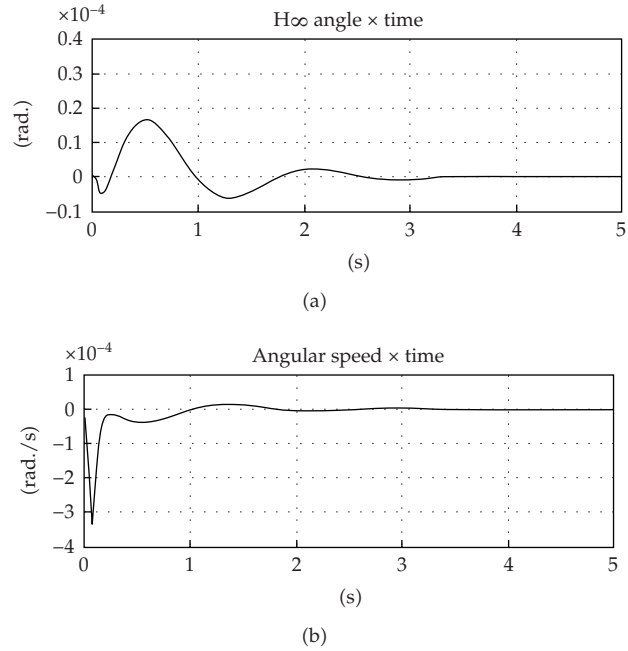


Figure 4: Angle and Angular Speed.

difference is the use of weighting functions W_1, W_2, and W_3, where $W_2 = 0$ and the others are given by

$$W_1^{-1} = \gamma^{-1} * \frac{0.1(1 + s/100)^2}{(1 + s/5000)^2} * I_{2x2}\ ,$$
$$W_3^{-1} = \frac{2000}{s} * I_{2x2}, \tag{5.1}$$

where W_1 punishes the error sign e, W_2 punish the control sign "u", and W_3 punishes the exit of the plant y; γ is a parameter obtained through successive attempts.

6. Results

First we analyze the open loop of the system through transmission zeros (TZs) and the close-loop with H∞ Control. The TZs are critical frequencies where signal transmission between input and output is stopped. The importance of use of the TZs is given by their application in robust control, because they are the zeros of a MIMO system. In Table 2 are represented the values.

Following, the performances of H∞ control in the ACS are observed in Figures 4 and 5.

Both graphs, in Figure 4, have existence of overshoot, in which they could commit the system; however, the time of stabilization of both was of approximately 3.5 seconds. In other words, in spite of the existence of overshoots, the control of the system, in a long time, was reached.

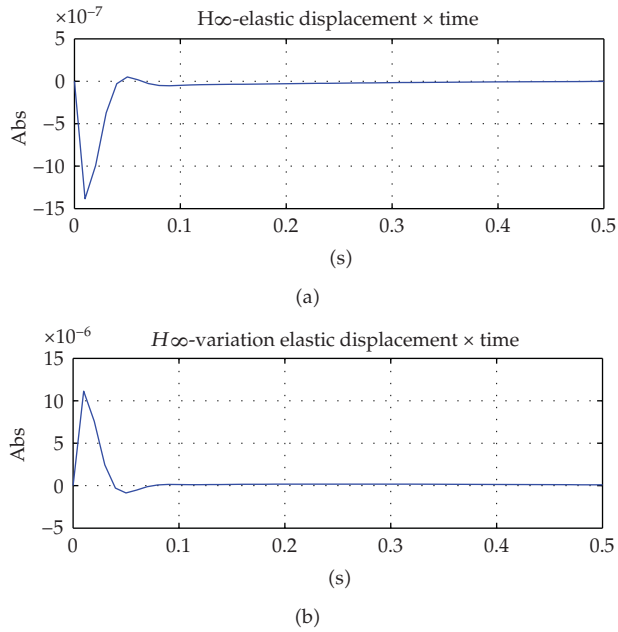


Figure 5: Vibration of the Panels.

Table 2: Transmission Zeros.

	Transmission Zeros
Open Loop	$-0.0244 \pm 5.7049i$
H∞ Control	−2318.10
	−433.3
	$-0.0 \pm 05.700i$
	$-0 \pm 0i$

In Figure 5 the behavior of the vibration of the panels is presented. The displacement of overshoot is of the order of 10^{-7}, in other words, very small. The time of stabilization in the first graph is about 0.5 seconds and for the second one is about 0.45 seconds. This demonstrates that the control H∞ possesses a good performance for angle and angular velocity, as well as to control the vibration of the panels.

7. Conclusions

The problem of attitude control of satellites is not new and has been addressed by several researchers using many different approaches. The H∞ control method is one of the most advanced techniques available today for designing robust controllers. One great advantage with this technique is that it allows the designer to tackle the most general form of control architecture wherein explicit accounting of uncertainties, disturbances, actuator/sensor noises, actuator constraints, and performance measures can be accomplished. The system is very different from the methods LQR and LQG, for example. However, a great disadvantage is the experience and necessary abilities to design the form of the weighting functions and

the fact that the plant can increase. Basically, the success of the method depends on the correct choice of the weight functions transfer.

Acknowledgments

The authors would like to thank CAPES and INPE/DMC. This work was supported by CAPES thought the Brasil—Portugal Cooperation Project PCT no. 241/09.

References

[1] G. Zames, "Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 301–320, 1981.

[2] L. C. G. De Souza and S. A. Silva, "Vibration control of a rigid-flexible satellite during attitude maneuver," in *Proceeding of the 17th ASME Biennial Conference on Mechanical Vibration and Noise (DETC '99)*, Las Vegas, Nev, USA, September 1999.

[3] L. Meirovitch, *Methods of Analytical Dynamics*, Dover, Mineola, NY, USA, 1998.

[4] J. L. Junkins and Y. Kim, *Introduction to Dynamics and Control of Flexible Structures*, AIAA, Washington, DC, USA, 1993.

[5] X. C. M. Cubillos, *Investigation of Multivariable Control Techniques for Attitude Control of a Rigid—Flexible Satellite*, INPE, São José dos Campos, Brazil, 2008.

[6] C. H. G. Hassmann, and A. Fenili, "Attitude and vibration control of a satellite with a flexible solar panel using Lqr tracking with infinite time," in *Proceedings of the ASME International Congress of Mechanical Engineering*, Seattle, Wash, USA, November 2007.

[7] M. G. Safonov, A. J. Laub, and G. L. Hartmann, "Feedback properties of multivariable systems: the role and use of the return difference matrix," *IEEE Transactions on Automatic Control*, vol. 26, no. 1, pp. 47–65, 1981.

[8] O. H. Bosgra and H. Kwakernaak, "Design methods for control systems," Notes for a Course of the Dutch Institute of Systems and Control, Winter term 2001-2002.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 753653, 18 pages
doi:10.1155/2009/753653

Research Article

Hill Problem Analytical Theory to the Order Four: Application to the Computation of Frozen Orbits around Planetary Satellites

Martin Lara[1] and Jesús F. Palacián[2]

[1] *Celestial Mechanics Division, Real Observatorio de la Armada, 11110 San Fernando, Spain*
[2] *Departamento de Ingeniería Matemática e Informática, Universidad Pública de Navarra, 31006 Pamplona, Spain*

Correspondence should be addressed to Martin Lara, mlara@roa.es

Received 7 June 2009; Accepted 24 August 2009

Recommended by Silvia Maria Giuliatti Winter

Frozen orbits of the Hill problem are determined in the double-averaged problem, where short and long-period terms are removed by means of Lie transforms. Due to the perturbation method we use, the initial conditions of corresponding quasi-periodic solutions in the nonaveraged problem are computed straightforwardly. Moreover, the method provides the explicit equations of the transformation that connects the averaged and nonaveraged models. A fourth-order analytical theory is necessary for the accurate computation of quasi-periodic frozen orbits.

1. Introduction

Besides its original application to the motion of the Moon [1], the Hill problem provides a good approximation to the real dynamics of a variety of systems, encompassing the motion of comets, natural and artificial satellites, distant moons of asteroids, or dynamical astronomy applications [2–4]. Specifically, the Hill model and its variations [5–9] are useful for describing the motion about planetary satellites. In addition, the Hill problem is an invariant model that does not depend on any parameter, thus, giving broad generality to the results, whose application to different systems becomes a simple matter of scaling. Note that Hill's case of orbits close to the smaller primary is a simplification of the restricted three-body problem, which in turn is a simplification of real models.

A classical result shows that low eccentricity orbits around a primary body are unstable for moderate and high inclinations due to third-body perturbations [10]. Almost circular orbits close to the central body remain with low eccentricity in the long-term only when the mutual inclination with the perturbing body is less than the critical inclination of the third-body perturbations $I = 39.2°$ (see [11] and references therein). Because of their

low eccentricity, high inclination orbits are precisely the candidate orbits for science missions around natural satellites. Therefore, a good understanding of the unstable dynamics of the Hill problem is required.

The study of the long-term dynamics is usually done in the double-averaged problem. After removing the short- and long-period terms, and truncating higher-order terms, the problem is reduced to one degree of freedom in the eccentricity and the argument of the periapsis. As the double-averaged problem is integrable and the corresponding phase space is a compact manifold, the solutions are closed curves and equilibria. The latter are orbits that, on average, have almost constant eccentricity and fixed argument of the periapsis, and are known as frozen orbits.

To each trajectory of the double-averaged problem it corresponds a torus of quasiperiodic solutions in the nonaveraged problem. The accurate computation of initial conditions on the torus requires the recovery of the short- and long-period effects that were eliminated in the averaging. This is normally done by trial and error, making iterative corrections on the orbital elements, although other procedures can be applied [12].

Our analytical theory is computed with Deprit's perturbation technique [13]. The procedure is systematic and has the advantage of providing the explicit transformation equations that connect the averaged analysis with proper initial conditions of the nonaveraged problem. A second-order truncation of the Hamiltonian shows that there are no degenerate equilibria and, therefore, it is sufficient to give the qualitative description of the reduced system. However, the second-order truncation introduces a symmetry between the direct and retrograde orbits that is not part of the original problem, and a third-order truncation is required to reveal the nonsymmetries of the problem.

While, in general, the third-order theory provides good results in the computation of quasiperiodic, frozen orbits, its solutions are slightly affected by long-period oscillations. This fact may adversely affect the long-term evolution of the frozen orbits and it becomes apparent in the computation of science orbits about planetary satellites, a case in which small perturbations are enough for the unstable dynamics to defrost the argument of the periapsis. Then, the orbit immediately migrates along the unstable manifold with an exponential increase in the eccentricity.

We find that a higher-order truncation is desirable if one wants to use the analytical theory for computing accurate initial conditions of frozen orbits. The computation of the fourth-order truncation removes almost all adverse effects from the quasiperiodic solutions, and shows a high degree of agreement between the averaged and nonaveraged models even in the case of unstable orbits.

Whereas the third-body perturbation is the most important effect in destabilizing science orbits around planetary satellites, the impact of the nonsphericity of the central body may be taken into account. The previous research including both effects has been limited up to third-order theories (see [14] and references therein), but from the conclusions of this paper it may worth to develop a higher-order theory including the inhomogeneities of the satellite's gravitational potential.

2. Double-Averaged Hill Problem to the Fourth-Order

The equations of motion of the Hill problem are derived from the Hamiltonian

$$\mathcal{H} = \left(\frac{1}{2}\right)(\mathbf{X}\cdot\mathbf{X}) - \boldsymbol{\omega}\cdot(\mathbf{x}\times\mathbf{X}) + W(\mathbf{x}), \qquad W = \left(\frac{\omega^2}{2}\right)\left(r^2 - 3x^2\right) - \frac{\mu}{r}, \tag{2.1}$$

where, in the standard coordinate system of Hill's model, $\mathbf{x} = (x, y, z)$ is the position vector, $\mathbf{X} = (X, Y, Z)$ is the vector of conjugate momenta, $r = ||\mathbf{x}||$, and both the rotation rate of the system $\omega = ||\boldsymbol{\omega}||$ and the gravitational parameter μ of the primary are set to 1 in appropriate units.

The problem is of three degrees of freedom, yet admitting the Jacobi constant $\mathscr{H} = -C/2$. Despite its nonintegrability, approximate solutions that explain the long-term dynamics can be found by perturbation methods. Close to the central body the Hill problem can be written as the perturbed two-body problem

$$\mathscr{H} = \left(\frac{1}{2}\right)\left(X^2 + Y^2 + Z^2\right) - \left(\frac{1}{r}\right) - \epsilon\ (x\ Y - y\ X) + \left(\frac{\epsilon^2}{2}\right)\left(r^2 - 3x^2\right), \tag{2.2}$$

where the first three terms of Hamiltonian (2.2) correspond to the Keplerian motion in the rotating frame and ϵ is a formal parameter introduced to manifest the importance of each effect. Thus, the Coriolis term is a first order effect and the third-body perturbation appears at the second-order .

To apply perturbation theory, we formulate the problem in Delaunay variables (ℓ, g, h, L, G, H), where ℓ is the mean anomaly, g is the argument of the periapsis, h the argument of the node in the rotating frame, $L = \sqrt{\mu\ a}$ is the Delaunay action, $G = L\ \sqrt{1 - e^2}$ is the modulus of the angular momentum vector, $H = G\cos I$ is its polar component, and a, e, I, are usual orbital elements: semimajor axis, eccentricity, and inclination.

Our theory is based on the use of Lie transforms as described by Deprit [13, 15]. It has the advantage of connecting the averaged and original problems through explicit transformation equations. After removing the short- and long-period terms we get the transformed Hamiltonian

$$\mathscr{K} = K_{0,0} + \varepsilon K_{0,1} + \left(\frac{\varepsilon^2}{2}\right)K_{0,2} + \left(\frac{\varepsilon^3}{6}\right)K_{0,3} + \left(\frac{\varepsilon^4}{24}\right)K_{0,4}, \tag{2.3}$$

where $\varepsilon = L^3$,

$$\begin{aligned}
K_{0,0} &= -\frac{1}{(2L^2)},\\
K_{0,1} &= K_{0,0}\ 2\sigma,\\
K_{0,2} &= K_{0,0}\left(\frac{1}{4}\right)\left[\left(2 + 3e^2\right)\ \left(2 - 3s^2\right) + 15e^2s^2\cos 2g\right],\\
K_{0,3} &= K_{0,0}\left(\frac{27}{32}\right)\sigma\left[2s^2 + \left(50 - 17s^2\right)e^2 + 15e^2s^2\cos 2g\right],\\
K_{0,4} &= K_{0,0}\left(-\frac{3}{512}\right)\left\{3285s^4e^4\cos 4g - 12s^2\left[3996 - 2940s^2 - \left(4582 - 4035s^2\right)e^2\right]e^2\cos 2g\right.\\
&\qquad + 8\left(784 - 708s^2 - 9s^4\right) - 144\left(926 - 941s^2 + 244s^4\right)e^2\\
&\qquad \left. + 9\left(10728 - 15208s^2 + 5007s^4\right)e^4\right\},
\end{aligned} \tag{2.4}$$

and $\sigma = H/L = c\eta$, $\eta = \sqrt{1-e^2}$ is the eccentricity function, and we use the abbreviations $s \equiv \sin I$, $c \equiv \cos I$. Details on the perturbation method and expressions to compute the transformation equations of the averaging are given in the appendix.

The double-averaged Hamiltonian (2.3) depends neither on the mean anomaly nor on the argument of the node. Therefore, the corresponding conjugate momenta, L and H, are integrals of the reduced problem and the Hamiltonian (2.3) represents a one degree of freedom problem in g and G. The equations of motion are computed from Hamilton equations $\mathrm{d}g/\mathrm{d}t = \partial\mathcal{K}/\partial G$, $\mathrm{d}G/\mathrm{d}t = -\partial\mathcal{K}/\partial g$,

$$\begin{aligned}
\frac{\mathrm{d}g}{\mathrm{d}t} &= \frac{6}{G}\left[5c^2-\eta^2-5\left(c^2-\eta^2\right)\cos 2g\right]+\frac{27\varepsilon\sigma}{4G}\left[5c^2+11\eta^2-5\left(c^2-\eta^2\right)\cos 2g\right]\\
&\quad+\frac{3\varepsilon^2}{128G}\Big\{2113c^2-3285c^4+\left(3915+9165c^4\right)\eta^2+\left(1581+7791c^2\right)\eta^4\\
&\qquad-4\left[802c^2-1095c^4+\left(19+2565c^4\right)\eta^2-\left(547+1744c^2\right)\eta^4\right]\cos 2g\\
&\qquad+1095e^2s^2\left(c^2-\eta^2\right)\cos 4g\Big\},\\
\frac{\mathrm{d}G}{\mathrm{d}t} &= -\frac{3}{4}e^2\,s^2\Bigg\{5(8+9\varepsilon\sigma)\sin 2g\\
&\qquad+\frac{\varepsilon^2}{32}\left(2\left[509-1095c^2+\left(547+4035c^2\right)\eta^2\right]\sin 2g-1095e^2s^2\sin 4g\right)\Bigg\}.
\end{aligned} \tag{2.5}$$

Once g and G are integrated for given initial conditions, the secular variations of ℓ and h are computed from simple quadratures derived from Hamilton equations $\mathrm{d}h/\mathrm{d}t = \partial\mathcal{K}/\partial H$, $\mathrm{d}\ell/\mathrm{d}t = \partial\mathcal{K}/\partial L$,

$$h = h_0 + \int \frac{\partial}{\partial H}\mathcal{K}\big(g(t),G(t);H,L\big)\mathrm{d}t, \qquad \ell = \ell_0 + \int \frac{\partial}{\partial L}\mathcal{K}\big(g(t),G(t);H,L\big)\mathrm{d}t. \tag{2.6}$$

3. Qualitative Dynamics

The flow can be integrated from the differential equations mentioned previously, (2.5). However, since the system defined by (2.5) is integrable, the flow is made of closed curves and equilibria, and it can be represented by contour plots of Hamiltonian (2.3). Thus, for given values of the dynamical parameters L and H—or ε and σ—we can plot the flow in different maps that are function of g, G. Figure 1 shows an example in semiequinoctial elements $(e\cos g, e\sin g)$, where we note a hyperbolic point corresponding to an unstable circular orbit, and two elliptic points corresponding to two stable elliptic orbits with $e = 0.2$ and periapsis at $g = \pm\pi/2$, respectively.

Delaunay variables are singular for zero eccentricity orbits, where the argument of the periapsis and the mean anomaly are not defined, and for equatorial orbits, where the

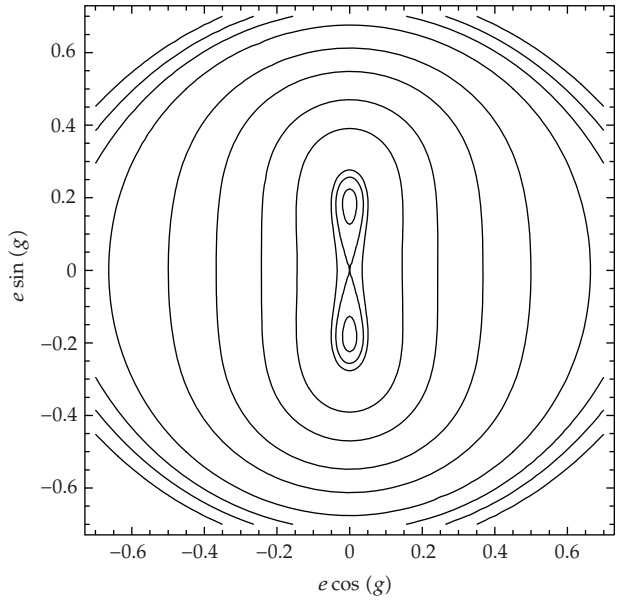


Figure 1: Flow in the doubly reduced phase space.

argument of the node is not defined. Hence, it is common to study the reduced phase space in the variables introduced by Coffey et al. [16], see also [8]:

$$\chi_1 = \eta e s \cos g, \qquad \chi_2 = \eta e s \sin g, \qquad \chi_3 = \eta^2 - \frac{1}{2}\left(1+\sigma^2\right), \tag{3.1}$$

that define the surface of a sphere

$$\chi_1^2 + \chi_2^2 + \chi_3^2 = \frac{1}{4}\left(1-\sigma^2\right)^2 \tag{3.2}$$

of radius $R = (1/2)\ (1-\sigma^2)$ (the sphere representation misses the case $G = H = 0$, irrelevant in astrodynamics.)

Then, after dropping constant terms and scaling, Hamiltonian (2.3) writes

$$\begin{aligned}\mathcal{K} = {} & -12\eta^2 - \frac{30\chi_2^2}{\eta^2} + \frac{9}{4}\,\varepsilon\sigma\left(25 - 24\eta^2 - \sigma^2 - 15\,\frac{\chi_2^2}{\eta^2}\right) \\ & + \frac{\varepsilon^2}{64}\left[3815 + 9528\sigma^2 + 9\sigma^4 - 6\left(343 + 1709\sigma^2\right)\eta^2 - 1824\eta^4\right. \\ & \left. + 6\left(293 - 821\eta^2 - 1470\sigma^2\right)\frac{\chi_2^2}{\eta^2} - 3285\,\frac{\chi_2^4}{\eta^4}\right].\end{aligned} \tag{3.3}$$

The flow on the sphere is obtained from Liouville equations $\dot{\chi}_i = \{\chi_i; \mathcal{K}\}$, $i = 1, 2, 3$, where the dot means derivative in the new time scale. Then,

$$\begin{aligned}\dot{\chi}_1 = \frac{3}{16\eta}\,\chi_2 \Bigg\{ & 64\left(5 - 8\eta^2 + 5\sigma^2\right) + 72\varepsilon\sigma\left(5 - 2\eta^2 + 5\sigma^2\right) \\ & + \frac{\varepsilon^2}{64}\Bigg[3815 - 1824\eta^4 + 9528\sigma^2 + 9\sigma^4 - 6\eta^2\left(343 + 1709\sigma^2\right) \\ & + 6\left(293 - 821\eta^2 - 1470\sigma^2\right)\left(\frac{\chi_2}{\eta}\right)^2 - 3285\left(\frac{\chi_2}{\eta}\right)^4\Bigg]\Bigg\},\end{aligned} \tag{3.4}$$

$$\begin{aligned}\dot{\chi}_2 = -\frac{3}{16\eta}\,\chi_1 \Bigg\{ & 608\varepsilon^2\eta^4 + \eta^2\left[128 + 576\varepsilon\sigma + \varepsilon^2\left(343 + 1709\sigma^2\right)\right] \\ & - \left[320 + 360\varepsilon\,\sigma - \varepsilon^2\left(293 - 1470\sigma^2\right)\right]\left(\frac{\chi_2}{\eta}\right)^2 - 1095\varepsilon^2\left(\frac{\chi_2}{\eta}\right)^4\Bigg\},\end{aligned} \tag{3.5}$$

$$\dot{\chi}_3 = \frac{3}{8\eta}\,\chi_1\,\chi_2\left\{320 + 360\varepsilon\sigma - \varepsilon^2\left[293 - 1470\sigma^2 - 821\eta^2 - 1095\left(\frac{\chi_2}{\eta}\right)^2\right]\right\}, \tag{3.6}$$

with the constraint $\chi_1\,\dot{\chi}_1 + \chi_2\,\dot{\chi}_2 + \chi_3\,\dot{\chi}_3 = 0$, derived from (3.2).

Equations (3.4)–(3.6) show that circular orbits ($\chi_1 = \chi_2 = 0$, $\chi_3 = R$, the "north" pole of the sphere) are always equilibria. Equations (3.5) and (3.6) vanish when $\chi_1 = 0$, $\chi_2 \neq 0$, but (3.4) vanishes only when

$$\begin{aligned}1095\varepsilon^2\sigma^4 - \sigma^2\left[320 + 360\varepsilon\sigma + \varepsilon^2\left(802 + 2565\sigma^2\right)\right]\eta^2 \\ + \left[192 - 216\varepsilon\sigma - \varepsilon^2\left(362 - 35\sigma^2\right)\right]\eta^6 - 61\varepsilon^2\eta^8 = 0.\end{aligned} \tag{3.7}$$

Equation (3.7) is a polynomial equation of degree 8 in η, therefore admitting eight roots. Note that, for the accepted values of $\varepsilon \ll 1$, (3.7) is of the form $A_1^2 - A_2^2\,x + A_3^2\,x^3 - A_4^2\,x^4 = 0$ that admits a maximum of three real roots, according to Descartes' rule of signs.

The real roots of (3.7) verified by the dynamical constraint $|\sigma| \leq \eta \leq 1$ are also equilibria. The root $\eta = 1$ marks a change in the number of equilibria due to a "bifurcation" ($\eta > 1$ could be a root but not an equilibrium). Then, the number of equilibria changes when crossing the line

$$\varepsilon = 4\,\frac{-27\sigma - 45\sigma^3 \pm \sqrt{5076 + 1473\sigma^2 + 4730\sigma^4 - 27375\sigma^6}}{423 + 767\sigma^2 + 1470\sigma^4} \tag{3.8}$$

obtained setting $\eta = 1$ in (3.7) that establishes a relation between the dynamical parameters ε and σ corresponding to bifurcations of circular orbits. Figure 2 shows that this line defines two regions in the parameters plane with different number of equilibria in phase space. Circular orbits in the outside region of the curve are stable. When crossing the line given by (3.8) the number of real roots of (3.7) with dynamical sense increases such that a pitchfork

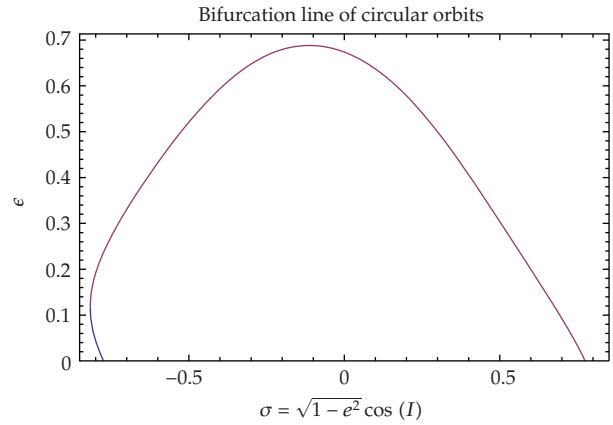


Figure 2: Regions in the parameters plane with different numbers of equilibria.

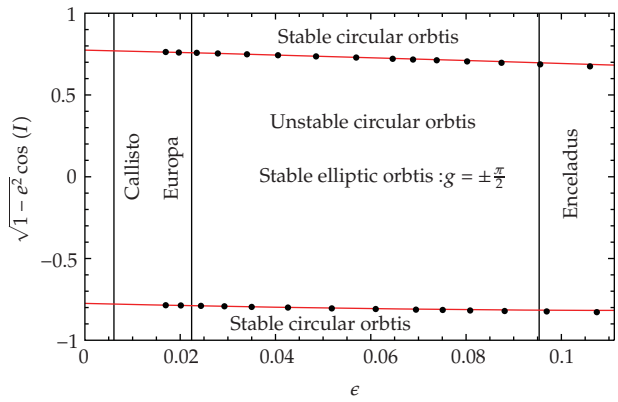


Figure 3: Bifurcation lines in the parameter plane.

bifurcation takes place: circular orbits change to unstable and two stable elliptic orbits appear with periapsis, respectively at $g = \pm\pi/2$, as in the example of Figure 1.

Note that the curve given by (3.8) notably modifies the classical inclination limit $\cos^2 I > 3/5$ for circular orbits' stability. However, we cannot extend the practical application of the analytical theory to any value of ε. It is common to limit the validity of the Hill problem approximation to one third of the Hill radius $r_H = 3^{-1/3}$. Then $\varepsilon < (r_H/3)^{3/2} = 1/9$, including most of the planetary satellites of interest. Figure 3 shows the bifurcation lines of circular orbits in the validity region of the parameters plane with the values of ε corresponding to low altitude orbits around different planetary satellites highlighted.

A powerful test for estimating the quality of the analytical theory is to check the degree of agreement of the bifurcation lines of the analytical theory with those computed numerically in the nonaveraged problem. To do that we compute several families of three-dimensional, almost circular, periodic orbits of the Hill (nonaveraged) problem that bifurcate from the family of planar retrograde orbits at different resonances. For variations of the Jacobi constant the almost circular periodic orbits evolve from retrograde to direct orbits through the 180 degrees of inclination. At certain critical points, almost circular orbits change from stable to unstable in a bifurcation phenomenon in which two new elliptic periodic orbits appear.

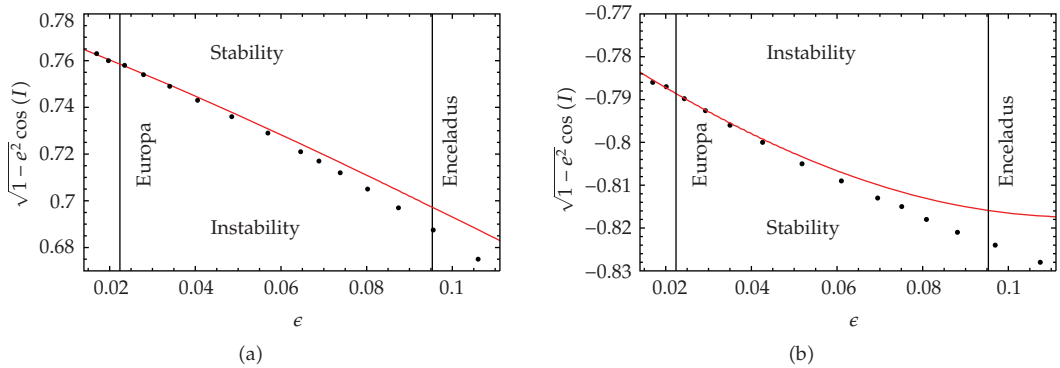


Figure 4: Comparison between the bifurcation line of circular, averaged orbits (full line), and the curve of critical periodic orbits (dots).

Table 1: Initial orbital elements of an elliptic frozen orbit for $\varepsilon = 0.0470573$, $\sigma = 0.422618$.

Theory	a (Hill units)	e	I (deg)	g (deg)	h	ℓ
Classical	0.130342	0.674094	55.0995	−90	0	0
2nd order	0.130342	0.648065	55.6915	−90	0	0
3rd order	0.130515	0.637316	56.1798	−90	0	0
4th order	0.130538	0.634803	56.2813	−90	0	0

The computation of a variety of these critical points helps in determining stability regions for almost circular orbits [17].

The tests done show that the fourth-order theory gives good results for $\varepsilon < 0.05$. As presented in Figure 4, the bifurcation line of retrograde orbits clearly diverges from the line of corresponding critical periodic orbits for higher values of ε, and it may be worth developing a higher-order theory that encompasses also the case of Enceladus.

4. Frozen Orbits Computation

Hill's case of orbits close to the smaller primary is a simplification of the restricted three-body problem, which in turn is a simplification of real models. Therefore, the final goal of our theory is not the generation of ephemerides but to help in mission designing for artificial satellites about planetary satellites, where frozen orbits are of major interest.

For given values of the parameters ε, σ, determined by the mission, a number of frozen orbits may exist. A circular frozen orbit, either stable or unstable, exists always and the computation of real roots $|\sigma| \leq \eta \leq 1$ of (3.7), if any, will provide the eccentricities of the stable elliptic solutions with frozen periapsis at $g = \pm\pi/2$. To each equilibrium of the doubly reduced phase space it corresponds a torus of quasiperiodic solutions in the original, nonaveraged model. In what follows we present several examples that justify the effort in computing a fourth-order theory to reach the quasiperiodicity condition in the Hill problem.

4.1. Elliptic Frozen Orbits

We choose $\varepsilon = 0.0470573$, $\sigma = 0.422618$. If we first try the classical double-averaged solution, the Hamiltonian (2.3) is simplified to $K_{0,0} + \varepsilon K_{0,1} + (\varepsilon^2/2)K_{0,2}$, and the existence of elliptic

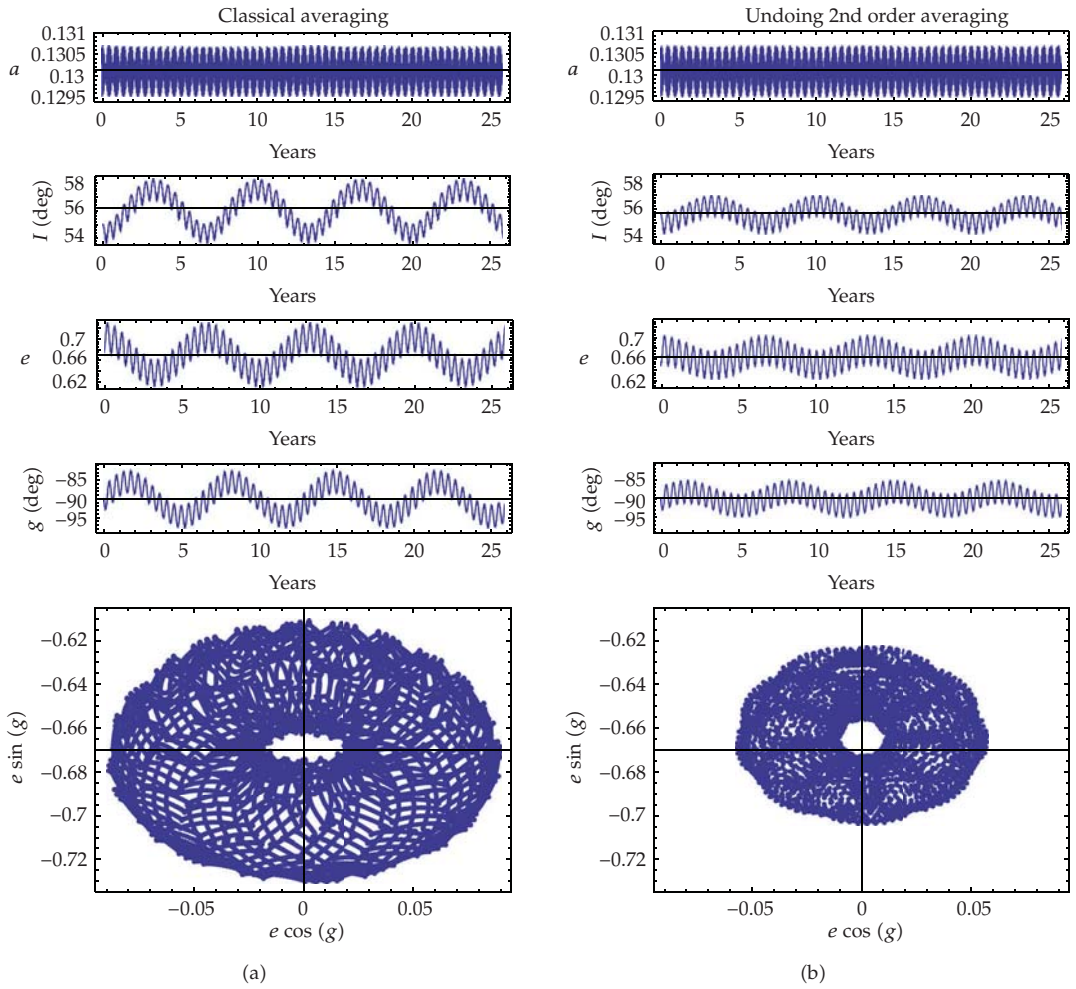


Figure 5: Long-term evolution of the orbital elements of the elliptic frozen orbit.

frozen orbits reduces to the case $\sigma^2 < 3/5$, $g = \pm\pi/2$. The eccentricity of the elliptic frozen solutions is then computed from $\eta = (5\sigma^2/3)^{1/4}$ —obtained by neglecting terms in ε in (3.7). Thus, for the given values of ε and σ, and taking into account that we are free to choose the initial values of the averaged angles ℓ, h, we get the orbital elements of the first row of Table 1. The left column of Figure 5 shows the long-term evolution of the instantaneous orbital elements for this case, that we call "classical averaging," in which we find long-period oscillations of more than four degrees in inclination, more than fifteen in the argument of periapsis, and a variation of ±0.06 in the eccentricity.

When computing a second-order theory with the Lie-Deprit perturbation method we arrive exactly at the classical Hamiltonian obtained by a simple removal of the short-period terms and the classical bifurcation condition that results in the critical inclination of the third-body perturbations $I = 39.2°$ [10, 11]. However, now we have available the transformation equations to recover the short- and long-period effects, although up to the first order only. After undoing the transformation equations we find the orbital elements of the second row of Table 1, where we see that all the elements remain unchanged except for

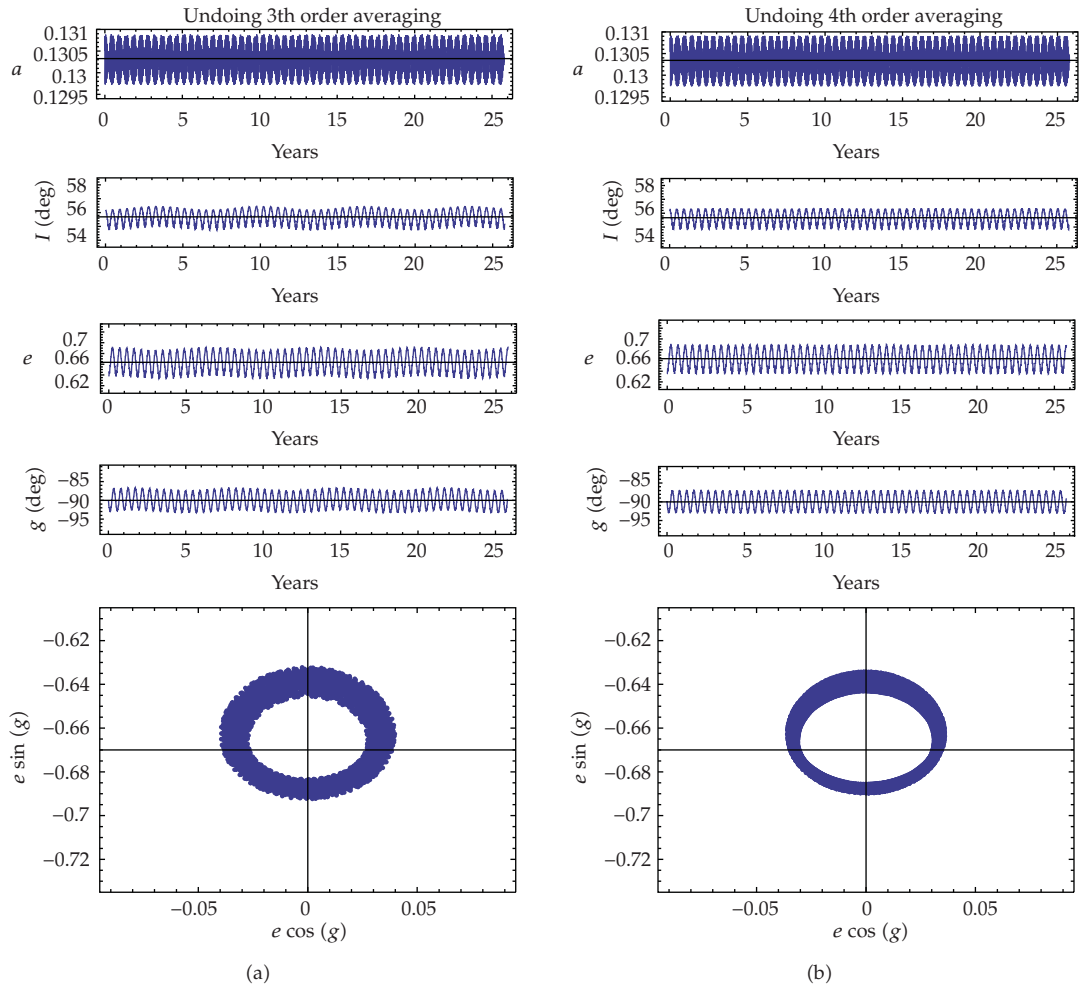


Figure 6: Long-term evolution of the orbital elements of the elliptic frozen orbit.

the eccentricity and inclination. The long-term evolution of these elements is presented in the right column of Figure 5, in which we notice a significant reduction in the amplitude of long-period oscillations: 2.5° in inclination, around 10° in the argument of the periapsis, and ±0.04 in eccentricity.

The results of the third- and fourth-order theories are presented in the last two rows of Table 1 and in Figure 6. The higher-order corrections drive slight enlargements in the semimajor axis. While both higher-order theories produce impressive improvements, we note a residual long-period oscillation in the elements computed from the third-order theory (left column of Figure 6). On the contrary, the orbital elements of the frozen orbit computed with the fourth-order theory are almost free from long-period oscillations and mainly show the short-period oscillations typical of quasiperiodic orbits.

4.2. Circular Frozen Orbits

If we choose the same value for ε but now $\sigma = 0.777146$, frozen elliptic orbits do not exist any longer and the circular frozen orbit is stable. Both the third and fourth-order theories provide

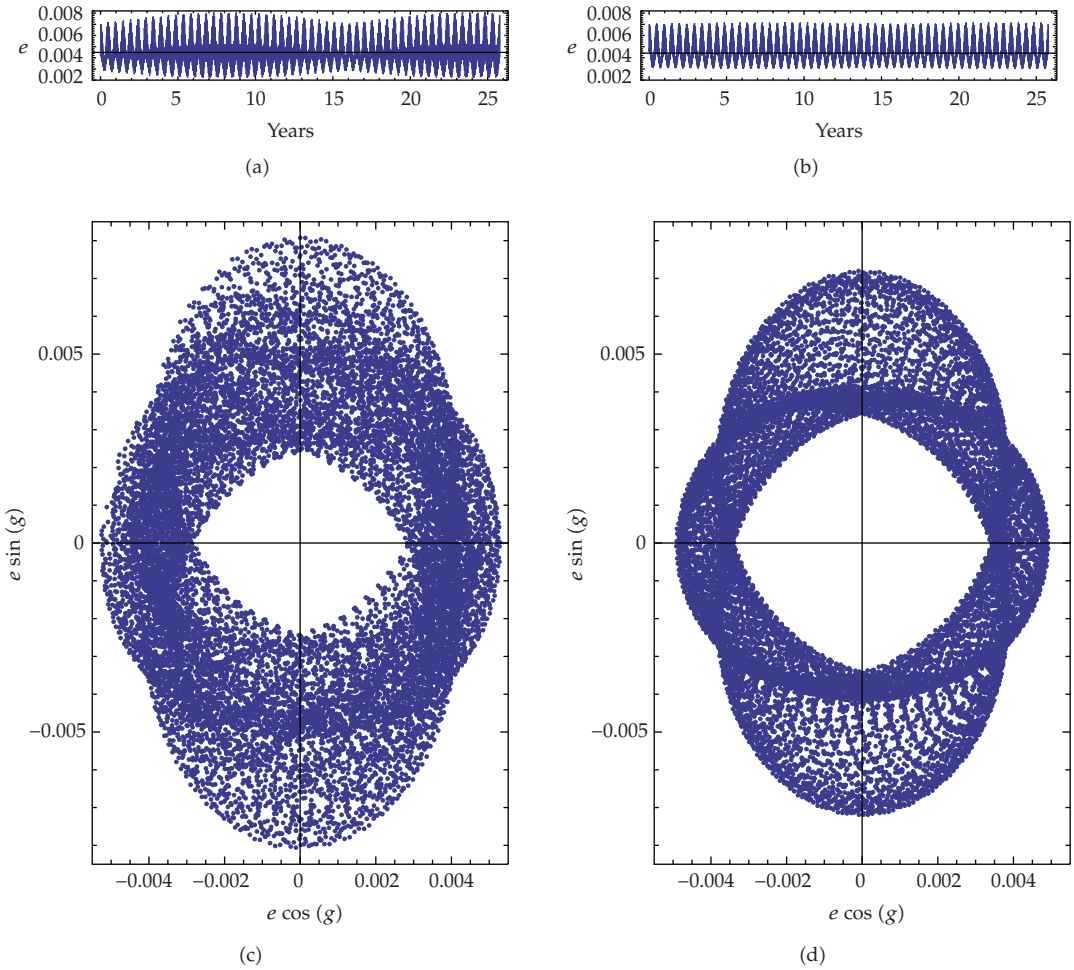


Figure 7: Long-term evolution of the orbital elements of the circular stable frozen orbit. (a) and (c) third-order theory. (b) and (d) fourth-order theory.

good results, but, again, the third-order theory provides small long-period oscillations in the eccentricity whereas the fourth-order theory leads to a quasiperiodic orbit (see Figure 7).

For $\varepsilon = 0.0339919$ and $\sigma = 0.34202$ the circular frozen orbit is unstable. Due to the instability, a long-term propagation of the initial conditions from either the third or the fourth theory shows that the orbit escapes following the unstable manifold with exponential increase in the eccentricity. But, as Figure 8 shows, the orbit remains frozen much more time when using the fourth-order theory. A variety of tests performed on science orbits close to Galilean moons Europa and Callisto showed that the fourth-order theory generally improves by 50% the lifetimes reached when using the third-order theory.

4.3. Fourier Analysis

Alternatively to the temporal analysis mentioned previously, a frequency analysis using the Fast Fourier Transform (FFT) shows how initial conditions obtained from different orders of the analytical theory can be affected of undesired frequencies that defrost the orbital elements.

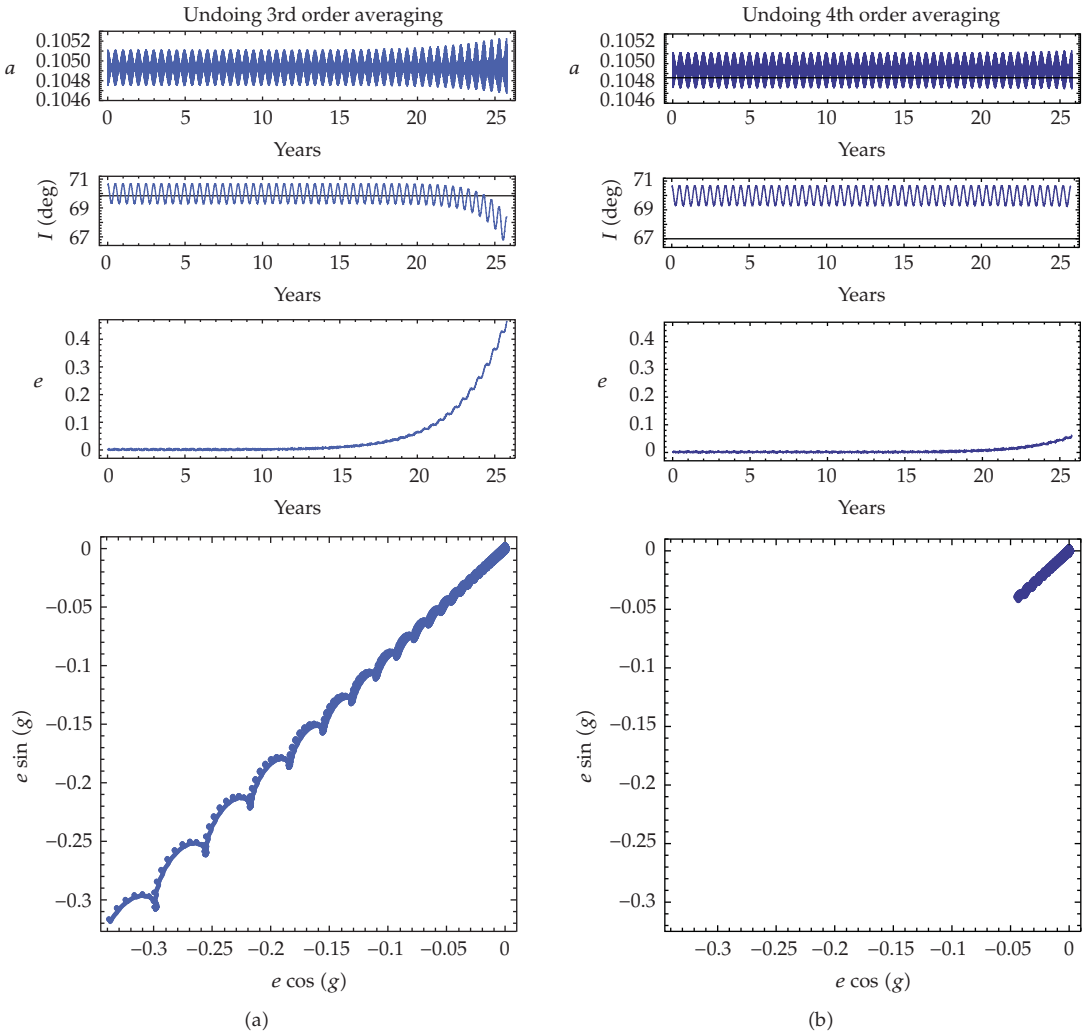


Figure 8: Long-term evolution of the orbital elements of the circular, unstable, frozen orbit.

Thus, Figure 9 shows the FFT analysis of the instantaneous argument of the periapsis of the elliptic orbit in the example mentioned previously. Dots correspond to initial conditions obtained from the double-averaged phase space after a classical analysis—that is equivalent to the second-order analytical theory—and the line corresponds to initial conditions obtained from the fourth-order analytical theory after undoing the transformation. While most of the frequencies match with similar amplitudes, in the magnification of the right plot we clearly appreciate a very low frequency of ~0.15 cycles/year with a very high amplitude in the classical theory that is almost canceled out with the fourth-order approach. The semiannual frequency remains in both theories because it is intrinsic to the problem. It is due to the third-body perturbation and it cannot be avoided.

Figure 10 shows a similar analysis for the instantaneous eccentricity of the stable circular orbit mentioned previously. Now, dots correspond to the fourth-order theory and the line to the third-order one (both after undoing the transformation equations). While the

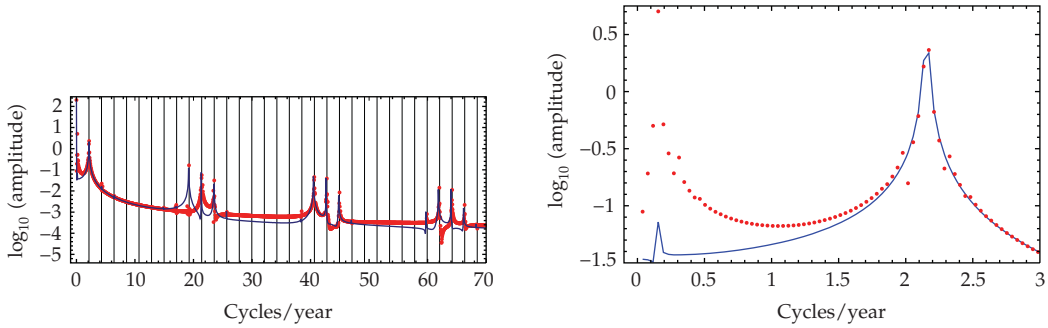


Figure 9: (a) FFT analysis of the instantaneous argument of the periapsis of the elliptic solution. (b) Magnification over the low frequencies region.

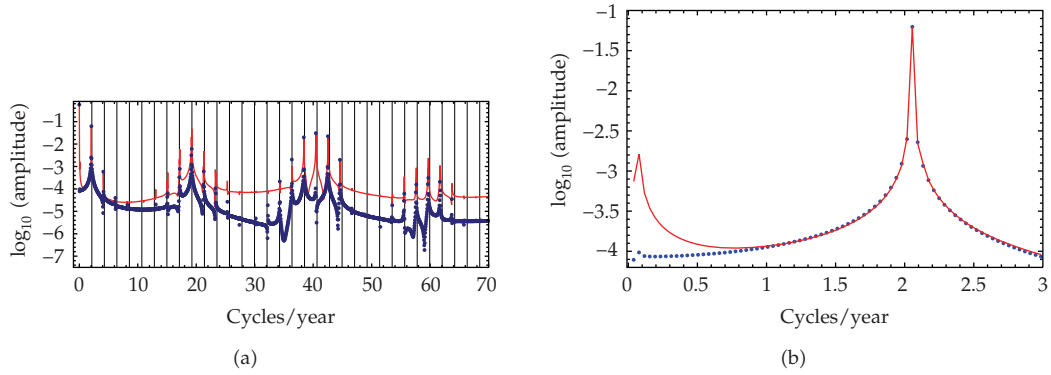


Figure 10: (a) FFT analysis of the instantaneous eccentricity of the elliptic solution. (b) Magnification over the low frequencies region.

third-order theory provides good results, reducing the amplitude of the undesired frequency to low values, the fourth-order theory practically cancels out that frequency.

An FFT analysis of unstable circular orbits has not much sense because of the time scale in which the orbit destabilizes.

5. Conclusions

Frozen orbits computation is a useful procedure in mission designing for artificial satellites. After locating the frozen orbit of interest in a double-averaged problem, usual procedures for computing initial conditions of frozen orbits resort to trial-and-error interactive corrections, or require involved computations. However, the explicit transformation equations between averaged and nonaveraged models can be obtained with analytical theories based on the Lie-Deprit perturbation method, which makes the frozen orbits computations straightforward.

Accurate computations of the initial conditions of frozen, quasiperiodic orbits can be reached with higher-order analytical theories. This way of proceeding should not be undervalued in the computation of science orbits around planetary satellites, a case in which third-body perturbations induce unstable dynamics.

Higher-order analytical theories are a common tool for computing ephemeris among the celestial mechanics community. They are usually developed with specific purpose, sophisticated algebraic manipulators. However, the impressive performances of modern computers and software allow us to build our analytical theory with commercial, general-purpose manipulators, a fact that may challenge aerospace engineers to use the safe, well-known techniques advocated in this paper.

Appendix

Let $\mathcal{T} : (\mathbf{x}, \mathbf{X}) \rightarrow (\mathbf{x}', \mathbf{X}')$, where $\mathbf{x}$ are coordinates and $\mathbf{X}$ their conjugate momenta, be a Lie transform from "new" (primes) to "old" variables. If $W = \sum_i (\epsilon^i / i!)\, W_{i+1}(\mathbf{x}, \mathbf{X})$ is its generating function expanded as a power series in a small parameter ϵ, a function $\mathcal{F} = \sum_i (\epsilon^i / i!)\, F_{i,0}(\mathbf{x}, \mathbf{X})$ can be expressed in the new variables as the power series $(\mathcal{T} : \mathcal{F}) = \sum_i (\epsilon^i / i!)\, F_{0,i}(\mathbf{x}', \mathbf{X}')$ whose coefficients are computed from the recurrence

$$F_{i,j} = F_{i+1,j-1} + \sum_{0 \le k \le i} \binom{i}{k} \{F_{k,j-1}; W_{i+1-k}\}, \tag{A.1}$$

where $\{F_{k,j-1}; W_{i+1-k}\} = \nabla_{\mathbf{x}} F_{k,j-1} \cdot \nabla_{\mathbf{X}} W_{i+1-k} - \nabla_{\mathbf{X}} F_{k,j-1} \cdot \nabla_{\mathbf{x}} W_{i+1-k}$, is the Poisson bracket. Conversely, the coefficients W_{i+1} of the generating function can be computed step by step from (A.1) once corresponding terms $F_{0,i}$ of the transformed function are chosen as desired. In perturbation theory it is common to chose the $F_{0,i}$ as an averaged expression over some variable, but it is not the unique possibility [18]. Full details can be found in the literature [19, 20].

To average the short-period effects we write Hamiltonian (2.2) in Delaunay variables as

$$\mathcal{H} = H_{0,0} + \epsilon H_{1,0} + \left(\frac{\epsilon^2}{2}\right) H_{2,0} + \left(\frac{\epsilon^3}{6}\right) H_{3,0} + \left(\frac{\epsilon^4}{24}\right) H_{4,0}, \tag{A.2}$$

where $H_{0,0} = -1/(2L^2)$, $H_{1,0} = -H$, $H_{2,0} = r^2\{1 - 3[\cos(f+g)\cos h - c\sin(f+g)\sin h]^2\}$, and $H_{3,0} = H_{4,0} = 0$. Note that the true anomaly f is an implicit function of ℓ.

Since the radius r never appears in denominators, it results convenient to express Hamiltonian (A.2) as a function of the elliptic—instead of the true—anomaly u by using the ellipse relations $r \sin f = \eta\ a \sin u$, $r \cos f = a(\cos u - e)$, $r = a(1 - e\cos u)$.

After applying the Delaunay normalization [21] up to the fourth-order in the Hamiltonian, we get

$$\mathcal{H}' = H_{0,0} + \epsilon\ H_{0,1} + \left(\frac{\epsilon^2}{2}\right) H_{0,2} + \left(\frac{\epsilon^3}{6}\right) H_{0,3} + \left(\frac{\epsilon^4}{24}\right) H_{0,4}, \tag{A.3}$$

where, omitting primes,

$$H_{0,0} = -\frac{1}{2L^2},$$

$$H_{0,1} = H_{0,0}\varepsilon 2c\eta,$$

$$
\begin{aligned}
H_{0,2} &= H_{0,0}\left(\frac{\varepsilon^2}{8}\right)\left\{\left(4+6e^2\right)\ \left(2-3s^2+3s^2\cos 2h\right)\right.\\
&\qquad \left.+\ 15e^2\left[2s^2\cos 2g+(1-c)^2\cos(2g-2h)+(1+c)^2\cos(2g+2h)\right]\right\},\\
H_{0,3} &= H_{0,0}\left(\frac{45\varepsilon^3}{8}\right)e^2\eta\left[(1+c)^2\cos(2g+2h)-(1-c)^2\cos(2g-2h)\right],\\
H_{0,4} &= H_{0,0}\left(-\frac{3\varepsilon^4}{512}\right)\\
&\quad\times\left\{16\left(47+282c^2+63c^4\right)-144\left(227+90c^2+59c^4\right)e^2\right.\\
&\qquad -18\left(227+610c^2-701c^4\right)e^4-24s^2\left[558+270c^2+\left(109-555c^2\right)e^2\right]e^2\cos 2g\\
&\qquad +24s^2\left[216+56c^2-8\ \left(161+59c^2\right)e^2-\left(11-701c^2\right)e^4\right]\cos 2h\\
&\qquad -48(1+c)^2\left[338-90c+90c^2-\left(91-185c+185c^2\right)e^2\right]e^2\cos(2g+2h)\\
&\qquad -48(1-c)^2\left[338+90c+90c^2-\left(91+185c+185c^2\right)e^2\right]e^2\cos(2g-2h)\\
&\qquad +6s^4\left(56-472e^2+701e^4\right)\cos 4h+1710s^4e^4\cos 4g\\
&\qquad -60s^2\left(18-37e^2\right)e^2\left[(1+c)^2\cos(2g+4h)+(1-c)^2\cos(2g-4h)\right]\\
&\qquad +1140s^2e^4\left[(1+c)^2\cos(4g+2h)+(1-c)^2\cos(4g-2h)\right]\\
&\qquad \left.+\ 285e^4\left[(1+c)^4\cos(4g+4h)+(1-c)^4\cos(4g-4h)\right]\right\}.
\end{aligned}
\tag{A.4}
$$

The generating function of the transformation is $W = W_2 + (1/2)W_3$, where

$$
\begin{aligned}
W_2 &= L\left(\frac{\varepsilon^2}{192}\right)\\
&\quad\times\left\{4\left(2-3s^2\right)\left[3e\left(5+3\eta^2\right)\ S_{1,0,0}-9e^2S_{2,0,0}+e^3S_{3,0,0}\right]\right.\\
&\qquad +6s^2e\left[3\ \left(5+3\eta^2\right)(S_{1,0,2}+S_{1,0-2})-9e(S_{2,0,2}+S_{2,0-2})+e^2(S_{3,0,2}+S_{3,0-2})\right]\\
&\qquad +6s^2(1+\eta)^2\left[15eS_{1,2,0}-(9-6\eta)S_{2,2,0}+eS_{3,2,0}\right]\\
&\qquad +6s^2(1-\eta)^2\left[15eS_{1,-2,0}-(9+6\eta)S_{2,-2,0}+eS_{3,-2,0}\right]\\
&\qquad +3(1+c)^2(1+\eta)^2\left[15eS_{1,2,2}-(9-6\eta)S_{2,2,2}+eS_{3,2,2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+3(1-c)^2(1+\eta)^2\left[15eS_{1,2,-2}-(9-6\eta)S_{2,2,-2}+eS_{3,2,-2}\right]\\
&+3(1-c)^2(1-\eta)^2\left[15eS_{1,-2,2}-(9+6\eta)S_{2,-2,2}+eS_{3,-2,2}\right]\\
&+3(1+c)^2(1-\eta)^2\left[15eS_{1,-2,-2}-(9+6\eta)S_{2,-2,-2}+eS_{3,-2,-2}\right]\Big\},\\
W_3 = L&\left(\frac{\varepsilon^3}{256}\right)\\
&\times\Big\{72es^2\left(13+3\eta^2\right)[S_{1,0,2}-S_{1,0,-2}]-24e^2s^2\left(17+4\eta^2\right)[S_{2,0,2}-S_{2,0,-2}]\\
&+88e^3s^2[S_{3,0,2}-S_{3,0,-2}]-6e^4s^2[S_{4,0,2}-S_{4,0,-2}]\\
&+36e(1+\eta)\left(13+\eta+8\eta^2\right)\left[(1+c)^2S_{1,2,2}-(1-c)^2S_{1,2,-2}\right]\\
&+36e(1-\eta)\left(13-\eta+8\eta^2\right)\left[(1-c)^2S_{1,-2,2}-(1+c)^2S_{1,-2,-2}\right]\\
&-12(1+\eta)^2\left(17-6\eta-8\eta^2\right)\left[(1+c)^2S_{2,2,2}-(1-c)^2S_{2,2,-2}\right]\\
&-12(1-\eta)^2\left(17+6\eta-8\eta^2\right)\left[(1-c)^2S_{2,-2,2}-(1+c)^2S_{2,-2,-2}\right]\\
&+4(1+\eta)^2e(11-6\eta)\left[(1+c)^2S_{3,2,2}-(1-c)^2S_{3,2,-2}\right]\\
&+4(1-\eta)^2e(11+6\eta)\left[(1-c)^2S_{3,-2,2}-(1+c)^2S_{3,-2,-2}\right]\\
&-3(1+\eta)^2e^2\left[(1+c)^2\ S_{4,2,2}-(1-c)^2S_{4,2,-2}\right]\\
&-3(1-\eta)^2e^2\left[(1-c)^2S_{4,-2,2}-(1+c)^2S_{4,-2,-2}\right]\Big\}.
\end{aligned}
\tag{A.5}
$$

We shorten notation calling $S_{i,j,k} \equiv \sin(i\ u + j\ g + k\ h)$.

The Lie transform of generating function W can be applied to any function of Delaunay variables $\mathcal{F} = \sum_i (\epsilon^i/i!)\ F_i(\ell', g', h', L', G', H')$. Since $W_1 = 0$, up to the third-order in the small parameter recurrence (A.1) gives

$$
\mathcal{F} = F_0 + \left(\frac{\epsilon^2}{2}\right)\{F_0; W_2\} + \left(\frac{\epsilon^3}{6}\right)\{F_0; W_3\}. \tag{A.6}
$$

Specifically, this applies to the transformation equations of the Delaunay variables themselves, where $F_0 \in (\ell', g', h', L', G', H')$ and $F_i \equiv 0$ for $i > 0$.

A new application of the recurrence (A.1) to the Hamiltonian $\mathcal{K} = \sum_{0\le i\le 4}\ (\epsilon^i/i!)\ K_{i,0}$, where $K_{i,0} \equiv H_{0,i}$ of (A.3), allows to eliminate the node up to the fourth-order, obtaining the double-averaged Hamiltonian (2.3). Note that $K_{0,4}$ corrects previous results in [22]. The generating function of the transformation is $V = V_1 + \varepsilon V_2 + (\varepsilon^2/2)\ V_3$, where, omitting double

primes,

$$
\begin{aligned}
V_1 &= L\left(\frac{3}{64}\right)\left[\left(4+6e^2\right)s^2\sin 2h + 5\ (1+c)^2e^2\sin(2g+2h) - 5\ (1-c)^2e^2\sin(2g-2h)\right],\\
V_2 &= L\left(-\frac{3}{128}\right)\eta\left[6c\left(2-17e^2\right)s^2\sin 2h + 5\ (2-9c)\ (1+c)^2e^2\sin(2g+2h)\right.\\
&\qquad \left. + 5(1-c)^2(2+9c)e^2\sin(2g-2h)\right],\\
V_3 &= L\left(-\frac{9}{32768}\right)\\
&\quad \times\left\{16s^2\left[456-104c^2-8\left(193+754c^2\right)e^2+\left(47+7831c^2\right)e^4\right]\sin 2h\right.\\
&\qquad + 2s^4\left(232+416e^2-1803e^4\right)\sin 4h\\
&\qquad - 32(1+c)^2e^2\left[2\left(323-285c+780c^2\right)-\left(527-1135c+2125c^2\right)e^2\right]\sin(2g+2h)\\
&\qquad + 32(1-c)^2e^2\left[2\left(323+285c+780c^2\right)-\left(527+1135c+2125c^2\right)e^2\right]\sin(2g-2h)\\
&\qquad + 220s^2e^2\left(4-11e^2\right)\left[(1+c)^2\sin(2g+4h)-(1-c)^2\sin(2g-4h)\right]\\
&\qquad + 4520s^2e^4\left[(1+c)^2\sin(4g+2h)-(1-c)^2\sin(4g-2h)\right]\\
&\qquad \left. - 385e^4\left[(1+c)^4\sin(4g+4h)-(1-c)^4\sin(4g-4h)\right]\right\}.
\end{aligned}
\tag{A.7}
$$

The new Lie transform of generating function V can be applied to any function of Delaunay variables, and, specifically, to the Delaunay variables themselves. For any $\xi'' \in (\ell'', g'', h'', L'', G'', H'')$ the transformation equations of the Lie transform are computed, up to the third-order, from

$$
\xi' = \xi'' + \varepsilon\delta_1 + \left(\frac{\varepsilon^2}{2}\right)\delta_2 + \left(\frac{\varepsilon^3}{6}\right)\delta_3, \tag{A.8}
$$

where

$$
\begin{aligned}
\delta_1 &= \{\xi''; V_1\},\\
\delta_2 &= \{\xi''; V_2\} + \{\delta_1; V_1\},\\
\delta_3 &= \{\xi''; V_3\} + \{\{\xi''; V_2\}; V_1\} + \{\delta_1; V_2\} + \{\delta_2; V_1\}.
\end{aligned}
\tag{A.9}
$$

Acknowledgments

This work was supported from Projects ESP 2007-64068 (the first author) and MTM 2008-03818 (the second author) of the Ministry of Science and Innovation of Spain is

acknowledged. Part of this work has been presented at 20th International Symposium on Space Flight Dynamics, Annapolis, Maryland, USA, September, 24–28 2007.

References

[1] G. W. Hill, "Researches in the lunar theory," *American Journal of Mathematics*, vol. 1, no. 2, pp. 129–147, 1878.

[2] M. L. Lidov, "The evolution of orbits of artifficial satellites of planets under the action of gravitational perturbations of external bodies," *Planetary and Space Science*, vol. 9, no. 10, pp. 719–759, 1962, translated from *Iskusstvennye Sputniki Zemli*, no. 8, p. 5, 1961.

[3] M. Hénon, "Numerical exploration of the restricted problem. VI. Hill's case: non-periodic orbits," *Astronomy and Astrophysics*, vol. 9, pp. 24–36, 1970.

[4] D. P. Hamilton and A. V. Krivov, "Dynamics of distant moons of asteroids," *Icarus*, vol. 128, no. 1, pp. 241–249, 1997.

[5] Y. Kozai, "Motion of a lunar orbiter," *Publications of the Astronomical Society of Japan*, vol. 15, no. 3, pp. 301–312, 1963.

[6] M. L. Lidov and M. V. Yarskaya, "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," *Kosmicheskie Issledovaniya*, vol. 12, pp. 155–170, 1974.

[7] D. J. Scheeres, M. D. Guman, and B. F. Villac, "Stability analysis of planetary satellite orbiters: application to the Europa orbiter," *Journal of Guidance, Control and Dynamics*, vol. 24, no. 4, pp. 778–787, 2001.

[8] M. Lara, J. F. San-Juan, and S. Ferrer, "Secular motion around synchronously orbiting planetary satellites," *Chaos*, vol. 15, no. 4, pp. 1–13, 2005.

[9] M. E. Paskowitz and D. J. Scheeres, "Orbit mechanics about planetary satellites including higher order gravity fields," in *Proceedings of the Space Flight Mechanics Meeting*, Copper Mountain, Colo, USA, January 2005.

[10] Y. Kozai, "Secular perturbations of asteroids with high inclination and eccentricity," *The Astronomical Journal*, vol. 67, no. 9, pp. 591–598, 1962.

[11] R. A. Broucke, "Long-term third-body effects via double averaging," *Journal of Guidance, Control and Dynamics*, vol. 26, no. 1, pp. 27–32, 2003.

[12] M. E. Paskowitz and D. J. Scheeres, "Design of science orbits about planetary satellites: application to Europa," *Journal of Guidance, Control and Dynamics*, vol. 29, no. 5, pp. 1147–1158, 2006.

[13] A. Deprit, "Canonical transformations depending on a small parameter," *Celestial Mechanics*, vol. 1, no. 1, pp. 12–30, 1969.

[14] M. Lara, "Simplified equations for computing science orbits around planetary satellites," *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 1, pp. 172–181, 2008.

[15] A. Deprit and A. Rom, "The main problem of artifficial satellite theory for small and moderate eccentricities," *Celestial Mechanics*, vol. 2, no. 2, pp. 166–206, 1970.

[16] S. Coffey, A. Deprit, and E. Deprit, "Frozen orbits for satellites close to an Earth-like planet," *Celestial Mechanis and Dynamical Astronomy*, vol. 59, no. 1, pp. 37–72, 1994.

[17] M. Lara and D. Scheeres, "Stability bounds for three-dimensional motion close to asteroids," *Journal of the Astronautical Sciences*, vol. 50, no. 4, pp. 389–409, 2002.

[18] A. Deprit, "The elimination of the parallax in satellite theory," *Celestial Mechanics*, vol. 24, no. 2, pp. 111–153, 1981.

[19] S. Ferrer and C. A. Williams, "Simplifications toward integrability of perturbed Keplerian systems," *Annals of the New York Academy of Sciences*, vol. 536, pp. 127–139, 1988.

[20] J. F. Palacián, "Dynamics of a satellite orbiting a planet with an inhomogeneous gravitational field," *Celestial Mechanis and Dynamical Astronomy*, vol. 98, no. 4, pp. 219–249, 2007.

[21] A. Deprit, "Delaunay normalisations," *Celestial Mechanics*, vol. 26, no. 1, pp. 9–21, 1982.

[22] J. F. San-Juan and M. Lara, "Normalizaciones de orden alto en el problema de Hill," *Monografías de la Real Academia de Ciencias de Zaragoza*, vol. 28, pp. 23–32, 2006.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 396267, 14 pages
doi:10.1155/2009/396267

Research Article

Collision and Stable Regions around Bodies with Simple Geometric Shape

A. A. Silva,[1,2,3] O. C. Winter,[1,2] and A. F. B. A. Prado[1,2]

[1] *Space Mechanics and Control Division (DMC), National Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil*
[2] *UNESP, Universidade Estadual Paulista, Grupo de Dinâmica Orbital & Planetologia, Av. Ariberto Pereira da Cunha, 333-Guaratinguetá 12516-410, Brazil*
[3] *Universidade do Vale do Paraíba (CT I / UNIVAP), São José dos Campos 12245-020, Brazil*

Correspondence should be addressed to A. A. Silva, aurea_as@yahoo.com.br

Received 29 July 2009; Accepted 20 October 2009

Recommended by Silvia Maria Giuliatti Winter

We show the expressions of the gravitational potential of homogeneous bodies with well-defined simple geometric shapes to study the phase space of trajectories around these bodies. The potentials of the rectangular and triangular plates are presented. With these expressions we study the phase space of trajectories of a point of mass around the plates, using the Poincaré surface of section technique. We determined the location and the size of the stable and collision regions in the phase space, and the identification of some resonances. This work is the first and an important step for others studies, considering 3D bodies. The study of the behavior of a point of mass orbiting around these plates (2D), near their corners, can be used as a parameter to understand the influence of the gravitational potential when the particle is close to an irregular surface, such as large craters and ridges.

1. Introduction

The aim of this paper is to study the phase space of trajectories around some homogeneous bodies with well-defined simple geometric shapes. Closed-form expressions derived for the gravitational potential of the rectangular and triangular plates were obtained from Kellogg [1] and Broucke [2]. They show the presence of two kinds of terms: logarithms and arc tangents. With these expressions we study the phase space of trajectories of a particle around two different bodies: a square and a triangular plates. The present study was made using the Poincaré surface of section technique which allows us to determine the location and size of the stable and chaotic regions in the phase space. We can find the periodic, quasiperiodic and chaotic orbits.

Some researches on this topic can be found in Winter [3] that study the stability evolution of a family of simply periodic orbits around the Moon in the rotating Earth-Moon-particle system. He uses the numerical technique of Poincaré surface of section to obtain the structure of the region of the phase space that contains such orbits. In such work it is introduced a criterion for the degree of stability. The results are a group of surfaces of section for different values of the Jacobi constant and the location and width of the maximum amplitude of oscillation as a function of the Jacobi constant. Another research was done by Broucke [4] that presents the Newton's law of gravity applied to round bodies, mainly spheres and shells. He also treats circular cylinders and disks with the same methods used for shells and it works very well, almost with no modifications. The results are complete derivations for the potential and the force for the interior case as well as the exterior case.

In Sections 2 and 3, following the works of Kellogg [1] and Broucke [2], we show the expressions for the potential of the rectangular and triangular plates, respectively. In Section 4 we use the Poincaré surface of section technique to study the phase space around the plates. In Section 5 we show the size and location of stable and collision regions in the phase space. In the last section, we have some final comments.

2. The Potential of the Rectangular Plate

Let us consider a homogeneous plane rectangular plate and an arbitrary point $P\,(0,0,Z)$, not on the rectangle. Take x and y axes parallel to the sides of the rectangle, and their corners referred to these axes are $A(b,c)$, $B(b',c)$, $C(b',c')$, and $D(b,c')$. Let ΔS_k denote a typical element of the surface, containing a point Q_k located in the rectangular plate with coordinates (x_k,y_k). See Figure 1.

The potential of the rectangular plate can be given by the expression

$$\begin{aligned} U &= \frac{G\sigma\Delta S_k}{r_k} = G\sigma\int_c^{c'}\int_b^{b'}\frac{dxdy}{\sqrt{x^2+y^2+Z^2}} \\ &= G\sigma\left\{\int_c^{c'}\ln\left(b'+\sqrt{b'^2+y^2+Z^2}\right)dy-\int_c^{c'}\ln\left(b+\sqrt{b^2+y^2+Z^2}\right)dy\right\}, \end{aligned} \tag{2.1}$$

where G is the Newton's gravitational constant, σ is the density of the material, and r_k is the distance between the particle and the point Q_k.

In evaluating the integrals we find

$$\begin{aligned} U = G\sigma\Bigg[&c'\ln\frac{(b'+d_3)}{(b+d_4)}+c\ln\frac{(b+d_1)}{(b'+d_2)}+b'\ln\frac{(c'+d_3)}{(c+d_2)}+b\ln\frac{(c+d_1)}{(c'+d_4)} \\ &+Z\left(\tan^{-1}\frac{b\cdot c'}{Z\cdot d_4}-\tan^{-1}\frac{b'\cdot c'}{Z\cdot d_3}+\tan^{-1}\frac{b'\cdot c}{Z\cdot d_2}-\tan^{-1}\frac{b\cdot c}{Z\cdot d_1}\right)\Bigg], \end{aligned} \tag{2.2}$$

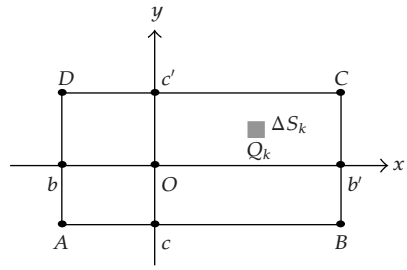


Figure 1: Rectangular plate is on the plane (x, y).

where

$$\begin{aligned} d_1^2 &= b^2 + c^2 + Z^2, \\ d_2^2 &= b'^2 + c^2 + Z^2, \\ d_3^2 &= b'^2 + c'^2 + Z^2, \\ d_4^2 &= b^2 + c'^2 + Z^2 \end{aligned} \tag{2.3}$$

are the distances from $P\ (0, 0, Z)$ to the corners A, B, C, and D, respectively.

3. The Potential of the Triangular Plate

We will give the potential at a point $P\ (0, 0, Z)$ on the Z-axis created by the triangle shown in Figure 2 located in the xy-plane. The side $\overline{P_1P_2}$ is parallel to the x-axis. The coordinates of P_1 and P_2 are $\vec{r}_1(x_1, y_1)$ and $\vec{r}_2(x_2, y_2)$, but we have that $y_1 = y_2$ and $x_1 > x_2 > 0$. The distances are given by $d_1^2 = x_1^2 + y_1^2 + Z^2$ and $d_2^2 = x_2^2 + y_1^2 + Z^2$, where d_1 is the distance from $P\ (0, 0, Z)$ to the corner $P_1(x_1, y_1)$ and d_2 is the distance from $P\ (0, 0, Z)$ to the corner $P_2(x_2, y_1)$. Using the definition of the potential, we have that the potential at $P\ (0, 0, Z)$ can be given by

$$U = G\sigma\left\{ y_1 \ln\left[\frac{x_1 + d_1}{x_2 + d_2}\right] + Z\tan^{-1}\left(\frac{\beta_1 Z}{d_1}\right) - Z\tan^{-1}\left(\frac{\beta_2 Z}{d_2}\right) - |Z|\alpha_{12}\right\}, \tag{3.1}$$

where $\beta_1 = x_1/y_1$, $\beta_2 = x_2/y_2$, and $\alpha_{12} = \alpha_1 - \alpha_2$ represent the angle of the triangle at the origin O and it is showed in Figure 2.

The potential of this triangle at the point $P\ (0, 0, Z)$ on the Z-axis must be invariant under an arbitrary rotation of the triangle around the same Z-axis. Therefore, (3.1) should also be invariant under this rotation and their four terms are individually invariant, where they can be expressed in terms of invariant quantities, such as the sides and the angle of the triangle. The potential of an arbitrary triangular plate $P_1P_2P_3$ (Figure 3) can be obtained by

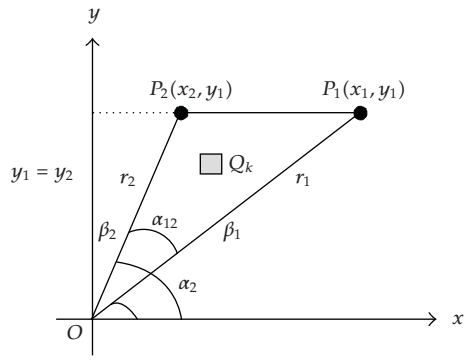


Figure 2: Triangle $\overline{OP_1P_2}$ located in the xy-plane (x, y).

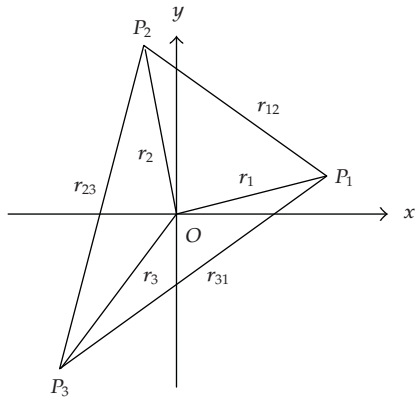


Figure 3: Triangular plate $\overline{P_1P_2P_3}$.

the sum of three special triangles of the type used in this section. So, the result will have three logarithmic terms and three arc-tangent terms. So, the potential is given by

$$U = G\sigma\left\{\frac{C_{12}}{r_{12}}\ln\left[\frac{d_1+d_2+r_{12}}{d_1+d_2-r_{12}}\right] + \frac{C_{23}}{r_{23}}\ln\left[\frac{d_2+d_3+r_{23}}{d_2+d_3-r_{23}}\right] + \frac{C_{31}}{r_{31}}\ln\left[\frac{d_3+d_1+r_{31}}{d_3+d_1-r_{31}}\right] + \tan^{-1}\left(\frac{Nd_1}{D_1}\right) + \tan^{-1}\left(\frac{Nd_2}{D_2}\right) + \tan^{-1}\left(\frac{Nd_3}{D_3}\right) + \text{Sign}(Z)\pi\right\}, \tag{3.2}$$

where the numerator and the denominators are

$$\begin{aligned} N &= -Z(C_{12}+C_{23}+C_{31}),\\ D_1 &= Z^2\left(r_1^2 + D_{23} - D_{31} - D_{12}\right) - C_{12}C_{31},\\ D_2 &= Z^2\left(r_2^2 + D_{31} - D_{12} - D_{23}\right) - C_{23}C_{12},\\ D_3 &= Z^2\left(r_3^2 + D_{12} - D_{23} - D_{31}\right) - C_{31}C_{23}. \end{aligned} \tag{3.3}$$

The symbol Sign(Z) in (3.2) is the sign of the variable Z and r_{ij} $(i, j = 1, 2, 3)$ is the distance between P_i and P_j. In particular, the dot-product $D_{ij} = r_i r_j \cos \alpha_{ij}$ and the cross-product $C_{ij} = r_i r_j \sin \alpha_{ij}$ of the vectors $\mathbf{r}_i$ and $\mathbf{r}_j$ are two invariants, where $\alpha_{ij} = \alpha_i - \alpha_j$. In order to obtain more details see Broucke [2]. This result is invariant with respect to an arbitrary rotation around the Z-axis.

An important generalization of (2.2) and (3.2) can be done, for the case where the point P has arbitrary coordinates (X, Y, Z) instead of being on the Z-axis. For the rectangle, this task is rather easy, because the four vertices A, B, C, and D have completely arbitrary locations. In order to have more general results it is sufficient to replace (b, b', c, c') by $(b - X,\ b' - X,\ c - Y,\ c' - Y)$ in (2.2). For the triangle, it is sufficient to replace all the vertex-coordinates (x_i, y_i) by $(x_i - X, y_i - Y)$ in (3.2). These generalizations give us expressions for the potential U of the rectangle and the triangle as a function of the variables (X, Y, Z). Then, it is possible to compute the components of the acceleration by taking the gradient of the potential. It is important to note that, in taking the partial derivatives of $U(X, Y, Z)$, the arguments of the logarithms and the arc tangents were treated as constants by Broucke [2]. This simplifies the work considerably. The general expressions of the acceleration allow us to study orbits around these plates, and this study is very important to obtain knowledge that will be necessary to study the cases of three-dimensional solids, such as polyhedra.

4. Study of the Phase Space around Simple Geometric Shape Bodies

With the potential determined, we study the phase space of trajectories of a point of mass around the plates. In this section we use the Poincaré surface of section technique to study the regions around the rectangular and triangular plates. In this part of the paper we explore the space of initial conditions. The results are presented in Poincaré sections $(x, \dot{x})$, from which one can identify the nature of the trajectories: periodic, quasiperiodic, or chaotic orbits. We can find the collision regions and identify some resonances.

In order to obtain the orbital elements of a particle at any instant it is necessary to know its position (x, y) and velocity $(\dot{x}, \dot{y})$ that corresponds to a point in a four dimensional phase space. The conservation of the total energy of the system implies in the existence of a three-dimensional surface in this phase space. For a fixed value of the total energy only three of the four quantities are needed, for example, x, y, and $\dot{x}$, since the other one $\dot{y}$ is determined, up to the sign, by the total energy. By defining a plane, that we choose $y = 0$, in the resulting three-dimensional space, the values of x and $\dot{x}$ can be plotted every time the particle has $y = 0$. The ambiguity of the sign of $\dot{y}$ is removed by considering only those crossing with a fixed sign of $\dot{y}$. The section is obtained by fixing a plane in the phase space and plotting the points when the trajectory intersects this plane in a particular direction (Winter and Vieira Neto, [5]). This technique is used to determine the regular or chaotic nature of the trajectory. In the Poincaré map, if there are closed well-defined curves then the trajectory is quasiperiodic. If there are isolated single points inside such islands, the trajectory is periodic. Any "fuzzy" distribution of points in the surface of section implies that the trajectory is chaotic (Winter and Murray [6]).

In order to study the regions around simple geometric shape bodies, two cases were considered. In the first case we used a square plate, where all of its sides are $\ell = 2$ and the initial position of the test particle is chosen arbitrarily in the right side of the plate. The motion of the particle is in the counterclockwise direction. In the second case we used an equilateral triangular plate, with sides $\ell = 1$ and one of its sides is parallel to the y-axis. For the triangle, the initial position of the test particle is in the right side of the plate and its motion is in the

counterclockwise direction. Both of the cases consider a baricentric system, where the plates are centered in the origin of the system. The constants used are the Newton's gravitational constant and the density of the material, respectively, $G = \sigma = 1$.

In this paper, the Poincaré surface of section describes a region of the phase space for two-body problem (plate-particle). The numerical study makes use of the Burlisch-Stoer method of integration with an accuracy of $O(10^{-12})$. The Newton-Raphson method was used to determine when the trajectory crosses the plane $y = 0$ with an accuracy of $O(10^{-11})$. There is always a choice of surfaces of section and, in our work, the values of x and $\dot{x}$ were computed whenever the trajectory crossed the plane $y = 0$ with $\dot{y} > 0$.

In this section, Figure 4 represents a set of Poincaré surfaces of section around the triangular plate with values of energy, E, varying from -0.41 to -0.46, at intervals of 0.01. We considered approximately seventy starting conditions for each surface of section, but most of them generated just a single point each one. In that case the particle collided with the plate.

Figures 4(a) to 4(f) show regions where the initial conditions generated well-defined curves. These sections represent quasiperiodic trajectories around the triangular plate. The center of the islands for each surface of section corresponds to one periodic orbit of the first kind (circular orbit). In Figure 4(a), for example, we can see a group of four islands corresponding to a single trajectory with $E = -0.41$ and $x_0 = -0.8422$. The center of each one of those islands corresponds to one periodic orbit of the second kind (resonant orbit). In the same figure there is a separatrix that represents the trajectory with $E = -0.41$ and $x_0 = -0.83176$ where the particle has a chaotic movement confined in a small region. Figure 4(d) shows that the group of four islands is disappearing and a new group of five islands appears. This new group of islands corresponds to a trajectory with $E = -0.44$ and $x_0 = -0.80281$. The region of single points outside the islands corresponds to a chaotic "sea" where the particle collides with the plate. This is a region highly unstable.

Figure 5(a) shows a Poincaré surface of section around the triangular plate for energy, $E = -0.4$. With a zoom in the region 1 we can see a fuzzy but confined distribution of points that corresponds to the separatrix (Figure 5(b)) generated with $x_0 = -0.80816$. With a zoom in the region 2 we can see islands (Figure 5(c)) that were generated with $x_0 = -0.81596$ and $x_0 = -0.83351$.

A similar structure of the phase space is found for the square plate when analyzed its Poincaré surface of section.

5. Collision and Stable Regions

In this section we show a global vision of the location and size of the stable and collision regions around the triangular and square plates, using Poincaré map. The values of x, when $\dot{x} = 0$, are measured with the largest island of stability (quasiperiodic orbit) for each value of energy (Winter [3]). With the diagrams (Figures 6 and 7) we can see the evolution of the stability for the family of periodic orbits for different values of energy constant.

The surfaces of section are plotted at intervals of 0.02 for the energy constant (E). For the triangular plate, the energy varies from -0.01 to -0.47 and, for the square plate, from -0.13 to -1.09. The values of the energy ($E = -0.47$) for the triangle and ($E = -1.09$) for the square represent the region that there is no island of stability within the adopted precision.

Figure 6(a) shows the stable and collision regions around the triangular plate. The white color represents the region that is considered stable, neglecting the small chaotic regions that appear in the separatrices. The collision region is represented by the dark gray color and it indicates the region where the particle collides with the plate due to its

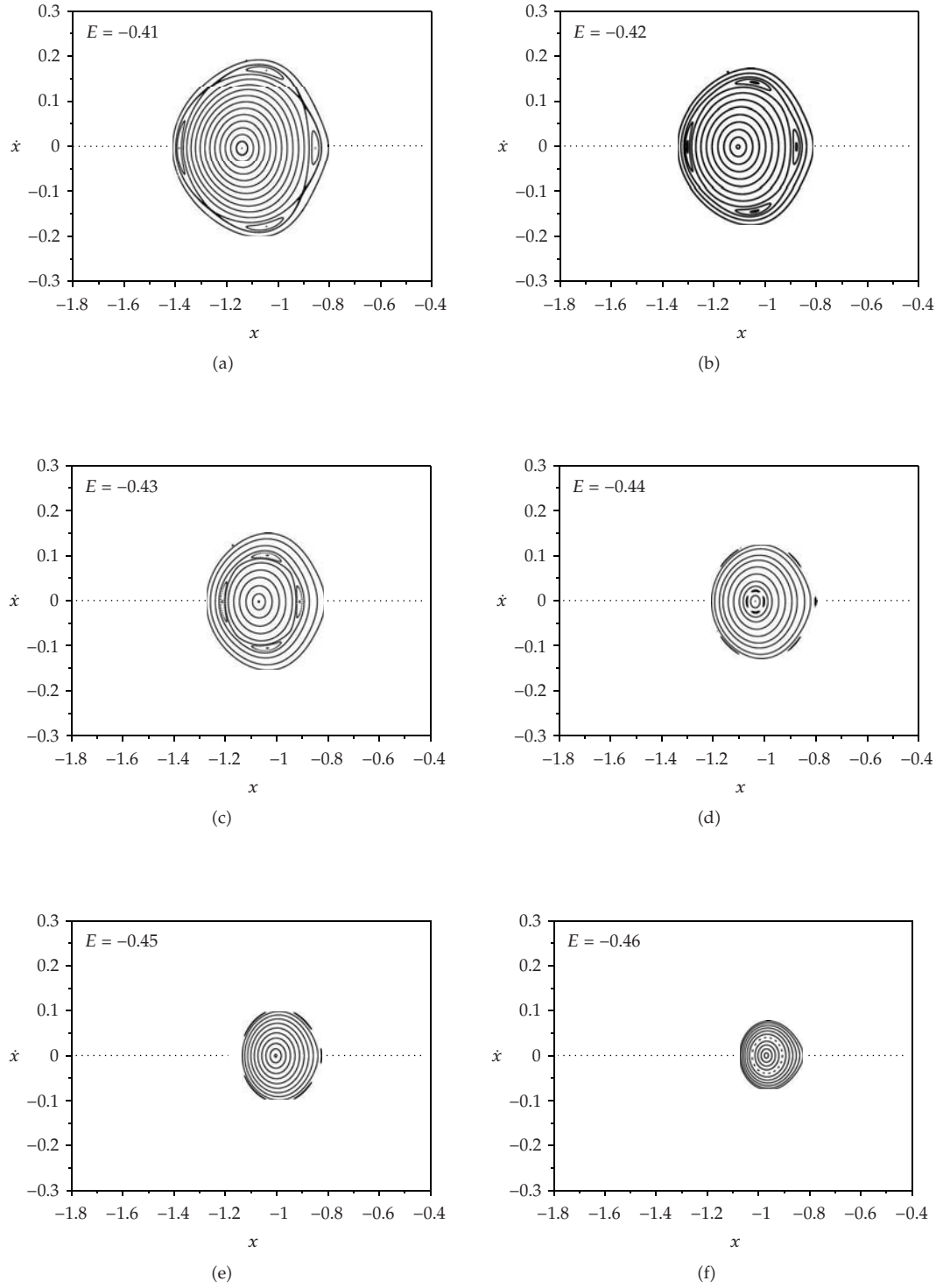


Figure 4: A group of Poincaré surfaces of section of trajectory around the triangular plate for different values of energy, $-0.46 \leq E \leq -0.41$ at intervals of 0.01.

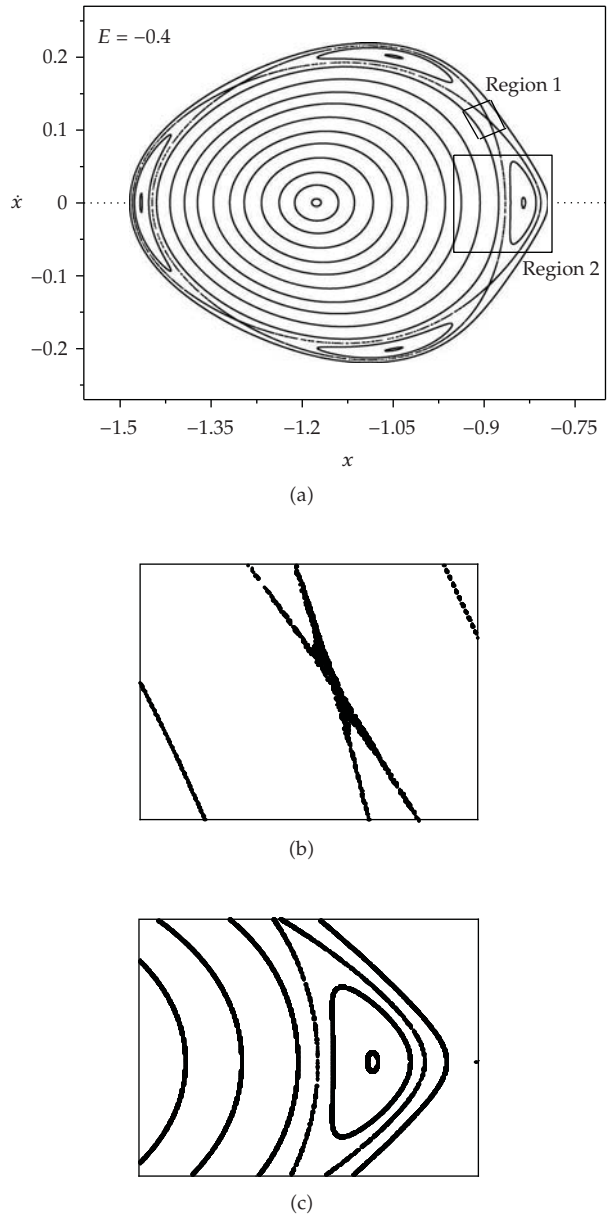


Figure 5: (a) Poincaré surfaces of section around the triangular plate for energy, $E = -0.4$; (b) a zoom in region 1 that represents the separatrix; (c) a zoom in region 2 that represents a group of islands.

gravitational attraction. The light gray color corresponds to a "prohibited" region where there are no starting conditions for those values of energy. The triangular plate is located to the right of the limit of the collision border. Figure 6(b) corresponds to a zoom of the previous one in order to have the best visualization of the regions.

Figure 7(a) summarizes the same studies around the square plate and the color codes are the same as mentioned for the triangle. The square plate is located to the left of the limit of the collision border. Figure 7(b) corresponds to a zoom of the previous one.

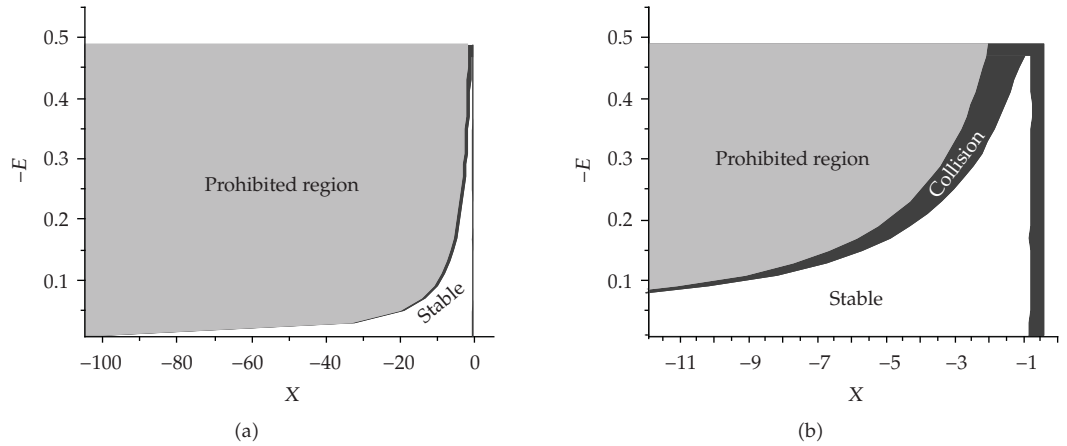


Figure 6: (a) Study of the regions around the triangular plate: stable area (white), collision area (dark gray), and "prohibited"area (light gray); (b) zoom of a region in (a).

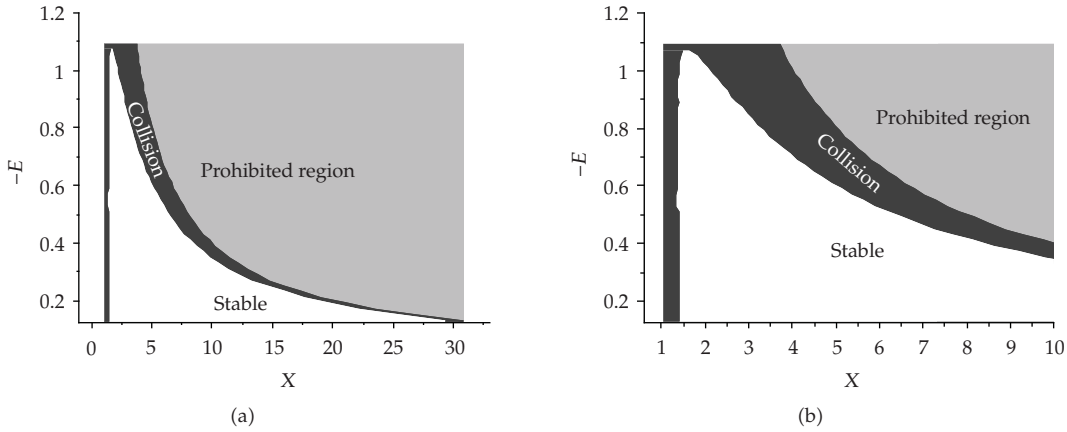


Figure 7: (a) Study of the regions around of the square plate: stable area (white), collision area (dark gray), and "prohibited" area (light gray); (b) zoom of a region in (a).

6. Trajectories around Triangular and Square Plates

In the last section was shown a global vision of a family of orbits around the geometric plates. In this section we will show some trajectories of a particle around triangular and square plates. The initial conditions (position and velocity) of the particle that orbit the plates are given by:

$$\begin{aligned}(x, y, z) &= (x_0, 0, 0),\\ (Vx, Vy, Vz) &= \left(0, -\sqrt{2(E+U)}, 0\right),\end{aligned} \tag{6.1}$$

where E is the total energy of the system and U is the potential energy generated by the plate.

Table 1: Value of the energy (E) and the initial position of the particle (x_0) for the triangular and square plates.

Figures	E	x_0
8	−0.40	−0.83
9	−0.41	−0.83176
10	−0.60	1.4

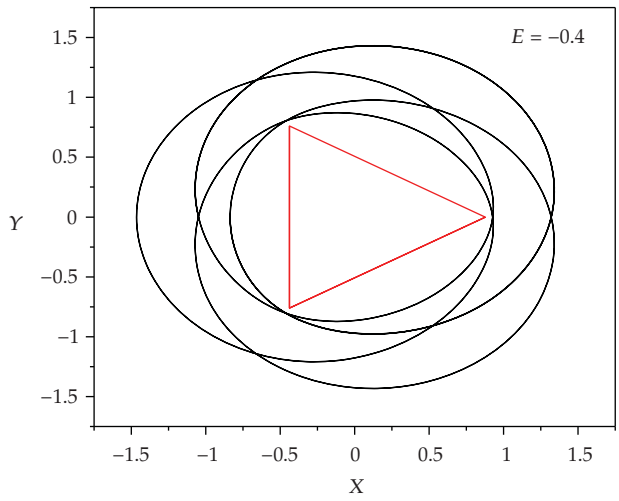


Figure 8: Trajectory around the triangular plate: quasiperiodic.

Table 1 shows the value of the initial position of the particle (x_0) and its correspondent value of the energy for each trajectory around the triangular and square plates. Figures 8 and 9 show a quasiperiodic orbit and a chaotic orbit (in a region of the separatrix), respectively, around the triangular plate. Figure 10 shows a quasiperiodic orbit around the square plate.

The trajectories in these figures show that the effect of the potential due to the plates can be compared with the effect of the Earth's flattening (J_2), generating trajectories as a precessing elliptic orbit.

In this section we will explore the regions very close to the vertex of the geometric plates and verify the behavior of the particle in this situation due to the potentials of the plates. So, it is showed an analysis of the semimajor axis and the eccentricity of the orbit.

In the next example is considered a triangular plate as a central body, the value of the energy is $E = -0.1$ and the initial position for the particle is $x_0 = -1.14$ and $y_0 = z_0 = 0$, that generate a quasiperiodic orbit. Figure 11 shows the trajectory around the plate, where A, B, and C are the vertexes of the triangular plate and Figure 12 shows the behavior of the semimajor axis versus time. The points in Figure 11 correspond to the enumerated regions in Figure 12. For better visualization of Figure 11, there is a zoom of the region around the plate. Figure 13 shows a zoom of Figure 12 with the details of the semimajor axis peaks along six orbital periods. The results show the following features.

The codes 1a, 2a, 3a, 4a, 5a, and 6a are the maximum values of the semimajor axis that occur when the particle is close to the vertex A during six complete orbits. The codes 1 to 6 correspond to the six close approaches of the particle to the vertex A. The results show that

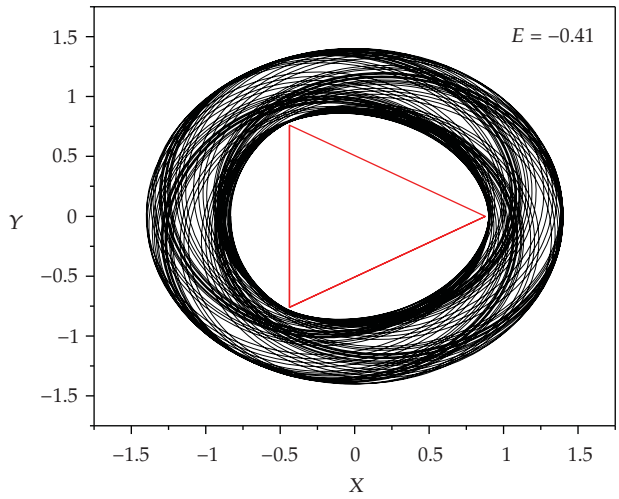


Figure 9: Trajectory around the triangular plate: chaotic orbit (separatrix).

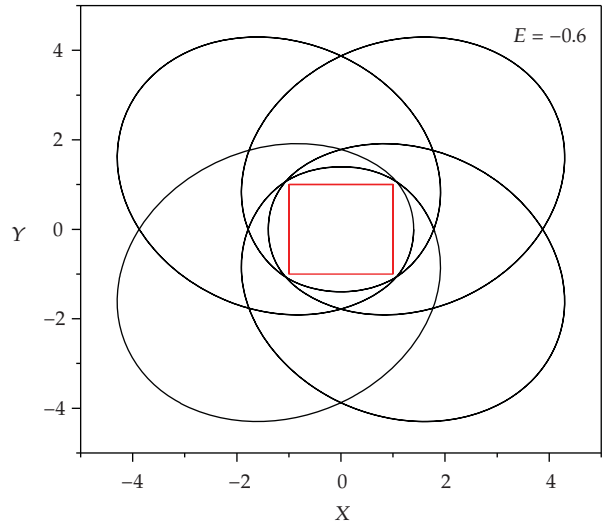


Figure 10: Trajectory around the square plate: quasiperiodic.

the semimajor axis increases as the particle approaches to the vertex A. From Figures 11 to 13 we see the same codes for passages of the particle close to the other two vertexes, B and C. Since the trajectory of the particle does not get so close to the vertexes, B and C, then the increase on the semimajor axis is not so large in these cases.

With these analyses we can verify that the proximity of the particle to the corners of the plate changes the behavior of the trajectory. In the studied cases, the orbits become eccentric and precess. For the eccentricity we could verify a similar behavior as that found for the semimajor axis.

The studies with the square plate show the same behavior as obtained for the triangle.

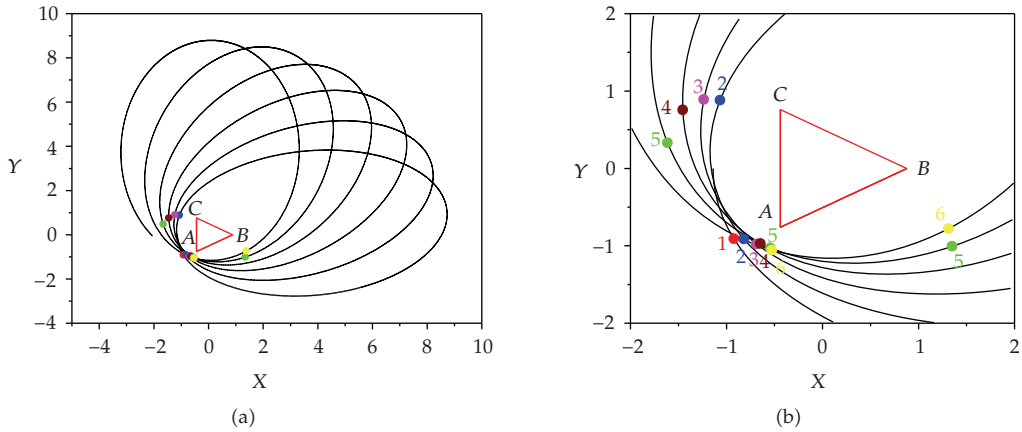


Figure 11: (a) Trajectory of the particle around the triangular plate. (b) Zoom of the trajectory near the plate.

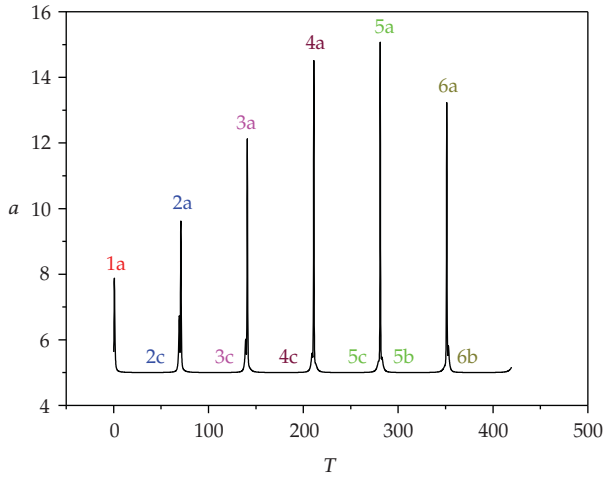


Figure 12: Semimajor axis of the orbit.

7. Final Comments

In this paper we used closed form solutions for the gravitational potential for two simple geometric shape bodies in order to study the behavior of a particle around each one of them. The development was applied to the square and to the triangular plates. We used the Poincaré surface of section technique to study the phase space of trajectories of a particle around those plates. We identified different kinds of orbits: periodic, quasiperiodic, and chaotic orbits. We found a collision region and identified some resonances.

The results showed that there is a region of starting conditions where it is possible to have orbits around the plates before the collision. The location and size of the stable and collision regions were measured for each value of energy and it showed us the evolution of the stable regions in the phase space. We have seen that the Poincaré surface of section technique is an easy and powerful way to identify the nature of the orbits.

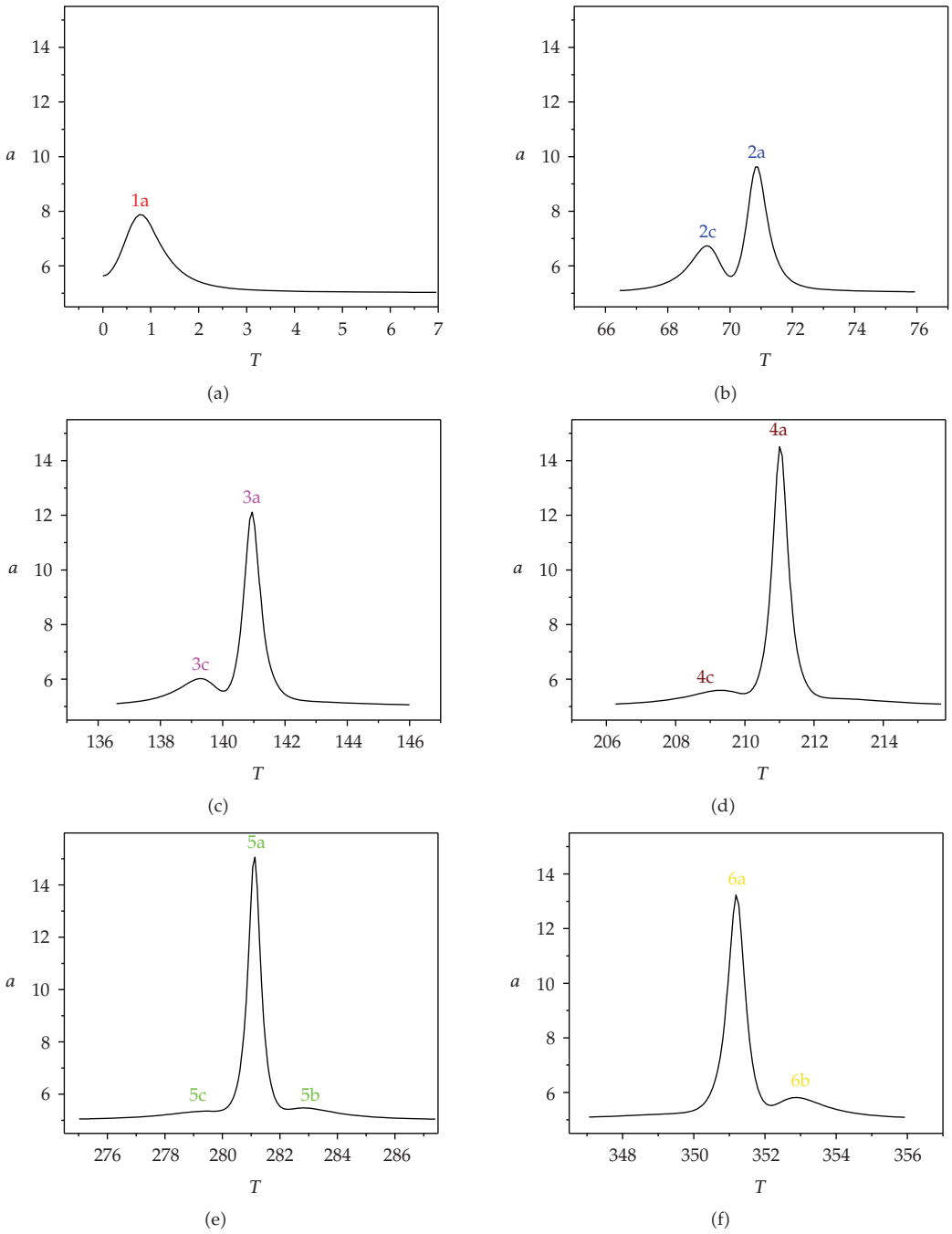


Figure 13: Zoom of Figure 12. Peak of semimajor axis in six orbital periods.

It was observed that the corners of the plates have a significant influence in the behavior of the trajectory, mainly when the particle has a close approach to them. The orbits become eccentric and precess due their gravitational field.

This study opens the way to obtain the potential and the trajectories around three-dimensional bodies with irregular shapes, such as asteroids and comets.

Acknowledgment

The authors wish to express their appreciation for the supports provided by the Brazilian agencies: FAPESP, CAPES, and CNPq. They are very grateful to Roger *A*. Broucke for the information and discussions.

References

[1] O. D. Kellogg, *Foundations of Potential Theory*, Dover, New York, NY, USA, 1929.
[2] R. A. Broucke, "Closed form expressions for some gravitational potentials: triangle, rectangle, pyramid and polyhedron," in *Proceedings of AAS/AIAA Spaceflight Mechanics Meeting*, Albuquerque, NM, USA, February 1995.
[3] O. C. Winter, "The stability evolution of a family of simply periodic lunar orbits," *Planetary and Space Science*, vol. 48, no. 1, pp. 23–28, 2000.
[4] R. A. Broucke, "The gravitational attraction of spheres and spherical shells," in *Advances in Space Dynamics*, O. C. Winter and A .F. B. A. Prado, Eds., vol. 2, pp. 190–233, INPE, São José dos Campo, Brazil, 2002.
[5] O. C. Winter and E. Vieira Neto, "Distant stable direct orbits around the Moon," *Astronomy and Astrophysics*, vol. 393, no. 2, pp. 661–671, 2002.
[6] O. C. Winter and C. D. Murray, *Atlas of the Planar, Circular, Restricted Three-Body Problem. I. Internal Orbit*, vol. 16 of *QMW Maths Notes*, Queen Mary and Westfield College, London, UK, 1994.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 181360, 23 pages
doi:10.1155/2009/181360

Research Article

Dynamical Aspects of an Equilateral Restricted Four-Body Problem

Martha Álvarez-Ramírez[1] and Claudio Vidal[2]

[1] *Departamento de Matemáticas, UAM-Iztapalapa, A.P. 55-534, 09340 Iztapalapa, México, Mexico*
[2] *Departamento de Matemáticas, Facultad de Ciencias, Universidad del Bío Bío, Casilla 5-C, Concepción, VIII-Región 4081112, Chile*

Correspondence should be addressed to Claudio Vidal, clvidal@ubiobio.cl

Received 26 July 2009; Revised 10 November 2009; Accepted 9 December 2009

Recommended by Tadashi Yokoyama

The spatial equilateral restricted four-body problem (ERFBP) is a four body problem where a mass point of negligible mass is moving under the Newtonian gravitational attraction of three positive masses (called the primaries) which move on circular periodic orbits around their center of mass fixed at the origin of the coordinate system such that their configuration is always an equilateral triangle. Since fourth mass is small, it does not affect the motion of the three primaries. In our model we assume that the two masses of the primaries m_2 and m_3 are equal to μ and the mass m_1 is $1 - 2\mu$. The Hamiltonian function that governs the motion of the fourth mass is derived and it has three degrees of freedom depending periodically on time. Using a synodical system, we fixed the primaries in order to eliminate the time dependence. Similarly to the circular restricted three-body problem, we obtain a first integral of motion. With the help of the Hamiltonian structure, we characterize the region of the possible motions and the surface of fixed level in the spatial as well as in the planar case. Among other things, we verify that the number of equilibrium solutions depends upon the masses, also we show the existence of periodic solutions by different methods in the planar case.

1. Introduction

Dynamical systems with few bodies (three) have been extensively studied in the past, and various models have been proposed for research aiming to approximate the behavior of real celestial systems. There are many reasons for studying the four-body problem besides the historical ones, since it is known that approximately two-thirds of the stars in our Galaxy exist as part of multistellar systems. Around one-fifth of these is a part of triple systems, while a rough estimate suggests that a further one-fifth of these triples belongs to quadruple or higher systems, which can be modeled by the four-body problem. Among these models, the configuration used by Maranhão [1] and Maranhão and Llibre [2], where three point masses form at any time a collinear central configuration (Euler configuration, see [3]), is

of particular interest not only for its simplicity but mainly because in the last 10 years, an increasing number of extrasolar systems have been detected, most of them consisting of a "sun" and a planet or of a "sun" and two planets.

We study the motion of a mass point of negligible mass under the Newtonian gravitational attraction of three mass points of masses m_1, m_2, and m_3 (called primaries) moving in circular periodic orbits around their center of mass fixed at the origin of the coordinate system. At any instant of time, the primaries form an equilateral equilibrium configuration of the three-body problem which is a particular solution of the three-body problem given by Lagrange (see [4] or [3]). Two of these primaries have equal masses and are located symmetrically with respect to the third primary.

We choose the unity of mass in such a way that $m_1 = 1 - 2\mu$ and $m_2 = m_3 = \mu$ are the masses of the primaries, where $\mu \in (0, 1/2)$. Units of length and time are chosen in such a way that the distance between the primaries is one.

For studying the position of the infinitesimal mass, m_4, in the plane of motion of the primaries, we use either the sideral system of coordinates, or the synodical system of coordinates (see [5] for details). In the synodical coordinates, the three point masses m_1, m_2, and m_3 are fixed at $(\sqrt{3}\mu, 0, 0)$, $(-(\sqrt{3}/2)(1-2\mu), 1/2, 0)$, and $(-(\sqrt{3}/2)(1-2\mu), -1/2, 0)$, respectively. In this paper, *the equilateral restricted four-body problem* (shortly, ERFBP) consists in describing the motion of the infinitesimal mass, m_4, under the gravitational attraction of the three primaries m_1, m_2, and m_3. Maranhão's PhD thesis [1] and the paper [2] by Maranhão and Llibre studied a restricted four body problem, where three primaries rotating in a fixed circular orbit define a collinear central configuration.

In the ERFBP, the equations of motion of m_4 in synodical coordinates (x, y, z) are

$$\begin{aligned} \ddot{x} - 2\dot{y} &= \Omega_x, \\ \ddot{y} + 2\dot{x} &= \Omega_y, \\ \ddot{z} &= \Omega_z, \end{aligned} \tag{1.1}$$

where

$$\begin{aligned} \Omega = \Omega(x, y, z) &= \frac{1}{2}\left(x^2 + y^2\right) + \frac{1-2\mu}{\rho_1} + \frac{\mu}{\rho_2} + \frac{\mu}{\rho_3}, \\ \rho_1 &= \sqrt{\left(x - \sqrt{3}\mu\right)^2 + y^2 + z^2}, \quad \rho_2 \\ &= \sqrt{\left(x + \frac{\sqrt{3}}{2}(1-2\mu)\right)^2 + \left(y - \frac{1}{2}\right)^2 + z^2}, \\ \rho_3 &= \sqrt{\left(x + \frac{\sqrt{3}}{2}(1-2\mu)\right)^2 + \left(y + \frac{1}{2}\right)^2 + z^2}. \end{aligned} \tag{1.2}$$

We remark that the ERFBP becomes *the central force problem* when $\mu = 0$, and $m_1 = 1$ is situated in the origin of the system, while $\mu = 1/2$ results in *the restricted three-body problem* with the bodies m_2 and m_3 of mass $1/2$.

Our paper is organized as follows: Section 2 is devoted to describing the most important dynamical phenomena that governs the evolution of asteroid movement and states the problem under consideration in the present study. In Section 3 reductions of the problem are discussed and a comprehensive treatment of streamline analogies is given. Section 4 is devoted to the principal qualitative aspect of the restricted problem—the surfaces and curves of zero velocity, several uses of which are discussed. The regions of allowed motion and the location and properties of the equilibrium points are established. We describe the Hill region. The description of the number of equilibrium points is given in Section 5, and in the symmetrical case (i.e., $\mu = 1/3$), we describe the kind of stability of each equilibrium. In Section 6, the planar case is considered. There, we prove the existence of periodic solutions as a continuation of periodic Keplerian orbits, and also when the parameter μ is small and when it is close to $1/2$. Finally, in Section 8 we present the conclusions of the present work.

Next, we will enunciate some four-body problem that has been considered in the literature. Cronin et al. in [6, 7] considered the models of four bodies where two massive bodies move in circular orbits about their center of mass or barycenter. In addition, this barycenter moves in a circular orbit about the center of mass of a system consisting of these two bodies and a third massive body. It is assumed that this third body lies in the same plane as the orbits of the first two bodies. The authors studied the motion of a fourth body of small mass which moves under the combined attractions of these three massive bodies. This model is called bicircular four-body problem. Considering this restricted four-body problem consisting of Earth, Moon, Sun, and a massless particle, this problem can be used as a model for the motion of a space vehicle in the Sun-Earth-Moon system. Several other authors have considered the study of this problem, for example, [8–11] and references therein. The quasi-bicircular problem is a restricted four body problem where three masses, Earth-Moon-Sun, are revolving in a quasi-bicircular motion (i.e. a coherent motion close to bicircular) also has been studied, see [12] and references therein. The restricted four-body problem with radiation pressure was considered in [13], while the photogravitational restricted four body problem was considered in [14].

2. Statement of the Problem

It is known that equilateral configurations of three-bodies with arbitrary masses m_1, m_2, and m_3 on the same plane, moving with the same angular velocity, form a relative equilibrium solution of the three-body problem (see e.g., [4] or [3]). More precisely, we consider three particles of masses m_1, m_2, and m_3 (called primaries) each describing, at any instant, a circle around their center of masses (which is fixed at the origin), with the same angular velocity ω and such that its configuration at any instant is an equilateral triangle (see Figure 1). Now, we consider an infinitesimal particle m_4 attracted by the primaries m_1, m_2, and m_3 according to Newton's gravitational law. Let $\mathbf{r}$ be the position vector of m_4.

The equations of motion can be written as

$$\mathbf{r}'' = \nabla U, \tag{2.1}$$

where $(\,)'$ denotes derivative with respect to t and

$$U = U(\mathbf{r}; t, m_1, m_2, m_3) = \frac{m_1}{\|\mathbf{r} - \mathbf{r}_1(t)\|} + \frac{m_2}{\|\mathbf{r} - \mathbf{r}_2(t)\|} + \frac{m_3}{\|\mathbf{r} - \mathbf{r}_3(t)\|}, \tag{2.2}$$

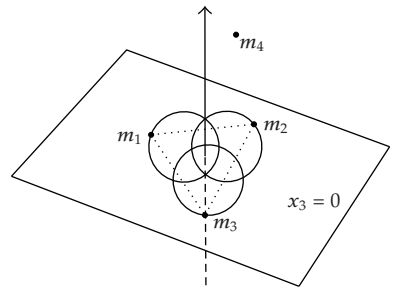


Figure 1: The equilateral restricted four body problem in inertial coordinates.

with $\mathbf{r}_1(t), \mathbf{r}_2(t)$, and $\mathbf{r}_3(t)$ representing the position of each primary, respectively. To remove the time dependence of the system (2.1), we consider the orthonormal moving frame in $\mathbb{R}^3$, given by $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where

$$\mathbf{e}_1 = \mathbf{e}_1(t) = e^{i\omega t}, \qquad \mathbf{e}_2 = \mathbf{e}_2(t) = i\mathbf{e}_1, \qquad \mathbf{e}_3 = \mathbf{e}_3(t) = (0,0,1) \tag{2.3}$$

with $i^2 = -1$. This orthonormal moving frame corresponds to the synodical system. Then, (2.1) can be written as

$$\begin{aligned} x_1'' - 2\omega x_2' - \omega^2 x_1 &= \frac{\partial U}{\partial x_1}, \\ x_2'' + 2\omega x_1' - \omega^2 x_2 &= \frac{\partial U}{\partial x_2}, \\ x_3'' &= \frac{\partial U}{\partial x_3}, \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} U = U(x_1, x_2, x_3) &= \frac{m_1}{d_1} + \frac{m_2}{d_2} + \frac{m_3}{d_3}, \\ d_1 &= \sqrt{(x_1 - \alpha_1)^2 + (x_2 - \beta_1)^2 + x_3^2}, \\ d_2 &= \sqrt{(x_1 - \alpha_2)^2 + (x_2 - \beta_2)^2 + x_3^2}, \\ d_3 &= \sqrt{(x_1 - \alpha_3)^2 + (x_2 - \beta_3)^2 + x_3^2}, \end{aligned} \tag{2.5}$$

where $\mathbf{r}_j(t) = e^{i\omega t}\zeta_j$, with $\zeta_j = \alpha_j + \mathrm{i}\beta_j$ for $j = 1, 2, 3$. Applying the above notation, we can write $\mathbf{r} = (x_1 + ix_2)\mathbf{e}_1 + x_3\mathbf{e}_3$, $\mathbf{r}_1 = \zeta_1\mathbf{e}_1$, $\mathbf{r}_2 = \zeta_2\mathbf{e}_1$, $\mathbf{r}_3 = \zeta_3\mathbf{e}_1$, and so $\|\mathbf{r} - \mathbf{r}_j\| = \|(x_1 + ix_2) + x_3\mathbf{e}_3 - \zeta_j\|$ for $j = 1, 2, 3$.

We perform the reparametrization of time $d\tau = \omega dt$, then the system (2.4) is transformed into

$$\begin{aligned} \ddot{x}_1 - 2\dot{x}_2 - x_1 &= \frac{1}{\omega^2}\frac{\partial W}{\partial x_1}, \\ \ddot{x}_2 + 2\dot{x}_1 - x_2 &= \frac{1}{\omega^2}\frac{\partial W}{\partial x_2}, \\ \ddot{x}_3 &= \frac{1}{\omega^2}\frac{\partial W}{\partial x_3}, \end{aligned} \tag{2.6}$$

where the dot denotes the derivative with respect to τ, and the potential W is given by

$$W = W(x_1, x_2, x_3) = \frac{m_1}{\rho_1} + \frac{m_2}{\rho_2} + \frac{m_3}{\rho_3} \tag{2.7}$$

with

$$\begin{aligned} \rho_1 &= \sqrt{(x_1 - \alpha_1)^2 + (x_2 - \beta_1)^2 + x_3^2}, \\ \rho_2 &= \sqrt{(x_1 - \alpha_2)^2 + (x_2 - \beta_2)^2 + x_3^2}, \\ \rho_3 &= \sqrt{(x_1 - \alpha_3)^2 + (x_2 - \beta_3)^2 + x_3^2}. \end{aligned} \tag{2.8}$$

If we define $\mu_1 = m_1/M$, $\mu_2 = m_2/M$, and $\mu_3 = m_3/M$, where $M = m_1+m_2+m_3$, the equations of motions (2.4) become

$$\begin{aligned} \ddot{x}_1 - 2\dot{x}_2 - x_1 &= \frac{M}{\omega^2}\frac{\partial W}{\partial x_1}, \\ \ddot{x}_2 + 2\dot{x}_1 - x_2 &= \frac{M}{\omega^2}\frac{\partial W}{\partial x_2}, \\ \ddot{x}_3 &= \frac{M}{\omega^2}\frac{\partial W}{\partial x_3}, \end{aligned} \tag{2.9}$$

where

$$W = W(x_1, x_2, x_3) = \frac{M\mu_1}{\rho_1} + \frac{M\mu_2}{\rho_2} + \frac{M\mu_3}{\rho_3}. \tag{2.10}$$

For simplicity, we will consider an equilateral triangle of side 1 and so we obtain that $M/\omega^2 = 1$.

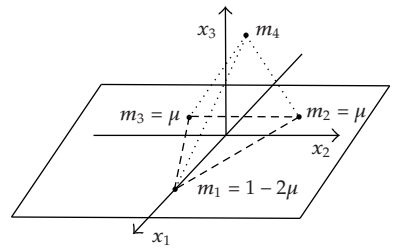


Figure 2: The equilateral restricted four body problem in a rotating frame.

3. Equations of Motion and Preliminary Results

From (2.9), we deduce that the equations of motion of the ERFBP in synodical coordinates are given by the system of differential equations

$$\begin{aligned} \ddot{x}_1 - 2\dot{x}_2 &= \Omega_{x_1}, \\ \ddot{x}_2 + 2\dot{x}_1 &= \Omega_{x_2}, \\ \ddot{x}_3 &= \Omega_{x_3}, \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} \Omega = \Omega(x_1, x_2, x_3) &= \frac{1}{2}\left(x_1^2 + x_2^2\right) + W(x_1, x_2, x_3), \\ W = W(x_1, x_2, x_3) &= \frac{1-2\mu}{\rho_1} + \frac{\mu}{\rho_2} + \frac{\mu}{\rho_3}, \end{aligned} \tag{3.2}$$

with

$$\begin{aligned} \rho_1 &= \sqrt{\left(x_1 - \sqrt{3}\mu\right)^2 + x_2^2 + x_3^2}, \\ \rho_2 &= \sqrt{\left(x_1 + \frac{\sqrt{3}}{2}(1-2\mu)\right)^2 + \left(x_2 - \frac{1}{2}\right)^2 + x_3^2}, \\ \rho_3 &= \sqrt{\left(x_1 + \frac{\sqrt{3}}{2}(1-2\mu)\right)^2 + \left(x_2 + \frac{1}{2}\right)^2 + x_3^2}. \end{aligned} \tag{3.3}$$

Analogously to the circular three-body problem, we can verify that the system (3.1) possesses a first Jacobi type integral given by

$$C = \frac{1}{2}\left(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2\right) - \Omega(x_1, x_2, x_3). \tag{3.4}$$

Thus we have the following result.

Proposition 3.1. *The Jacobi-type function* (3.4) *is a first integral of the ERFBP for any value of* μ.

Proof. Differentiating (3.4) with respect to the time, we get

$$\frac{dC}{dt} = \dot{x}_1\ddot{x}_1 + \dot{x}_2\ddot{x}_2 + \dot{x}_3\ddot{x}_3 - x_1\dot{x}_1 - x_2\dot{x}_2 - \frac{\partial W}{\partial x_1}\dot{x}_1 - \frac{\partial W}{\partial x_2}\dot{x}_2 - \frac{\partial W}{\partial x_3}\dot{x}_3, \tag{3.5}$$

and using (2.9) we can reduce the obtained expression to

$$\begin{aligned}\frac{dC}{dt} = {} & \dot{x}_1\left(x_1 + \frac{\partial W}{\partial x_1} + 2\dot{x}_2\right) + \dot{x}_2\left(x_2 + \frac{\partial W}{\partial x_2} - 2\dot{x}_1\right) + \dot{x}_3\frac{\partial W}{\partial x_3} \\ & - x_1\dot{x}_1 - x_2\dot{x}_2 - \frac{\partial W}{\partial x_1}\dot{x}_1 - \frac{\partial W}{\partial x_2}\dot{x}_2 - \frac{\partial W}{\partial x_3}\dot{x}_3 = 0.\end{aligned} \tag{3.6}$$

Hence C is a constant of motion. □

In order to write the Hamiltonian formulation of the ERFBP we introduce the new variables

$$\begin{aligned} x = x_1, \quad & y = x_2, \quad z = x_3, \\ X = \dot{x} - y, \quad & Y = \dot{y} + x, \quad Z = \dot{z}. \end{aligned} \tag{3.7}$$

Hence, it is verified that system (3.1) is equivalent to an autonomous Hamiltonian system with three degrees of freedom with Hamiltonian function given by

$$H = H(x, y, z, X, Y, Z) = \frac{1}{2}\left(X^2 + Y^2 + Z^2\right) + (yX - xY) - W. \tag{3.8}$$

Therefore, the Hamiltonian system associated is

$$\begin{aligned} \dot{x} = y + X, \quad & \dot{X} = Y + W_x \\ \dot{y} = -x + Y, \quad & \dot{Y} = -X + W_y \\ \dot{z} = Z, \quad & \dot{Z} = W_z. \end{aligned} \tag{3.9}$$

Of course, the phase space where the equations of motion are well defined is

$$\begin{aligned}\mathcal{M} = \Bigg\{ (x, y, z, X, Y, Z) \in \Bigg(\mathbb{R}^3 \setminus \Bigg\{ & \left(\sqrt{3}\mu, 0, 0\right), \left(-\frac{\sqrt{3}}{2}(1 - 2\mu), \frac{1}{2}, 0\right), \\ & \left(-\frac{\sqrt{3}}{2}(1 - 2\mu), -\frac{1}{2}, 0\right) \Bigg\} \Bigg) \times \mathbb{R}^3 \Bigg\},\end{aligned} \tag{3.10}$$

where the points that have been removed correspond to binary collisions between the massless particle and one of the primaries.

Additionally, the spatial ERFBP admits the planar case as a subproblem, that is, $z = Z = 0$ is invariant under the flow defined by (3.9).

On the other hand, we see that there are two limiting cases in the ERFBP, which we described below.

(a) If $\mu = 0$, we obtain a *central force problem*, with the body of mass $m_1 = 1$ at the origin of the coordinates.

(b) If $\mu = 1/2$, we obtain *the circular restricted three-body problem*, with masses $m_2 = m_3 = 1/2$.

Note that $\mu = 1/3$ corresponds to the symmetric case, that is, where the masses of the primaries are all equal to $1/3$.

It is easily seen that the equations of motion (3.9) are invariant by the symmetry

$$S: \ (x, y, z, X, Y, Z, \tau) \longrightarrow (x, -y, z, -X, Y, -Z, -\tau). \tag{3.11}$$

This means that if $\psi(\tau) = (x(\tau), y(\tau), z(\tau), X(\tau), Y(\tau), Z(\tau))$ is a solution of the system (3.9), then $\varphi(t) = (x(-\tau), -y(-\tau), z(-\tau), -X(-\tau), Y(-\tau), -Z(-\tau))$ is also a solution. We note that this symmetry corresponds to a symmetry with respect to the xz-plane. In the planar case, the symmetry corresponds to symmetry with respect to the x-axis.

4. Permitted Regions of Motion

In this section, we will see that the function $\Omega(x, y, z)$ allows us to establish regions in the (x, y, z) space, where the motion of the infinitesimal particle could take place. We will use similar ideas to those developing in [15, 16].

By using (3.4), the surface of zero velocity is defined by the set

$$\mathcal{R}_C : (x, y, z) \in \mathbb{R}^3 \quad \text{such that } \Omega(x, y, z) = -C, \quad \text{for any level } C. \tag{4.1}$$

This set corresponds to the so-called Hill region. We note that $C \geq 0$ implies $\mathcal{R}_C = \mathbb{R}^3 \setminus \{(\sqrt{3}\mu, 0, 0), (-\sqrt{3}/2(1-2\mu), 1/2, 0), (-\sqrt{3}/2(1-2\mu), -1/2, 0)\}$. That is, the region of all possible motions is given by the whole phase space and so the infinitesimal particle is free to move; in particular escape solutions are permitted.

In the spatial case, the surfaces that separate allowed and nonallowed motions are called zero-velocity surfaces, and for the planar case the set that separates the allowed and nonallowed motions is called zero-velocity curve. The shape and size of zero velocity sets $-C = \Omega(x, y, z)$ depend on C and μ. They correspond to the boundary of the Hill regions. The zero-velocity set $(\partial\mathcal{R}_C)$ is defined by the equation

$$\Omega = \frac{1}{2}\left(x^2 + y^2\right) + \frac{1-2\mu}{\sqrt{\left(x - \sqrt{3}\mu\right)^2 + y^2 + z^2}} + \frac{\mu}{\sqrt{\left(x + (\sqrt{3}/2)(1-2\mu)\right)^2 + (y - 1/2)^2 + z^2}} + \frac{\mu}{\sqrt{\left(x + (\sqrt{3}/2)(1-2\mu)\right)^2 + (y + 1/2)^2 + z^2}} = -C, \tag{4.2}$$

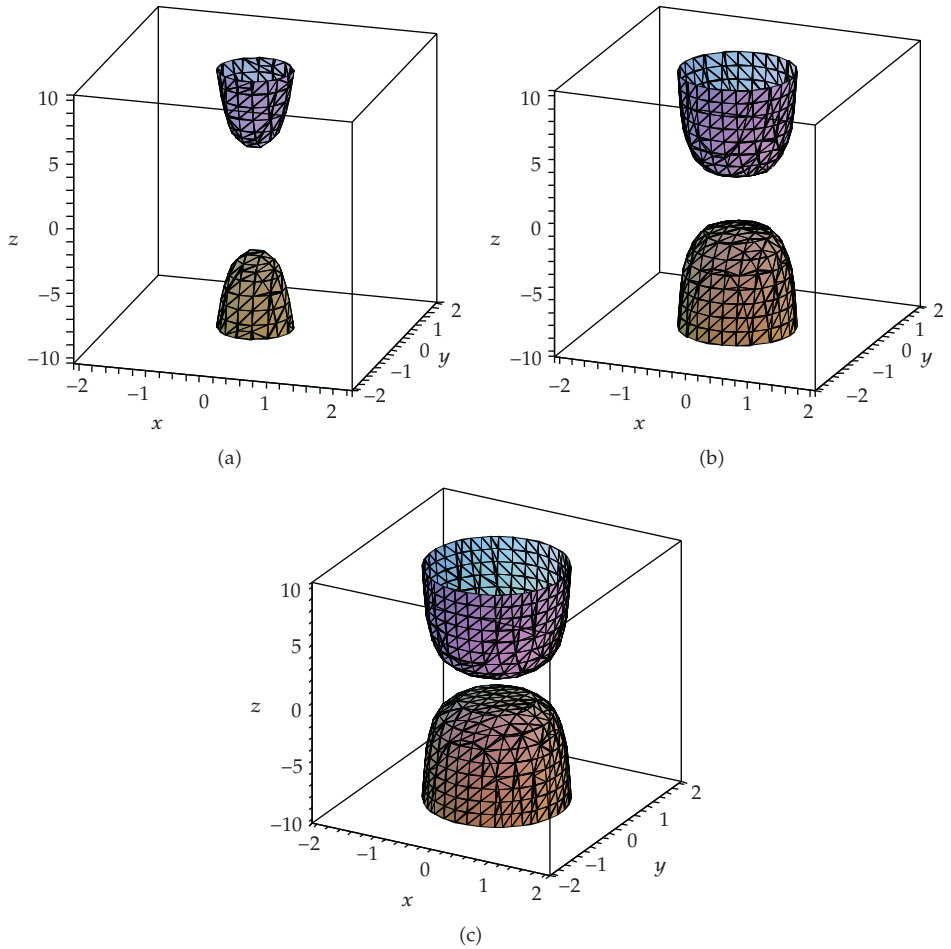


Figure 3: Evolution of zero-velocity surface in the three-dimensional ERFBP for $\mu = 1/3$. (a) $C = -1/4$. (b) $C = -1/2$. (c) $C = -3/4$.

only for $C < 0$ and any value of μ. Next, we give a list of all possible situations that may appear when this condition is fulfilled.

(1) $z \to \pm\infty$ on the $\partial\mathcal{R}_C$ in which case $x^2 + y^2 \to -2C$, this means, that around the z-axis the variables (x, y) must be asymptotic to a circle of radius $\sqrt{-2C}$.

(2) $x \to \infty$ or $-\infty$ (resp., $y \to \infty$ or $-\infty$) on the $\partial\mathcal{R}_C$, when $C \to -\infty$.

(3) For $|C|$ very large this implies that (x, y) can be sufficiently close to one of the primaries, or the infinitesimal mass is close to infinity.

(4) Since $x^2 + y^2$ is a factor of Ω on $\partial\mathcal{R}_C$, then small values for $-C$ are not allowed.

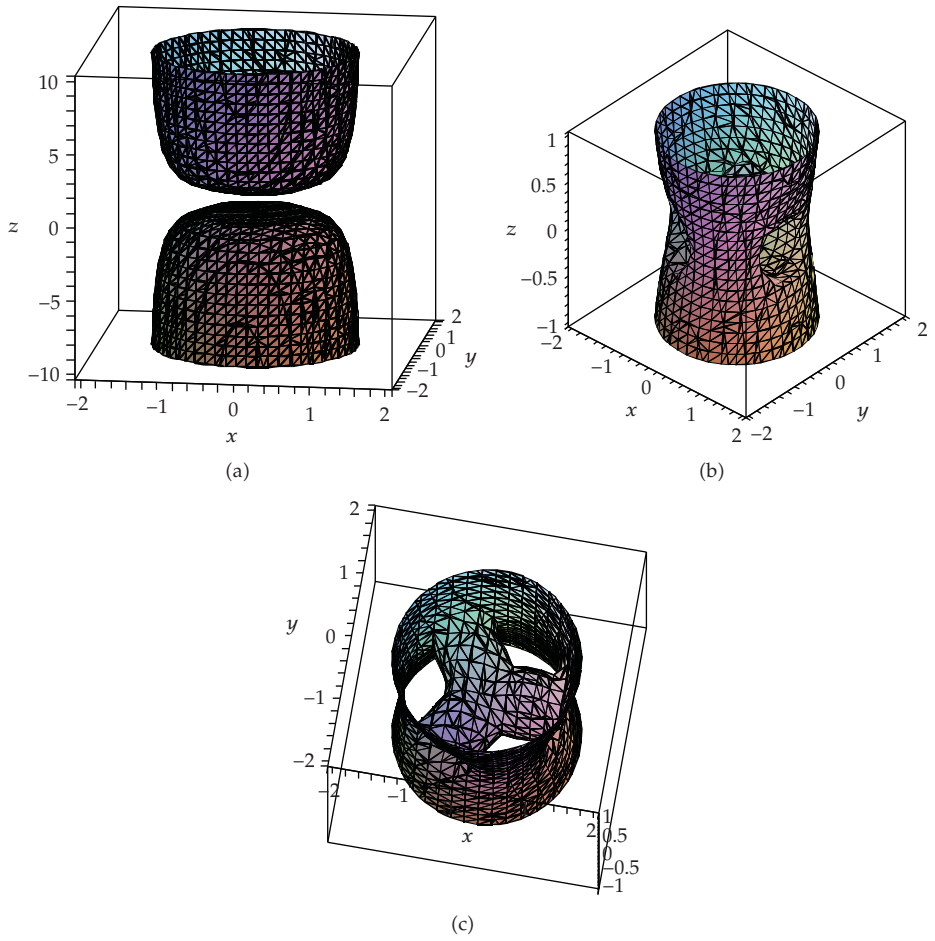


Figure 4: Evolution of zero-velocity surface in the three dimensional ERFBP for $\mu = 1/3$. (a) $C = -1$. (b) and (c) $C = -1.6$ under different points of view.

By simplicity, we will only show zero-velocity surfaces for the case $\mu = 1/3$ and different values of the integral of motion C. Figures 3, 4, 5, and 6 show evolution of zero-velocity surfaces for several C values.

4.1. The Planar Case

As we mentioned in last section, the set $\{z = Z = 0\}$ is invariant under the flow, and so the motion of the infinitesimal body lies on the xy plane that contains the primaries. In Figure 7, we show the evolution of the function Ω in the planar case for different values of the parameter μ.

Next we show the evolution of the Hill's regions as well as the zero velocity curves, for $\mu = 1/3$ and many values of the Jacobian constant C; the permissible areas are shown on Figures 8, 9, 10, and 11 shading.

In Figure 12, we show the behavior of level curves in the planar case for some values of μ and for different energy levels.

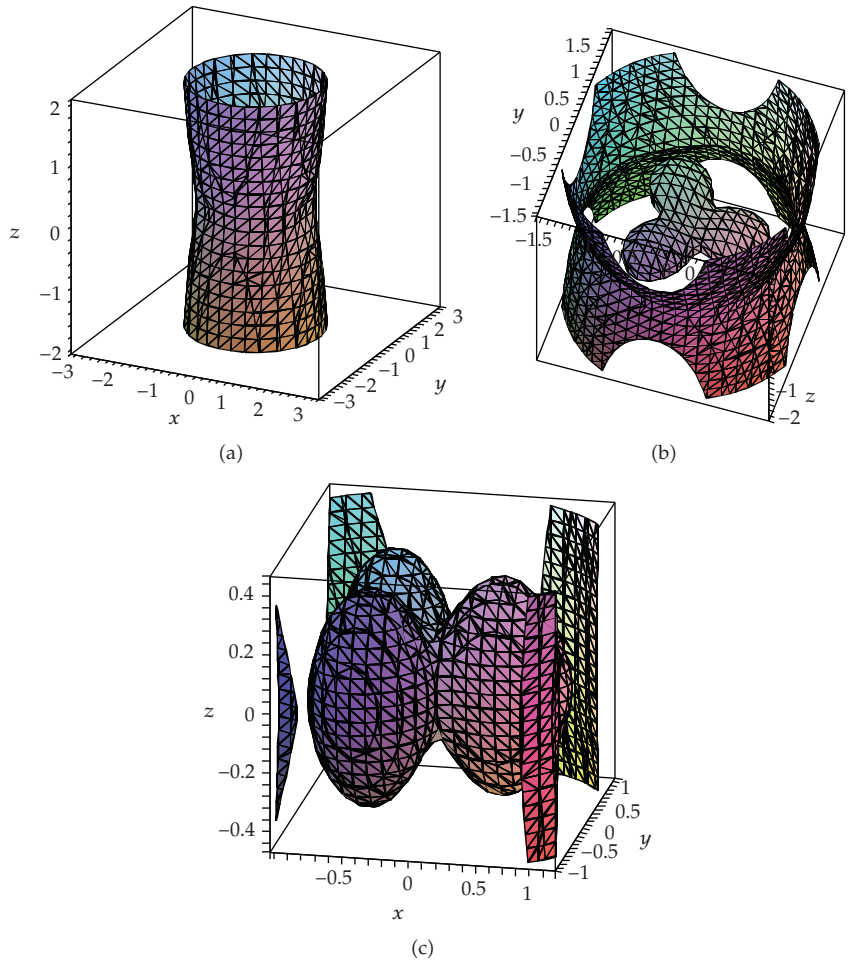


Figure 5: Evolution of zero-velocity surface in the three dimensional ERFBP for $\mu = 1/3$. All cases correspond to $C = -1.7$ under different points of view.

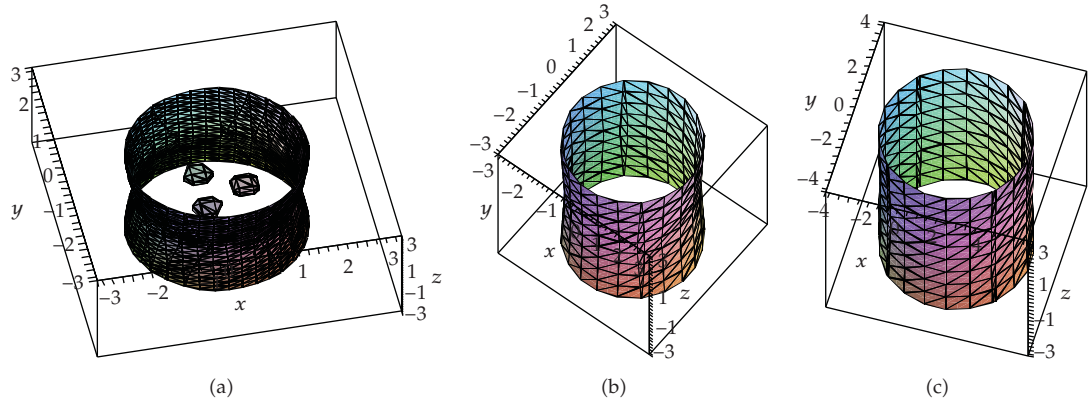


Figure 6: Evolution of zero-velocity surface in the three dimensional ERFBP for $\mu = 1/3$. (a) $C = -2$. (b) $C = -3$. (c) $C = -5$.

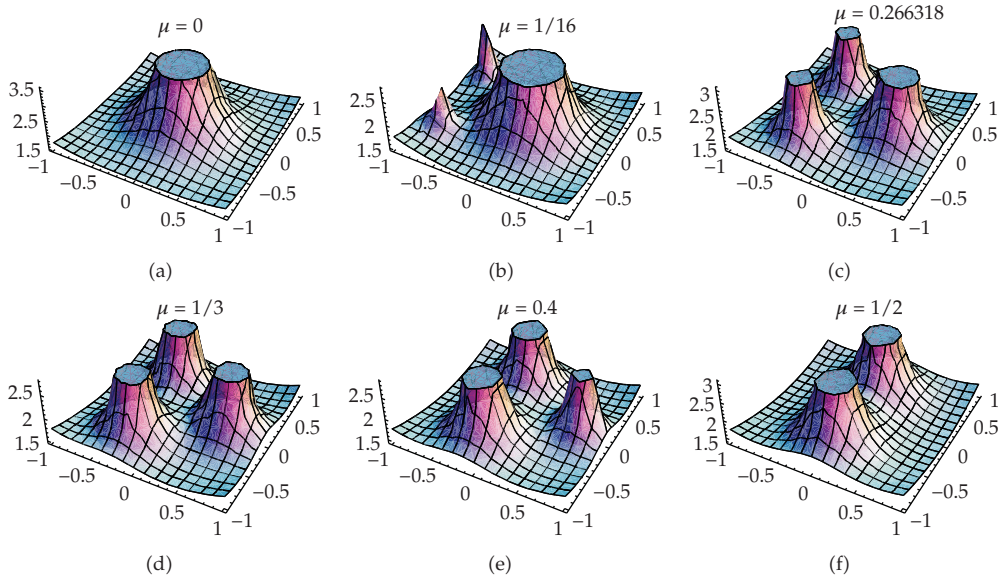


Figure 7: Evolution of the graph of $\Omega(x, y)$ on the xy plane for different values of μ.

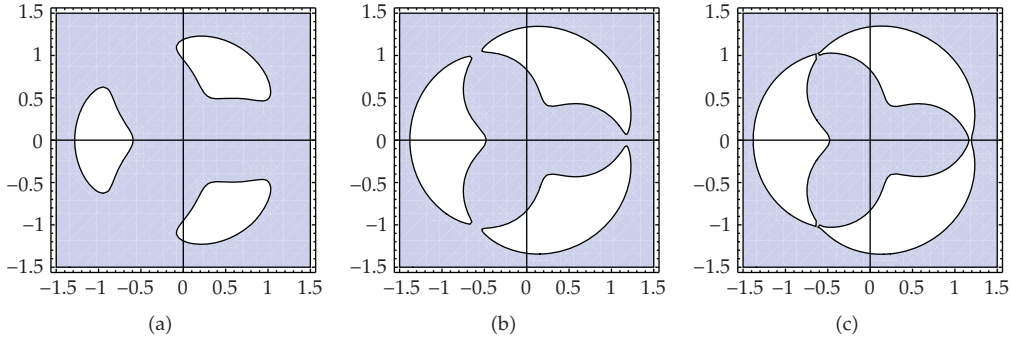


Figure 8: Evolution of zero-velocity curves and Hill's region in the planar ERFBP for $\mu = 1/3$, where shading represents permissible areas. (a) $C = -1.6$, (b) $C = -1.6775$, (c) $C = -1.6795$.

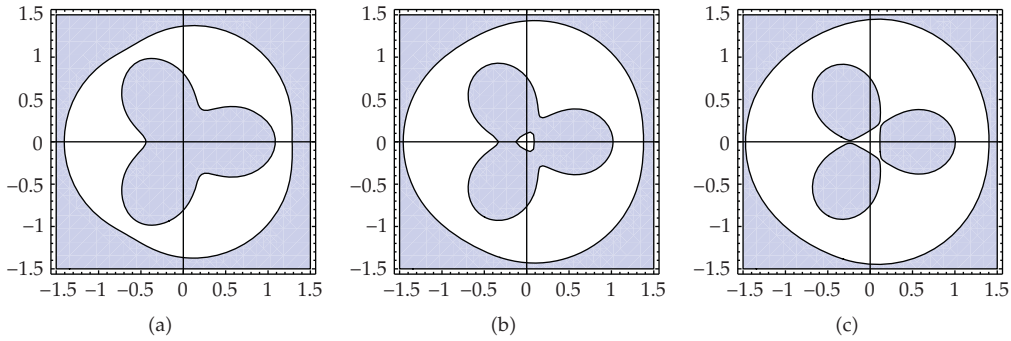


Figure 9: Evolution of zero-velocity curves and Hill's region in the planar ERFBP for $\mu = 1/3$, where shading are permissible areas. (a) $C = -1.7$, (b) $C = -1.75$, (c) $C = -1.765$.

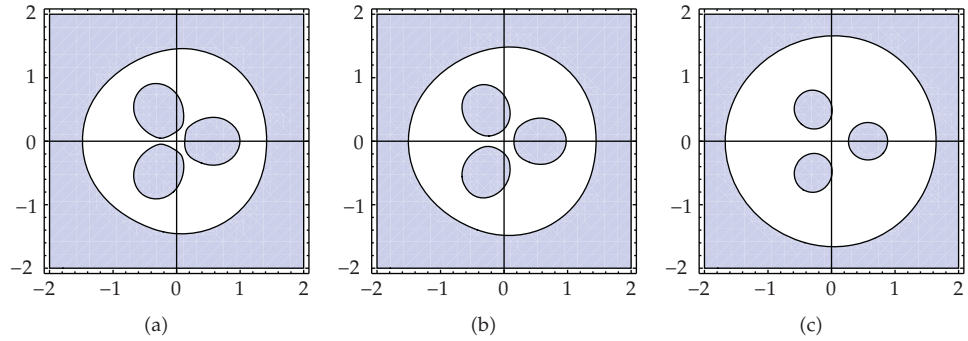


Figure 10: Evolution of zero-velocity curves and Hill's region in the planar ERFBP for $\mu = 1/3$, where shading are permissible areas. (a) $C = -1.775$. (b) $C = -1.8$. (c) $C = -2$.

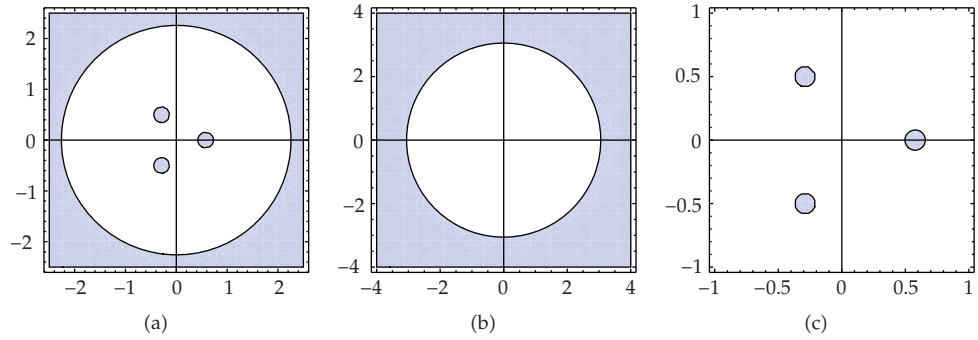


Figure 11: Evolution of zero-velocity curves and Hill's region in the planar ERFBP for $\mu = 1/3$, where shading are permissible areas. (a) $C = -3$, (b) and (c) both figures correspond to $C = -5$ with different window size.

5. Equilibrium Solutions

It is verified that the equilibrium solutions of the system (3.9) or equivalently (3.1) are given by the critical points of the function $\Omega = \Omega(x, y, z)$ or simply they are the solutions of the following system of equations:

$$\begin{aligned}
&(1-2\mu)\frac{x-\sqrt{3}\mu}{\rho_1^3} + \mu\left(x + \frac{\sqrt{3}}{2}(1-2\mu)\right)\left(\frac{1}{\rho_2^3} + \frac{1}{\rho_3^3}\right) = x,\\
&(1-2\mu)\frac{y}{\rho_1^3} + \mu\left(\frac{y-1/2}{\rho_2^3} + \frac{y+1/2}{\rho_3^3}\right) = y,\\
&-\left(\frac{1-2\mu}{\rho_1^3} + \frac{\mu}{\rho_2^3} + \frac{\mu}{\rho_3^3}\right)z = 0.
\end{aligned} \tag{5.1}$$

From the last equation we see that the coordinate z must be zero, so the critical points are restricted to the plane xy, and are given by the solutions of the first two equations.

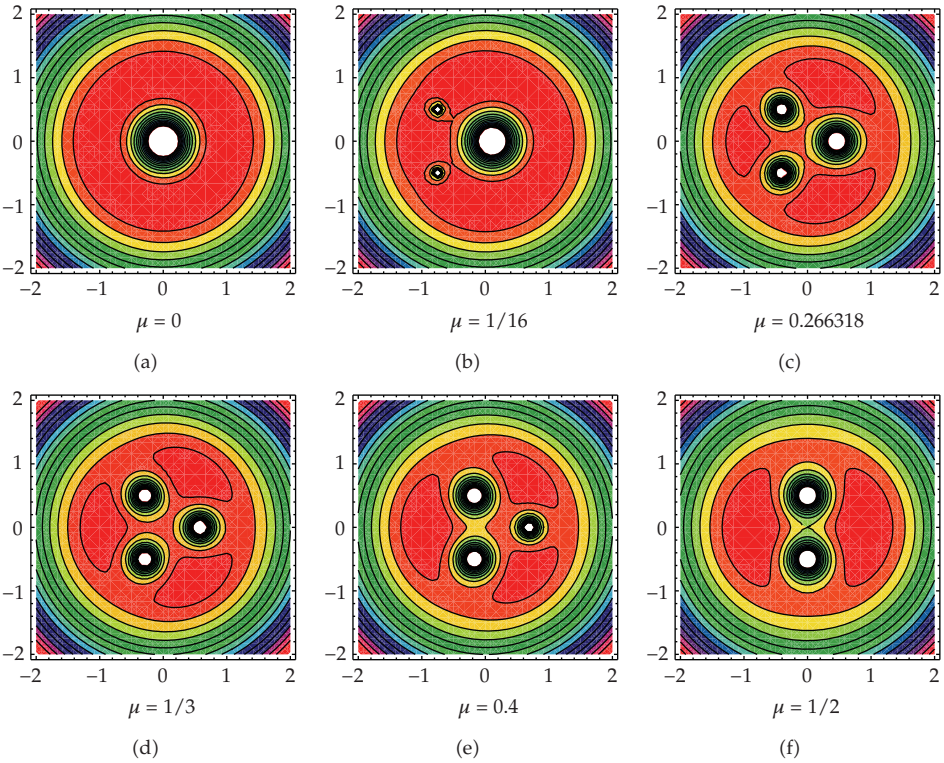


Figure 12: Energy level curves for some values of the parameter μ in the planar case.

It is known (see [17]) that the number of equilibrium solutions of the system (5.1) is 8, 9 or 10 depending on the values of the masses, m_1, m_2 and m_3 which must be positive. Six of them are out of the symmetry axis (i.e., out of the x-axis), therefore on the axis of symmetry we must have 2, 3 or 4. From the analysis done it follows that the number of the equilibrium solutions depends on the parameter μ. This implies that finding the critical points is a non-trivial problem, and this is one of the main differences with the problem studied by Maranhão in his doctoral thesis [1], because there, the number of critical points did not depend on the parameter μ.

The critical points on the axis $y = 0$ are the zeros of the function

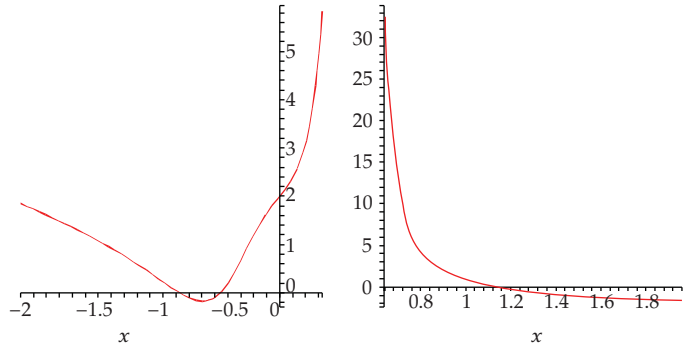
$$F_\mu(x) = (1-2\mu)\frac{\left|x-\sqrt{3}\mu\right|}{\left(x-\sqrt{3}\mu\right)^3} + 2\mu\frac{x+\left(\sqrt{3}/2\right)(1-2\mu)}{\rho_2^3} - x, \tag{5.2}$$

where

$$\rho_1 = \left|x-\sqrt{3}\mu\right|, \qquad \rho_2 = \rho_3 = \sqrt{\left(x+\frac{\sqrt{3}}{2}(1-2\mu)\right)^2 + \frac{1}{4}}. \tag{5.3}$$

Table 1: Number of critical points on the x-axis.

$\mu = 0$	$0 < \mu < \mu^*$	$\mu = \mu^*$	$\mu^* < \mu < 1/2$	$\mu = 1/2$
2	2	3	4	3



(a) (b)

Figure 13: Graph of $F_{1/3}$. (a) $x < \sqrt{3}/3$. (b) $x > \sqrt{3}/3$.

An explicit computation shows that in the limit case problems the number of equilibrium points corresponding to the system (5.1) is as follows.

(a) The function (5.2) with $\mu = 0$ results in

$$F_0(x) = \frac{x}{|x|^3} - x \tag{5.4}$$

whose zeros are $x = -1$ and $x = 1$, and so there are two equilibrium points.

(b) Taking $\mu = 1/2$ in (5.2) becomes

$$F_{1/2}(x) = \frac{x}{\left(x^2 + (1/4)\right)^{3/2}} - x \tag{5.5}$$

with zeros given by $x = -(\sqrt{3}/2)$, $x = 0$ and $x = (\sqrt{3}/2)$. We conclude that there are three equilibrium points.

From numerical simulations we get that the number of critical points along the x–axis is given in Table 1. Observe that $\mu^* := 0.266318$ is the bifurcation value.

In the symmetric case when all the masses are equals (i.e., $\mu = 1/3$) we have that the graph of $F_{1/3}$ is similar to the one shown in Figure 13. As a consequence, there are exactly 4 equilibrium solutions on the x-axis, and therefore there are exactly 10 equilibrium solutions. Of course, $(0, 0, 0)$ is an equilibrium solution.

In general for any equilibrium solution of the form $(x_0, y_0, 0)$, the linearized system (3.9) in the planar case give us that the characteristic polynomial is

$$\begin{aligned} C_A(\lambda) =& \Big(\lambda^2 - W_{zz}(x_0, y_0, 0)\Big)\lambda^4 + (2 - W_{xx}(x_0, y_0, 0) - W_{yy}(x_0, y_0, 0))\lambda^2 \\ &+ \Big(1 + W_{xx}(x_0, y_0, 0) + W_{yy}(x_0, y_0, 0) + W_{xx}(x_0, y_0, 0)W_{yy}(x_0, y_0, 0) \\ &- W_{xy}^2(x_0, y_0, 0)\Big), \end{aligned} \tag{5.6}$$

whose roots are

$$\lambda = \pm\sqrt{W_{zz}(x_0, y_0, 0)}, \qquad \lambda = \pm\frac{1}{2}\sqrt{\rho^{\pm}}, \tag{5.7}$$

where $\rho^{\pm}$is given by

$$\rho^{\pm} = -4 + 2(a + c) \pm 2\sqrt{(a - c)^2 + 4b^2 - 8(a + c)} \tag{5.8}$$

with $a = W_{xx}(x_0, y_0, 0)$, $b = W_{xy}(x_0, y_0, 0)$ and $c = W_{yy}(x_0, y_0, 0)$. A very simple result is the following.

Lemma 5.1. *The roots of $p(\rho) = \rho^2 + A\rho + B$ are real and negative if and only if $A > 0$, $B > 0$ and $\Delta = A^2 - 4B \geq 0$.*

Associating to our characteristic polynomial (5.6) we have

$$\begin{aligned} A &= 2 - W_{xx}(x_0, y_0, 0) - W_{yy}(x_0, y_0, 0), \\ B &= 1 + W_{xx}(x_0, y_0, 0) + W_{yy}(x_0, y_0, 0) + W_{xx}(x_0, y_0, 0)W_{yy}(x_0, y_0, 0) - W_{xy}^2(x_0, y_0, 0). \end{aligned} \tag{5.9}$$

Now, we remark that

$$W_{zz}(x_0, y_0, 0) = -\left[\frac{1 - 2\mu}{\rho_1^3} + \frac{\mu}{\rho_2^3} + \frac{\mu}{\rho_3^3}\right] < 0. \tag{5.10}$$

Consequently we have the following result:

Corollary 5.2. *In the spatial ERFBP for any equilibrium solution $(x_0, y_0, 0)$ we have at least two pure imaginary eigenvalues associated to the linear part, which are given by $\lambda = \pm\sqrt{-W_{zz}(x_0, y_0, 0)}i$.*

From this corollary we deduce that to study the nonlinear stability in the Lyapunov sense of each equilibrium solution of the spatial ERFBP is not a simple problem, because we need to take into account the existence or not of resonance in each situation. Leandro in [17] studied the spectral stability in some situations (according to the localization of the equilibrium solution along the symmetry-axis). In a future work we intend to study the Lyapunov stability of each equilibrium.

5.1. Analysis of the Symmetrical Case, $\mu = 1/3$

As we have said previously in the symmetrical case (i.e., $\mu = 1/3$) there are 10 equilibrium solutions and one of them is $(0,0,0)$. Here we have $\rho_1 = \rho_2 = \rho_3 = 1/\sqrt{3}$, $a = b = 3\sqrt{3}/2$ and $c = 0$. Consequently, the characteristic roots are

$$\begin{aligned} \lambda_1 &= -\sqrt{3\sqrt{3}i}, & \lambda_2 &= \sqrt{3\sqrt{3}i}, \\ \lambda_3 &= -\frac{1}{2}\sqrt{6\sqrt{3} - 4 + 4\sqrt{-6\sqrt{3}}}, & \lambda_4 &= \frac{1}{2}\sqrt{6\sqrt{3} - 4 + 4\sqrt{-6\sqrt{3}}}, \\ \lambda_5 &= -\frac{1}{2}\sqrt{6\sqrt{3} - 4 - 4\sqrt{-6\sqrt{3}}}, & \lambda_6 &= \frac{1}{2}\sqrt{6\sqrt{3} - 4 - 4\sqrt{-6\sqrt{3}}}. \end{aligned} \tag{5.11}$$

Therefore, we have the following result.

Corollary 5.3. *In the symmetrical spatial ERFBP the equilibrium solution* $(0,0,0)$ *is unstable in the Lyapunov sense.*

In general, it is possible to prove that the equilibrium solutions on the x–axis are $x_1 = -0.9351859666722429$, $x_2 = -0.23895830919534947$ and $x_3 = 1.1799984048894328$, and by symmetry it follows:

Corollary 5.4. *In the symmetrical spatial ERFBP all the equilibrium solutions are unstable in the Lyapunov sense.*

According to [17] we have the following corollary.

Corollary 5.5. *In the symmetrical planar ERFBP all the equilibrium solutions are unstable in the Lyapunov sense.*

6. Continuation of Periodic Solutions in the Planar Case

In this section we prove the existence of periodic solutions in the ERFBP for μ sufficiently small in the planar case and by the use of the Lyapunov Center Theorem when μ is close to $1/2$. In order to find periodic orbits of our problem we will use the continuation method developed by Poincaré which is one of the most frequently used methods to prove the existence of periodic orbits in the planar circular restricted three-body problem (see [15]). This method has been also used by other authors in different problems. In Meyer and Hall [5], we find a good discussion of the Poincaré continuation method to different n-body problem (see also [18]).

In our approach we will continue circular and elliptic solutions of the Kepler problem with the body fixed in the origin of the system with mass 1. We know that all the orbits of the Kepler problem with angular momentum zero are collision orbits with the origin. We assume that the angular momentum is not zero and we study the orbits that have positive distance of $(-\sqrt{3}/2, 1/2)$ and $(-\sqrt{3}/2, -1/2)$. In the following lemma we resume the kind of orbits that we will consider.

Lemma 6.1. *Fixed $a > 0$ there exists a finite number of elliptic orbits with semi-major axis a, such that its trajectories are periodic in the rotating system and pass through the singularity of the other primaries $(-\sqrt{3}/2, 1/2)$ or $(-\sqrt{3}/2, -1/2)$.*

The proof of this lemma can be found in [19].

6.1. Continuation of Circular Orbits

In this section we show that circular solutions of the unperturbed Kepler problem can be continued to periodic solutions of the ERTBP for small values of μ. We introduce the polar coordinates given as $x = r\cos\theta, y = r\sin\theta$, thus $\dot{x} = \dot{r}\cos\theta - r\dot{\theta}\sin\theta$ and $\dot{y} = \dot{r}\sin\theta + r\dot{\theta}\cos\theta$. So, $\dot{X} = \dot{r}\cos\theta - r(\dot{\theta}+1)\sin\theta$ and $\dot{Y} = \dot{r}\sin\theta + r(\dot{\theta}+1)\cos\theta$, consequently $X^2 + Y^2 = \dot{r}^2 + r^2(\dot{\theta}+1)^2$ and $yX - xY = -r^2(\dot{\theta}+1)$. Thus, the Hamiltonian (3.8) now is

$$H = \frac{\dot{r}^2 + r^2(\dot{\theta}+1)^2}{2} - r^2(\dot{\theta}+1) - V(r,\theta), \tag{6.1}$$

where

$$V(r,\theta) = \frac{1-2\mu}{\rho_1} + \mu\left(\frac{1}{\rho_2} + \frac{1}{\rho_3}\right), \tag{6.2}$$

where

$$\begin{aligned} \rho_1 &= \sqrt{\left(r\cos\theta - \sqrt{3}\mu\right)^2 + r^2\sin^2\theta}, \\ \rho_2 &= \sqrt{\left(r\cos\theta + \sqrt{3}/2(1-2\mu)\right)^2 + (r\sin\theta - 1/2)^2}, \\ \rho_3 &= \sqrt{\left(r\cos\theta + \sqrt{3}/2(1-2\mu)\right)^2 + (r\sin\theta + 1/2)^2}. \end{aligned} \tag{6.3}$$

The new coordinates are not symplectic. In order to obtain a set of symplectic coordinates (r, θ, R, Θ) we define $R = \dot{r}$ (radial velocity in the sideral system) and $\Theta = r^2(\dot{\theta}+1)$ (angular momentum in the sideral system), then H is

$$H = \frac{R^2 + \Theta/r^2}{2} - \Theta - V(r,\theta). \tag{6.4}$$

When $\mu = 0$ we have that

$$H = \frac{R^2 + \Theta^2/r^2}{2} - \Theta - \frac{1}{r}, \tag{6.5}$$

is the Hamiltonian of the Kepler problem in polar coordinates. So, if μ is a small parameter, the Hamiltonian (3.8) assumes the form

$$H = \frac{R^2 + \Theta^2/r^2}{2} - \Theta - \frac{1}{r} + \mathcal{O}(\mu). \tag{6.6}$$

For $\mu = 0$, the Hamiltonian system associated is

$$\begin{aligned} \dot{r} &= R, \qquad \dot{R} = \frac{\Theta^2}{r^3} - \frac{1}{r^2}, \\ \dot{\theta} &= \frac{\Theta}{r^2} - 1, \qquad \dot{\Theta} = 0. \end{aligned} \tag{6.7}$$

Let $\Theta = c$ be a fixed constant. For $c \neq 1$, the circular orbit $R = 0$, $r = c^2$ is a periodic solution with period $|2\pi c^3/(1 - c^3)|$. Linearizing the r and R equations about this solution gives

$$\dot{r} = R, \qquad \dot{R} = -c^{-6} r, \tag{6.8}$$

which has solutions of the form $\exp(\pm i t/c^3)$, and so the nontrivial multipliers of the circular orbits are $\exp(\pm i 2\pi/(1 - c^3))$ which are not +1, provided $1/(1 - c^3)$ is not an integer. Thus we have proved the following theorem (see details in [5]).

Theorem 6.2. *If $c \neq 1$ and $1/(1 - c^3)$ is not an integer, then the circular orbits of the Kepler problem in rotating coordinates with angular momentum c can be continued into the equilateral restricted four body problem for small values of μ.*

6.2. Continuation of Elliptic Orbits

In Section 3, we saw that the ERFBP has the S-symmetry which when exploited properly proves that some elliptic orbits can be continued from the Kepler problem. The main idea is given in the following lemma, which is a consequence of the uniqueness of the solution of the differential equations and the symmetry of the problem.

Lemma 6.3. *A solution of the equilateral restricted problem which crosses the line of syzygy (the x-axis) orthogonally at a time $t = 0$ and later at a time $t = T/2 > 0$ is T-periodic and symmetric with respect to the line syzygy.*

That is, if $x(t)$ and $y(t)$ is a solution of the equilateral restricted four body problem such that $y(0) = \dot{x}(0) = y(T/2) = \dot{x}(T/2) = 0$, where $T > 0$, then this solution is T-periodic and symmetric with respect to the x-axis.

In Delaunay variables (l, g, L, G), an orthogonal crossing of the line of sizygy at a time t_0 is

$$l(t_0) = n\pi, \quad g(t_0) = m\pi, \quad n, m \text{ integers}. \tag{6.9}$$

These equations will be solved using the Implicit Function theorem to yield the following theorem (see details in [5]).

Theorem 6.4. *Let m, k be relatively prime integers and $T = 2\pi m$. Then the elliptic T-periodic solution of the Kepler problem in rotating coordinates which satisfies*

$$l(0) = \pi, \qquad g(0) = \pi, \qquad L^3(0) = \frac{m}{k} \tag{6.10}$$

and does not go through $(-\sqrt{3}/2, 1/2)$ and $(-\sqrt{3}/2, -1/2)$ can be continued into the equilateral restricted four body problem for small μ. This periodic solution is symmetric with respect to the line of syzygy.

Proof. The Hamiltonian of the ERFBP in Delaunay coordinates for μ sufficiently small is

$$H = -\frac{1}{2L^2} - G + \mathcal{O}(\mu), \tag{6.11}$$

and the equations of motion are

$$\begin{aligned} \dot{l} &= \frac{1}{L^3} + \mathcal{O}(\mu), \qquad \dot{L} = 0 + \mathcal{O}(\mu), \\ \dot{g} &= -1 + \mathcal{O}(\mu), \qquad \dot{G} = 0 + \mathcal{O}(\mu). \end{aligned} \tag{6.12}$$

Let $L_0^3 = m/k$, and let $l(t, \Lambda, \mu)$, $g(t, \Lambda, \mu)$, $L(t, \Lambda, \mu)$ and $G(t, \Lambda, \mu)$ be the solution which goes through $l = \pi$, $g = \pi$, $L = \Lambda$, G arbitrary at $t = 0$; so, it is a solution with an orthogonal crossing of the line of syzygy at $t = 0$.

From (6.12) $l(t, \Lambda, 0) = t/\Lambda^3 + \pi$, $g(t, \Lambda, 0) = -t + \pi$. Thus, $l(T/2, L_0, 0) = (1 + k)\pi$ and $g(T/2, L_0, 0) = (1 - m)\pi$, and so when $\mu = 0$, this solution has another orthogonal crossing at time $T/2 = m\pi$. Also,

$$\det \begin{pmatrix} \dfrac{\partial l}{\partial t} & \dfrac{\partial l}{\partial \Lambda} \\ \dfrac{\partial g}{\partial t} & \dfrac{\partial g}{\partial \Lambda} \end{pmatrix}_{t=T/2,\ L=L_0,\ \mu=0} = \det \begin{pmatrix} \dfrac{k}{m} & -3\pi \left(\dfrac{k^4}{m}\right)^{1/3} \\ -1 & 0 \end{pmatrix} \neq 0. \tag{6.13}$$

Thus, the theorem follows by the Implicit Function theorem. □

6.3. Application of the Lyapunov Center Theorem

For $\mu = 1/2$, we have three equilibrium solutions on the x-axis which are $P_1 = (-\sqrt{3}/2, 0)$, $P_2 = (0, 0)$ and $P_3 = (\sqrt{3}/2, 0)$. At the point P_2, the associated eigenvalues are $\pm\sqrt{75 + 8\sqrt{2}}$ and $\pm\sqrt{75 - 8\sqrt{2}}$. Therefore, this equilibrium is unstable and by Lyapunov's Center Theorem (see [5]), we obtain the following theorem.

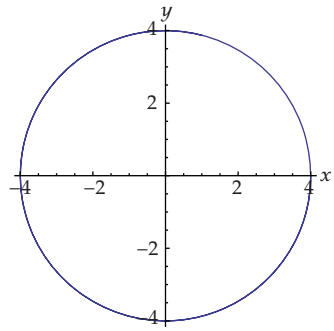


Figure 14: Circular orbit for $c = 2$ with $\mu = 10^{-2}$.

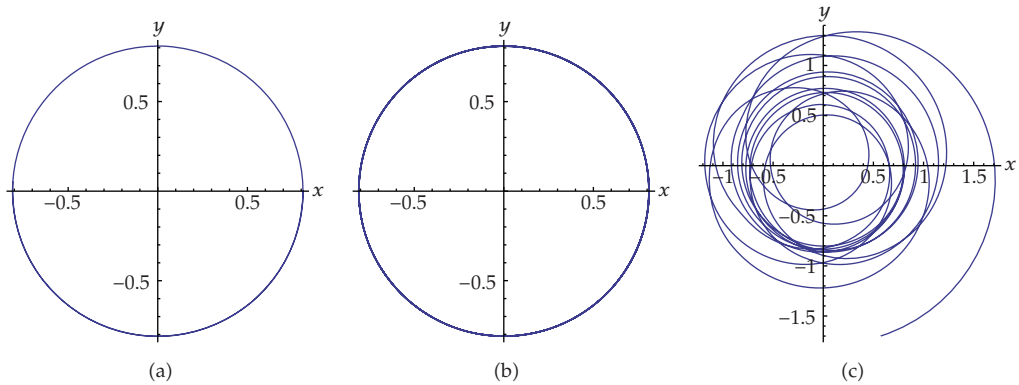


Figure 15: The circular orbit associated to $c = 9/10$ for $\mu = 0$ continued to $\mu = 10^{-4}$ which is not circular for $\mu = 10^{-2}$.

Theorem 6.5. *There exists a one-parameter family of periodic orbits of the ERFBP emanating from the Euler equilibrium (for $\mu = 1/2$). Moreover, when approaching the equilibrium point along the family, the periods tend to $2\pi/\sqrt{-3+8\sqrt{2}}$.*

7. Numerical Results

In the Section 8, we established theorems on the continuation of periodic solutions from the Kepler's problem in rotating coordinates to the ERFBP. In this section, we present some numerical experiments that illustrates the thesis of Theorem 6.2.

To find those circular orbits we first selected an angular momentum c such that $c \neq 1$ and $1/(1-c^3) \notin \mathbb{Z}$. By varying c we generated a set of initial conditions for Kepler problem in rotating coordinates given by the system (3.9) taking $\mu = 0$. We have chosen $y_0 = 0$ and $X_0 = 0$ for all orbits, ensuring that we were following a family of symmetric orbits; we have taken into account the fact that circular orbits satisfy $r = c^2$.

We have noticed that for values of $c = 2, 3, 4, 5, 6, 7, 8, 9, 10$ with $\mu = 10^{-2}$ the orbit is close to the circular orbit, see Figure 14.

However, the circular orbits associated to $c \approx 0,1$ is close to the circular orbit if $\mu \leq 10^{-4}$, for instance $c = 9/10$ can be continued for μ small and of the order 10^{-4} but not for higher values. The orbits obtained as a consequence of numerical simulations are shown in Figure 15.

8. Conclusions and Final Remarks

The spatial equilateral restricted four-body problem (ERFBP) is considered. This model of four-body problem, we have that three masses, moving in circular motion such that their configuration is always an equilateral triangle, the fourth mass being small and not influencing the motion of the three primaries. In our model we assume that two masses of the primaries m_2 and m_3 are equal to μ and the mass m_1 is $1 - 2\mu$. In a synodical systems of coordinates the dynamics obeys to the system of differential equations

$$\begin{aligned} \ddot{x} - 2\dot{y} &= \Omega_x, \\ \ddot{y} + 2\dot{x} &= \Omega_y, \\ \ddot{z} &= \Omega_z, \end{aligned} \tag{8.1}$$

where

$$\begin{aligned} \Omega = \Omega(x,y,z) &= \frac{1}{2}\left(x^2 + y^2\right) + \frac{1-2\mu}{\rho_1} + \frac{\mu}{\rho_2} + \frac{\mu}{\rho_3}, \\ \rho_1 &= \sqrt{\left(x - \sqrt{3}\mu\right)^2 + y^2 + z^2}, \\ \rho_2 &= \sqrt{\left(x + \frac{\sqrt{3}}{2}(1-2\mu)\right)^2 + \left(y - \frac{1}{2}\right)^2 + z^2} \\ \rho_3 &= \sqrt{\left(x + \frac{\sqrt{3}}{2}(1-2\mu)\right)^2 + \left(y + \frac{1}{2}\right)^2 + z^2}. \end{aligned} \tag{8.2}$$

In Section 4 it is devoted to give the principal qualitative aspect of the restricted problem—the surfaces and curves of zero velocity, several uses of which are discussed. The regions of permitted motion and the location and properties of the equilibrium points are established. We describe the Hill region. The description of the number of equilibrium points is given in Section 5, and in the symmetrical case (i.e., $\mu = 1/3$) we are describing the kind of stability of each equilibrium. In Section 6 the planar case is considered. Here, we prove the existence of periodic solutions as continuation of periodic Keplerian orbits, when the parameter μ is small and when it is close to $1/2$. Finally, in Section 7 we present some numerical experiments that illustrates the thesis of theorem concerning with the continuation of circular orbits of the Kepler problem to the ERFBP with μ small enough.

In a work in progress we intend to continue the study of the ERFBP in different aspects of its dynamics. For example, the behavior of the flow near the singularities (collisions). The study of the escapes solutions (i.e., the unbounded solutions). Existence of chaos under the

construction of a shift map. We desired to get periodic solutions under the use of numerical methods.

Acknowledgments

The first author was partially supported by CoNaCyT México Grant 32167-E. The authors would like to thank the anonymous referees for their very careful reviews of the paper that include many important points and will improve significantly the clarity of this paper.

References

[1] D. Maranhão, *Estudi del flux d'un problema restringt de quatre cossos*, Ph.D. thesis, UAB, Barcelona, Spain, 1995.
[2] D. L. Maranhão and J. Llibre, "Ejection-collision orbits and invariant punctured tori in a restricted four-body problem," *Celestial Mechanics & Dynamical Astronomy*, vol. 71, no. 1, pp. 1–14, 1998/99.
[3] A. Wintner, *The Analytical Foundations of Celestial Mechanics*, vol. 5 of *Princeton Mathematical Series*, Princeton University Press, Princeton, NJ, USA, 1941.
[4] R. H. Battin, *An Introduction to the Mathematics and Methods of Astrodynamics*, AIAA Education Series, AIAA, Reston, Va, USA, 1999.
[5] K. R. Meyer and G. R. Hall, *Introduction to Hamiltonian Dynamical Systems and the* N*-Body Problem*, vol. 90 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1992.
[6] J. Cronin, P. B. Richards, and L. H. Russell, "Some periodic solutions of a four-body problem. I," *Icarus*, vol. 3, pp. 423–428, 1964.
[7] J. Cronin, P. B. Richards, and I. S. Bernstein, "Some periodic solutions of a four-body problem. II," *Icarus*, vol. 9, pp. 281–290, 1968.
[8] C. Simó, G. Gómez, À. Jorba, and J. Masdemont, "The bicircular model near the triangular libration points of the RTBP," in *From Newton to Chaos (Cortina d'Ampezzo, 1993)*, vol. 336 of *NATO Advanced Science Institutes Series B: Physics*, pp. 343–370, Plenum, New York, NY, USA, 1995.
[9] A. Jorba, "On practical stability regions for the motion of a small particle close to the equilateral points of the real earth-moon system," in *Proceedings of the 3rd International Symposium (HAMSYS-98) Held at Pátzcuaro*, vol. 6 of *World Scientific Monograph Series in Mathematics*, pp. 197–213, World Scientific, River Edge, NJ, USA, 2000.
[10] A. F. B. De Almeida Prado, "Numerical and analytical study of the gravitational capture in the bicircular problem," *Advances in Space Research*, vol. 36, no. 3, pp. 578–584, 2005.
[11] A. L. Machuy, A. F. B. A. Prado, and T. J. Stuchi, "Numerical study of the time required for the gravitational capture in the bi-circular four-body problem," *Advances in Space Research*, vol. 40, no. 1, pp. 118–124, 2007.
[12] M. A. Andreu, "Dynamics in the center manifold around L_2 in the quasi-bicircular problem," *Celestial Mechanics & Dynamical Astronomy*, vol. 84, no. 2, pp. 105–133, 2002.
[13] T. J. Kalvouridis, M. Arribas, and A. Elipe, "Parametric evolution of periodic orbits in the restricted four-body problem with radiation pressure," *Planetary and Space Science*, vol. 55, no. 4, pp. 475–493, 2007.
[14] T. J. Kalvouridis and K. G. Hadjifotinou, "Bifurcations from planar to three-dimensional periodic orbits in the photo-gravitational restricted four-body problem," *International Journal of Bifurcation and Chaos*, vol. 18, no. 2, pp. 465–479, 2008.
[15] V. Szebehely, *Theory of Orbits*, Academic Press, New York, NY, USA, 1967.
[16] J. Lundberg, V. Szebehely, R. S. Nerem, and B. Beal, "Surfaces of zero velocity in the restricted problem of three bodies," *Celestial Mechanics*, vol. 36, no. 2, pp. 191–205, 1985.
[17] E. S. G. Leandro, "On the central configurations of the planar restricted four-body problem," *Journal of Differential Equations*, vol. 226, no. 1, pp. 323–351, 2006.
[18] C. L. Siegel and J. K. Moser, *Lectures on Celestial Mechanics*, vol. 18 of *Die Grundlehren der mathematischen Wissenschaften*, Springer, New York, NY, USA, 1971.
[19] R. F. Arenstorf, "Periodic solutions of the restricted three-body problem representing analytic continuations of Keplerian elliptic motions," *American Journal of Mathematics*, vol. 85, pp. 27–35, 1963.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 740460, 24 pages
doi:10.1155/2009/740460

Research Article

Nonsphericity of the Moon and Near Sun-Synchronous Polar Lunar Orbits

Jean Paulo dos Santos Carvalho,[1] Rodolpho Vilhena de Moraes,[1] and Antônio Fernando Bertachini de Almeida Prado[2]

[1] *UNESP—Universidade Estadual Paulista, CEP 12516-410, Guaratinguetá-SP, Brazil*
[2] *Division of Space Mechanics and Control—INPE, CEP 12227-010, São José dos Campos, SP, Brazil*

Correspondence should be addressed to Rodolpho Vilhena de Moraes, rodolpho@feg.unesp.br

Received 30 July 2009; Revised 28 October 2009; Accepted 3 December 2009

Recommended by Tadashi Yokoyama

Herein, we consider the problem of a lunar artificial satellite perturbed by the nonuniform distribution of mass of the Moon taking into account the oblateness (J_2) and the equatorial ellipticity (sectorial term C_{22}). Using Lie-Hori method up to the second order short-period terms of the Hamiltonian are eliminated. A study is done for the critical inclination in first and second order of the disturbing potential. Coupling terms due to the nonuniform distribution of mass of the Moon are analyzed. Numerical simulations are presented with the disturbing potential of first and second order is. It an approach for the behavior of the longitude of the ascending node of a near Sun-synchronous polar lunar orbit is presented.

1. Introduction

In this paper, we consider the problem of a lunar artificial satellite of low altitude taking into account the oblateness (J_2) and the equatorial ellipticity (sectorial term C_{22}) of the Moon. The Lie-Hori [1] perturbation theory method up to the second order is applied to eliminate the short-period terms of the disturbing potential. The perturbation method up to the second order is applied to analyze coupling terms. In this work, the long-period term of the disturbing potential is analyzed. A formula is developed to compute the critical inclination when the perturbations due to the nonsphericity of the Moon as a function of the terms of the zonal and sectorial harmonics occur.

An approach is done for a special type of orbit, denominated Sun-synchronous orbit of Moon's artificial satellites. The Sun-synchronous orbit is a particular case of an almost polar orbit. The satellite travels from the North Pole to the South Pole and vice versa, but its orbital

Table 1: Magnitude orders for J_2 and C_{22}.

	$C_{20} \equiv -J_2$	C_{22}
Earth	-10^{-3}	2×10^{-6}
Moon	-2×10^{-4}	2×10^{-5}

plane is always fixed for an observer that is posted in the Sun. Thus the satellite always passes approximately on the same point of the surface of the Moon every day in the same hour. In such a way the satellite can transmit all the data collected for a lunar fixed antenna, during its orbits. An analysis of Sun-synchronous orbits considering the nonuniform distribution of mass of the Moon is done for the longitude of the ascending node with an approach based on Park and Junkins [2].

In [3–5] an analytical theory, is developed to study the orbital motion of lunar artificial satellites using the method of transformation of Lie [6, 7] as a perturbation method. The main perturbation is due to the nonspherical gravitational field of the Moon and the attraction of the Earth. The disturbing body is in circular orbit with the disturbing function developed in polynomial of Legendre up to the second order. In [8–11] an analytical theory, is developed with numerical applications taking into account the nonuniform distribution of mass of the Moon and the perturbation of the third body in elliptical orbit (Earth is considered). The disturbing function is expanded in Legendre associated functions up to the fourth order.

This paper is developed based on [2, 4], where the perturbation theory method of Lie-Hori up to the second order is used. Our contribution is characterized by (a) We developed of a new formula for the critical inclination of second order; (b) we fix g and h to assure the condition of frozen orbits; (c) we showed that the coefficients J_2 and C_{22} affect the variation of the eccentricity strongly (it affects the eccentricity directly) in the second order contributing (especially the C_{22} term) to increasing the variation of the eccentricity mainly for small inclinations; (d) the coupled perturbations (nonuniform distribution of mass of the Moon (J_2 and C_{22}) and third-body ($P2$)) help to control the variation of the eccentricity for low-altitude polar orbits; (e) we presented a new formula to compute inclinations for Sun-synchronous orbits when it is taking into account the harmonic J_2 and C_{22} in the first-order potential.

This paper has seven sections. In Section 2, the terms due to nonsphericity of the Moon are presented while Section 3 is devoted to the Hamiltonian of the system. In Section 4, an approach concerning the critical inclinations is used. In Section 5, an approach concerning the Sun-synchronous lunar orbits is used. Numeric applications are done in Section 6. Section 7 is devoted to the conclusions.

2. Nonsphericity of the Moon

Besides the fact that the Moon is much less flattened than the Earth, it also causes perturbations in space vehicles. Table 1 presents orders of magnitude for some zonal and sectorial harmonics compared with the same parameters for the Earth. The term C_{20} describes the equatorial bulge of the Moon, often referred to as the oblateness. The coefficient C_{22} measures the ellipticity of the equator.

The space vehicle is a point of mass in a three-dimensional orbit with orbital elements: a (semimajor axis), e (eccentricity), i (inclination), ω (argument of periapsis), Ω (longitude of the ascending node), and n (mean motion) given by the third Kepler's law $n^2a^3 = Gm_0$, where m_0 is the mass of the Moon.

Then, we will present the Hamiltonian formalism using Delaunay canonical variables [7] defined as $L = \sqrt{\mu a}$, $G = L\sqrt{1-e^2}$, $H = G\cos i$, $l = M$ mean anomaly, $g = \omega$ argument of periapsis, and $h = \Omega$ longitude of the ascending node.

The force function, the negative of the total energy as used in physics, is given by

$$F = H = V - T, \tag{2.1}$$

here, V is the negative of the potential energy, and T is the kinetic energy. The force function can be put as [12]:

$$H = \frac{\mu}{r} + R - T = \frac{\mu}{2a} + R, \tag{2.2}$$

or

$$H = \frac{\mu^2}{2L^2} + R. \tag{2.3}$$

the function R, comprising all terms of V except the central term, is known as the disturbing potential. The term due to the unperturbed potential is given by

$$H_0 = \frac{\mu^2}{2L^2}. \tag{2.4}$$

Considering the lunar equatorial plane as the reference plane, the disturbing potential can be written in the form [13] of

$$V_M = -\frac{\mu}{r}\left[\sum_{n=2}^{5}\left(\frac{R_0}{r}\right)^n J_n P_n(\sin\phi) - \left(\frac{R_0}{r}\right)^2 C_{22}P_{22}(\sin\phi)\cos 2\lambda - \left(\frac{R_0}{r}\right)^3 C_{31}(\sin\phi)\cos\lambda\right], \tag{2.5}$$

where μ is the Lunar gravitational constant, R_0 is the equatorial radius of the Moon ($R_0 = 1738\,\text{km}$), P_n represent the Legendre polynomial, P_{nm} represent the associated Legendre polynomial, the angle ϕ is the latitude of the orbit with respect to the equator of the Moon, the angle λ is the longitude measured from the direction of the longest axis of the Moon, where $\lambda = \lambda' - \lambda_{22}$, since λ' is the longitude reckoned from any fixed direction, and λ_{22} is the longitude of the Moon's longest meridian from the same fixed direction. However, λ_{22} will contain the time explicitly (see [4, 13], for a detailed discussion). Using spherical trigonometry, we have $\sin\phi = \sin i \sin(f + g)$, where f is the true anomaly.

The following assumptions have been made [4, 14] and will be used in this work: (a) the motion of the Moon is uniform (librations are neglected); (b) the lunar equator lies in the ecliptic (we neglect the inclination of about 1.5° of the lunar equator to the ecliptic, and the inclination of the lunar orbit to the ecliptic of about 5°); (c) the longitude of the lunar longest meridian is equal to the mean longitude of the Earth (librations are neglected); (d) the mean longitude of the Earth, $\lambda_{\otimes}$, is equal to λ_{22}. See De Saedeleer [4] for a detailed discussion.

Since the variables Ω and $\lambda_{\otimes}$ appear only as a combination of $\Omega - \lambda_{\otimes}$, where $\lambda_{\otimes} = n_M t + \text{const}$, with n_M being the lunar mean motion, the degree of freedom can be reduced by choosing as a new variable $h = \Omega - \lambda_{\otimes}$. A new term must then be added to the Hamiltonian in order to get $\dot{h} = -\partial H/\partial H = n_M$. The Hamiltonian is still time-dependent through $\lambda_{\otimes}$. Since the longest meridian is always pointing toward the Earth, it is possible to choose a rotating system whose x-axis passes through this meridian.

Regarding the Earth's potential, the dominant coefficient is J_2. The rest are of higher-order terms [15]. In contrast to the Earth, the first harmonics of the Lunar potential are all almost of the same order (see Table 1). This fact complicates the choice of the harmonic where the potential can be truncated and this makes its choice a little arbitrary. The influence of the Earth and of the nonsphericity of the Moon on the stability of lunar satellites was also pointed out by [16] but sectorial harmonics were not considered. In terms of the orbital elements, the Legendre associated functions for zonal up to J_5 and sectorial terms C_{22} and C_{31}, where the values for the spherical harmonic coefficients are given in the Appendix C, can be written in the following form [13, 14]:

$$
\begin{aligned}
P_2(\sin\phi) &= \frac{1}{2}\left\{3\sin^2(i)\sin^2(f+g) - 1\right\},\\
P_3(\sin\phi) &= \frac{5}{2}\sin^3(i)\sin^3(f+g) - \frac{3}{2}\sin(i)\sin(f+g),\\
P_4(\sin\phi) &= \frac{35}{8}\sin^4(i)\sin^4(f+g) - \frac{15}{4}\sin^2(i)\sin^2(f+g) + \frac{3}{8},\\
P_5(\sin\phi) &= \frac{63}{8}\sin^5(i)\sin^5(f+g) - \frac{35}{4}\sin^2(i)\sin^3(f+g) + \frac{15}{8}\sin(i)\sin(f+g),\\
P_{22}(\sin\phi) &= 3 - 3\sin^2(i)\sin^2(f+g),\\
P_{22}(\sin\phi)\cos 2\lambda &= 6\left(\xi^2\cos^2(f) + \chi^2\sin^2(f) + 2\xi\chi\sin(2f) - 3\left(1 - \sin^2(i)\sin^2(f+g)\right)\right),\\
P_{31}(\sin\phi)\cos\lambda &= \left(\frac{15}{8}s^2 - \frac{3}{2}\right)\cos(h)\cos(f+g) + \left(\frac{-3}{4} - \frac{15}{8}s^2\right)c\sin(h)\sin(f+g)\\
&\quad -\frac{15}{8}s^2\cos(h)\cos(3f+3g) + \frac{15}{8}s^2 c\sin(h)\sin(3f+3g),
\end{aligned}
\tag{2.6}
$$

where

$$
\begin{aligned}
\xi &= \cos(g)\cos(h) - \cos(i)\sin(g)\sin(h),\\
\chi &= -\sin(g)\cos(h) - \cos(i)\cos(g)\sin(h),\\
s &= \sin(i), \qquad c = \cos(i).
\end{aligned}
\tag{2.7}
$$

In this paper, are taken into only the terms due to J_2 and C_{22} accout. The zonal perturbation due to the oblateness J_2 is defined by [3] $H_{20} = \varepsilon(\mu/r^3)P_{20}(\sin\phi)$, where $\varepsilon = J_2R_0^2$. However, the disturbing potential is

$$H_{20} = \varepsilon\frac{\mu}{4r^3}\Big(1 - 3\cos^2(i) - 3\sin^2(i)\cos(2f + 2g)\Big). \tag{2.8}$$

Substituting the relation $\mu = n^2a^3$, using the Cayley's tables [17] to express the true anomaly in terms of the mean anomaly, and with some algebraic manipulations, we get

$$\begin{aligned} H_{20} = \frac{3}{8}\varepsilon\Big(\Big(\Big(5e^2 - 2\Big)\cos^2(i) + 2 - 5e^2\Big)\cos(2g + 3l) + \Big(7e - 7e\cos^2(i)\Big)\cos(2g + 3l) \\ + \Big(-17e^2\cos^2(i) + 17e^2\Big)\cos(2g + 4l) + \Big(-e + e\cos^2(i)\Big)\cos(2g + l) \\ +9\Big(\cos^2(i) - \frac{1}{3}\Big)\Big(\frac{2}{9} + \frac{2}{3}e\cos(l) + \frac{1}{3}e^2 + e^2\cos(2l)\Big)\Big)n^2. \end{aligned} \tag{2.9}$$

For the sectorial perturbation, we get [13, 14]

$$H_{22} = \delta\frac{\mu}{r^3}\Big(6\xi^2\cos^2(f) + 6\chi^2\mathrm{sen}^2(f) + 12\xi\chi\mathrm{sen}(2f) - 3 + 3s^2\mathrm{sen}^2(f + g)\Big), \tag{2.10}$$

where $\delta = C_{22}R_0^2$ (R_0 is the equatorial radius of the Moon; $R_0 = 1738\,\mathrm{km}$).

With some manipulations, we get

$$\begin{aligned} H_{22} = -\frac{45}{16}\delta n^2 \\ \times \Big[\Big(e^2 - \frac{2}{5}\Big)(\cos i - 1)^2\cos(2l + 2g - 2h) - \frac{1}{3}(\cos i + 1)^2\Big(e^2 - \frac{2}{5}\Big)\cos(2l - 2g - 2h) \\ + (\cos i + 1)^2\Big(e^2 - \frac{2}{5}\Big)\cos(2l + 2g + 2h) - \frac{1}{3}\Big(e^2 - \frac{2}{5}\Big)(\cos i - 1)^2\cos(2l - 2g + 2h) \\ - \frac{17}{5}e^2(\cos i - 1)^2\cos(4l + 2g - 2h) - \frac{7}{5}e(\cos i - 1)^2\cos(3l + 2g - 2h) \\ + \frac{17}{15}e^2(\cos i + 1)^2\cos(4l - 2g - 2h) + \frac{7}{15}e(\cos i + 1)^2\cos(3l - 2g - 2h) \\ - \frac{7}{5}e(\cos i + 1)^2\cos(3l + 2g + 2h) - \frac{17}{5}e^2(\cos i + 1)^2\cos(4l + 2g + 2h) \\ + \frac{7}{15}e(\cos i - 1)^2\cos(3l - 2g + 2h) + \frac{17}{15}e^2(\cos i - 1)^2\cos(4l - 2g + 2h) \\ - \frac{1}{15}e(\cos i + 1)^2\cos(l - 2g - 2h) + \frac{1}{5}e(\cos i - 1)^2\cos(l + 2g - 2h) \end{aligned}$$

$$+\frac{1}{5}e(\cos i+1)^2\cos(l+2g+2h)-\frac{1}{15}e(\cos i-1)^2\cos(l-2g+2h)$$
$$-\frac{2}{3}\sin^2 i\left(e^2-\frac{2}{5}\right)\cos(-2g+2l)+\frac{9}{5}e^2\cos(2l-2h)$$
$$-\frac{7}{5}e\cos(-2g+3l)+\frac{1}{5}e\cos(-2g+l)+\frac{9}{5}e^2\cos(2l+2h)+\frac{6}{5}\cos(2h)e^2$$
$$+\frac{4}{5}\cos(2h)-\frac{17}{5}e^2\cos(-2g+4l)+\frac{6}{5}e\cos(l+2h)+\frac{6}{5}e\cos(l-2h)\Bigg], \tag{2.11}$$

where the disturbing potential is written in the form $R = H_{20} + H_{22}$.

3. The Hamiltonian System

We find in the literature several papers that use the method of the average to calculate perturbations of long-period on artificial satellites of the Moon, such as [18–23]. However, our objective here is to compute analytically secular and periodic perturbations up to the second order and, using this, to analyze the coupling terms relating the harmonic coefficients. The Lie-Hori [1] perturbation method is applied to eliminate short-period terms of the Hamiltonian.

In [24], a different approach is proposed for the canonical version of Hori method. The reference [24] showed that the ordinary differential equation with an auxiliary parameter t^* as independent variable, introduced through Hori auxiliary system, can be replaced by a partial differential equation in time t.

In what follows, first the Lie-Hori [1] method will be shortly presented and then applied to the problem of the orbital motion of the satellite around the Moon.

Consider the mth order equation of the algorithm of the perturbation method proposed by Hori [1]:

$$\{H_0, S_m\} + \Psi_m = H_m^*, \tag{3.1}$$

where braces stand for the Poisson brackets, H_0^* is the undisturbed Hamiltonian, and Ψ_m is a function obtained from the preceding orders, involving H_0^*, H_m, S_k, H_k^*, and H_k, $k = 1, \ldots, m-1$. All of these functions are written in terms of the new set of canonical variables (ξ, η) and defined through the following equations:

$$H(x,y) = H_0(x,y) + \sum_{k=1} \varepsilon^k H_k(x,y),$$
$$H^*(\xi,\eta) = H_0^*(\xi,\eta) + \sum_{k=1} \varepsilon^k H_k(\xi,\eta), \tag{3.2}$$
$$\varepsilon S(\xi,\eta) = \sum_{k=1} \varepsilon^k S_k(\xi,\eta),$$

where (x, y) is the original set of canonical variables, $H(x, y)$ is the original Hamiltonian, $H^*(\xi, \eta)$ the new Hamiltonian and $S(\xi, \eta)$ is the generating function of the canonical transformation, $(x, y) \rightarrow (\xi, \eta)$. The transformation is such that the new canonical system has some advantages for the solution.

In order to determine the functions S_m and H_m^*, Hori introduces an auxiliary parameter t^* through the following system of canonical equations [1]:

$$\frac{d\xi_i}{dt^*} = \frac{\partial H_0^*}{\partial \eta_i}, \qquad \frac{d\eta_i}{dt^*} = -\frac{\partial H_0^*}{\partial \xi_i}, \quad i = 1, \ldots, n. \tag{3.3}$$

Accordingly, (3.1) reduces to

$$\frac{dS_m}{dt^*} = \Psi_m - H_m^*, \tag{3.4}$$

with Ψ_m written in terms of the general solution of the system (3.3), involving $2n$ constants of integration. Equation (3.4) has two unknown functions: S_m and H_m^*.

The Poisson brackets are defined as

$$\{x, y\} = \frac{\partial x}{\partial L}\frac{\partial y}{\partial l} - \frac{\partial x}{\partial l}\frac{\partial y}{\partial L} + \frac{\partial x}{\partial G}\frac{\partial y}{\partial g} - \frac{\partial x}{\partial g}\frac{\partial y}{\partial G} + \frac{\partial x}{\partial H}\frac{\partial y}{\partial h} - \frac{\partial x}{\partial h}\frac{\partial y}{\partial H} \tag{3.5}$$

with respect to the classical Delaunay variables set l, g, h, L, G, H. Since only l, g, h, L are explicitly present in the Hamiltonian, the partial derivatives with respect to L, G, H are computed as $\partial/\partial L = (\partial/\partial L) + (\eta^2/eL)(\partial/\partial e)$, $(\partial/\partial G) = -(\eta/eL)(\partial/\partial e) + (c/L\eta s)(\partial/\partial i)$, $(\partial/\partial H) = -(1/L\eta s)(\partial/\partial i)$ where the bracket indicates the derivative with respect to L occurring explicitly, and c, s, η are $c = \cos i$, $s = \sin i$, $\eta = \sqrt{1 - e^2}$.

To solve (3.4), we separated in secular and periodic part as it is done by the principle of the mean [7]. Using (3.3) and (3.4), Hori [1] writes the equations that define the integration theory based on average principle to determine the new Hamiltonian and the generating function as follows.

Order zero:

$$H_0^* = H_0. \tag{3.6}$$

Order one:

$$\begin{aligned} H_1^* &= H_{1s}, \\ S_1 &= \int H_{1p}\, d\tau. \end{aligned} \tag{3.7}$$

Order two:

$$H_2^* = H_{2s} + \frac{1}{2}\{H_1 + H_1^*, S_1\}_s,$$
$$S_2 = \int \left(F_{2p} + \frac{1}{2}\{H_1 + H_1^*, S_1\}_p \right) d\tau, \tag{3.8}$$
$$\cdots$$

where, at each function, the index s represents the secular part of the function and the index p the periodic part of the function. See Hori [1] for a detailed discussion.

The Hamiltonian of the dynamical system associated to the problem of the orbital motion of the satellite around the Moon can be written in the following form:

$$H = H_0 + H_{20} + H_{22}, \tag{3.9}$$

where

$$H_0 = \frac{\mu^2}{2L^2} + n_M H, \tag{3.10}$$

The term $n_M H$ is added to reduce the degree of freedom, since the mean longitude of the Earth is time-dependent [14]. Here, the term $n_M H$ is taken as order zero as suggested by Breiter [25].

Now, doing

$$H_{20} = \varepsilon H_1,$$
$$H_{22} = \delta H_2, \tag{3.11}$$

we get,

$$H = H_0 + \varepsilon H_1 + \delta H_2. \tag{3.12}$$

With the purpose of applications of the perturbation method, the terms of the Hamiltonian are written in the following form:

$$H_0^{(0)} = \frac{\mu^2}{2L^2} + n_M H,$$
$$H_1^{(1)} = \varepsilon H_1 + \delta H_2. \tag{3.13}$$

The disturbing terms are represented in the first order of the applied method. To eliminate the short-period terms of equation (3.13), the method of Lie-Hori [1] perturbation theory is applied. In this work long-period terms are calculated, substituting the result in the planetary equations of Lagrange [26]. The equations of motion are integrated and finally

the results analyzed. With a simplified model for the disturbing potential it is possible to do analyses for the orbital motion of the satellite.

Applying the method of Hori [1] to our problem to eliminate the terms of short period, we get the following:

Order zero:

$$H_0^* = H_0 = \frac{\mu^2}{2L^2} + n_M H. \tag{3.14}$$

Order one:

$$\begin{aligned} H_1^* &= \left(H_1^{(1)}\right)_s = \frac{1}{2\pi}\int_0^{2\pi} H_1^{(1)} dl, \\ S_1 &= \int H_{1p} d\tau = -\frac{1}{\beta}\int \left(H_1^{(1)} - H_1^*\right) dl, \end{aligned} \tag{3.15}$$

where $\beta = \partial H_0^{(0)}/\partial L$.

Order two:

$$H_2^* = \frac{1}{2}\left\{H_1^{(1)} + H_1^*, S_1\right\}_s = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{2}\left\{H_1^{(1)} + H_1^*, S_1\right\} dl. \tag{3.16}$$

4. Critical Inclination

We consider now the problem of a lunar artificial satellite with low altitude taking into account the oblateness (J_2) and the equatorial ellipticity (sectorial term C_{22}) of the Moon. The first order long-period disturbing potential (order of the method of perturbation theory) obtained by the Hori method algorithm can be written as

$$\begin{aligned} k1 = \frac{1}{8}n^2\Big(&6\varepsilon\cos^2(i) - 3\varepsilon e^2 - 2\varepsilon - 18\delta\cos(2h)e^2 + 18\delta\cos(2h)e^2\cos^2(i) - 12\delta\cos(2h) \\ &+12\delta\cos(2h)\cos^2(i) + 9\varepsilon\cos^2(i)e^2\Big), \end{aligned} \tag{4.1}$$

where $\varepsilon = J_2 R_0^2 e\delta = C_{22}R_0^2$ and $k1 = H_1^*$.

We observe that, at the order considered, the disturbing potential has the terms due to the oblateness (ε), that are secular, and terms due to the equatorial ellipticity of the Moon (δ), that appear multiplied for $\cos(2h)$.

Taking into account (4.1) a formula for the critical inclination is found. In fact, substituting (4.1) in the planetary equations of Lagrange [26] and solving the equation $dg/dt = 0$, we get

$$\cos^2(i) = \frac{-\varepsilon + 6\delta\cos(2h)}{5(-\varepsilon + 2\delta\cos(2h))}. \tag{4.2}$$

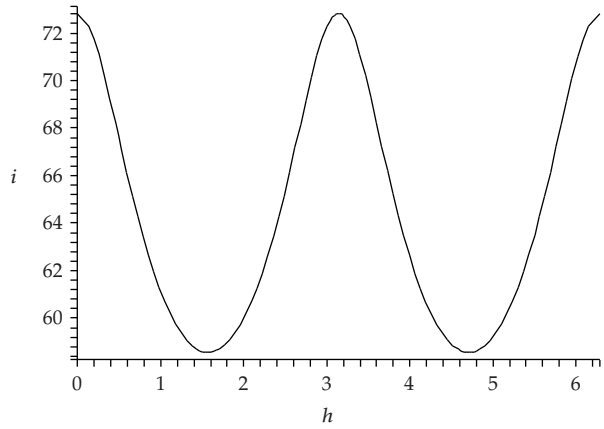


Figure 1: Variation of the critical inclination with respect to the longitude of the ascending node where *i*-degree and *h*-rad.

Table 2: Critical inclination for the potential of first order, where: $\varepsilon = 613.573$; $\delta = 67.496$.

Longitude of the ascending node (h)	Critical inclination for direct orbits (i_c)	Critical inclination for retrograde orbits (i_c)
1 rad	61.10°	118.90°
2 rad	59.98°	120.02°
$\pi/2$	58.56°	121.45°
$\pi/3$	60.69°	119.31°
π	72.83°	107.17°

this formula was already obtained by De Saedeleer and Henrard [5] and was here derived in an independent way, observing that in [5] $\delta = -C_{22}R_0^2$. Thus, when we consider the terms due to the oblateness (J_2) and the equatorial ellipticity of the Moon (C_{22}), the critical inclination depends on the longitude of the ascending node. Figure 1 represents the variation between the inclination and the longitude of the ascending node. Table 2 represents the values of the critical inclination for some values of the ascending node.

Now, let us consider the second-order disturbing potential $k2 = H_2^*$ where is given in the Appendix A (order of the method of perturbation theory); the potential $k2$ presents:

(a) coefficients of second order (J_2^2, C_{22}^2),

(b) coupling terms between J_2 and C_{22}.

Plugging the equations for the potential in the planetary equations of Lagrange and solving the equation $dg/dt = 0$, we present a new formula to compute the critical inclination taking into account the J_2 and C_{22} terms of the second-order disturbing potential. It is a function of two variables: the argument of the periapsis (g) and the longitude of the ascending node (h). When the sectorial term C_{22} is considered, the first order disturbing potential is a function of the longitude of the ascending node and of both longitude of the ascending node and argument of the periapsis to the second order potential. The new formula is given in the Appendix B.

We observe that, at the second order, the disturbing potential presents terms due to the oblateness (ε) and to the equatorial ellipticity of the Moon (δ), that also appear multiplied by periodic functions. Here, terms of couplings between the oblateness and the equatorial ellipticity of the Moon (J_2, C_{22}) and terms of second order appear. Several scenarios can be considered. For instance frizzing orbits with particular values of h and g, let us say $h = \pi/2$ and $g = 3\pi/2$, we get a value of 53.46° for the critical inclination taking into account equation given in the Appendix B. Therefore, the critical inclination taking into account equation (4.2) is 58.56° (see Table 2).

5. Sun-Synchronous Lunar Orbit

Now, an approach is presented for a Lunar Sun-synchronous orbit. The Moon rotates with angular rate about 360° for 27, 32 days, while the Earth rotates with angular rate about 360° by day. The perturbation caused by the orbital precession has been studied historically for orbits centered in the Earth because of near polar orbits the precession is about one degree per day and to provide attractive Sun-synchronous orbits for many missions around the Earth. Considering Sun-synchronous orbits for lunar satellites we show that it is not possible to produce near polar orbits. The precession of the ascending node due to the nonsphericity of the Moon, when only the effect of the J_2 is considered, in (3.3) is well known in Brouwer theory [27] that the precession of the longitude of the ascending node is given by

$$\frac{d\Omega}{dt} = -\frac{3}{2}\frac{J_2 R_0^2 n \cos i}{a^2(1-e^2)^2}. \tag{5.1}$$

The Moon's orbital period is about 27, 32 days and the Earth's orbital period is about 365, 26 days. Then, for a Sun-synchronous orbit, we have, in lunar day [2]:

$$\frac{d\Omega}{dt} = \left(\frac{27,35}{365,26}\right)360^\circ/\text{lunar day} = 26,92657^\circ/\text{lunar day}. \tag{5.2}$$

An inclination for a Sun-synchronous orbit was presented by Park and Junkins [2] using (5.1) and (5.2), with the following initial conditions: $a = 1837.63\,\text{km}$; $e = 0$. The calculated inclination is $i_s = 144.82^\circ$. This inclination is not feasible for producing near-polar orbits, because this orbit does not pass sufficiently near the poles. Considering the disturbing potential given by (4.1) and substituting in the Lagrange planetary equations [26] to calculate the variation of the longitude of the ascending node, we get

$$\frac{d\Omega}{dt} = -\frac{3}{2}\frac{\varepsilon n}{a^2(1-e^2)^2}\cos(i) + \frac{3n\delta}{2a^2\left(\sqrt{1-e^2}\right)}\cos(i)\left[\left(2+3e^2\right)\cos(2h)\right], \tag{5.3}$$

where $\varepsilon = J_2 R_0^2$, $\delta = C_{22} R_0^2$. This new equation (5.3) gives the precession of the longitude of the ascending node due to nonsphericity of the Moon when the effect of the J_2 and the C_{22} are considered. In first approximation the periodic terms due to the J_2 are negligible, however, for the C_{22} term in the first approximation appears the periodic term $\cos(2h)$. When $\delta = 0$, we obtain the classic solution given by (5.1).

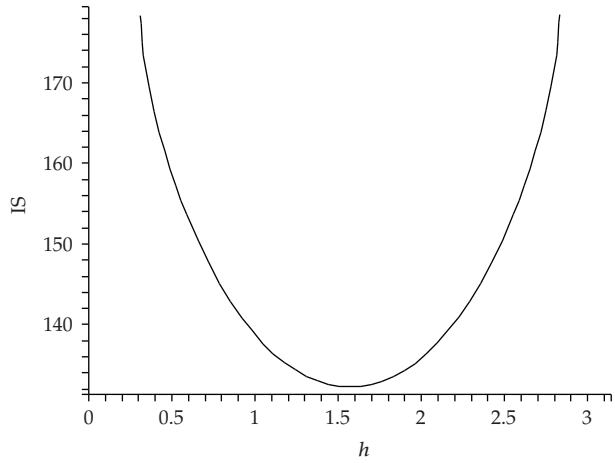


Figure 2: Variation of the inclination Sun-synchronous (IS in degree) with longitude of the ascending node (h-rad), $a_0 = 1838\,\text{km}$, $e_0 = 0.038$.

In what follows, the variation of the longitude of the ascending node will be analyzed to obtain polar or near-polar orbits for some special cases. Using (5.2) and (5.3) we get a new formula to compute the inclination of Sun-synchronous orbits for Moon's satellites of low altitude:

$$i_s = \pi - \arccos\left(\frac{1.32730740910^{-7} a^2 \left(1-e^2\right)^{(3/2)}}{n\left(\varepsilon + \delta\left(-2-e^2+3e^4\right)\cos(2h)\right)}\right), \tag{5.4}$$

where n is given in rad/s, $\delta = 67.496\,\text{km}^2$; $\varepsilon = 613.573\,\text{km}^2$. Equation (5.4) gives the inclination depending on the semi-major axis, the eccentricity and on the longitude of the ascending node. For Sun-synchronous orbits, considering $a = 1837.63\,\text{km}$; $e = 0$; $h = \pi/2$ the calculated inclination is $i_s = 132.35°$. This inclination is not also ideal for near-polar orbits. Thus, the obtained results are still distant from a polar Sun-synchronous orbit, but it is important to consider the term due to the Moon's equatorial ellipticity to get more realistic results. Figure 2 represents the variation of Sun-synchronous inclinations with respect to the longitude of the ascending node where $a_0 = 1838\,\text{km}$ and $e_0 = 0.038$. It can be observed in Figure 2 that, fixing $h = \pi/2$, we get an inclination of about 132°.

6. Applications for Low Altitude Satellites

The disturbing potential (first order ($k1$) and second order ($k2$)) is substituted in the Lagrange's planetary equations [26] and numerically integrated. Considering the disturbing potential due to the nonsphericity of the Moon (J_2 and C_{22}), numerical applications (long-period potential) are performed to analyze the variation of the eccentricity for different initial conditions of the inclination.

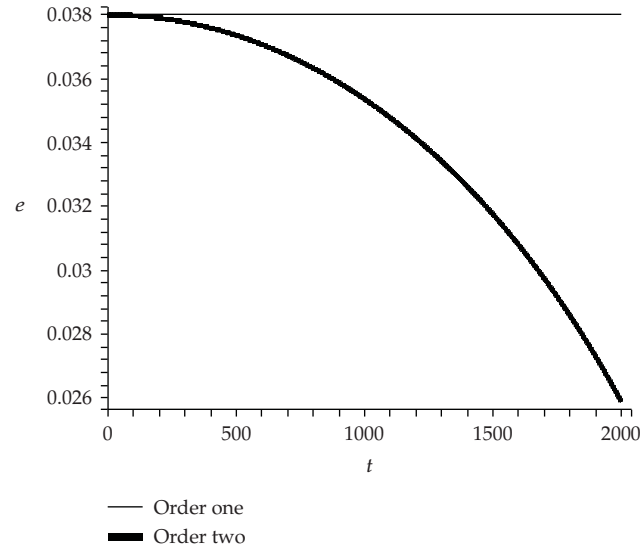


Figure 3: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038$, $i_0 = 30^\circ$, $g_0 = 3\pi/2$, $h_0 = \pi/2$ and t-days.

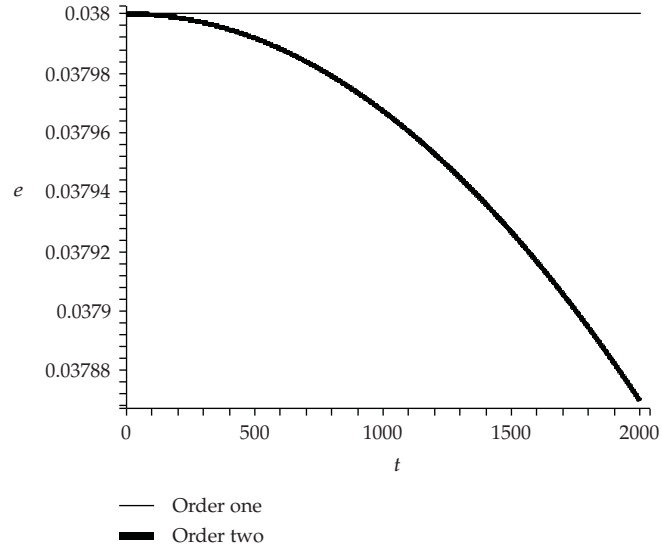


Figure 4: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038$, $i_0 = 90^\circ$, $g_0 = 3\pi/2$, $h_0 = \pi/2$ and t-days.

Figures 3 and 4 represent the comparison for different orders of the disturbing potential. The variation of the eccentricity for lunar satellites in low altitude is constant at first order. This happens because the coefficients J_2 and C_{22} do not affect the variation rate of the eccentricity (as we can verify in Lagrange's planetary equations). Therefore, it is important to insert more terms in the potential to get more realistic results as, for example, to study the lifetimes of low altitude Moon artificial satellites [28], considering the zonal terms J_2, J_3 and J_5 and the sectorial terms C_{22} and C_{31} in the disturbing potential. When taking into account the disturbing potential of second order considering the effect

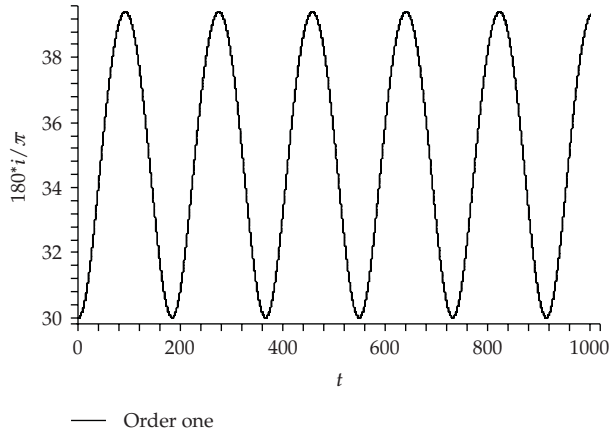


Figure 5: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038, i_0 = 30°, g_0 = 3\pi/2, h_0 = \pi/2$ and t-days. Fixing a value $h = \pi/2$ we find a value of 58.56° for the critical inclination.

of the nonuniform distribution of mass of the Moon (J_2 and C_{22}), the coefficients J_2 and C_{22} affect the variation rate of the eccentricity (as we can verify in the Lagrange's planetary equations). For the second order the results shows a small variation of the eccentricity for larger inclinations and an accentuated increase for small inclinations. This is due to the C_{22} term that affects the eccentricity of the satellite directly in second order.

The expression of the eccentricity is presented in the following form $de/dt = \cdots \varepsilon\delta \sin(2g) \cdots + \varepsilon^2 \sin(2g) + \cdots + \delta^2 \sin(2g + 2h) \cdots - \cdots \varepsilon\delta \sin(-2g + 2h) \cdots + \cdots$ where the terms due to the oblateness (ε) and the equatorial ellipticity (δ) appear multiplied by periodic functions, terms of couplings between J_2 and C_{22} and terms of second order of the type J_2^2 and C_{22}^2. Figures 3 and 4 shows the temporal variation of the eccentricity for cases where initial conditions are obtained from the frozen orbits condition. For instance frizzing orbits with particular values of h and g, let us say $h = \pi/2$ and $g = 3\pi/2$. Another factor that also contributes for the variation of the eccentricity, using the potential up to the second order, is the presence of the coefficients of second order terms (J_2^2, C_{22}^2) and the coupling terms between J_2 and C_{22}. The choice of the initial inclination is very important to assure a frozen orbit when it is taken into account the second order disturbing potential.

Figures 5 and 6 show the inclination suffering a periodic variation that depends of the longitude of the ascending node, as we can verify by (4.2). Figure 5 shows a variation of the inclination for a value below of the critical inclination around 8 degrees in 100 days, while Figure 6 presents a more accentuated variation for values of the inclination above of the critical inclination around 50 degrees in 200 days.

Considering the second-order disturbing potential, Figures 7 and 8 represent the variation rate for the inclination. The variation represented is due to the initial values given for the argument of the periapsis and for the longitude of the ascending node. The given initial values for g and h are for the condition of frozen orbits. For instance frizzing orbits with particular values of h and g, let us say $h = \pi/2$ and $g = 3\pi/2$. The same comments done for the eccentricity, including those for the coupling terms (zonal and sectorial), are valid for the variation of the orbital inclination.

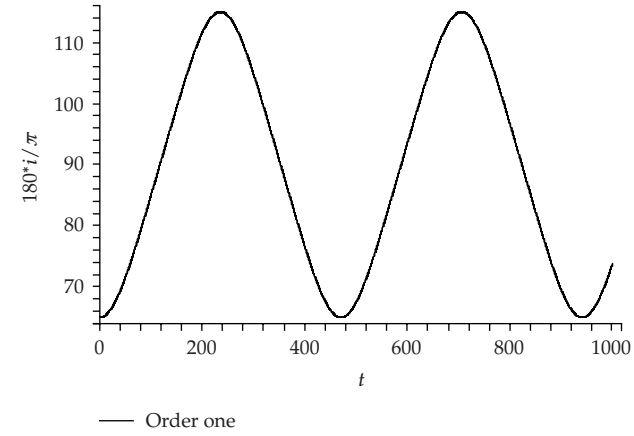


Figure 6: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038$, $i_0 = 65^\circ$, $g_0 = 3\pi/2$, $h_0 = \pi/2$ and t-days. Fixing a value $h = \pi/2$ we find a value of 58.56° for the critical inclination.

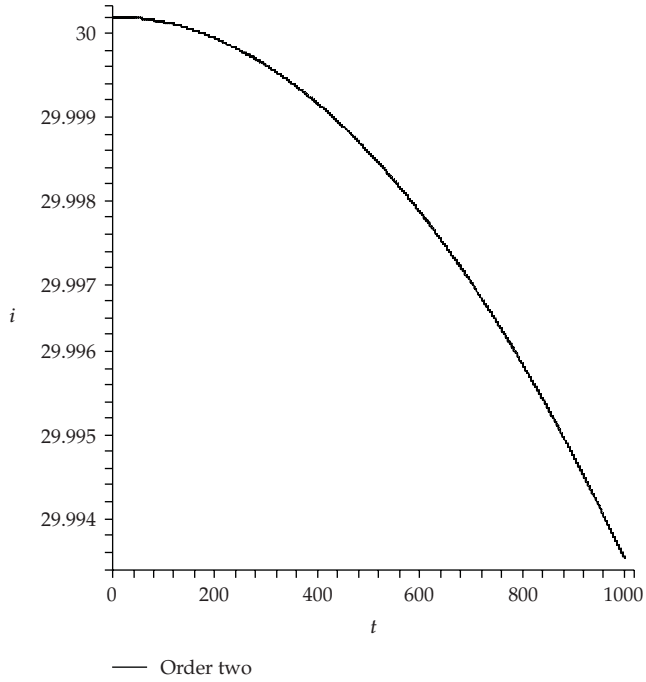


Figure 7: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038$, $i_0 = 30^\circ$, $g_0 = 3\pi/2$, $h_0 = \pi/2$ and t-days.

Numerical applications with the first and second order disturbing potential are done taking into account the nonsphericity of the Moon and perturbations from the third-body in elliptical orbit (Earth is considered) considering the term P_2 of the Legendre polynomial and the eccentricity of the disturbing body up to the second order. Figures 9, 10 and

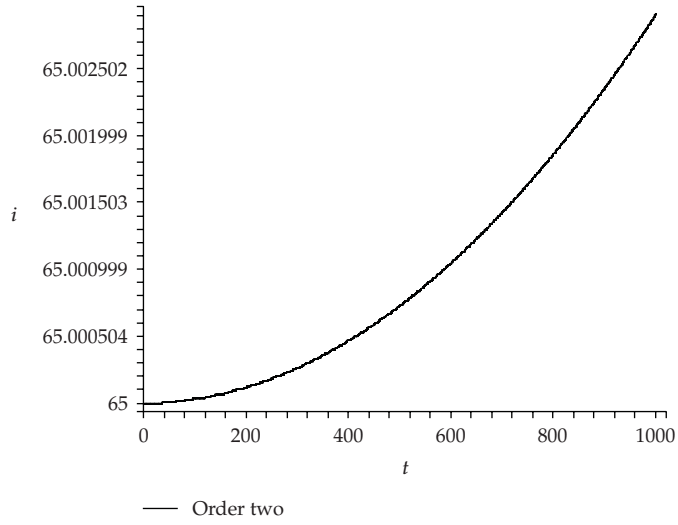


Figure 8: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038$, $i_0 = 65°$, $g_0 = 3\pi/2$, $h_0 = \pi/2$ and t-days.

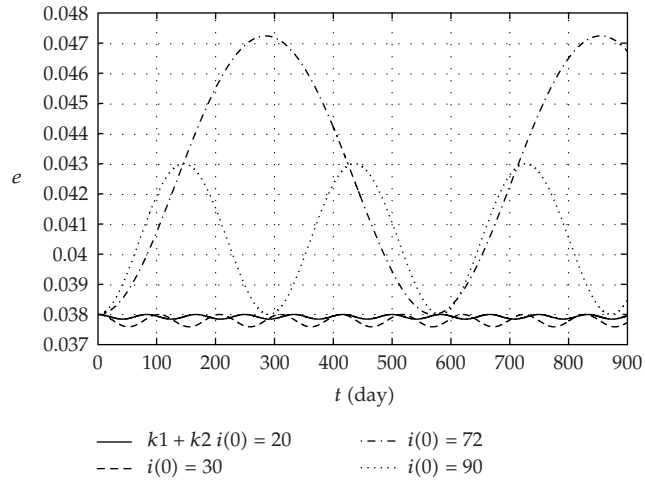


Figure 9: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038$, $g = 3\pi/2$, $h = \pi/2$, $e_{\max} = 0.05$. Considered Perturbations: $P_2, J_2, C_{22} = 0$.

11 represent the sum of the potential of the first order with the second order ($k1 + k2$). Figure 9 shows that, if $J_2 \neq 0$ and $C_{22} = 0$ the small inclinations cause small oscillations in the variation of the eccentricity and Figure 10 shows the effect caused by the C_{22} term, when $J_2 = 0$, where the small inclinations cause a large increase in the variation of the eccentricity, as well as we can visualize in Figure 11, where it is taken into account the P_2, J_2, and C_{22} terms. Considering the disturbing potential up to second order, the terms J_2 and C_{22} that does not cause perturbation in the eccentricity in first order, appear as

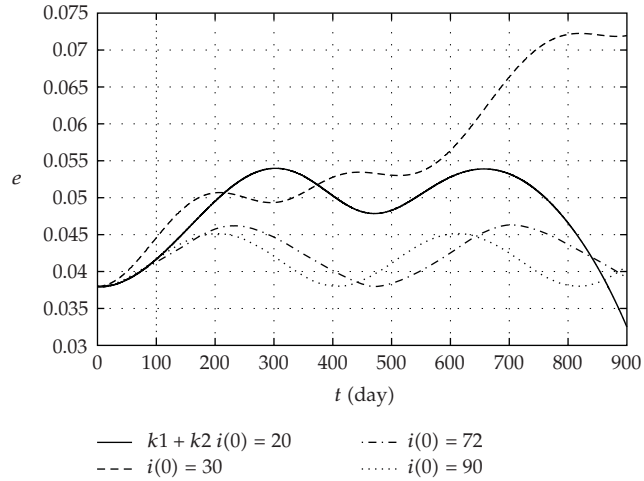


Figure 10: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038, g = 3\pi/2, h = \pi/2, e_{\max} = 0.05$. Considered Perturbations: P_2, J_2, C_{22}.

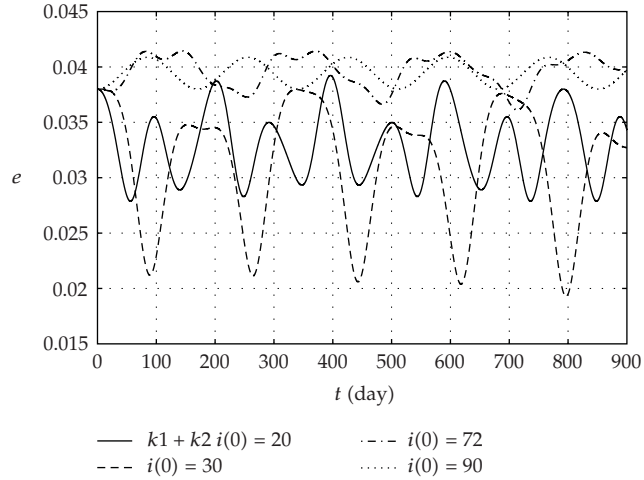


Figure 11: Initial conditions: $a_0 = 1838\,\text{km}$, $e_0 = 0.038, g = 3\pi/2, h = \pi/2, e_{\max} = 0.05$. Considered Perturbations: P_2, J_2, C_{22}.

disturbing term of the type J_2^2 and C_{22}^2 and terms of coupling of the type J_2C_{22} that affects the eccentricity of the satellite directly. The terms appear due to the perturbation method used.

The temporal variation of the eccentricity is strongly affected by the initial inclination (i_0). As it can be observed by Figure 11, for $i_0 < 48.6°$ the variation of the eccentricity presents great amplitude but, for $i_0 > 48.6°$, the variation has small amplitude.

Figures 12 and 13 represent the variation of the eccentricity for a lunar satellite of low altitude considering different terms in the disturbing potential. Figure 12 considered

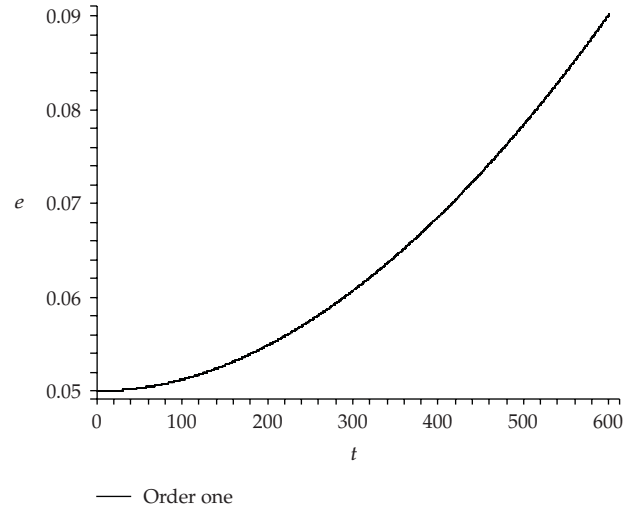


Figure 12: Initial conditions: $a_0 = 1935.79\,\text{km}$, $e_0 = 0.05, i = 90^\circ, g = 3\pi/2, h = \pi/2$, $e_{\max} = 0.09$ and t-days. Perturbations P_2.

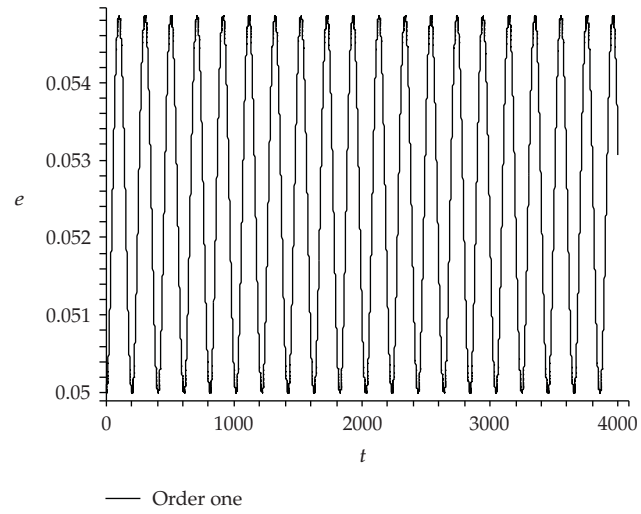


Figure 13: Initial conditions: $a_0 = 1935.79\,\text{km}$, $e_0 = 0.05, i = 90^\circ, g = 3\pi/2, h = \pi/2$, $e_{\max} = 0.09$ and t-days. Perturbations $P_2 + J_2 + C_{22}$.

third body perturbation and Figure 13 considered nonsphericity of the Moon and third body (P_2) perturbation. A comparison is done between the perturbations for the case where the inclination is 90° (polar orbit). Therefore we can conclude that, besides the term due to the J_2, the sectorial term C_{22} should also be considered in the disturbing potential to get more realistic results. For lunar satellites of low altitude it is impracticable to consider real applications taking into account only the perturbation of the third body (P_2). In fact, taking into account only the perturbation of the third body (P_2), the eccentricity of the satellite increases causing escape from the Moon, or crashing to the Moon in 600 days (see Figure 12).

We observe that, for a Moon's artificial satellite orbiting in low altitude, the combination of the two perturbations help to control the variation of the eccentricity, mainly for inclinations of 90° (see Figure 13).

These results agree with those obtained by d'Avanzo et al. [29] when it is considered just the effects of the zonal harmonics J_2–J_5 and a first order potential.

7. Conclusions

Using Lie-Hori method, the disturbing potential due to the nonsphericity of the gravitational field of the Moon is obtained up to first and second order. The disturbing potential is substituted in Lagrange's planetary equations and they are numerically integrated. Analyses for the variations of the orbital elements are done. Terms of couplings between the oblateness and the equatorial ellipticity of the Moon (J_2, C_{22}) and terms of second order of type J_2^2 and C_{22}^2 are obtained. A formula is developed to compute the critical inclination when the effect of the C_{22} (equatorial ellipticity) term is considered in the Hamiltonian in first and second order. The critical inclination can be strongly affected by the coefficient due to the equatorial ellipticity of the Moon and by the longitude of the ascending node. The formula for the critical inclination for the second order is a function of two variables: the argument of the periapsis and the longitude of the ascending node. At the first order this formula is a function of the longitude of the ascending node only. For Lunar low altitude satellites (LLAS), it is important to take into account both, the terms due to the oblateness and terms with the equatorial ellipticity of the Moon to get more realistic results.

The variation of the longitude of the ascending node is analyzed looking Lunar Sun-synchronous orbits and near-polar orbits. A new formula is obtained to compute inclinations of Lunar Sun-synchronous orbits when the terms due to the oblateness of the Moon (J_2 and C_{22}) are taken into account. The presented formula to the inclination depends on the semi-major axis, eccentricity and on the longitude of the ascending node of the satellite. The term due to the effect of the C_{22} must be considered for the case of a lunar satellite to analyze the precession of the longitude of the ascending node.

For a LLAS, when it is considered only the nonsphericity of the Moon in the disturbing potential, and at the first order, the orbital eccentricity of the satellite is constant along the time. This happens since the coefficients J_2 and C_{22}, at this order, do not affect the variation rate of the eccentricity directly, therefore it is important to insert more terms in the potential as, for example, the zonal terms J_3, J_5 and the sectorial term C_{31}, to get more realistic results. At the second order, small variations are present. In this case, the small variations of the eccentricity are due to a combination of the following factors: (a) initial conditions (given to get frozen orbits) (b) to couplings terms between the oblateness and the equatorial ellipticity of the Moon (J_2, C_{22}) and (c) terms of second order of type J_2^2 and C_{22}^2. For small inclinations, second-order terms (including coupled terms) are greater than 1st order terms. This happens because the coefficients J_2 and C_{22}, at second order, affect the variation of the eccentricity directly.

To study lifetime of LLAS, due to the characteristics of the mass distribution of the Moon, it is necessary to take into account up the second order of the disturbing potential and develop up to the second order the Lie-Hori algorithm. In fact, at first order, the coefficients do not affect the eccentricity directly while at the second order the coefficients J_2 and C_{22} affects the eccentricity directly and thus contributing efficiently to more complete studies.

Appendices

A. Appendix A

Let us consider the second-order disturbing potential $k2 = H_2^*$ (order of the method of perturbation theory), where $k2 = H_2^*$, $c = \cos i$, $s = \sin i$:

$$
\begin{aligned}
k2 = -\frac{1}{16a^2}\Bigg(27\Bigg(-\frac{15}{2}\Bigg(\Bigg(-\frac{4}{15}\delta + e^2\delta - \frac{11}{30}e^2\varepsilon - \frac{2}{15}\varepsilon\Bigg)c^2 + \Bigg(-\frac{14}{15}e^2\varepsilon + \frac{32}{15}e^2\delta - \frac{8}{15}\delta\Bigg)c \\
+\frac{2}{45}\varepsilon + \frac{17}{15}e^2\delta - \frac{13}{30}e^2\varepsilon - \frac{4}{15}\delta\Bigg)\delta(c-1)^2\cos(2g-2h) \\
-\frac{15}{2}\Bigg(\Bigg(-\frac{4}{15}\delta + e^2\delta - \frac{11}{30}e^2\varepsilon - \frac{2}{15}\varepsilon\Bigg)c^2 + \Bigg(\frac{14}{15}e^2\varepsilon - \frac{32}{15}e^2\delta + \frac{8}{15}\delta\Bigg)c \\
+\frac{2}{45}\varepsilon + \frac{17}{15}e^2\delta - \frac{13}{30}e^2\varepsilon - \frac{4}{15}\delta\Bigg)(c+1)^2\delta\cos(2g+2h) \\
-\frac{643}{6}s^2\Bigg(\frac{36}{643} + e^2\Bigg)\delta\Bigg(\delta + \frac{1}{4}\varepsilon\Bigg)(c-1)^2\cos(4g-2h) \\
-\frac{643}{6}(c+1)^2\Bigg(\frac{36}{643} + e^2\Bigg)\delta\Bigg(\delta + \frac{1}{4}\varepsilon\Bigg)s^2\cos(4g+2h) \\
+\frac{643}{16}\Bigg(\frac{36}{643} + e^2\Bigg)\delta^2(c-1)^4\cos(4g-4h) + \frac{643}{16}(c+1)^4\Bigg(\frac{36}{643} + e^2\Bigg) \\
\times\delta^2\cos(4g+4h) + \frac{1}{2}\delta^2s^2(c-1)^2\Big(-2+e^2\Big)\cos(2g-4h) \\
+\frac{1}{2}\delta^2s^2(c+1)^2\Big(-2+e^2\Big)\cos(4h+2g) \\
+\Bigg(\Bigg(\Big(-2+e^2\Big)\delta^2 + \Bigg(\frac{15}{2}e^2\varepsilon - 2\varepsilon\Bigg)\delta - \frac{13}{4}e^2\varepsilon^2\Bigg)c^2 \\
+\Big(5e^2-2\Big)\delta^2 + \Bigg(-\frac{17}{6}e^2\varepsilon + \frac{2}{3}\varepsilon\Bigg)\delta\frac{11}{12}e^2\varepsilon^2\Bigg) \\
\times s^2\cos(2g) + \frac{3215}{24}s^4\Bigg(\frac{36}{643} + e^2\Bigg)\delta\Bigg(\delta + \frac{2}{5}\varepsilon\Bigg)\cos(4g) \\
+\frac{307}{2}\delta s^2\Bigg(\Bigg(\frac{56}{921}\varepsilon + \frac{148}{2763}\delta + e^2\delta + \frac{199}{307}e^2\varepsilon\Bigg)c^2 \\
+\frac{407}{921}e^2\varepsilon + \frac{172}{2763}\delta + \frac{20}{921}\varepsilon + \frac{2863}{2763}e^2\delta\Bigg)\cos(2h) \\
-\frac{705}{8}\Bigg(e^2 + \frac{484}{6345}\Bigg)\delta^2s^4\cos(4h) \\
+\Bigg(\Bigg(-\frac{3401}{24}e^2 - \frac{175}{18}\Bigg)\delta^2 + \Bigg(-\frac{139}{6}e^2\varepsilon - \frac{10}{9}\varepsilon\Bigg)\delta - \frac{229}{12}e^2\varepsilon^2 - \frac{37}{18}\varepsilon^2\Bigg)c^4 \\
+\Bigg(\Bigg(-\frac{4019}{12}e^2 - \frac{55}{3}\Bigg)\delta^2 + \Bigg(\frac{26}{9}\varepsilon + \frac{442}{9}e^2\varepsilon\Bigg)\delta + \frac{266}{9}e^2\varepsilon^2 + \frac{22}{9}\varepsilon^2\Bigg)c^2 \\
+\Bigg(-\frac{11003}{72}e^2 - \frac{223}{18}\Bigg)\delta^2 + \Bigg(-\frac{467}{18}e^2\varepsilon - \frac{16}{9}\varepsilon\Bigg)\delta - \frac{19}{18}\varepsilon^2 - \frac{497}{36}e^2\varepsilon^2\Bigg)n^2\Bigg).
\end{aligned}
\tag{A.1}
$$

B. Appendix B

Plugging the equations for the potential in the planetary equations of Lagrange and solving the equation $dg/dt = 0$, we present a new formula to compute the critical inclination taking into account the J_2 and C_{22} terms of the second-order disturbing potential. We get

$$
\begin{aligned}
&\cos(i)\\
&= \Big(\operatorname{sqrt}\Big(-2840856\varepsilon^3\delta\cos(-2h+2g) + 56754360\delta^3\varepsilon\cos(6g+2h) + 7864578\delta^2\varepsilon^2\cos(8g+4h)\\
&\quad - 3390336\varepsilon^3\delta\cos(6g) - 1294704\delta^3\varepsilon\cos(6g+4h) + 225172432\delta^2\varepsilon^2\cos(2h)\\
&\quad - 2840856\varepsilon^3\delta\cos(2h+2g) - 115917276\delta^3\varepsilon\cos(4g+2h) + 9664416\delta^2\varepsilon^2\cos(2h+2g)\\
&\quad + 86411214\delta^3\varepsilon\cos(8g+2h) - 17987760\varepsilon^3\delta\cos(4g+2h) - 1157400\varepsilon^3\delta\cos(6g+2h)\\
&\quad - 176827952\delta^3\varepsilon\cos(4h) + 50121840\delta^2\varepsilon^2\cos(-4h+2g) - 115917276\delta^3\varepsilon\cos(4g-2h)\\
&\quad + 22974624\delta^3\varepsilon\cos(6g) - 266643072\delta^2\varepsilon^2\cos(4g) - 76280160\delta^3\varepsilon\cos(4g+4h)\\
&\quad + 86411214\delta^3\varepsilon\cos(8g-2h) + 142040508\delta^3\varepsilon\cos(4g+6h) + 18851659\delta^3\varepsilon\cos(8g)\\
&\quad - 68147040\varepsilon^3\delta\cos(4g) - 31388328\delta^3\varepsilon\cos(8g-4h) + 7864578\delta^2\varepsilon^2\cos(8g-4h)\\
&\quad - 23541246\delta^3\varepsilon\cos(8g-6h) + 31434984\delta^2\varepsilon^2\cos(8g-2h) + 16378220\delta^3\varepsilon\cos(2h)\\
&\quad + 5387328\delta^2\varepsilon^2\cos(6g+2h) - 180979560\delta^3\varepsilon\cos(2g+6h) + 50121840\delta^2\varepsilon^2\cos(4h+2g)\\
&\quad - 4835676\delta^3\varepsilon\cos(2h+2g) - 157708224\delta^2\varepsilon^2\cos(4g+2h) - 23541246\delta^3\varepsilon\cos(8g+6h)\\
&\quad + 2482920\delta^2\varepsilon^2\cos(6g+4h) - 136912332\delta^3\varepsilon\cos(6h) - 180635040\delta^3\varepsilon\cos(-4h+2g)\\
&\quad - 4835676\delta^3\varepsilon\cos(-2h+2g) + 33807840\varepsilon^3\delta\cos(2h) - 6238188\delta^3\varepsilon\cos(6g-6h)\\
&\quad + 7617120\delta^3\varepsilon\cos(2g) - 358413130\delta^4 + 9200304\varepsilon^4 + 2482920\delta^2\varepsilon^2\cos(6g-4h)\\
&\quad - 17987760\varepsilon^3\delta\cos(4g-2h) - 4235136\varepsilon^3\delta\cos(2g) - 157708224\delta^2\varepsilon^2\cos(4g-2h)\\
&\quad - 180635040\delta^3\varepsilon\cos(4h+2g) + 142040508\delta^3\varepsilon\cos(4g-6h) - 1298448\delta^2\varepsilon^2\cos(2g)\\
&\quad + 56754360\delta^3\varepsilon\cos(6g-2h) - 1157400\varepsilon^3\delta\cos(6g-2h) + 3657744\delta^2\varepsilon^2\cos(6g)\\
&\quad + 47140812\delta^2\varepsilon^2\cos(8g) - 94164984\delta^4\cos(8g-6h) - 94164984\delta^4\cos(8g+6h)\\
&\quad - 27747576\delta^4\cos(6g+2h) + 516167880\delta^4\cos(4g-2h) + 16582968\delta^4\cos(6g-6h)\\
&\quad + 16582968\delta^4\cos(6g+6h) + 206320632\delta^4\cos(4g-6h) + 516167880\delta^4\cos(4g+2h)\\
&\quad + 46013544\delta^4\cos(2g-6h) + 46013544\delta^4\cos(2g+6h) + 206320632\delta^4\cos(4g+6h)\\
&\quad + 572713584\delta^4\cos(4g-4h) + 81338580\delta^2\varepsilon^2 + 31906368\varepsilon^3\delta + 1319976\delta^4\cos(6g-8h)\\
&\quad - 35311869\delta^4\cos(8g-8h) + 1319976\delta^4\cos(6g+8h) - 35311869\delta^4\cos(8g+8h)
\end{aligned}
$$

$$
\begin{aligned}
&-1099291328^4\cos(8g-4h)-1099291328^4\cos(8g+4h)+280975688^4\cos(-4h+2g)\\
&+280975688^4\cos(4h+2g)+3550428968^3\varepsilon+23033528^4\cos(2g-8h)\\
&+23033528^4\cos(2g+8h)+572713584\delta^4\cos(4g+4h)-277233128^4\cos(2g)\\
&-180428526\delta^4\cos(8h)-754078536\delta^4\cos(4h)-624087968^4\cos(6h)\\
&+378941832\delta^4\cos(4g)-212477742\delta^4\cos(8g)-41964048^4\cos(6g)\\
&-381690336\delta^4\cos(2h)+153606924\delta^4\cos(4g-8h)+153606924\delta^4\cos(4g+8h)\\
&-4408992\delta^4\cos(6g+4h)-277475768^4\cos(6g-2h)-4408992\delta^4\cos(6g-4h)\\
&+39600\varepsilon^4\cos(4g)+1801056\varepsilon^4\cos(2g)-190262168^4\cos(-2h+2g)\\
&-190262168^4\cos(2h+2g)-1573148888^4\cos(8g-2h)-1573148888^4\cos(8g+2h)\\
&-180979568^3\varepsilon\cos(2g-6h)+314349848^2\varepsilon^2\cos(8g+2h)\\
&+710935648^2\varepsilon^2\cos(4h)+569589128^2\varepsilon^2\cos(4g-4h)+569589128^2\varepsilon^2\cos(4g+4h)\\
&+53873288^2\varepsilon^2\cos(6g-2h)-623818888^3\varepsilon\cos(6g+6h)-3138832888^3\varepsilon\cos(8g+4h)\\
&-1294704\delta^3\varepsilon\cos(6g-4h)-1855092488^3\varepsilon\cos(4g)+9664416\delta^2\varepsilon^2\cos(-2h+2g)\\
&-762801608^3\varepsilon\cos(4g-4h)\Big)+\left(-60\delta\varepsilon-72\delta^2\right)\cos(-2h+2g)\\
&+\left(60\delta\varepsilon+72\delta^2\right)\cos(2h+2g)+11898\delta^2\cos(4g-4h)\\
&-11898\delta^2\cos(4g+4h)+\left(15864\delta^2+3966\delta\varepsilon\right)\cos(4g-2h)\\
&+\left(-15864\delta^2-3966\delta\varepsilon\right)\cos(4g+2h)\Big)/\Big(\left(-1056\delta\varepsilon+1728\delta^2\right)\cos(-2h+2g)\\
&+\left(-1056\delta\varepsilon+1728\delta^2\right)\cos(2h+2g)+36666\delta^2\cos(4g-4h)+36666\delta^2\cos(4g+4h)\\
&+\left(-1104\delta\varepsilon+600\varepsilon^2+576\delta^2\right)\cos(2g)+\left(-5424\delta\varepsilon+992\delta^2\right)\cos(2h)\\
&-40740\delta\left(\frac{2}{5}\varepsilon+\delta\right)\cos(4g)+27316\delta^2\cos(4h)+7488\delta\varepsilon-50868\delta^2+4608\varepsilon^2\Big).
\end{aligned}
\tag{B.1}
$$

C. Appendix C

Model given by Akim and Golikov [30] for the spherical harmonic coefficients is

$$
\begin{aligned}
&J_2=2.09006496\times10^{-4},\quad J_4=2.32815\times10^{-6},\quad C_{22}=2.447305\times10^{-5},\\
&J_3=5.48445\times10^{-6},\quad J_5=-3.169113\times10^{-6},\quad C_{31}=2.871327\times10^{-5}.
\end{aligned}
\tag{C.1}
$$

Acknowledgment

This work was accomplished with support of the FAPESP under the contract no. 2007/04413-7 and 2006/04997-6, SP-Brazil, and CNPQ (300952/2008-2).

References

[1] G. I. Hori, "Theory of general perturbations with unspecified canonical variables," *Publications of the Astronomical Society of Japan*, vol. 18, no. 4, pp. 287–296, 1966.

[2] S.-Y. Park and J. L. Junkins, "Orbital mission analysis for a lunar mapping satellite," *Journal of the Astronautical Sciences*, vol. 43, no. 2, pp. 207–217, 1995.

[3] B. De Saedeleer and J. Henrard, "Orbit of a lunar artificial satellite: analytical theory of perturbations," in *Transits of Venus: New Views of the Solar System and Galaxy*, IAU Colloquium, no. 196, pp. 254–262, Cambridge University Press, Cambridge, Mass, USA, 2005.

[4] B. De Saedeleer, "Analytical theory of a lunar artificial satellite with third body perturbations," *Celestial Mechanics & Dynamical Astronomy*, vol. 95, no. 1–4, pp. 407–423, 2006.

[5] B. De Saedeleer and J. Henrard, "The combined effect of J_2 and C_{22} on the critical inclination of a lunar orbiter," *Advances in Space Research*, vol. 37, no. 1, pp. 80–87, 2006.

[6] A. Deprit, "Canonical transformations depending on a small parameter," *Celestial Mechanics*, vol. 1, pp. 12–30, 1969.

[7] S. Ferraz Mello, *Canonical Perturbation Theories*, Springer, New York, NY, USA, 2007.

[8] J. P. dos Santos Carvalho, R. V. de Moraes, and A. F. B. de Almeida Prado, "Semi-analytic theory of a moon artificial satellite considering lunar oblateness and perturbations due to a third-body in elliptic orbit," in *Proceedings of the 7th Brazilian Conference on Dynamics, Control and Applications*, pp. 51–57, Presidente Prudente, Brazil, 2008.

[9] J. P. dos Santos Carvalho, R. V. de Moraes, and A. F. B. de Almeida Prado, "Moon artificial satellites: lunar oblateness and earth perturbations," in *Proceedings of the International Conference on Mathematical Problems in Engineering, Aerospace and Sciences (ICNPAA '08)*, pp. 1095–1106, Cambrige Scientific, Genoa, Italy, 2009.

[10] J. P. dos Santos Carvalho, R. V. de Moraes, and A. F. B. de Almeida Prado, "A study on resonance for a lunar artificial satellite," in *Proceedings of the 8th Brazilian Conference on Dynamics, Control and Applications*, pp. 1–6, Bauru, Brazil, 2009.

[11] J. P. dos Santos Carvalho, R. V. de Moraes, and A. F. B. de Almeida Prado, "Non-sphericity of the moon and critical inclination," in *Proceedings of the 32nd Congresso Nacional de Matemática Aplicada e Computacional*, Cuiabá, Brazil, 2009.

[12] W. M. Kaula, *Theory of Satellite Geodesy: Applications of Satellites to Geodesy*, Dover, New York, NY, USA, 2000.

[13] M. Radwan, "Analytical approach to the motion of a lunar artificial satellite," *Astrophysics and Space Science*, vol. 283, no. 2, pp. 133–150, 2003.

[14] G. E. O. Giacaglia, J. P. Murphy, and T. L. Felsentreger, "A semi-analytic theory for the motion of a lunar satellite," *Celestial Mechanics*, vol. 3, no. 1, pp. 3–66, 1970.

[15] A. Elipe and M. Lara, "Frozen orbits about the moon," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 2, pp. 238–243, 2003.

[16] Z. Knežević and A. Milani, "Orbit maintenance of a lunar polar orbiter," *Planetary and Space Science*, vol. 46, no. 11-12, pp. 1605–1611, 1998.

[17] A. Cayley, "Tables of the developments of functions in the theory of elliptic motion," *Memoirs of the Royal Astronomicai Society*, vol. 29, p. 191, 1861.

[18] R. A. Broucke, "Long-term third-body effects via double averaging," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 1, pp. 27–32, 2003.

[19] I. V. Costa and A. F. B. de Almeida Prado, "Orbital Evolution of a Satellite Perturbed by a Third Body," in *Advances in Space Dynamics*, A. F. Bertachini and A. F. B. de Almeida Prado, Eds., p. 176, 2000.

[20] R. C. Domingos, R. V. de Moraes, and A. F. B. de Almeida Prado, "Third-body perturbation in the case of elliptic orbits for the disturbing body," *Advances in the Astronautical Sciences*, vol. 129, part 3, pp. 2639–2652, 2008.

[21] A. F. B. de Almeida Prado, "Third-body perturbation in orbits around natural satellites," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 1, pp. 33–40, 2003.

[22] C. R. H. Solórzano, *Third-body perturbation using a single averaged model*, M.S. dissertation, National Institute for Space Research (INPE), São José dos Campos, Brazil, 2002.
[23] C. R. H. Solórzano and A. F. B. de Almeida Prado, "Third-body perturbation using a single averaged model," in *Advances in Space Dynamics 4: Celestial Mechanics and Astronautics*, H. K. Kuga, Ed., pp. 47–56, 2004.
[24] S. Da Silva Fernandes, "Notes on Hori method for canonical systems," *Celestial Mechanics & Dynamical Astronomy*, vol. 85, no. 1, pp. 67–77, 2003.
[25] S. Breiter, "Second-order solution for the zonal problem of satellite theory," *Celestial Mechanics and Dynamical Astronomy*, vol. 67, pp. 237–249, 1997.
[26] J. Kovalevsky, *Introduction to Celestial Mechanics*, Bureau dês Longitudes, Paris, France, 1967.
[27] D. Brouwer, "Solution of the problem of artificial satellite theory without drag," *The Astronomical Journal*, vol. 64, pp. 378–397, 1959.
[28] W. K. Meyer, J. J. Buglia, and P. N. Desai, "Lifetimes of Lunar Satellite Orbits," Tech. Rep. N-TP-3394 94, p. 27771, NASA STI/Recon, 1994.
[29] P. D'Avanzo, P. Teofilatto, and C. Ulivieri, "Long-term effects on lunar orbiter," *Acta Astronautica*, vol. 40, no. 1, pp. 13–20, 1997.
[30] E. Akim and A. Golikov, "Combined model of the lunar gravity field," in *Proceedings of the 12th International Symposium on Space Flight Dynamics (ESOC '97)*, pp. 357–360, Darmstadt, Germany, June 1997.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 347835, 12 pages
doi:10.1155/2009/347835

Research Article

GPS Satellites Orbits: Resonance

Luiz Danilo Damasceno Ferreira[1] and Rodolpho Vilhena de Moraes[2]

[1] *Department of Geomatic, Federal University of Paraná (UFPR), 81531-909 Curitiba, PR, Brazil*
[2] *UNESP-Univ Estadual Paulista, Guaratinguetá, CEP 12516-410, SP, Brazil*

Correspondence should be addressed to Rodolpho Vilhena de Moraes, rodolpho@feg.unesp.br

Received 29 July 2009; Accepted 28 September 2009

Recommended by Antonio Prado

The effects of perturbations due to resonant geopotential harmonics on the semimajor axis of GPS satellites are analyzed. For some GPS satellites, secular perturbations of about 4 m/day can be obtained by numerical integration of the Lagrange planetary equations considering in the disturbing potential the main secular resonant coefficients. Amplitudes for long-period terms due to resonant coefficients are also exhibited for some hypothetical satellites orbiting in the neighborhood of the GPS satellites orbits. The results are important to perform orbital maneuvers of GPS satellites such that they stay in their nominal orbits. Also, for the GPS satellites that are not active, the long-period effects due to the resonance must be taken into account in the surveillance of the orbital evolutions of such debris.

1. Introduction

The period of the orbits of the GPS satellites is about 12 hours, and the main perturbations acting on their orbits are caused by the nonuniform distribution of the Earth's mass, by the lunar and solar gravitational attractions and by the solar radiation pressure. In this paper, it is analyzed just some perturbations due to resonant terms of the geopotential coefficients. The resonance considered here is the $2:1$ commensurability between the orbital period of the GPS satellites and the period of the Earth's rotation. As it is pointed out by Hugentobler [1] the resonance leads a daily drift rate in semimajor axis of up to 7 m/day.

Lagrange planetary equations, describing the temporal variation of the orbital elements, are used here to analyze the orbital perturbations of the GPS satellites under the influence of resonant coefficients.

2. Disturbing Potential

The geopotential acting on Earth's artificial satellite can be expressed as [2]

$$R = \frac{GM}{a}\sum_{n=2}^{\infty}\sum_{m=0}^{n}\left(\frac{a_e}{a}\right)^n (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\sin\varphi). \tag{2.1}$$

Here GM is the geogravitational constant, r is geocentric distance of the satellite, a_e is the semimajor axis of the adopted Earth's ellipsoid, a is the orbital semimajor axis, C_{nm} and S_{nm} are the spherical harmonic coefficients. P_{nm} denotes the associated Legendre functions, φ, λ are the satellite's geocentric latitude and the longitude, n and m are, respectively, the degree and order of the harmonic coefficients.

The geocentric distance, the latitude, and the longitude of the satellite can be expressed in terms of the orbital elements and (2.1) becomes [2, 3]

$$R = \sum R_{nmpq} = \frac{GM}{a}\sum_{n=2}^{\infty}\sum_{m=0}^{n}\left(\frac{a_e}{a}\right)^n \sum_{p=0}^{n} F_{nmpq}(i)\sum_{q=-\infty}^{\infty} G_{npq}(e)S_{nmpq}(\omega,\Omega,M,\theta_T), \tag{2.2}$$

where $F_{nmpq}(i)$ and $G_{npq}(e)$ are, respectively, functions of the satellite's orbital inclination i and eccentricity e, Ω represents the right ascension of the orbital ascending node, ω is the argument of perigee, M is the mean anomaly, and θ_T is the Greenwich sidereal time. The function $S_{nmpq}(\omega,\Omega,M,\theta_T)$ can be expressed as

$$S_{nmpq}(\omega,\Omega,M,\theta_T) = \left[\begin{pmatrix} C_{nm} \\ S_{nmn} \end{pmatrix}\cos A + \begin{pmatrix} C_{nm} \\ S_{nmn} \end{pmatrix}\sin A\right]_{n-m,\text{odd}}^{n-m,\text{even}}, \tag{2.3}$$

and the argument A is given by

$$A = (n-2p)\omega + (n-2p+q)M + m(\Omega-\theta_T). \tag{2.4}$$

The functions $F_{nmpq}(i)$ and $G_{npq}(e)$ are presented as tables [2] and can be adapted to be used in computers.

3. Resonance in GPS Satellites

Resonance is associated with small divisors. For artificial Earth satellites whose orbital periods are in commensurability with the period of the Earth's rotation resonance can occur when [1, 3, 4]

$$\dot{A} = (n-2p)\dot{\omega} + (n-2p+q)\dot{M} + m(\dot{\Omega}-\dot{\theta}_T) \cong 0, \tag{3.1}$$

where $\dot{\theta}_T$ is the Earth's sidereal rotation.

For the GPS satellites where the commensurability is $2:1$, there are bounds among the parameters n, m, p, q.

Table 1: Resonant parameters.

Degree n	Order m	p	q
3	2	1	0
4	4	1	0
2	2	1	1
4	2	2	1

Taking into account that $\dot{\omega}, \dot{\Omega} \ll \dot{\theta}_T$, (3.1) can be put as

$$(n - 2p + q)\dot{M} - m\dot{\theta}_T = 0, \tag{3.2}$$

since $\overline{n} = \dot{M} \cong 2\dot{\theta}_T$, we get

$$(n - 2p + q)2 \cong m. \tag{3.3}$$

Considering harmonics of order and degree up to order 4, Table 1, presents some values for the parameters satisfying the resonance condition for the GPS satellites.

Taking into account in the summations the conditions $(n - m)$ even and $(n - m)$ odd and putting [5]

$$\begin{aligned} C_{nm} &= K_{nm} \cos m\lambda_{nm}, \\ S_{nm} &= K_{nm} \sin m\lambda_{nm}, \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} K_{nm} &= \sqrt{C_{nm}^2 + S_{nm}^2}, \\ \lambda_{nm} &= \frac{1}{m} \tan^{-1}\left(\frac{S_{nm}}{C_{nm}}\right), \end{aligned} \tag{3.5}$$

equation (2.2) can be written in general form as [1]

$$R_{nmpq} = \frac{GM}{a}\left(\frac{a_e}{a}\right)^n F_{nmp}(i) G_{npq}(e) \begin{pmatrix} \cos \mathrm{A} \\ \sin \mathrm{A} \end{pmatrix}_{n-m,\mathrm{odd}}^{n-m,\mathrm{even}} K_{nm}, \tag{3.6}$$

where

$$A = (n - 2p)\omega + (n - 2p + q)M + m(\Omega - \theta_T - \lambda_{nm}). \tag{3.7}$$

In order to analyze the orbital perturbations of GPS satellites due to the resonant coefficients presented in (Table 1), the Lagrange planetary equations will be used. However,

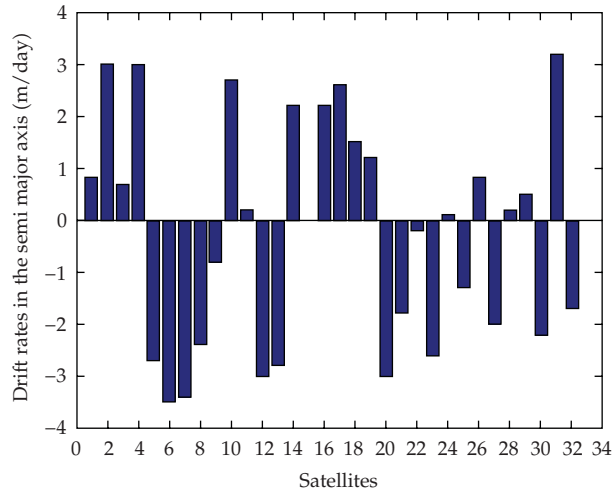


Figure 1: Drift rates in semimajor axis due to the resonant geopotential coefficient 32.

Table 2: Maximum values.

Resonants coefficients	Maximum drift rates m/day
C_{32}, S_{32}	−4.0
C_{44}, S_{44}	1.5
C_{22}, S_{22}	1.5
C_{42}, S_{42}	0.002

here this analysis will be concentrated initially in the secular effects of such perturbation on the orbital semimajor axis. Therefore, we have [2]

$$\frac{da}{dt} = \frac{2}{\overline{n}a}\frac{\partial R_{nmpq}}{\partial M}, \tag{3.8}$$

with $\overline{n}$ being the mean motion of the satellite.

Special care must be taken to compute $\partial R_{nmpq}/\partial M$. In fact, the parameters n, m, p, q given by Table 1 must be considered and each set of them gives a unique solution for (3.8).

A numerical integration of (3.8) was performed for a period of one day using IGS/POE (The International GNSS Service Precise Orbital Ephemeredes) given on June, 07, 2007. Transformations were performed to get the corresponding orbital elements.

Figures 1, 2, 3, and 4 show the daily variations of the semimajor axis a for all GPS satellites.

Figure 1 gives the variation due to the coefficients C_{32}, S_{32}, which are responsible by the greatest variation in the semimajor axis that, in this case, corresponds to the satellite PRN 06. Figures 2, 3, and 4 present, respectively, the daily variations as function of the coefficients $C_{44}, S_{44}, C_{22}, S_{22}, C_{42}, S_{42}$. Figure 5 represents the total daily variation due to the above considered coefficients.

Table 2 shows the maxima and minima values for the daily variation of the semimajor axis according to the resonant coefficients.

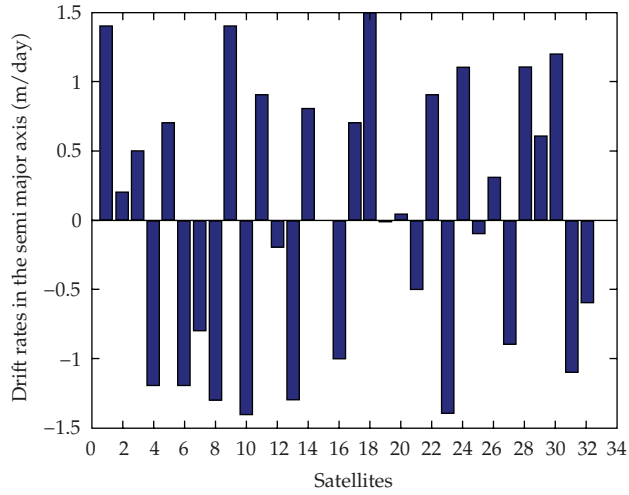


Figure 2: Drift rates in semimajor axis due to the resonant geopotential coefficient 44.

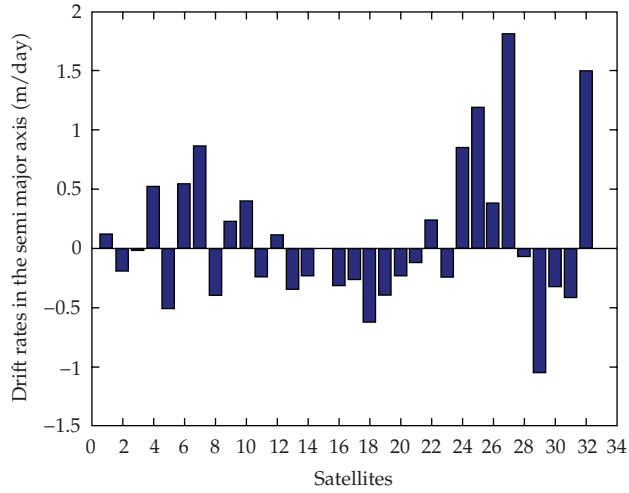


Figure 3: Drift rates in semimajor axis due to the resonant geopotential coefficient 22.

Figures 6 and 7 show the semimajor axis variation for the satellites PRN 02 and PRN 06 for a time interval of 200 days and taking into account the harmonic coefficients C_{32}, S_{32}. During this period and for the considered initial condition, it can be observed variations of about 600 m and 680 m, respectively.

4. Long-Period Perturbations

The effects of the resonance are enhanced when long periods are considered. Table 3 and Table 4, Figures 8, 9, 10, 11, 12, and 13 show some simulations using hypothetical satellites and a particular method to study effects of resonance on the orbits of artificial satellites [6–9].

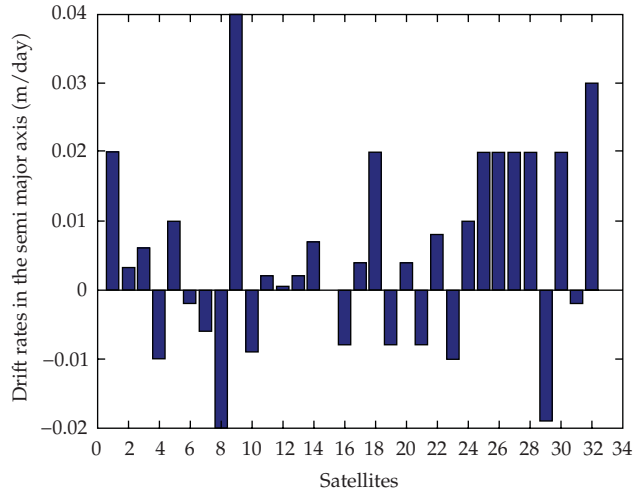


Figure 4: Drift rates in semimajor axis due to the resonant geopotential coefficient 42.

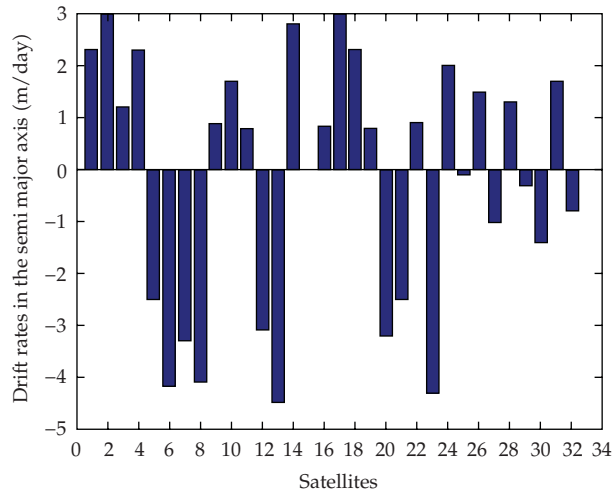


Figure 5: Drift rates in semimajor axis due to all resonant geopotential coefficients, Table 1.

Table 3 contains the amplitude and period of the variations of orbital elements for hypothetical satellites considering low eccentricity, small and high inclination, and the influence of the harmonics J_{20} and J_{22}

Figures 8 and 9 represent, respectively, the temporary variation of the semimajor axis and of the eccentricity of artificial satellites of the GPS type in the neighborhood of the 2 : 1 resonance region when the harmonics J_{20} and J_{22} are considered. By Figure 10, assuming several values for the semimajor axes in the neighborhood of the 2 : 1 resonance, it can be observed distinct behavior for their temporary variations.

Figure 11 represents the temporary variation for the semimajor axis in the neighborhood of the resonance 2 : 1 considering the influence of the harmonics J_{20} and J_{32} for a satellite of the GPS type with inclination of about 55°.

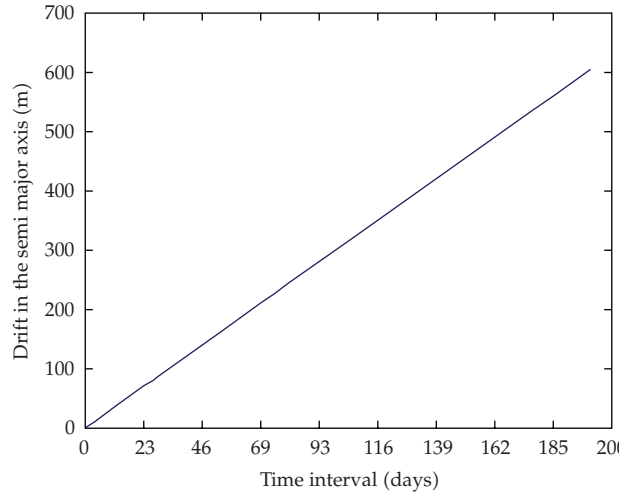


Figure 6: Semimajor axis variation (PRN 02) due to the resonant geopotential coefficient 32, for a time interval of 200 days.

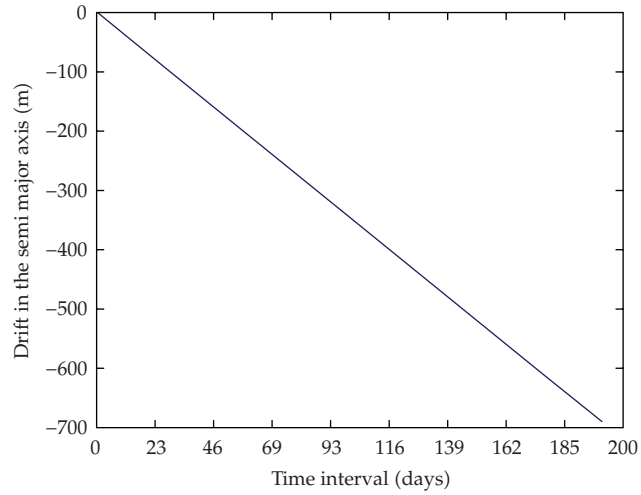


Figure 7: Semimajor axis variation (PRN 06) due to the resonant geopotential coefficient 32, for a time interval of 200 days.

Table 4 gives the influence of the resonance due to the harmonics J_{20} and J_{32} considering different inclinations, including that of the GPS type satellite. It can be observed that the amplitudes of the variations are smaller when compared with those when the harmonics J_{20} and J_{22} are taken into account. The more the satellites approach the region that was defined as a resonant region, the more the variations increase.

Figure 12 represents the temporary variation of the semimajor axis of a satellite of the GPS type considering resonance due to the harmonics J_{20}, J_{22}, and J_{32}.

Table 3: Amplitude and period of perturbations due to the 2 : 1 resonance: $J_2 + J_{22}$.

Orbital elements			Amplitude			Period
$a_o = 26561.770$ km	e	i	$\Delta a_{\max}$	$\Delta e_{\max}$	$\Delta i_{\max}(^\circ)$	t (days)
$a_o + 0.430$	0.01	4°	12 km	0.014	8.7×10^{-4}	2500
$a_o + 0.718$	0.01	4°	12.5 km	0.0145	8.6×10^{-4}	7000
$a_o + 0.290$	0.01	4°	4.5 km	0.0065	3.6×10^{-4}	2000
$a_o + 0.500$	0.01	4°	3 km	0.0045	2.8×10^{-4}	1000
$a_o - 2.5$	0.01	4°	6.5 km	0.0085	4×10^{-4}	1700
$a_o - 5$	0.01	4°	1.4 km	0.0023	1.3×10^{-6}	500
a_o	0.05	55°	1.8 km	0.005	2×10^{-3}	3200
$a_o + 0.718$	0.005	55°	2.75 km	0.006	2.8×10^{-3}	3900
$a_o - 1.770$	0.005	55°	1.2 km	0.0045	1.4×10^{-3}	2000
$a_o + 5.129$	0.005	55°	900 m	0.0033	10^{-3}	1500
$a_o + 3.729$	0.005	55°	1.8 km	0.0055	1.9×10^{-3}	2500
a_o	0.05	63.4°	7.5 km	0.003	10.8×10^{-3}	1000
$a_o - 2$	0.05	63.4°	4 km	0.0015	5.7×10^{-3}	1000
a_o	0.05	87°	7 km	0.0037	13.9×10^{-3}	2000
$a_o - 2$	0.05	87°	3.5 km	0.002	10.3×10^{-3}	1900

Table 4: Amplitude and period of perturbations due to the 2 : 1 resonance: $J_2 + J_{32}$.

Orbital elements			Amplitude			Period
$a_o = 26561.770$ km	e	i	$\Delta a_{máx}$	$\Delta e_{máx}$	$\Delta i_{máx}(^\circ)$	t (days)
$a_o - 3.77$	0.01	4°	120 m	0.0004	1.78×10^{-3}	500
$a_o + 0.718$	0.01	4°	245 m	0.0009	3.9×10^{-3}	1000
$a_o + 0.929$	0.01	4°	110 m	0.00045	1.78×10^{-3}	500
$a_o + 5$	0.01	4°	100 m	0.0003	1.15×10^{-3}	250
$a_o - 2.5$	0.005	4°	100 m	0.0009	4.6×10^{-5}	500
$a_o + 0.429$	0.005	4°	700 m	0.006	2.4×10^{-4}	1500
$a_o + 0.929$	0.005	55°	1.9 km	0.0085	6.8×10^{-4}	16000
$a_o + 1.73$	0.005	55°	250 m	0.0017	1.13×10^{-3}	16000
$a_o + 5$	0.005	55°	30 m	0.00055	2.3×10^{-4}	1500
$a_o - 3.77$	0.005	55°	47 m	0.00035	3.43×10^{-5}	500
$a_o - 0.429$	0.005	87°	110 m	0.00085	2.29×10^{-4}	1800
$a_o - 2.23$	0.005	87°	45 m	0.00035	8.59×10^{-5}	800
$a_o - 3.77$	0.005	87°	20 m	0.00014	4×10^{-5}	500

Figure 13 represents the time variation of the orbital semimajor axis of a satellite with eccentricity about 0.01 and inclination about 4° orbiting in a region near the 2 : 1 resonance. It is remarkable the oscillation of the semimajor axis in the region between by $a = 26560.0$ km and $a = 26562.48$ km. For instance, a variation of about 10 m in the initial semimajor axis ($a = 26562.48$ km) brings up a variation of about more than 10 km in its amplitude in a period of about 2000 days.

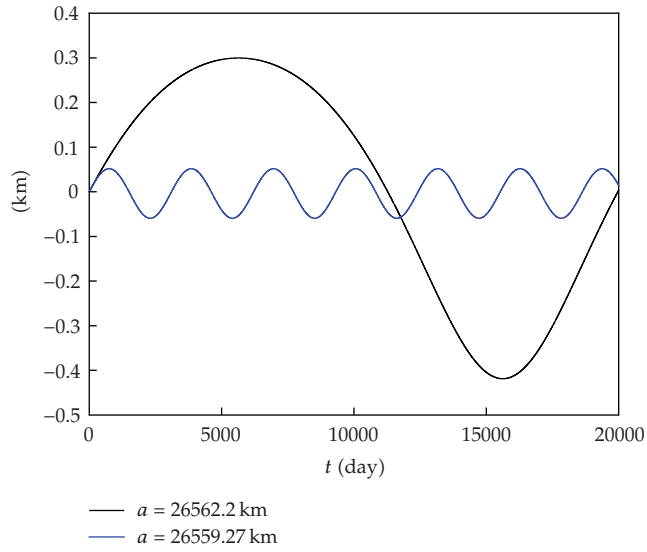


Figure 8: Δa versus time, $e = 0.005$, $i = 55°$.

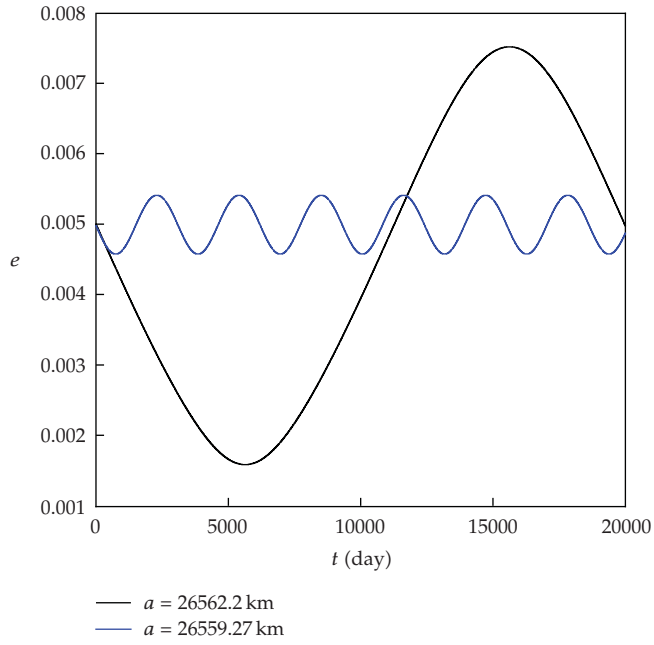


Figure 9: Eccentricity versus time, $e = 0.005$.

5. Conclusions

From the obtained results it can be seen that for the GPS satellites no negligible perturbations are provoked by resonant tesseral coefficients. The daily variation of the semimajor axis due to the C_{32}, S_{32} is less than 4 m/day and those values found for a time interval of 200 days

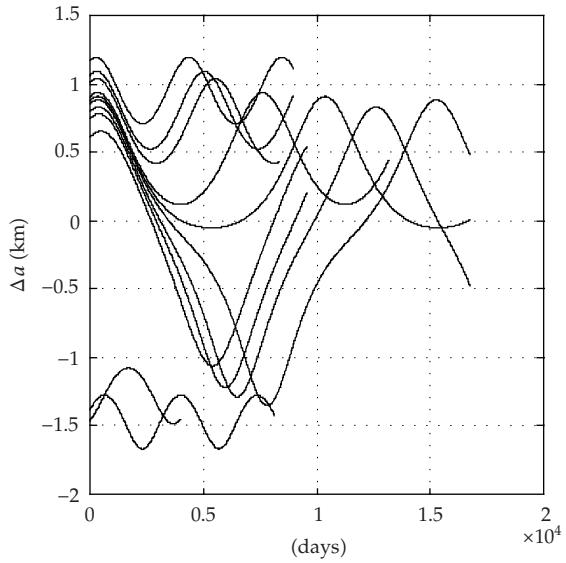


Figure 10: Δa versus time, $e = 0.005$, $i = 55°$, $a_0 = 26561.18$ km.

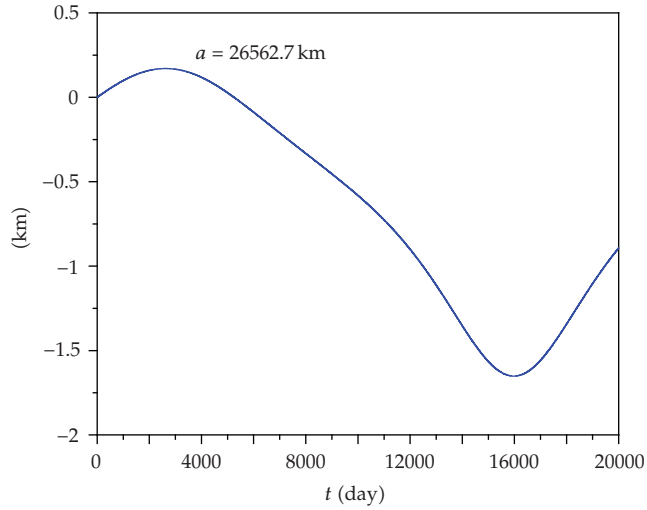


Figure 11: Δa versus time, $e = 0.005$, $i = 55°$, considering the harmonics J_{20} and J_{32}.

(~680 m) are compatible with the results presented by specialized literature (~7 m/day and ~1200 m) about this subject [1].

Taking into account the total effects of daily resonant perturbations, it can be observed that for some GPS satellites these effects are enhanced and for another satellites are attenuated but all of these effects are smaller than the amplitudes mentioned above. It was shown also that the effects of the resonance are very important for the analysis of long-period behavior of the GPS satellites orbits.

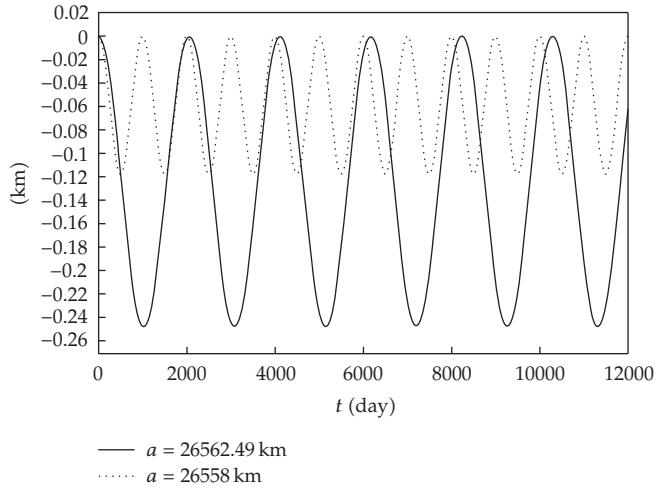


Figure 12: Time variation of semimajor axis considering the harmonics J_{20}, J_{22}, and $J_{3,2}$.

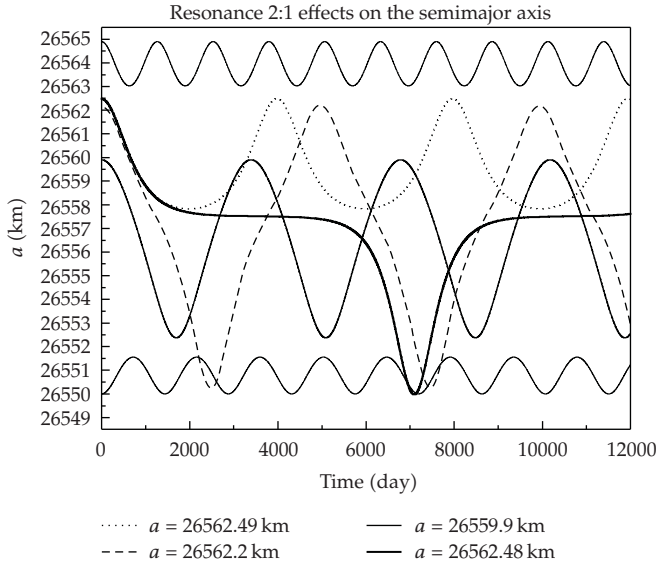


Figure 13: Time variation of the semimajor axis considering the harmonics J_{20} and J_{22} for $e = 0.005$, $i = 4°$.

An important aspect to be considered is the necessity to perform orbital maneuvers of GPS satellites in such way that they stay in their nominal orbits. Also, for the GPS satellites that are not active, the long-term effects due to the resonance must be taken into account in the surveillance of the orbital evolutions of such debris.

Acknowledgment

This work was partially supported by CNPq under the contracts No. 305147/2005-6 and 300952/2008-2

References

[1] U. Hugentobler, "Astrometry and satellite orbits: theoretical consideration and typical applications," *Geodätisch-geophysikalische Arbeiten in der Schweiz*, vol. 57, 1998.
[2] W. M. Kaula, *Theory of Satellite Geodesy*, Blaisdell, Waltham, Mass, USA, 1966.
[3] G. Seeber, *Satellite Geodesy: Foundations, Methods and Applications*, Walter de Gruyter, Berlin, Germany, 2003.
[4] D. Ineichen, G. Beutler, and U. Hugentobler, "Sensitivity of GPS and GLONASS orbits with respect to resonant geopotential parameters," *Journal of Geodesy*, vol. 77, no. 7-8, pp. 478–486, 2003.
[5] D. Delikaraoglu, *On Principles, Methods and Recent Advances in Studies Towards a GPS-Based Control System for Geodesy and Geodynamics*, NASA Technical Memorandum 100716, Greenbelt, Md, USA, 1989.
[6] P. H. C. L. Lima Jr., *Sistemas ressonantes a altas excentricidades no movimento de satélites artificiais*, Doctoral thesis, ITA, São José dos Campos, 1998.
[7] P. H. C. L. Lima Jr., R. Vilhena de Moraes, and S. S. Fernandes, "Semi analytical method to study geopotential perturbations conssidering high eccentric resonant orbits," in *Dinamics of Natural and Artificial Celestial Bodies*, H. Pretka-Ziomek, E. Wnuk, P. K. Seidelmann, and D. Richardson, Eds., pp. 407–413, Kluwer, Dordrecht, The Netherlands, 2001.
[8] J. K. S. Formiga, *Study of resonances in the orbital motion of artificial satellites*, Master dissertation, Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, Guaratinguetá, Brazil, 2005.
[9] R. Vilhena de Moraes, J. K. S. Formiga, P. H. C. L. Lima Jr., and H. K. Kuga, "Orbital perturbations: resonance effects," in *Proceedings of 6th International Symposium of the IAA on Small Satellites for Earth Observation*, p. 8, Berlin, Germany, April 2007, paper IAA-B6-0717P.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 510759, 22 pages
doi:10.1155/2009/510759

Research Article

Some Initial Conditions for Disposed Satellites of the Systems GPS and Galileo Constellations

Diogo Merguizo Sanchez,[1] Tadashi Yokoyama,[1] Pedro Ivo de Oliveira Brasil,[2] and Ricardo Reis Cordeiro[3]

[1] *Departamento de Estatística, Matemática Aplicada e Computação, IGCE, UNESP campus Rio Claro, Caixa Postal 178, CEP: 13506-900 Rio Claro, SP, Brazil*
[2] *Departamento de Física, IGCE, UNESP campus Rio Claro, CEP: 13506-900 Rio Claro, SP, Brazil*
[3] *Departamento de Física, Centro de Ciências Exatas e Tecnológicas, Universidade Federal de Viçosa, UFV, CEP: 36571-000 Viçosa, MG, Brazil*

Correspondence should be addressed to Diogo Merguizo Sanchez, sanchezfisica@gmail.com

Received 31 July 2009; Accepted 20 October 2009

Recommended by Silvia Maria Giuliatti Winter

Through the averaged equations we revisit theoretical and numerical aspects of the strong resonance that increases the eccentricity of the disposed objects of GPS and Galileo Systems. A simple view of the phase space of the problem shows that the resonance does not depend on the semi-major axis. Therefore, usual strategies of changing altitude (raising perigee) do not work. In this problem we search for a set of initial conditions such that the deactivated satellites or upper-stages remain at least for 250 years without penetrating in the orbits of the operational satellites. In the case that Moon's perturbation is not significant, we can identify, in the phase space, the regions where eccentricity reaches maximum and minimum values so that possible risks of collision can be avoided. This is done semi-analytically through the averaged system of the problem. Guided by this idea, we numerically found the (ω, Ω) values of the real unaveraged problem. In particular, for the Galileo case, the theoretical results predicted in the averaged system are in good agreement with numerical results. We also show that initial inclination of the Moon plays an important role in the search of these conditions.

1. Introduction

Broadly speaking, the GPS, GLONASS, and Galileo [1] systems are satellite constellations which were designed mainly for positioning and navigation purposes. The first members of GPS (block I) originally were designed to have inclination of 63.4 degrees with respect to the equator, distributed in three orbital planes, each one separated from 120 degrees in the longitude of the node. The altitude is 20,200 km. The GLONASS members are similar, with slightly lower altitude (19,100 km, orbital period = 11:15 h). The European GALILEO

system is still in construction, and the inclination of the satellites will be 55 or 56 degrees, with altitude of 23,615 km. All the three systems have rather similar altitudes. In order to avoid risks of collision and following American Govern instructions about debris mitigation, there are some recommendations that the disposal satellites and upper-stages should be deposited at least 500 km above or below the semisynchronous orbit [2].

In a constellation of a navigation system, the members must be kept under precise requirements of functionality. However, after some time, they have to be deactivated since some level of these requirements cannot be fulfilled for long time. The destination of these deactivated objects is a problem since they must be moved into some disposal regions in order to preclude collisions with operational members of the constellation. While these vehicles can be designed a priori to transport additional propellant (at some nonnegligible cost) to be used in some planned maneuvers to insert them in the disposal regions; the same is not true for the upper-stage. In some cases (block IIF of GPS system), due to design restrictions, this upper-stage cannot be easily guided to the disposal region. It must perform several operations after the satellite is injected in the constellation. All these operations change its final parameters [3]. Since the inclination of these vehicles is near to 55 or 56 degrees, the eccentricity suffers strong variations and even an initially circular orbit can become highly eccentric so that they can cross very easily the orbit of the operational satellites. What is interesting and also problematic is the fact that the rate of growing of the eccentricity is very sensitive to the initial parameters of the disposal orbit (eccentricity, argument of the perigee, and longitude of the node). In this work, based on a theoretical framework, we present a set of initial conditions (argument of the perigee, longitude of the node) for GPS and Galileo systems such that the disposed objects can remain at the least 250 years with small eccentricity (0.01 or 0.02) without causing any risk to the operational satellites.

The above strategy of keeping small eccentricity can generate some additional problem: after some time, the disposed vehicles will accumulate and a graveyard of these objects will be created. Therefore, a risk of collisions amongst themselves is a crucial problem, since the products of these extra collisions are almost untrackable fragments that may offer more risks to the operational elements of the constellation.

According to Jenkin and Gick [4], the strategy in the opposite direction, that is, exploiting the growth of the eccentricity in order to dilute disposal orbit collision risk has some interesting points to be considered: the percentage of disposed vehicles that will reenter in the atmosphere can be increased. Another advantage observed is, although eccentricity growth strategy increases the collision risk in the constellation, that in some cases this risk can be reversed with proper choice of the initial disposal eccentricity.

In this sense, we briefly started the investigation of some initial conditions that can cause large increase of the eccentricity, for a minimum time interval, considering also different initial inclinations of the Moon's orbit (see Appendix B). Our calculations show that the growth of the eccentricity is rather sensitive to the Moon's position (inclination and semi-major axis).

2. Methods

2.1. Disturbing Function of the Sun

As we want to highlight some theoretical aspects, it is instructive to write the main disturbing forces in terms of the orbital elements.

In this section we obtain the averaged disturbing function of the Sun. Following the classical procedure [5], in a reference center fixed in the Earth equator, the disturbing function of the Sun is

$$R_\odot = k^2 M_\odot \left(\frac{1}{|\vec{r} - \vec{r}_\odot|} - \frac{\vec{r} \cdot \vec{r}_\odot}{|\vec{r}_\odot|^3} \right), \tag{2.1}$$

where $M_\odot$ is the mass of the Sun, k^2 is the gravitational constant, and $\vec{r}$, $\vec{r}_\odot$ are position vector of the satellite and the Sun, respectively.

Expanding (2.1) in powers of $(r/r_\odot)$ up to order 2, we have

$$R_\odot = k^2 M_\odot \frac{r^2}{r_\odot^3} \left(-\frac{1}{2} + \frac{3}{2}\cos^2(S) \right), \tag{2.2}$$

where S is the angular distance between the satellite and the Sun. We use the classical notations a, e, I, f, ω, and Ω, for semi-major axis, eccentricity, inclination, true anomaly, argument of the perigee, and longitude of the node. The same set is used for the Sun's elements, adding the index $\odot$.

In order to write $\cos(S)$ in terms of orbital elements, we have (see Figure 1)

$$\cos(S) = \frac{x}{r}\frac{x_\odot}{r_\odot} + \frac{y}{r}\frac{y_\odot}{r_\odot} + \frac{z}{r}\frac{z_\odot}{r_\odot}. \tag{2.3}$$

Considering classical relations of the two-body problem, we write $\cos(S)$ in terms of f, $f_\odot$, Ω, $\Omega_\odot$, ω, $\omega_\odot$, I, $I_\odot$ as follows:

$$\begin{aligned}
\cos(S) = {} & \frac{1}{4}(1+\cos(I))(1-\cos(I_\odot))\cos\left(f+\omega+f_\odot+\omega_\odot+\Omega-\Omega_\odot\right) \\
& + \frac{1}{4}(1-\cos(I))(1+\cos(I_\odot))\cos\left(f+\omega+f_\odot+\omega_\odot-\Omega+\Omega_\odot\right) \\
& + \frac{1}{4}(1+\cos(I))(1+\cos(I_\odot))\cos\left(f+\omega-f_\odot-\omega_\odot+\Omega-\Omega_\odot\right) \\
& + \frac{1}{4}(1-\cos(I))(1-\cos(I_\odot))\cos\left(f+\omega-f_\odot-\omega_\odot-\Omega+\Omega_\odot\right) \\
& + \frac{1}{2}\sin(I)\sin(I_\odot)\left[\cos\left(f+\omega-f_\odot-\omega_\odot\right)-\cos\left(f+\omega+f_\odot+\omega_\odot\right)\right],
\end{aligned} \tag{2.4}$$

or in a compact form as follows:

$$\cos(S) = Aa + Bb + Cc + Dd + Ee, \tag{2.5}$$

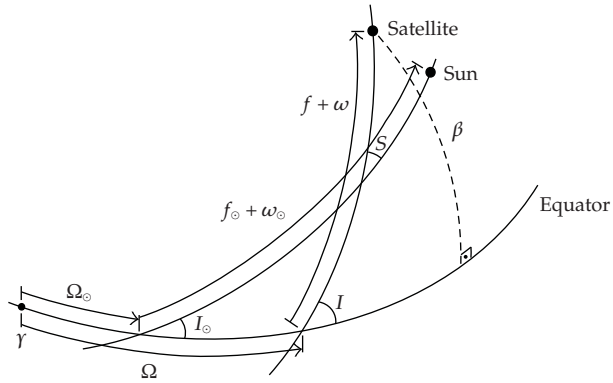


Figure 1: Geometry of the problem.

where

$$
\begin{aligned}
A &= \frac{1}{4}(1+\cos(I))(1-\cos(I_\odot)),\\
a &= \cos\left(f+\omega+f_\odot+\omega_\odot+\Omega-\Omega_\odot\right),\\
B &= \frac{1}{4}(1-\cos(I))(1+\cos(I_\odot)),\\
b &= \cos\left(f+\omega+f_\odot+\omega_\odot-\Omega+\Omega_\odot\right),\\
C &= \frac{1}{4}(1+\cos(I))(1+\cos(I_\odot)),\\
c &= \cos\left(f+\omega-f_\odot-\omega_\odot+\Omega-\Omega_\odot\right),\\
D &= \frac{1}{4}(1-\cos(I))(1-\cos(I_\odot)),\\
d &= \cos\left(f+\omega-f_\odot-\omega_\odot-\Omega+\Omega_\odot\right),\\
E &= \frac{1}{2}\sin(I)\sin(I_\odot),\\
e &= \cos\left(f+\omega-f_\odot-\omega_\odot\right)-\cos\left(f+\omega+f_\odot+\omega_\odot\right).
\end{aligned}
\tag{2.6}
$$

In order to get rid of the short period variations, we have to obtain the averaged system and the rigorous procedure is to apply the classical von-Zeipel or Hori's method [5, 6]. In the present case, as our interest is only to examine the long-term behavior, without

retrieving the contribution of the short period terms eliminated during the averaging process, the secular disturbing function can be found simply considering [7, 8]

$$\langle R_\odot \rangle = \frac{1}{2\pi}\int_0^{2\pi} R_\odot dl, \tag{2.7}$$

where l is the mean anomaly of the satellite. Consider

$$\begin{aligned}
R_\odot^* &= \langle R_\odot \rangle \\
&= \frac{k^2 M_\odot a^2}{2r_\odot^3} \\
&\quad \times \left[\frac{3}{2}P\left(A^2 + B^2 + C^2 + D^2 + 2E^2 - \frac{2}{3}\right)\right. \\
&\quad + \frac{3}{2}A^2 Z\cos\left(2\omega + 2f_\odot + 2\omega_\odot + 2\Omega - 2\Omega_\odot\right) + \frac{3}{2}C^2 Z\cos\left(2\omega - 2f_\odot - 2\omega_\odot + 2\Omega - 2\Omega_\odot\right) \\
&\quad + \frac{3}{2}B^2 Z\cos\left(2\omega + 2f_\odot + 2\omega_\odot - 2\Omega + 2\Omega_\odot\right) + \frac{3}{2}D^2 Z\cos\left(2\omega - 2f_\odot - 2\omega_\odot - 2\Omega + 2\Omega_\odot\right) \\
&\quad + \frac{3}{2}Z\left(E^2 + 2CD\right)\cos\left(2\omega - 2f_\odot - 2\omega_\odot\right) + \frac{3}{2}Z\left(E^2 + 2AB\right)\cos\left(2\omega + 2f_\odot + 2\omega_\odot\right) \\
&\quad + 3Z\left(-E^2 + AD + BC\right)\cos(2\omega) + 3P\left(-E^2 + AC + BD\right)\cos\left(2f_\odot + 2\omega_\odot\right) \\
&\quad + 3P(AB + CD)\cos(2\Omega - 2\Omega_\odot) + 3ACZ\cos(2\omega + 2\Omega - 2\Omega_\odot) \\
&\quad + 3ADP\cos\left(2f_\odot + 2\omega_\odot + 2\Omega - 2\Omega_\odot\right) + 3EP(A - D)\cos\left(2f_\odot + 2\omega_\odot + \Omega - \Omega_\odot\right) \\
&\quad + 3EP(-A - B + C + D)\cos(\Omega - \Omega_\odot) + 3EZ(A - C)\cos(2\omega + \Omega - \Omega_\odot) \\
&\quad - 3AEZ\cos\left(2\omega + 2f_\odot + 2\omega_\odot + \Omega - \Omega_\odot\right) + 3BCP\cos\left(2f_\odot + 2\omega_\odot - 2\Omega + 2\Omega_\odot\right) \\
&\quad + 3BDZ\cos(2\omega - 2\Omega + 2\Omega_\odot) + 3EP(B - C)\cos\left(2f_\odot + 2\omega_\odot - \Omega + \Omega_\odot\right) \\
&\quad + 3EZ(B - D)\cos(2\omega - \Omega + \Omega_\odot) - 3BEZ\cos\left(2f_\odot + 2\omega + 2\omega_\odot - \Omega + \Omega_\odot\right) \\
&\quad \left. + 3CEZ\cos\left(2\omega - 2f_\odot - 2\omega_\odot + \Omega - \Omega_\odot\right) + 3DEZ\cos\left(2\omega - 2f_\odot - 2\omega_\odot - \Omega + \Omega_\odot\right)\right],
\end{aligned} \tag{2.8}$$

where $P = 1 + (3/2)e^2$, $Z = (5/2)e^2$.

Performing a second and similar average with respect to the mean anomaly of the Sun, we get

$$\begin{aligned}\widehat{R}_{\odot} = {} & \frac{k^2 M_{\odot} a^2}{4 r_{\odot}^3} \\ & \times \left[\frac{P}{4}\left(1 - 3\cos^2(I) - 3\cos^2(I_{\odot}) + 9\cos^2(I)\cos^2(I_{\odot})\right)\right. \\ & + \frac{3}{2} Z \sin^2(I)\left(-1 + 3\cos^2(I_{\odot})\right)\cos(2\omega) \\ & + \frac{3}{2} P \sin^2(I)\sin^2(I_{\odot})\cos(2\Omega - 2\Omega_{\odot}) \\ & + \frac{3}{8} Z (1 + \cos(I))^2 \sin^2(I_{\odot})\cos(2\omega + 2\Omega - 2\Omega_{\odot}) \\ & - \frac{3}{2} Z \sin(I)\sin(I_{\odot})(1 + \cos(I))\cos(I_{\odot})\cos(2\omega + \Omega - \Omega_{\odot}) \\ & + 3P \sin(I)\cos(I)\sin(I_{\odot})\cos(I_{\odot})\cos(\Omega - \Omega_{\odot}) \\ & + \frac{3}{8} Z \left(1 + \cos^2(I)\right)^2 \sin^2(I_{\odot})\cos(2\omega - 2\Omega + 2\Omega_{\odot}) \\ & \left. + \frac{3}{2} Z \sin(I)(1 - \cos(I))\sin(I_{\odot})\cos(I_{\odot})\cos(2\omega - \Omega + \Omega_{\odot})\right]. \end{aligned} \tag{2.9}$$

In the above expression, the orbit of the Sun is assumed to be a Keplerian circular orbit. The elliptic case is briefly discussed in Appendix A.

2.2. Oblateness Disturbing Function

For the oblateness, the disturbing function truncated at second order of R_P/r is

$$U_2 = \frac{k^2 M_T R_P^2}{r^3} J_2 \left(\frac{1}{2} - \frac{3}{2}\sin^2(\beta)\right), \tag{2.10}$$

where R_P, J_2, and β are equatorial radius of the planet, oblateness coefficient, and geocentric latitude of the satellite, respectively. If we proceed in the exact same way as we did before, we have from the geometry of Figure 1

$$\sin(\beta) = \sin(I)\sin(f + \omega). \tag{2.11}$$

Once β is eliminated, the average of U_2 with respect to the mean anomaly of satellite is [9, 10]

$$\begin{aligned} \langle U_2 \rangle &= \frac{1}{2\pi}\int_0^{2\pi} U_2 dl, \\ R_{J_2} = \langle U_2 \rangle &= \frac{1}{4} n^2 J_2 R_P^2 \left(3\cos^2(I) - 1\right)\left(1 - e^2\right)^{-3/2}, \end{aligned} \tag{2.12}$$

where n is the mean motion of the satellite.

2.3. Some Special Resonances

For close satellites, usually the oblateness is the dominant part. In this case, the main frequencies of the system are given by

$$\begin{aligned} \dot{\omega} &\approx \frac{3nJ_2R_P^2}{4a^2(1-e^2)^2}\left(5\cos^2(I) - 1\right), \\ \dot{\Omega} &\approx -\frac{3nJ_2R_P^2}{2a^2(1-e^2)^2}\cos(I). \end{aligned} \tag{2.13}$$

The ratio of these two frequencies is

$$\frac{\dot{\Omega}}{\dot{\omega}} \approx \frac{2\cos(I)}{1 - 5\cos^2(I)} = k. \tag{2.14}$$

Note that for k = integer, we have the special resonances which do not depend on the semi-major axis. These resonances usually affect the eccentricity [8]. For $k = -2$, we have $2\dot{\omega} + \dot{\Omega} \approx 0$ for $I = 56.06°$ or $I = 110.99°$. Another classical resonance occurs when $I = 63.4°$, so that $\dot{\omega} \approx 0$.

3. Effects of $2\dot{\omega} + \dot{\Omega}$ and $\dot{\omega}$ Resonances

In order to see the effects of the resonances which affect GPS and Galileo satellites, the osculating equations of a satellite will be integrated. For the moment, as disturbers, we consider only the Sun and the oblateness (the complete Cartesian equations involving the remaining disturbers will be given in Section 4). Note that the resonant conditions to be used this time are extracted from the averaged system as presented in the precedent section ($I = 56.06°$, $I = 63.4°$). Figures 2 and 3 show the effects of both resonances on the eccentricity and on the resonant angles. Note that an initial small eccentricity reaches a significant increase.

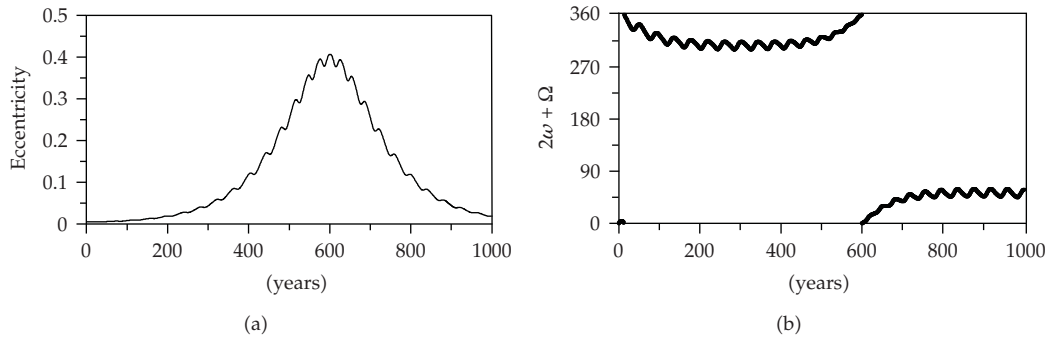


Figure 2: Time evolution of the eccentricity (a) and the critical angle (b). Initial conditions: $a = 4.805\, R_T$ (30,647 km), $e = 0.005$, $I = 56.06°$, and other elements equal to zero. In the simulations, we consider only Sun and oblateness as disturbers.

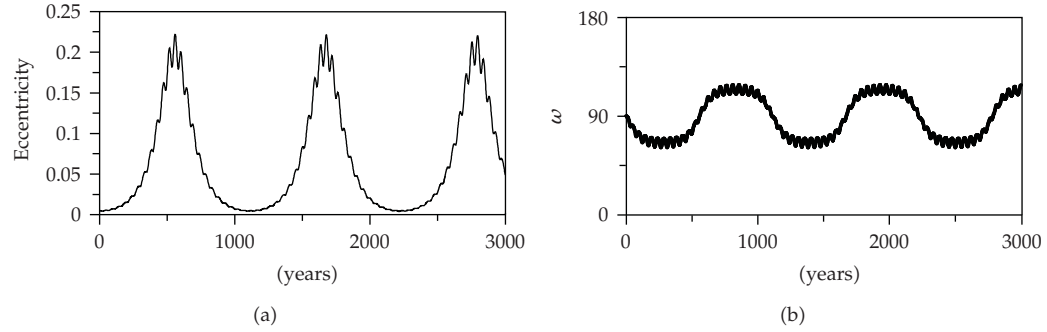


Figure 3: Same as Figure 1. Initial conditions: $a = 4.7\, R_T$ ($\approx$ 29,977 km), $e = 0.005$, $I = 63.4°$, $\omega = 90°$, and other elements equal to zero.

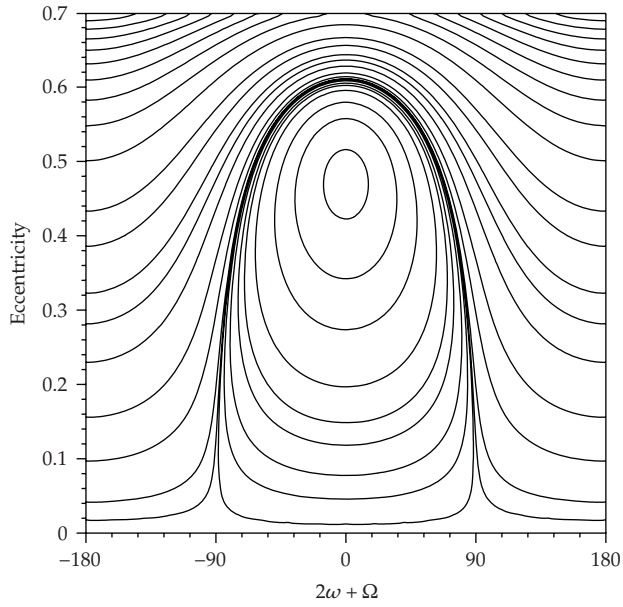


Figure 4: Level curves of Hamiltonian $(\widetilde{R}_{\odot} + \widetilde{R}_{J_2})$, showing the eccentricity variation versus resonant angle.

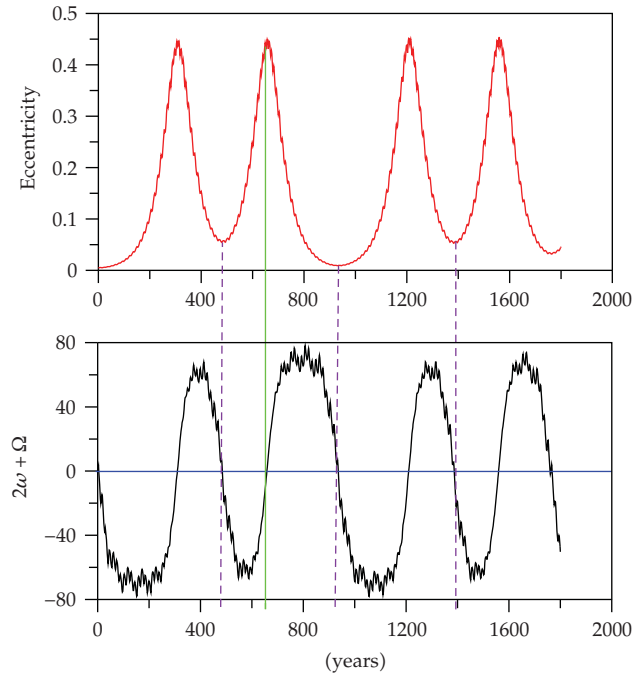


Figure 5: Time evolution of the osculating eccentricity (top) and the osculating critical angle (bottom) for a disposal GPS satellite. Note that the minimum of the eccentricity occurs when $2\omega + \Omega$ is crossing zero in decreasing direction (from positive to negative value) while maximum occurs when $2\omega + \Omega$ is crossing zero, but in increasing direction (from negative to positive value). Initial conditions: $a = 3.5\ \mathrm{R}_T$ (22,324 km), $e = 0.005$, $I = 56.06°$, and other elements equal to zero; $I_L = 28.58°$.

Let us pay more attention to the case $I = 56.06°$ which is the inclination of the members of the Galileo constellation. For this inclination, the dominant term in the $R^*_\odot$ is $\cos(2\omega + \Omega - \Omega_\odot)$. Neglecting the remaining terms of $\widehat{R}_\odot$, the Hamiltonian of the problem is

$$F = R_{J_2} + \frac{k^2 M_\odot a^2}{2r_\odot^3}\left[\frac{P}{8}\left(1 - 3\cos^2(I) - 3\cos^2(I_\odot) + 9\cos^2(I)\cos^2(I_\odot)\right)\right.$$
$$\left. -\frac{3}{4}Z\sin(I)\sin(I_\odot)(1+\cos(I))\cos(I_\odot)\cos(2\omega + \Omega - \Omega_\odot)\right]. \tag{3.1}$$

Let us take $L = \sqrt{k^2(M_T + m)a}$, $G = L\sqrt{1-e^2}$, $H = G\cos(I)$, l = mean anomaly, $\omega = g$, and $\Omega = h$ the set of the Delaunay variables. After a trivial Mathieu canonical transformation [11],

$$(G, H, \omega, \Omega) \longrightarrow (P_1, P_2, \theta_1, \theta_2), \tag{3.2}$$

where

$$\theta_1 = 2\omega + \Omega, \qquad P_1 = \frac{G}{2}, \qquad \theta_2 = \Omega, \qquad P_2 = H - \frac{G}{2}, \tag{3.3}$$

then we have

$$\widetilde{R}_{\odot} = \frac{k^2 M_{\odot} a^2}{2 r_{\odot}^3} \left[\frac{P}{8} \left(1 - 3\frac{(P_1+P_2)^2}{4P_1^2} - 3\cos^2(I_{\odot}) + 9\frac{(P_1+P_2)^2}{4P_1^2}\cos^2(I_{\odot}) \right) \right.$$
$$\left. - \frac{3}{4} Z \left(1 - \frac{(P_1+P_2)^2}{4P_1^2} \right)^{1/2} \sin(I_{\odot}) \left(1 + \frac{P_1+P_2}{2P_1} \right) \cos(I_{\odot}) \cos(\theta_1) \right], \tag{3.4}$$
$$\widetilde{R}_{J_2} = \frac{1}{4} n^2 J_2 R_P^2 \left(3\frac{(P_1+P_2)^2}{4P_1^2} - 1 \right) \left(1 - \frac{L^2 - 4P_1^2}{L^2} \right)^{-3/2}.$$

Since Sun's orbit is a Keplerian one, we also considered $\omega_{\odot} = 0$, $\Omega_{\odot} = 0$. In (P_i, θ_i) variables, our Hamiltonian is a one degree of freedom problem, whose dynamics is very similar to the very well-known Lidov-Kozai resonance. In Figure 4, we consider an initial eccentricity $e_0 = 0.005$ and semi-major axis $a = 4.805\, R_T$ (30,647 km). This figure is very instructive: note that in the bottom part there is a large region where the satellite remains some finite time with very small eccentricity. These are the exactly region we are looking for. It corresponds to the region where $2\omega + \Omega \approx 0$. On the other hand, we have the counterpart of this situation at the top of the figure: very high eccentricity, which occurs again for $2\omega + \Omega \approx 0$. We can separate these two configurations and have a close view of these two cases. Only to confirm our reasoning, let us integrate the problem in Cartesian coordinates. We also have to decrease the effect of the Moon's perturbation since in the averaged analysis we considered only $\widetilde{R}_{\odot}$ and $\widetilde{R}_{J_2}$. To do that, we consider convenient value for the semi-major axis. Figure 5 (initial conditions: $a = 3.5\, R_T$ ($\approx$ 22323 km), $e = 0.005$, $I = 56.06°$, and other elements equal to zero; Moon inclination $I_L = 18.28°$) shows that the minimum of eccentricity occurs when $2\omega + \Omega$ crosses zero from positive to negative values (decreasing direction), while maximum occurs when $2\omega + \Omega$ crosses zero from negative to positive values (increasing direction). It is worth noting that if the semi-major axis is high, then the effect of the Moon cannot be neglected, so that the problem is no more a one degree of freedom problem. In this case the search of the (ω, Ω) pair, such that eccentricity remains small, must be done integrating the complete equations of the motion.

4. Special (ω, Ω) Initial Conditions for Galileo Case

In the previous section, we considered only the effects of the Sun and of the oblateness. Moreover, in the presence of the resonance, the main effects are governed by the long term variations, so that we eliminated the short period terms. From a theoretical point of view, this averaged system is quite efficient to highlight the basic dynamics that affects the eccentricity of the GPS and Galileo satellites. However, for a more complete and realistic study, we need to include more disturbers.

In this section we want to find some special initial conditions such that the satellites can remain stable for at least 250 years with very small eccentricity without causing any risk of collision to the operational elements of the constellation. The strategy to search these

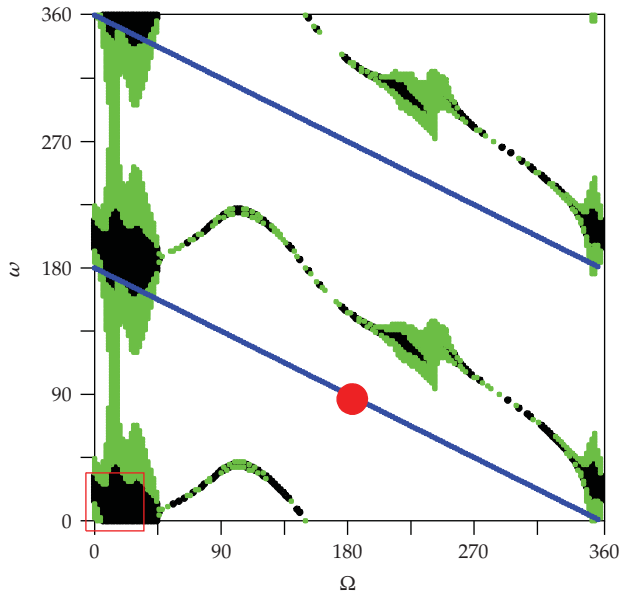


Figure 6: Black dots: represent (ω, Ω) values such that a satellite with $a = 30{,}647$ km remains at least 250 years with $e_{\text{MAX}} \leq 0.01$. Green dots: the same, but $e_{\text{MAX}} \leq 0.02$. Blue dots: curve satisfying $2\omega + \Omega = 0$. Moon's inclination: $I_L = 18.28°$. Note that most of the "stable" (black dots) (ω, Ω) points satisfy $2\omega + \Omega = 0$, 2π with $\Omega \approx 0$, $\omega = \pi$.

particular initial conditions is guided from the theoretical approach described in the previous section.

In this section we integrate the osculating elements of a disposal satellite of the Galileo system under the effect of the Sun, Moon, and the oblateness.

The equations for the osculating elements (exact system), including Moon, are

$$\ddot{\vec{r}} = -\frac{k^2(M+m)}{r^3}\,\vec{r} - k^2 M_\odot\left(\frac{\vec{r}-\vec{r}_\odot}{|\vec{r}-\vec{r}_\odot|^3} - \frac{\vec{r}_\odot}{|\vec{r}_\odot|^3}\right) - k^2 M_L\left(\frac{\vec{r}-\vec{r}_L}{|\vec{r}-\vec{r}_L|^3} - \frac{\vec{r}_L}{|\vec{r}_L|^3}\right) + \vec{P}_{J_2}, \tag{4.1}$$

$$P_{J_x} = -k^2 M J_2 R_P^2\left[\frac{3x}{2r^5} - \frac{15}{2}\frac{z^2 x}{r^7}\right], \tag{4.2}$$

$$P_{J_y} = -k^2 M J_2 R_P^2\left[\frac{3y}{2r^5} - \frac{15}{2}\frac{z^2 y}{r^7}\right], \tag{4.3}$$

$$P_{J_z} = -k^2 M J_2 R_P^2\left[\frac{9z}{2r^5} - \frac{15}{2}\frac{z^3}{r^7}\right], \tag{4.4}$$

where $\vec{P}_{J_2}$ is the acceleration due to the oblateness, whose Cartesian components are given by P_{J_x}, P_{J_y}, P_{J_z} [10]. The second and third terms in (3.3) are the contributions of the Sun and the Moon, respectively, and M, m, M_L are the masses of the Earth, satellite, and Moon,

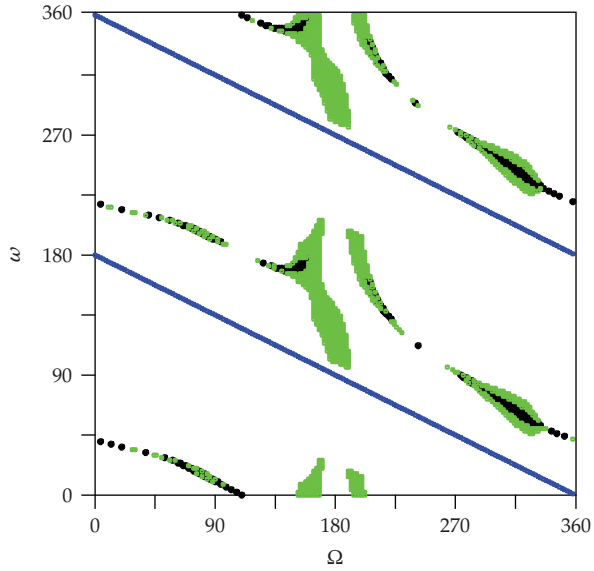


Figure 7: Same of Figure 6, but now $I_L = 28.58°$. Note that the distribution of the stable (ω, Ω) is very sensitive under changes in I_L.

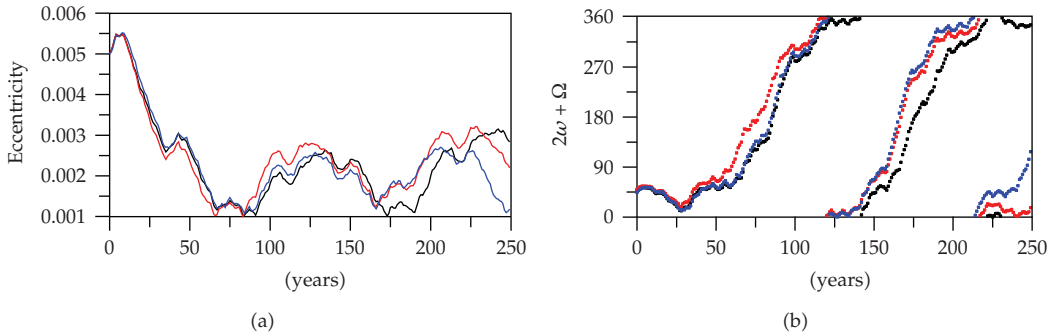


Figure 8: Time evolution of the eccentricity (a) and the critical angle (b) obtained from integration of (4.1). Initial conditions: $a = 4.805\ R_T$, $e = 0.005$, $I = 56.06°$. Initial (ω, Ω): (24°, 0°): black, (23°, 2°): red, (18°, 8°): blue. These initial conditions were extracted from the red square shown in Figure 6. Initial conditions of the Moon: $a_L = 380$, 367.2 km, $e_L = 0.0276$, $I_L = 18.28°$, $\Omega_L = 12.11°$, $\omega_L = 92°$, and $l_L = 337°$. As expected, the eccentricity remains less than 0.005 for at least 250 years.

respectively. The position vectors of the satellite, Sun, and Moon are indicated by $\vec{r}$, $\vec{r}_\odot$, and $\vec{r}_L$.

As we said before, we take 500 km above of the nominal altitude of the constellation. The initial elements are fixed to $a = 4.805 R_T$ (30,647 km), $e = 0.005$, $l = 0°$, and $I = 56.06°$. We consider two cases for the Moon's inclination, $I = 18.28°$ and 28.58°. We show that the initial value of the inclination is important as shown in Figures 6 and 7. In these figures, we show the pair (ω, Ω) such that the disposal object remains at least 250 years with eccentricity smaller than 0.01, so that there is no risk of collision with any member of the constellation. The black region corresponds to initial conditions such that the maximum eccentricity is less than 0.01. In the green region, the maximum eccentricity is less than 0.02. The two

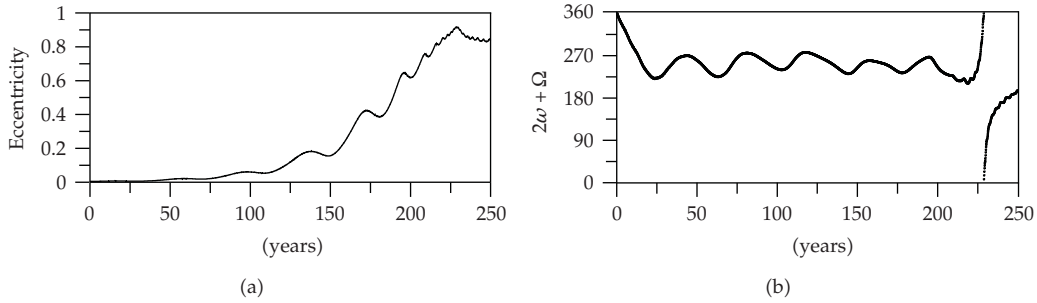


Figure 9: Same as Figure 8 but considering initial conditions from the red small circle of Figure 6. Initial conditions: $a = 4.805\, R_T$ (30,647 km), $e = 0.005$, $I = 56.06°$, $\omega = 90°$, $\Omega = 180°$. This time eccentricity grows fastly since the corresponding initial condition was not taken from the black region of Figure 6.

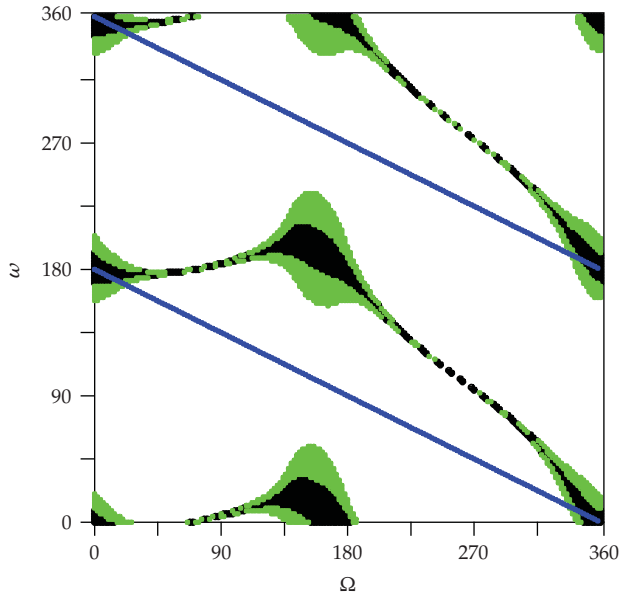


Figure 10: Same of Figure 6, but now $a = 26{,}060$ km.

straight lines represent the exact condition $2\omega + \Omega = k\pi$ (in particular we only plot the case $k = 0$). Note that, in special, the black dots (Figure 6) are in fact located in places predicted from the previous theoretical model. For the remaining figures, the black dots are slightly shifted (upward) from the line $2\omega + \Omega = 0$. We believe that this is caused by the strong perturbation of the Moon. Figure 8 shows the time evolution of the eccentricity for integration whose initial conditions are obtained from Figure 6 (small square in the bottom). As expected, the eccentricity remains very low, while if we take (ω, Ω) outside the marked regions in Figure 6 or Figure 7, a significant increase is verified as shown in Figure 9. The initial conditions (ω, Ω) used in this case correspond to the red circle given in Figure 6.

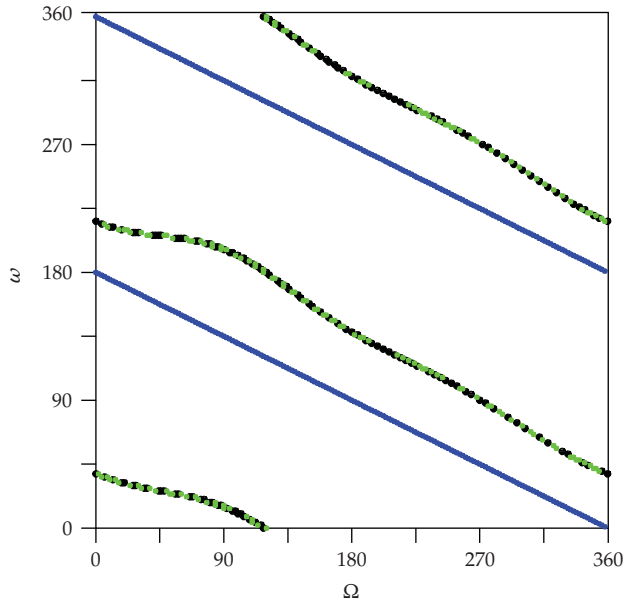


Figure 11: Same of Figure 10, but now $I_L = 28.58°$.

5. (ω, Ω) Conditions for GPS Case

Here let us consider the GPS system. Again we take $I = 18.28$ and $I = 28.58$ for the Moon's inclination. As before, the importance of the Moon's inclination is clear.

6. Tesseral and Sectorial Harmonics

Up to now, we have not considered tesseral and sectorial harmonics. Since GPS satellites have orbital period near to 12:00 h, the inclusion of such harmonics must be examined when drawing the figures of Section 5. While the numerical values of the coefficients of these harmonics are very small compared to the zonal harmonics, due to the $1n : 2\gamma$ resonance (γ is the rotation velocity of the Earth), this contribution could cause some nonnegligible effects. We show very briefly these perturbations. There are several models of the geopotential, including most sophisticated recursive formulae to generate very high-order (JGM-3 [12], EGM96S, EGM96 [13], etc.). In this work, we do not need high order model, so that we consider only few terms. The disturbing function for the general geopotential can be written in the form [14, 15]

$$V = \frac{k^2 M_T}{r} + \frac{k^2 M_T}{r}\left\{-\sum_{n=2}^{\infty}\left(\frac{R_P}{r}\right)^n J_n P_{n0}(\sin(\phi)) + \sum_{n=2}^{\infty}\sum_{m=1}^{n}\left(\frac{R_P}{r}\right)^n J_{nm}\cos m(\lambda - \lambda_{mn}) P_{nm}(\sin(\phi))\right\}, \tag{6.1}$$

where J_{nm}, λ_{nm} are numerical coefficients and P_{nm} are the associated functions of Legendre.

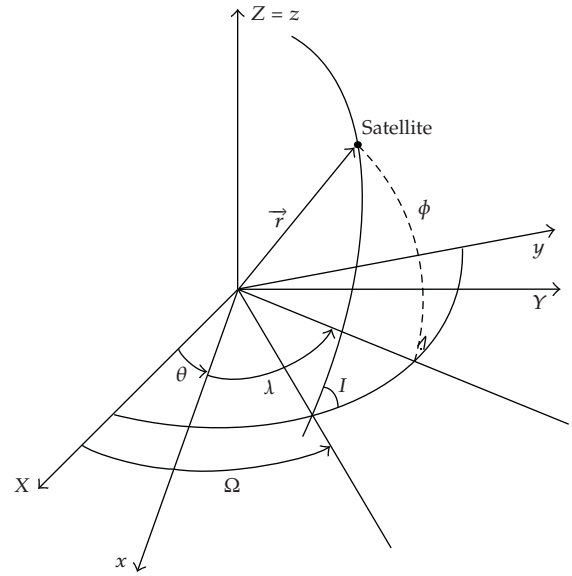


Figure 12: Geometry of the problem.

Figure 12 describes the fundamental axes of the following reference: the potential (6.1) is referred to the (x, y, z) which is an equatorial system fixed on the Earth; therefore, it rotates with respect to (X, Y, Z) which is an inertial system, as follows:

$\vec{r}$: the position vector of the satellite.

I: inclination.

ϕ: geocentric latitude.

λ: longitude.

Ω: longitude of node.

According to Figure 12, some simple geometrical relations can be written as

$$\begin{aligned} x &= r\cos(\phi)\cos(\lambda), \\ y &= r\cos(\phi)\sin(\lambda), \\ z &= r\sin(\phi), \\ \cos(2\lambda) &= \frac{x^2 - y^2}{x^2 + y^2}, \qquad \sin(2\lambda) = \frac{2xy}{x^2 + y^2}, \\ \cos^2(\lambda) &= \frac{x^2}{x^2 + y^2}, \qquad \sin^2(\lambda) = \frac{y^2}{x^2 + y^2}, \\ x^2 + y^2 &= r^2\cos^2(\phi). \end{aligned} \tag{6.2}$$

Let us define $\varphi_{nm} = (k^2 M_P/r)(R_P/r)^n J_{nm} \cos m(\lambda - \lambda_{nm}) P_{nm}(\sin(\phi))$.

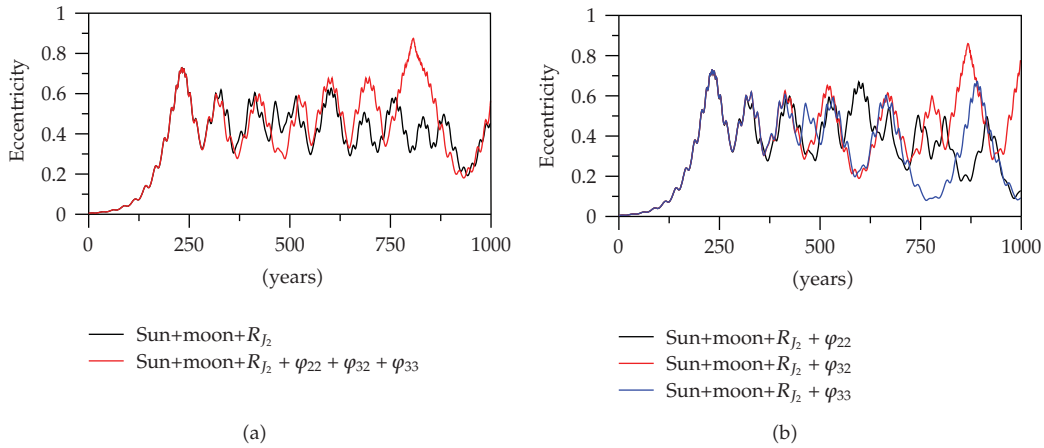


Figure 13: Time evolution of eccentricity of GPS satellite. Initial conditions: a = 27,059.74 km, e = 0.005, ω = 58°, Ω = 154°. The initial conditions of the Moon correspond to the epoch August 1, 2001. The contribution of individual and all φ_{nm} is shown through different colors. The differences appear only after 370 years.

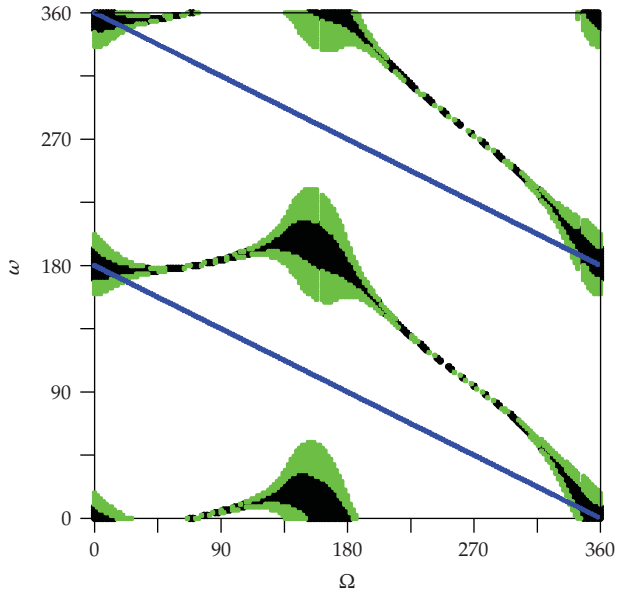


Figure 14: Same of Figure 10, including φ_{22}, φ_{32}, and φ_{33}.

Therefore,

$$\varphi_{22} = \frac{3k^2 M_P R_P^2}{r^5} J_{22}\left[\left(x^2 - y^2\right)\cos(2\lambda_{22}) + 2xy\sin(2\lambda_{22})\right]. \tag{6.3}$$

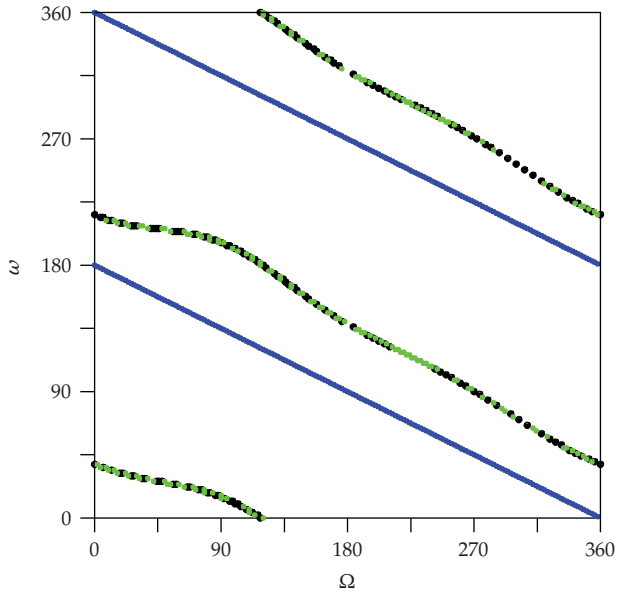


Figure 15: Same of Figure 14, but now $I_L = 28.58°$.

Proceeding in the similar way, we obtain

$$\varphi_{32} = \frac{15k^2 M_P R_P^3}{r^7} J_{32}\left[\left(x^2 - y^2\right) z\cos(2\lambda_{32}) + 2xyz\sin(2\lambda_{32})\right],$$
$$\varphi_{33} = \frac{15k^2 M_P R_P^3}{r^7} J_{33}\left[x\left(x^2 - 3y^2\right)\cos(3\lambda_{33}) + y\left(3x^2 - y^2\right)\sin(3\lambda_{33})\right]. \tag{6.4}$$

Note that the zonal terms ($m = 0$) were already considered before. In Figure 12 the system (x, y, z) is fixed on the Earth while (X, Y, Z) is an inertial system, so that (x, y, z) rotates with respect to (X, Y, Z). Therefore, to have φ_{nm} referred to (X, Y, Z), we consider the trivial rotation:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \tag{6.5}$$

where $\theta = \gamma t + \theta_0$, $\gamma = 2\pi/\text{day}$.

Therefore considering only φ_{22}, φ_{32}, and φ_{33}, the force to be added in (4.1) is

$$P_\varphi = \text{grad}_{XYZ}\left(\varphi_{22} + \varphi_{32} + \varphi_{33}\right). \tag{6.6}$$

Once we have introduced XYZ system, it is clear that several angular combinations like $2\theta - \lambda$ appear in the above φ_{nm}, when (6.3) and (6.4) are expressed in the classical orbital

elements. Due to the nominal semi-major axis of the GPS satellites, $2\theta - \lambda$ will generate long-term variations due to their proximity of the 2 : 1 exact resonance. In fact, the orbital period of GPS satellite is about 12:00 h.

To express φ_{nm} in terms of the orbital elements is trivial; however, the best way to see the effects is to keep these terms in Cartesian coordinates without any expansion in eccentricity or inclination, as the former increases to high values, while the second is essentially high from the beginning.

The effects of these additional terms are shown in Figure 13. The interaction of several arguments coming from $\varphi_{22}, \varphi_{32}, \varphi_{33}$, and also those related to the lunar disturbing function, give rise to new resonant combinations. It is interesting that the presence of different φ_{nm} cause significative differences in the behavior of the eccentricity, however, in the beginning (up to t = 370 years), all the curves are almost coincident. In particular, our numerical experiments show that if eccentricity remains small, the effects of φ_{nm} are not significant.

Figures 14 and 15 show that the (Ω, ω) initial conditions are similar to those figures of Section 4. Note that differences when φ_{nm} are included are almost negligible as expected according to what we learned from Figure 13. However, in the strategy of exploiting the growth of the eccentricity, the initial (Ω, ω) can be sensitive, depending on the time integration and on the number of the harmonics considered in the geopotential. In Appendix B, we show similar figures, where (Ω, ω) values are the initial conditions that cause fast increase of the eccentricity in less than 250 years.

7. Conclusion

With the averaged equations, we showed the dynamics of the $2\omega + \Omega$ resonance. The reason of the increase of the eccentricity is essentially due to this resonance which does not depend on the value of the semi-major axis. Therefore, any change of the semi-major axes (raising the perigee) of the decommissioned object will not remove from the resonance. After showing the existence of some initial conditions in the (ω, Ω) domain where the eccentricity can remain very small based on the averaged simplified model, we used the complete set of equations to search this pair in (ω, Ω) plane. The importance of the Moon's inclination becomes clear as shown in Figures 6, 7, 10, and 11. We obtained these initial values for GALILEO and GPS systems. For completeness, we also derived a first-order averaged system in the eccentricity of the third body (Appendix A). Then several additional resonances appear although their effects are not so relevant for the navigation system. The search of the (ω, Ω) pair for the maximum increase of the eccentricity can be done straightforward following the same procedure we used for minimum eccentricity. For completeness, in the disturbing function of the geopotential we also included terms coming from J_{22}, J_{32}, and J_{33}. However, their contributions for the first 370 years are not significant. In a separate paper, we intend to show the corresponding Figures 10 and 11 considering the second strategy of exploiting the increase of the eccentricity. Our experiments show that the effect of J_{22}, J_{32}, and J_{33} terms becomes more visible only after some hundred years. Therefore, their contributions in Figures 14 and 15 are almost negligible. However, considering long-time integration, their effects and interaction with solar and lunar perturbations become important.

Appendices

A. Solar Disturbing Function (Up to First Order in e)

Here we give the complete expression of the averaged disturbing function up to first order in eccentricity of the third body:

$$
\begin{aligned}
R_{\odot}^{1} = {} & \frac{k^2 M_{\odot}^3 a^3}{2a_{\odot}^3} e_{\odot} \\
& \times \left[\frac{9}{2} P \left(A^2 + B^2 + C^2 + D^2 + 2E^2 - \frac{2}{3} \right) \cos l_{\odot} + \frac{9}{4} A^2 Z \cos(2\omega + l_{\odot} + 2\omega_{\odot} + 2\Omega - 2\Omega_{\odot}) \right. \\
& + \frac{9}{4} A^2 Z \cos(2\omega + 3l_{\odot} + 2\omega_{\odot} + 2\Omega - 2\Omega_{\odot}) + \frac{9}{4} B^2 Z \cos(2\omega + l_{\odot} + 2\omega_{\odot} - 2\Omega + 2\Omega_{\odot}) \\
& + \frac{9}{4} B^2 Z \cos(2\omega + 3l_{\odot} + 2\omega_{\odot} - 2\Omega + 2\Omega_{\odot}) + \frac{9}{4} C^2 Z \cos(2\omega - 3l_{\odot} - 2\omega_{\odot} + 2\Omega - 2\Omega_{\odot}) \\
& + \frac{9}{4} C^2 Z \cos(2\omega - l_{\odot} - 2\omega_{\odot} + 2\Omega - 2\Omega_{\odot}) + \frac{9}{4} D^2 Z \cos(2\omega - 3l_{\odot} - 2\omega_{\odot} - 2\Omega + 2\Omega_{\odot}) \\
& + \frac{9}{4} D^2 Z \cos(2\omega - l_{\odot} - 2\omega_{\odot} - 2\Omega + 2\Omega_{\odot}) + \frac{9}{4} Z \left(E^2 + 2CD \right) \cos(2\omega - 3l_{\odot} - 2\omega_{\odot}) \\
& + \frac{9}{4} Z \left(E^2 + 2CD \right) \cos(2\omega - l_{\odot} - 2\omega_{\odot}) + \frac{9}{4} Z \left(E^2 + 2AB \right) \cos(2\omega + l_{\odot} + 2\omega_{\odot}) \\
& + \frac{9}{4} Z \left(E^2 + 2AB \right) \cos(2\omega + 3l_{\odot} + 2\omega_{\odot}) + \frac{9}{2} Z \left(-E^2 + AD + BC \right) \cos(l_{\odot} - 2\omega) \\
& + \frac{9}{2} Z \left(-E^2 + AD + BC \right) \cos(l_{\odot} + 2\omega) + \frac{9}{2} P \left(-E^2 + AC + BD \right) \cos(l_{\odot} + 2\omega_{\odot}) \\
& + \frac{9}{2} P \left(-E^2 + AC + BD \right) \cos(3l_{\odot} + 2\omega_{\odot}) + \frac{9}{2} P(AB + CD) \cos(l_{\odot} - 2\Omega + 2\Omega_{\odot}) \\
& + \frac{9}{2} P(AB + CD) \cos(l_{\odot} + 2\Omega - 2\Omega_{\odot}) + \frac{9}{2} ACZ \cos(l_{\odot} - 2\omega - 2\Omega + 2\Omega_{\odot}) \\
& + \frac{9}{2} ACZ \cos(l_{\odot} + 2\omega + 2\Omega - 2\Omega_{\odot}) + \frac{9}{2} ADP \cos(l_{\odot} + 2\omega_{\odot} + 2\Omega - 2\Omega_{\odot}) \\
& + \frac{9}{2} ADP \cos(3l_{\odot} + 2\omega_{\odot} + 2\Omega - 2\Omega_{\odot}) + \frac{9}{2} EP(A - D) \cos(l_{\odot} + 2\omega_{\odot} + \Omega - \Omega_{\odot}) \\
& + \frac{9}{2} EP(A - D) \cos(3l_{\odot} + 2\omega_{\odot} + \Omega - \Omega_{\odot}) + \frac{9}{2} EP(-A - B + C + D) \cos(l_{\odot} - \Omega + \Omega_{\odot}) \\
& + \frac{9}{2} EP(-A - B + C + D) \cos(l_{\odot} + \Omega - \Omega_{\odot}) + \frac{9}{2} EZ(A - C) \cos(l_{\odot} - 2\omega - \Omega + \Omega_{\odot}) \\
& + \frac{9}{2} EZ(A - C) \cos(l_{\odot} + 2\omega + \Omega - \Omega_{\odot}) - \frac{9}{2} AEZ \cos(2\omega + l_{\odot} + 2\omega_{\odot} + \Omega - \Omega_{\odot}) \\
& \left. - \frac{9}{2} AEZ \cos(2\omega + 3l_{\odot} + 2\omega_{\odot} + \Omega - \Omega_{\odot}) + \frac{9}{2} BCP \cos(l_{\odot} + 2\omega_{\odot} - 2\Omega + 2\Omega_{\odot}) \right.
\end{aligned}
$$

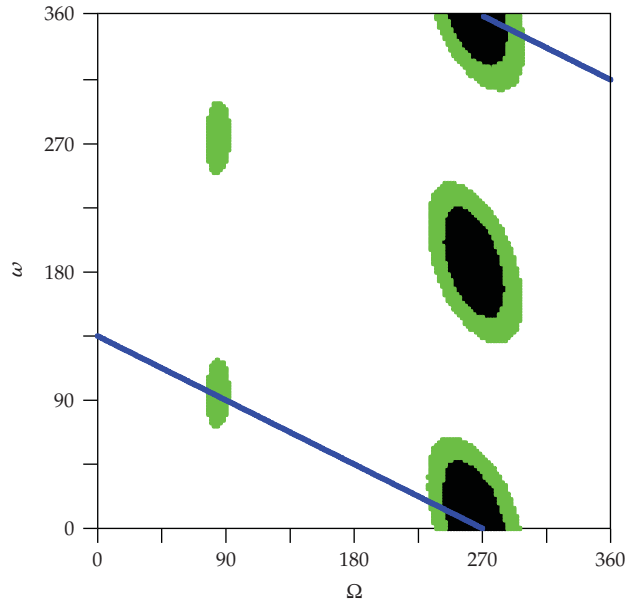


Figure 16: Black dots represent (ω, Ω) values such that a satellite with $a = 26{,}559.74$ km reaches $e \geq 0.6$ in 250 years. Green dots: the same, but eccentricity reaches $e \geq 0.5$ in 250 years. Blue dots: curve satisfying $2\omega + \Omega = 270$. Moon's inclination: $I_L = 18.28°$.

$$
\begin{aligned}
&+\frac{9}{2}BCP\cos(3l_\odot + 2\omega_\odot - 2\Omega + 2\Omega_\odot) + \frac{9}{2}BDZ\cos(l_\odot - 2\omega + 2\Omega - 2\Omega_\odot)\\
&+\frac{9}{2}BDZ\cos(l_\odot + 2\omega - 2\Omega + 2\Omega_\odot) + \frac{9}{2}EP(B-C)\cos(l_\odot + 2\omega_\odot - \Omega + \Omega_\odot)\\
&+\frac{9}{2}EP(B-C)\cos(3l_\odot + 2\omega_\odot - \Omega + \Omega_\odot) + \frac{9}{2}EZ(B-D)\cos(l_\odot - 2\omega + \Omega - \Omega_\odot)\\
&+\frac{9}{2}EZ(B-D)\cos(l_\odot + 2\omega - \Omega + \Omega_\odot) - \frac{9}{2}BEZ\cos(l_\odot + 2\omega + 2\omega_\odot - \Omega + \Omega_\odot)\\
&-\frac{9}{2}BEZ\cos(3l_\odot + 2\omega + 2\omega_\odot - \Omega + \Omega_\odot) + \frac{9}{2}CEZ\cos(2\omega - 3l_\odot - 2\omega_\odot + \Omega - \Omega_\odot)\\
&+\frac{9}{2}CEZ\cos(2\omega - \ l_\odot - 2\omega_\odot + \Omega - \Omega_\odot) + \frac{9}{2}DEZ\cos(2\omega - 3l_\odot - 2\omega_\odot - \Omega + \Omega_\odot)\\
&+\frac{9}{2}DEZ\cos(2\omega - \ l_\odot - 2\omega_\odot - \Omega + \Omega_\odot)\Big].
\end{aligned}
\tag{A.1}
$$

Note that all the cosines in the above relations have $l_\odot$ in the argument. Since the inclination of satellite is fixed about 55° or 56°, each possible resonance can occur for one particular value of the semi-major axis. However, considering the nominal altitude of the GPS and Galileo satellites, and since $e_\odot$ is small, none of the above combinations of angles is significant.

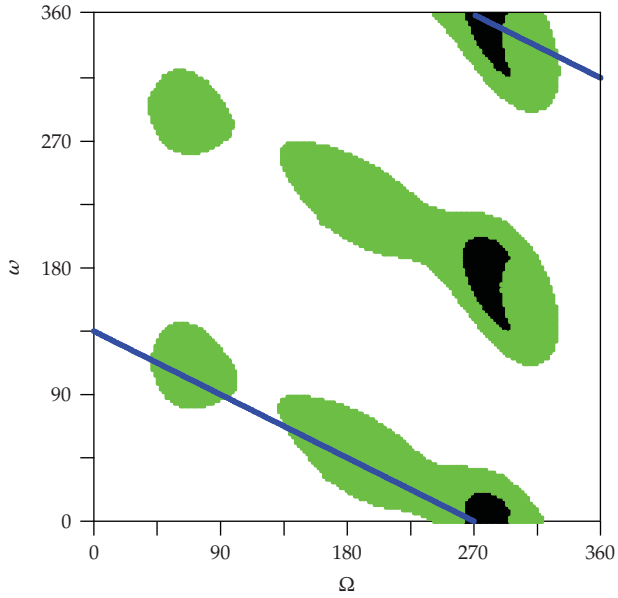


Figure 17: Same of Figure 16, but now $I_L = 28.58°$.

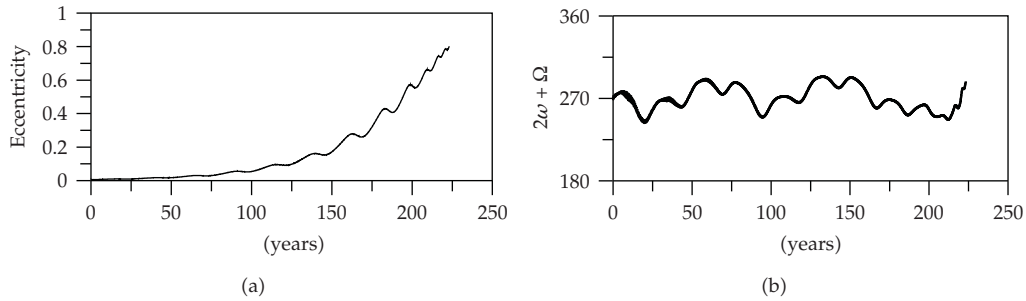


Figure 18: Time evolution of eccentricity (a) and critical angle (b) of a GPS satellite with $a = 26{,}559.74$ km whose eccentricity reaches the value of 0.6 after 200 years. Initial conditions: $e = 0.005$, $I = 56.06°$, $\omega = 0°$, $\Omega = 270°$, $I_L = 18.28°$.

B. The Eccentricity Increasing Strategy

Following the same strategy to obtain Figures 10, 11, 14, and 15, we obtain similar figures, but now the pair (ω, Ω) represents the initial condition of an orbit whose eccentricity attains $e \geq 0.5$ in less than 250 years. As before, all orbits start with initial $a = 26{,}559.74$ km, $e = 0.005$, and $I = 56.06°$. The maximum eccentricity usually is reached after $t = 200$ years.

Acknowledgments

The authors thank CNPQ, FAPESP, and FUNDUNESP. An anonymous referee is gratefully thanked for very useful comments.

References

[1] ESA, "Galileo: Mission High Level Definition," 2002, http://ec.europa.eu/dgs/energy_transport/galileo/doc/galileo_hld_v3_23_09_02.pdf.

[2] NASA, "Guidelines and assessment procedures for limiting orbital debris," NASA Safety Standard 1740.14, Office of Safety and Mission Assurance, Washington, DC, USA, 1995.

[3] A. B. Jenkin and R. A. Gick, "Collision risk posed to the global positioning system by disposed upper stages," *Journal of Spacecraft and Rockets*, vol. 43, no. 6, pp. 1412–1418, 2006.

[4] A. B. Jenkin and R. A. Gick, "Dilution of disposal orbit collision risk for the medium earth orbit constellations," *The Astronomical Journal*, vol. 126, pp. 398–429, 2005.

[5] D. Brouwer and G. M. Clemence, *Methods of Celestial Mechanics*, Academic Press, New York, NY, USA, 1961.

[6] G. Hori, "Theory of general perturbation with unspecified canonical variable," *Publications of the Astronomical Society of Japan*, vol. 18, pp. 287–296, 1966.

[7] H. Kinoshita and H. Nakai, "Secular perturbations of fictitious satellites of uranus," *Celestial Mechanics and Dynamical Astronomy*, vol. 52, no. 3, pp. 293–303, 1991.

[8] T. Yokoyama, "Dynamics of some fictitious satellites of Venus and Mars," *Planetary and Space Science*, vol. 47, no. 5, pp. 619–627, 1999.

[9] Y. Kozai, "The motion of a close earth satellite," *The Astronomical Journal*, vol. 64, pp. 367–377, 1959.

[10] T. Yokoyama, "Possible effects of secular resonances in Phobos and Triton," *Planetary and Space Science*, vol. 50, no. 1, pp. 63–77, 2002.

[11] C. Lánczos, *The Variational Principles of Mechanics*, Mathematical Expositions, no. 4, University of Toronto Press, Toronto, Canada, 4th edition, 1970.

[12] B. D. Tapley, M. M. Watkins, J. C. Ries, et al., "The joint gravity model 3," *Journal of Geophysical Research*, vol. 101, no. 12, pp. 28029–28049, 1996.

[13] F. G. Lemoine, et al., "The development of the joint NASA GSFC and NIMA geopotential model EGM96," Tech. Rep. NASA/TP-1998-206891, NASA/GSFC, 1998.

[14] P. R. Escobal, *Methods of Orbit Determination*, John Wiley & Sons, New York, NY, USA, 1965.

[15] O. Montenbruck and E. Gill, *Satellite Orbits: Models, Methods and Applications*, Springer, Berlin, Germany, 2005.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 794578, 16 pages
doi:10.1155/2009/794578

Research Article

Quality of TEC Estimated with Mod_Ion Using GPS and GLONASS Data

Paulo de Oliveira Camargo

Department of Cartography, Faculty of Science and Technology, São Paulo State University (UNESP), Rua Roberto Simonsen 305, 19060-900 Presidente Prudente, SP, Brazil

Correspondence should be addressed to Paulo de Oliveira Camargo, paulo@fct.unesp.br

Received 1 July 2009; Accepted 28 September 2009

Recommended by Silvia Maria Giuliatti Winter

One of the largest sources of error in positioning and navigation with GNSS is the ionosphere, and the associated error is directly proportional to the TEC and inversely proportional to the square of the signal frequency that propagates through the ionosphere. The equatorial region, especially in Brazil, is where the highest spatial and temporal value variations of the TEC are seen, and where these various features of the ionosphere, such as the equatorial anomaly and scintillation, can be found. Thus, the development and assessments of ionospheric models are important. In this paper, the quality of the TEC was evaluated, as well as the systematic error in the L1 carrier and the inter-frequency biases of satellites and receivers estimated with the Mod_Ion, observable from GPS and integration with the GLONASS, collected with dual frequency receivers.

1. Introduction

The Global Navigation Satellite System (GNSS) is one of the most advanced technologies and has revolutionized the activities related to navigation and positioning from space technology. A main component of GNSS is the Global Positioning System (GPS), developed by the United States, and Global'naya Navigatsionnaya Sputnikowaya Sistema (GLONASS) of responsibility of the Republic of Russia. A relevant fact is that, in December 2005, the first GALILEO satellite was launched, which is being developed by the European Union and must come into operation in 2013. The GNSS is composed by the so-called (Satellite Based Augmentation System) SBAS, such as the Wide Area Augmentation System (WAAS) in USA, European Geostationary Navigation Overlay Service (EGNOS) in Europe, Multi-functional Satellite-Based Augmentation Service (MSAS) in Japan, and Satellite Navigation Augmentation System (SNAS) in China [1].

In general, most users simply use the GNSS system to get their coordinates, without being committed to details, but for the quality of information (precision) provided by it, except aviation. This quality information is very optimistic, demanding attention from users. However, certain applications require the knowledge of the various processes related to the system. The mitigation of the effects of the atmosphere (troposphere and ionosphere) over GNSS observables requires knowledge of the signal analysis and its behavior in the atmosphere, requiring interaction with other sciences such as Aeronomy, Meteorology, among others. Accordingly, this interaction allows studies related to the behavior of the ionosphere and the troposphere to be made from GNSS observables. In Brazil, the ionosphere shows a very complex behavior, for being located near the geomagnetic equator, requiring the development of models and appropriate studies for the region [1–12].

With the Selective Availability (SA) deactivation, in the case of GPS, the error due to the ionosphere has become a major source of systematic error in positioning, especially in periods of high solar activity for one frequency GNSS users, in the conventional point positioning as well as relative positioning. Another effect that affects considerably the GNSS signals is ionosphere scintillation, a result of propagating the signal through a region in which there are irregularities in the density of electrons.

The error due to the ionosphere depends on the Total Electron Content (TEC) present in the ionosphere and in signal frequency. Users of, at least, dual frequency receivers can make corrections of this effect, using the ionospheric free linear combination. This observable eliminates first-order ionospheric effect. Users of single frequency receivers, however, need to correct the systematic effect observables due to the ionosphere. The quantification of this effect can be done by [3, 4, 12]: coefficients transmitted by navigation messages, using the Klobuchar model; observations collected with one or dual frequency GNSS receivers of (Ionosphere map Exchange format) IONEX archives obtained from Global Ionospheric Maps (GIM), which provide values of vertical TEC (VTEC) in a grid with spatial resolution of $5^\circ \times 2.5^\circ$ in longitude and latitude, respectively, and temporal resolution of 2 hours [13].

In the geodetic community, one of the models used and implemented in commercial software, to minimize the effects of the ionosphere on GPS observables, is the Klobuchar model. This model, also called Broadcast model, estimates the systematic error due to the ionosphere to one frequency receivers [14] and its coefficients are transmitted by GPS satellites in navigation messages. However, this model removes around 50% to 60% of the total effect [15, 16]. Being more appropriate for use in regions of middle latitudes, which is the more predictable ionospheric region, where the ionosphere has a more regular behavior. However, this is not an appropriate model to be used in Brazil, where there is high variation in the density of electrons as well as in South America. So with the need to have a more effective correction strategy of the ionosphere effect, several models were developed by various research centers and universities, using observations collected with dual frequency GPS receivers. In terms of South America, we can quote the (La Plata Ionospheric Model) LPIM model, developed at (Astronomical and Geophysical Sciences Faculty of Universidad Nacional de La Plata) FCAG/UNLP, Argentine [2] and the Regional Model of Ionosphere (Mod_Ion) developed in FCT/UNESP, Brazil [3].

The accuracy of VTEC values in the final IONEX files grid (~11 days of latency) is 2–8 TEC units (TECUs) and for rapid files (<24 hours) of 2–9 TECU [17]. Ciraolo et al. [18], in a calibration process, determined the interfrequency bias (IFB) of a pair of receivers, which ranged from 1.4 to 8.8 TECU. This paper aims to assess the quality of TEC and the error in the L_1 carrier estimated with the Mod_Ion, from GPS observables and integration with the ones from GLONASS, collected with dual frequency receivers.

The paper is organized as follows. In Section 2 there is a brief description of impact of the ionosphere on the propagation of GNSS signals; Section 3 describes the equations, based on the geometry-free linear combination of observables collected with dual frequency receivers, used in the Mod_Ion, as well as gets the TEC and the systematic error due to the ionosphere in the L_1 carrier, and some aspects of adjustment by the least squares; the results and analysis of the experiments in order to verify the quality of the TEC provided by model, as well as the IFB of the satellites and receivers are presented in Section 4; based on the experiments, conclusions and future works will be presented in Section 5.

2. Impact of the Ionosphere on the Propagation of GNSS Signals

The terrestrial atmosphere, for practical purposes, can be considered as a set of gas layers, spherical and concentric to the Earth. Its structure is related to various thermal, chemical, and electromagnetic elements. These combined parameters vary depending on the time, latitude, longitude, time of year, and solar activity.

With respect to the propagation of electromagnetic waves, the Earth's atmosphere is divided into ionosphere and troposphere. In this division, the troposphere is the layer between the Earth's surface up to 50 km in height. It is composed of neutral particles, and the highest concentration of gas is found on up to a height of 12 km, consisting of nitrogen, oxygen, carbon dioxide, argon, water vapor, among others. The propagation of the signal in the troposphere depends mainly on the water vapor content, air pressures, and temperature. For frequencies below 30 GHz, the refraction does not depend on the frequency of the signal transmitted [16].

The ionosphere is defined as the portion of the upper atmosphere, where there is sufficient ionization to affect the propagation of radio waves [19]. Unlike the troposphere, it is a dispersive medium; that is, in this case, signal propagation depends on the frequency. It is characterized mainly by the formation of ions and electrons, and it starts at around 50 km, extending to approximately 1000 km in height.

In the region covered by the ionosphere, the electron density is sufficient to alter the propagation of electromagnetic waves. The ions and free electrons in the ionosphere are mainly created by the process of photo ionization. The ionospheric photo ionization is the absorption of solar radiation, predominantly in the range of extreme ultraviolet and X-rays by neutral atmospheric elements [19–21]. The ionosphere as a dispersive means affects the modulation and phase of the carrier, causing, respectively, a delay and an advance [16]. The delay is also referred to as ionospheric delay and increases the apparent length of the path traveled by the signal.

The troposphere effects on GNSS signals are usually reduced by processing techniques or determined directly by models. Since it is not possible to assess the atmospheric pressure and temperature along the route of the signal through the neutral layer, there are several models available, which correct for 92% to 95% of this effect [22]. In contrast, the ionosphere effect, which depends on frequency and, hence, on the refractive index proportional to the TEC, that is, to the number of electrons present along the path between the satellite and the receiver. If the TEC values were constant, the effects caused by the ionosphere would be easy to determine. The problem is that the TEC varies in time and space, in relation daytime, season, solar cycle, geographical location of the receiver and Earth's magnetic field, and so forth. Besides the refraction effect, these variations can cause the receiver to go out of tune with the satellite, by weakening the signal strength, the specific case of the phenomenon known as scintillation.

Table 1: Maximum vertical ionospheric range error (m).

Frequency	1st-order effect $(1/f^2)$	2nd-order effect $(1/f^3)$	3rd-order effect $(1/f^4)$
L_1	32.5	0.036	0.002
L_2	53.5	0.076	0.007
L_0	0.0	0.026	0.006

The ionosphere effects are divided into effects of 1st, 2nd, and 3rd order. Table 1 shows the maximum error in the vertical direction, which can be expected for the GPS L_1, L_2 carriers and for the ionospheric free linear combination (L_0). For inclined directions, the influence increases [1].

The error or effect of 1st order, due to the ionosphere in phase (I^s_{fr}) and pseudorange (I^s_{gr}) along the satellite direction (s) and receiver antenna (r), is given according to the TEC and the frequency of the signal (f) [3, 16]:

$$I^s_{fr} = -\frac{40.3}{f^2}\text{TEC}, \tag{2.1}$$

$$I^s_{gr} = \frac{40.3}{f^2}\text{TEC}. \tag{2.2}$$

According to (2.1) and (2.2), we can see that the errors due to the ionosphere for the phase and pseudorange have the same magnitude but opposite signs. Both are proportional to the TEC and inversely proportional to the square of the frequency of the carrier. The TEC unit (TECU) is given in electrons per square meter (el/m^2) and the constant $40.3\,\text{m}\,\text{Hz}^2(\text{el}/\text{m}^2)^{-1}$. The effect of first order can be obtained from the free geometry linear combination using observables collected with GPS receivers and/or dual frequency GPS/GLONASS, and the remaining error represents a few centimeters [1].

The effect of second order of the ionosphere depends on, besides the TEC and the frequency, geomagnetic induction at the point where the signal passes through the layer of the ionosphere and the angle of the signal in the geomagnetic induction vector. Unlike the effect of first order which is the same and has opposite signals to the phase and pseudorange, the one of the second order of the phase is half of the second-order effect of the group [23].

But the effect of third order does not depend on the magnetic field, but is a function of maximum density of electrons, at the phase the effect is equivalent to one third of the pseudorange effect [23].

2.1. Regular Variations of the TEC

The regular temporal changes of the TEC include daytime and seasonal variations and cycles of long periods. The daytime variation is mainly due to Sunlight, that is, solar radiation. Throughout the day, the density of electrons depends on the local time, with its peak occurring between 12 and 16 local times [24]. In the low latitude equatorial region, a second peak occurs in the hours preceding midnight, especially in periods close to the equinoxes and to the summer and during periods of high solar activity.

Seasons also influence variation in electron density, due to the change in the zenithal angle of the sun and the intensity of the ionization flow, characterizing seasonal variations. During the equinoxes, the effects of the ionosphere are bigger, whereas in the solstices, they are smaller [5].

Changes in long-period cycles, with cycles of approximately 11 years, are associated with the occurrence of sunspots and the increase of ionization and thus the TEC is proportional to the number of spots.

The geographic location also influences the variation of the density of electrons in the ionosphere, because the overall structure of the ionosphere is not homogeneous. It changes with latitude, due to the variation of the zenithal angle of the Sun, which influences directly, the level of radiation, which changes, in turn, the density of electrons in the ionosphere. The equatorial regions are characterized by a high density of electrons and have a high spatial variation. The regions of middle latitudes, however, are considered relatively free from ionospheric abnormalities, presenting a more regular behavior, close to that described by theoretical models. The ionosphere over the north and south poles, alternatively, known as polar or high latitudes ionosphere, is extremely unstable [20]. More details on the changes of regular TEC can be obtained, for example, from [19, 20].

3. Regional Ionosphere Model (Mod_Ion)

The Mod_Ion was developed in FCT/UNESP to represent the ionosphere in an analytical way [3]. The parameters of the model are estimated from data collected with dual frequency GNSS receivers. With the introduction of several receivers it was possible to also estimate the systematic error due to satellites and receivers, called Differential Code Bias (DCB) or IFB, caused by the signal route on the hardware of satellites, until it was spread out on space, and on antenna cables and hardware of receivers, until the signal decorrelation.

The adjustment by the Least Squares Method (LSM) with constraints is used in bath in the process of estimating the parameters of the model. The GNSS observable used in the calculation of the TEC or the systematic error due to the ionosphere in the L_1 carrier is the pseudorange filtered by the carrier phase [25]. The original observable can also be used as well as the carrier phase.

3.1. Ionospheric Model

Models that use GNSS data are based on the geometry-free linear combination of observables collected with dual frequency receivers. In the derivation of the model, errors due to nonsynchronism of the satellite and receiver, ephemeredes and the tropospheric refraction are neglected, since their effects contaminate both frequencies the same way and do not affect the validity of results.

The model is based on the difference between the pseudoranges of the carriers L_2 and L_1, with frequencies f_2 and f_1 of signals generated by the satellites that are part of the GNSS [3]:

$$P_{2r}^{s} - P_{1r}^{s} = I_{2r}^{s} - I_{1r}^{s} + \left(S_{p2}^{s} - S_{p1}^{s}\right) + \left(R_{p2} - R_{p1}\right) + \varepsilon_{p21}. \tag{3.1}$$

From (2.2) we have

$$I_{2r}^{s} - I_{1r}^{s} = 40.3\,\mathrm{TEC}^{s}\frac{f_1^2 - f_2^2}{f_1^2 f_2^2} = I_{1r}^{s}\frac{f_1^2 - f_2^2}{f_2^2}, \tag{3.2}$$

thus

$$F^{\mathrm{TEC}}\left(P_{2r}^{s} - P_{1r}^{s}\right) = \mathrm{TEC}_{r}^{s} + F^{\mathrm{TEC}}\left[\left(S_{p2}^{s} - S_{p1}^{s}\right) + \left(R_{p2} - R_{p1}\right)\right] + F^{\mathrm{TEC}}\varepsilon_{p21} \tag{3.3}$$

or

$$F^{I_1}\left(P_{2r}^{s} - P_{1r}^{s}\right) = I_{1r}^{s} + F^{I_1}\left[\left(S_{p2}^{s} - S_{p1}^{s}\right) + \left(R_{p2} - R_{p1}\right)\right] + F^{I_1}\varepsilon_{p21}. \tag{3.4}$$

Equation (3.3) is the observation equation of Mod_Ion used to calculate the TEC in the satellite/receiver direction. The unknowns $(S_{p2}^{s} - S_{p1}^{s})$ and $(R_{p2} - R_{p1})$ represent, respectively, the IFBs of satellites and receivers, and ε_{P21} represents another differential remaining errors (multipath, receiver noise, etc.), where $F^{\mathrm{TEC}} = f_1^2 f_2^2/40.3(f_1^2 - f_2^2)$, in general representing a constant for the GPS satellites and particularly for each of the GLONASS satellites.

By (3.4), one can calculate the ionospheric delay, that is, ionospheric error (I_{1r}^{s}) in the L_1 carrier, in the satellite/receiver direction, with $F^{I_1} = f_2^2/(f_1^2 - f_2^2)$.

The TEC or the ionospheric delay along the path of the satellite/receiver can be obtained according to the VTEC or the vertical ionospheric delay (I_1^{v}), by the expression, assigned as standard geometric mapping function $(1/\cos z')$, which provides the slant factor, like this

$$\mathrm{TEC}_{r}^{s} = \frac{\mathrm{VTEC}}{\cos z_{r}^{\prime s}} \tag{3.5}$$

or

$$I_{1r}^{s} = \frac{I_1^{v}}{\cos z_{r}^{\prime\prime s}} \tag{3.6}$$

for a receiver (r), z'^{s} is the zenithal angle of the signal path from the satellite (s) to a ionospheric point located in a ionospheric layer, for example, of 400 km of height. Then

$$F^{\mathrm{TEC}}\left(P_{2r}^{s} - P_{1r}^{s}\right) = \frac{\mathrm{VTEC}}{\cos z_{r}^{\prime s}} + F^{\mathrm{TEC}}\left[\left(S_{p2}^{s} - S_{p1}^{s}\right) + \left(R_{p2} - R_{p1}\right)\right] + F^{\mathrm{TEC}}\varepsilon_{p21} \tag{3.7}$$

or

$$F^{I_1}\left(P_{2r}^{s} - P_{1r}^{s}\right) = \frac{I_1^{v}}{\cos z_{r}^{\prime s}} + F^{I_1}\left[\left(S_{p2}^{s} - S_{p1}^{s}\right) + \left(R_{p2} - R_{p1}\right)\right] + F^{I_1}\varepsilon_{p21}. \tag{3.8}$$

Due to the periodic nature of the effect, to model the diurnal behavior of the VTEC or the error in the L_1 carrier [10] use the series

$$\begin{aligned}\text{VTEC or } I_1^{\text{V}} = a_1 + a_2 \text{B}^s + \sum_{\substack{i=1 \\ j=2i+1}}^{n=4} \left\{ a_j \cos(i\text{B}^s) + a_{j+1} \sin(i\text{B}^s) \right\} \\ + a_{n^*2+3} h^2 + \sum_{\substack{i=1 \\ j=2i+10}}^{m=4} \left\{ a_j \cos(ih^s) + a_{j+1} \sin(ih^s) \right\}.\end{aligned} \tag{3.9}$$

The variable B^s represents geographic latitude of the subionospheric point (projection of a point on ionospheric layer on the earth surface) and variable h^s is given as

$$h^s = \frac{2\pi}{T}\left(t - 14^h\right), \tag{3.10}$$

where T represents the 24-hour period and t the local time of the subionospheric point.

The total number of parameters of the model is given by $4^*4 + 3 + r + s$, where the 4^*4+3 represent the coefficients of the series, **r** is the receivers IFB, a total equal to the number of receivers used in the network, and **s** is the satellites IFB, which is equal to the number of satellites tracked to determine the parameters of the model.

In adjustment by least squares, matrix A shows rank deficiency, equal to two. This implies that satellite or receivers IFBs have to be determined for two of them, one regarding GPS and the other regarding GLONASS. Thus, the constraints were imposed in one of the GPS/GLONASS receivers.

4. Experiments, Results, and Analysis

The experiments were performed at the Laboratory of Space Geodesy of the FCT/UNESP, where 4 dual frequency GPS/GLONASS receivers were connected to a TRM 55971.00 Zephyr GNSS Geodetic Model 2 Antenna, using a splitter with 4 outputs. Data were collected for 15 days in the year 2007 (132 to 137, 153 to 157, and 173 to 177) using 2 Topcon TPS HYPER GGD (H826 and H819), TRIMBLE NTR5, and LEICA GRX1200 GGPRO receivers. Two experiments were conducted, the first using only GPS observables, and the second aiming the integration of GPS and GLONASS systems. The experiments were conducted using data in RINEX format, with observables collected every 15 seconds, with 20 degrees elevation. The precise ephemerides and satellite clocks of the International GNNS Service (IGS) were used. It is worth mentioning that all experiments passed the quality control of adjustment and, according to the Dst geomagnetic index, the observables were collected in condition of weak geomagnetic storm (−30 nT to −50 nT), and that on days in question they did not exceed −25 nT.

Receiver H826 was chosen as a reference for estimation of IFBs, and relatively constrained as zero and weight tending to infinity, since the value of the receivers IFB is unknown. Some GLONASS satellites did not participate in some days, for being under maintenance in the quoted period (http://gge.unb.ca/Resources/GLONASSConstellationPlot.pdf).

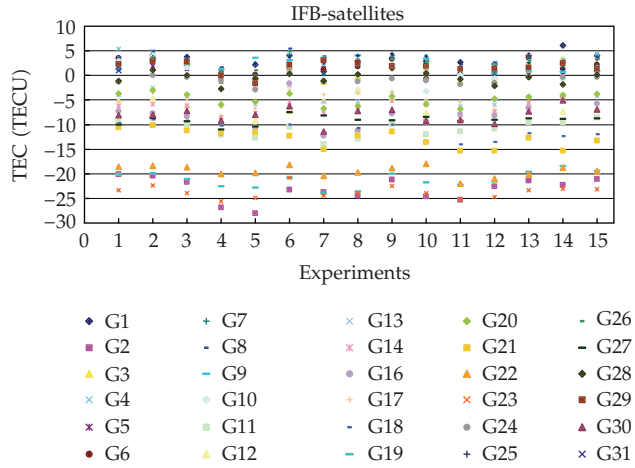


Figure 1: GPS satellite IFB—Error in TEC.

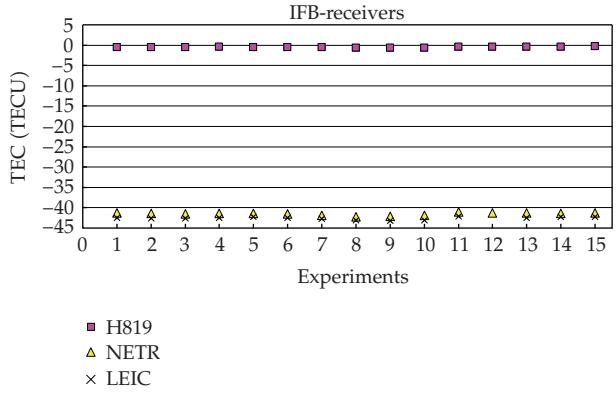


Figure 2: GPS receiver IFB—Error in TEC.

It was yet adopted as a criterion for rejection of the observables, the standard one, which represents the upper limit of change in the TEC for static users, given by 0.1×10^{16} (el/m^2) per second [26]. This value represents 0.085 Hz in L_1 (GPS) and corresponds to 0.0163 m/s of change in pseudoranges due to the ionospheric effect. The differences between consecutive linear combinations bigger than 0.0163 m/s imply the rejection of the observables used to estimate the parameters of the model.

4.1. Satellites and Receivers IFBs Obtained with GPS Observables

Using established procedures, the satellite and receiver IFBs were estimated, as well as the coefficients of the series that allows the calculation of the VTEC considering only the GPS observables. On Figures 1 and 2, the satellites and receivers IFBs are presented, in TECU. Experiments 1 to 5 correspond to the days of the year 133 to 137 of 2007, 6 to 10 correspond to days 153 to 157/2007, and 11 to 15 correspond to days 173 to 177/2007.

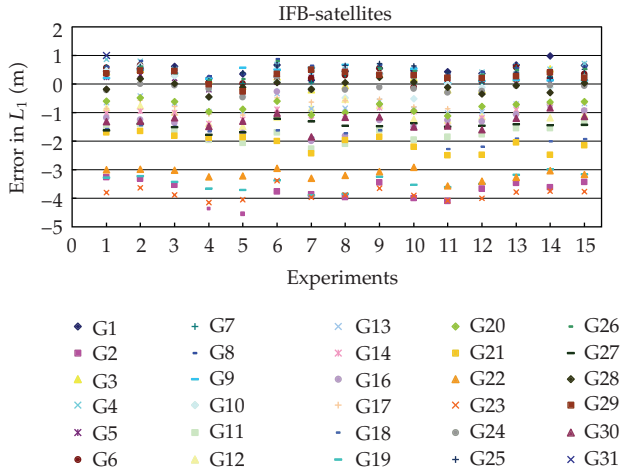


Figure 3: GPS satellite IFB—Error in L_1.

Analyzing Figure 1, we find that the satellites IFBs showed a similar behavior, except for satellite 2 (G2) in the 4th and 5th experiments, and satellite 16 (G16), whose variation was 10.658 TECU. This behavior shows that the systematic errors of the satellites are not stable. It is worth mentioning that the IFBs of satellites include the IFB of receiver H826 which was adopted as reference. The Root Mean Square (RMS) error indicates that you can estimate the IFBs of satellites with precision, better than 2.735 TECU.

The receiver IFB of H826 was constrained as being zero, the values receivers IFBs were estimated in relation to the receiver. On Figure 2, we observed that receivers showed a behavior very similar and stable, but with different values for each. Receiver H819 features an IFB very close to the one adopted as reference, with an average of −0.500 TECU. For receivers NETR and LEIC, the values were, for the trial period, respectively, −41.510 and −42.564 TECU. The variation of the IFB of receivers was around 9 times less than the ones of satellites, indicating the stability of the receivers. The RMS indicates that one can estimate the IFBs of receivers with accuracy better than 0.329 TECU.

Regarding IFBs in L_1 carrier, which represent the systematic error that affects GPS observables made in L_1, values can be obtained using (2.2) or through the Mod_Ion. To determine the error in GPS L_2 carrier can also be used by the same equation or multiply the error in L_1 by the constant 1.64694. Figures 3 and 4 show IFBs due to satellites and receivers for L_1 carrier, in units of m, respectively.

The error in the L_1 carrier due to satellites showed RMS better than 0.444 m and the receivers better than 0.054 m.

4.2. VTEC Obtained with GPS Observables

From geometry-free linear combination (see (3.3)), applying the correction of IFBs due to satellites and receivers can get a set of values of VTEC for each of the 4 receivers. To calculate the differences of VTEC, they took as reference the value obtained from the 19 coefficients estimated for the series (see (3.9)), which analytically represents the ionosphere. Figures 5 to 7 show the discrepancies of VTEC in the quoted period. For each of the experiments it is also presented the values of VTEC determined analytically, and used as reference.

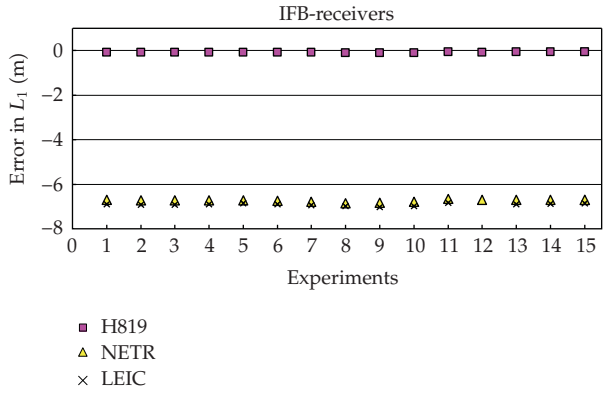


Figure 4: GPS receiver IFB—Error in L_1.

Analyzing Figures 5, 6, and 7, we see that the values of VTEC did not exceed 30 units, because the experiments were conducted over a period of low solar activity. The receivers of the same manufacturer (H826 and H819) have the same behavior for the discrepancies of VTEC, and the daily average is less than −0.227 VTEC units, the RMS indicates that the precision with which the VTEC is estimated is better than 2.365 units, and the biggest discrepancy was −10.917 units. The behavior of receiver NTER is noisier and the daily average and RMS of discrepancies are better than, respectively, −0.244 and 2.396 VTEC units. Regarding receiver LEIC, there is a little higher value than the other, with daily average of discrepancies of up to −0.337 units and RMS of 2.713 VTEC units. The modeling also shows a systematic error, as all the daily average of discrepancies show the same bias, that is, the same signal.

4.3. Satellites and Receivers IFBs Obtained with GPS/GLONASS Observables

In this experiment for the weight of GLONASS observables was assigned a scale factor of 1/2 on the GPS, because fundamental frequency is half the frequency of the GPS. The experiment was conducted only with data collected between the days of the year from 133 to 137 of 2007, and receiver LEIC did not participate because it did not collect GLONASS observables in this period. In Figure 8, it is presented the IFBs due to GPS and GLONASS satellites and, in Figure 9, the ones due to receivers.

On Figure 8, we find that the GPS satellites IFBs had a very similar behavior, contrary to what occurred with the most part of GLONASS satellites. For GPS satellites, in relation to IFBs determined only with GPS observables, the biggest difference to the RMS was 1.675 TECU (G14). The GLONASS satellites showed the biggest variation in the determination of IFBs, reaching 36.191 TECU, with the RMS value of 12.727 TECU (R14).

GPS receiver IFBs (Figure 9) are much more stable than those of the GLONASS receivers. In the previous experiment, the biggest difference of the averages for the GPS did not exceed 0.567 TECU, with RMS of 0.164 TECU. The IBFs related to GLONASS observables have much variation in the order of up to 13.167 TECU, with values very dispersed in relation to the average, with RMS being 3.994 and 4.364 TECU, respectively, for H819 and NETR.

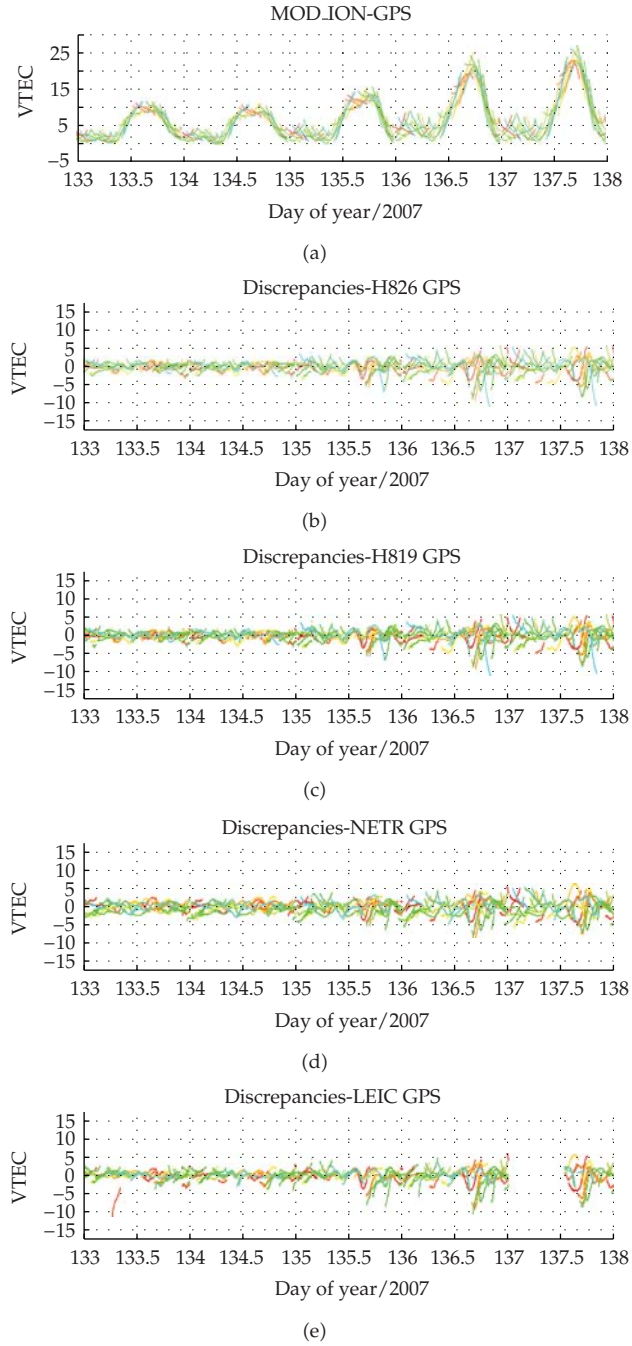


Figure 5: VTEC and discrepancies (GPS: 133 to 137/2007).

4.4. VTEC Obtained with GPS/GLONASS Observables

To evaluate the quality of the TEC obtained with Mod_Ion, a modeling was conducted using the GPS and GLONASS observables simultaneously. Figure 10 shows the daily difference in the quoted period, including the modeled VTEC.

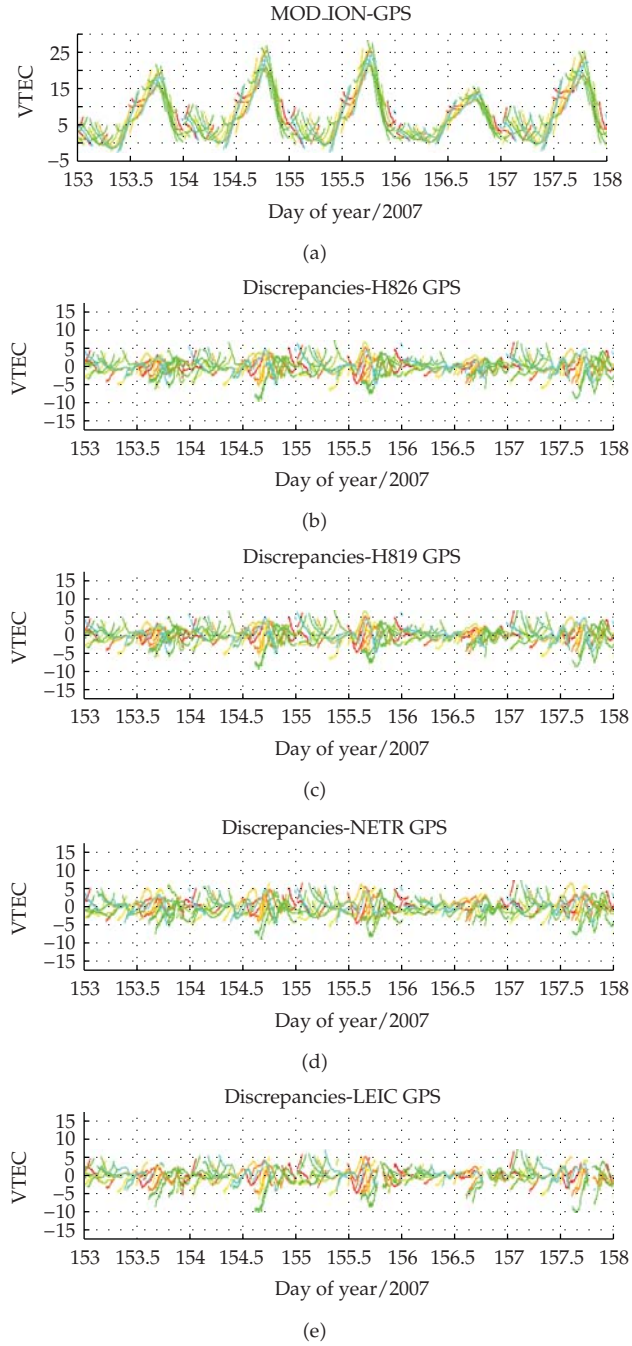


Figure 6: VTEC and discrepancies (GPS: 153 to 157/2007).

The behavior of the VTEC discrepancies (Figure 10) regarding receivers H826, H819, and NETR are similar, the daily average is less than −0.891 VTEC units and RMS is better than 4.929 VTEC. The biggest discrepancy was −30.409 VTEC units, related to the influences of GLONASS satellites.

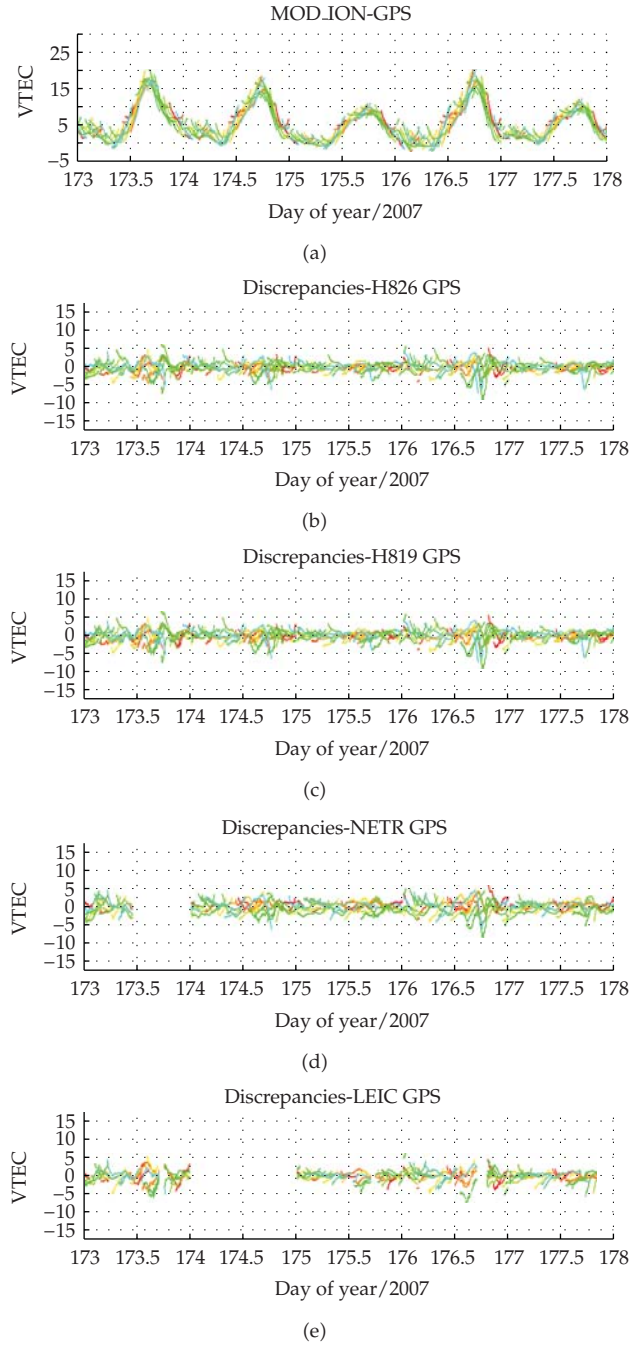


Figure 7: VTEC and discrepancies (GPS: 173 to 177/2007).

5. Conclusions and Future Works

Results showed that, when using only GPS observables, you get the estimation of satellites IFBs with RMS better than 2.735 TECU, and better than 0.329 TECU for those of receivers.

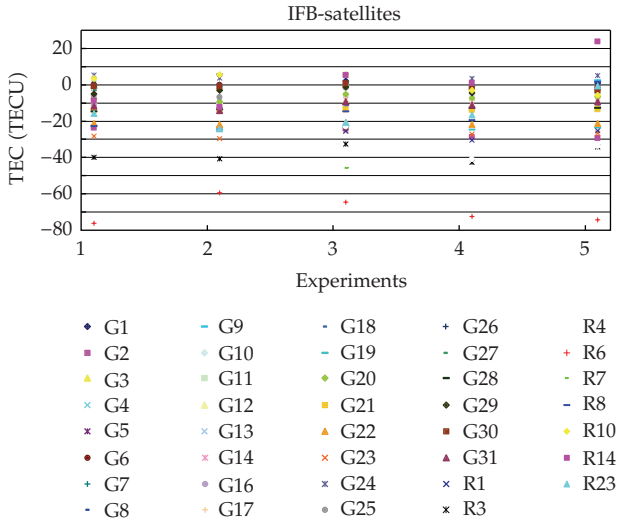


Figure 8: GPS and GLONASS satellite IFB—Error in TEC.

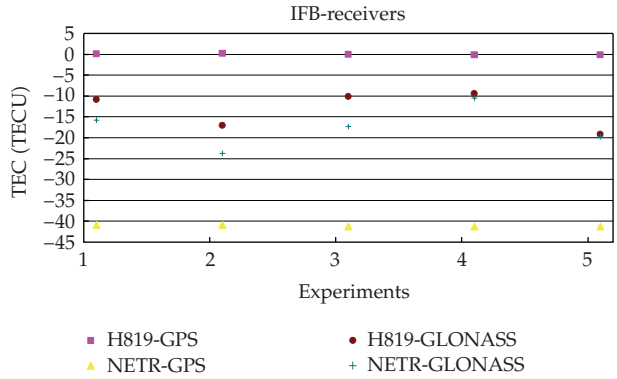


Figure 9: GPS and GLONASS receiver IFB—Error in TEC.

It represents, respectively, in L_1 carrier, errors of 0.444 m and 0.054 m, respectively, for satellites and receivers used. The RMS obtained for the estimation of the VTEC is better than 2.713 units, which for L_1 carrier is an error of 0.441 m. The RMS obtained by integrating GPS and GLONASS was better than, respectively, for the satellites and receivers IFBs and VTEC of 12.727, 4.364 TECU, and 4.929 VTEC units, representing an error of 6.568 m in L_1 carrier, 0.709 m, and 0.800 m respectively. Out of experiments conducted, it is concluded that the GPS observables show better quality than when combined with GLONASS, and compatible with the final values determined with GIM, which is about 2–8 TECU.

Also, new experiments will be conducted using data collected in times of minimum and maximum solar activity, because the solar cycle is the period of minimal activity and periods of irregularities in the ionosphere, having the performance of the model analyzed in the result of absolute and relative positioning with one frequency receivers. In the case of relative positioning, the resolution of ambiguities will also be evaluated.

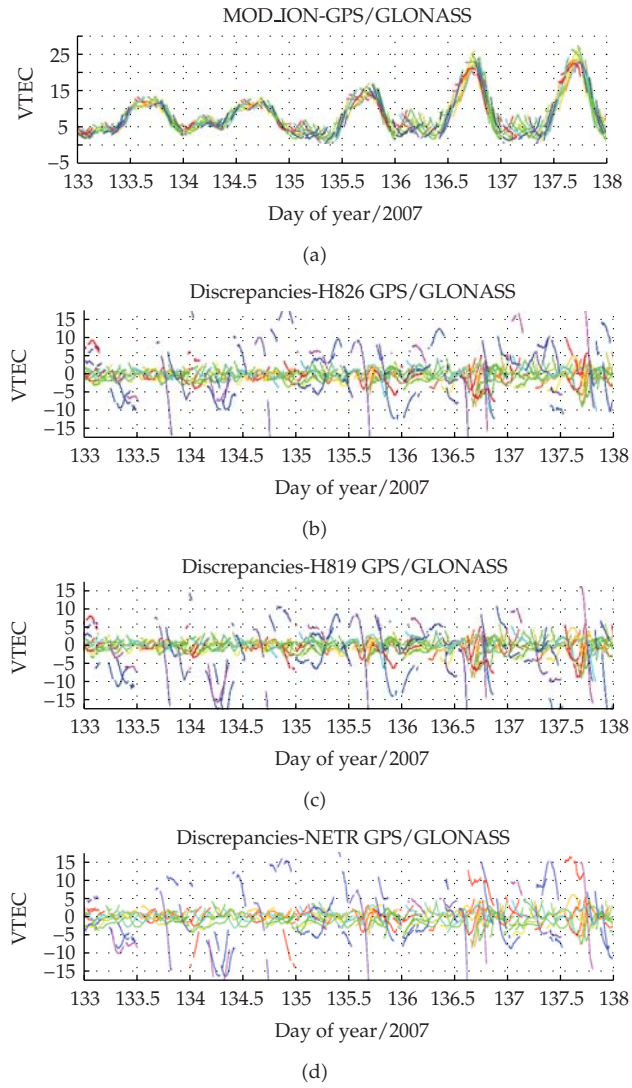


Figure 10: VTEC and discrepancies (GPS/GLONASS: 133 to 137/2007).

About receivers IFBs are necessary to develop methodologies to calibrate and constraint them in the modeling process. It will also be implemented the calculations of effects of 2nd and 3rd orders in the model, in order to provide all the effects of the ionosphere to the users of GPS, GLONASS and, in the future, GALILEO.

And finally, with the modernization and expansion of GNSS networks in Brazil, it is possible to produce maps of the ionosphere in terms of TEC and/or effects on L_1 carrier.

Acknowledgments

The author thanks CNPq for the financial support (Process: 482052/2007-4) and Research Productivity Grant—PQ (Process: 309332/2007-9).

References

[1] G. Seeber, *Satellite Geodesy*, Walter de Gruyter, Berlin, Germany, 2003.

[2] C. A. Brunini, *Global ionospheric model from GPS measurement*, Ph.D. thesis, Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, La Plata, Argentina, 1998.

[3] P. de Oliveira Camargo, *Modelo regional da ionosfera para uso em posicionamento com receptores de uma freqüência*, Ph.D. thesis, Universidade Federal do Paraná, Curitiba, Brazil, 1999.

[4] P. de Oliveira Camargo, J. F. G. Monico, and L. D. D. Ferreira, "Application of ionospheric corrections in the equatorial region for L1 GPS users," *Earth, Planets and Space*, vol. 52, no. 11, pp. 1083–1089, 2000.

[5] E. S. Fonseca Jr., *O sistema GPS como ferramenta para avaliação da refração ionosférica no Brasil*, Ph.D. thesis, Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil, 2002.

[6] M. Fedrizzi, *Estudo do efeito das tempestades magnéticas sobre a ionosfera utilizando dados do GPS*, Ph.D. thesis, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil, 2003.

[7] M. T. Matsuoka and P. de Oliveira Camargo, "Evaluation of functions for modelling of the effect of the ionospheric refraction in the propagation of the GPS signals," in *Proceedings of the 16th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS/GNSS '03)*, pp. 1736–1741, Portland, Ore, USA, 2003.

[8] M. T. Matsuoka, P. de Oliveira Camargo, and W. R. Dal Poz, "Declínio do número de manchas solares do ciclo 23: redução da atividade ionosférica e melhora da performance do posicionamento com GPS," *Boletim de Ciências Geodésicas*, vol. 10, no. 2, pp. 141–157, 2004.

[9] W. R. Dal Poz and P. de Oliveira Camargo, "Conseqüências de uma tempestade geomagnética no posicionamento relativo com receptores GPS de simples freqüência," *Boletim de Ciências Geodésicas*, vol. 12, no. 2, pp. 275–294, 2006.

[10] C. R. de Aguiar and P. de Oliveira Camargo, "Modelagem em tempo real do erro sistemático das observáveis GPS devido à ionosfera," *Boletim de Ciências Geodésicas*, vol. 12, no. 1, pp. 101–120, 2006.

[11] I. S. Batista, M. A. Abdu, J. R. Souza, et al., "Unusual early morning development of the equatorial anomaly in the Brazilian sector during the Halloween magnetic storm," *Journal of Geophysical Research A*, vol. 111, no. 5, Article ID A05307, 2006.

[12] M. T. Matsuoka, *Influência de diferentes condições da ionosfera no posicionamento com GPS: avaliação na região brasileira*, Ph.D. thesis, Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente, Brazil, 2007.

[13] S. Schaer, *Mapping and predicting the earth's ionosphere using the Global Positioning System*, Ph.D. dissertation, University of Bern, Bern, Switzerland, 1999.

[14] J. A. Klobuchar, "Ionospheric time-delay algorithm for single-frequency GPS users," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 23, no. 3, pp. 325–331, 1987.

[15] S. P. Newby and R. B. Langley, "Three alternative empirical ionospheric models—are they better than GPS broadcast model?" in *Proceeding of the 6th International Geodetic Symposium on Satellite Positioning*, vol. 1, pp. 240–244, Columbus, Ohio, USA, 1992.

[16] A. Leick, *GPS Satellite Surveying*, John Wiley & Sons, New York, NY, USA, 1995.

[17] IGS, International GNSS Service, 2009, http://igscb.jpl.nasa.gov/components/prods.html.

[18] L. Ciraolo, F. Azpilicueta, C. Brunini, A. Meza, and S. M. Radicella, "Calibration errors on experimental slant total electron content (TEC) determined with GPS," *Journal of Geodesy*, vol. 81, no. 2, pp. 111–120, 2007.

[19] K. Davies, *Ionospheric Radio*, Peter Peregrinus, London, UK, 1990.

[20] L. F. McNamara, *The Ionosphere: Communications, Surveillance, and Direction Finding*, Krieger, Malabar, Fla, USA, 1991.

[21] V. W. J. H. Kirchhoff, *Introdução à Geofísica Espacial*, EDUSP, São Paulo, Brazil, 1991.

[22] D. Wells, N. Beck, D. Delikaraoglou, et al., *Guide to GPS Positionin*, Canadian GPS Associates, Frederiction, Canada, 1986.

[23] D. Odijk, *Fast Precise GPS Positioning in the Presence of Ionospheric Delays*, Netherlands Geodetic Commission, Delft, The Netherlands, 2002.

[24] I. Webster, *A regional model for prediction of ionospheric delay for single frequency users of the global positioning system*, M.S. thesis, University of New Brunswick, New Brunswick, Canada, 1993.

[25] X. X. Jin, *Theory of carrier adjusted DGPS positioning approach and some experimental results*, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 1996.

[26] J. A. Klobuchar, "Ionospheric effects on GPS," in *Global Positioning System: Theory and Applications*, vol. 1, chapter 12, pp. 485–515, American Institute of Aeronautics and Astronautics, Washington, DC, USA, 1996.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 856138, 9 pages
doi:10.1155/2009/856138

Research Article

The Impact on Geographic Location Accuracy due to Different Satellite Orbit Ephemerides

Claudia C. Celestino,[1] Cristina T. Sousa,[2] Wilson Yamaguti,[1] and Helio Koiti Kuga[3]

[1] *Space Systems Division (DSE), Instituto Nacional de Pesquisas Espaciais (INPE), 12227-010 São José dos Campos, SP, Brazil*
[2] *Mathematics Department (DMA), University of São Paulo State (UNESP-FEG), 12516-410 Guaratinguetá, SP, Brazil*
[3] *Space Mechanics and Control Division (DMC), Instituto Nacional de Pesquisas Espaciais (INPE), 12227-010 São José dos Campos, SP, Brazil*

Correspondence should be addressed to Helio Koiti Kuga, hkk@dem.inpe.br

Received 30 July 2009; Accepted 29 October 2009

Recommended by Tadashi Yokoyama

The current Brazilian System of Environmental Data Collection is composed of several satellites (SCD-1 and 2, CBERS-2 and 2B), Data Collection Platforms (DCPs) spread mostly over the Brazilian territory, and ground reception stations located in Cuiabá and Alcântara. An essential functionality offered to the users is the geographic location of these DCPs. The location is computed by the in-house developed "GEOLOC" program which processes the onboard measured Doppler shifts suffered by the signal transmitted by the DCPs. These data are relayed and stored on ground when the satellite passes over the receiving stations. Another important input data to GEOLOC are the orbit ephemeris of the satellite corresponding to the Doppler data. In this work, the impact on the geographic location accuracy when using orbit ephemeris which can be obtained through several sources is assessed. First, this evaluation is performed by computer simulation of the Doppler data, corresponding to real existing satellite passes. Then real Doppler data are used to assess the performance of the location system. The results indicate that the use of precise ephemeris can improve the performance of GEOLOC by reducing the location errors, and such conclusion can then be extended to similar location systems.

1. Introduction

The Brazilian System of Environmental Data Collection (SBCDA) can be organized in three subsystems: the space subsystem, the data collection ground subsystem, and the tracking and control ground segment [1, 2]. The space subsystem is composed of SCD-1, SCD-2, CBERS-2, and CBERS-2B satellites. The ground subsystem of data collection is composed of

Figure 1: Typical meteorological data collection platform.

hundreds of Data Collection Platforms (DCPs) that are deployed on ground, fixed or mobile (see Figure 1). The INPE's tracking and control ground segment is composed of the Satellite Control Center and Ground Stations in Cuiabá and Alcântara.

In this system, the satellite works as a message retransmitter, that is, providing a communication link between data collection platform (DCP) and a reception ground station. One of the functionalities offered by the system is the Geographic Location of the DCPs. The location is computed by the in-house developed "GEOLOC" program which processes the Doppler shifts suffered by the signal transmitted by the DCPs, together with a statistical least-squares method [3, 4]. The DCPs transmit data signals to the satellites in the UHF frequency band. Aboard the satellite, the DCP is identified and the payload data and the received frequency data are relayed to the tracking reception station. Then, the Geographical Location software GEOLOC developed by INPE is fed with Doppler shift data and the corresponding satellite orbit ephemeris. In general the most common format of orbit ephemeris exchange is the Two Line Elements (TLE) set [5, 6]. The TLEs can be obtained through the Satellite Control Center at INPE or through Internet, for example, [5]. The TLE format is composed of seven parameters and time and is basically defined by orbital mean motion, eccentricity, inclination, right ascension of the ascending node, perigee argument, mean anomaly, and a modified ballistic coefficient [5, 6]. Figure 2 shows a representation of Two Line Elements format.

To evaluate the impacts on geographic localization errors generated by the GEOLOC software, satellite orbit ephemerides from several sources were used such as: accurate PVT (Position, Velocity, Time), TLEs from INPE's Control Center, and TLEs from Internet. Then actual passes of the satellite were used to compare the errors considering such different orbital data sources for the location software. To foresee the magnitude of the expected errors, a comparison with ideal (simulated) case is also performed as detailed in the next section.

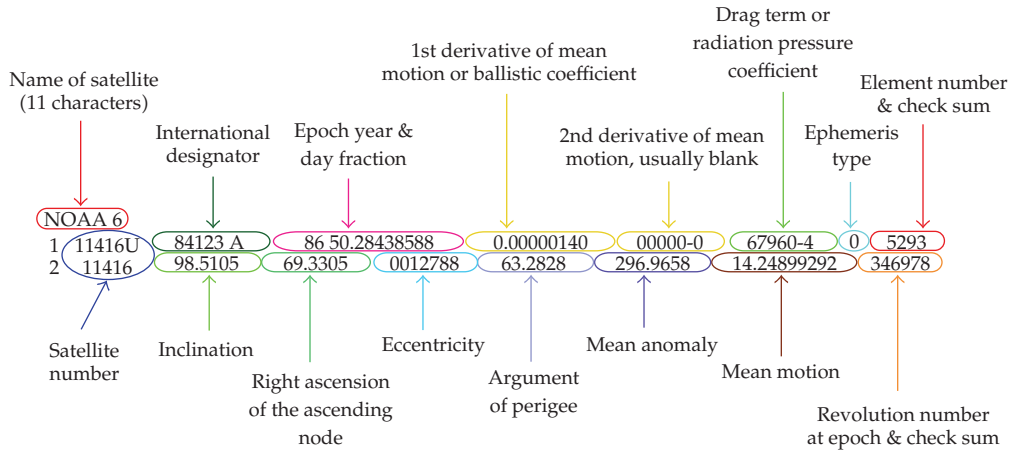


Figure 2: Presentation of the two line elements. Source: [5]

It is expected that conclusions arising from this work can be extended to similar systems using Doppler data and artificial satellites for location purposes.

2. Data and Work Outline

In order to get a benchmark to allow the analysis of location errors due to orbit ephemeris, the work followed the following steps.

(i) A survey on the satellite passes (samples) and the feasibility of obtaining the accurate orbit ephemeris (PVT format) were made to select the test period.

(ii) Three reference DCPs (no. 113, no. 32590, and no. 109) were selected whose locations are: DCP113 (12.0960°S, 77.0400°W), DCP32590 (15.5550°S, 56.0698°W), and DCP109 (5.1860°N, 52.6870°W). Such locations are accurate to the GPS level, that is, 10 to 30 m. The nominal frequency is 401.650 MHz (UHF). The Doppler shift data corresponding to the passes were obtained by the reception stations of Cuiaba and/or Alcântara, consistent with the location of the DCPs. The test period was from November, 21st to 27th, 2008.

(iii) The Doppler shift data of SCD-2 satellite passes were considered. The orbit inclination of this satellite is 25°, that is, a near equatorial orbit. The orbit ephemerides in Two Line Element (TLE) format were obtained from both the Satellite Control Center (CCS) at INPE [7–9] and Internet at [5].

(iv) The accurate orbit ephemerides (PVT format) were provided by the INPE CCS Orbit Determination System [7] with steps of one minute for each considered day. The GEOLOC location software was executed with this accurate ephemerides. Then, the accuracy of the SCD2 satellite orbit ephemerides was verified.

(v) To foresee the magnitude of the error when using TLEs instead of accurate PVTs, another program simulated the errors for the ideal case.

(vi) The GEOLOC location software was run using the PVT, TLEs from CCS, and TLEs from Internet, covering 4 SCD2 satellite actual passes. For each satellite pass, considering the established period above, the reference DCP locations are done using the Doppler shift measurements and the ephemerides obtained from two different sources (TLEs from both CCS and Internet). The location errors are compared with the ones obtained using accurate PVT state vector, considered as reference.

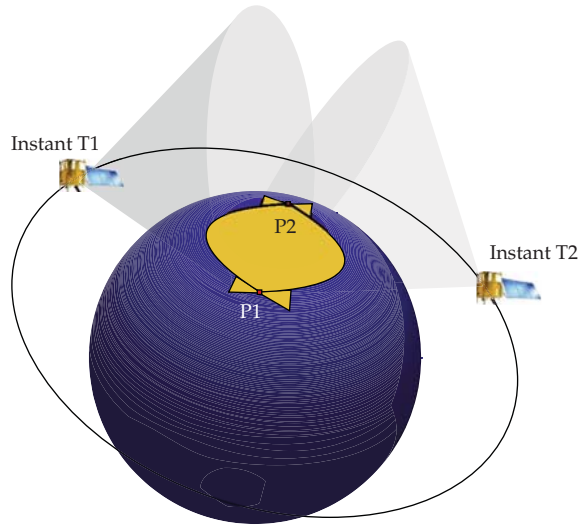


Figure 3: Location cones.

3. The System for Geographic Location of DCP's

During the pass of a satellite, the signals transmitted from the DCPs are immediately relayed, through the satellite, to the reception stations. In the reception station, the signals are received and the Doppler shifted frequencies, computed on board, are collected. The difference between the frequency of the received signal and the nominal frequency supplies the Doppler shift. For each Doppler measure it corresponds to a solid cone of location, whose intersection with the terrestrial sphere represents the possible positions of the transmitter. The intersection of the two cones in the altitude sphere supplies two possible position solutions, in a single pass (Figure 3).

With the apriori knowledge of an approximate position, it is possible to distinguish which is the correct one. However, because the Doppler measurement is corrupted by several error sources, a direct solution is not possible. Therefore, one should use statistical methods to solve the problem.

To do so, the parameters of position location and the ephemeris of the satellite are related through expression [3, 4]**:**

$$h(\mathbf{x}) = \frac{(x - X)(\dot{x} - \dot{X}) + (y - Y)(\dot{y} - \dot{Y}) + (z - Z)(\dot{z} - \dot{Z})}{\sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2}} + b_0 + b_1 \Delta t, \tag{3.1}$$

where $x = (x, y, z)$ and (X, Y, Z) are the position coordinates of the transmitter and the satellite respectively; $(\dot{x}, \dot{y}, \dot{z})$ and $(\dot{X}, \dot{Y}, \dot{Z})$ represent the velocity vectors of the transmitter and the satellite; b_0 and b_1 are the bias and drift associated with each Doppler curve; Δt is time interval from the time of closest approach.

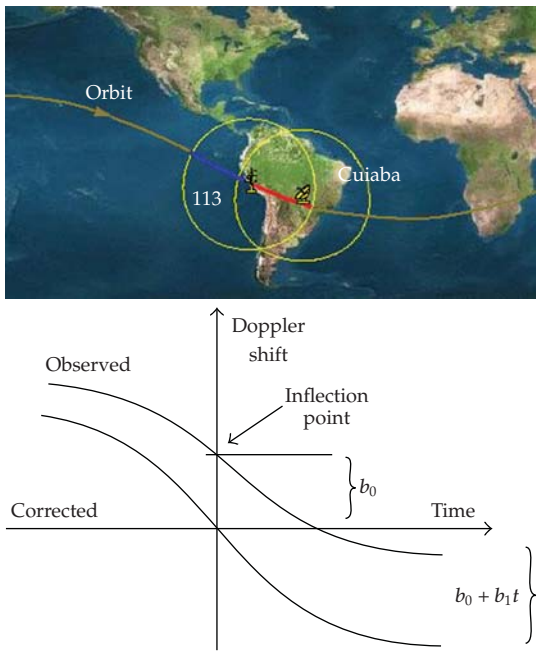


Figure 4: Typical SCD-2 pass and Doppler curve.

Figure 4 shows an example of a typical pass of satellite SCD-2 over DCP 113 and Cuiaba reception station. In this figure, the blue line represents positive Doppler shifts, and the red line the negative ones, with respect to DCP 113. The yellow circles are visibility circles where the contact with satellite is possible. Also a Typical Doppler curve is depicted.

4. TLE-S Accuracy: CCS and Internet

With the aim at studying the impact of ephemeris accuracy of SCD2 satellite on the location system, some tests were made to compare the accuracy of different sources of orbit ephemeris. Four satellite passes on November 22-23, 2008 were used, and these ephemerides were compared with TLEs from CCS and Internet.

The CCS carries out precise orbit determination based on ground tracking data of type "ranging" and Doppler [7]. The orbit determination using "ranging" is called "ranging" solution, and the orbit determination using Doppler data is called Doppler solution. The differences between both solutions agree to the level of 15 m [10]. The orbit ephemerides are provided as Inertial True of Date ("True Of Date") PVTs (Position, Velocity, Time), equally spaced at one-minute intervals. To recover the ephemeris to the desired time, the PVTs from CCS are interpolated using a sixth-degree Lagrange polynomial.

CCS provides also TLEs of SCD2. The TLEs obtained from CCS are generated using an approximate period of one week data. Analyses have shown that the errors, compared to the PVTs, are zero mean with standard deviations around 300 m, however, sometimes presenting peaks of errors of 1000 to 1500 m in along-track (transversal) components [8, 10, 11]. Such accuracies are consistent with the ones presented elsewhere [12].

Table 1: Satellite SCD2 passes considered in the comparisons of the orbit ephemeris differences.

Pass	Date Nov. 2008	Initial time (UTC)	End time (UTC)
1	22	17:27:44	17:35:18
2	22	19:13:49	19:25:56
3	23	16:40:25	16:52:32
4	23	18:27:00	18:40:08

The TLEs obtained by Internet are refreshed with a not well-defined periodicity but, in general, fresh TLEs are available every 2 to 3 days. It is claimed that accuracy [6, 12] enough for tracking purposes is provided. The orbital ephemerides of many satellites are broadcast by NASA in the TLE format in the electronic addresses of Internet, for example, [5]. To recover the ephemerides for the desired time, the model SGP4 [6] is conventionally used.

All the ephemerides are in Inertial True of Date ("True Of Date") system. The inertial coordinates are transformed to the ECEF system ("Earth Centered Earth Fixed") of WGS-84 system, taking into account the polar motion and equation of the equinoxes, which are the most relevant corrections. The SCD2 satellite passes considered in the tests are presented in Table 1.

From the former considerations, the PVTs from CCS are taken as reference to the comparisons. Figures 5 (a) and (b) show the differences between the ephemerides using TLE from Internet (NORAD) and from CCS, respectively, in terms of the radial (R), transversal or along-track (T), and normal (N) components.

Considering the different ephemerides from PVTs and TLEs of CCS, one observes that all R, N, and T component differences are inferior to around 450 m. On the other hand, the differences between the PVTs and Internet TLEs present values similar for the R and N components; however, the transversal component T presents much higher values near 1000 m.

Therefore, for the TLEs from Internet, there is a more pronounced difference in the transversal component T. It is very likely that it can cause a biased error on the geographic location, because of the usage of different source of orbit ephemeris.

5. Test Results

In the tests, SCD-2 satellite and three DCPs in different locations were used. The SCD-2 satellite is in a quasi-circular low orbit around of Earth, with 25° inclination, and altitude of 750 km. The 3 DCPs are DCP109, DCP113, and DCP32590.

5.1. Ideal Case

Considering a fictitious ideal case, the Doppler measurements for DPCs 109 and 32590 were simulated computationally using the more precise PVTs provided by the CCS Satellite Control Center as orbit ephemeris, with SCD2 satellite relayed by the reception station of Cuiabá. On the other hand, TLEs from Internet and from CCS were used as orbit ephemeris in the GEOLOC location software, and the results are presented in Table 2. This test gives an idea of the errors magnitude expected when using ephemeris of different accuracies. For days 21 and 22, of November 2008, the Table shows that these errors are mostly near 1 km, which, as expected, is a rather direct translation of ephemerides differences as shown in Figure 5.

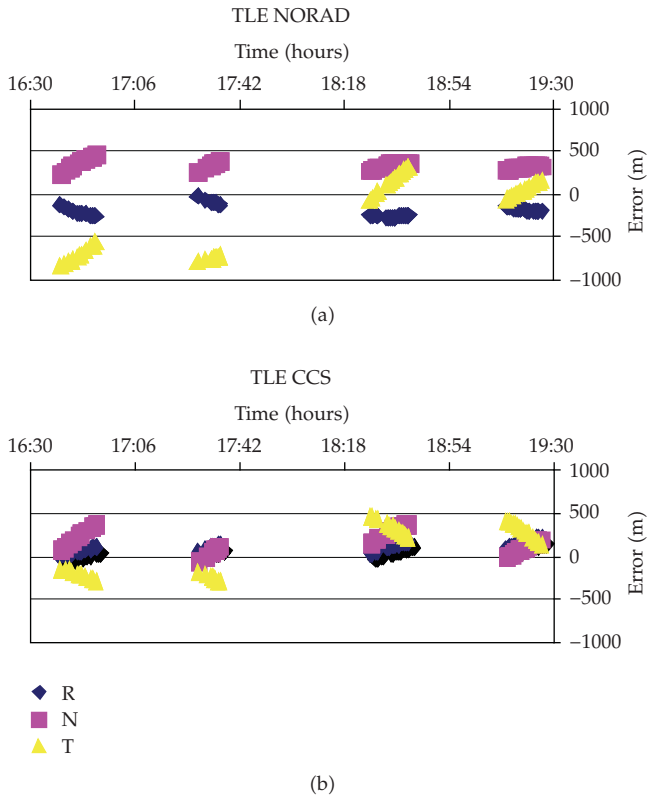


Figure 5: Radial (R), Normal (N), and Transversal (T) deviations between CCS PVTs and TLEs from Internet (a) and from CCS (b).

Table 2: Location errors for a simulated ideal case.

DPC no.	Date Nov. 2008	Time (UTC)	Internet TLE error (km)	CCS TLE error (km)
109	21	16:29:50	1.07	0.59
109	21	18:14:20	0.67	0.42
109	22	08:31:30	0.99	0.28
109	22	10:19:50	0.83	0.20
109	22	17:27:20	0.88	0.32
109	22	19:13:50	0.85	0.29
109	22	21:02:00	0.61	0.66
32590	21	09:16:40	0.59	0.38
32590	21	11:02:50	0.49	0.27
32590	21	18:09:40	0.94	0.23
32590	22	08:29:50	0.64	0.46
32590	22	10:15:40	1.04	0.32
32590	22	15:36:10	1.25	0.51
32590	22	17:22:30	1.28	0.17
32590	22	19:09:40	0.24	0.42

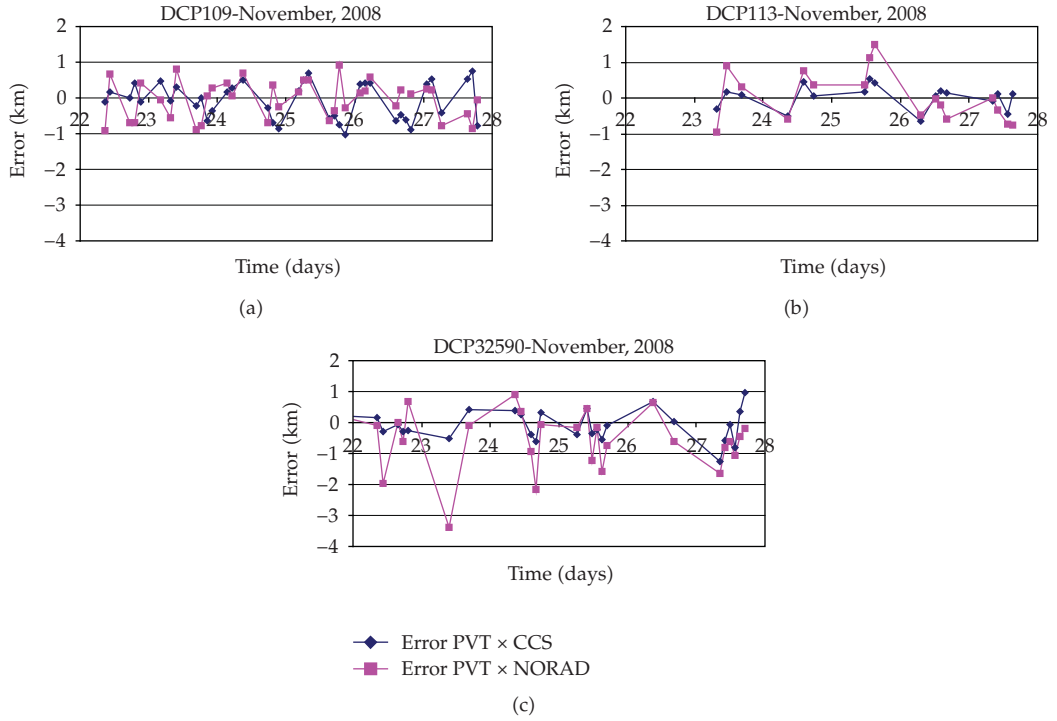


Figure 6: Difference between location errors using the PVT location compared to locations from CCS TLE and NORAD TLE.

5.2. Real Case

In this case, Doppler data relayed by satellite SCD2 and corresponding to days from November 22–28, 2008, were retrieved from the SBCDA archives for the three DCPs 109, 113, and 32590. For this actual test case, Doppler data were processed by the GEOLOC software, and the results are presented in Figure 6. Figure 6 shows the difference in location errors using all three sources of orbit ephemeris: PVTs from CCS, TLEs from CCS, and TLEs from Internet. The figure contains, for all 3 DCPs, the location error differences between CCS and PVT, and TLE and PVT. That is, they show the differences in location when using 3 different sources of orbital ephemeris. Notice that DCP 32590 presents the highest differences when using TLEs from Internet (NORAD), similar to what occurred in the simulated ideal case.

6. Conclusions

This article presents the impact on the accuracy when different sources of orbit ephemeris with heterogeneous accuracies are used to perform location of ground transmitters using the Doppler shifted data recorded on board satellites. Three different sources were analyzed (i) precise PVTs (Position, Velocity, Time) arising from INPE's Control Center orbit determination system, (ii) TLEs computed by INPE's Control Center on weekly basis, and (iii) TLEs obtained freely from Internet. For a restricted period of 2-day comparison, analyses show the differences in accuracy which could arise between the different orbit ephemeris sources. The comparisons showed that the ephemerides have minor and similar differences

in radial and normal components. Nevertheless, extending the comparison to the transverse (along-track) component, the difference ranges from 500 to 1000 meters. Therefore, it is straightforward to conclude that this is the major contributing error to the location system. A simulated location session using ideal measurements, but with orbit ephemeris of different accuracy, showed clearly the along track error being transposed to the location error at similar levels. Actual data confirmed the same behavior. A future work can make use of more precise orbit ephemerides available via on-board GPS, to confirm the location errors steaming from ephemeris with different accuracies. At the end, the results point that usage of precise ephemeris can improve the performance the GEOLOC location system, and such conclusion can then be extended to similar systems.

Acknowledgment

The authors thank CNPq (Grant no. 382746/2005-8) and FAPESP (Grant no. 2005/04497-0) for the financial support.

References

[1] W. Yamaguti, et al., "The Brazilian environmental data collecting system: status, needs, and study of continuity of the Data Collecting Mission," (SCD-ETD-002), Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil, 2006.

[2] W. Yamaguti, V. Orlando, and S. P. Pereira, "Brazilian data collecting system: status and future plans," in *Proceedings of the 14th Remote Sensing Brazilian Symposium (SBSR '09)*, pp. 1633–1640, Natal, Brazil, 2009.

[3] C. T. de Sousa, *Geolocation of transmitters by satellites using Doppler shifts in near real time*, Ph.D. dissertation, Space Engineering and Technology, Space Mechanics and Control Division, INPE—Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil, 2000.

[4] C. T. de Sousa, H. K. Kuga, and A. W. Setzer, "Geo-location of transmitters using real data, Doppler shifts and least squares," *Acta Astronautica*, vol. 52, no. 9–12, pp. 915–922, 2003.

[5] Celestrak, November 2008, http://www.celestrak.com/.

[6] F. R. Hoots and R. L. Roehrich, "Models for propagation of NORAD element sets," Spacetrack Report no. 3, Peterson AFB, Colorado Springs, Colo, USA, December 1980.

[7] H. K. Kuga, "Flight dynamics at INPE," in *Nonlinear Dynamics, Chaos, Control and Their Applications to Engineering Sciences*, J. M. Balthazar, D. T. Mook, and J. M. Rosario, Eds., vol. 1, pp. 306–311, AAM—American Academy of Mechanics, 1997.

[8] R. V. de Moraes, H. K. Kuga, and D. Y. Campos, "Orbital propagation for Brazilian satellites using NORAD models," *Advances in Space Research*, vol. 30, no. 2, pp. 331–335, 2002.

[9] H. K. Kuga, "On using the two-lines ephemerides from NORAD in the CBERS-1 satellite orbit model," in *Proceedings of the Brazilian Conference on Dynamics, Control and Their Applications (DINCON '02)*, vol. 1, pp. 925–930, São José do Rio Preto, Brazil, 2002.

[10] H. K. Kuga, A. R. Silva, and R. V. F. Lopes, "Analysis of accuracy of on-board CBERS-2B satellite GPS orbital ephemerides," in *Proceedings of the 14th Remote Sensing Brazilian Symposium (SBSR '09)*, pp. 2057–2064, Natal, Brazil, 2009.

[11] H. K. Kuga and V. Orlando, "Analysis of on-board orbit ephemeris impact on CBERS-2 image processing," in *Proceedings of the 18th International Symposium on Space Flight Dynamics (ISSFD '04)*, pp. 543–546, Munich, The Netherlands, October 2004, European Space Agency, (Special Publication) ESA SP.

[12] R. Wang, J. Liu, and Q. M. Zhang, "Propagation errors analysis of TLE data," *Advances in Space Research*, vol. 43, no. 7, pp. 1065–1069, 2009.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 147326, 19 pages
doi:10.1155/2009/147326

Research Article

Simulations under Ideal and Nonideal Conditions for Characterization of a Passive Doppler Geographical Location System Using Extension of Data Reception Network

Cristina Tobler de Sousa,[1] Rodolpho Vilhena de Moraes,[1] and Hélio Koiti Kuga[2]

[1] *UNESP-University, Estadual Paulista, Guaratinguetá, CEP 12516-410, SP, Brazil*
[2] *INPE-DMC, CP 515, S.J. Campos, CEP 12245-970, Brazil*

Correspondence should be addressed to Hélio Koiti Kuga, hkk@dem.inpe.br

Received 30 July 2009; Accepted 5 November 2009

Recommended by Maria Zanardi

This work presents a (Data Reception Network) DRN software investigation considering simulated conditions inserting purposely errors into the Doppler measurements, satellites ephemeris, and time stamp, to characterize the geographical location software (GEOLOC) developed by Sousa (2000) and Sousa et al. (2003). The extension of reception stations in Brazilian territory can result in more precise locations if the network is considered in the GEOLOC. The results and analyses were first obtained considering the ground stations separately, to characterize their effects in the geographical location (GL) result. Six conditions were investigated: ideal simulated conditions, random and bias errors in the Doppler measurements, errors in the satellite ephemeris, and errors in the time stamp in order to investigate the DRN importance to get more accurate locations; an analysis was performed considering the random errors of 1 Hz in the Doppler measurements. The results are quite satisfactory and also show good compatibility between the simulator and the GEOLOC using the DRN.

1. Introduction

This work presents the validation, through software simulations, of the passive Doppler GEOLOC system using extended DRN. The Doppler measurements data of a single satellite pass over a (Data Collecting Platform) DCP, considering a network of ground reception stations, is the rule of the DRN. The DRN uses an ordering selection method that merges the collected Doppler shift measurements through the stations network in a single file. The preprocessed and analyzed measurements encompass the DCP signal transmission time and

the Doppler shifted signal frequency received on board the satellite. Thus, the assembly to a single file of the measurements collected, considering a given satellite pass, will contain more information about the full Doppler effect behavior while decreasing the amount of measurement losses, as a consequence of an extended visibility between the relay satellite and the reception stations.

The simulations software was developed to produce simulated Doppler measurements according to several error simulation scenarios. A test scenario composed by SCD-2 (Brazilian Data Collecting satellite) and NOAA-17 (National Oceanic Atmospheric Administration) satellite passes, a single DCP and five ground receiving stations were used. This software is integrated to the developed GEOLOC system that uses the method of near real-time (just after data reception) LOCATION of transmitters through satellites [1].

Nowadays there are more than 600 (fixed and moving) DCPs transmitting several types of payload data (meteorological, hydrological, agricultural, deforestation, CO_2 gas concentration, etc.) through (Low Earth Orbit) LEO satellites. In Brazil, near real-time LOCATION of transmitters and its monitoring through satellites is particularly useful for monitoring moving DCPs [2] (drifting buoys in sea or rivers), to track displacements and habits of animals by fixing minitransmitters on them [3], for checking if the DCP is still in place, or for insuring that goods reaches the destiny, to monitor emergency location and rescue of aircraft and ships [4], and others.

The relay satellite measures the Doppler shift suffered by the DCP transmitted signal, which in turn, together with the payload data, downlinks the Doppler measurements to ground receiving stations. Such Doppler shift measurements are freely available (passive), being further processed to compute the DCP location through the location software.

The results and analysis with table and graphics are represented under six conditions:

(1) ideal conditions from simulated Doppler shift measurements;

(2) random Gaussian errors in the Doppler measurements;

(3) bias errors in the Doppler measurements;

(4) errors in the satellite ephemeris;

(5) errors in the time stamp;

(6) errors of 1 Hz in the Doppler measurements using DRN.

For the Doppler shift measurements simulations we considered two transmission intervals: either 600 Doppler measurements per pass (good statistical condition with lots of measurements, approximately one measurement every second) or 7 Doppler measurements per pass (realistic condition and not so good statistical scenario, approximately one measurement every 90 seconds).

2. Geographical Location with Doppler Shift Measurements

Basically the satellite receives the UHF signals from DCPs and relays such signals to ground reception stations in range. The Doppler shift measurements are computed in the ground station. In the Data Collection Mission Center, the Doppler measurements are sorted and merged to input them to the geographic location software, which provides the DCPs location.

When the transmitter and the reception stations are inside the satellite visibility circle of around 5000 km diameter for 5° minimum elevation angle, the nominal UHF frequency signals periodically sent by the transmitter are received by the satellite and immediately

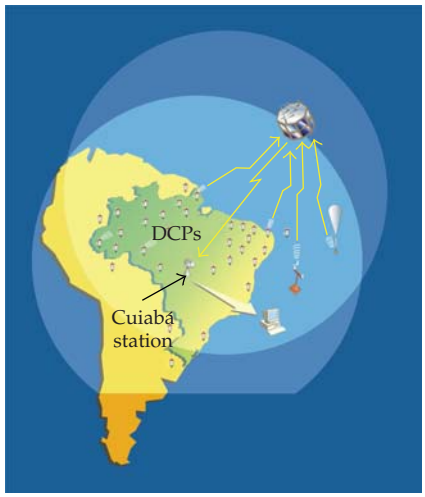


Figure 1: Brazilian Environmental data collection system and Cuiabá and Alcântara station visibility circles.

(realtime) sent down to the reception station. The platforms installed on ground (fixed or mobile) are configured for transmission intervals of between 40 to 220 seconds. In a typical condition, in which both transmitter and receiver are close enough, this period can last up to 10 minutes. The DCP messages retransmited by the satellites and received by the Cuiabá and Alcântara stations are sent to the Data Collection Mission Center located at Cachoeira Paulista, for processing, storage, and dissemination to the users, as seen in Figure 1.

The difference between the received signal frequency and the nominal frequency supplies the Doppler shift. The basic principle of transmitter location considers that for each signal transmitted a location cone is obtained (Figure 2). The satellite is in the cone vertex and its velocity vector v lies in the symmetry axis. Two different cones of location intercept the surface and its intersection contains two possible transmitter positions. To find which of the two ambiguous positions is the correct one, additional information is required, as for example, the knowledge of an initial approximate position. A second satellite overpass removes any uncertainties.

The transmitter geographic location can be determined by means of the Doppler shift of the transmitted frequency due to the relative velocity between the satellite and the transmitter. The satellite velocity relative to the transmitter ($V\cos\alpha$) in vacuum conditions, denoted by $\dot{\rho}$, is given by the Doppler effect [5] equation

$$\dot{\rho} = \left[\frac{(f_r - f_t)}{f_t}\right] c, \tag{2.1}$$

where f_r is the frequency value as received by the satellite; f_t is the reference frequency sent by the transmitter; $(f_r - f_t)$ is the Doppler shift due to the relative velocity satellite transmitter; c is the speed of light; α is the angle between the satellite velocity vector V and the transmitter position relative to the satellite.

The satellite ephemeris generator uses the SGP8 (Special General Perturbation) model of (North American Aerospace Defense Command) NORAD [6, 7], for obtaining the satellite

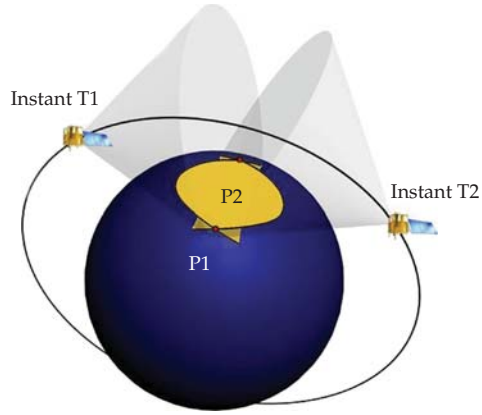


Figure 2: Location cones.

orbit at the measured Doppler shift times. The updated ephemeris are used within the Doppler effect equation to model the observations. Given the observations modeled as

$$y = h(x) + v, \tag{2.2}$$

where y is the set of Doppler shifts measured data; v is a noise vector assumed zero mean Gaussian; $h(x)$ is the nonlinear function relating the measurements to the location parameters and function of the satellite ephemeris:

$$h(x) = \frac{\left[(x-X)(\dot{x}-\dot{X}) + (y-Y)(\dot{y}-\dot{Y}) + (z-Z)(\dot{z}-\dot{Z})\right]}{\sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2}} + b_0 + b_1 \Delta t, \tag{2.3}$$

where (x, y, z) and (X, Y, Z) are the satellite and transmitter coordinates position; b_0 (drift) and b_1 (drift rate) are constants associated with each Doppler curve to account for unknown bias in the Doppler measurements and a possible drift in the transmitter oscillator; Δt is the diference time between the first Doppler measurement and the current one.

The nonlinear least squares solution [8] is

$$H_1 \delta\hat{x} = \delta y_1, \tag{2.4}$$

where $\delta\hat{x} = \hat{x} - \overline{x}$, and H_1 is a triangular matrix. The method turns out to be iterative as we take the estimated value $\hat{x}$ as the new value of the reference $\overline{x}$ successively until $\delta\hat{x}$ goes to zero. The H_1 matrix is the result of the Householder orthogonal [9] transformation T such that

$$\begin{bmatrix} H_1 \\ 0 \end{bmatrix} = T \begin{bmatrix} S_0^{1/2} \\ W^{1/2} H \end{bmatrix}, \tag{2.5}$$

where H is the partial derivatives matrix $[\partial h/\partial x]_{x=\overline{x}}$ of the observations relative to the state parameters (latitude, longitude, height, bias, drift, drift rate) around the reference values;

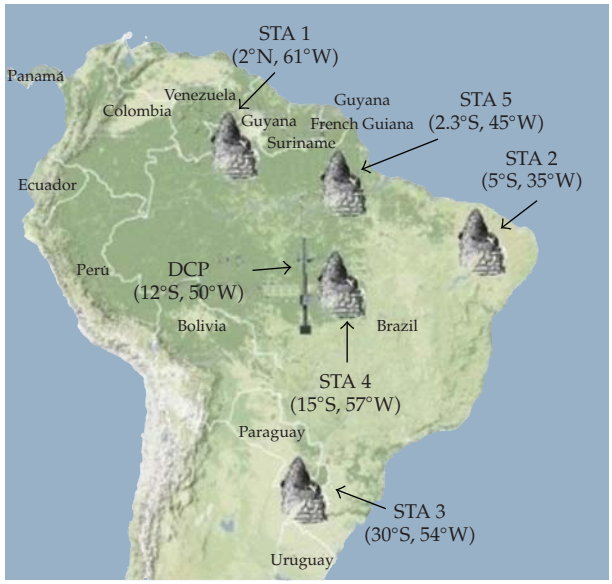


Figure 3: Reception stations configuring ideal geometry.

$W^{1/2}$ is the square root of the measurements weight matrix, and $S_0^{1/2}$ is the square root of the information matrix.

The δy_1 is such that

$$\begin{bmatrix} \delta y_1 \\ \delta y_2 \end{bmatrix} = T \begin{bmatrix} S_0^{1/2} \delta \hat{x}_0 \\ W^{1/2} \delta y \end{bmatrix}, \tag{2.6}$$

where δy is the residuals vector. The final cost function is

$$J = \left\| \delta y_1 - H_1 \delta \hat{x} \right\|^2 + \left\| \delta y_2 \right\|^2 \tag{2.7}$$

with $\|\delta y_2\|^2 = J_{\min}$, where $J_{\min}$ is the minimum cost. The whole detailed procedure is fully described in [10]

3. Results

In this section we show the results of the simulated Doppler shift measurements for ideal conditions, inserting random errors in the Doppler measurements, bias errors in the Doppler measurements, errors in the satellite ephemeris, errors in the time stamp and realistic random errors in the Doppler measurements using DRN.

For this analysis we initially considered the results using ideal conditions from simulated Doppler shift measurements of a transmitter located in the center of Brazil at latitude 12.1200°S and longitude 310.1100°W and five simulated ground reception stations configuring the ideal geometry as illustrated in Figure 3.

Table 1: Ideal conditions.

Satellite	Number of samples per pass	Mean location error (km)				
		STA 1	STA 2	STA 3	STA 4	STA 5
SCD-2	600	7.E-5 ± 3.E-5	4.E-5 ± 1.E-5	6.E-5 ± 2.E-5	4.E-5 ± 2.E-5	6.E-5 ± 2.E-5
	7	3.E-4 ± 1.E-4	7.E-4 ± 4.E-4	3.E-4 ± 1.E-4	5.E-4 ± 7.E-4	6.E-4 ± 4.E-4
NOAA-17	600	6.E-5 ± 2.E-5	1.E-5 ± 4.E-5	1.E-5 ± 3.E-5	3.E-5 ± 3.E-5	3.E-5 ± 1.E-5
	7	2.E-4 ± 1.E-4	2.E-4 ± 1.E-4	3.E-4 ± 1.E-4	1.E-4 ± 7.E-4	2.E-4 ± 8.E-4

The data were gathered from March 10 to 19, 2008. The SCD-2 ephemerides were provided by the Control Center of INPE and the NOAA-17 ephemerides were obtained from Internet at "http://www.celestrak.com/".

3.1. Simulated Doppler Shift Measurements in Ideal Conditions

Table 1 shows results under ideal conditions, without any errors (ideal).

In the third column (mean location error) we have the geographical location mean errors from the transmitter nominal position. They are listed for the 5 reception stations (STA1, STA2, STA3, STA4, STA5) of Figure 3, in terms of mean and standard deviations. From Table 1, we can observe that the mean location errors and their deviation standards for both SCD-2 and NOAA-17 satellites, in the case of 600 samples (Doppler measurements) per pass, are around 10^{-5} km. This is an evidence that the developed GEOLOC software (Sousa, 2000) using data that simulates ideal conditions provides precise results.

The files corresponding to high sampling rate (600 samples), which represent a transmission rate of one burst per second yield results one order of magnitude more precise than the files resulting that from low sampling rate (7 samples or around one transmission burst each 90 s). This emphasizes that a large number of measurements imply a better statistical result as expected.

3.2. Random Errors in the Simulated Doppler Measurements

In this section we present the analyses and results of simulated files under nonideal conditions inserting purposely errors in the Doppler shifts.

With the aim of verifying the effect of inserting Doppler measurements errors, the estimator of "bias" was turned off. Thus, the GEOLOC assumes that there is no "bias" in the Doppler measurements.

The random Gaussian errors (zero mean) with 1 Hz, 10 Hz, and 100 Hz (standard deviations) were inserted in the Doppler measurements. The corresponding results are presented in Tables 2(a) and 2(b) as follows.

From Tables 2(a) and 2(b) we can observe that as the random errors inserted in the Doppler measurements increase, the corresponding mean location errors increase.

We can also verify in the mean errors column that the resulting errors for STA 4 were smaller than for the other stations. This is a consequence of a higher number of measurements obtained by this station per each satellite pass and occurs because the fourth reception station is nearby the transmitter, and so they are almost inside the same satellite visibility circle of around 5000 km diameter for 5° minimum elevation angle.

Table 2

(a) Random errors in the doppler measurements for satellite SCD-2.

Satellite	Number of samples per pass	Random errors in the simulated Doppler measurements (Hz)	Locations processed (sum of the 5 stations)	Mean location error for reception stations STA 1 to STA 5 (km)
SCD-2	600	1	394	0.06 ± 0.02 0.03 ± 0.01 0.03 ± 0.05 0.02 ± 0.04 0.04 ± 0.06
	7	1	408	0.36 ± 0.27 0.44 ± 0.22 0.39 ± 0.10 0.31 ± 0.14 0.33 ± 0.22
	600	10	385	0.16 ± 0.12 0.14 ± 0.11 0.23 ± 0.05 0.13 ± 0.04 0.34 ± 0.06
	7	10	386	2.54 ± 1.51 3.39 ± 2.87 2.43 ± 1.84 2.33 ± 1.78 2.91 ± 1.82
	600	100	385	1.68 ± 0.98 1.43 ± 0.03 2.53 ± 1.81 1.12 ± 0.29 2.62 ± 0.48
	7	100	408	34.15 ± 23.39 30.74 ± 22.31 24.59 ± 11.46 23.67 ± 12.97 27.37 ± 17.40

(b) Random errors in the Doppler measurements for Satellite NOAA-17.

Satellite	Number of samples per pass	Random errors in the simulated Doppler measurements (Hz)	Locations processed (sum of the 5 stations)	Mean Location error for reception stations STA 1 to STA 5 (km)
NOAA-17	600	1	173	0.04 ± 0.09 0.06 ± 0.06 0.06 ± 0.06 0.09 ± 0.08 0.08 ± 0.01
	7	1	157	0.31 ± 0.24 0.30 ± 0.17 0.47 ± 0.14 0.29 ± 0.18 0.30 ± 0.15
	600	10	173	0.19 ± 0.01 0.18 ± 0.02 0.10 ± 0.02 0.12 ± 0.02 0.28 ± 0.14
NOAA-17	7	10	154	2.87 ± 1.00 3.87 ± 2.06 3.58 ± 1.60 2.00 ± 1.11 3.56 ± 1.18
	600	100	173	1.44 ± 0.57 1.69 ± 0.81 1.69 ± 0.31 1.24 ± 0.62 2.71 ± 0.71
	7	100	153	34.71 ± 24.55 35.97 ± 25.32 37.86 ± 20.07 21.80 ± 11.69 24.25 ± 15.20

Also when we consider the low sampling (7 measurements) compared to the high sampling (600 measurements) cases, for the same reason the location errors were about one order of magnitude higher. Thus, greater measurements quantity during one satellite pass will provide a better location result.

The conclusion is as expected: to improve the quality of the presented geographical location results, considering the reception stations independently, it would be necessary to install more receptions stations around Brazil territory, nearby the transmitters, to get full Doppler curve reconstitutions without loss of Doppler shift measurements.

3.3. Bias (Systematic) Errors in the Simulated Doppler Measurements

This section presents the results of purposely inserted bias errors effect in the simulated Doppler measurements. For the simulation we used the SCD-2 and NOAA-17 satellites and bias errors of 1 Hz, 10 Hz, and 100 Hz. The results are presented in Table 3 as follows.

We can observe in Tables 3(a) and 3(b) that, due to the inserted errors in the simulated Doppler measurements being systematic errors and not random errors nature, the results considering high and low sampling are in the same order of magnitude. Thus we conclude that the location errors resulting from bias errors are influenced in a lesser extent by the rate of the sampling.

Also, considering the differences among the random and biased simulations results, we can conclude from to Tables 2 and 3 that the random errors are filtered out. They produce mean locations results that tend to zero, mainly for high sampling rate (600 measures per pass). The location errors maintain some consistency and proportionality; thus, high sampling random errors of 1, 10, and 100 Hz result in location errors of 0.02, 0.2, and 2 km, respectively. For low sampling rate (7 measures per pass) the obtained results are also consistent: random errors of 1, 10, and 100 Hz produce location errors of 0.2, 2, and 20 km, respectively.

These results obtained were expected, due to the errors intrinsic characteristics. We can conclude that the smaller the measured Doppler error, the smaller becomes the location error, especially for high sampling rate. The inclusion of biased errors produces more degradation in the location results than that for random errors.

The GEOLOC location process with tendency (bias) in the Doppler measurements was calculated without estimating the drift in the measurements; however, these tendencies may be removed if the bias estimator is turned on in the location algorithm.

3.4. Errors in the Satellite Ephemerides

In this section we present the transmitters location error results considering inserted errors of 10 km in the satellites ephemerides. This analysis was performed to verify how the accuracy in the satellites ephemerides impacts the location accuracy. We also included errors of observation (Doppler measurements) of 10 Hz random and/or bias. The results considering ephemerides errors adding and not adding observation errors are described in Table 4.

Observing Tables 4(a) and 4(b) we verify that for the obtained locations, both using SCD-2 and NOAA-17, the largest errors appear when we insert all errors types simultaneously (see the last line for each satellite). From the table, we can see that the largest error contribution is due the errors inserted in the ephemerides. The random errors are filtered by the least squares algorithm and produce marginal inaccuracy. The systematic errors (bias) result in a greater final inaccuracy, and the added error levels are similar to those obtained in Table 3 for bias errors.

We can then conclude that the precision in the "two-lines" elements (ephemerides) produces direct impact in the location accuracy, that is, an error of 10 km in the ephemerides results in location errors of the same order of magnitude. Therefore considering the analyzed satellites, we can claim that, in low sampling rate conditions, precisions of 1-2 km order can only be obtained if the ephemerides errors are less than 2 km and the random and biased errors lower than 10 Hz.

Table 3

(a) Bias errors in the simulated Doppler measurements for satellite SCD-2.

Satellite	Number of samples per pass	Bias errors in the simulated Doppler measurements (Hz)	Locations processed (sum of the 5 stations)	Mean location error for reception stations STA1 to STA5 (km)
SCD-2	600	1	385	0.30 ± 0.03
				0.27 ± 0.01
				0.25 ± 0.05
				0.21 ± 0.05
				0.22 ± 0.07
	7	1	410	0.38 ± 0.09
				0.44 ± 0.07
				0.30 ± 0.08
				0.23 ± 0.02
				0.39 ± 0.04
	600	10	385	2.99 ± 0.21
				2.78 ± 0.52
				2.46 ± 0.21
				2.10 ± 0.48
				2.27 ± 0.70
	7	10	419	3.47 ± 0.34
				3.34 ± 0.97
				2.87 ± 0.46
				2.77 ± 0.19
				2.90 ± 0.14
	600	100	385	21.37 ± 4.88
				23.76 ± 6.07
				23.82 ± 8.74
				20.00 ± 4.80
				24.02 ± 8.64
	7	100	420	28.72 ± 7.91
				30.37 ± 10.05
				26.14 ± 9.80
				25.48 ± 5.40
				26.63 ± 6.83

(b) Bias errors in the simulated Doppler measurements for satellite NOAA-17.

Satellite	Number of samples per pass	Bias errors in the simulated Doppler measurements (Hz)	Locations processed (sum of the 5 stations)	Mean location error for reception stations STA1 to STA5 (km)
NOAA-17	600	1	173	0.27 ± 0.01 0.24 ± 0.02 0.31 ± 0.06 0.21 ± 0.05 0.22 ± 0.05
	7	1	155	0.32 ± 0.07 0.28 ± 0.07 0.28 ± 0.07 0.24 ± 0.05 0.34 ± 0.05
NOAA-17	600	10	173	2.71 ± 0.89 2.41 ± 0.43 2.49 ± 0.39 2.12 ± 0.24 2.19 ± 0.47
	7	10	161	3.69 ± 0.59 3.64 ± 0.52 3.64 ± 0.52 3.22 ± 0.18 3.49 ± 0.91
	600	100	173	26.52 ± 2.48 23.56 ± 3.11 27.75 ± 7.81 20.42 ± 2.36 21.994 ± 9.72
	7	100	158	28.79 ± 9.93 29.05 ± 5.05 28.23 ± 8.71 21.97 ± 5.72 31.82 ± 10.88

Comparing the presented Tables 2, 3, and 4, we can observe that the mean location error from the five simulated reception stations considering a transmission burst every 1s (600 samples per pass) is lower than for a burst every 90 s (7 samples per pass). This shows the importance of having a high rate DCP transmission burst to gather more Doppler data and a better location result.

Also, the obtained location results using NOAA-17 satellite are better than those using the SCD-2 satellite. This occurs because we simulate random errors of 0.1 Hz to the NOAA-17 and 1 Hz to the SCD-2 in the Doppler shift measurements, which is more consistent with their respective on board oscillator accuracies [10].

Table 4

(a) Satellite ephemeris and observed error for satellite SCD-2.

Satellite	Number of samples per pass	Errors in the satellite ephemeris (km)	Random errors in the simulated Doppler measurements (Hz)	Locations processed (sum of the 5 stations)	Mean location error for reception stations STA 1 to STA 5 (km)
SCD-2	600	10	No	357	8.48 ± 0.14 8.57 ± 0.13 8.53 ± 0.11 8.38 ± 0.13 8.67 ± 0.14
	7	10	“	410	8.84 ± 0.18 8.72 ± 0.11 8.83 ± 0.18 8.59 ± 0.11 8.75 ± 0.07
	600	10	10 random	357	8.79 ± 0.23 8.71 ± 0.26 8.74 ± 0.18 8.67 ± 0.18 8.68 ± 0.27
	7	10	“	410	9.19 ± 0.35 9.01 ± 0.27 9.64 ± 0.14 9.77 ± 0.12 9.23 ± 0.32
	600	10	10 biased	357	10.53 ± 0.26 10.96 ± 2.42 10.60 ± 0.38 10.76 ± 0.24 10.78 ± 0.43
	7	10	“	412	11.41 ± 0.41 11.10 ± 0.32 11.17 ± 0.21 11.58 ± 0.12 11.50 ± 0.31

(a) Continued.

Satellite	Number of samples per pass	Errors in the satellite ephemeris (km)	Random errors in the simulated Doppler measurements (Hz)	Locations processed (sum of the 5 stations)	Mean location error for reception stations STA 1 to STA 5 (km)
					10.89 ± 0.32
					10.89 ± 0.39
			10 random		10.63 ± 0.24
	600	10	+ 10 biased	357	10.76 ± 0.26
					10.81 ± 0.30
					11.30 ± 0.37
					11.58 ± 0.33
	7	10	"	410	11.99 ± 0.31
					11.73 ± 0.23
					11.88 ± 0.34
					10.65 ± 0.36
					10.62 ± 0.37
	600	10	10 biased	174	10.44 ± 0.22
					10.79 ± 0.26
					10.68 ± 0.33
					11.72 ± 0.33
					11.82 ± 0.32
	7	10	"	155	11.02 ± 0.28
					11.04 ± 0.23
					11.84 ± 0.36
SCD-2					10.69 ± 0.48
					10.62 ± 0.37
			10 random	174	10.45 ± 0.14
	600	10	+ 10 biased		10.21 ± 0.12
					10.86 ± 0.34
					11.97 ± 0.32
					11.47 ± 0.28
	7	10	"	148	11.10 ± 0.36
					11.05 ± 0.31
					11.36 ± 0.39

(b) Satellite ephemeris and observed error for satellite NOAA-17.

Satellite	Number of samples per pass	Errors in the satellite ephemeris (km)	Random errors in the simulated Doppler measurements (Hz)	Locations processed (sum of the 5 stations)	Mean location error for reception stations STA 1 to STA 5 (km)
NOAA-17	600	10	No	174	8.32 ± 0.10 8.22 ± 0.10 8.57 ± 0.11 8.12 ± 0.10 8.58 ± 0.12
	7	10	"	164	8.77 ± 0.14 8.87 ± 0.12 8.70 ± 0.19 8.66 ± 0.11 8.78 ± 0.12
	600	10	10 random	174	8.62 ± 0.24 8.69 ± 0.27 8.67 ± 0.15 8.55 ± 0.14 8.71 ± 0.28
	7	10	"	160	9.56 ± 0.21 9.15 ± 0.21 9.11 ± 0.31 9.58 ± 0.10 9.19 ± 0.31
	600	10	10 biased	174	10.65 ± 0.36 10.62 ± 0.37 10.44 ± 0.22 10.79 ± 0.26 10.68 ± 0.33
	7	10	"	155	11.72 ± 0.33 11.82 ± 0.32 11.02 ± 0.28 11.04 ± 0.23 11.84 ± 0.36
	600	10	10 random + 10 biased	174	10.69 ± 0.48 10.62 ± 0.37 10.45 ± 0.14 10.21 ± 0.12 10.86 ± 0.34
	7	10	"	148	11.97 ± 0.32 11.47 ± 0.28 11.10 ± 0.36 11.05 ± 0.31 11.36 ± 0.39

The SCD-2 satellite ephemerides used are computed just once a week by the INPE Control Center of São José of Campos. In order to improve the quality of the obtained results it would be necessary that the ephemerides be calculated with a higher precision, that is, it is recommended a daily computation instead of a week one. The NOAA satellite ephemerides used are obtained with high precision by (Collecte Localisation Satellites) CLS/Argos using their own method and not supplied to the public. The two-line format NOAAs ephemerides supplied by Internet site, like SCD-2's, are also imprecise, according to [1].

From the tables' results, we conclude that it is fundamental to minimize the errors in the satellite ephemerides, because the location algorithm cannot compensate for them. The errors inserted in the satellites ephemeris are approximately similar in magnitude to the resulting errors in the transmitters' location.

3.5. Errors in the Time Stamp

In this section we present the results and analysis of inserted errors of 1 and 0.1 seconds ahead in each transmitted signal frequency time stamp. This analysis was accomplished to verify how the accuracy in time impacts the location accuracy. The results are described in Table 5.

Observing Table 5 we verify that the mean location error values increase in the same order the errors in the time stamp increase. We can also conclude that a time delay of up to 1 s regarding the uplink signal transmitted to the satellite can cause an inaccuracy of up to 10 km in the location.

Thus, to avoid losing the quality on the transmitters geographic location results the measurements time tagging should be very stable; otherwise we will have one more unexpected error in the final location result.

4. Simulation Using the (Data Reception Network) DRN

In Section 3 we showed the analyses of the location results using collected measurements of several reception stations during a satellite pass and the impact of inserting errors on the transmitters geographical locations. Now the main goal of this present section is to show the location results when merging the Doppler measurements of all reception stations during the satellite pass in the location algorithm. We notice that using this approach, the final location results enhance considerably.

In this section the simulations consider a (Data Reception Network) DRN composed of five reception stations according to Figure 3. We insert random errors of 1 Hz in the Doppler shift measurements and consider one transmission each 90 s for a period of 10 minutes of the SCD-2 and the NOAA-17 satellite passes.

Tables 6 and 7 present the location errors considering five reception stations separately and as a network (DRN) during March 10 and 11, 2008.

In Tables 6 and 7, the first and second columns of the tables present the day and hour of each satellite pass. The next six columns present the location errors inserting random error of 1Hz in the simulated Doppler measurements, over (sign /) the respective measurements amount obtained during the satellite pass.

The last column shows the location errors using the DRN network. In the cells without results, the symbol "—" means that there were no data to compute the location. As shown in the last column, when we use the DRN to merge data from the five reception stations, we can obtain results that were not present before, as well as improved results.

Table 5: Errors inserted in the time stamp.

Satellite	Number of samples per pass	Error in the time (seg)	Locations processed (sum of the 5 stations)	Mean location error for reception stations STA 1 to STA 5 (km)
SCD-2	600	1	385	1.54 ± 0.13
				1.16 ± 0.30
				1.02 ± 0.46
				0.71 ± 0.23
				0.90 ± 0.23
	7	1	411	9.35 ± 0.16
				8.96 ± 0.41
				8.88 ± 0.36
				7.48 ± 0.17
				8.28 ± 0.18
	600	0.1	385	0.18 ± 0.39
				0.11 ± 0.19
				0.10 ± 0.13
				0.08 ± 0.05
				0.09 ± 0.09
	7	0.1	411	1.03 ± 0.14
				1.05 ± 0.50
				1.11 ± 0.02
				0.89 ± 0.09
				0.97 ± 0.29
NOAA-17	600	1	173	1.29 ± 0.10
				1.68 ± 0.20
				1.68 ± 0.30
				2.13 ± 0.13
				2.33 ± 0.23
	7	1	166	11.81 ± 0.17
				8.87 ± 0.14
				7.94 ± 0.19
				9.51 ± 0.13
				9.51 ± 0.13
	600	0.1	173	0.47 ± 0.19
				0.62 ± 0.21
				0.62 ± 0.31
				0.64 ± 0.16
				0.22 ± 0.24
	7	0.1	170	1.26 ± 1.28
				0.85 ± 0.17
				1.09 ± 0.29
				0.65 ± 0.17
				0.65 ± 0.27

Table 6: Location error considering the DRN and SCD-2.

March 2008		Location error (km) with simulated Doppler measurements (with 1 Hz random error and transmission burst of 90 s), using the following reception stations					/Number of samples per pass:
Day	Hour	STA 1	STA 2	STA 3	STA 4	STA 5	All 5 stations
10	10	0.28/6	0.31/7	0.16/7	0.39 /9	0.46/9	0.14/38
10	12	0.12/6	0.08/6	0.16/9	0.12/9	0.11/8	0.05/38
10	14	0.36/3	0.41/4	0.15/8	0.30/8	0.22/6	0.05/29
10	16	—	0.61/5	0.29/8	0.16/7	0.17/6	0.13/26
10	17	0.31/5	0.51/5	0.18/7	0.13/8	—	0.11/25
10	19	0.30/8	0.31/7	0.25/6	0.14/8	0.25/8	0.13/37
10	21	0.15/7	0.03/5	—	0.47/6	0.38/7	0.02/25
11	08	0.07/6	0.21/7	—	0.09/6	0.14/7	0.07/26
11	09	0.26/7	0.08/7	0.35/7	0.17/9	0.35/9	0.08/39
11	11	0.08/5	0.07/6	0.06/8	0.19/8	0.18/8	0.03/35
11	13	1.36/4	0.95/4	0.13/8	0.15/8	0.19/6	0.12/30
11	15	—	0.19/5	0.06/8	0.20/8	0.09/6	0.04/27
11	17	0.61/4	0.35/6	0.06/8	0.16/8	0.32/7	0.03/33
11	18	1.79/8	0.13/6	0.99/7	1.55/9	0.69/7	0.03/37
11	20	0.17/8	0.17/6	0.61/3	0.30/7	0.14/7	0.12/31

Table 7: Location error considering the DRN and NOAA-17.

March 2008		Location error (km) with simulated Doppler measurements (with 1 Hz random error and transmission burst of 90 s), using the following reception stations					/Number of samples per pass:
Day	Hour	STA 1	STA 2	STA 3	STA 4	STA 5	All 5 stations
10	00	0.09/6	0.05/9	0.06/6	0.05/9	0.11/8	0.02/38
10	02	0.14/4	—	0.10/5	0.38/5	— /1	0.07/15
10	12	—/—	0.08/7	0.26/4	0.14/6	0.08/7	0.06/24
10	13	0.17/7	0.20/5	0.25/6	0.43/8	0.12/6	0.07/32
11	00	0.13/4	0.50/7	0.48/4	0.19/7	0.41/7	0.28/29
11	02	0.52/6	—	0.66/5	0.22/7	0.55/5	0.03/23
11	11	—	0.49/5	0.57/3	0.64/4	0.34/4	0.13/16
11	13	0.11/7	0.17/7	0.15/6	0.26/8	0.10/8	0.05/36

In most of the results the locations obtained for satellite passes with a larger number of measurements are better than those for smaller number of measurements.

As we see in the last column, the quality of the obtained results was improved when considering measurements collected by all reception stations compared to one single station alone. Therefore, there is an enhancement in the location results when generated using the DRN. For example we notice in Table 7 that Stations 2 and 5 (STA 2 and STA 5) on day 10

at 02 o'clock have only 0 and 1 data points, respectively, insufficient for a location. When we merge the whole data set (all other stations), the Doppler curve is fully recovered and the location error drops (improves) drastically to 0.07 km.

5. Conclusions

We conclude that the developed GEOLOC is operating appropriately as we can see by the obtained results under ideal conditions, say, without errors in the Doppler measurements, ephemeris, and time.

Considering the random and biased simulations results, we conclude that the random errors are filtered out by the least squares algorithm. They produce mean locations results that tend to zero error, mainly for high sampling rate (600 measures per pass). The simulation results considering biased errors yield errors that equally degrade the location for both sampling rates (600 and 7). The inclusion of biased errors degrades the location results more than the random errors. From the tables with results inserting errors in the satellite ephemerides, we can conclude that it is fundamental to minimize such errors, because the location system cannot compensate for them. The satellites ephemeris errors are approximately similar in magnitude to their resulting transmitters' location errors.

The simulation results using the DRN showed that to improve the location results quality it would be necessary to have more Reception Stations than the existing Cuiabá, Cachoeira Paulista, and Alcântara, spread over the Brazilian territory, to increase the data amount. Then, on the other hand, it improves the geometrical coverage between satellite and DCPs, and recovers better the full Doppler curves, yielding as a consequence more valid and improved locations.

Acknowledgment

The authors thank FAPESP (Process no. 2005/04497-0) for the financial support.

References

[1] C. T. Sousa, H. K. Kuga, and A. W. Setzer, "Geo-Location of transmitters using real data, Doppler shifts and Least Squares," *Acta Astronautica*, vol. 52, no. 9, pp. 915–922, 2003.

[2] M. Kampel and M. R. Stevenson, "Heat transport estimates in the surface layer of the Antarctic polar front using a satellite tracked drifter—first results," in *Proceedings of the 5th International Congress of the Brazilian Geophysical Society*, São Paulo, Brazil, September 1997.

[3] C. M. M. Muelbert, et al., "Movimentos sazonais de elefantes marinhos do sul da ilha elefante, Shetland do sul, antártica, observações através de telemetria de satélites," in *Proceedings of the 7th Seminars on Antartica Research*, p. 38, USP/IG, São Paulo, Brazil, 2000, program and abstracts.

[4] Techno-Sciences, COSPAR/SARSAT, October 2000, http://www.technosci.com/.

[5] H. Resnick, *Introduction to Special Relativity*, John Wiley & Sons, New York, NY, USA, 1968.

[6] F. R. Hoots and R. L. Roehrich, "Models for propagation of NORAD element sets," Spacetrack Report 3, Aerospace Defense Command, Peterson Air Force Base, Colorado Springs, Colo, USA, 1980.

[7] R. Vilhena de Moraes, A. Silva, and H. K. Kuga, "Low cost orbit determination using GPS," in *Proceedings of the 5th International Conference on Mathematical Problems in Engineering and Aerospace Sciences*, S. Sivasundaram, Ed., pp. 495–502, Cambridge Scientific, Cambridge, UK, 2006.

[8] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, NY, USA, 1977.

[9] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1974.

[10] C. T. Sousa, *Geolocalização de transmissores com satélites usando desvio doppler em tempo quase real*, PhD dissertation, Space Engineering and Technology, Space Mechanics and Control Divison, INPE-Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil, 2000.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 140963, 12 pages
doi:10.1155/2009/140963

Research Article

A Discussion Related to Orbit Determination Using Nonlinear Sigma Point Kalman Filter

Paula Cristiane Pinto Mesquita Pardal,[1] Helio Koiti Kuga,[1] and Rodolpho Vilhena de Moraes[2]

[1] *INPE (National Institute for Space Research), DMC, Avenue dos Astronautas, 1.758. Jd. Granja, CEP 12227-010, São José dos Campos, SP, Brazil*
[2] *UNESP (Univ Estadual Paulista), Guaratinguetá, CEP 12516-410, SP, Brazil*

Correspondence should be addressed to Paula Cristiane Pinto Mesquita Pardal, paulacristiane@gmail.com

Received 30 July 2009; Accepted 9 November 2009

Recommended by Tadashi Yokoyama

Herein, the purpose is to present a Kalman filter based on the sigma point unscented transformation development, aiming at real-time satellite orbit determination using GPS measurements. First, a brief review of the extended Kalman filter will be done. After, the sigma point Kalman filter will be introduced as well as the basic idea of the unscented transformation, in which this filter is based. Following, the unscented Kalman filter applied to orbit determination will be explained. Such explanation encloses formulations about the orbit determination through GPS; the dynamic model; the observation model; the unmodeled acceleration estimation; also an application of this new filter approaches on orbit determination using GPS measurements discussion.

1. Introduction

In orbit determination of artificial satellites, the dynamic system and the measurements equations are of nonlinear nature. It is a nonlinear problem in which the disturbing forces are not easily modeled. The problem consists of estimating variables that completely specify the body trajectory in the space, processing a set of information (pseudorange measurements) related to this body. A tracking network on Earth or through sensors, like the GPS receiver onboard Topex/Poseidon (T/P) satellite, can collect such observations.

The Global Positioning System (GPS) is a powerful and low cost means to allow computation of orbits for artificial Earth satellites by means of redundant measurements. The T/P is an example of using GPS for space positioning.

The Extended Kalman Filter (EKF) implementation in orbit estimation, under inaccurate initial conditions and scattered measurements, can lead to unstable or diverging solutions. For solving the problem of nonlinear nature, convenient extensions of the Kalman filter have been sought. In particular, the unscented transformation was developed as a method to propagate mean and covariance information through nonlinear transformations. The Sigma Point Kalman Filter (SPKF) appears as an emerging estimation algorithm applied to nonlinear system, without needing linearization steps.

2. Extended Kalman Filter

The real-time estimators are a class of estimators which fulfill the real time requirements. They are recursive algorithms and produce sequentially the state to be estimated. Among them, the Kalman filter and its variations are outstanding.

The Kalman filter is the recursive estimator most used nowadays because it is easy to implement and to use on digital computers. Its recursiveness leads to lesser memory storage, which makes it ideal for real-time applications. The EKF is a nonlinear version of the Kalman filter that generates reference trajectories which are updated at each measurement processing, at the corresponding instant. Details can be found in Brown and Hwang [1].

Due to the complexity of modeling the artificial satellites orbit dynamics accurately, the EKF is generally used in works of such nature. The EKF algorithm always brings up to date reference trajectory around the most current available estimate.

The KF filter consists of phases of time and measurement updates. In the first, state and covariance are propagated from one previous instant to a later one, meaning that they are propagated between discrete instants of the system dynamics model. In the second one, state and covariance are corrected for the later instant corresponding to the measurement time, through the observations model. This method has, therefore, recursive nature and it does not need to store the measurements previously in large matrices.

Exploiting the assumption that all transformations are quasilinear, the EKF simply linearizes all nonlinear transformations and substitutes the Jacobian matrices for the linear transformations in the Kalman filter equations. Although the EKF maintains the elegant and computationally efficient recursive update form of the Kalman Filter, it suffers a number of serious limitations.

The first limitation is that linearized transformations are only reliable if the error propagation can be matched approximated by a linear function. If this condition does not hold, the linearized approximation can be extremely poor. At best, this undermines the performance of the filter. The second is that linearization can be applied only if the Jacobian matrix exists. However, this is not always the case. Some systems contain discontinuities, singularities, and the states themselves are inherently discrete. And the last is that calculating Jacobian matrices can be a very difficult and error-prone process. The Jacobian equations frequently produce many pages of dense algebra that must be converted to code. This introduces numerous opportunities for human coding errors that may degrade the performance of the final system in a manner that cannot be easily identified and debugged. Regardless of whether the obscure code associated with a linearized transformation is or is not correct, it presents a serious problem for subsequent users who must validate it for use in any high integrity system.

Summarizing, the Kalman filter can be applied to nonlinear systems if a consistent set of predicted quantities can be calculated. These quantities are derived by projecting a prior

estimate through a nonlinear transformation. Linearization, as applied in the EKF, is widely recognized to be inadequate, but the alternatives incur substantial costs in terms of derivation and computational complexity. Therefore, there is a strong need for a method that is probably more accurate than linearization but does not incur the implementation nor computational costs of other higher-order filtering schemes. The Unscented Transformation was developed to meet these needs.

3. Sigma Point Kalman Filters

If the dynamics system and the observation model are linear, the conventional Kalman filter can be used fearlessly. Although, not rarely, the dynamic systems and the measurement equations are nonlinear, convenient extensions of the Kalman Filter have been sought.

The SPKF is a new estimator that allows similar performance than the Kalman filter for linear systems and it elegantly extends to nonlinear systems, without the linearization steps. These filters are a new approach to generalize the Kalman filter for nonlinear process and observation models. A set of weighted samples, called sigma points, is used for normalizing mean and covariance of a probability distribution. This technique is claimed to lead to a filter more accurate and easier to implement than the EKF or a second-order Gaussian filter.

The SPKF approach is described by Van Merwe et al. [2] as follows.

(1) A set of weighted samples is deterministically calculated, based on mean and covariance decomposition of a random variable. One minimum need is that the first- and second-order momentums are known.

(2) The sigma points are propagated through the real nonlinear function, using only functional estimation, that is, analytical derivatives are not used to generate a posteriori set of sigma points.

(3) The later statistics are calculated using propagated sigma points functions and weights. In general, they assume the form of a simple weighted average of the mean and the covariance.

3.1. Basic Idea: The Unscented Transformation

A recent method to calculate the statistics of a random variable that passes through a nonlinear transformation is the unscented transform (UT). The UT builds on the principle that it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function, which is detailed in Julier and Uhlmann [3]. The approach is simple: a set of points (sigma points) are chosen so that their mean and covariance are $\overline{x}$ and $\mathbf{P}_{xx}$, according to Julier et al. [4, 5], and Julier and Uhlmann [6]. The nonlinear function is applied to each point, in turn, to yield a cloud of transformed points. The statistics of the transformed points, mean $\overline{\mathbf{y}}$, and covariance $\mathbf{P}_{yy}$ predicted, can then be calculated to form an estimate of the nonlinearly transformed mean and covariance.

There are several fundamental differences to the particle filters, although this method bears resemblance it. First, the sigma points are deterministically chosen so that they exhibit certain specific properties (like a given mean and covariance), and are not drawn at random. Second, sigma points can be weighted in ways that are inconsistent with the distribution interpretation of sample points in a particle filter.

The n-dimensional random variable $\mathbf{x}$, where n is the state vector dimension, with $\overline{\mathbf{x}}$ mean and $\mathbf{P}_{\mathbf{xx}}$ covariance, is approximated by $2n + 1$ weighted points, given by

$$\begin{aligned} \chi_0 &= \overline{\mathbf{x}}, \\ \chi_i &= \overline{\mathbf{x}} + \left(\sqrt{(n+\kappa)\mathbf{P}_{\mathbf{xx}}}\right)_i, \\ \chi_{i+n} &= \overline{\mathbf{x}} - \left(\sqrt{(n+\kappa)\mathbf{P}_{\mathbf{xx}}}\right)_i \end{aligned} \tag{3.1}$$

in which $\kappa \in \Re$, $\left(\sqrt{(n+\kappa)\mathbf{P}_{\mathbf{xx}}}\right)_i$ is the ith row or column of the root square matrix of $(n+\kappa)\mathbf{P}_{\mathbf{xx}}$, and W_i is the weight associated to the ith point

$$\begin{aligned} W_0 &= \frac{\kappa}{(n+\kappa)}, \\ W_i &= \frac{1}{2(n+\kappa)}, \quad i = 1, \ldots, n, \\ W_{i+n} &= \frac{1}{2(n+\kappa)}, \quad i = 1, \ldots, n. \end{aligned} \tag{3.2}$$

The transformation in the prediction step of the EKF occurs as follows.

(1) Transform each point through the function to yield the set of transformed sigma points

$$y_i = \mathbf{f}\left[\chi_i\right]. \tag{3.3}$$

(2) The observations mean is given by the weighted average of the transformed points

$$\overline{\mathbf{y}} = \sum_{i=0}^{2n} W_i y_i. \tag{3.4}$$

(3) The covariance is the weighted outer product of the transformed points

$$\mathbf{P}_{yy} = \sum_{i=0}^{2n} W_i \left[y_i - \overline{y}\right]\left[\mathbf{y}_i - \overline{\mathbf{y}}\right]^{\mathrm{T}}. \tag{3.5}$$

As the algorithm works with a finite number of sigma points, it naturally lends itself to being used in a "black box" filtering library. Given a model with defined inputs and outputs, a standard routine can be used to calculate the predicted quantities as necessary for any given transformation. The computational cost of the algorithm is the same order of magnitude as the EKF. The most expensive operations are to calculate the matrix square root and the outer products which are required to compute the covariance of the projected sigma points.

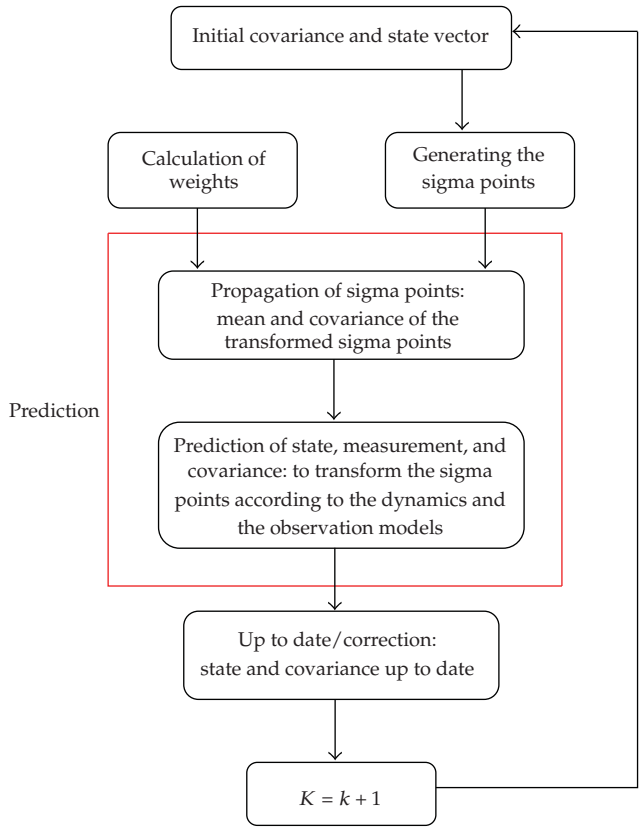


Figure 1: Modified EKF, leading to UKF.

Any set of sigma points that encodes the mean and covariance correctly calculates the projected mean and covariance correctly to the second order. Therefore, the estimate implicitly includes the second-order "bias correction" term of the truncated second-order filter, but without the need to calculate any derivatives.

The algorithm can be used with discontinuous transformations. Sigma points can pass over a discontinuity and, thus, can approximate the effect of a discontinuity on the transformed estimate.

3.2. The Unscented Kalman Filter

Using UT, the Kalman filter processes, as summarized in the following steps.

The steps shown in Figure 1 are detailed next.

(1) To predict the new state system $\widehat{\mathbf{x}}(k+1 \mid k)$ and its associated covariance $\mathbf{P}(k+1 \mid k)$, taking into account the effects of the process white gaussian noise.

(2) To predict the expected observation $\widehat{\mathbf{z}}(k+1 \mid k)$ and its residual covariance (innovation) $\mathbf{P}_{vv}(k+1 \mid k)$, considering the effects of the observation noise.

(3) To predict the cross correlation matrix $\mathbf{P}_{xz}(k+1 \mid k)$.

These steps are put in order in the Kalman filter with the restructuring of dynamics, state vector, and observations models. First, the state vector is added of the noise vector $\mathbf{w}_k$, with dimension $q \times 1$, in order to obtain a vector of dimension $n^a = n + q$,

$$\mathbf{x}^a(k) = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{w}_k \end{bmatrix}. \tag{3.6}$$

The dynamics model is rewritten in function of $\mathbf{x}^a(k)$ as

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}^a(k)], \tag{3.7}$$

and the UT uses $2n^a + 1$ sigma points, generated by

$$\widehat{\mathbf{x}}^a(k \mid k) = \begin{pmatrix} \widehat{\mathbf{x}}(k \mid k) \\ \mathbf{0}_{q\times 1} \end{pmatrix}, \qquad \mathbf{P}^a(k \mid k) = \begin{bmatrix} \mathbf{P}(k \mid k) & \mathbf{P}_{xv}(k \mid k) \\ \mathbf{P}_{xv}(k \mid k) & \mathbf{Q}_{xv}(k \mid k) \end{bmatrix}. \tag{3.8}$$

The matrices in the principal diagonal of $\mathbf{P}^a(k \mid k)$ are the variances, and the ones out of it are the correlations between the state dynamic errors and the Gaussian process noises.

There are several extensions and modifications that can be done in this basic method to consider specific details for one given application. In the next section, it will be presented a discussion of the orbit determination, in real time, using UKF.

3.3. Comparing EKF and UKF

The conventional nonlinear filters, such as the linearized or the extended Kalman Filter, many times have a poor performance when applied to nonlinear problems, due to two known difficulties.

The linearization (of the dynamic and the measurements models) can lead to a highly instable performance of the filter if the time discretization is not enough small.

The derivation of the Jacobian is not simple in most applications, and usually it makes the implementation difficult.

The UKF has more advantages, when compared to the EKF, as Lee and Alfriend [7, 8] wrote, in the following aspects.

(i) It allows more stable and accurate estimates of mean and covariance.

(ii) It can estimate discontinuous functions.

(iii) No explicit derivation of the Jacobian and/or Hessian matrix is necessary.

(iv) It is suitable for parallel processing.

4. Using UKF on Orbit Determination

Before presenting a discussion of the orbit determination, in real time, using UKF, some points need to be outlined: the orbit determination; the dynamic model; the observations model.

4.1. The Orbit Determination

The orbit determination will be based on GPS technology, whose working principle is based on the geometric method. In such method, the observer knows the set of satellites position in the reference system, obtaining its own position in the same reference frame. Figure 2 presents the basic parameters used by GPS for user position determination.

Here, $\vec{R}_{\text{GPSi}}$ is the position of the ith GPS satellite in the reference system; $\vec{\rho}_i$ is the pseudorange, the user satellite position in respect to the ith GPS satellite; $\vec{r}_u$ is the user satellite position in the reference system.

4.2. The Dynamic Model

In the case of orbit determination via GPS, the ordinary differential equations that represent the dynamic model are as follows:

$$\begin{aligned} \dot{\vec{r}} &= \vec{v}, \\ \dot{\vec{v}} &= -\mu \frac{\vec{r}}{r^3} + \vec{a} + \vec{w}, \\ \dot{b} &= d, \\ \dot{d} &= 0 + w_d \end{aligned} \tag{4.1}$$

with variables given in the inertial reference frame. In the equations above, $\vec{r}$ is the vector containing the position components (x, y, z); $\vec{v}$ is the vector of velocity components; $\vec{a}$ represents the modeled perturbations; $\vec{w}$ is the white noise vector with covariance Q; b is the user clock bias; d is the user clock drift; wd is the white noise on the drift rate with variance Q_d.

4.3. The Observations Model

The nonlinear equation of the observations model is given by:

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, t) + \boldsymbol{\nu}_k, \tag{4.2}$$

where $\mathbf{z}_k$ is the vector of m observations; $\mathbf{h}_k(\mathbf{x}_k)$ is the nonlinear function of state $\mathbf{x}_k$, with dimension m; $\boldsymbol{\nu}_k$ is the vector of observation errors with dimension m.

4.4. Estimation of the Unmodeled Accelerations

Some spacecraft missions require precise orbit knowledge to support payload experiments. Sometimes after launch, ground-based orbit determination solutions do not provide the level of accuracy expected. After verifying all known dynamic models, there may be a residual signature in the orbit as result of unmodeled accelerations. This leads to attempt to estimate anomalous accelerations during the orbit fit, if sufficient data exist. If successful, the acceleration estimates can improve the fit residuals, and also results in better orbital position estimates, as can be seen in Soyka and Davis [9].

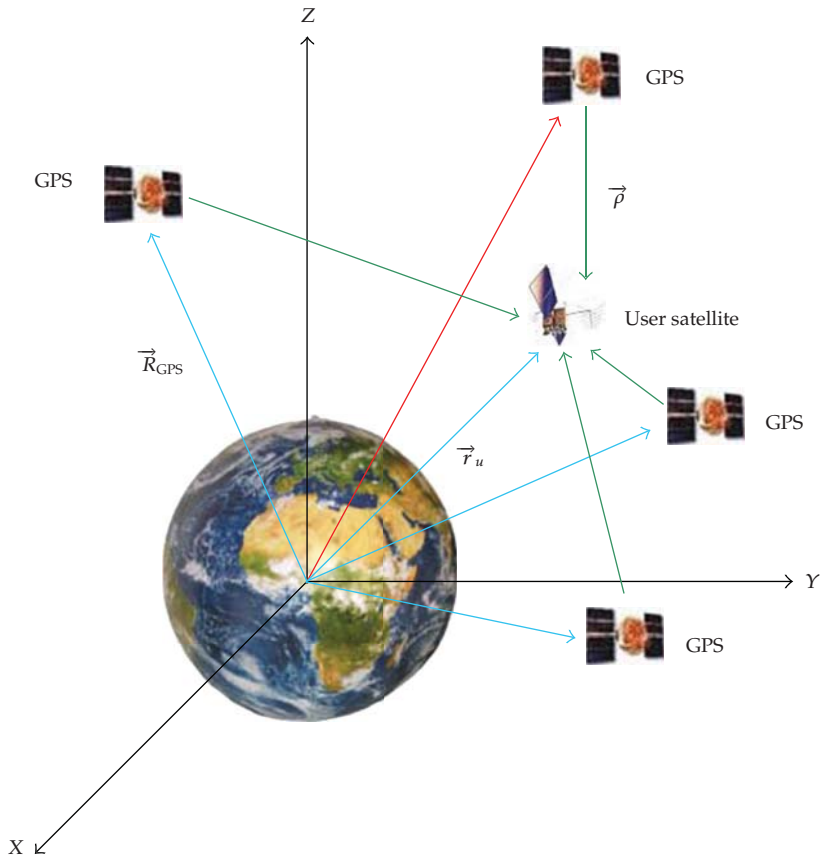


Figure 2: The geometric method.

Unmodeled accelerations may have many reasons: truncation of geopotential field; limitations of modeling solar pressure, Earth albedo, Earth infrared radiation, drag, and others. Some of these accelerations can be corrected through the use of higher fidelity dynamic and physical modeling, while others require postlaunch calibration.

The use of periodic accelerations, with a period near once per revolution of the satellite orbit, has been used within precision orbit determination programs to improve the accuracy of the derived ephemeris.

4.4.1. Anomalous Accelerations Modeling

When defining an anomalistic or periodic acceleration, one must consider three aspects: the subarc interval, the type of function, and the coordinate frame.

Subarc Interval

The subarc interval is the time of duration or number of revolutions for a given acceleration to be active. As its name implies, it is usually a subset of the total arc. A reason to break an arc into a subarc is to allow for better overall fits.

Type of Function

The underlying mathematical function of an acceleration function is usually a constant, a sine, or a cosine function.

The constant function is the most basic: a constant force in a specific direction. And the periodic functions (sine or cosine) have amplitude, frequency, and phase associated with them. The periodic functions are written as

$$\text{acc} = A\sin(\omega t + \phi_A) \quad \text{or} \quad \text{acc} = B\cos(\omega t + \phi_B), \tag{4.3}$$

where A and B are amplitudes, ω is the frequency, t is the time elapsed since the start of the periodic function reference point or subarc interval, ϕ_A and ϕ_B are the phase offsets. Either of these accelerations can be rewritten as:

$$\text{acc} = A'\sin(\omega t) + B'\cos(\omega t), \tag{4.4}$$

where for a sine acceleration with phase, $A' = +A\cos\phi_A$; $B' = +B\sin\phi_A$, and, for a cosine acceleration with phase, $A' = -B\cos\phi_B$; $B' = +B\sin\phi_B$. When estimated, the amplitudes A' and B' will adjust themselves to produce an effective phase offset.

Coordinate Frame

The selection of the start of the subarc can be important, especially for noncircular orbits. Conventionally, equator crossings, argument of perigee, mean anomaly, or orbit angle have been used as reference point.

4.5. Discussion of the Application

Now that the purpose of the discussion is known, it is possible to present a discussion about the subject.

In order to generate the UKF, it is necessary to rewrite the Kalman filter from UT. First, the state vector is increased, with the measurements noise vector $\mathbf{w}_k$, yielding a vector with dimension $n^a = n + q$. The increased versions of the state and the covariance are

$$\mathbf{x}^a(k) = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{w}_k \end{bmatrix}, \qquad \mathbf{P}_k^a = \begin{bmatrix} \mathbf{P}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_k \end{bmatrix}. \tag{4.5}$$

This increase can still contain $\boldsymbol{\upsilon}_k$, the Gaussian process noise, of dimension $l \times 1$. The new covariance matrix would have $\mathbf{R}_k$ in the principal diagonal, the covariance of such noise, and the new state vector dimension would be $n^a = n + q + l$, according to Lee and Alfriend [10].

Next, the increased set of sigma points is built:

$$\begin{aligned}
\mathcal{X}_{0,k}^{a} &= \overline{\mathbf{x}}_{k}^{a}, \\
\mathcal{X}_{i,k}^{a} &= \overline{\mathbf{x}}_{k}^{a} + \left(\sqrt{(n^{a}+\lambda)\mathbf{P}_{k}^{a}}\right)_{i}, \quad i = 1, \ldots, n^{a}, \\
\mathcal{X}_{i,k}^{a} &= \overline{\mathbf{x}}_{k}^{a} - \left(\sqrt{(n^{a}+\lambda)\mathbf{P}_{k}^{a}}\right)_{i}, \quad i = n^{a}+1, \ldots, 2n^{a}
\end{aligned} \tag{4.6}$$

with $\lambda = \alpha^{2}\ (n^{a}+\kappa) - n^{a}$, where α control the sigma points scattered about the mean $\overline{\mathbf{x}}_{k}^{a}$, and it is usually chosen small, in the interval $10^{-4} \leq \alpha \leq 1$, as Jwo and Lai [11] wrote; κ provides an extra degree of freedom; $\left(\sqrt{(n^{a}+\lambda)\mathbf{P}_{k}^{a}}\right)_{i}$ is the ith row or column of the root square matrix of $(n^{a}+\lambda)\mathbf{P}_{k}^{a}$.

In the propagation step, the state vector and covariance predicted are calculated based on the mean and the covariance of the propagated sigma points, transformed from the state vector and dynamical noises

$$\begin{aligned}
\overline{\mathbf{x}}_{k+1}^{-} &= \sum_{i=0}^{2n_{a}} W_{i}\mathcal{X}_{i,k+1}^{x}, \\
\overline{\mathbf{P}}_{k+1}^{-} &= \sum_{i=0}^{2n_{a}} W_{i}\left[\mathcal{X}_{i,k+1}^{x} - \overline{\mathbf{x}}_{k+1}^{-}\right]\left[\mathcal{X}_{i,k+1}^{x} - \overline{\mathbf{x}}_{k+1}^{-}\right]^{\mathrm{T}}.
\end{aligned} \tag{4.7}$$

The prediction of the observation vector and the innovation matrix, $\mathbf{P}_{k+1}^{vv}$, is done the same way. That means, the observation and the innovation are predicted from mean and covariance of the transformed sigma points:

$$\overline{\mathbf{y}}_{k+1}^{-} = \sum_{i=0}^{2n_{a}} W_{i}^{(m)}\ \mathrm{y}_{i,k+1}, \tag{4.8}$$

where $\mathrm{y}_{i,k+1}$ represents the sigma vectors propagated through the nonlinear equation of the observation model, yielding the transformed sigma points from the state vector and the dynamics noise, shown before.

In the up to date (correction) step of measurement, the Kalman gain, $\mathcal{K}_{k+1}$, is calculated based on the correlation matrix between the measurement and the observation, $\mathbf{P}_{k+1}^{xy}$, and the innovation matrix, both predicted

$$\mathcal{K}_{k+1} = \mathbf{P}_{k+1}^{xy}\left(\mathbf{P}_{k+1}^{vv}\right)^{-1}, \text{with} \quad \begin{aligned}
\mathbf{P}_{k+1}^{vv} &= \sum_{i=0}^{2n_{a}} W_{i}^{(\mathrm{c})}\left[\mathrm{y}_{i,k+1} - \overline{\mathbf{y}}_{k+1}^{-}\right]\left[\mathrm{y}_{i,k+1} - \overline{\mathbf{y}}_{k+1}^{-}\right]^{\mathrm{T}}, \\
\mathbf{P}_{k+1}^{xy} &= \sum_{i=0}^{2\,n_{a}} W_{i}^{(\mathrm{c})}\left[\mathcal{X}_{i,k+1}^{x} - \overline{\mathbf{x}}_{k+1}^{-}\right]\left[\mathrm{y}_{i,k+1} - \overline{\mathbf{y}}_{k+1}^{-}\right]^{\mathrm{T}}.
\end{aligned} \tag{4.9}$$

Finally, the up to date state and covariance are

$$\begin{aligned}\overline{\mathbf{x}}_{k+1}^{+} &= \overline{\mathbf{x}}_{k+1}^{-} + \mathcal{K}_{k+1}\left(\mathbf{y}_{k+1} - \overline{\mathbf{y}}_{k+1}^{-}\right),\\ \mathbf{P}_{k+1}^{+} &= \mathbf{P}_{k+1}^{-} - \mathcal{K}_{k+1}\mathbf{P}_{k+1}^{vv}\mathcal{K}_{k+1}^{\mathrm{T}},\end{aligned} \tag{4.10}$$

where $\mathbf{y}$ is the vector effectively measured in the instant $k+1$.

The process is repeated for the next instant, and the up to date mean (from state $\overline{\mathbf{x}}_{k+1}^{+}$) and covariance will be used to deterministically generate the sigma points of the next instant.

5. Final Comments

The SPKF estimation technique has been investigated in very many different applications. After analyzing the investigations, some comments may be done.

(i) The main advantages of the SPKF are: easy to implement; computationally strong; high accuracy.

(ii) The motivation for applying such technique herein: recent nonlinear state estimate techniques will be applied to the specific problem of orbit determination using real data from GPS, instead of simulated data. This can improve the results, when compared to the EKF, and get better the estimated state variables accuracy.

(iii) alman filter may be used as the estimation algorithm. However, not rarely, the dynamic systems and the measurements equations are of nonlinear nature. For solving such a problem, convenient extensions of the Kalman filter have been sought. The EKF implementation in orbit estimation, under inaccurate initial conditions and scattered measurements, can lead to unstable solutions. Nevertheless, the unscented transformation was developed as a method to propagate mean and covariance information through nonlinear transformations. The SPKF appears as an emerging estimation algorithm applied to nonlinear system, without linearization steps.

Acknowledgment

The authors wish to express their appreciation for the support provided by FAPESP (The State of São Paulo Research Foundation), under Contract 07/53256-1.

References

[1] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1985.

[2] R. D. Van Merwe, E. A. Wan, and S. I. Julier, "Sigma-point kalman filters for nonlinear estimation and sensor-fusion—applications to integrated navigation," in *Collection of Technical Papers—AIAA Guidance, Navigation, and Control Conference*, vol. 3, pp. 1735–1764, Providence, RI, USA, August 2004.

[3] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Signal Processing, Sensor Fusion, and Target Recognition VI*, vol. 3068 of *Proceedings of SPIE*, pp. 182–193, Orlando, Fla, USA, April 1997.

[4] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "New approach for filtering nonlinear systems," in *Proceedings of the American Control Conference*, vol. 3, pp. 1628–1632, Seattle, Wash, USA, June 1995.
[5] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, 2000.
[6] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
[7] D.-J. Lee and K. T. Alfriend, "Precise real-time orbit estimation using the unscented Kalman filter," *Advances in the Astronautical Sciences*, vol. 114, supplement, pp. 1835–1854, 2003.
[8] D.-J. Lee and K. T. Alfriend, "Adaptive sigma point filtering for state and parameter estimation," in *Collection of Technical Papers—AIAA/AAS Astrodynamics Specialist Conference*, vol. 2, pp. 897–916, Providence, RI, USA, August 2004.
[9] M. T. Soyka and M. A. Davis, "Estimation of periodic accelerations to improve orbit ephemeris accuracy," *Advances in the Astronautical Sciences*, vol. 108, pp. 1123–1140, 2001.
[10] D.-J. Lee and K. T. Alfriend, "Sigma point filtering for sequential orbit estimation and prediction," *Journal of Spacecraft and Rockets*, vol. 44, no. 2, pp. 388–398, 2007.
[11] D.-J. Jwo and C.-N. Lai, "Unscented Kalman filter with nonlinear dynamic process modeling for GPS navigation," *GPS Solutions*, vol. 12, no. 4, pp. 249–260, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 537256, 11 pages
doi:10.1155/2009/537256

Research Article

Orbital Dynamics of a Simple Solar Photon Thruster

Anna D. Guerman,[1] Georgi V. Smirnov,[2] and Maria Cecilia Pereira[3]

[1] *University of Beira Interior, 6201-001 Covilhã, Portugal*
[2] *University of Porto, 4169-007 Porto, Portugal*
[3] *Aeronautical Institute of Technology, 12.228-900 Sao José dos Campos, Brazil*

Correspondence should be addressed to Anna D. Guerman, anna@ubi.pt

Received 29 July 2009; Accepted 13 November 2009

Recommended by Maria Zanardi

We study orbital dynamics of a compound solar sail, namely, a Simple Solar Photon Thruster and compare its behavior to that of a common version of sailcraft. To perform this analysis, development of a mathematical model for force created by light reflection on all sailcraft elements is essential. We deduce the equations of sailcraft's motion and compare performance of two schemes of solar propulsion for two test time-optimal control problems of trajectory transfer.

1. Introduction

The use of solar pressure to create propulsion can minimize spacecraft on-board energy consumption during a mission [1, 2]. Modern materials and technologies made this propulsion scheme feasible, and many projects of solar sails are now under development, making solar sail dynamics the subject of numerous studies.

So far, the most extensively studied problem is the orbital maneuver of a Flat Solar Sail (FSS, Figure 1). In this case, the control is performed by turning the entire sail surface with respect to the Sun direction. This changes the radiation pressure and results in evolution of the vehicle trajectory. Some of the many missions studied are described in [3–13].

The use of a compound solar sail, or Solar Photon Thruster (SPT), was proposed by Tsander long ago [1, 2], but the study of this spacecraft began quite recently [14–18]. The SPT consists of a parabolic surface which concentrates the solar radiation pressure on a system of smaller mirrors. The control effort in such system is produced by displacement of a small mirror with respect to the parabolic surface. The sail axis is supposed to be oriented along

the Sun-sailcraft direction. There exist several versions of compound solar sails. Forward [14] described two types of a compound sail, namely, Simple Solar Photon Thruster (SSPT) and Dual Reflection Solar Photon Thruster (DRSPT). The few existent studies on SPT dynamics consider the latter scheme.

In order to assess the dynamical characteristics and to compare the SPT performance to the most studied version, FSS, one should study the application of these propulsion schemes for orbital transfer and/or maintenance for various missions. To perform this analysis, a coherent mathematical model for force acting on such a structure due to solar radiation pressure is essential. There are many results concerning radiation pressure force and torque models for a sunlit body, and in many studies they are applied successively to develop a force model for an FSS. However, the usual approach cannot be applied for a Solar Photon Thruster due to multiple light reflections on the SPT elements.

Some attempts to develop a mathematical description for SPT force have been made before, mostly for a Dual Reflection Solar Photon Thruster. In [2] the model for an ideally reflecting DRSPT is described. In [19] this model is extended for nonideal DRSPT. Meanwhile, both of these models are based on the supposition that all the incoming light flux is reflected consequently on each one of the DRSPT elements and then leaves the system, which is not true [20]. Moreover, the results of [20] show that the existing shadowing and related energy dissipation diminish significantly the DRSPT efficiency, making dubious the advantages of this propulsion scheme compared to FSS.

In the present article, we focus on the other version of the compound scheme of solar propulsion, Simple Solar Photon Thruster (Figure 2). We develop a mathematical model for a solar radiation force acting on SSPT and provide a comparative study of trajectory dynamics and control for the FSS and SSPT schemes. In our analysis, we use the following assumptions.

(i) Solar radiation pressure follows inverse-square variation law.

(ii) The only gravitational field is that of the Sun, and this field is central Newtonian.

(iii) The sails are ideal reflectors (all photons are perfectly reflected).

We derive the equations of motion for the SPT and compare the orbital behavior of FSS and SPT studying two test problems: Earth-Mars transfer and Earth-Venus transfer.

2. Equations of Motion

To write down the equations of motion for a solar sail spacecraft, we introduce two right-oriented Cartesian frames with their origin in the center of mass of the spacecraft O as follows:

(i) $Oxyz$ is the coordinate frame attached to the spacecraft; the axes Ox, Oy, and Oz are the central principal axes of the spacecraft.

(ii) $OXYZ$ is the orbital frame, its axis OZ is directed along the radius vector of the point O with respect to the center of mass of the Sun, and the axis OY is orthogonal both to OZ and to the velocity of the point O.

We determine the position of the coordinate frame $Oxyz$ attached to the spacecraft with respect to the orbital frame using the transition matrix between these frames, $\|a_{ij}\|$.

We use a set of canonical units which implies that the radius of the Earth's orbit is 1 AU, and the period of its revolution is 2π. The equations of orbital motion can be written in the form

$$\ddot{\vec{r}} = -\frac{\vec{r}}{r^3} + \vec{a}_s, \tag{2.1}$$

where a_s is the acceleration due to the radiation pressure.

To complete the equations of motion, we have to calculate also the force produced by the Sun radiation pressure.

2.1. Flat Solar Sail

The interaction of the solar radiation flow with a flat perfectly reflecting surface has been studied earlier [1, 2]. By the symmetry of falling and reflected flows, the total solar radiation force is directed along the symmetry axis of the sail and produces no torque with respect to any point of this axis, including the sail's center of mass (Figure 1). This force can be expressed as [1, 2]

$$\vec{P} = -2\vec{n}\left(\vec{\sigma}, \vec{n}\right)^2 S\frac{\Phi}{r^2}, \tag{2.2}$$

where $\vec{n}$ is the normal to the sail surface and points to the Sun, $\vec{\sigma}$ is the unit vector of the parallel light flow (i.e., it opposes the Sun-sailcraft direction), its coordinates in the $Oxyz$ frame are $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z) = (a_{31}, a_{32}, a_{33})$, S is the total area of the sail, and $\Phi = 4.563 \cdot 10^{-6}\,\mathrm{N/m^2}$ is the nominal solar radiation pressure constant at 1 AU. The solar radiation force projections onto the spacecraft-connected and orbital coordinate frames are

$$\begin{gathered} P_x = P_y = 0, \qquad P_z = 2S\frac{\Phi}{r^2}a_{33}^2 \operatorname{sign} a_{33}, \\ P_X = 2S\ \frac{\Phi}{r^2}a_{13}a_{33}^2 \operatorname{sign} a_{33}, \qquad P_Y = 2S\frac{\Phi}{r^2}a_{32}a_{33}^2 \operatorname{sign} a_{33}, \qquad P_Z = 2S\frac{\Phi}{r^2}|a_{33}|a_{33}^2, \end{gathered} \tag{2.3}$$

respectively.

If $\vec{n}$ lies in the OXZ plane, then the components of the radiation force are

$$P_X = -S\frac{\Phi}{r^2}\sin 2\theta|\cos\theta|, \qquad P_Y = 0, \qquad P_Z = 2S\frac{\Phi}{r^2}\cos^2\theta|\cos\theta|, \tag{2.4}$$

where θ is the angle between the vector $\vec{n}$ and the OZ axis.

The above expressions are standard and appear in numerous studies of the propulsion effort of a Flat Solar Sail.

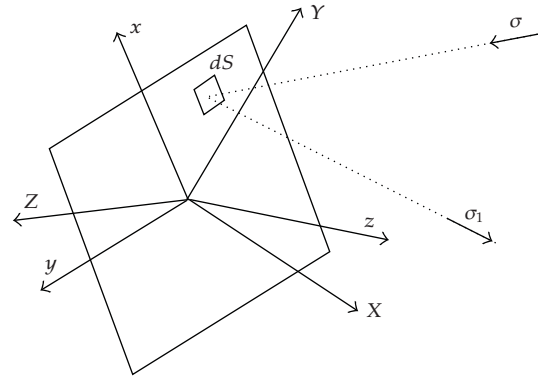


Figure 1: Flat Solar Sail (FSS).

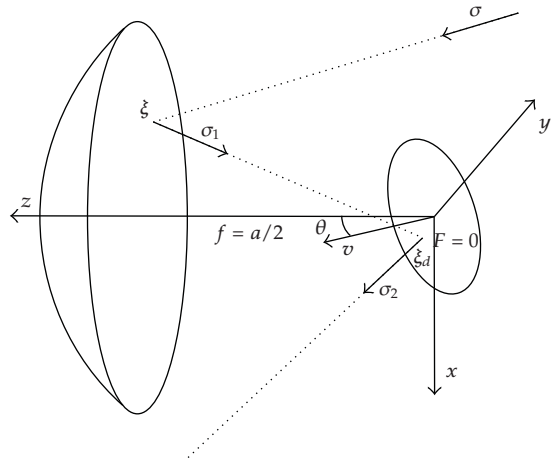


Figure 2: Solar Photon Thruster (SPT).

2.2. Solar Photon Thruster

We consider here another system that is shown in Figure 2. It consists of a parabolic collector and a control mirror (director). When reflection is ideal and the collector axis is exactly aligned with the Sun-sailcraft direction, the collector concentrates the sunlight in the center of the director. In order to minimize the solar radiation torque that causes perturbations of the sailcraft orientation, the director should be located at the sailcraft's center of mass. This scheme of solar propulsion seems to be more reliable with respect to small misalignments of the sail axis than the DRSPT scheme studied in [15–17] which uses a collimator.

In the analysis, we assume the control mirror small enough to disregard the influence of its shadow. We also suppose that the SPT axis is aligned exactly along the Sun direction.

We consider the parabolic surface described in the reference frame $Oxyz$ by the equation

$$x^2 + y^2 + 2a(z - f) = 0, \quad \left(x^2 + y^2 \leq R^2\right), \tag{2.5}$$

where a is the parameter of the paraboloid, f is the focal distance, and R is the radius of the sail's projection on the plane Oxy. Since $f = a/2$, equation (2.5) reduces to

$$z = \frac{1}{2a}\left(a^2 - x^2 - y^2\right). \tag{2.6}$$

The sunlight is directed along the vector $\vec{\sigma} = (0, 0, 1)$. Suppose that it is reflected on the element of the parabolic surface dS, containing the point $\vec{\xi} = x\vec{i} + y\vec{j} + z\vec{k}$ with z satisfying (2.6). The force produced by the falling light is given by

$$d\vec{P}_1 = -\rho\vec{\sigma}(\vec{\sigma}, \vec{n})dS, \tag{2.7}$$

where $\vec{n}$ is the normal to the sunlit side of the sail surface

$$\vec{n} = -\frac{x\vec{i} + y\vec{j} + a\vec{k}}{\sqrt{x^2 + y^2 + a^2}}, \tag{2.8}$$

and ρ is the intensity of the light flow at the current point of the orbit

$$\rho = \frac{\Phi}{r^2}. \tag{2.9}$$

The ray reflected from the element dS of the collector's surface has the direction $\vec{\sigma}_1$ satisfying

$$\vec{\sigma}_1 = \vec{\sigma} - 2(\vec{\sigma}, \vec{n})\vec{n}. \tag{2.10}$$

Reflection of light from the element of surface dS produces the force

$$d\vec{P}_2 = \rho\vec{\sigma}_1(\vec{\sigma}, \vec{n})dS. \tag{2.11}$$

Finally, this ray is reflected at the focus on the director's surface with the normal $\vec{\nu} = (\nu_x, \nu_y, \nu_z)$. The force produced by the reflected light can be written as

$$d\vec{P}_3 = -d\vec{P}_2, \qquad d\vec{P}_4 = \rho\vec{\sigma}_2(\vec{\sigma}, \vec{n})dS. \tag{2.12}$$

Here

$$\vec{\sigma}_2 = \vec{\sigma}_1 - 2(\vec{\sigma}_1, \vec{\nu})\,\vec{\nu} \tag{2.13}$$

is the direction of the ray reflected from the control mirror.

The reflection of the light on the parabolic surface is unique if the normal to the director does not cross this surface, so the control angle must be greater than half the angular aperture, that is,

$$\theta \geq \tan^{-1}\left(\frac{R}{z_R}\right). \tag{2.14}$$

Here θ is the angle between vector $\overrightarrow{\nu}$ and the sail axis ($\cos\theta = \nu_z$), and $z_R = (1/2a)(a^2 - R^2)$ is the z-coordinate of the collector's border. Finally we arrive at the restriction

$$|\tan\theta| \geq \frac{2aR}{a^2 - R^2}. \tag{2.15}$$

Multiple reflections on the collector destroy the collector's film and produce a considerable disturbing torque, and so have to be avoided. Therefore condition (2.15) has to be satisfied during the orbital maneuver.

The elementary force created by interaction of light with parabolic surface and mirror is

$$\begin{aligned} d\overrightarrow{P} &= d\overrightarrow{P}_1 + d\overrightarrow{P}_2 + d\overrightarrow{P}_3 + d\overrightarrow{P}_4 = d\overrightarrow{P}_1 + d\overrightarrow{P}_4 \\ &= -\rho\overrightarrow{\sigma}\ (\overrightarrow{\sigma}, \overrightarrow{n})dS + \rho\overrightarrow{\sigma}_2(\overrightarrow{\sigma}, \overrightarrow{n})d = \rho(\overrightarrow{\sigma}_2 - \overrightarrow{\sigma})(\overrightarrow{\sigma}, \overrightarrow{n})dS. \end{aligned} \tag{2.16}$$

After integration, we obtain

$$\begin{aligned} P_x &= 2\frac{\Phi}{r^2}\pi R^2 \nu_x \nu_z \left[1 - 2\frac{a^2}{R^2}\ln\left(1 + \frac{R^2}{a^2}\right)\right], \\ P_y &= 2\frac{\Phi}{r^2}\pi R^2 \nu_y \nu_z \left[1 - 2\frac{a^2}{R^2}\ln\left(1 + \frac{R^2}{a^2}\right)\right], \\ P_z &= 2\frac{\Phi}{r^2}\pi R^2 \left[\nu_z^2 + \left(1 - 2\nu_z^2\right)\ \frac{a^2}{R^2}\ln\left(1 + \frac{R^2}{a^2}\right)\right]. \end{aligned} \tag{2.17}$$

If the control mirror moves in the OXZ plane ($\nu_x = \sin\theta$, $\nu_y = 0$, $\nu_z = \cos\theta$), then the components of the light pressure force in the orbital coordinate frame are

$$\begin{aligned} P_X &= \frac{\Phi}{r^2}\pi R^2 \left[1 - 2\frac{a^2}{R^2}\ln\left(1 + \frac{R^2}{a^2}\right)\right]\ \sin 2\theta, \qquad P_Y = 0, \\ P_Z &= 2\frac{\Phi}{r^2}\pi R^2 \left[\cos^2\theta - \frac{a^2}{R^2}\ln\left(1 + \frac{R^2}{a^2}\right)\cos 2\theta\right]. \end{aligned} \tag{2.18}$$

If $\chi = R/a \ll 1$, then it is possible to simplify (2.18). One can use Taylor's formula to obtain the expressions

$$P_X = -\frac{\Phi}{r^2}\pi R^2\left[1 - \frac{R^2}{a^2} + o\left(\frac{R^3}{a^3}\right)\right]\sin 2\theta, \qquad P_Y = 0,$$
$$P_Z = 2\frac{\Phi}{r^2}\pi R^2\left[\sin^2\theta + \frac{R^2}{2a^2}\cos 2\theta + o\left(\frac{R^3}{a^3}\right)\right]. \tag{2.19}$$

If $\chi^2 = R^2/a^2$ is negligible (i.e., the sail is almost plane), we get

$$P_X = -\frac{\Phi}{r^2}S\sin 2\theta, \qquad P_Y = 0, \qquad P_Z = 2\frac{\Phi}{r^2}S\sin^2\theta, \tag{2.20}$$

where $S = \pi R^2$ is the effective sail area, that is, the area of the sail projection on the plane Oxy. Formulas (2.20) are similar to those used in [11–13] for a different scheme of SPT, so one can expect qualitative similarity of the results for small χ, at least for the maneuvers that require control angles within limits (2.15).

3. In-Plane Orbital Motion

To compare the principal characteristics of SPT and FSS we studied two test time-optimal control problems of solar sail dynamics, namely, the time-optimal Earth-Mars and Earth-Venus transfers [3, 21] for both systems. Since our goal is to compare qualitative behavior of the above systems, we choose the simplest formulation for orbital transfer problem. In both cases, we assume that the planet orbits are circular and coplanar and that the spacecraft moves in the ecliptic plane, starting from the Earth-orbit at 1 AU with Earth-orbital velocity. We find the control law that guarantees the fastest transfer to the planet's orbit.

This model of orbital dynamics results in the following equations of motion in the orbit plane [3]:

$$\dot r = u, \qquad \dot\varphi = \frac{w}{r}, \qquad \dot u = \frac{w^2}{r} - \frac{1}{r^2} + a_{sZ}, \qquad \dot w = -\frac{uw}{r} + a_{sX}. \tag{3.1}$$

Here φ is the polar angle, and u and w are the radial and transversal components of sail velocity, respectively.

For the FSS, the components of the light pressure acceleration onto the axis of orbital coordinate frame $OXYZ$ are

$$a_{sX} = \frac{\Phi}{mr^2}S|\cos\theta|\sin 2\theta, \qquad a_{sZ} = 2\frac{\Phi}{mr^2}S\left|\cos^3\theta\right|. \tag{3.2}$$

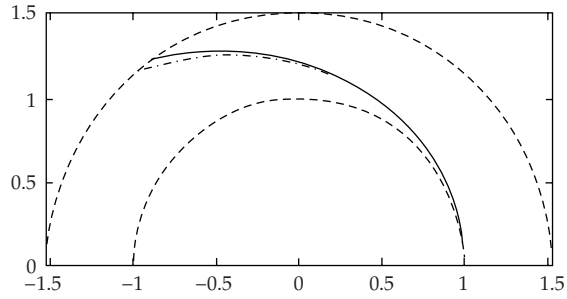


Figure 3: Earth-Mars transfer trajectories.

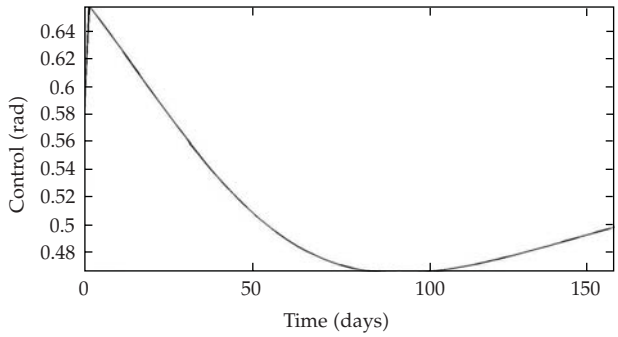


Figure 4: Earth-Mars transfer control effort for SSPT for $\chi = 0.125$.

For the SPT the light pressure acceleration is given by

$$a_{sX} = \frac{\Phi}{mr^2}\pi R^2\left[1 - 2\frac{a^2}{R^2}\ln\left(1 + \frac{R^2}{a^2}\right)\right]\sin 2\theta,$$
$$a_{sZ} = 2\frac{\Phi}{mr^2}\pi R^2\left[\cos^2\theta - \frac{a^2}{R^2}\ln\left(1 + \frac{R^2}{a^2}\right)\cos 2\theta\right]. \tag{3.3}$$

The control angle θ is limited by condition (2.15). In this case the sail surface has to follow the Sun direction.

4. Results

The time-optimal problems for Earth-Mars transfer and Earth-Venus transfer are studied numerically using the interactive software from [22]. This optimization software developed for personal computers running under MS Windows operating systems is based on the penalty function approach and offers to the user a possibility to effectively solve optimal control problems. During the interactive problem-solving process, the user can change the penalty coefficients, change the precision influencing the stopping rule, and choose/change the optimization algorithms. The system includes various gradient-free algorithms used at

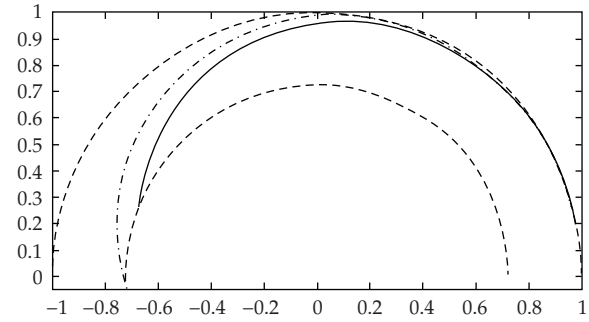


Figure 5: Earth-Venus transfer trajectories.

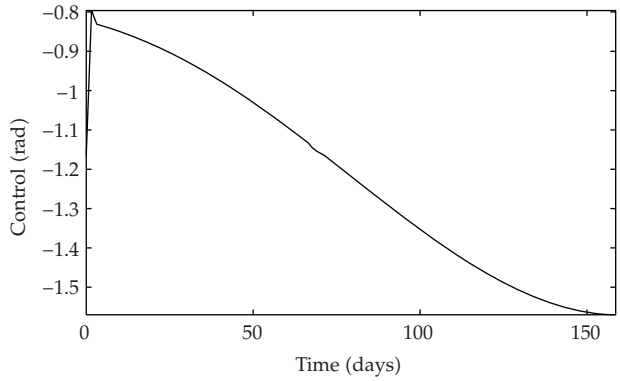


Figure 6: Earth-Venus transfer control effort for SSPT for $\chi = 0.125$.

the beginning of the optimization, as well as more precise conjugate gradient and Newton methods applied at the final stage in order to obtain a precise solution.

The sail parameter is assumed to be $\Phi S/m = 0.0843$ for both systems (it corresponds to the acceleration due to solar radiation pressure equal to 1 mm/s^{2}at the Earth-orbit).

The Earth-Mars transfer trajectories for the FSS and the SPT are shown in Figure 3. The continuous line corresponds to the SPT trajectory and the dot-dashed line to the FSS trajectory. We consider the ratio $\chi = R/a = 0.125$. The best possible transfer time for the FSS is $T^M_{\text{FSS}} = 2.87$ (166.7 days), and for the SPT it is $T^M_{\text{SPT}} = 2.71$ (157.5 days), so SSPT maneuver is 5.6% faster than that of the FSS one. Figure 4 shows the variation of the SSPT control angle for the optimal transfer; for FSS the respective control is well known [3].

For the Earth-Mars problem, we also study the influence of SPT sail ratio χ. The increase of χ results in longer maneuver time T_{SPT}: for $\chi = 0.25$ it is $T^M_{\text{SPT}} = 2.76$, and for $\chi = 0.5$ the maneuver time is $T^M_{\text{SPT}} = 2.96$. For greater values of χ, the control angle θ attains the limits described by restriction (2.15) more frequently.

Analyzing the maneuver to Venus orbit for these two sailcraft schemes (Figures 5 and 6), we have established that FSS reaches the objective in 181.2 days ($T^V_{\text{FSS}} = 3.12$), while the SPT performs this maneuver in 158.6 days ($T^V_{\text{SPT}} = 2.73$). In this case, the efficiency of SPT is more significant; SPT reaches Venus orbit 12.5% faster than FSS.

5. Conclusions

The problems of orbital dynamics and control are studied for two systems of solar propulsion: a Flat Solar Sail (FSS) and a Simple Solar Photon Thruster (SSPT). We develop a mathematical model for force acting on SSPT due to solar radiation pressure, taking into account multiple reflections of the light flux on the sailcraft elements. We derive the SSPT equations of motion. For in-plane motions of an almost flat sail with negligible attitude control errors, these equations are similar to those used in the previous studies of DRSPT.

For these two solar propulsion schemes, FSS and SSPT, we compare the best time response in two test problems (Earth-Mars transfer and Earth-Venus transfer). Our analysis showed a better performance of SSPT in terms of response time. The result was more pronounced for Earth-Venus transfer that can be explained by the greater values of the transversal component of the acceleration developed by SSPT compared to those of FSS.

Acknowledgments

The authors are grateful to Vladimir Bushenkov for fruitful discussions and his help in use of the optimization software [22]. This work was supported by FCT-Portuguese Foundation for Science and Technology (project CODIS-PTDC/CTE-SPA/64123/2006) and project CoDMoS (Portugal-Brazil collaboration program, FCT-CAPES).

References

[1] E. H. Polyahova, *Space Flight with Solar Sail*, Nauka, Moscow, Russia, 1986.
[2] C. R. McInnes, *Solar Sailing: Technology, Dynamics and Mission Applications*, Praxis Series, Springer, Berlin, Germany, 1999.
[3] B. Ya. Sapunkov, V. A. Egorov, and V. V. Sazonov, "Optimization of Earth-Mars solar sail spacecraft trajectories," *Cosmic Research*, vol. 30, no. 2, pp. 155–162, 1992.
[4] M. Otten and C. R. McInnes, "Near minimum-time trajectories for solar sails," *Journal of Guidance, Control, and Dynamics*, vol. 24, no. 3, pp. 632–634, 2001.
[5] G. Colasurdo and L. Casalino, "Optimal control law for interplanetary trajectories with nonideal solar sail," *Journal of Spacecraft and Rockets*, vol. 40, no. 2, pp. 260–265, 2003.
[6] D. N. Sharma and D. J. Scheeres, "Solar-system escape trajectories using solar sails," *Journal of Spacecraft and Rockets*, vol. 41, no. 4, pp. 684–687, 2004.
[7] B. Dachwald, "Minimum transfer times for nonperfectly reflecting solar sailcraft," *Journal of Spacecraft and Rockets*, vol. 41, no. 4, pp. 693–695, 2004.
[8] G. Mengali and A. A. Quarta, "Optimal three-dimensional interplanetary rendezvous using nonideal solar sail," *Journal of Guidance, Control, and Dynamics*, vol. 28, no. 1, pp. 173–177, 2005.
[9] B. Dachwald, "Optimal solar-sail trajectories for missions to the outer solar system," *Journal of Guidance, Control, and Dynamics*, vol. 28, no. 6, pp. 1187–1193, 2005.
[10] B. Dachwald, W. Seboldt, and L. Richter, "Multiple rendezvous and sample return missions to near-Earth objects using solar sailcraft," *Acta Astronautica*, vol. 59, no. 8–11, pp. 768–776, 2006.
[11] M. Macdonald, G. W. Hughes, C. McInnes, A. Lyngvi, P. Falkner, and A. Atzei, "GeoSail: an elegant solar sail demonstration mission," *Journal of Spacecraft and Rockets*, vol. 44, no. 4, pp. 784–796, 2007.
[12] B. Wie, "Thrust vector control analysis and design for solar-sail spacecraft," *Journal of Spacecraft and Rockets*, vol. 44, no. 3, pp. 545–557, 2007.
[13] J. Bookless and C. McInnes, "Control of Lagrange point orbits using solar sail propulsion," *Acta Astronautica*, vol. 62, no. 2-3, pp. 159–176, 2008.
[14] R. L. Forward, "Solar photon thrustor," *Journal of Spacecraft and Rockets*, vol. 27, no. 4, pp. 411–416, 1990.
[15] C. R. McInnes, "Payload mass fractions for minimum-time trajectories of flat and compound solar sails," *Journal of Guidance, Control, and Dynamics*, vol. 23, no. 6, pp. 1076–1078, 2000.

[16] G. Mengali and A. A. Quarta, "Earth escape by ideal sail and solar-photon thrustor spacecraft," *Journal of Guidance, Control, and Dynamics*, vol. 27, no. 6, pp. 1105–1108, 2004.
[17] G. Mengali and A. A. Quarta, "Time-optimal three-dimensional trajectories for solar photon thruster spacecraft," *Journal of Spacecraft and Rockets*, vol. 42, no. 2, pp. 379–381, 2005.
[18] B. Dachwald and P. Wurm, "Design concept and modeling of an advanced solar photon thruster," in *Proceedings of the 19th AAS/AIAA Space Flight Mechanics Meeting*, Savannah, Ga, USA, February 2009, paper no. AAS 09-147.
[19] G. Mengali and A. A. Quarta, "Compound solar sail with optical properties: models and performance," *Journal of Spacecraft and Rockets*, vol. 43, no. 1, pp. 239–245, 2006.
[20] A. D. Guerman and G. Smirnov, "Comment on "compound solar sail with optical properties: models and performance"," *Journal of Spacecraft and Rockets*, vol. 44, no. 3, pp. 732–734, 2007.
[21] P. A. Tychina, V. A. Egorov, and V. V. Sazonov, "Optimization of the flight of a spacecraft with a solar sail from Earth to Mars with a perturbation maneuver near Venus," *Cosmic Research*, vol. 40, no. 3, pp. 255–263, 2002.
[22] G. Smirnov and V. Bushenkov, *Curso de Optimização: Programação Matemática, Cálculo de Variações, Controlo Óptimo*, Escolar Editora, Lisbon, Portugal, 2005.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 303604, 12 pages
doi:10.1155/2009/303604

Research Article

Alternative Transfers to the NEOs 99942 Apophis, 1994 WR12, and 2007 UW1 via Derived Trajectories from Periodic Orbits of Family G

C. F. de Melo,[1] E. E. N. Macau,[2] and O. C. Winter[3]

[1] *UFABC, Universidade Federal do ABC, Santo André, 09210-170, SP, Brazil*
[2] *INPE, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 12227-010, SP, Brazil*
[3] *UNESP, Universidade Estadual Paulista, Guaratinguetá, 12516-410, SP, Brazil*

Correspondence should be addressed to C. F. de Melo, cristianofiorilo@terra.com.br

Received 30 July 2009; Accepted 11 December 2009

Recommended by Maria F. P. S. Zanardi

Swing-by techniques are extensively used in interplanetary missions to minimize fuel consumption and to raise payloads of spacecrafts. The effectiveness of this type of maneuver has been proven since the beginning of space exploration. According to this premise, we have explored the existence of a natural and direct link between low Earth orbits and the lunar sphere of influence to get low-energy transfer trajectories to the Near Earth Objects (NEOs) 99942 Apophis, 1994 WR12, and 2007 UW1 through swing-bys with the Moon. The existence of this link is related to a family of retrograde periodic orbits around the Lagrangian equilibrium point L1 predicted for the circular, planar, restricted three-body Earth-Moon-particle problem. The trajectories in this link are sensitive to small disturbances. This enables them to be conveniently diverted reducing so the cost of the swing-by maneuver. These maneuvers allow a gain in energy sufficient for the trajectories to escape from the Earth-Moon system and to stabilize in heliocentric orbits between the Earth and Venus or Earth and Mars. Therefore, the trajectories have sufficient reach to intercept the NEOs' orbits.

1. Introduction

The dynamics of the circular restricted three-body Earth-Moon-particle problem predicts the existence of the retrograde periodic orbits around the Lagrangian equilibrium point L1. Such orbits belong to the so-called Family G [1], and starting from them it is possible to define a set of trajectories that form a natural round trip link between the Earth and Moon and a link between the Earth, the Moon and the distant space of the Earth sphere of influence. These links occur even for more complex dynamical systems as the complete Sun-Earth-Moon-particle problem. In the last dynamical system, many kinds of transfers can be exploited.

For example, the round trip link is useful for transfers between Low Earth Orbits (LEOs) and Low Lunar Orbits (LLOs), including polar LLOs, since some of its trajectories pass a few hundreds of kilometers from the Earth's surface and a few dozens of kilometers from the Moon's surface (Figure 1) [2]. Moreover, the round trip link is also useful for transfer between two LEOs with different altitudes and inclinations [3]. In these cases the Moon's gravitational field, which is near the Earth, works as an extra impulse and provides an efficient method to minimize the fuel consumption to be used in plane change maneuvers. This is possible, because the trajectories of the links can lead spacecrafts in a natural way to accomplish swing-bys with the Moon. In this paper, we use the link between the Earth, the Moon and the distant space of the Earth sphere of influence and the energy gain of the swing-by with the Moon to get escape trajectories and to design missions to NEOs.

In general, a swing-by, or a flyby, consists of the alteration of small celestial bodies orbits (comets, asteroids, or spacecrafts) when they have a close approach with a planet or a moon. Literature presents many studies about this subject with general descriptions, applications in interplanetary missions and in the Earth-Moon System. For example, Lawden [4], Minovitch [5], and Broucke [6] present general descriptions of the mechanics of gravity-assisted maneuvers. Successful examples in interplanetary missions were the Voyager 1 and 2 which accomplished several swing-bys with the visited planets to gain energy [7], and Ulysses solar probe which accomplished a swing-by with Jupiter to get a perpendicular orbit to the ecliptic and observe the Sun's Poles [8]. Among the many studies on swing-bys in the Earth-Moon system, we can highlight Dunham and Davis [9], who designed missions with multiple swing-bys with the Moon; Uphoff [10], who gave a description about lunar gravity-assisted maneuvers, and Prado [11], who presented a study to use flyby with the Moon to accomplish transfers between Earth orbits with the same altitudes and different inclinations. Among the missions that accomplished swing-bys with the Moon, we should mention the ISEE-3 which after completing its mission in 1982 was renamed ICE (International Cometary Explorer) and accomplished several lunar encounters to gain energy and intercept the comets Giacobini-Zinner in September 1985 and Halley, in March 1986 [12]. On the other hand, also in general, the literature does not present works focused on getting the trajectory that will accomplish the swing-by. In this work we do not only use the swing-by to obtain transfer trajectories, but we have included a study that involves the use of trajectories derived from periodic orbits to accomplish the swing-by in an efficient way. This allows the reduction of the swing-by maneuver costs.

Our results, considering trajectories derived from periodic orbits of Family G and the quasiperiodic orbits that oscillate around them, have allowed us to present the parameters to design Earth-NEOs transfers. The trajectories studied in this work can offer a saving up to 4% in relation to the conventional methods used to send spacecrafts to these asteroids.

This article is laid out according to the following order: in Section 2, the dynamical systems and the orbits of Family G are described; in Section 3, the set of trajectories of interest are defined. In Section 4, the studies about the mechanism of the proposed transfers between Earth and NEOs are shown and, in Section 5, we present and discuss the results. The conclusions on the work are presented in Section 6.

2. Earth-Moon Link

The equations of motion of a small particle of negligible mass, m, moving under the gravitational influence of two bodies with preponderant masses, written in a normalized

rotating coordinate system (x, y), also-called synodic system, are

$$\ddot{x} - 2n\dot{y} = \frac{\partial U}{\partial x}, \qquad \ddot{y} + 2n\dot{x} = \frac{\partial U}{\partial y}, \tag{2.1}$$

where n is the mean motion and U is a scalar function given by

$$U = \frac{n^2}{2}\left(x^2 + y^2\right) + \frac{\mu_1}{r_1} + \frac{\mu_2}{r_2}, \tag{2.2}$$

μ_1 and μ_2 are the reduced masses of the two preponderant bodies, called primaries. $r_1^2 = (x + \mu_2)^2 + y^2$ and $r_2^2 = (x - \mu_1)^2 + y^2$ are the distances between the primaries and m. Equation (2.1) define the well-known restricted three-body problem in which the primaries have circular orbits around their common centre of mass [1, 13].

Equation (2.1) presents a special number of solutions that can be achieved searching for points in synodic frame where $\ddot{x} = \ddot{y} = \dot{x} = \dot{y} = 0$. There are five points with these features and they are called of Lagrangiam equilibrium points, L_i, $i = 1, \ldots, 5$. The locations of these points are obtained solving the simultaneous nonlinear equations $\partial U/\partial x = \partial U/\partial y = 0$. Figure 1 shows the locations of these points for Earth-Moon-particle problem. In this case, $\mu_1 = \mu_{\text{Earth}} = 0.9878494$, $\mu_2 = \mu_{\text{Moon}} = 0.0121506$, and $n = 1$.

In the Earth-Moon-particle problem, groups of periodic orbits around the Earth, the Moon and the L_i points are known. We have considered a particular family of periodic orbits around L1 known as Family G [1–3]. In the normalized synodic frame, one of the groups of initial conditions that allow us to obtain the Family G orbits has the form

$$\left(x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0\right) = \left(x_0, 0, 0, 0, \dot{y}_0, 0\right), \tag{2.3}$$

where $(-\mu_{\text{Moon}} + R_E^*) < x_0 < x(\text{L1})$, with $R_E^* = R_E/384{,}400\,\text{km} = 0.016572$, $R_E = 6{,}370\,\text{km}$ is the average radius of the Earth, $x(\text{L1}) = 0.836893$ is the abscissa of the Lagrangian equilibrium point L1, and $-9.389476 \leq \dot{y}_0 \leq 601.045381$ [1], which corresponds to a variation between $-9.607\,\text{km/s}$ and $614.964\,\text{km/s}$. The average Earth-Moon distance (384,400 km) is chosen as the unit of length of the normalized system. According to (2.3), for any orbit of Family G, in $t = 0$, the Earth, the particle and the Moon are aligned in this order, that is, in inferior conjunction. All orbits in Family G are unstable and this feature is very important for the transfers here exploited. Figure 1 also shows a typical orbit of Family G.

Periodic orbits as the one shown in Figure 1 and the quasiperiodic orbits that oscillate around them are tangents to the Earth and Moon low orbits. In other words, they define a link between the Earth and the lunar sphere of influence. This property allied with the instability of Family G orbits allows the design of controlled swing-bys with the Moon to gain sufficient energy to escape from the Earth-Moon system [2]. Moreover, we remark that these properties allow the reduction of the swing-by maneuver costs besides providing the necessary energy to win the Earth's and the Moon's gravitational fields [3, 14]. Controlled swing-bys also allow the use of this energy to obtain trajectories inclination changes [3].

In order to exploit these properties and to design more efficient and realistic Earth-NEOs transfers, the Sun's gravitational field must be introduced. That is, new equations of motion must be considered, this time, taking into account the Sun, the Earth and the Moon

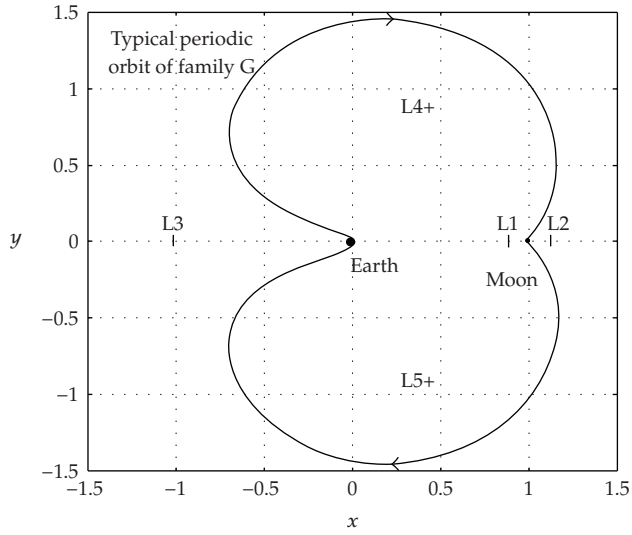


Figure 1: Locations of Lagrangian equilibrium points and a typical periodic orbit of Family G in synodic frame (x, y).

mutual gravitational attractions, besides the characteristics of the Earth's orbit around the Sun (eccentricity) and the Moon's orbit around the Earth (eccentricity and inclination). Thus, we define the full four-body Sun-Earth-Moon-particle problem, whose 12 equations of motion in a Cartesian coordinates system (X, Y, Z) with origin in a fixed point of the space are

$$\ddot{\mathbf{R}}_i = \sum_{\substack{j=1 \\ j \neq i}}^{4} \frac{\mu_j}{R_{ji}^3} \left(\mathbf{R}_j - \mathbf{R}_i\right), \tag{2.4}$$

where $R_{ij} = |\mathbf{R}_j - \mathbf{R}_i| = \left[(X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2\right]^{1/2}$, with $j \neq i$, are the distances between ith and jth bodies. The eccentricity of the Earth's orbit and the eccentricity and inclination of the Moon's orbit are introduced through initial conditions.

The link between the Earth and the lunar sphere of influence continues existing, even when the dynamical system given by (2.4) is considered. This can be seen in Figure 2(a) that shows a trajectory found considering the full four-body problem. Figure 2(b) shows the inclination gain relative to the ecliptic after the swing-by with the Moon. Note that the trajectory inclination change is about 52 degrees. As a first conclusion, we can say that the Moon gravitational field acts like a natural propeller, without fuel consumption. In the next section, we are going to exploit this property in a controlled way. That is, taking advantage of the existing natural link between the Earth and the lunar sphere of influence to obtain a change in inclination and the energy gain needed for a determined sort of mission.

3. Escape Trajectories

The existence of the link between the Earth and the lunar sphere of influence for the full four-body problem is essential to the design of the Earth-Moon system escape trajectories.

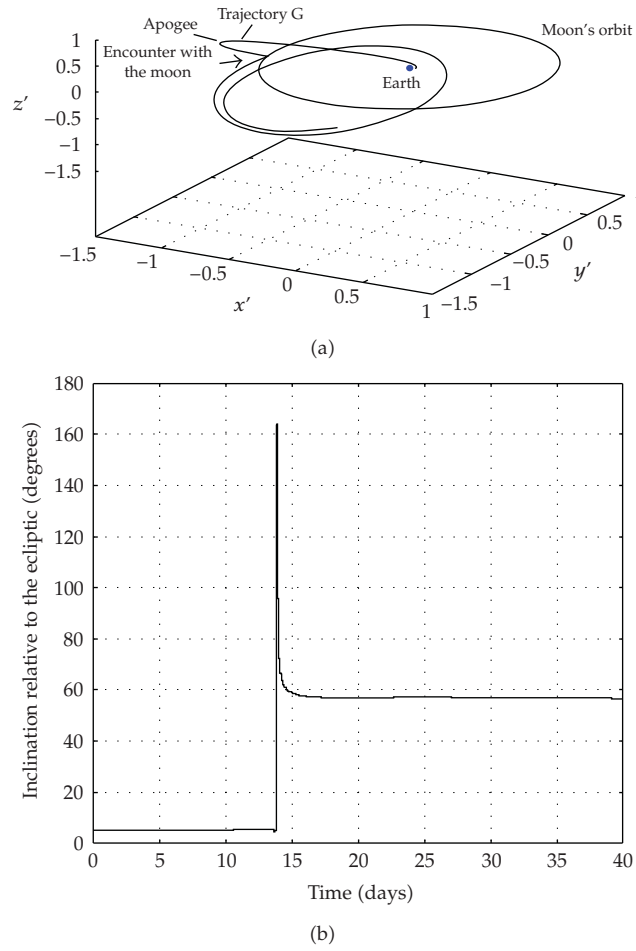


Figure 2: (a) Derived trajectory of a Family G periodic orbit considering the full four-body problem in geocentric frame (x', y', z'). (b) Trajectory inclination change as function of the time.

For this dynamical system, the initial conditions of an escape trajectory in a geocentric frame (x', y', z') are [3]

$$\left(x_0', y_0', z_0', \dot{x}_0', \dot{y}_0', \dot{z}_0'\right) = \left(x_0', 0, z_0', 0, \dot{y}_0', 0\right), \tag{3.1}$$

where

$$\begin{aligned} x_0' &= (R_E + h_0)\cos(i_{\text{Moon}}), \\ y_0' &= 0, \\ z_0' &= (R_E + h_0)\sin(i_{\text{Moon}}), \\ \dot{x}_0' &= 0, \\ \dot{y}_0' &= V_I, \\ \dot{z}_0' &= 0. \end{aligned} \tag{3.2}$$

Quantity V_I is called injection velocity. Considering a circular LEO, $V_I = V_{\text{CE}} + \Delta V_1$, where V_{CE} is the LEO's velocity, $V_{\text{CE}} = [GM_{\text{Earth}}/(R_E + h_0)]^{1/2}$, G is the gravitational constant, M_{Earth} is the Earth's mass, and h_0 is the LEO's altitude. i_{Moon} is the inclination of the Moon's orbit. ΔV_1 is the velocity increment that must be provided to the spacecraft. ΔV_1 is a function of the h_0 and V_{CE} and its value is given by

$$\Delta V_1 = \frac{V_{CE}}{2}\left(-2.3340 \times 10^{-6} h_0 + 0.8085 \pm 0.0001\right) + \delta + \vartheta. \tag{3.3}$$

Quantity δ assumes values that will take into account the relative position among the Sun, the Earth, and the Moon. In other words, the values assumed by δ reflect the influence of the eccentricities of the Earth's orbit and, mainly, of the Moon's orbit in the distance between them. Equations to estimate the values assumed by δ as a function of the true anomalies of the Earth's and Moon's orbits are given by de Melo et al. [3].

Here, we are interested in Earth-Moon escape trajectories. Then, we can consider the Earth and the Moon in circular orbits, with the average radius of the Moon's orbit equal to 384,400 km. This does not represent a significant loss of precision. This way, we can assume that $\delta = 0$ in (3.3) and, then, we just analyze the quantity θ whose values will determine the spacecraft position relative to the Moon and the Earth during the passage by the lunar sphere of influence and the swing-by. For $-1.50 \times 10^{-4}\,\text{km/s} \leq \vartheta \leq 1.50 \times 10^{-4}\,\text{km/s}$, we will have collision trajectories with the Moon, for $\vartheta < -1.50 \times 10^{-4}\,\text{km/s}$, the trajectories will have periselenium in the region close to the Moon, between itself and the Earth (anterior region). For $\vartheta > 1.50 \times 10^{-4}\,\text{km/s}$, the periselenium will be in the posterior region to the Moon relative to the Earth. For $\vartheta < -1.50 \times 10^{-4}\,\text{km/s}$, it is possible to design transfers between LEOs and LLOs of any altitudes and also between two Earth orbits of different inclinations and altitudes, both at a low cost [2, 3]. Our interest, however, is the trajectories generated for $\vartheta > 1.50 \times 10^{-4}\,\text{km/s}$. In those cases, the gain of energy with the swing-by guarantees the necessary velocity increment to win the Earth's and the Moon's gravitational fields.

The spacecraft must reach the velocity $\mathbf{V}_I$ when the Earth, the spacecraft and the Moon are aligned in this order. This is necessary because the trajectory is derived from a Family G orbit, and this is the only situation in which the orbits of Family G are tangent to the circular LEOs. Note that this condition does not impose significant restrictions in practical purpose. For instance, if a satellite is in a LEO with an altitude of 200 km, this condition will occur every 1.48 hour, approximately, which is the period of the LEO. Figure 3 shows a typical escape trajectory derived from Family G orbits. From now on, we will call them of escape trajectories G.

It is important to note that the trajectories evolution of the Earth-Moon and escape links shown here (Figures 2 and 3) do not suffer any intervention since they start from the LEO. That is, the inclination and the energy gain occur naturally during the passage through the lunar sphere of influence. However, it is possible to take advantage of the inherent instability of these trajectories to control the spacecraft passage through the lunar sphere of influence [2, 15]. This idea will be exploited in the next section to generate specific escape trajectories for determined missions.

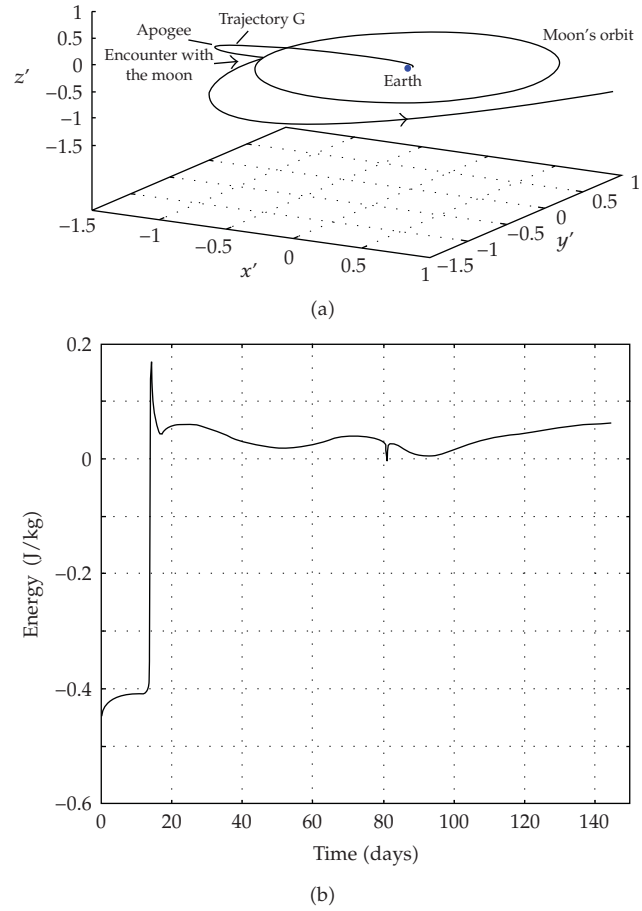


Figure 3: (a) Typical escape trajectory from the Earth-Moon system. (b) Energy gain relative to the Earth as function of the time for the trajectory. Note that it is a typical swingy-by energy gain.

4. Escape Trajectories to Design Earth-NEOs Transfers

The basic idea is to find transfer trajectories between the Earth and the NEOs starting from escape trajectories derived from Family G periodic orbits. This will be done to three NEOs: the 99942 Apophis, the 1994 WR12, and the 2007 UW1. All these are NEOs of the Aten asteroids class and they are defined by having semimajor axes of less than one astronomical unit (1 AU is equal to 149.6×10^6 km). Some data about these objects are provided in Tables 1 and 2. Table 1 shows the orbital elements: semimajor axis (a), eccentricity (e), inclination (i), longitude of ascending node (Ω), argument of perihelion (w), and the orbital period (P). Table 2 brings other information as average orbital velocity (V_{ave}), Escape velocity (V_{esc}), Aphelion, Perihelion, Mass, and diameter. In Table 3, the three next Closest Point Approaches (CPA) with the Earth are presented for 99942 Apophis, the 1994 WR12, and the 2007 UW1.

Once the mission target is determined, it is necessary to find the point in the asteroid's orbit where the spacecraft will reach it. However, our goal here is not to find an optimal trajectory exactly. We intend to show that transfer trajectories derived from periodic orbits around L1, generated after swing-bys with the Moon, can reach an NEO with ΔVs smaller

Table 1: NEOs orbital elements.

Asteroid	a (UA)	e	i** (deg)	Ω (deg)	w (deg)	P (yr)
Apophis	0.992	0.191	3.331	304,5	126,7	0.89
1994WR12	0.757	0.397	6.864	62,85	205,9	0.66
2007UW1	0.907	0.121	8.224	26,04	146,5	0.86

**Inclination relative to the ecliptic. Source: http://neo.jpl.nasa.gov/.

Table 2: Some characteristics of the NEOs orbits.

Asteroid	V_{ave} (km/s)	V_{esc} (m/s)	Aphelion (UA)	Perihelion (UA)	Mass (kg)	Diameter (m)
Apophis	30.728	0.144	1.099	0.746	2.7×10^{10}	270
1994WR12	34.706	0.050	1.058	0.455	2.0×10^{9}	110–260
2007UW1	31.509	0.047	1.017	0.798	1.3×10^{9}	80–190

Source: http://neo.jpl.nasa.gov/.

than the ones required by trajectories that leave from the Earth directly. In order to do so, we only considered trajectories that are tangents to the NEOs's orbits. In order to find an escape trajectory that will reach an NEO's orbit, for example, an algorithm integrates a number of them starting from the interception point backward in time until they reach the closest point approach with the Moon during the swing-bys. These integrations are supplied for initial conditions determined starting from previous numerical simulations of trajectories that escape naturally from the Earth-Moon system, as the one shown in Figure 3. In these simulations, a group of trajectories that have a point closer to the NEO's orbit is selected. The positions and the velocities of these trajectories are analyzed and adjusted to form a group of initial conditions for the integrations beginning in the interception point of the NEO's orbit. When the closest point approach with the Moon is found, the algorithm calculates a small $\Delta\mathbf{V}$ to determine which escape trajectory G is more suitable for the transfer. In other words, it calculates the components values of the spacecraft's initial condition in the LEO (3.1). The algorithm provides a group of solutions with ΔV_{Total}, ΔV_1, and $\Delta V_{\text{intermediary}}$, transfer time and other quantities of interest. Besides the $\Delta\mathbf{V}_{\text{intermediary}}$ applied in the closest point approach with the Moon, the algorithm also calculates another $\Delta\mathbf{V}_{\text{intermediary}}$ applied in the first apogee of trajectory G. This apogee is reached about eight days after the departure from the LEO and prior to the passage by the lunar sphere of influence (Figure 3). This is the best point to do velocity and inclination changes. In this point, the trajectories G velocities are very small; so small $\Delta\mathbf{V}$s can provide significant velocity and inclination changes.

A similar procedure is used to find the transfer trajectory between the Earth and the NEO for Patched-conic approximation. In this method, the transfers are designed seeking for a heliocentric conic arc that links an LEO to an orbit around the asteroid [16].

5. Results

The results of the intercept missions to the NEOs Apophis, 1999 WR12, and 2007 UW1 are in Tables 4 and 5. Table 4 presents the values found for the missions conceived by escape trajectories G, and Table 5 presents the correspondent values found for the missions conceived by Patched-conic approach. For each sort of mission, the ΔVs applied in the departure from the LEO, for occasional velocity and inclination changes, the spacecraft velocities relative to the asteroids in the closest point approach (CPA), and the transfer times

Table 3: Three next closest point approaches.

	Date	Apophis	Asteroid 1994WR12	2007UW1
1st	Date	2013/01/09.48850	2017/11/22.55300	2020/04/24.60600
	Minimum possible distance (UA)	0.096662	0.001781	0.137462
2nd	Date	2029/03/06.05209	2019/11/24.23939	2026/10/07.03002
	Minimum possible distance (UA)	0.112651	0.001731	0.002622
3rd	Date	2036/04/14.60516	2021/12/05.05848	2039/06/0.832120 2039/10/22.06710
	Minimum possible distance (UA)	0.000254	0.001540	0.099407 0.001451

Sources: http://neo.jpl.nasa.gov/.

Table 4: Missions' data of the Earth-NEOs transfer by escape trajectories G.

Asteroid	ΔV_1^* (km/s)	$\Delta V_{\text{int}1}^{**}$ (km/s)	$\Delta V_{\text{int}2}^{***}$ (km/s)	ΔV_{Total} (km/s)	$T1$ (days)	V_1^{****} (km/s)	$T2$ (days)	V_2^{*****} (km/s)
Apophis	3.148	0.010	0.018	3.176	81	0.800	225	0.550
1994WR12	3.147	0.018	—	3.165	53	3.700	103	3.600
2007UW1	3.149	0.018	0.012	3.169	153	0.490	—	—

*Departure, **1st apogee, ***CPA with the Moon, ****spacecraft velocity relative to the asteroid in $T1$, and *****spacecraft velocity relative to the asteroid in $T2$.

Table 5: Missions' data of the Earth-NEOs transfer by Patched-conic approach.

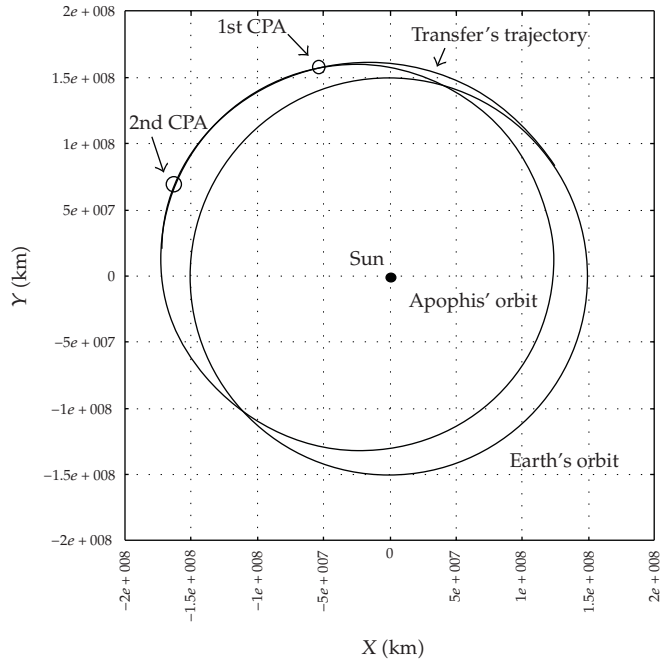
Asteroid	ΔV_1^* (km/s)	$\Delta V_{\text{int}1}^{**}$ (km/s)	ΔV_{Total} (km/s)	$T1$ (days)	V_1^{***} (km/s)	$T2$ (days)	V_2^{****} (km/s)
Apophis	3.280	0.020	3.308	155	1.200	198	1.350
1994WR12	3.225	0.035	3.260	131	4.100	—	—
2007UW1	3.250	0.028	3.278	111	0.690	—	—

*Departure, **middle curse, ***spacecraft velocity relative to the asteroid in $T1$, and ****spacecraft velocity relative to the asteroid in $T2$.

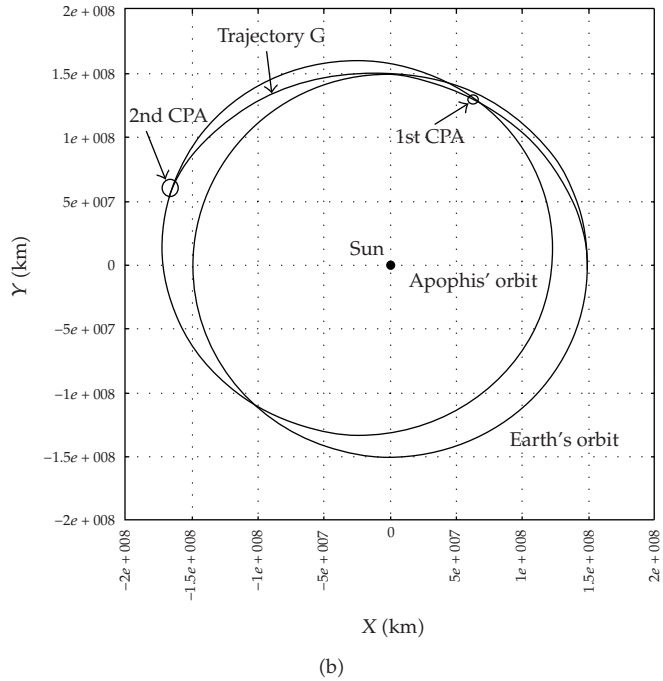
were calculated. The altitude of the circular departure LEO was considered equal to 200 km, and the altitude of the CPA with the asteroids was considered equal to 1 km. For the transfers between the Earth and Apophis and the Earth and 1994 WR12, it was possible to find two CPAs between the spacecraft and the asteroids. In Tables 4 and 5, we call these times of T1 for the 1st CPA and T2 for the 2nd CPA. It is interesting to note that spacecraft velocities relative to the asteroids are different in each CPA, and, depending on the mission goals, the choice of one of them could represent an extra fuel saving in an occasional insertion maneuver of the spacecraft into an orbit around the asteroid.

Figure 4(a) shows the Earth's and the Apophis's orbits and the heliocentric ellipse arc generated by the Patched-conic approach. The times of the 1st and 2nd CPA are 155 and 198 days after the departure from the LEO, respectively, and the distances from the Earth to these points are 11.2×10^6 km and 113.6×10^6 km, respectively. The values of $T1$ and $T2$ of the other transfers are in Tables 4 and 5.

By analogy, Figure 4(b) shows the escape trajectory G found for the Earth-Apophis transfer. The 1st and the 2nd CPAs take place 86 and 225 days after the departure from the LEO, and the distances from the Earth to these points are 52.27×10^6 km and 204.45×10^6 km, respectively.



(a)



(b)

Figure 4: (a) Earth's and Apophis's orbits and the transfer trajectory found for Patched-conic transfer. (b) Earth's and Apophis's orbits and transfer trajectory G (heliocentric frame).

The analysis of Tables 4 and 5 shows that the procedure described in Section 4 to generate transfer trajectories between the Earth and NEOs, derived from periodic orbits of Family G, requires ΔVs of 2% up to 4% less than the transfers conceived by Patched-conic approach. Besides, the spacecraft velocities relative to the asteroids in the CPAs, for the trajectories G, are the smallest. This can also provide fuel saving in occasional insertion maneuvers into an orbit around the asteroids. However, the smallest relative velocities always correspond to the longest transfer times.

6. Conclusions

In this work, a procedure capable of generating transfer trajectories between the Earth and the NEOs has been presented. These trajectories are derived from the periodic orbits around the Lagrangian point L1 and escape from the Earth-Moon system after a swing-by with the Moon.

In terms of ΔV_{Total}, those required by escape trajectories G are, in general, fewer than the ones required by conventional transfers (Patched-conic), between 2% up to 4%. Besides, the spacecraft velocities relative to the asteroids are also, in general, less than those found by the conventional methods.

With regard to the transfer time, we verify that in two cases (Apophis and 1994 WR12) it was possible to find two CPAs. The time longest always corresponds to the smallest relative velocity in CPA for trajectories G. We also verify that there are not any discrepancies between the transfer times of the two considered methods.

This way, we can conclude that the escape trajectories G presented in this work are a good alternative to design future missions destined to the NEOs.

Finally, future studies of techniques to accomplish a swing-by with the Earth and, thus, to gain more energy and a larger reach could provide the planning of transfers to more distant objects, including the internal planets.

Acknowledgments

The authors are grateful to the FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), processes 2005/05169-7 and 2006/04997-6, and the CNPq (Conselho Nacional para Desenvolvimento Científico e Tecnológico) Brazil, for the financial support.

References

[1] R. A. Broucke, "Periodic orbits in the restricted three-body problem with Earth-Moon masses," Tech. Rep. 32-1168, JPL, Los Angeles, Calif, USA, 1968.

[2] C. F. de Melo, E. E. N. Macau, O. C. Winter, and E. V. Neto, "Alternative paths for insertion of probes into high inclination lunar orbits," *Advances in Space Research*, vol. 40, no. 1, pp. 58–68, 2007.

[3] C. F. de Melo, E. E. N. Macau, and O. C. Winter, "Strategies for plane change of Earth orbits using lunar gravity and derived trajectories of family *G*," *Celestial Mechanics & Dynamical Astronomy*, vol. 103, no. 4, pp. 281–299, 2009.

[4] D. Lawden, "Perturbation maneuvers," *Journal of the British Interplanetary Society*, vol. 13, no. 5, 1954.

[5] M. A. Minovitch, "A method for determining interplanetary free-fall reconnaissance trajectories," Technical Memo TM-312-130, JPL, 1961.

[6] R. A. Broucke, "The celestial mechanics of gravity assist," in *Proceedings of AIAA/AAS Astrodynamics Conference*, Minneapolis, Minn, USA, August 1988.

[7] C. E. Kohlhase and P. A. Penzo, "Voyager mission description," *Space Science Reviews*, vol. 21, no. 2, pp. 77–101, 1977.
[8] R. Carvell, "Ulysses-the Sun from above and below," *Space*, vol. 1, pp. 18–55, 1985.
[9] D. Dunham and S. Davis, "Optimization of a multiple lunar swing-by trajectory sequence," *Journal of the Astronautical Sciences*, vol. 33, no. 3, pp. 275–288, 1985.
[10] C. Uphoff, "The art and science of lunar gravity assist, orbital mechanics and mission design," *Advances in the Astronautical Sciences*, vol. 69, pp. 333–346, 1989.
[11] A. F. B. A. Prado, "Orbital control of a satellite using the gravity of the moon," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 28, no. 1, pp. 105–110, 2006.
[12] R. Farquhar, "Detour a Comet, Journey of the International Cometary Explorer," *Planetary Report*, vol. 5, no. 3, pp. 4–6, 1985.
[13] C. D. Murray and S. F. Dermott, *Solar System Dynamics*, Cambridge University Press, Cambridge, Mass, USA, 1999.
[14] K. Deb, N. Padhye, and G. Neema, "Interplanetary trajectory optimization with swing-bys using evolutionary multi-objective optimization," in *Proceedings of the 2nd International Symposium on Intelligence Computation and Applications (ISICA '07)*, L. Kang, Y. Liu, and S. Zeng, Eds., vol. 4683 of *Lecture Notes in Computer Science*, pp. 26–35, Wuhan, China, September 2007.
[15] E. E. N. Macau and C. Grebogi, "Control of chaos and its relevancy to spacecraft steering," *Philosophical Transactions of the Royal Society of London. Series A*, vol. 364, no. 1846, pp. 2463–2481, 2006.
[16] V. A. Chobotov, *Orbital Mechanics*, AIAA, 1991.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 159287, 10 pages
doi:10.1155/2009/159287

Research Article

Controlling the Eccentricity of Polar Lunar Orbits with Low-Thrust Propulsion

O. C. Winter,[1,2] D. C. Mourão,[1] C. F. de Melo,[3] E. N. Macau,[2] J. L. Ferreira,[4] and J. P. S. Carvalho[1]

[1] *Grupo de Dinâmica Orbital & Planetologia, UNESP, Universidade Estadual Paulista, Avenida Ariberto Pereira da Cunha, 333, CEP, 12516-410, Guaratinguetá, SP, Brazil*
[2] *Instituto Nacional de Pesquisas Espaciais—INPE, Avenida dos Astronautas 1752, CEP, 12227-010, São José dos Campos-SP, Brazil*
[3] *Universidade Federal do ABC, UFAB, Santo André, 09210-170, SP, Brazil*
[4] *Instituto de Física, Universidade de Brasília—UnB, CP 04455, CEP, 70919-970, Brasília, DF, Brazil*

Correspondence should be addressed to O. C. Winter, ocwinter@feg.unesp.br

Received 30 July 2009; Accepted 31 December 2009

Recommended by A. F. B. De Almeida Prado

It is well known that lunar satellites in polar orbits suffer a high increase on the eccentricity due to the gravitational perturbation of the Earth. The final fate of such satellites is the collision with the Moon. Therefore, the control of the orbital eccentricity leads to the control of the satellite's lifetime. In the present work we study this problem and introduce an approach in order to keep the orbital eccentricity of the satellite at low values. The whole work was made considering two systems: the 3-body problem, Moon-Earth-satellite, and the 4-body problem, Moon-Earth-Sun-satellite. First, we simulated the systems considering a satellite with initial eccentricity equals to 0.0001 and a range of initial altitudes between 100 km and 5000 km. In such simulations we followed the evolution of the satellite's eccentricity. We also obtained an empirical expression for the length of time needed to occur the collision with the Moon as a function of the initial altitude. The results found for the 3-body model were not significantly different from those found for the 4-body model. Secondly, using low-thrust propulsion, we introduced a correction of the eccentricity every time it reached the value 0.05.

1. Introduction

Recently, several nations presented plans to reach the Moon. Satellites have been launched and many more are planned for following years (see, e.g., [1]). The expectations are that in the near future there will be a lunar base. The lunar poles are particularly of interest since it seems to be where water can be found. Therefore, long living satellites in polar lunar orbits will be needed. It is well known that lunar satellites in polar orbits suffer a strong gravitational perturbation from the Earth. That effect is a natural consequence of the Lidov-Kozai resonance.

It is well known that the Lidov-Kozai resonance introduces equilibrium configurations. In the case of lunar polar orbits disturbed by the Earth's gravitational field, this can be used as an advantage to implement constellations of satellites with elliptic highly inclined orbits [2, 3]. On the other hand it causes instability for near circular highly inclined orbits. Wytrzyszczak et al. [4] studied the regular and chaotic motion of geosynchronous satellites disturbed by the Moon's gravitational field. They found that the chaotic nature of high inclination satellites is caused due to the significant eccentricity growth caused by the Lidov-Kozai resonance.

Similarly, the final fate of polar lunar near circular satellites is the collision with the Moon. Therefore, the control of the orbital eccentricity leads to the control of the satellite's lifetime.

In this paper we propose the control of the eccentricity using an electrical thruster, similar to the one that is in development at the University of Brasília. Electric propulsion is basically a technique of space propulsion which involves the conversion of electrical power into the kinetic power or thrust of the exhaust beam of ionized particles. The ability to obtain high exhaust velocities with ionized particles enables plasma thrusters to perform high specific impulse mission in space [5]. The main innovation of the thruster that is being developed at the University of Brasília is the use of a permanent magnet, saving energy during the mission. Preliminary results in the laboratory show that it is possible to obtain more than 100 mN with this technology. Inspired on this thruster project, in this work, we assume a constant exhaust velocity, and we can control the switch of the thruster during the mission.

In the present work we introduce an approach in order to keep the orbital eccentricity of lunar polar satellites at low values. The approach is based on the use of low-thrust propulsion in order to introduce a correction of the eccentricity.

In the next section we introduce the Lidov-Kozai resonance. In Section 3 we show the evolution of the eccentricities form our numerical integrations. The approach proposed to control the eccentricity and its application is presented in Section 3. In the final section we present our final comments.

2. The Lidov-Kozai Resonance

Lidov [6], studying the dynamics of artificial satellites, and Kozai [7], studying the dynamics of asteroids, independently discovered what is now called the Lidov-Kozai resonance. Following, we introduce the basic features of such resonance.

In this section we adopted a simple model (see, e.g., [2]) for the orbital evolution of an artificial satellite disturbed by a third body in circular equatorial orbit around the primary. It was obtained by double averaging the system [8] taking into account the disturbing function expanded in Legendre polynomials up to second order and the eccentricity of the disturbing body also up to the second order. The disturbing function of the problem is averaged independently over the mean longitudes of the satellite and the third body. The standard definition for average used in this work is

$$\langle F \rangle = \frac{1}{2\pi} \int_0^{2\pi} (F) dM, \tag{2.1}$$

where M is the mean anomaly, which is proportional to the time.

Following such approach one can find the double averaged disturbing function given by (see, e.g., [9])

$$R = \frac{3G(m_1+m_2)a^2}{16a_E^2}\left(2\left(e^2-\sin^2 i\right)+e^2(5\cos 2\omega-3)\sin^2 i\right), \tag{2.2}$$

where a, e, ω, and i are, respectively, the semimajor axis, eccentricity, argument of pericenter and inclination, G is the gravitational constant, a_E is the semimajor axis of the Earth with respect to the Moon, and m_1 and m_2 are the masses of the Earth and Moon respectively. Substituting R in Lagrange's planetary equations (see, e.g., [10]), we find

$$\frac{da}{dt} = 0, \tag{2.3}$$

$$\frac{de}{dt} = \frac{15e\gamma}{16n}\sqrt{1-e^2}n_E^2\left(2+3e_E^2\right)\sin(2\omega)\sin^2 i, \tag{2.4}$$

$$\frac{dw}{dt} = -\frac{3\gamma\left(2+3e_E^2\right)n_E^2}{16n\sqrt{1-e^2}}\left[1-5\cos^2 i-e^2+5\cos 2\omega\left(\cos^2 i+e^2\right)-1\right] \tag{2.5}$$

where n is the mean motion and $\gamma = m_1/(m_1+m_{12})$.

Considering the case when $de/dt = 0$ and $d\omega/dt = 0$ one can find three first integrals:

$$a = a_0, \tag{2.6}$$

$$\left(1-e^2\right)\cos^2 i = k_1, \tag{2.7}$$

$$e^2\left(\frac{2}{5}-\sin^2 i\sin^2 w\right) = k_2, \tag{2.8}$$

where a, e, i, and w are the semi-major axis, eccentricity, inclination, and argument of pericenter of the satellite, and a_0, k_1, and k_2 are the constants of motion. This system has a set of fixed points given by

$$w = 90^\circ \text{ or } 270^\circ, \qquad e^2+\left(\frac{5}{3}\right)\cos^2 i = 1. \tag{2.9}$$

Therefore, for a system with e, i, and ωsatisfying conditions (2.10), the satellite would be in what can be called a frozen orbit; that is, apart from short-period oscillations, the orbit would be kept fixed in size and location.

A simple analysis of (2.2) and (2.3) shows that [9]

(i) for $k_2 > 0$ and any value of k_1, w circulates,

(ii) for $k_2 < 0$ and $k_1 < 3/5, w$ librates around 90° or 270°,

(iii) for $k_2 = 0$ and $w = 90^\circ$ or 270°, $i = i^* \sim 39.2^\circ$,

where $i^* = 39.2^\circ$ is the critical value, which corresponds to the frozen orbit.

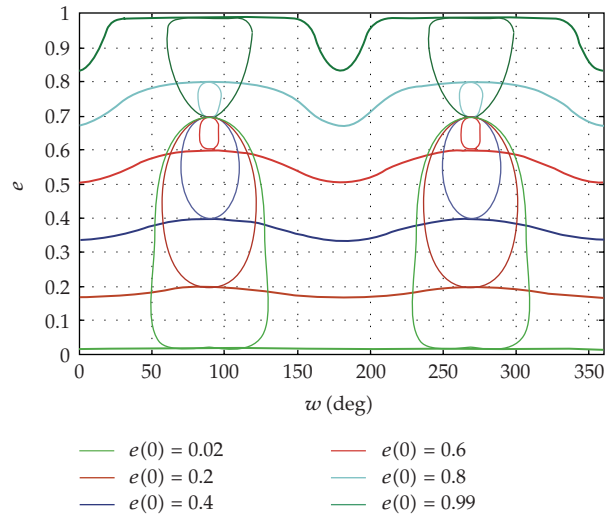


Figure 1: Sample of the satellite's orbital evolution in a diagram e versus w. There are two sets of initial values of eccentricity and inclination (e_o, i_o) : One for low inclination, $i_o = 20°$ (thick lines) and other for high inclination, $i_o = 56.2°$, (thin lines). The values of the initial eccentricities are represented by the colour code in the bottom of the figure. The argument of pericentre circulates in the case of initial low inclination, while librates in the case of initial high inclination.

So, that is the Lidov-Kozai resonance. When $i > i^*$, the system behaves like a pendulum, with stable fixed points, librations around such points and circulation.

Following another approach let us consider a satellite orbiting the Moon and disturbed by the Earth in an elliptical orbit with respect to the Moon. Taking into account only the term P_2 of the Legendre polynomial, the disturbing potential is given by

$$R = \frac{45}{32}a^2\mu_E\left[\left(e^2\sin^2 i\right)\cos 2w + \frac{3}{5}\left(\frac{2}{3}+e^2\right)\left(\cos^2 i - \frac{1}{3}\right)\right]n_E^2\left(\frac{2}{3}+e_E^2\right), \tag{2.10}$$

where e_E and n_E are the eccentricity, and mean motion of the Earth with respect to the Moon, and $\mu_E = m_E/(m_E + m_M)$, where m_E and m_M are the masses of the Earth and Moon, respectively.

One can identify the Lidov-Kozai resonant features by numerically integrating Lagrange's planetary equations for the temporal variation of the argument of pericentre, w, and the eccentricity, e, with the disturbing potential given by (2.10). In Figure 1 we present a sample of these numerical integrations in a diagram e versus w. There are two sets of initial values of eccentricity and inclination (e_o, i_o). One for low inclination $(i_o = 20°)$ and other for high inclination $(i_o = 56.2°)$. This figure shows a clear dependence of the eccentricity on the argument of pericentre for an orbit with high inclination. All satellites inclined to the orbital plane of the third body (the Earth) by more than 39.2°, the critical inclination, experience a considerable growth of eccentricity. The Earth causes the Lidov-Kozai resonance driving the eccentricity growth. For all trajectories from the set with initial inclination higher than the critical value, the argument of pericentre librates, while it circulates for trajectories from the set with initial inclination lower than the critical value.

3. Eccentricity Growth

In this section we present numerical integrations considering two dynamical systems: the 3-body problem, Moon-Earth-satellite, and the 4-body problem, Moon-Earth-Sun-satellite. In all simulations the satellite is initially in polar orbit ($i = 90°$).

In the case of the 3-body problem, considering a coordinate system centered in the barycenter of the Earth-Moon system (X, Y, Z), the equations of motion of the satellite are given by

$$\begin{aligned}
\ddot{\vec{x}} &= \sum_{i=1}^{2} G\frac{m_i}{|x_i - x|^3}(\vec{x}_i - \vec{x}),\\
\ddot{\vec{y}} &= \sum_{i=1}^{2} G\frac{m_i}{|y_i - y|^3}(\vec{y}_i - \vec{y}),\\
\ddot{\vec{z}} &= \sum_{i=1}^{2} G\frac{m_i}{|z_i - z|^3}(\vec{z}_i - \vec{z}),
\end{aligned} \tag{3.1}$$

where m is mass and the index $i = 1$ refers to the Earth and $i = 2$ refers to the Moon. In this system, the equations of motion for the moon and the Earth are given by

$$\begin{aligned}
\ddot{\vec{x}}_i &= \sum_{j=1,\, j\neq i}^{2} G\frac{m_i}{|x_j - x_i|^3}(\vec{x}_j - \vec{x}_i),\\
\ddot{\vec{y}}_i &= \sum_{j=1,\, j\neq i}^{2} G\frac{m_i}{|y_j - y_i|^3}\left(\vec{y}_j - \vec{y}_i\right),\\
\ddot{\vec{z}}_i &= \sum_{j=1,\, j\neq i}^{2} G\frac{m_i}{|z_j - z_i|^3}(\vec{z}_j - \vec{z}_i).
\end{aligned} \tag{3.2}$$

First, we simulated the system, integrating numerically (3.1) and (3.2), considering a satellite with initial eccentricity equals to 0.0001 and a range of initial altitudes between 100 km and 5000 km. Figure 2 shows the evolution of the satellite's eccentricity for the 3-body simulations, considering altitudes $h = 100$, 200, 500, 1000, and 5000 km. The plots show an exponential evolution of the eccentricity. We computed the time needed in order to reach the eccentricity that corresponds to the collision of the satellite with the Moon. A fit of the collision time, $T_{\text{collision}}$, in Earth days, as a function of the altitude, h, is given by the expression:

$$T_{\text{collision}} = 2693e^{0.062h - 3.92\times10^{-4}h^2}. \tag{3.3}$$

The same set of simulations was performed considering the 4-body problem, adding the perturbations of the Sun. Now we considered a new coordinate system $(X'Y'Z')$, centered at the barycenter of Sun-Earth-Moon system. In this case we integrate numerically the equations similar (using prime in the variables) to equations (3.1) and (3.2), but we add the index $i = 3$, where the fourth term refers to Sun.

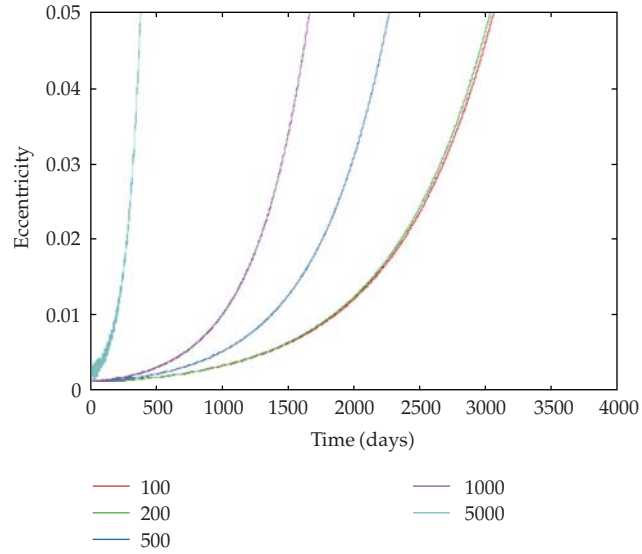


Figure 2: Time evolution of the eccentricity for the 3-body problem. The colour code indicates the initial altitude in kilometers. The time is in Earth days.

However, the results found for the 4-body model were not significantly different from those found for the 3-body model. The empirical expression for the length of time needed to occur the collision with the Moon as a function of the initial altitude is given by

$$T_{\text{collision}} = 2494e^{0.063h-3.94\times10^{-4}h^2}. \tag{3.4}$$

A comparison of the two sets of simulations and (2.5) and (2.6) is shown in Figure 3.

4. Controlling the Eccentricity

In order to control the satellite's eccentricity we will use low-thrust propulsion. Following the work of Sukhanov [11], we use the locally optimal thrust for each orbital element. This development is based on the performance index, through the minimization of a functional in the direction of the orbital element to be changed. In our case, the eccentricity is the parameter to be minimized. The result is a vector, called Lawden's primer vector, P, which gives the direction of the thruster to be turned on.

The eccentricity of the satellite relative to the Moon is given by

$$e = \sqrt{1 + \frac{c^2}{Gm_2}h}, \tag{4.1}$$

where c is the magnitude of the angular momentum, and h is the integral energy. The primer vector is given by

$$\overrightarrow{p} = \frac{1}{Gm_2 e}\left(P\overrightarrow{v} - \frac{r^2}{a}\overrightarrow{v}_n\right), \tag{4.2}$$

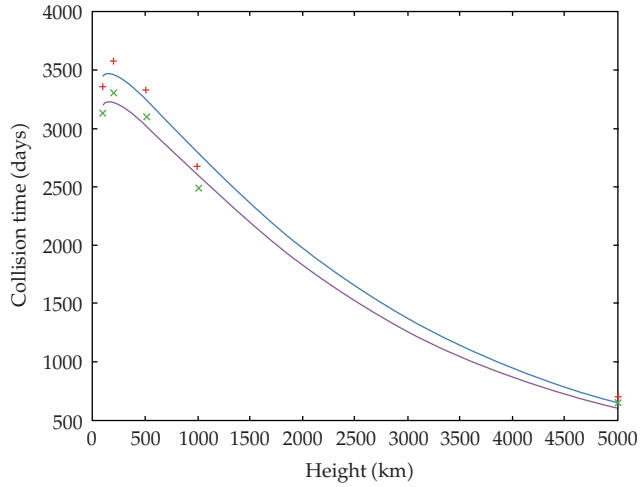


Figure 3: Collision time as a function of the initial altitude. The red crosses are for the 3-body simulations and green are for the 4-body simulations. The blue curve corresponds to (2.5) and the purple curve corresponds to (2.6).

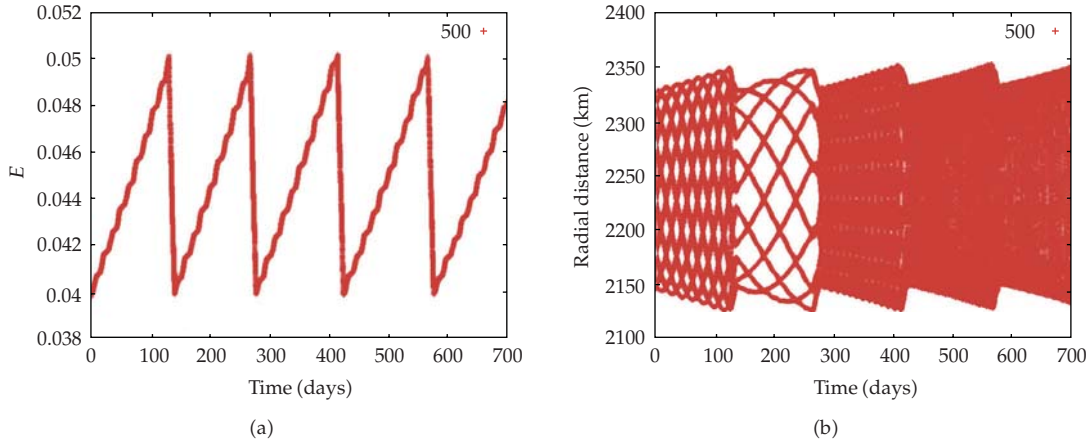


Figure 4: Temporal evolution of the eccentricity (a) and of the orbital radius (b). In this simulation the initial altitude was 500 km and the thrust value used was 0.2 N.

where $\vec{r}$, $\vec{v}$, and $\vec{v}_n$are the vector position, velocity, and tangential velocity relative to the Moon, and a and P are the semimajor axis and semilatus rectum relative to the Moon.

Then, we have that the acceleration components to change eccentricity are given by

$$p_r = \frac{1}{\mu e} p_a, \qquad p_n = \frac{1}{\mu e}\left(p_a - \frac{r_p^2}{a}\right), \tag{4.3}$$

where $p_a = c^2/\mu$, p_r, and p_n are the radial and the tangential components, respectively.

The approach we are proposing is a very simple one. The idea is to introduce a correction on the eccentricity every time it reaches a certain limit. The procedure is as follows.

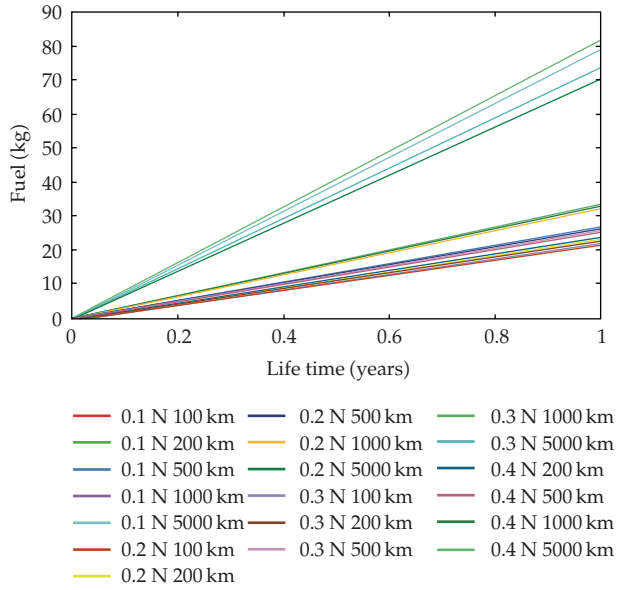


Figure 5: The propellant consumption per year of lifetime for the whole set of simulations, that is, different initial altitudes and different values of the thruster.

First fix the nominal eccentricity, e_o, and the maximum acceptable increase in eccentricity, Δe, according to the mission design. Then, turn on the thruster every time the condition

$$e > (e_o + \Delta e) \tag{4.4}$$

is satisfied and turn off the thruster when $e > e_o$.

Following, we present the results of some simulations assuming $e_o = 0.04$ and $\Delta e = 0.01$. These simulations were made considering a set of different thrust values, from 0.1 N up to 0.4 N. In each run we measured the length of time T_{Thruster}, needed to correct the eccentricity value (from $e = 0.04$ to $e = 0.05$). From these results we obtained empirical expressions of T_{Thruster} as a function of the initial altitude and as a function of the thrust value. As an example, in Figure 4 is shown the temporal evolution of the eccentricity and of the orbital radius for a satellite with an initial altitude of 500 km and using a thruster of 0.2 N.

In Figure 5 we present the propellant consumption per year of lifetime for the whole set of simulations, that is, different initial altitudes and different values of the thruster. The time intervals that the thrusters are turned on and off are shown in Figures 6 and 7.

5. Final Comments

In the present work we have studied the problem of polar lunar satellites in near circular orbits under the gravitational perturbations of the Earth and the Sun. The problem is dominated by the Lidov-Kozai resonance, which forces satellites with near circular orbits to have an exponential growth of its eccentricity. In order to keep the satellite with low eccentricity we proposed to use low-thrust propulsion every time the eccentricity reaches a limiting eccentricity, acceptable by the requirements of the satellite mission. The results show

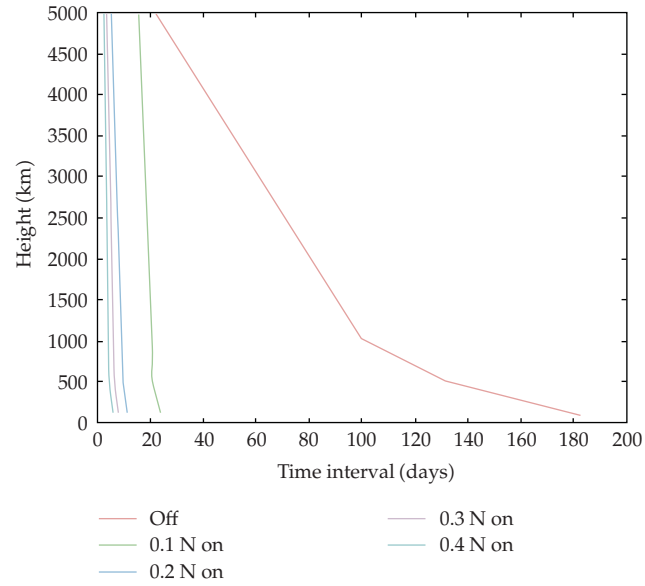


Figure 6: Time interval that the thrusters are turned on and off, for the whole set of simulations.

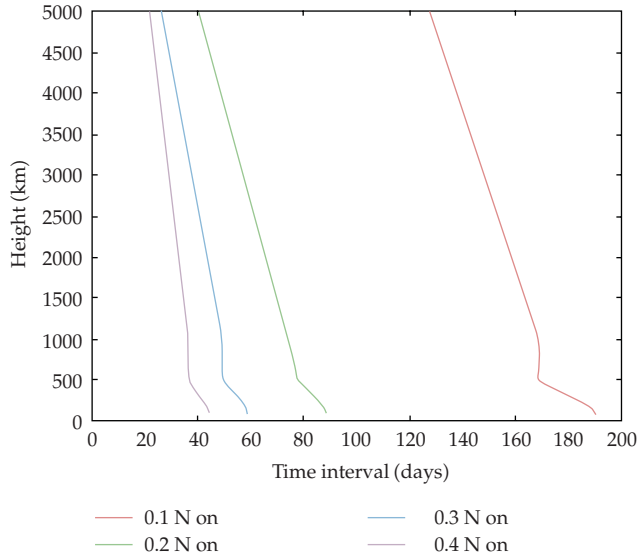


Figure 7: Zoom of Figure 6.

that the satellite's lifetime can be reasonably extended (several years) at a not so expensive cost. Therefore, it is shown that low-thrust propulsion is very adequate for this kind of purpose.

Acknowledgment

This research was supported by CNPq, Fapesp, and the Brazilian Space Agency.

References

[1] B. H. Foing and P. Ehrenfreund, "Journey to the Moon: recent results, science, future robotic and human exploration," *Advances in Space Research*, vol. 42, no. 2, pp. 235–237, 2008.
[2] T. A. Ely, "Stable constellations of frozen elliptical inclined lunar orbits," *Journal of the Astronautical Sciences*, vol. 53, no. 3, pp. 301–316, 2005.
[3] T. A. Ely and E. Lieb, "Constellations of elliptical inclined lunar orbits providing polar and global coverage," *Journal of the Astronautical Sciences*, vol. 54, no. 1, pp. 53–67, 2006.
[4] I. Wytrzyszczak, S. Breiter, and W. Borczyk, "Regular and chaotic motion of high altitude satellites," *Advances in Space Research*, vol. 40, no. 1, pp. 134–142, 2007.
[5] J. H. Souza, *Estudo da dinâmica de partículas em um propulsor a plasma do tipo Hall com ímãs permanentes*, M.S. thesis, Universidade de Brasília, Brasília, Brazil, 2006.
[6] M. L. Lidov, "Evolution of artificial planetary satellites under the action of gravitational perturbations due to external bodies," *Iskusstviennye Sputniki Zemli*, vol. 8, pp. 5–45, 1961 (Russian).
[7] Y. Kozai, "Secular perturbations of asteroids with high inclination and eccentricity," *Astrophysical Journal*, vol. 67, no. 9, pp. 591–598, 1962.
[8] A. F. B. A. Prado, "Third-body perturbation in orbits around natural satellites," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 1, pp. 33–40, 2003.
[9] M. A. Vashkov'yak and N. M. Teslenko, "Peculiarities of the orbital evolution of the Jovian satellite J34 (Euporie)," *Astronomy Letters*, vol. 33, no. 11, pp. 780–787, 2007.
[10] J. Kovalevsky, *Introduction to Celestial Mechanics*, Bureau dês Longitudes, Paris, France, 1967.
[11] A. A. Sukhanov, *Lectures on Astrodynamics*, INPE, São José dos Campos, Brazil, 5th edition, 2007.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 471804, 36 pages
doi:10.1155/2009/471804

Research Article

Internal Loading Distribution in Statically Loaded Ball Bearings Subjected to an Eccentric Thrust Load

Mário César Ricci

Space Mechanics and Control Division (DMC), Brazilian Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil

Correspondence should be addressed to Mário César Ricci, mariocesarricci@uol.com.br

Received 29 July 2009; Accepted 18 November 2009

Recommended by Antonio Prado

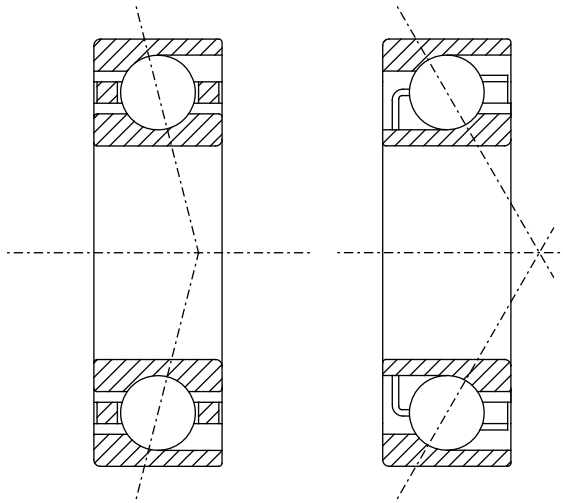
Rolling-element bearings are simple machine elements of great utility used both in simple commercial devices as in complex engineering mechanisms. Because of being a very popular machine element, there is a lot of literature on the subject. With regard to the behavior of internal loading distribution, elastic deformations at point or line contacts, and geometric parameters under loading, although there are many works describing the parameters variation models, few works show such variations in practice, even under simple static loadings. In an attempt to cover this gap some studies are being developed in parallel. Particularly in this work, a *new*, iterative computational procedure is introduced which calculates internal normal ball loads in statically loaded single-row, angular-contact ball bearings, subjected to a known thrust load which is applied to a variable distance (lever arm or eccentricity) from the geometric bearing center line. Numerical examples results for a 218 angular-contact ball bearing have been compared with those from the literature. Fifty figures are presented showing geometrical features and the following parameters variations as functions of the thrust load and eccentricity: contact angle, contact ellipse parameters, normal ball loads, distances between groove curvature centers, normal and axial deflections, and loading zones.

1. Introduction

Ball and roller bearings, generically called *rolling bearings*, are commonly used machine elements. They are employed to permit rotary motions of, or about, shafts in simple commercial devices such as bicycles, roller skates, and electric motors. They are also used in complex engineering mechanisms such as aircraft gas turbines, rolling mils, dental drills, gyroscopes, and power transmissions.

The standardized forms of ball or roller bearings permit rotary motion between two machine elements and always include a complement of ball or rollers that maintain the shaft and a usually stationary supporting structure, frequently called a *housing*, in a radially or axially spaced-apart relationship. Usually, a bearing may be obtained as a unit, which

Figure 1: An angular-contact ball bearing (courtesy of SKF Industries).



(a) Small angle (b) Large angle

Figure 2: Angular-contact ball bearings.

includes two steel rings; each of which has a hardened raceway on which hardened balls or rollers roll. The balls or rollers, also called *rolling elements*, are usually held in an angularly spaced relationship by a *cage*, also called a *separator* or *retainer*.

There are many different kinds of rolling bearings. This work is concerned with *single-row angular-contact ball bearings* (Figure 1) that are designed to support combined radial and thrust loads or heavy thrust loads depending on the contact angle magnitude. The bearings having large contact angle can support heavier thrust loads. Figure 2 shows bearings having small and large contact angles. The bearings generally have groove curvature radii in the range of 52%–53% of the ball diameter. The contact angle does not usually exceed 40^o.

This work is devoted to the study of the internal loading distribution in statically loaded ball bearings. Several researchers have studied the subject as, for example, Stribeck [1], Sjoväll [2], Jones [3], and Rumbarger [4], to cite a few. The methods developed by

them to calculate distribution of load among the balls and rollers of rolling bearings can be used in most bearing applications because rotational speeds are usually slow to moderate. Under these speed conditions, the effects of rolling-element centrifugal forces and gyroscopic moments are negligible. At high speeds of rotation these body forces become significant, tending to alter contact angles and clearance. Thus, they can affect the static load distribution to a great extension.

Harris [5] described methods for internal loading distribution in statically loaded bearings addressing pure radial, pure thrust (centric and eccentric loads), combined radial and thrust load, which uses radial and thrust integrals introduced by Sjoväll [2], and for ball bearings under combined radial, thrust, and moment load, initially due to Jones [3].

The method described by Harris for eccentric thrust load, initially due to Rumbarger [4], is an approximate, direct method, based in a single-row, 90^o thrust bearing and in thrust and moment integrals whose values are obtained from tables and graphics, as functions of eccentricity and pitch diameter. The maximum ball load is given directly and no computer is necessary. Although it is not entirely appropriate, the method was used by Harris to find approximations for the maximum ball load magnitude and for the extension of the loading zone in the 218 angular-contact ball bearing.

We can see that there are many works describing the parameters variation models under static loads but few show such variations in practice, even under simple static loadings. The author believes that the lack of practical examples in the literature is mainly due to the inherent difficulties of the numerical procedures that, in general, deal with the resolution of several nonlinear algebraic equations that must be solved simultaneously.

In an attempt to cover this gap studies are being developed in parallel [6–14]. Particularly in this work is described a new, precise method for internal load distribution computation in statically loaded, single-row, angular-contact ball bearings subjected to a known external thrust load which is applied to a variable distance (lever arm or eccentricity) from the geometric bearing center line. It must be solved iteratively using a digital computer and can be thought as a particular case of the Jones method, with null external radial load and external moment load given by the product of the thrust load by the eccentricity. Unlike Rumbarger's method, it is adequate to angular-contact bearings, and theoretically and numerically more precise. The novelty of the method is in the choice of the set of the nonlinear equations, which must be solved simultaneously. The author did not find in the literature the resolution of this problem using the same set of equations.

The difference between the method described here and the method described by Harris for eccentric thrust load mainly comes from the fact that Rumbarger's method, for sake of simplicity, makes use of the pitch radius, $d_e/2$, as lever arm, instead of the inner contact radius, $d_{cj}/2$, in the r.h.s. of the moment equation—see (4.19) for comparison— and secondarily by the fact that it uses the pitch radius instead of the locus of the centers of the inner ring raceway groove curvature radii, R_i, in the computations of the load distribution factor, ε, in (4.10) and of the extension of load zone, φ_l, in (4.11). These approximations are guarantee of the straightforwardness but obviously they introduce errors in the normal ball loads determination. However, at first glance appears that the method for thrust bearing is more attractive than the method of this paper because it supplies results more directly whereas no computer is necessary. But, despite the simplicity of the former, comparative analyses between the results show significant differences in the magnitudes of the maximum ball load and extension of the loading zone.

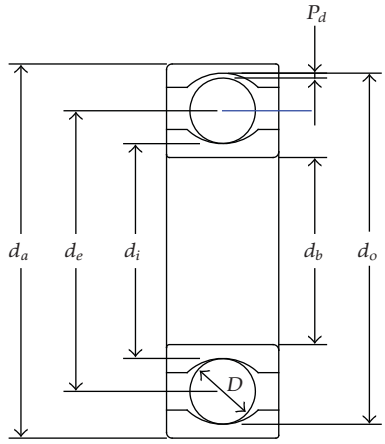


Figure 3: Radial cross-section of a single-row ball bearing.

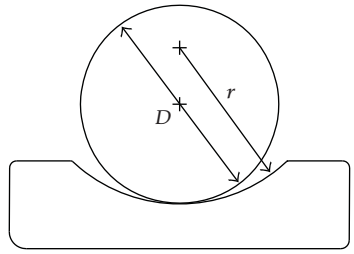


Figure 4: Cross-section of a ball and an outer race showing race conformity.

2. Geometry of Ball Bearings

In this section, the principal geometrical relationships for an unloaded ball bearing are summarized. The radial cross section of a single-row ball bearing shown in Figure 3 depicts the *diametral clearance* and various diameters. The *pitch diameter*, d_e, is the mean of the inner- and outer-race diameters d_i and d_o, respectively, and is given by

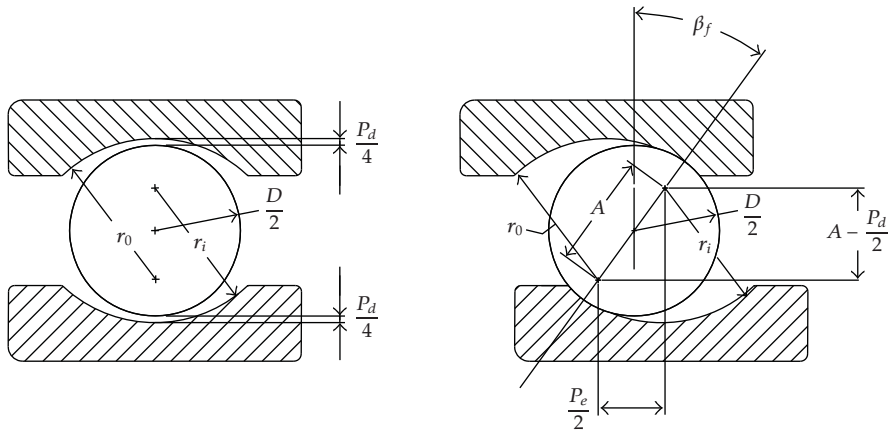
$$d_e = \frac{1}{2}(d_i + d_o). \tag{2.1}$$

The diametral clearance, P_d, can be written as

$$P_d = d_o - d_i - 2D. \tag{2.2}$$

Race conformity is a measure of the geometrical conformity of the race and the ball in a plane passing through the bearing axis (also named center line or rotation axis), which is a line passing through the center of the bearing perpendicular to its plane and transverse to the race. Figure 4 depicts a cross section of a ball bearing showing race conformity expressed as

$$f = \frac{r}{D}. \tag{2.3}$$



(a) Initial position

(b) Shifted position

Figure 5: Cross-section of a radial ball bearing showing ball-race contact due to axial shift of inner and outer rings.

Radial bearings have some axial play since they are generally designed to have a diametral clearance, as shown in Figures 5(a) and 5(b), that shows a radial bearing with contact due to the axial shift of the inner and outer rings when no measurable force is applied. The radial distances between the curvature centers of the two races are the same in Figures 5(a) and 5(b). Denoting quantities which referred to the inner and outer races by subscripts i and o, respectively, this radial distance value can be expressed as $A - P_d/2$, where $A = r_o + r_i - D$ is the curvature centers distance in the shifted position given by Figure 5(b). Using (2.3) we can write A as

$$A = BD, \tag{2.4}$$

where $B = f_o + f_i - 1$ is known as the *total conformity ratio* and is a measure of the combined conformity of both the outer and inner races to the ball.

The *contact angle*, β, is defined as the angle made by a line, which passes through the curvature centers of both the outer and inner raceways and that lies in a plane passing through the bearing rotation axis, with a plane perpendicular to the bearing axis of rotation. The *free-contact angle*, β_f, (Figure 5(b)) is the contact angle when the line also passes through the points of contact of the ball and both raceways and no measurable force is applied. From Figure 5(b), the expression for the free-contact angle can be written as

$$\cos\beta_f = \frac{A - P_d/2}{A}. \tag{2.5}$$

From (2.5), the diametral clearance, P_d, can be written as

$$P_d = 2A\left(1 - \cos\beta_f\right). \tag{2.6}$$

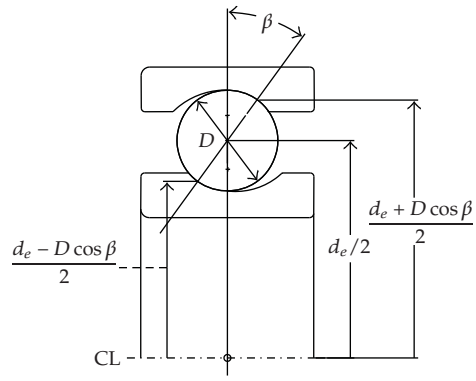


Figure 6: Cross-section of a ball bearing.

Free endplay, P_e, is the maximum axial movement of the inner race with respect to the outer when both races are coaxially centered and no measurable force is applied. Free endplay depends on total curvature and contact angle, as shown in Figure 5(b), and can be written as

$$P_e = 2A \sin \beta_f. \tag{2.7}$$

Considering the geometry of two contacting solids (ellipsoids) in a ball bearing, we can arrive at the two quantities of some importance in the analysis of contact stresses and deformations. The curvature sum, $1/R$, and curvature difference, Γ, are defined as

$$\begin{aligned} \frac{1}{R} &= \frac{1}{R_x} + \frac{1}{R_y}, \\ \Gamma &= R\left(\frac{1}{R_x} - \frac{1}{R_y}\right), \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} \frac{1}{R_x} &= \frac{1}{r_{ax}} + \frac{1}{r_{bx}}, \\ \frac{1}{R_y} &= \frac{1}{r_{ay}} + \frac{1}{r_{by}}, \end{aligned} \tag{2.9}$$

with r_{ax}, r_{bx}, r_{ay}, and r_{by} being the radii of curvature for the ball-race contact.

A cross section of a ball bearing operating at a contact angle β is shown in Figure 6. Equivalent radii of curvature for both inner- and outer-race contacts in, and normal to, the direction of rolling can be calculated from this figure. Considering x the direction of the

motion and y the transverse direction, the radii of curvature for the ball-inner-race contact are

$$\begin{aligned} r_{ax} &= r_{ay} = \frac{D}{2}, \\ r_{bx} &= \frac{d_e - D\cos\beta}{2\cos\beta}, \\ r_{by} &= -f_i D = -r_i. \end{aligned} \tag{2.10}$$

The radii of curvature for the ball-outer-race contact are

$$\begin{aligned} r_{ax} &= r_{ay} = \frac{D}{2}, \\ r_{bx} &= -\frac{d_e + D\cos\beta}{2\cos\beta}, \\ r_{by} &= -f_o D = -r_o. \end{aligned} \tag{2.11}$$

Let

$$\gamma = \frac{D\cos\beta}{d_e}. \tag{2.12}$$

Then

$$\begin{aligned} r_{bx} &= \frac{D}{2}\frac{1-\gamma}{\gamma}, \\ \left.\frac{1}{R}\right|_i &= \frac{1}{r_{ax}} + \frac{1}{r_{bx}} + \frac{1}{r_{ay}} + \frac{1}{r_{by}} = \frac{1}{D}\left(4 - \frac{1}{f_i} + \frac{2\gamma}{(1-\gamma)}\right), \\ \Gamma_i &= R\left(\frac{1}{r_{ax}} + \frac{1}{r_{bx}} - \frac{1}{r_{ay}} - \frac{1}{r_{by}}\right) = \frac{1/f_i + 2\gamma/(1-\gamma)}{4 - 1/f_i + 2\gamma/(1-\gamma)}, \end{aligned} \tag{2.13}$$

for the ball-inner-race contact, and

$$\begin{aligned} r_{bx} &= -\frac{D}{2}\frac{1+\gamma}{\gamma}, \\ \left.\frac{1}{R}\right|_o &= \frac{1}{r_{ax}} + \frac{1}{r_{bx}} + \frac{1}{r_{ay}} + \frac{1}{r_{by}} = \frac{1}{D}\left(4 - \frac{1}{f_o} - \frac{2\gamma}{1+\gamma}\right), \\ \Gamma_o &= R\left(\frac{1}{r_{ax}} + \frac{1}{r_{bx}} - \frac{1}{r_{ay}} - \frac{1}{r_{by}}\right) = \frac{1/f_o\ - 2\gamma/(1+\gamma)}{4 - 1/f_o - 2\gamma/(1+\gamma)}, \end{aligned} \tag{2.14}$$

for the ball-outer-race contact.

3. Contact Stress and Deformations

When two elastic solids are brought together under a load, a contact area develops; the shape and size of which depend on the applied load, the elastic properties of the materials, and the curvatures of the surfaces. For two ellipsoids in contact the shape of the contact area is elliptical, with a being the semimajor axis in the y direction (transverse direction) and b being the semiminor axis in the x direction (direction of motion).

The *elliptical eccentricity parameter*, k, is defined as

$$k = \frac{a}{b}. \tag{3.1}$$

From Harris [5], k can be written in terms of the curvature difference, Γ, and the *elliptical integrals of the first and second kind*s $\mathbf{K}$ and $\mathbf{E}$, as

$$J(k) = \sqrt{\frac{2\mathbf{K} - \mathbf{E}(1+\Gamma)}{\mathbf{E}(1-\Gamma)}}, \tag{3.2}$$

where

$$\begin{aligned} \mathbf{K} &= \int_0^{\pi/2} \left[1 - \left(1 - \frac{1}{k^2}\right)\sin^2\varphi\right]^{-1/2} d\varphi, \\ \mathbf{E} &= \int_0^{\pi/2} \left[1 - \left(1 - \frac{1}{k^2}\right)\sin^2\varphi\right]^{1/2} d\varphi. \end{aligned} \tag{3.3}$$

A one-point iteration method which has been used successfully in the past [15] is used, where

$$k_{n+1} = J(k_n). \tag{3.4}$$

When the *ellipticity parameter*, k, the *elliptic integrals of the first* and *second kinds*, $\mathbf{K}$ and $\mathbf{E}$, respectively, the normal applied load, Q, Poisson's ratio, ν, and the modulus of elasticity, E, of the contacting solids, are known, we can write the semimajor and semiminor axes of the contact ellipse and the maximum deformation at the center of the contact, from the analysis of Hertz [16], as

$$\begin{aligned} a &= \left(\frac{6k^2\mathbf{E}QR}{\pi E'}\right)^{1/3}, \\ b &= \left(\frac{6\mathbf{E}QR}{\pi k E'}\right)^{1/3}, \end{aligned} \tag{3.5}$$

$$\delta = \mathbf{K}\left[\frac{9}{2\mathbf{E}R}\left(\frac{Q}{\pi k E'}\right)^2\right]^{1/3}, \tag{3.6}$$

where

$$E' = \frac{2}{\left(1 - \upsilon_a^2\right)/E_a + \left(1 - \upsilon_b^2\right)/E_b}. \tag{3.7}$$

4. Static Load Distribution under Eccentric Thrust Load

Methods to calculate distribution of load among the balls and rollers of rolling bearings statically loaded can be found in various papers [5, 17]. The methods have been limited to, at most, three degrees of freedom in loading and demand the solution of a simultaneous nonlinear system of algebraic equations for higher degrees of freedom. Solution of such equations generally necessitates the use of a digital computer. In certain cases, however—for example, applications with pure radial, pure thrust, or radial and thrust loading with nominal clearance—the simplified methods will probably provide sufficiently accurate calculational results.

Having defined a simple analytical expression for the deformation in terms of load in the previous section, it is possible to consider how the bearing load is distributed among the rolling elements. Most rolling-element bearing applications involve steady-state rotation of either the inner or outer race or both; however, the speeds of rotation are usually not so great as to cause ball or roller centrifugal forces or gyroscopic moments of significant magnitudes. In analyzing the loading distribution on the rolling elements, it is usually satisfactory to ignore these effects in most applications. In this section the load-deflection relationships for ball bearings are given, along with a specific load distribution consisting of an eccentric thrust load of statically loaded rolling elements.

4.1. Load-Deflection Relationships for Ball Bearings

From (3.6) it can be seen that for a given ball-raceway contact (point loading)

$$Q = K\delta^{3/2}, \tag{4.1}$$

where

$$K = \pi k E' \sqrt{\frac{2\mathbf{E}R}{9\mathbf{K}^3}}. \tag{4.2}$$

The total normal approach between two raceways under load separated by a rolling element is the sum of the approaches between the rolling element and each raceway. Hence

$$\delta_n = \delta_i + \delta_o. \tag{4.3}$$

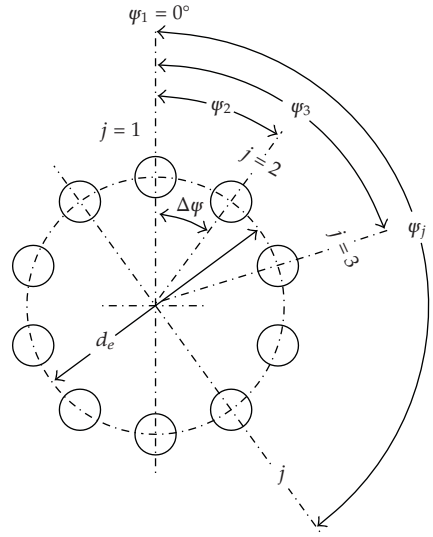


Figure 7: Ball angular positions in the radial plane that is perpendicular to the bearing's axis of rotation; $\Delta\psi = 2\pi/Z, \psi_j = 2\pi/Z(j-1)$.

Therefore,

$$K_n = \left[\frac{1}{1/{K_i}^{2/3} + 1/{K_o}^{2/3}}\right]^{3/2}, \tag{4.4}$$

$$Q = K_n \delta_n^{3/2}. \tag{4.5}$$

4.2. Ball Bearings under Eccentric Thrust Load

Let a ball bearing with a number of balls, Z, symmetrically distributed about a pitch circle according to Figure 7, be subjected to an eccentric thrust load. Then, a *relative axial displacement*, δ_a, and a *relative angular displacement*, θ, between the inner and outer ring raceways may be expected. Let $\psi = 0$ be the angular position of the maximum loaded ball.

Figure 8 shows the initial and final curvature centers positions at angular position ψ, before and after loading, whereas the centers of curvature of the raceway grooves are fixed with respect to the corresponding raceway. If δ_a and θ are known, then the *total axial displacement*, δ_t, at angular position ψ, is given by

$$\delta_t(\psi) = \delta_a + R_i \theta \cos\psi, \tag{4.6}$$

where

$$R_i = \frac{d_e}{2} + (f_i - 0.5) D \cos\beta_f \tag{4.7}$$

expresses the locus of the centers of the inner ring raceway groove curvature radii.

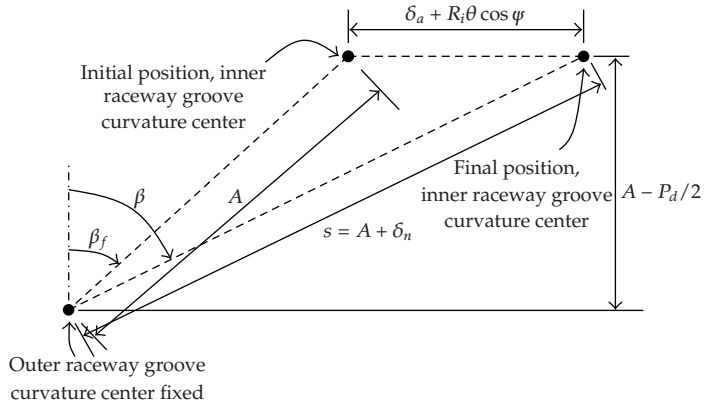


Figure 8: Initial and final curvature centers positions at angular position ψ, with and without applied load.

Also,

$$\delta_{\max} \equiv \delta_t(0) = \delta_a + R_i\theta. \tag{4.8}$$

From (4.6) and (4.8), one may develop the following relationship:

$$\delta_t = \delta_{\max}\left[1 - \frac{1}{2\varepsilon}(1 - \cos\psi)\right] \tag{4.9}$$

in which

$$\varepsilon = \frac{1}{2}\left(1 + \frac{\delta_a}{R_i\theta}\right). \tag{4.10}$$

The extension of the *loading zone* is defined by

$$\psi_l = \cos^{-1}\left(\frac{-\delta_a}{R_i\theta}\right). \tag{4.11}$$

From Figure 8,

$$\beta(\psi) = \cos^{-1}\left(\frac{A - P_d/2}{A + \delta_n}\right), \tag{4.12}$$

$$\delta_t(\psi) = (A + \delta_n)\sin\beta - A\sin\beta_f. \tag{4.13}$$

From (2.5) and (4.12), the total normal approach between two raceways at angular position ψ, after the thrust load has been applied, can be written as

$$\delta_n(\psi) = A\left(\frac{\cos\beta_f}{\cos\beta} - 1\right). \tag{4.14}$$

From Figure 8 and (4.14) it can be determined that s, the distance between the centers of the curvature of the inner and outer ring raceway grooves at any rolling-element position ψ, is given by

$$s(\psi) = A + \delta_n = A\frac{\cos\beta_f}{\cos\beta}. \tag{4.15}$$

From (4.6), (4.13), and (4.14), yields, for $\psi = \psi_j$,

$$\delta_a + R_i\theta\cos\psi_j - A\frac{\sin(\beta_j - \beta_f)}{\cos\beta_j} = 0, \quad j = 1,\ldots,Z. \tag{4.16}$$

From (4.5), and (4.14) one yields, for $\psi = \psi j$,

$$Q_j = K_{nj}A^{3/2}\left(\frac{\cos\beta_f}{\cos\beta_j} - 1\right)^{3/2}, \quad j = 1,\ldots,Z. \tag{4.17}$$

If the external thrust load, F_a, is applied at a point distant e from the bearing's axis of rotation, then for static equilibrium to exist

$$F_a = \sum_{j=1}^{Z} Q_j\sin\beta_j, \tag{4.18}$$

$$M = eF_a = \frac{1}{2}\sum_{j=1}^{Z} d_{cj}Q_j\sin\beta_j\cos\psi_j, \tag{4.19}$$

where $d_{cj} \equiv d_e - D\cos\beta_j$.

Substitution of (4.17) into (4.18) yields

$$F_a - A^{3/2}\sum_{j=1}^{Z} K_{nj}\sin\beta_j\left(\frac{\cos\beta_f}{\cos\beta_j} - 1\right)^{3/2} = 0. \tag{4.20}$$

Similarly,

$$eF_a - \frac{A^{3/2}}{2}\sum_{j=1}^{Z} K_{nj}d_{cj}\cos\psi_j\sin\beta_j\left(\frac{\cos\beta_f}{\cos\beta_j} - 1\right)^{3/2} = 0. \tag{4.21}$$

Equations (4.16), (4.20), and (4.21) are $Z + 2$ simultaneous nonlinear equations with unknowns δ_a, θ, and β_j, $j = 1,\ldots,Z$. Since K_{nj} and d_{cj} are functions of final contact angle, β_j, the equations must be solved iteratively to yield an exact solution for δ_a, θ, and β_j.

5. Numerical Results

A numerical method (the Newton-Rhapson method) was chosen to solve the simultaneous nonlinear equations (4.16), (4.20), and (4.21). Choosing the rolling bearing, input must be given the geometric parameters d_i, d_o, D, Z, r_i, and r_o, in accordance with Figures 3 and 5, and the elastic properties E_a, E_b, ν_a, and ν_b. Next, the following parameters must be obtained: f_i, f_o, B, A, $\psi_j (j = 1, \ldots, Z)$, E', d_e, P_d, β_f, and R_i.

The interest here is to observe the behavior of an angular-contact ball bearing under a known thrust load which is to be applied statically to a *variable* distance (lever arm or eccentricity), e, from the geometric bearing center line. Then, given a thrust load and the initial estimates for δ_a, θ, and β_j, $j = 1, \ldots, Z$, for each distance e, varying from zero up to a given maximum eccentricity, the values $1/R|_i$, $1/R|_o$, Γ_i, Γ_o, k_i, k_o, $\mathbf{K}_i$, $\mathbf{K}_o$, $\mathbf{E}_i$, $\mathbf{E}_o$, K_i, K_o, and K_n are calculated for each ball, according to previous sections, and new values for δ_a, θ, and β_j are obtained. The new β_j values are compared with old ones, and if the difference is greater than a minimal error, then new values for $1/R|_i$, $1/R|_o$, Γ_i, Γ_o, k_i, k_o, $\mathbf{K}_i$, $\mathbf{K}_o$, $\mathbf{E}_i$, $\mathbf{E}_o$, K_i, K_o, and K_n are calculated for each ball, and again new values for δ_a, θ, and β_j are obtained. If the difference is lesser than the error then a new value for e is taken. If e is the last valid value, then a new thrust load value is acquired and the procedure is repeated up to the last valid thrust load value, when the program ends.

To show an application of the theory developed in this work, a numerical example is presented here. It was chosen the 218 angular-contact ball bearing that was also used by Harris[5]. Thus, the results generated here can be compared to a certain degree with Harris results. The input data for this rolling bearing were the following:

inner raceway diameter: $d_i = 0.10279\,\text{m}$,
outer raceway diameter: $d_o = 0.14773\,\text{m}$,
ball diameter: $D = 0.02223\,\text{m}$,
ball number: $Z = 16$,
inner groove radius: $r_i = 0.01163\,\text{m}$,
outer groove radius: $r_o = 0.01163\,\text{m}$,
modulus of elasticity for both balls and races: $E = 2.075 \times 10^{11}\,\text{N/m}^2$,
poisson's ratio for both balls and races: $\upsilon = 0.3$.

The remaining parameters have been calculated yielding:

inner race conformity: $f_i = 0.523166891587944$,
outer race conformity: $f_o = 0.523166891587944$,
total conformity ratio: $B = 0.046333783175888$,
initial curvature centers distance: A=0.00103 m,
effective elastic modulus: $E' = 228021978021.978\,\text{N/m}^2$,
angular spacing between rolling elements: $\Delta\psi = 22.5°$,
angular position of rolling elements: $\psi_j = 22.5°(j - 1)$, $j = 1, \ldots, 16$,
bearing pitch diameter: $d_e = 0.12526\,\text{m}$,
diametral clearance: $P_d = 0.00048\,\text{m}$,
free-contact angle: $\beta_f = 39.915616407992260°$,
radius of locus of inner raceway groove curvature centers: $R_i = 0.063025\,\text{m}$.

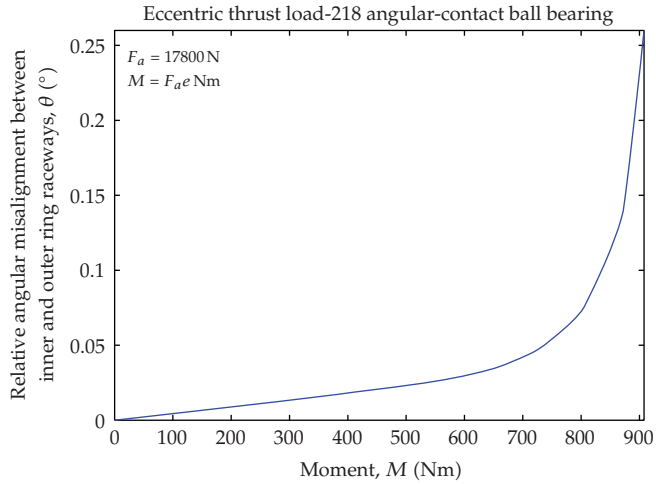


Figure 9: Relative angular misalignment, θ, for 17,800 N thrust load, as a function of the Moment, M.

For each thrust load value, the initial estimates for δ_a, θ, and β_j were the following:

axial deflection: $\delta^a = 10^{-5}$ m,

misalignment angle: $\theta = 10^{-2}$ rd,

contact angle: $\beta_j = 1.1\beta_f$, $j = 1, \ldots, 16$.

5.1. Numerical Results for a 17,800 N Thrust Load

Since it is the qualitative behavior of solutions that is the interest, the results are presented here in graphical form.

Initially, for comparative purposes with the Harris work, a specific thrust load $F_a = 17,800$ N was chosen to be applied, and the following graphical results are presented as functions of the moment, $M = F_a e$:

(i) relative angular displacement, θ (Figure 9),

(ii) partial axial displacement, $R_i\theta\cos\psi$ (Figure 10),

(iii) relative axial displacement, δ_a (Figure 11),

(iv) total relative axial deflection, δ_t (Figure 12),

(v) loading zone, ψ_l (Figure 13),

(vi) distance between loci of inner and outer raceway groove curvature centers, s (Figure 14),

(vii) maximum elastic compression at the ball/inner-race contact, δ_i (Figure 15),

(viii) maximum elastic compression at the ball/outer-race contact, δ_o (Figure 16),

(ix) total normal ball deflection, δ_n (Figure 17),

(x) ball-raceway normal load, Q (Figure 18),

(xi) contact angle, β (Figure 19),

(xii) semimajor axis of the ball/inner-race contact area, a_i (Figure 20),

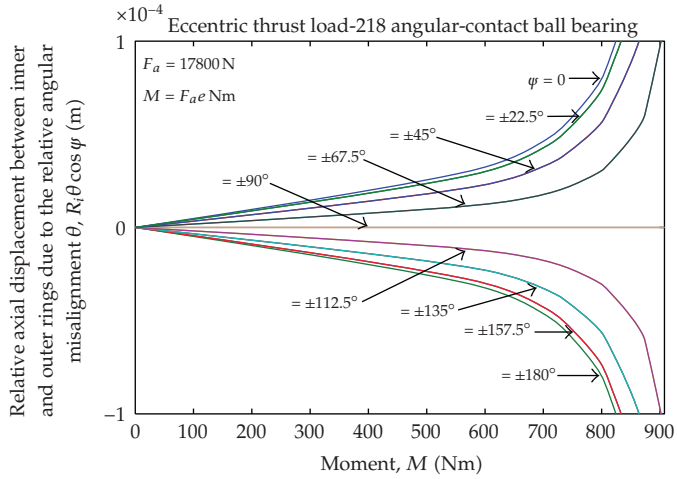


Figure 10: Partial axial deflection, $R_i\theta\cos\psi$, for each ball and 17,800 N thrust load, as a function of the Moment, M.

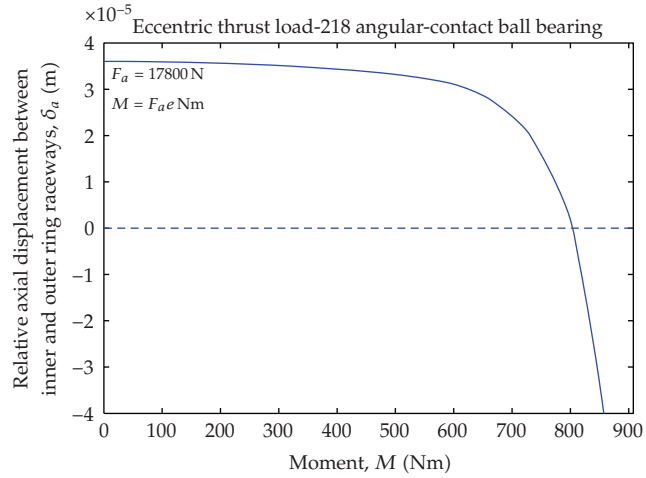


Figure 11: Axial deflection, δ_a, for 17,800 N thrust load, as a function of the Moment, M.

(xiii) semiminor axis of the ball/inner-race contact area, b_i (Figure 21),

(xiv) semimajor axis of the ball/outer-race contact area, a_o (Figure 22),

(xv) semiminor axis of the ball/outer-race contact area, b_o (Figure 23),

(xvi) elliptical eccentricity parameter for ball/inner-race contact, k_i (Figure 24),

(xvii) elliptical eccentricity parameter for ball/outer-race contact, k_o (Figure 25).

The graphics above, with exception of Figures 9, 11, and 13, show one curve for each ball angular position.

Figures 9 and 10 show the relative angular misalignment, θ, and the partial axial deflection for each ball, $R_i\theta\cos\psi$, respectively. It is observed that there is an approximately linear relationship between the misalignment angle, θ, and applied moment, M, for moment

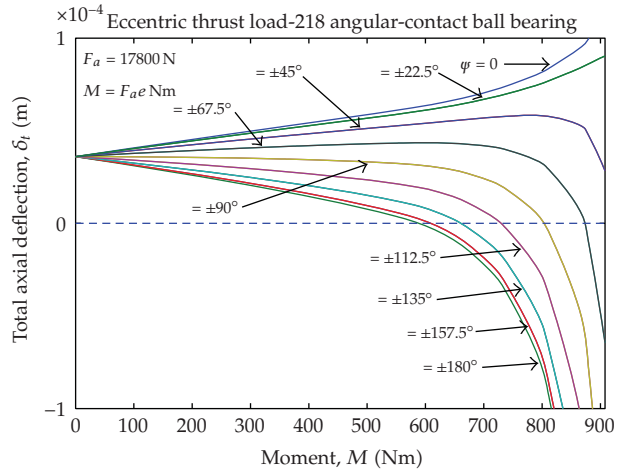


Figure 12: Total axial deflection, δ_t, for 17,800 N thrust load, as a function of the Moment, M.

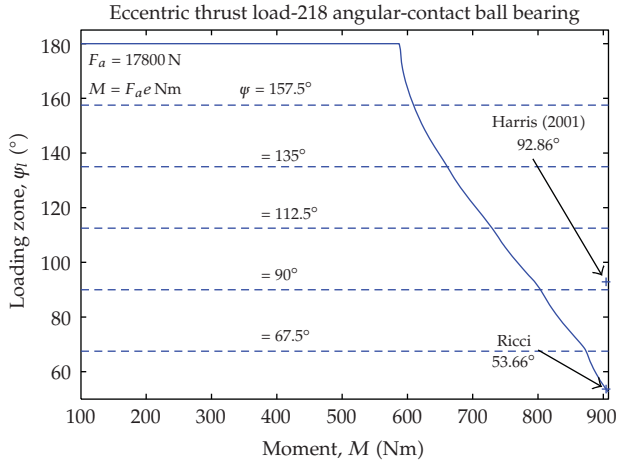


Figure 13: Loading zone, ψ_l, for 17,800 N thrust load, as a function of the Moment, M.

values ranging from zero up to about 600 Nm, which corresponds to a distance e of approximately 33.7 mm. Keeping the load constant and increasing the lever arm, e, above this value, it can be observed a deeper increase in the misalignment angle and, therefore, in the resultant axial deflection, $R_i\theta$.

From Figure 9 it can be observed that for an applied moment of 900 Nm (e ≅ 50.6 mm) the angular misalignment can be as high as a quarter of degree.

As already been waited for, from Figure 10 it can be observed that the partial axial deflection is symmetrical with respect to the horizontal axis (null displacement) and that the displacement is null for the balls located at $\psi = \pm 90°$.

Figure 11 shows the axial deflection, δ_a. It is observed that the axial deflection, δ_a, is approximately constant for moment values where the relationship between θ and M is approximately linear, that is, from zero up to about 600 Nm (e ≅ 33.7 mm). For higher moment values the axial deflection falls abruptly and becomes negative in the vicinity of

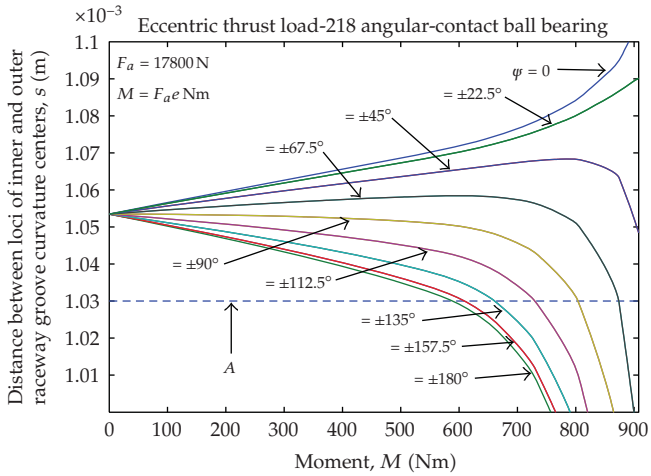


Figure 14: Distance between curvature centers, s, for 17,800 N thrust load, as a function of the Moment, M.

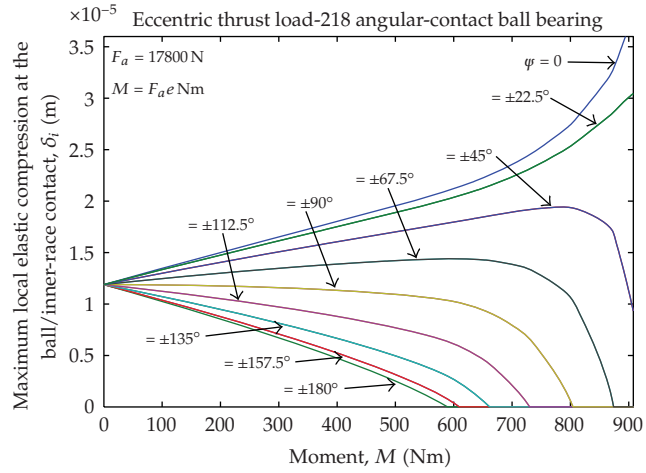


Figure 15: Maximum normal elastic compression at the ball/inner race contact, δ_i, for 17,800 N thrust load, as a function of the Moment, M.

800 Nm (e ≅ 44.9 mm). The deeper increase in θ due to the increase in the lever arm forces the decrease of δ_a to preserve the force and moment static balances.

Figure 12 shows the total axial deflection, δ_t. It can be observed that the total axial deflection, δ_t, is the axial deflection, δ_a, in two situations: under centric thrust load ($e = 0$), where all balls have the same axial deflection ($3.6011095400455 \times 10^{-5}$ m), and under eccentric thrust load for balls located at $\psi = \pm 90°$. Increasing from zero the lever arm, an almost linear increase (decrease) in the total axial deflection is observed for the balls whose angular positions satisfy $|\psi| < 90°$ ($|\psi| > 90°$). This relation is approximately linear up to vicinity of $M = 600$ Nm when the ball located at $\psi = 180°$ occurs to be unloaded, that is, $\delta_t(\psi = 180°) = 0$ for $M = 588.9687$ Nm ($e = 3.3088 \times 10^{-2}$ m).

From Figure 12 it is observed that for eccentricity of about 50 mm the total axial deflection of the most heavily loaded ball can reach one tenth of millimeter.

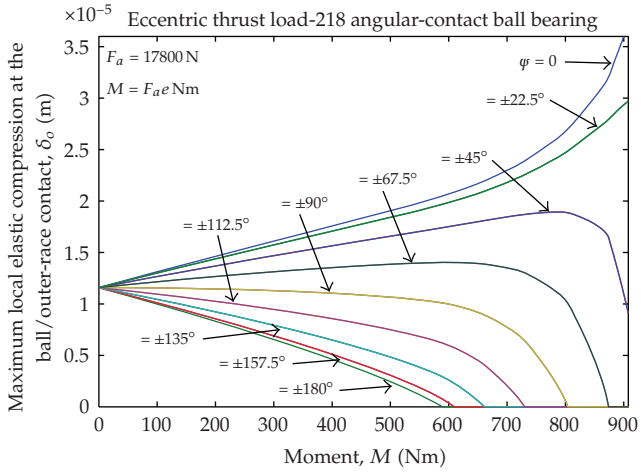


Figure 16: Maximum normal elastic compression at the ball/outer-race contact, δ_o, for 17,800 N thrust load, as a function of the Moment, M.

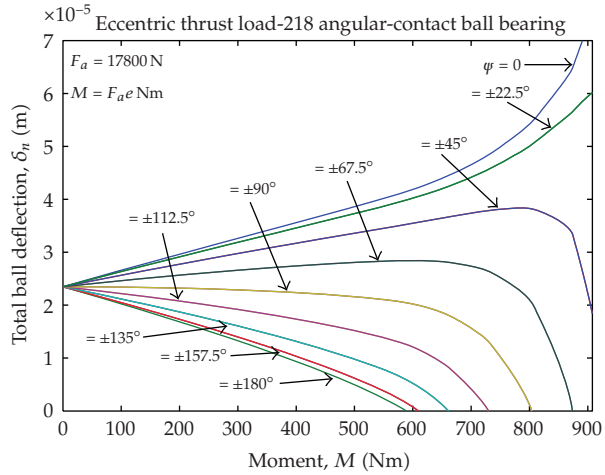


Figure 17: Total ball deflection, δ_n, for 17,800 N thrust load, as a function of the Moment, M.

The Figure 13 shows the loading zone, ψ_l. The increase of the moment above 587.4 Nm (or lever arm above 3.3×10^{-2} m) causes the decrease of the loading zone from initial value $\psi_l = \pm 180°$, with the successive unloading of the balls pairs located at $\psi = \pm 157.5°$ (M = 609.448 Nm), $\psi = \pm 135°$ (M = 661.1407 Nm), $\psi = \pm 112.5°$ (M = 729.9584 Nm), $\psi = \pm 90°$ (M = 803.9741 Nm), and $\psi = \pm 67.5°$ (M = 873.7125 Nm), respectively. Going ahead cause the unloading of the balls pair located at $\psi = \pm 45°$. However, it is not advisable to go beyond M = 900 Nm, once the radial displacements between curvature centers start to acquire micrometer order values and they cannot more be disregarded.

Figure 13 shows a substantial difference between results found in this work and those found by Harris. While Harris found a loading zone of 92.86° (p. 252) for an eccentricity of 50.8 mm, this work found a loading zone of 53.66°. Considering the last result as reference, this represents an error of +73% in the loading angle, meaning that Harris calculation has underestimated the effect of the moment M.

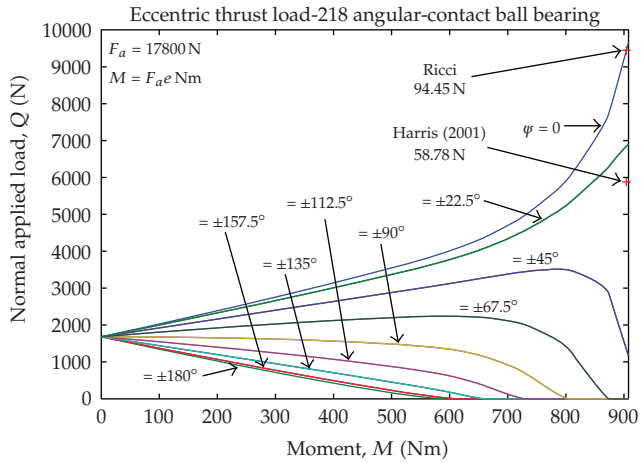


Figure 18: Normal ball load, Q, for 17,800 N thrust load, as a function of the Moment, M.

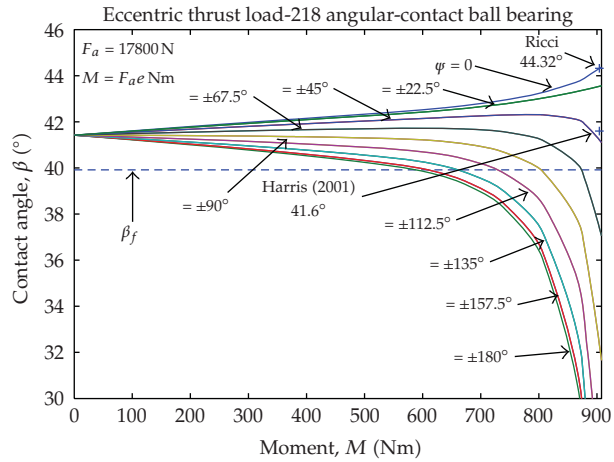


Figure 19: Contact Angle, β, for 17,800 N thrust load, as a function of the Moment, M.

Figure 14 shows the distance between loci of inner and outer raceway groove curvature centers, s. It can be observed that the distance, s, under centric thrust load ($e = 0$), is the same for all balls ($1.053468971830 \times 10^{-3}$ m). Increasing from zero the lever arm, an almost linear increase (decrease) in the distance, s, is observed for the balls whose angular positions satisfy $|\psi| < 90°$ ($|\psi| > 90°$). This relation is approximately linear up to vicinity of $M = 600$ Nm when the ball located at $\psi = 180°$ occurs to be unloaded, that is, $s(\psi = 180°) = A$ for $M = 588.9687$ Nm ($e = 3.3088 \times 10^{-2}$ m).

The increase of the moment above 588.9687 Nm (or lever arm above 3.3088×10^{-2} m) causes the decrease of the loading zone, as already explained, with the successive unloading of the ball pairs. At the points where the unloading occurs it is observed that the distance s falls below of the distance between centers of curvature, A, for the unloaded bearing.

Figures 15 and 16 show the maximum normal elastic compressions at the ball/inner-race and ball/outer-race contacts, δ_i and δ_o, respectively. It can be observed that δ_i and δ_o, under centric thrust load ($e = 0$), are the same for all balls ($1.18852986717367 \times 10^{-5}$ m for δ_i

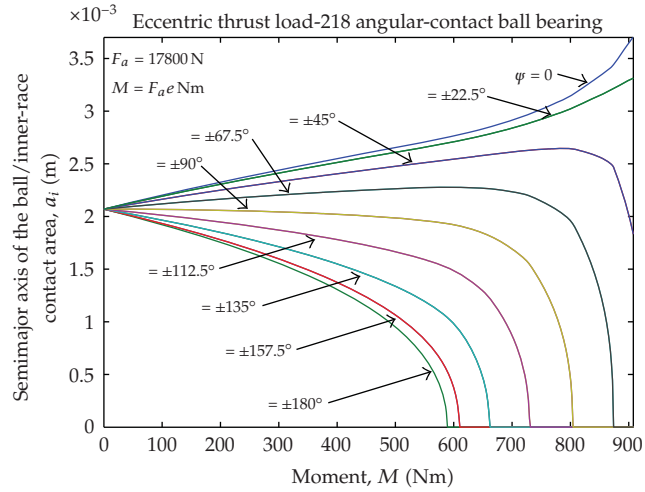


Figure 20: Semimajor axis of the ball/inner-race contact area, a_i, for 17,800 N thrust load, as a function of the Moment, M.

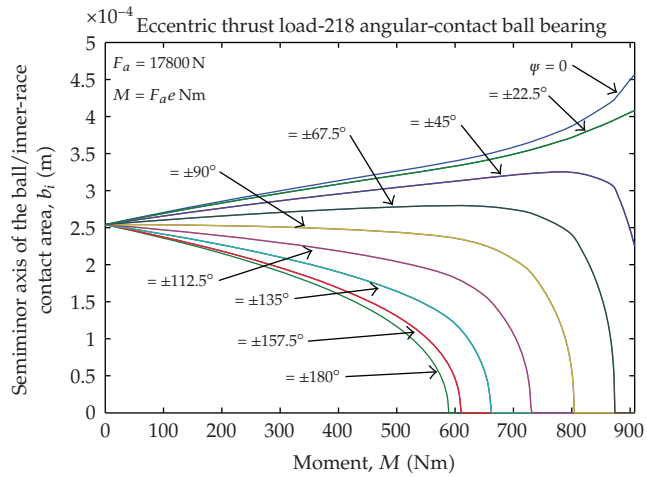


Figure 21: Semiminor axis of the ball/inner-race contact area, b_i, for 17,800 N thrust load, as a function of the Moment, M.

and $1.15836731583185 \times 10^{-5}$ m for δ_o) and that the deformation for the maximum loaded ball, in both cases, can reach values as high as 36 μm for moment about 900 Nm.

Figure 17 shows the total normal ball deflection, δ_n, that can be obtained by summing the maximum normal elastic compressions on the inner and outer races, δ_i and δ_o, or by subtracting A from s, once $\delta_n = s - A > 0$ also. It can be observed that δ_n, under centric thrust load ($e = 0$), is the same for all balls ($2.3468971830055 \times 10^{-5}$ m) and that the total normal elastic deformation for the maximum loaded ball can reach values as high as 70 μm for moment about 900 Nm.

Figure 18 shows the normal ball load, Q. It can be observed that the normal ball load, Q, under centric thrust load ($e = 0$), is the same for all balls (1, 681.663561507042 N). Increasing from zero the lever arm, an almost linear increase (decrease) in the normal ball

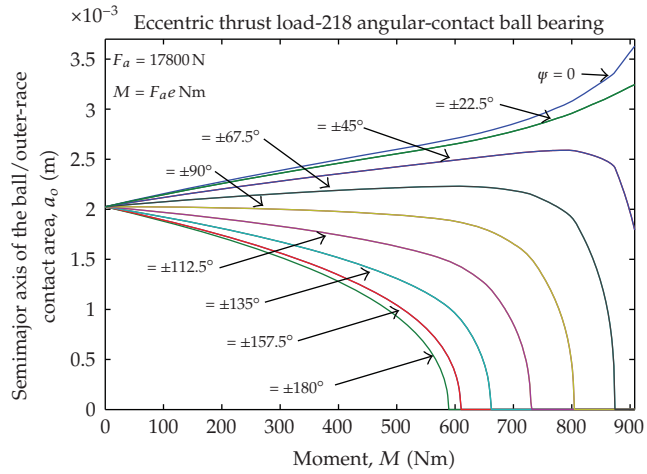


Figure 22: Semimajor axis of the ball/outer-race contact area, a_o, for 17,800 N thrust load, as a function of the Moment, M.

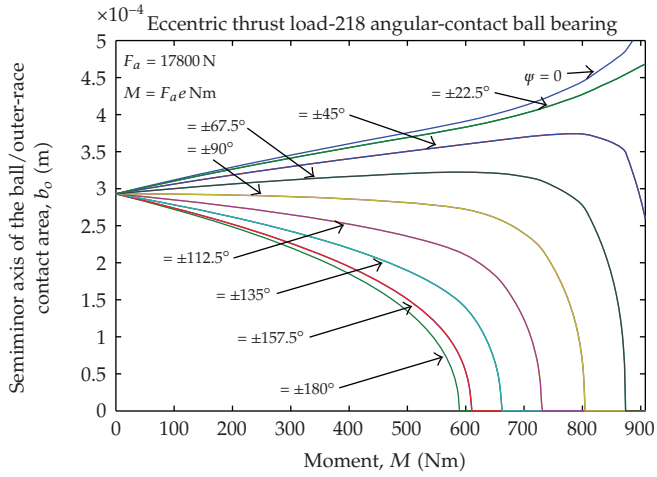


Figure 23: Semiminor axis of the ball/outer-race contact area, b_o, for 17,800 N thrust load, as a function of the Moment, M.

load is observed for the balls whose angular positions satisfy $|\psi| < 90°$ ($|\psi| > 90°$). This relation is approximately linear up to vicinity of $M = 600$ Nm when the ball located at $\psi = 180°$ occurs to be unloaded, that is, $Q(\psi = 180°) = 0$ for $M = 589.18$ Nm ($e = 3.31 \times 10^{-2}$ m).

Figure 18, as well as Figure 13, shows a substantial difference between results found in this work and those found by Harris. While Harris found a 5, 878 N magnitude for the maximum normal ball load (p. 252), for an applied load eccentricity of 50.8 mm, this work found a 9, 445 N maximum normal ball load. This represents an error of −62.2% in the normal load, meaning that the Harris calculation has underestimated the normal load for the maximum loaded ball.

Figure 19 shows the contact angle, β. It can be observed that the contact angle, β, under centric thrust load ($e = 0$), is the same for all balls (41.417986227161386^o). Increasing from zero the lever arm, an almost linear increase (decrease) in the contact angle is observed for

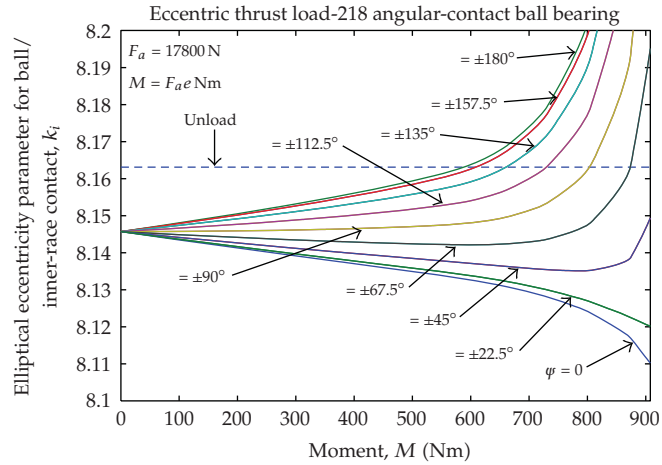


Figure 24: Elliptical eccentricity parameter for ball/inner-race contact, k_i, for 17,800 N thrust load, as a function of the Moment, M.

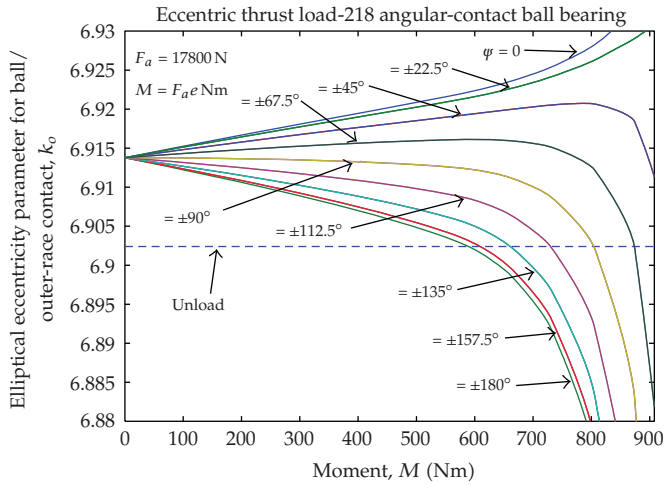


Figure 25: Elliptical eccentricity parameter for ball/outer-race contact, k_o, for 17,800 N thrust load, as a function of the Moment, M.

the balls whose angular positions satisfy $|\psi| < 90°$ ($|\psi| > 90°$). This relation is approximately linear up to vicinity of $M = 600$ Nm when the ball located at $\psi = 180°$ occurs to be unloaded, that is, $\beta(\psi = 180°) = \beta_f$ for $M = 589.18$ Nm ($e = 33.1$ mm).

Figure 19, as well as Figure 18 and Figure 13, shows a substantial difference between results found in this work and to those found by Harris. While Harris has assumed a contact angle magnitude of 41.6^o for all balls (p. 252), under a 50.8 mm applied load eccentricity, contact angles ranging from 44.31727851159821^o to 16.16919216282055^o were found in this work while ψ were varied from $\psi = 0°$ to $\pm 180^o$, respectively. This represents errors between −6.1% and +157.3% in the contact angles determination, meaning that the Harris calculation has underestimated (strongly overestimated) the contact angles for balls located at angular positions satisfying $|\psi| < 45°$($|\psi| > 45°$).

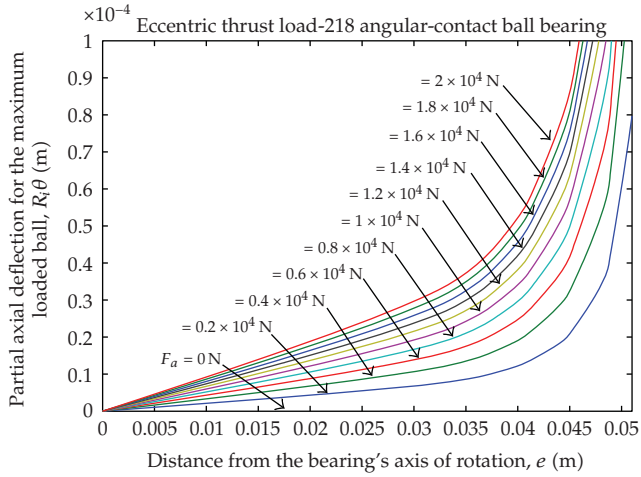


Figure 26: Partial axial displacement for the maximum loaded ball, $R_i\theta$, as a function of lever arm, e.

Figures 20 and 22 show the semimajor axes of the ball/inner-race and ball/outer-race contact areas, a_i and a_o, respectively. It can be observed that a_i and a_o, under centric thrust load ($e = 0$), are the same for all balls (2.069901480072 mm for a_i and 2.025827993682 mm for a_o) and that the major axes for the maximum loaded ball, in both cases, can reach values as high as 7.4 mm for moment about 900 Nm.

Figures 21 and 23 show the semiminor axes of the ball/inner-race and ball/outer-race contact areas, b_i and b_o, respectively. It can be observed that b_i and b_o, under centric thrust load ($e = 0$), are the same for all balls (0.254108993896064 mm for b_i and 0.293013306181356 mm for b_o) and that the major axes for the maximum loaded ball, in both cases, can reach values as high as 0.9 mm for moment about 900 Nm.

Figures 24 and 25 show the elliptical eccentricity parameters for ball/inner-race and ball/outer-race contact, k_i and k_o, respectively. It can be observed that k_i and k_o, under centric thrust load ($e = 0$), are the same for all balls (8.1457 for k_i and 6.9138 for k_o). Increasing from zero the lever arm, an almost linear increase (decrease) in the parameter k_i is observed for the balls whose angular positions satisfy $|\psi| > 90^\circ$ ($|\psi| < 90^\circ$), and an almost linear increase (decrease) in the parameter k_o is observed for the balls whose angular positions satisfy $|\psi| < 90^\circ$ ($|\psi| > 90^\circ$). These relations are approximately linear up to vicinity of $M = 600$ Nm when the ball located at $\psi = 180^\circ$ occurs to be unloaded. It can be observed that $k_i(\psi = 180^\circ) \cong 8.1631$ and $k_o(\psi = 180^\circ) \cong 6.9024$ when $M = 588.9687$ Nm ($e = 3.3088 \times 10^{-2}$ m).

The increase of the moment above 588.9687 Nm (or lever arm above 3.3088×10^{-2} m), causes the successive unloading of the ball pairs. At the points where the unloading occurs the values of the parameters k_i and k_o remain roughly equal to those indicated in the preceding paragraph for $\psi = 180^\circ$. So, it can be observed that the contact ellipse of the inner race is slightly more eccentric than that of the contact ellipse of the outer race. For $M = 900$ Nm, for example, while k_i varies numerically from 8.11, for the most heavily loaded ball, to 8.37, for the minimum loaded ball, k_o varies from 6.83 to 6.37, respectively.

5.2. Numerical Results for Thrust Load Ranging from 0 up to 20,000 N

Graphics for various thrust loads also are shown. The following graphics present curves for thrust loads ranging from 0 up to 20, 000 N as functions of lever arm, e:

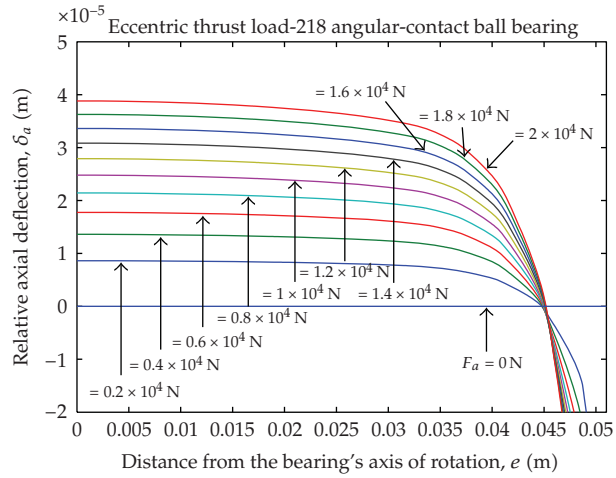


Figure 27: Relative axial displacement, δ_a, as a function of lever arm, e.

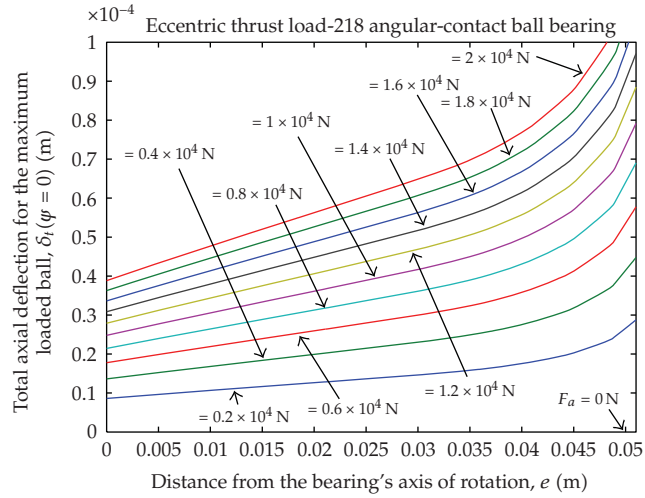


Figure 28: Total relative axial deflection for the maximum loaded ball, $\delta_t(\psi = 0)$, as a function of lever arm, e.

(i) partial axial displacement for the maximum loaded ball, $R_i\theta$ (Figure 26),

(ii) relative axial displacement, δ_a (Figure 27),

(iii) total relative axial deflection for the maximum loaded ball, $\delta_t(\psi = 0)$ (Figure 28),

(iv) total relative axial deflection for the minimum loaded ball, $\delta_t(\psi = 180°)$ (Figure 29),

(v) loading zone, ψ_l (Figure 30),

(vi) distance between loci of inner and outer raceway groove curvature centers for the maximum loaded ball, $s(\psi = 0)$ (Figure 31),

(vii) distance between loci of inner and outer raceway groove curvature centers for the minimum loaded ball, $s(\psi = 180°)$ (Figure 32),

(viii) total normal ball deflection for the maximum loaded ball, $\delta_n(\psi = 0)$ (Figure 33),

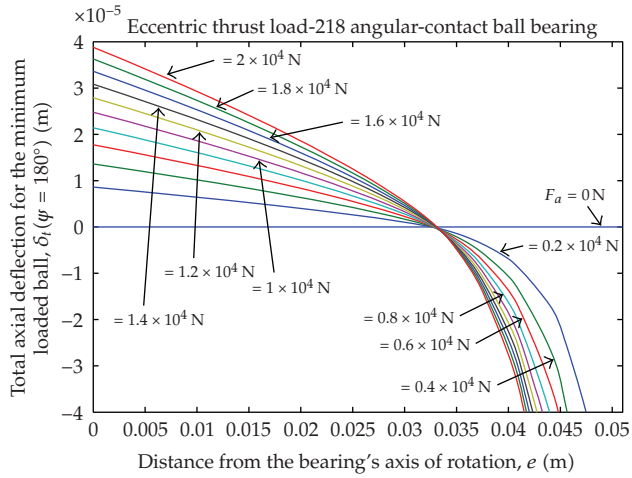


Figure 29: Total relative axial deflection for the minimum loaded ball, $\delta_t(\psi = 180°)$, as a function of lever arm, e.

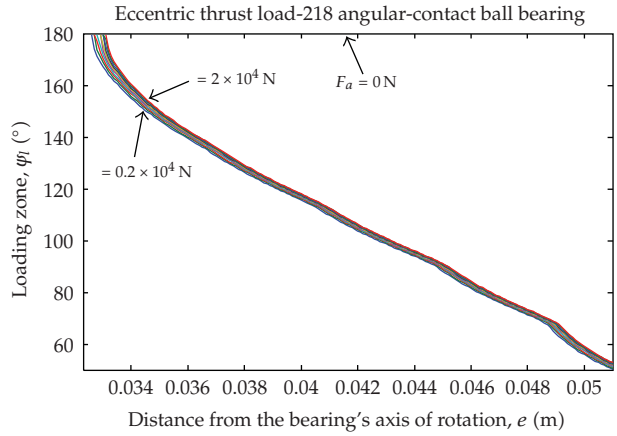


Figure 30: Loading zone, ψ_l, as a function of lever arm, e.

(ix) total normal ball deflection for the minimum loaded ball, $\delta_n(\psi = 180°)$ (Figure 34),

(x) ball-raceway normal load for the maximum loaded ball, $Q(\psi = 0)$ (Figure 35),

(xi) ball-raceway normal load for the minimum loaded ball, $Q(\psi = 180°)$ (Figure 36),

(xii) contact angle for the maximum loaded ball, $\beta(\psi = 0)$ (Figure 37),

(xiii) contact angle for the minimum loaded ball, $\beta(\psi = 180°)$ (Figure 38),

(xiv) semimajor axis of the ball/inner-race contact area for the maximum loaded ball, $a_i(\psi = 0)$ (Figure 39),

(xv) semiminor axis of the ball/inner-race contact area for the maximum loaded ball, $b_i(\psi = 0)$ (Figure 40),

(xvi) semimajor axis of the ball/outer-race contact area for the maximum loaded ball, $a_o(\psi = 0)$ (Figure 41),

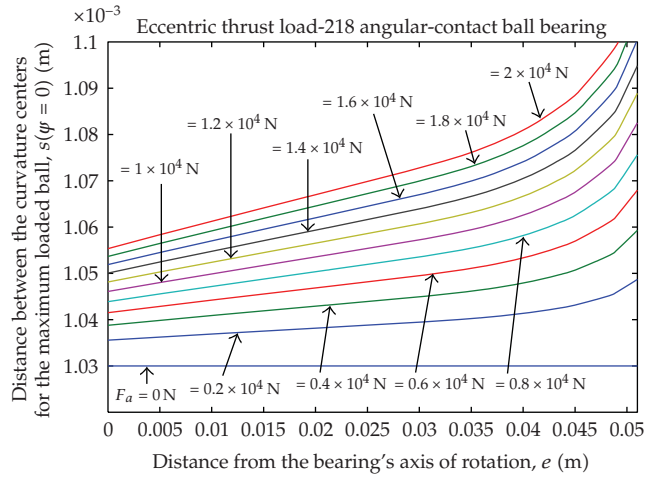


Figure 31: Distance between loci of inner and outer raceway groove curvature centers for the maximum loaded ball, $s(\psi = 0)$, as a function of lever arm, e.

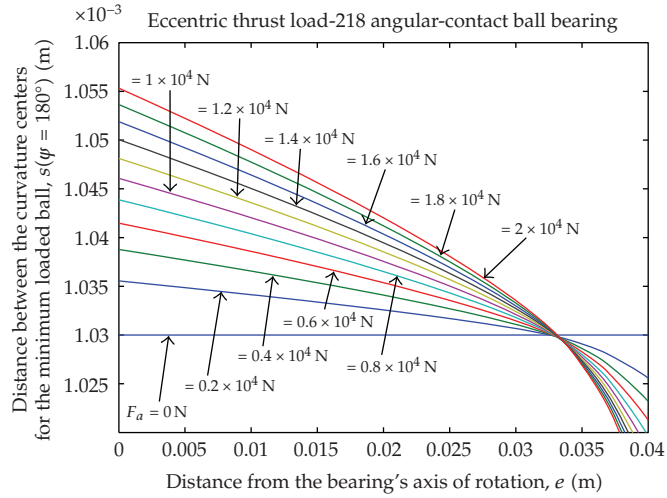


Figure 32: Distance between loci of inner and outer raceway groove curvature centers for the minimum loaded ball, $s(\psi = 180°)$, as a function of lever arm, e.

(xvii) semiminor axis of the ball/outer-race contact area for the maximum loaded ball, $b_o(\psi = 0)$ (Figure 42),

(xviii) semimajor axis of the ball/inner-race contact area for the minimum loaded ball, $a_i(\psi = 180°)$ (Figure 43),

(xix) semiminor axis of the ball/inner-race contact area for the minimum loaded ball, $b_i(\psi = 180°)$ (Figure 44),

(xx) semimajor axis of the ball/inner-race contact area for the minimum loaded ball, $a_o(\psi = 180°)$ (Figure 45),

(xxi) semiminor axis of the ball/inner-race contact area for the minimum loaded ball, $b_o(\psi = 180°)$ (Figure 46),

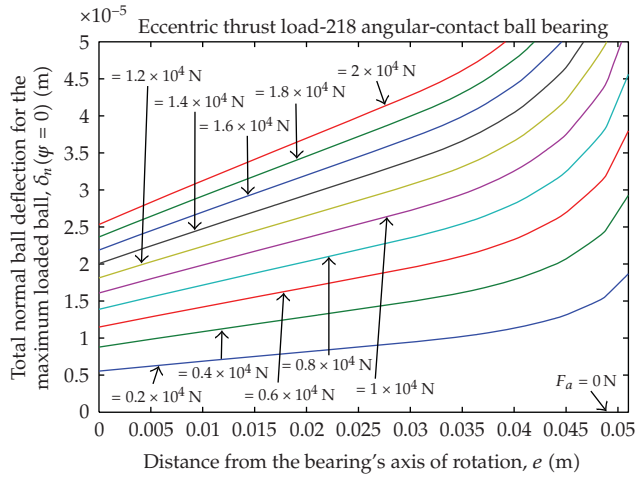


Figure 33: Total normal ball deflection for the maximum loaded ball, $\delta_n(\psi = 0)$, as a function of lever arm, e.

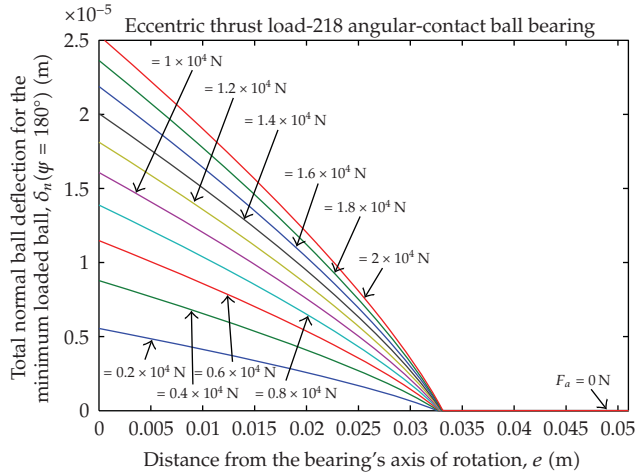


Figure 34: Total normal ball deflection for the minimum loaded ball, $\delta_n(\psi = 180°)$, as a function of lever arm, e.

(xxii) elliptical eccentricity parameter of the ball/inner-race contact area for the maximum loaded ball, $k_i(\psi = 0)$ (Figure 47),

(xxiii) elliptical eccentricity parameter of the ball/outer-race contact area for the maximum loaded ball, $k_o(\psi = 0)$ (Figure 48),

(xxiv) elliptical eccentricity parameter of the ball/inner-race contact area for the minimum loaded ball, $k_i(\psi = 180°)$ (Figure 49),

(xxv) elliptical eccentricity parameter of the ball/outer-race contact area for the minimum loaded ball, $k_o(\psi = 180°)$ (Figure 50).

Due to the size quite extensive of the paper comments about the figures will be omitted from now on.

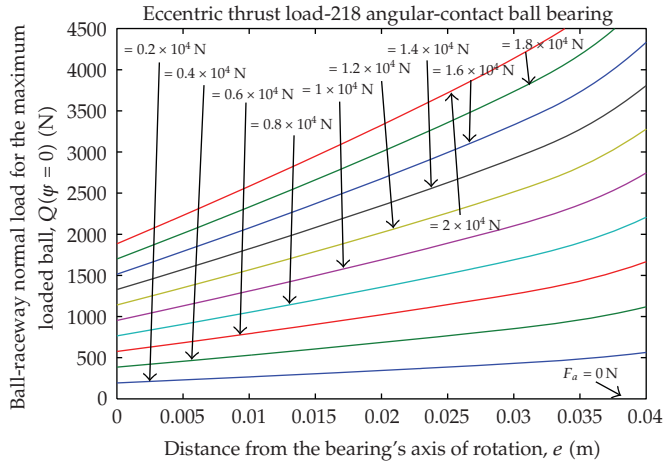


Figure 35: Ball-raceway normal load for the maximum loaded ball, $Q(\psi = 0)$, as a function of lever arm, e.

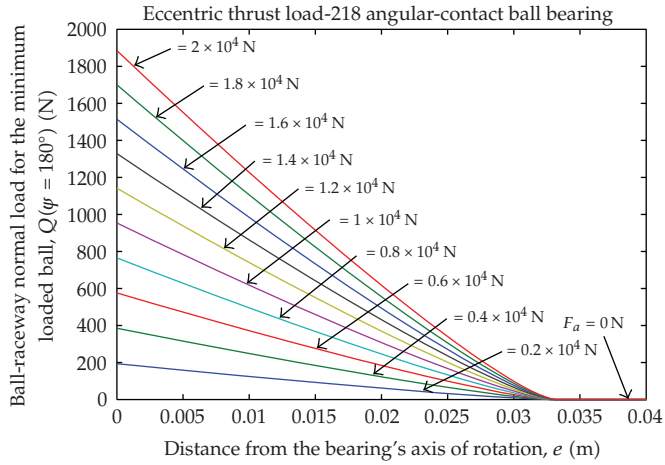


Figure 36: Ball-raceway normal load for the minimum loaded ball, $Q(\psi = 180°)$, as a function of lever arm, e.

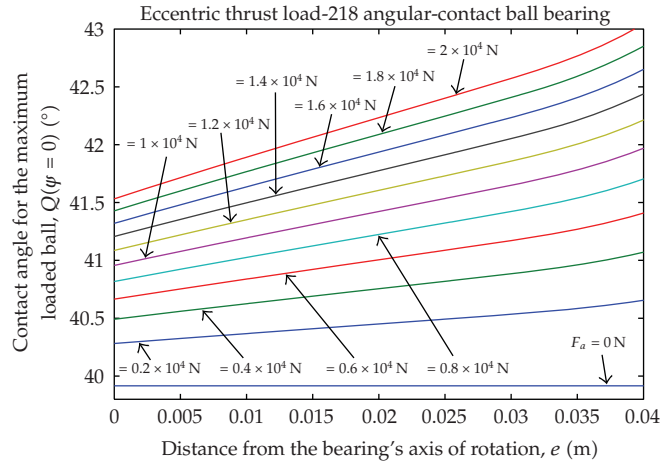


Figure 37: Contact angle for the maximum loaded ball, $\beta(\psi = 0)$, as a function of lever arm, e.

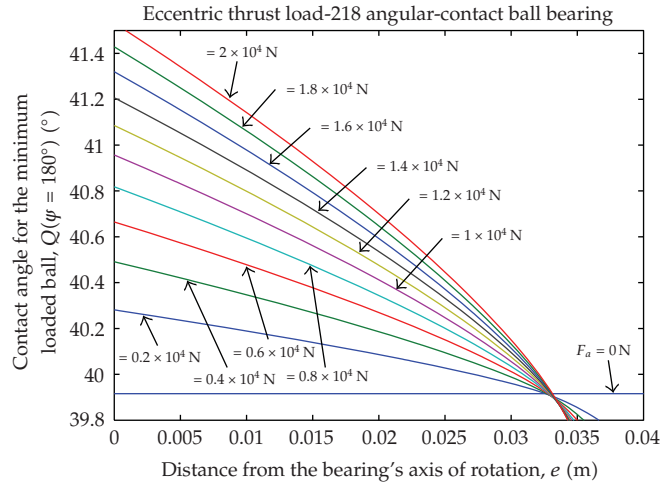


Figure 38: Contact angle for the minimum loaded ball, $\beta(\psi = 180°)$, as a function of lever arm, e.

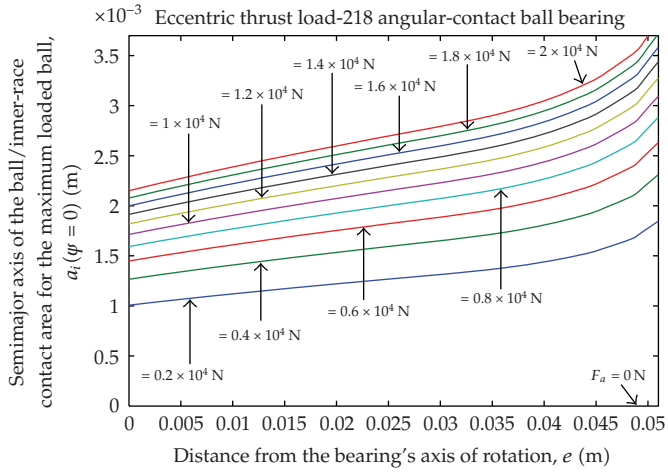


Figure 39: Semimajor axis of the ball/inner-race contact area for the maximum loaded ball, $a_i(\psi = 0)$, as a function of lever arm, e.

6. Conclusions

The importance of this work lies in the fact that it uses a new procedure for gettinnng numerically, accurately, and quickly the static load distribution of a single-row, angular-contact ball bearings, subjected to a known thrust load which is applied to a variable distance from the geometric bearing center line. Precise applications, as for example, space applications, require a precise determination of the static loading. Models available in literature are approximate and often are not compatible with the desired degree of accuracy. This work can be extended to determine the loading on high-speed bearings where centrifugal and gyroscopic forces are not discarded. The results of this work can be used in the accurate determination of the friction torque of the ball bearings, under any operating condition of temperature and speed.

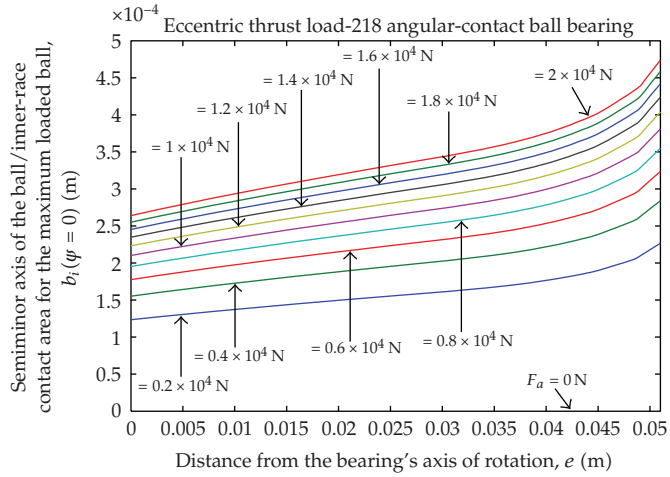


Figure 40: Semiminor axis of the ball/inner-race contact area for the maximum loaded ball, $b_i(\psi = 0)$, as a function of lever arm, e.

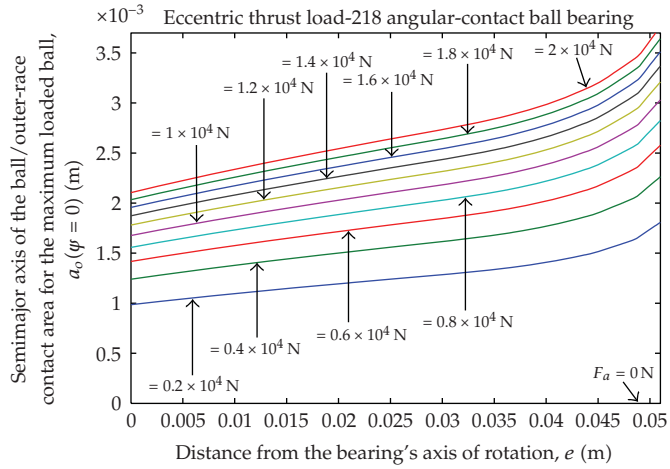


Figure 41: Semimajor axis of the ball/outer-race contact area for the maximum loaded ball, $a_o(\psi = 0)$, as a function of lever arm, e.

Symbols

a: Semimajor axis of the projected contact, m
A: Distance between raceway groove curvature centers, m
b: Semiminor axis of the projected contact, m
B: $f_o + f_i - 1$, total curvature
d: Raceway diameter, m
d_a: Bearing outer diameter, m
d_b: Bearing inner diameter, m
d_c: Contact diameter, m
d_e: Bearing pitch diameter, m
D: Ball diameter, m

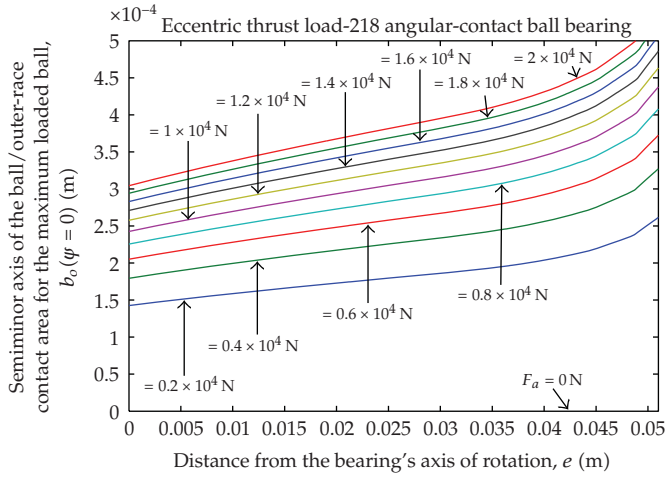


Figure 42: Semiminor axis of the ball/outer-race contact area for the maximum loaded ball, $b_o(\psi{=}0)$, as a function of lever arm, e.

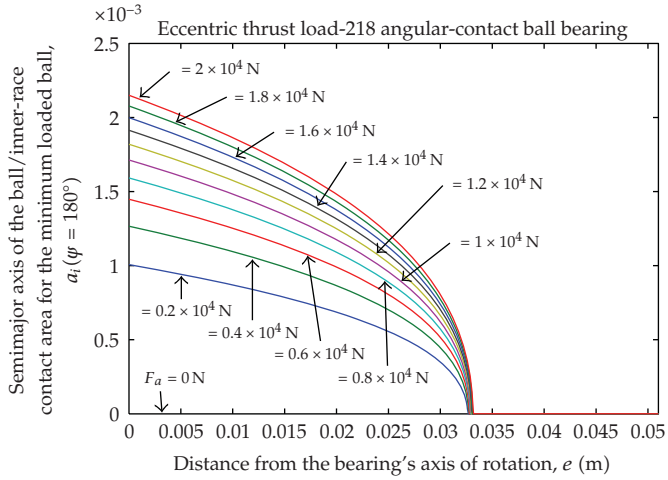


Figure 43: Semimajor axis of the ball/inner-race contact area for the minimum loaded ball, $a_i(\psi = 180°)$, as a function of lever arm, e.

e: Eccentricity of loading, m
E: Modulus of elasticity, N/m^2
E': Effective elastic modulus, N/m^2
$\mathbf{E}$: Elliptic integral of second kind
f: Raceway groove radius $\div D$
F: Applied load, N
k: a/b
K: Load-deflection factor, N/m$^{3/2}$
$\mathbf{K}$: Elliptic integral of first kind
M: eF_a
P_d: Diametral clearance, m
P_e: Free endplay, m

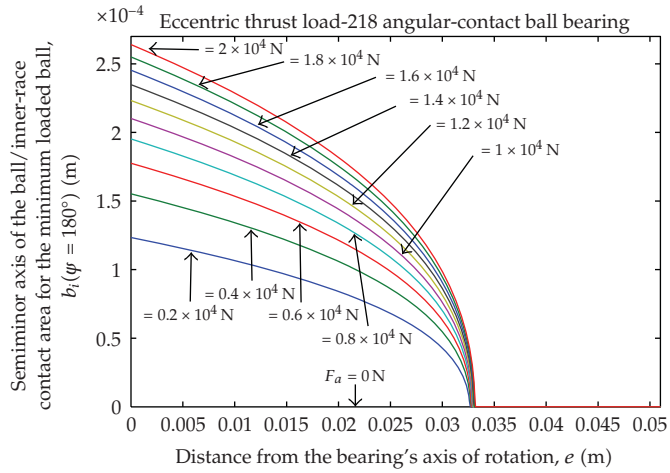


Figure 44: Semiminor axis of the ball/inner-race contact area for the minimum loaded ball, $b_i(\psi = 180°)$, as a function of lever arm, e.

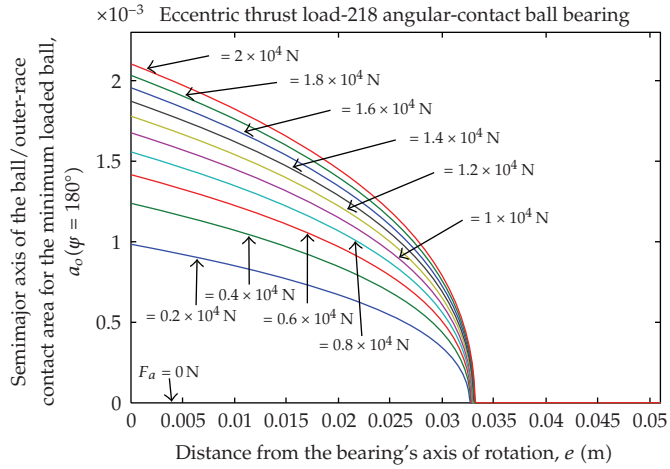


Figure 45: Semimajor axis of the ball/outer-race contact area for the minimum loaded ball, $a_o(\psi = 180°)$, as a function of lever arm, e.

Q: Ball-raceway normal load, N
r: Raceway groove curvature radius, solids curvature radius, m
s: Distance between loci of inner and outer raceway groove curvature centers, m
R: Curvature radius, radius of locus of raceway groove curvature centers, m
Z: Number of rolling elements
β: Contact angle, rad, ^{o}
β_f: Free-contact angle, rad, ^{o}
γ: $D \cos \beta / d_e$
Γ: Curvature difference
δ: Deflection or contact deformation, m
$\Delta\psi$: Angular spacing between rolling elements, rad, ^{o}
ε: Load distribution factor

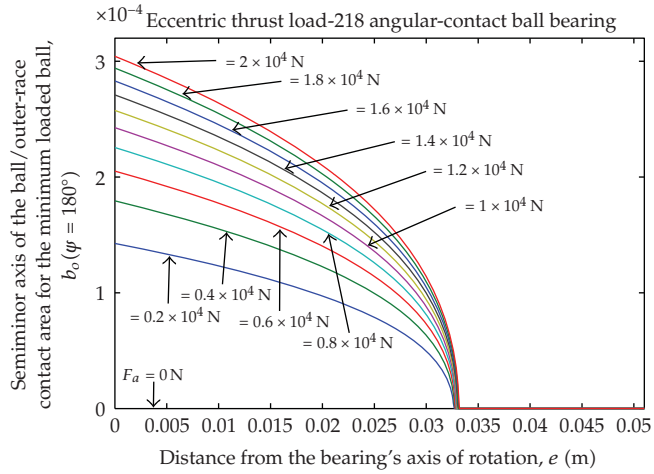


Figure 46: Semiminor axis of the ball/outer-race contact area for the minimum loaded ball, $b_o(\psi = 180°)$, as a function of lever arm, e.

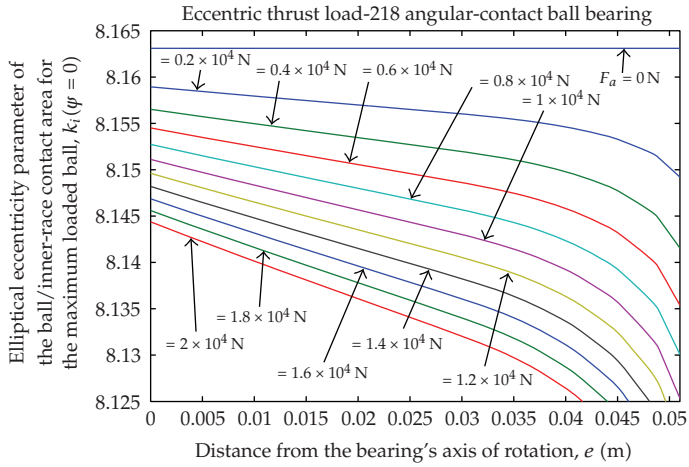


Figure 47: Elliptical eccentricity parameter of the ball/inner-race contact area for the maximum loaded ball, $k_i(\psi = 0)$, as a function of lever arm, e.

θ: Bearing misalignment angle, rad, ^{o}
υ: Poisson's ratio
φ: Auxiliary angle
ψ: Azimuth angle, rad, $^\circ$.

Subscripts:

a refers to solid a or axial direction.
b refers to solid b.
x, y refers to coordinate system.
i refers to inner raceway.

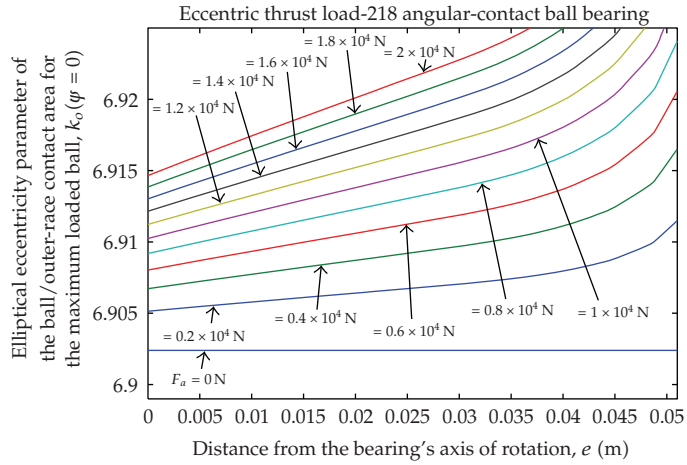


Figure 48: Elliptical eccentricity parameter of the ball/outer-race contact area for the maximum loaded ball, $k_o(\psi = 0)$, as a function of lever arm, e.

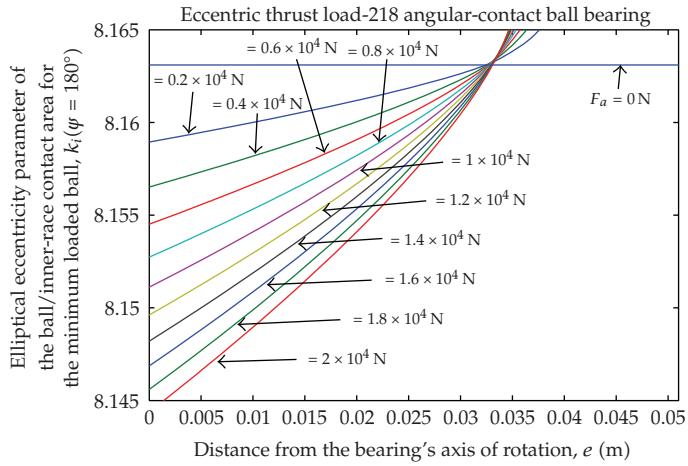


Figure 49: Elliptical eccentricity parameter of the ball/inner-race contact area for the minimum loaded ball, $k_i(\psi = 180°)$, as a function of lever arm, e.

j refers to rolling-element position.
n refers to direction collinear with normal load, integer number.
o refers to outer raceway.
t refers to total axial deformation.

Acknowledgments

The author thanks the financial support provided by the Brazilian Institute for Space Research (INPE), the Brazilian Scientific and Technological Development Council (CNPq), The State of

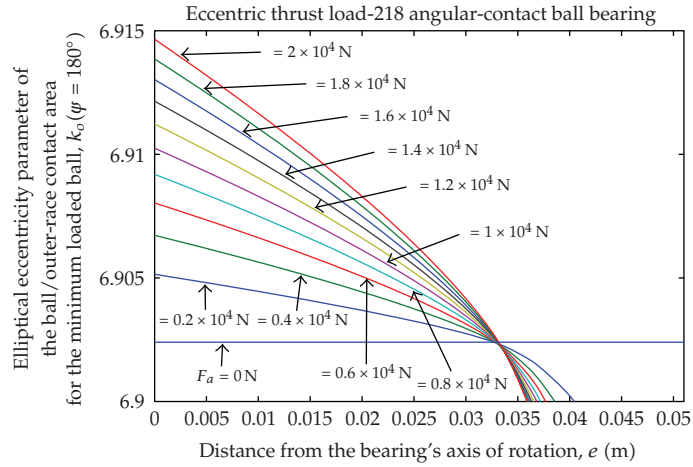


Figure 50: Elliptical eccentricity parameter of the ball/outer-race contact area for the minimum loaded ball, $k_o(\psi = 180°)$, as a function of lever arm, e.

São Paulo Research (FAPESP), and The Coordination of Development of Higher-Level Staff (CAPES). And the author thanks also José Pelógia da Silva for doing Figures 2 up to 7.

References

[1] R. Stribeck, "Ball bearings for various loads," *Transactions of the ASME*, vol. 29, pp. 420–463, 1907.
[2] H. Sjoväll, "The load distribution within ball and roller bearings under given external radial and axial load," *Teknisk Tidskrift, Mekaniska*, p. 9, 1933.
[3] A. Jones, *Analysis of Stresses and Deflections*, New Departure Engineering Data, Bristol, Conn, USA, 1946.
[4] J. Rumbarger, "Thrust bearings with eccentric loads," *Machine Design*, 1962.
[5] T. Harris, *Rolling Bearing Analysis*, John Wiley & Sons, New York, NY, USA, 4th edition, 2001.
[6] M. C. Ricci, "Ball bearings subjected to a variable eccentric thrust load," in *Booklet of Abstracts of the 8th Brazilian Conference on Dynamics, Control and Applications (DINCON '09)*, Bauru, Brazil, May 2009.
[7] M. C. Ricci, "Ball bearings subjected to a variable eccentric thrust load," in *Proceedings of the 8th Brazilian Conference on Dynamics, Control and Applications (DINCON '09)*, Bauru, Brazil, May 2009.
[8] M. C. Ricci, "Internal loading distribution in statically loaded ball bearings," in *Program and Abstracts of the 1st International Conference on Computational Contact Mechanics (ICCCM '09)*, pp. 21–22, Lecce, Italy, September 2009.
[9] M. C. Ricci, "Internal loading distribution in statically loaded ball bearings subjected to a combined radial and thrust load, including the effects of temperature and fit," in *Proceedings of World Academy of Science, Engineering and Technology (WCSET '09)*, vol. 57, Amsterdam, The Netherlands, September 2009.
[10] M. C. Ricci, "Internal loading distribution in statically loaded ball bearings subjected to a combined radial and thrust load," in *Book of Abstracts of the 6th International Congress of Croatian Society of Mechanics (ICCSM '09)*, p. 163, Dubrovnik, Croatia, September 2009.
[11] M. C. Ricci, "Internal loading distribution in statically loaded ball bearings subjected to a combined radial and thrust load," in *Proceedings of the 6th International Congress of Croatian Society of Mechanics (ICCSM '09)*, Dubrovnik, Croatia, September 2009.
[12] M. C. Ricci, "Internal loading distribution in statically loaded ball bearings subjected to a combined radial, thrust, and moment load," in *Abstracts of the 60th International Astronautical Congress*, Daejeon, South Korea, October 2009.
[13] M. C. Ricci, "Internal loading distribution in statically loaded ball bearings subjected to a combined radial, thrust, and moment load," in *Proceedings of the 60th International Astronautical Congress*, p. 163, Daejeon, South Korea, October 2009.

[14] M. C. Ricci, "Internal loading distribution in statically loaded ball bearings subjected to a combined radial, thrust, and moment load, including the effects of temperature and fit," in *Proceedings of the 11th Pan-American Congress of Applied Mechanics*, Foz do Iguaçu, Brazil, January 2010.
[15] B. J. Hamrock and W. J. Anderson, "Arched-outer-race ball-bearing considering centrifugal forces," NASA Report TN D-6765, NASA, Moffett Field, Calif, USA, 1972.
[16] H. Hertz, *On the contact of elastic solids and on hardness, in: D.E. Jones, G.A. Schott (Eds.)*, Miscellaneous Papers by Heinrich Hertz, Macmillan, London, 1896.
[17] B. J. Hamrock and W. J. Anderson, "Rolling-element bearings," NASA Report RP 1105, NASA, Moffett Field, Calif, USA, 1983.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 384071, 16 pages
doi:10.1155/2009/384071

Research Article

The Determination of the Velocities after Impact for the Constrained Bar Problem

André Fenili,[1] Luiz Carlos Gadelha de Souza,[2] and Bernd Schäfer[3]

[1] *Center for Engineering, Modeling and Applied Social Sciences (CECS), Federal University of ABC (UFABC), Aerospace Engineering, Av. dos Estados, 5001. Bloco B/Sala 936, 09210-580 Santo André, SP, Brazil*
[2] *National Institute for Space Research—INPE, Av. dos Astronautas, 1758, 12201-940 São José dos Campos, SP, Brazil*
[3] *German Aerospace Center—DLR, Institute of Robotics and Mechatronics, 82234 Wessling, Germany*

Correspondence should be addressed to Luiz Carlos Gadelha de Souza, gadelha@dem.inpe.br

Received 27 May 2009; Accepted 23 November 2009

Recommended by Antonio Prado

A simple mathematical model for a constrained robotic manipulator is investigated. Besides the fact that this model is relatively simple, all the features present in more complex problems are similar to the ones analyzed here. The fully plastic impact is considered in this paper. Expressions for the velocities of the colliding bodies after impact are developed. These expressions are important in the numerical integration of the governing equations of motion when one must exchange the set of unconstrained equations for the set of constrained equation. The theory presented in this work can be applied to problems in which robots have to follow some prescribed patterns or trajectories when in contact with the environment. It can also de applied to problems in which robotic manipulators must handle payloads.

1. Introduction

There are several ways to deal with the problem of interaction between bodies. Impact dynamics and continuous contact between bodies can both be included in the mathematical model of the constrained problem, or just one of these effects can be considered. It depends, obviously, on the characteristics of the studied problem.

The investigations about the contact between bodies include (at least) two different kind of analysis [1]: one associated with the beginning of contact and one associated with its termination. In the first analysis, the distance between the bodies must be checked in order to know when contact occurs; in the second analysis, once the contact is established, the reaction

(normal; compression) force between the bodies must be checked. In the second analysis, contact finishes when the contact force is equal to zero.

One of the hardest parts in the study of contact problems involves the different models that must be developed for contact and noncontact situations and the switching between these models when integrating the equations of motion [2, 3]. The unconstrained problem and the constrained problem do not have the same number of degrees of freedom. Dynamic systems when constrained have less degrees of freedom than when unconstrained.

The transition between constrained and unconstrained motions is sometimes called contact (including impact) and sometimes called just impact (mostly when the bodies separate after the collision). When contact occurs, the new velocities of the bodies involved must be known in order to generate the initial conditions to the second part (constrained problem) of the numerical integration. In the constrained problem, the concept of coefficient of restitution is very important [4].

2. Geometric Model of the System and Governing Equations of Motion

The problem discussed here is depicted in Figure 1. According to this figure, in a part of its trajectory, the free end of the bar moves along the constraint represented by the mass named m_w. All the movements occur in the horizontal plane. When contact occurs, impact and bouncing are also allowed to occur.

The mass in which the rigid bar is pivoted (m_s) oscillates when excited by the movement of the bar (free and constrained). In the axis Z, passing through the connection between the bar and m_s (perpendicular to the paper sheet), there is a prescribed moment, M_θ, acting to turn the bar.

The dashed lines represent the position of the masses in which the springs and dampers are free of forces. The dotted line represents the position from which one starts to count the angular displacement, θ.

In physical terms, this system may represent a robot with a translational joint and a rotational joint; m_w can be thought as an obstructing wall on the robot's trajectory (or some object this robot must handle or interact with), and M_θ can be thought as an external torque provided by a dc motor.

According to [5], the constrained governing equations of motion for this system are given by

$$\begin{aligned}
&(m_b + m_s)\ddot{y}_s + c_s\dot{y}_s + k_s y_s - m_b\ d_{Acmb}\dot{\theta}^2 \sin\theta + m_b d_{Acmb}\ddot{\theta}\cos\theta + F_N = 0,\\
&m_w\ddot{y}_w + c_w\dot{y}_w + k_w y_w - F_N = 0,\\
&\left(I_{b,\mathrm{cm}} + m_b d_{Acmb}^2\right)\ddot{\theta} + m_b d_{Acmb}\ddot{y}_s\cos\theta + F_N\ell\cos\theta = M_\theta,
\end{aligned} \tag{2.1}$$

and the constraint condition is given by

$$d - y_s + y_w - \ell\sin\theta = 0, \tag{2.2}$$

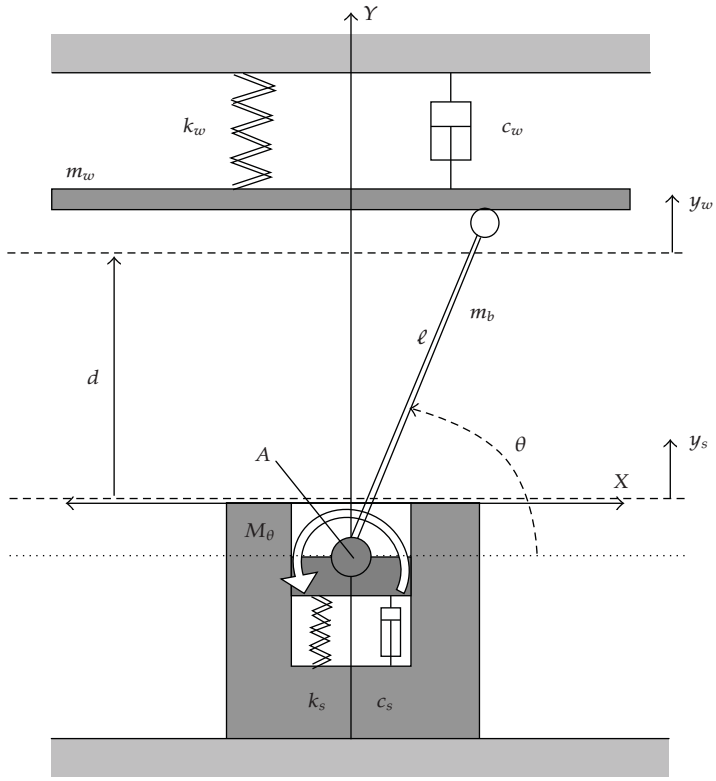


Figure 1: Oscillating constrained bar.

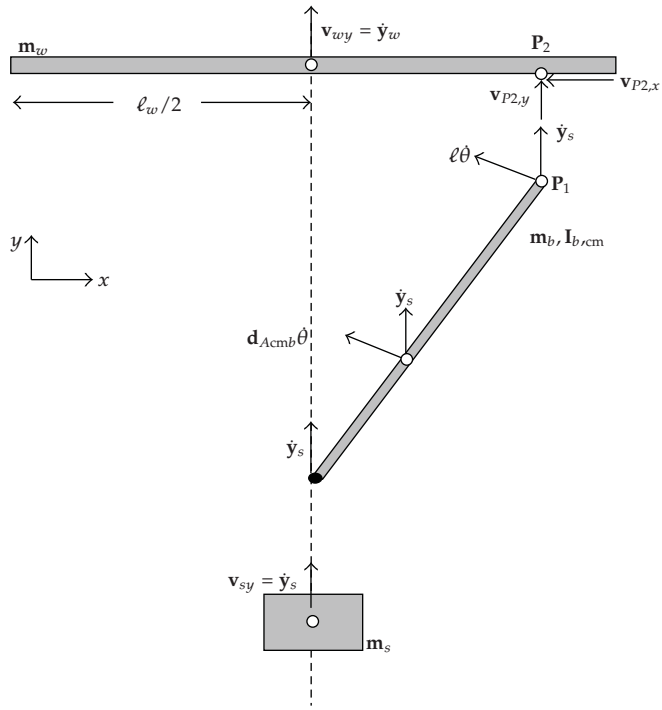


Figure 2: Velocities.

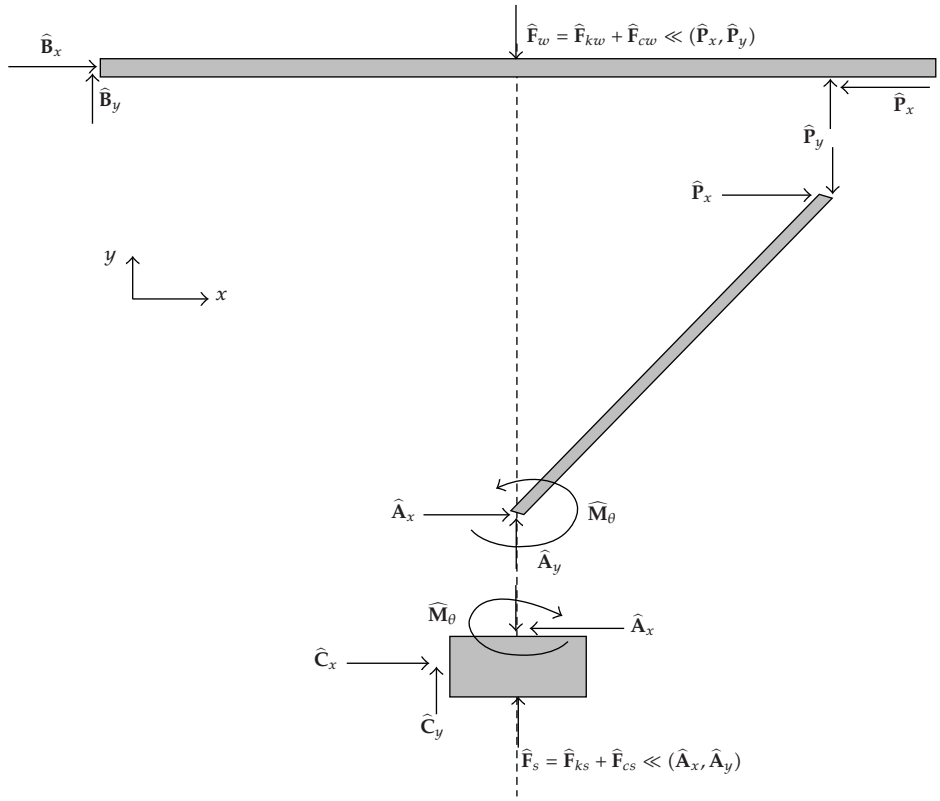


Figure 3: Impulses.

where $I_{b,\mathrm{cm}}$ represents the bar moment of inertia around its center of mass, m_b represents the mass of the bar, d_{Acmb} represents the distance from A to the cm of the bar, c_w represents the damping coefficient of m_w, c_s represents the damping coefficient associated with mass m_s, k_w represents the stiffness coefficient of mass m_w, k_s represents the stiffness coefficient associated with m_s, and F_N represents the amplitude of the normal force. It is assumed the there are no friction forces involved and ℓ represents the total length of the bar.

Equations (2.1) are the equations of motion for y_s, y_w, and θ. Equation (2.2) is an additional relationship between the generalized coordinates y_s, θ and y_w when contact occurs. Equations from (2.1) to (2.2) provide four equations and four unknowns (y_s, θ, y_w, and F_N) considering the constrained problem and three equations and three unknowns (y_s, θ, and y_w) considering the unconstrained problem. In the unconstrained case, (2.2) does not apply and $F_N = 0$.

3. The Contact Case

In contact, for this problem, there is the loss of one degree of freedom. In other words, one of the variables is dependent on all the others. The best choice is the elimination of the generalized coordinate y_w, which is not always present into the system represented by the

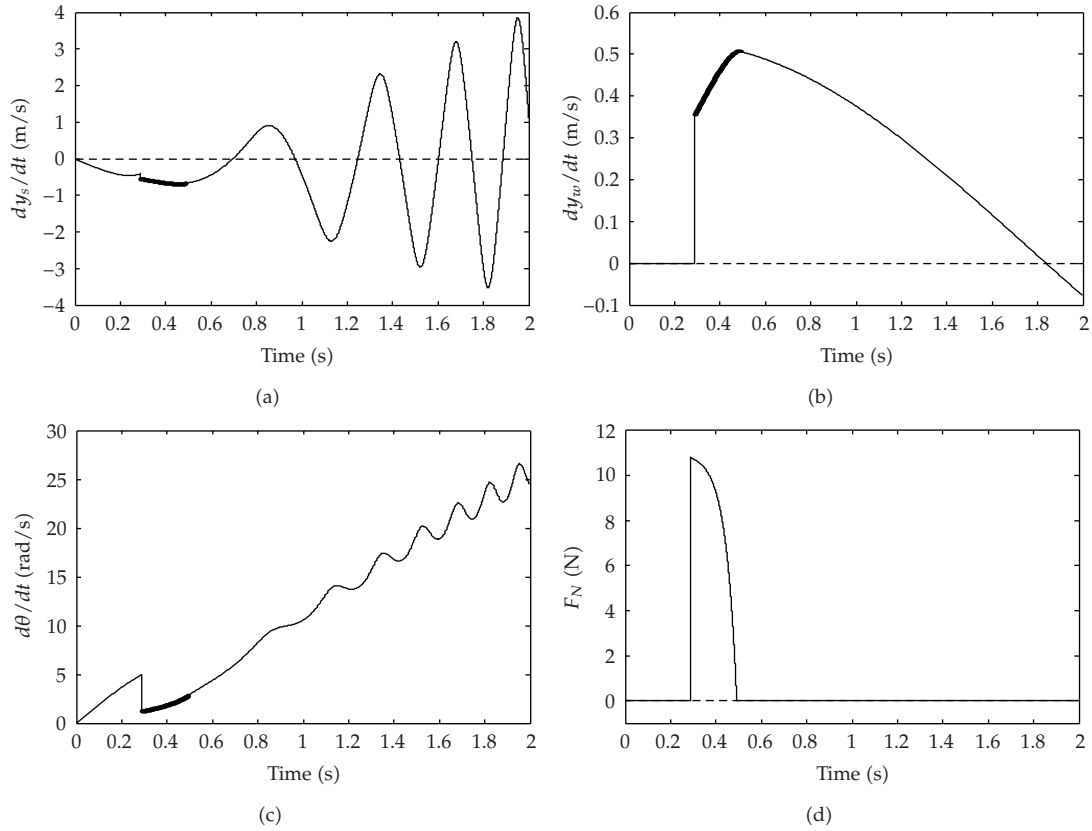


Figure 4: $\dot{y}_s$, $\dot{y}_w$, $\dot{\theta}$, and F_N considering $k_w = 10\,\text{Nm}$.

oscillating bar [6]. The new set of equations [5] is given by

$$\begin{aligned}
&\ddot{y}_s + \frac{1}{a_1 m_t + a_3 \cos^2\theta}\big(a_1(c_s + c_w)\dot{y}_s + a_1(k_s + k_w)y_s + a_1 c_w \ell\dot{\theta}\cos\theta + a_1 k_w \ell \sin\theta - a_1 a_2 \dot{\theta}^2 \sin\theta \\
&\quad - m_b c_w \ell^2 d_{Acmb}\dot{\theta}\cos^3\theta - m_b k_w \ell^2 d_{Acmb}\sin\theta\cos^2\theta \\
&\quad + m_b k_w d\ell d_{Acmb}\cos^2\theta + \ell(m_w \ell c_s - m_b d_{Acmb} c_w)\dot{y}_s\cos^2\theta \\
&\quad + \ell(m_w \ell k_s - m_b d_{Acmb} k_w)y_s\cos^2\theta - a_1 k_w d\big) = -\frac{a_2\cos\theta}{a_1 m_t + a_3\cos^2\theta} M_\theta, \\
&\ddot{\theta} + \frac{1}{a_1 m_t + a_3\cos^2\theta}\big(c_w \ell(m_t\ell - a_2)\dot{\theta}\cos^2\theta + k_w \ell(m_t\ell - a_2)\sin\theta\cos\theta - k_w d(m_t\ell - a_2)\cos\theta \\
&\quad + (a_2 m_b d_{Acmb} - m_w \ell(m_t\ell - a_2))\dot{\theta}^2\sin\theta\cos\theta + (k_w(m_t\ell - a_2) - a_2 k_s)y_s\cos\theta \\
&\quad + (c_w(m_t\ell - a_2) - a_2 c_s)\dot{y}_s\cos\theta\big) = \frac{m_t}{a_1 m_t + a_3\cos^2\theta} M_\theta.
\end{aligned} \tag{3.1}$$

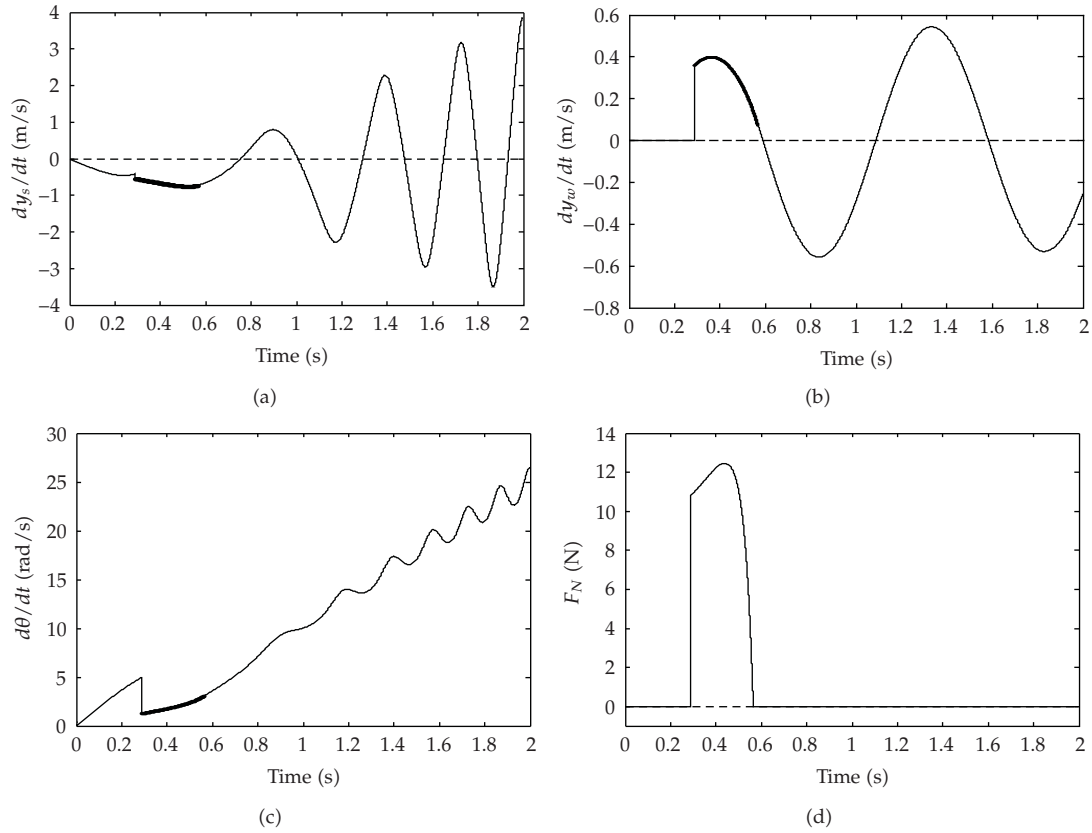


Figure 5: $\dot{y}_s$, $\dot{y}_w$, $\dot{\theta}$, and F_N considering $k_w = 400\,\text{Nm}$.

The fully plastic impact case is considered here for the calculation of the velocities immediately after contact. Separation will take place when the normal force is zero.

As soon as these two variables are known, the remaining variable, y_w, is also known through (2.2). Equations (3.1) represent, respectively, the time behavior of the generalized coordinates y_s and θ during the contact condition. In [5], an analytical expression to the reaction force, F_N, is also presented.

4. The Determination of the Velocities after Contact (Impact)

The equations for the impact are formulated for point P (see Figure 2 for the representation of the velocities of the three bodies) where, for sake of clarity, it is distinguished between Point P_1 belonging to the wall and point P_2 belonging to the bar. Figure 3 shows the free body diagram for the three rigid bodies indicating not the forces at the points of connection or contact but rather indicating the equivalent linear impulses due to impact. All these quantities are marked with an overhead symbol "hat", for example, $\widehat{P}_x$, which is the linear impulse of the equivalent force P_x. The physical dimension is the same as the linear momentum, that is, $N \cdot s$, except for the angular impulse $\widehat{M}_\theta$ whose unit is Nm · s.

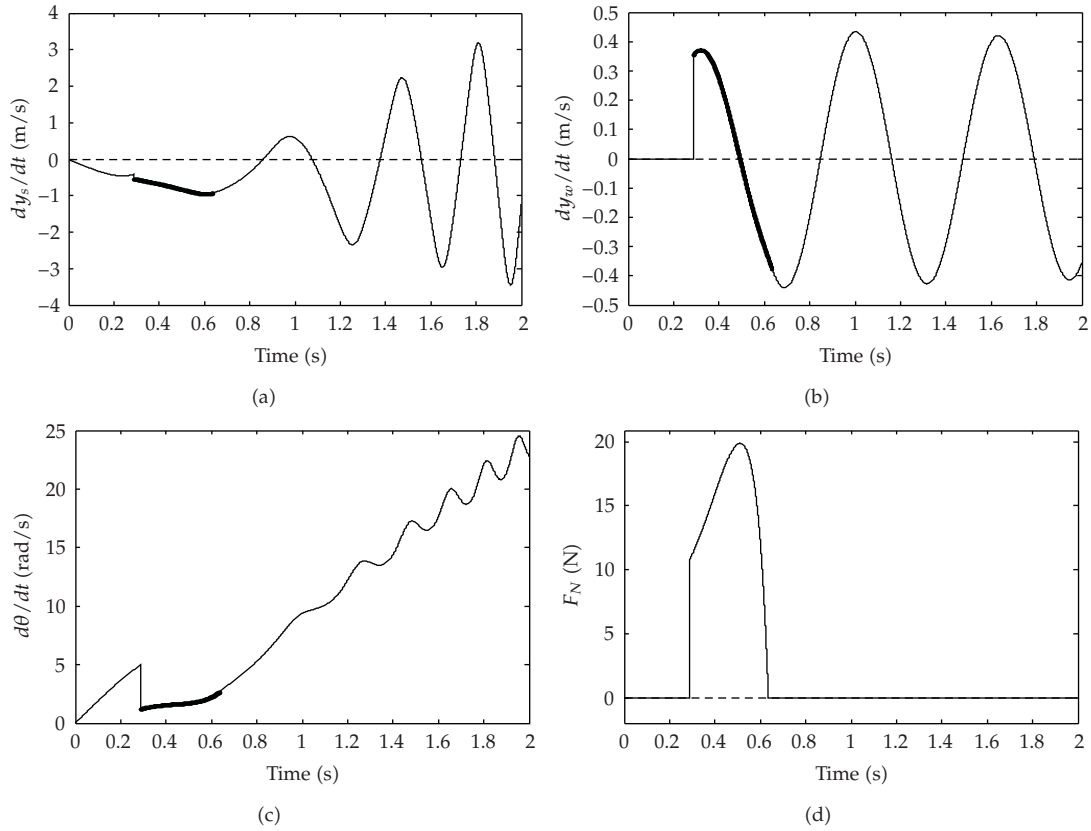


Figure 6: $\dot{y}_s$, $\dot{y}_w$, $\dot{\theta}$, and F_N considering $k_w = 1000\,\text{Nm}$.

For each of the three rigid bodies, we can formulate now the linear impulse/linear momentum equations in the two directions x and y. Additionally, for the rotating bodie(s), we have the equivalent angular impulse/angular momentum equation in z-direction, formulated w.r.t. to the respective centre of mass.

To better distinguish between velocities right before and right after impact, they are denoted with superscripts "+" (after) and "−" (before). Their two components in x- and y-directions are indicated by corresponding subscripts "x" and "y".

And, to be more general, it is also allowed initially for the rigid bodies with masses m_s and m_w to rotate as well. The respective angular velocities therefore will be denoted by ω with appropriate indices. Later, this additional degree of freedom will be kinematically constrained.

For the wall, it is obtained that

$$\begin{aligned} m_w v_{wy}^{+} - m_w v_{wy}^{-} &= \widehat{P}_y - \widehat{F}_w, \\ m_w v_{wx}^{+} - m_w v_{wx}^{-} &= -\widehat{P}_x + \widehat{B}_x, \\ I_w \omega_w^{+} - I_w \omega_w^{-} &= -\widehat{B}_y \frac{\ell_w}{2} + \widehat{P}_y \ell \sin\theta. \end{aligned} \tag{4.1}$$

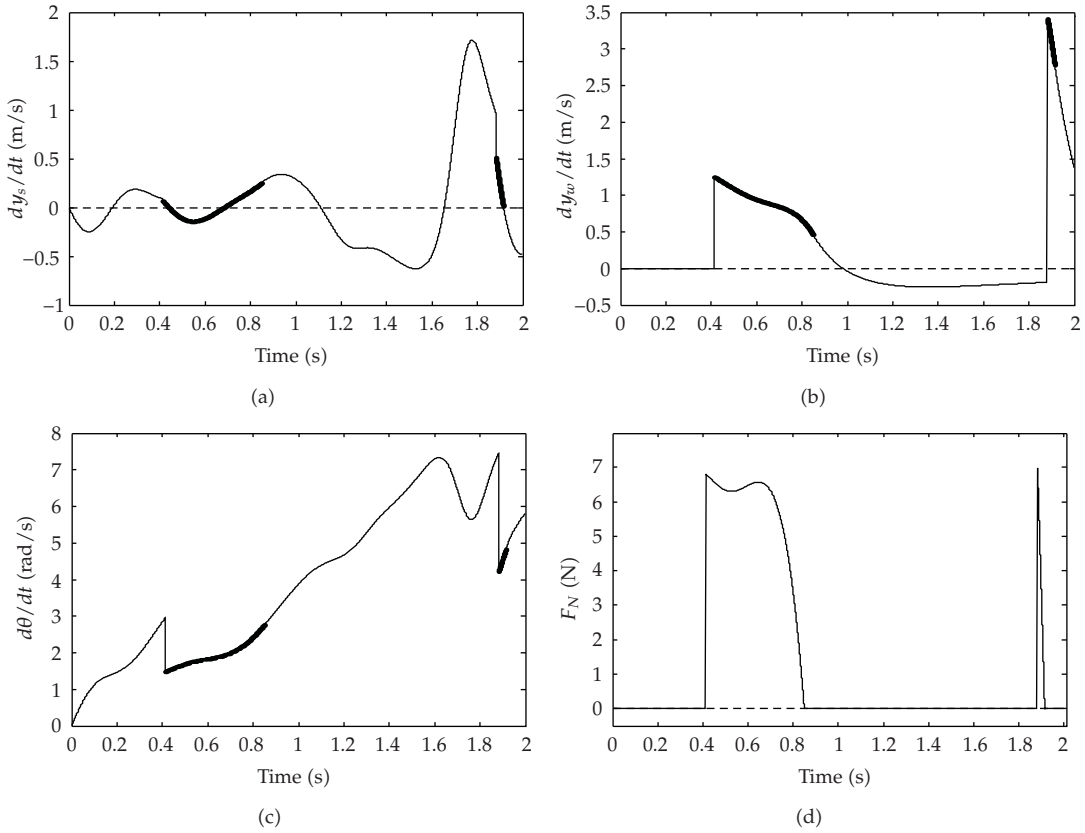


Figure 7: $\dot{y}_s$, $\dot{y}_w$, $\dot{\theta}$, and F_N considering $M_\theta = 5\text{Nm}$.

For the bar, it is obtained ($\omega_b \equiv \dot{\theta}$) that

$$\begin{aligned} m_b v_{by}^+ - m_b v_{by}^- &= \widehat{A}_y - \widehat{P}_y, \\ m_b v_{bx}^+ - m_b v_{bx}^- &= \widehat{A}_x + \widehat{P}_x, \\ I_{b,\mathrm{cm}}\omega_b^+ - I_{b,\mathrm{cm}}\omega_b^- &= \widehat{M}_\theta + \widehat{A}_x d_{Acmb}\sin\theta - \widehat{A}_y d_{Acmb}\cos\theta \\ &\quad - \widehat{P}_x(\ell - d_{Acmb})\sin\theta - \widehat{P}_y(\ell - d_{Acmb})\cos\theta. \end{aligned} \tag{4.2}$$

And, finally, for the lower rigid body with mass m_s, it is obtained that

$$\begin{aligned} m_s v_{sy}^+ - m_s v_{sy}^- &= -\widehat{A}_y + \widehat{C}_y + \widehat{F}_s, \\ m_s v_{sx}^+ - m_s v_{sx}^- &= -\widehat{A}_x + \widehat{C}_x, \\ I_s\omega_s^+ - I_s\omega_s^- &= -\widehat{M}_\theta + \widehat{A}_x b_{sy} - \widehat{C}_y b_{sx}, \end{aligned} \tag{4.3}$$

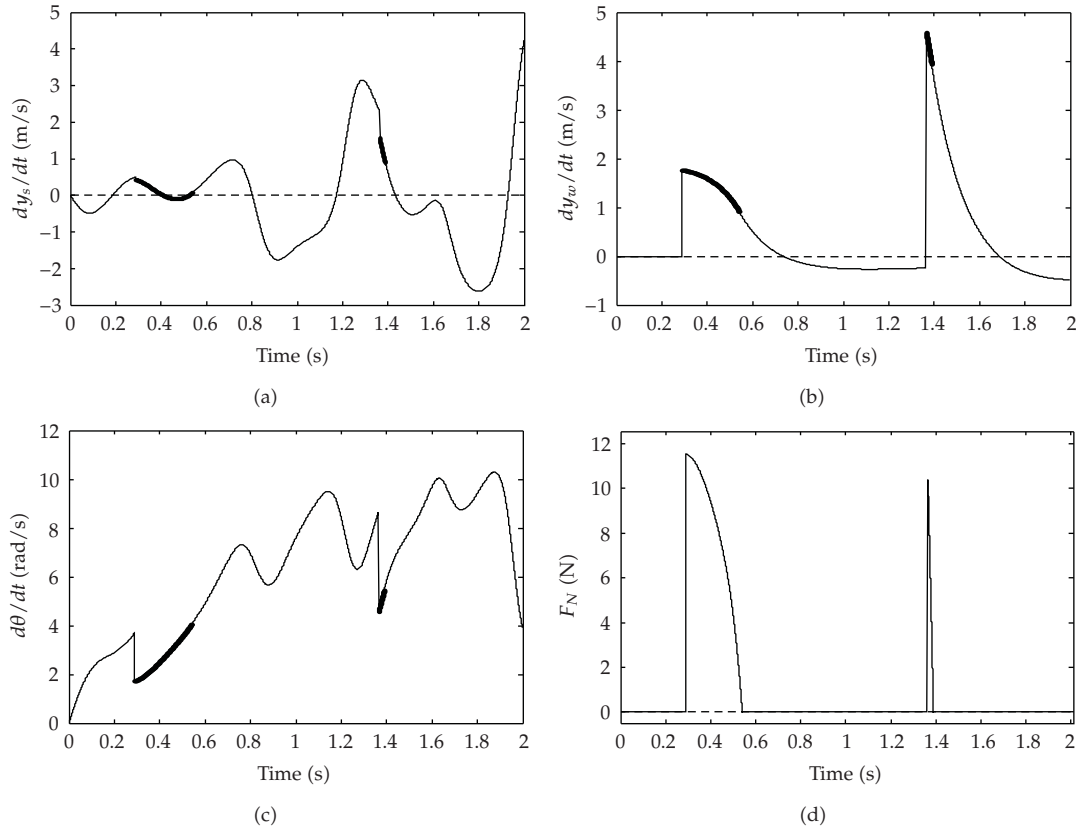


Figure 8: $\dot{y}_s$, $\dot{y}_w$, $\dot{\theta}$, and F_N considering $M_\theta = 10\,\text{Nm}$.

assuming that the directions of $\widehat{C}_x$ and $\widehat{F}_s$ are going through the center of mass. The geometric quantities b_{sx} and b_{sy}, not shown in Figure 3, denote the distances of the respective linear impulses measured from the center of mass.

These equations simplified if the following assumptions are made.

(1) The external two linear impulses $\widehat{F}_w$ and $\widehat{F}_s$, and the angular impulse $\widehat{M}_\theta$ are small compared with the internal impulses; therefore, they can be neglected.

(2) The rotational motion of both, the wall and the lower rigid body, is omitted; therefore, one has $\omega_w = 0$ and $\omega_s = 0$.

(3) The wall is allowed to move only in the vertical direction, as well as the lower rigid body; therefore, $v_{wx} = 0$ and $v_{sx} = 0$.

(4) The contact surface between the lower rigid body and the left or right vertical guiding surface (not shown in the figures) is assumed ideally smooth; therefore, $\widehat{C}_y = 0$.

(5) The contact zone between the free end of the bar and the wall surface is also assumed ideally smooth; therefore, $\widehat{P}_x = 0$. Otherwise, if this surface is rough, we have to account for an additional velocity relationship, for example, given by the definition of the coefficient of restitution in x-direction.

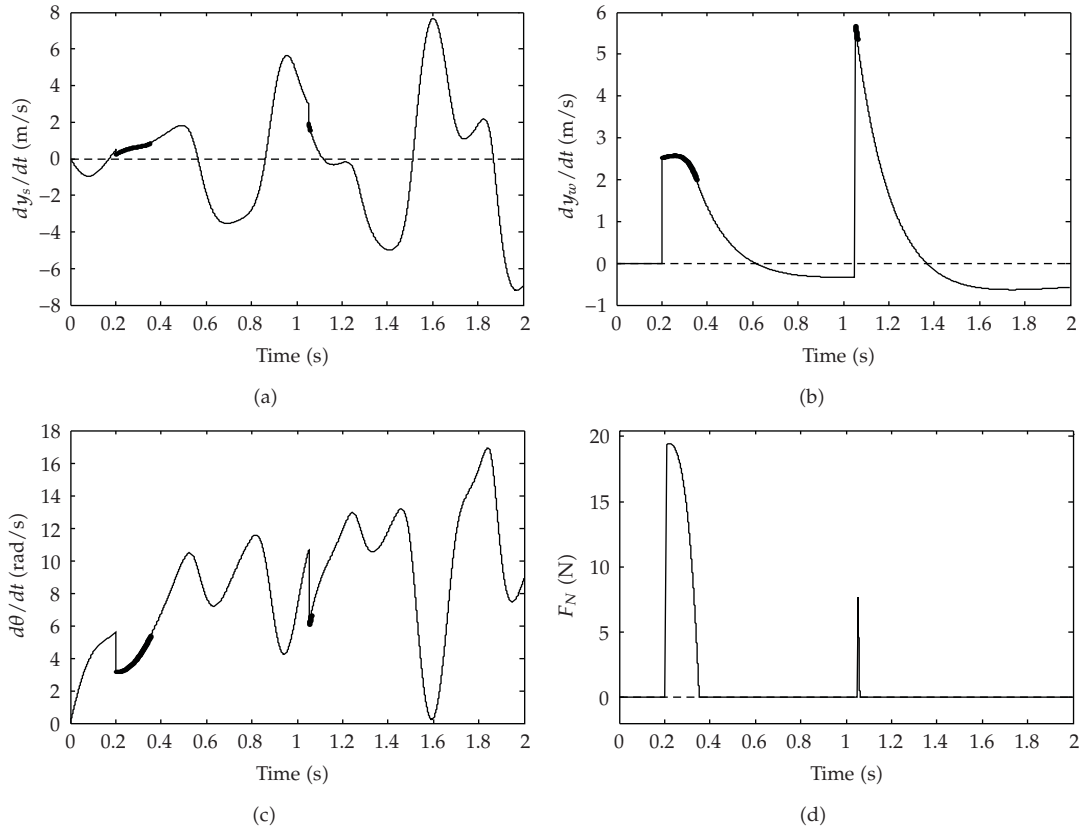


Figure 9: $\dot{y}_s$, $\dot{y}_w$, $\dot{\theta}$, and F_N considering $M_\theta = 20\,\text{Nm}$.

Applying these assumptions, the following set of equations is obtained:

$$m_w v_{wy}^{+} - m_w v_{wy}^{-} = \widehat{P}_y, \tag{4.4}$$

$$\widehat{B}_x = \widehat{P}_x = 0, \tag{4.5}$$

$$\widehat{B}_y = \frac{2\ell}{\ell_w} \sin\theta \cdot \widehat{P}_y \ , \tag{4.6}$$

$$m_b v_{by}^{+} - m_b v_{by}^{-} = \widehat{A}_y - \widehat{P}_y, \tag{4.7}$$

$$m_b v_{bx}^{+} - m_b v_{bx}^{-} = \widehat{A}_x + \widehat{P}_x = \widehat{A}_x, \tag{4.8}$$

$$I_{b,\text{cm}}\omega_b^{+} - I_{b,\text{cm}}\omega_b^{-} = \widehat{A}_x d_{Acmb} \sin\theta - \widehat{A}_y d_{Acmb} \cos\ \theta - \widehat{P}_y(\ell - d_{Acmb}) \cos\ \theta, \tag{4.9}$$

$$m_s v_{sy}^{+} - m_s v_{sy}^{-} = -\widehat{A}_y \ , \tag{4.10}$$

$$\widehat{C}_x = \widehat{A}_x, \tag{4.11}$$

$$b_{sy} = 0. \tag{4.12}$$

In order to calculate the velocities at the point of impact, P, only (4.4) and (4.7) to (4.10) are of interest. Additionally, it is needed to establish some kinematic relationships. For the bar center of mass, one has

$$\begin{aligned}\mathbf{v}_b &= \mathbf{v}_s + \boldsymbol{\omega}_b \times \mathbf{r}_{ASb} \\ &= \left(-\omega_b d_{Acmb} \sin\theta, v_{sy} + \omega_b d_{Acmb} \cos\ \theta\right)^T \\ &= \left(v_{bx}, v_{by}\right)^T,\end{aligned} \tag{4.13}$$

where the length of the vector $\mathbf{r}_{Acmb}$ is just $|\mathbf{r}_{Acmb}| = d_{Acmb}$. Equation (4.13) is valid for the velocity right before and after impact. For the free end of the bar, it is obtained equivalently

$$\begin{aligned}\mathbf{v}_{P1} &= \mathbf{v}_s + \boldsymbol{\omega}_b \times \mathbf{r}_{AP1} \\ &= \left(-\omega_b \ell \sin\theta, v_{sy} + \omega_b \ell \cos\theta\right)^T \\ &= \left(v_{P1x}, v_{P1y}\right)^T\end{aligned} \tag{4.14}$$

with $|\mathbf{r}_{AP1}| = \ell$. In the same way as (4.13), equation (4.14) is valid for the velocity right before impact and right after. During impact, one has the additional equation, which relates the velocities before and after impact at point P, in the direction normal to the contact surface, that is, in y-direction:

$$\varepsilon_y = -\frac{v_{P1y}^+ - v_{P2y}^+}{v_{P1y}^- - v_{P2y}^-} \tag{4.15}$$

with

$$\begin{aligned}v_{P1y}^- &= v_{sy}^- + \omega_b^- \ell \cos\theta, \\ v_{P1y}^+ &= v_{sy}^+ + \omega_b^+ \ell \cos\theta, \\ v_{P2y}^- &= v_{wy}^- = y_w^-, \\ v_{P2y}^+ &= v_{wy}^+ = y_w^+.\end{aligned} \tag{4.16}$$

In the following, it is assumed that there is a fully plastic impact, that is, the impacting bodies maintain steady contact as far as the contact force is repulsive (otherwise, they will separate). This leaves $\varepsilon_y = 0$, and hence

$$v_{P1y}^+ = v_{P2y}^+ \tag{4.17}$$

or

$$v_{wy}^+ = v_{sy}^+ + \omega_b^+ \ell \cos\theta = v_{by}^+ + \omega_b^+ (\ell - d_{Acmb}) \cos\theta. \tag{4.18}$$

With these equations, it is possible to calculate all the velocities right after impact, given the velocities before impact. Additionally, but not needed here, it is also possible to calculate the appropriate linear impulses. To summarize, one has the following eight equations to determine all the five velocities right after impact $(v_{sy}^{+}, v_{bx}^{+}, v_{by}^{+}, v_{wy}^{+}, \omega_b^{+})$, as well as the impulses $(\widehat{A}_x, \widehat{A}_y, \widehat{P}_y)$:

$$m_w v_{wy}^{+} - m_w v_{wy}^{-} = \widehat{P}_y, \tag{4.19}$$

$$m_b v_{by}^{+} - m_b v_{by}^{-} = \widehat{A}_y - \widehat{P}_y, \tag{4.20}$$

$$m_b v_{bx}^{+} - m_b v_{bx}^{-} = \widehat{A}_x, \tag{4.21}$$

$$I_{b,\mathrm{cm}}\omega_b^{+} - I_{b,\mathrm{cm}}\omega_b^{-} = \widehat{A}_x d_{Acmb}\sin\theta - \widehat{A}_y d_{Acmb}\cos\theta - \widehat{P}_y(\ell - d_{Acmb})\cos\theta, \tag{4.22}$$

$$m_s v_{sy}^{+} - m_s v_{sy}^{-} = -\widehat{A}_y, \tag{4.23}$$

$$v_{bx}^{+} = -\omega_b^{+} d_{Acmb}\sin\theta, \tag{4.24}$$

$$v_{by}^{+} = v_{sy}^{+} + \omega_b^{+} d_{Acmb}\cos\theta, \tag{4.25}$$

$$v_{wy}^{+} = v_{sy}^{+} + \omega_b^{+}\ell\cos\theta = v_{by}^{+} + \omega_b^{+}(\ell - d_{Acmb})\cos\theta. \tag{4.26}$$

Initially, all the impulses are obtained. $\widehat{A}_y$ is simply obtained from (4.23) or by adding the two (4.19) and (4.20), giving

$$\widehat{A}_y = -\left(m_s v_{sy}^{+} - m_s v_{sy}^{-}\right) = m_b v_{by}^{+} - m_b v_{by}^{-} + m_w v_{wy}^{+} - m_w v_{wy}^{-}. \tag{4.27}$$

$\widehat{A}_x$ also goes simply with (4.21),

$$\widehat{A}_x = m_b v_{bx}^{+} - m_b v_{bx}^{-} \tag{4.28}$$

and $\widehat{P}_y$ is simply obtained directly from (4.19) or by adding (4.20) and (4.23)

$$\widehat{P}_y = m_w v_{wy}^{+} - m_w v_{wy}^{-} = -\left(m_b v_{by}^{+} - m_b v_{by}^{-}\right) - \left(m_s v_{sy}^{+} - m_s v_{sy}^{-}\right). \tag{4.29}$$

Comparing (4.27) with (4.29), it is observed that both equations yield the same result for the linear momenta before and after impact. To determine now the velocities right after impact, one can rely on (4.22), (4.24), (4.25), (4.26), and (4.27) (or (4.29), which is the same).

Replacing v_{bx}^{+}, v_{by}^{+} and v_{wy}^{+}, one arrives at the two equations for the unknown velocities v_{sy}^{+} and ω_b^{+}:

$$\begin{aligned}
&(m_s + m_b + m_w)v_{sy}^{+} = m_s v_{sy}^{-} + m_b v_{by}^{-} + m_w v_{wy}^{-} - \omega_b^{+}(m_b d_{Acmb} + m_w \ell)\cos\theta,\\
&\left[I_{b,\mathrm{cm}} + m_b d_{Acmb}^2 \sin^2\theta + m_w \ell(\ell - d_{Acmb})\cos^2\theta\right]\omega_b^{+}\\
&\quad = v_{sy}^{+}[m_s d_{Acmb} - m_w(\ell - d_{Acmb})]\cos\theta + I_{b,\mathrm{cm}}\omega_b^{-} - m_s v_{sy}^{-} d_{Acmb}\cos\theta\\
&\quad - m_b v_{bx}^{-} d_{Acmb}\sin\theta + m_w v_{wy}^{-}(\ell - d_{Acmb})\cos\theta.
\end{aligned} \tag{4.30}$$

And with $v_{by}^{-} = v_{sy}^{-} + \omega_b^{-} d_{\mathrm{Acmb}}\cos\theta$, and $v_{bx}^{-} = -\omega_b^{-} d_{\mathrm{Acmb}}\sin\theta$, these equations can finally be expressed by means of the independent velocities, v_{sy}^{-}, v_{wy}^{-}, and ω_b^{-}, right before impact:

$$\begin{aligned}
&(m_s + m_b + m_w)v_{sy}^{+} = (m_s + m_b)v_{sy}^{-} + m_w v_{wy}^{-} + m_b \omega_b^{-} d_{Acmb}\cos\theta - \omega_b^{+}(m_b d_{Acmb} + m_w \ell)\cos\theta.\\
&\left[I_{b,\mathrm{cm}} + m_b d_{Acmb}^2 \sin^2\theta + m_w \ell(\ell - d_{Acmb})\cos^2\theta\right]\omega_b^{+}\\
&\quad = v_{sy}^{+}[m_s d_{Acmb} - m_w(\ell - d_{Acmb})]\cos\theta + \left(I_{b,\mathrm{cm}} + m_b d_{Acmb}^2 \sin^2\theta\right)\omega_b^{-}\\
&\quad - m_s v_{sy}^{-} d_{Acmb}\cos\theta + m_w v_{wy}^{-}(\ell - d_{Acmb})\cos\theta
\end{aligned} \tag{4.31}$$

With the abbreviations

$$\begin{aligned}
m_{\mathrm{tot}} &= m_s + m_b + m_w,\\
I_{\mathrm{tot}} &= I_{b,\mathrm{cm}} + m_b d_{Acmb}^2 \sin^2\theta + m_w \ell(\ell - d_{Acmb})\cos^2\theta,\\
r_1 &= (m_s + m_b)v_{sy}^{-} + m_w v_{wy}^{-} + m_b \omega_b^{-} d_{Acmb}\cos\theta,\\
r_2 &= -m_s v_{sy}^{-} d_{Acmb}\cos\theta + m_w v_{wy}^{-}(\ell - d_{Acmb})\cos\theta + \left(I_{b,\mathrm{cm}} + m_b d_{Acmb}^2 \sin^2\theta\right)\omega_b^{-},\\
\alpha_1 &= (m_b d_{Acmb} + m_w \ell)\cos\theta,\\
\alpha_2 &= [m_s d_{Acmb} - m_w(\ell - d_{Acmb})]\cos\theta,
\end{aligned} \tag{4.32}$$

one finally obtains

$$\begin{aligned}
v_{sy}^{+} &= \frac{r_1 I_{\mathrm{tot}} - r_2 \alpha_1}{\alpha_1 \alpha_2 + m_{\mathrm{tot}} I_{\mathrm{tot}}},\\
\omega_b^{+} &= \frac{r_1 \alpha_2 + r_2 m_{\mathrm{tot}}}{\alpha_1 \alpha_2 + m_{\mathrm{tot}} I_{\mathrm{tot}}}.
\end{aligned} \tag{4.33}$$

The denominator of these two equations then is written as

$$\alpha_1\alpha_2 + m_{\text{tot}}I_{\text{tot}} = m_s m_b d_{Acmb}^2 + m_{\text{tot}}I_{b,\text{cm}} + m_b m_w\left[\ell(\ell - 2d_{Acmb})\cos^2\theta + d_{Acmb}^2\right] + m_s m_w \ell^2\cos^2\theta + m_b^2 d_{Acmb}^2\sin^2\theta. \tag{4.34}$$

In order to check (4.33), one case is investigated; that is, for $\theta = 90°$, we should maintain the simple translational impact between the combined rigid body consisting of the two masses m_s and m_b and the wall with mass m_w. For the fully plastic impact, one then obtains from (4.33) with $\alpha_1 = 0$ and $\alpha_2 = 0$:

$$v_{sy}^+(\theta = 90°) = \frac{r_1}{m_{\text{tot}}} = \frac{(m_s + m_b)v_{sy}^- + m_w v_{wy}^-}{m_{\text{tot}}},$$
$$\omega_b^+(\theta = 90°) = \frac{r_2}{I_{\text{tot}}} = \frac{I_{b,\text{cm}} + m_b d_{Acmb}^2}{I_{\text{tot}}}\omega_b^-, \tag{4.35}$$

where the first equation for the translational motion coincides with the result governed from simple impact of two rigid bodies.

5. Numerical Results

The values for the parameters used in the numerical simulations that follow are presented in Tables 1 and 2. The time step considered in the integration of the governing equations of motion is kept constant and equal to 0.0001 s. The fourth-order Runge-Kutta is the numerical integrator used. Two different classes of simulation are investigated.

The constant torque (with different amplitudes) was chosen because it is the simplest one, and in order to make the bar rotate always in the same direction and fulfill 360°. Any other kind of excitation (e.g., like a sinusoidal one with maximum amplitude of 180º, for instance) can be chosen without problem. In the simulation runs, the motion of the bar starts always in its horizontal position to the right, that is, with $\theta = 0°$.

The very beginning of contact is considered here as a fully plastic impact with impact time $\Delta t \approx 0$ and with $e = 0$, where e represents the coefficient of restitution. Contact finishes when $F_N = 0$. No friction or contact is considered, up to this point of the investigation, between m_s and the guide it slides through or between m_b and m_w.

5.1. Considering Different Values of k_w

When first contact takes place, m_w is at rest. The second contact (only shown here for the simulations varying M_θ) will happen with m_w presenting some velocity. The bar is able to develop many turns and, in fact, there are possibilities for it to reach many contact conditions as the time evolves.

According to Figures 4, 5, 6, 7, 8, and 9, the amplitude of F_N jumps at the beginning of contact, from zero (no contact) to a value associated with the impact force between the bodies. The contact force evolves with time according to the system states and properties. The value of F_N at the instant of impact does not necessarily represent the biggest value for

Table 1: Numerical values considered in the numerical simulations for different values of k_w.

Parameter	Value	Unity
m_b	2.00	Kg
m_s	5.00	Kg
m_w	10.00	Kg
k_s	5.00	Nm
	10.00	
k_w	400.00	Nm
	1000.00	
c_s	7.00	Ns/m
c_w	1.00	Ns/m
ℓ	1.00	m
d	0.60	m
d_{Acmb}	0.50	m
M_θ	10.00	Nm
$I_{b,\text{cm}}$	0.1667	Kg/m^2

Table 2: Numerical values considered in the numerical simulations for different values of M_θ.

Parameter	Value	Unity
m_b	2.00	Kg
m_s	1.00	Kg
m_w	1.00	Kg
k_s	400.00	Nm
k_w	5.00	Nm
c_s	7.00	Ns/m
c_w	7.00	Ns/m
ℓ	1.00	m
d	0.60	m
d_{Acmb}	0.50	m
	5	
M_θ	10	Nm
	20	
$I_{b,\text{cm}}$	0.1667	Kg/m^2

the contact force, as can be seen in these figures. A sudden change in velocity, when collision takes place, can be verified clearly in these figures.

5.2. Considering Different Values of M_θ

Table 2 shows numerical values considered in the numerical simulations for different values of M_θ.

6. Conclusions

To conclude, it is important to say that the time step used in the numerical integration and the choice of the integrator are very important aspects to be considered. New numerical

integrators can be tested in the course of this investigation and results compared to the ones presented here.

An important consideration not to be forgotten when dealing with problems presenting some sort of constraint is that more than one set of governing equations of motion must be integrated to cover all the system dynamics. The set of equations that governs the system dynamics when the constraint condition is active is different from the one that governs the unconstrained movement of the system. One of these sets is always generating the states for the other.

In this context, the determination of the velocities after contact (impact) is very important. The velocity expressions presented in (4.33) are the necessary corrections one must do when considering the fully plastic impact case. If this correction is not taken into consideration in the numerical integration of the governing equations, the system will gain energy after impact, which is not true.

It is important to realize also that the number of degrees of freedom involved changes from one set of equations to the other. The necessity for changing from one set of governing equations to another (according to the system's requirements of contact or noncontact conditions) represents a source of integration errors, since the integrator is faced with singularities.

The problem presented in this paper and the procedures developed for its analysis can be extended to many other systems and situations (including more complex ones). The theory presented here can be applied to problems in which robots have to follow some prescribed patterns or trajectories when in contact with the environment (like in painting activities, for instance, or the ROKVISS experiment at DLR).

The next steps are the development of the analytical expressions for the velocities after impact considering any value for the coefficient of restitution and the inclusion of friction forces between m_s and the left and right vertical guiding surfaces; and between the free end of the bar and m_w.

References

[1] F. Pfeiffer and C. Glocker, *Multibody Dynamics with Unilateral Contacts*, Wiley Series in Nonlinear Science, John Wiley & Sons, New York, NY, USA, 1996.

[2] C. Lánczos, *The Variational Principles of Mechanics*, Mathematical Expositions, no. 4, University of Toronto Press, Toronto, Canada, 4th edition, 1970.

[3] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, Cambridge, UK, 1965.

[4] N. A. Fufaev and J. I. Neimark, *Dynamics of Nonholonomic Systems*, American Mathematical Society, 1972.

[5] A. Fenili, L. C. G. Souza, and B. Schäfer, "A mathematical model to investigate contact dynamics in constrained robots," in *Proceedings of the 6th International Symposium on Dynamic Problems of Mechanics (DINAME '05)*, D. A. Rade and V. Steffen Jr., Eds., Ouro Preto, Brazil, February-March 2005.

[6] B. Schäfer, B. Rebele, and A. Fenili, "Space robotics contact dynamics investigations and numerical simulations: ROKVISS," in *Proceedings of the 15th CISM-IFToMM Symposium on Robot Design, Dynamics and Control*, 2004.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 897570, 11 pages
doi:10.1155/2009/897570

Research Article

Gravitational Capture of Asteroids by Gas Drag

E. Vieira Neto and O. C. Winter

FE-Guaratinguetá, Departamento de Matemática, UNESP—Universidade Estadual Paulista, 12.516-410 SP, Brazil

Correspondence should be addressed to E. Vieira Neto, ernesto@feg.unesp.br

Received 30 July 2009; Revised 6 November 2009; Accepted 1 December 2009

Recommended by Silvia Maria Giuliatti Winter

Several irregular satellites of the giant planets were found in the last years. Their orbital configuration suggests that these satellites were asteroids captured by the planets. The restricted three-body problem can explain the dynamics of the capture, but the capture is temporary. It is necessary some kind of dissipative effect to turn the temporary capture into a permanent one. In this work we study an asteroid suffering a gas drag at an extended atmosphere of a planet to turn a temporary capture into a permanent one. In the primordial Solar System, gas envelopes were created around the planet. An asteroid that was gravitationally captured by the planet got its velocity reduced and could been trapped as an irregular satellite. It is well known that, depending on the time scale of the gas envelope, an asteroid will spiral and collide with the planet. So, we simulate the passage of the asteroid in the gas envelope with its density decreasing along the time. Using this approach, we found effective captures, and have a better understanding of the whole process. Finally, we conclude that the origin of the irregular satellites cannot be attributed to the gas drag capture mechanism alone.

1. Introduction

The giant planets of the Solar System have two kinds of satellites, the regular and the irregular ones [1]. This definition was based on the orbits of the satellites. The regular satellites have their orbits near the equatorial plane, the eccentricity is near zero, and they are close to their planets. The irregular ones are out of plane, many are retrograde, the eccentricities are above 0.1, and they are far from their planets.

Since the discovery of two new Uranus' moons in 1997 by Gladman et al. [2], 98 other satellites were found. These new satellites, plus the old irregulars satellites, are almost all retrograde, as we can see in Table 1.

The orbits of the irregular satellites suggest that they were captured by their planets. That is, they were formed elsewhere in the Solar System like any other asteroid, and later they might have had a close encounter with a planet and could have been captured.

Table 1: Number of irregular satellites of the giant planets [3].

Planet	Prograde	Retrograde	Total
Jupiter	6	49	55
Saturn	8	27	35
Uranus	1	8	9
Neptune	3	4	7

Using the dynamics of the three-body problem it is possible to explain the gravitational capture of these satellites. But the gravitational capture is temporary (see, for example, [4]). There are works that studied the capture time [5] or the directions of capture [6] in the case of the three-body problem. Thus, it is necessary some kind of additional effect to turn the temporary capture into a permanent one.

In the literature we can find several mechanisms which were proposed in order to turn the capture permanent. It could be a pull-down mechanism due to the mass growth of the planet [7–9], or a capture through n-body interactions [10–12]. In this work we will explore the gas drag mechanism. In the early history of the solar system, in the last stage of the giant planets formation, a gas envelope was formed around each one of them [13]. This gas envelope could make a flying by asteroid to loose enough energy in order to became a prisoner of the planet. Other authors already discussed this possibility (see, for example, [14]). Although the theory and numerical simulation have some success in explaining the prograde satellites, the retrograde ones are not well explained. This is because the retrograde satellites feel a stronger headwind which make them to spiral inward faster and collide with the planet sooner than the prograde ones. This suggests distinct conditions for the gas drag capture of the prograde and retrograde satellites.

Ćuk and Burns [14] studied the case of a prograde satellite capture, using the orbital and physical parameters of Himalia Jupiter's satellite. They used a static surface density gas envelope with edges to modify the size of the envelope. They vary the position of the edge, but maintaining the surface density of the gas envelope constant. They have discussed the effect of a gas envelope with decaying density in their work, but made no simulation. Also in the work of Ćuk and Burns [14], particles were started inside the gas envelope, mainly using Himalia's initial conditions, and integrated backwards in time.

In the work of Pollack et al. [15] they simulate the formation of a giant planet and characterized the phases: (1) accretion of solid material; (2) accretion of solid material and gas with a small rate; (3) runaway gas accretion. They concluded that the overall time scale is determined by the second phase, and the first and third phase are very quick. They suppose that the envelope did not collapse in the phase of runaway gas accretion as long as the solar nebula could supply the planet with gas rapidly enough. In their work they do not estimate a time interval for the third phase, but their Figure 1(a) give us a clue of this time interval and it seems to last less than a thousand years. In our work we will consider the very last stage, when there is no more gas in the solar nebula to feed the planetary gas envelope and this envelope collapses. We suppose the duration of this stage to be about a hundred years and in this stage the density of the gas envelope will vary until vanish.

In this work we simulate a gas envelope with no edges, but its surface density varies linearly with time. We started the integration of the asteroid orbit outside the gas envelope in a proper region as an heliocentric orbit, and integrated it forward in time until it reaches the gas envelope around the planet. We also point out that this work is in two dimensions, while the work of Ćuk and Burns [14] was three-dimensional.

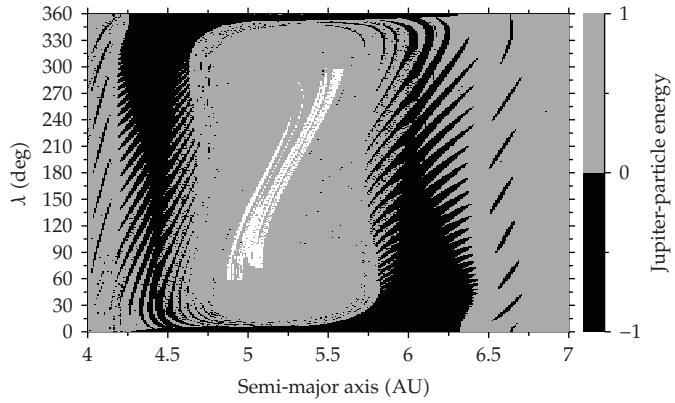


Figure 1: Grid of initial conditions of the asteroid with respect to the Sun. In the plot we can see which initial conditions that lead to negative two-body energy w.r.t. Jupiter, that is, the asteroid was temporarily captured by Jupiter. The white region in the middle of the plot are trajectories which lead to a two-body energy greater than one. For each asteroid initial condition in the $a \times \lambda$ space, Jupiter begins its orbit at 5.2 AU from the Sun with $\lambda = 0°$.

In our work we will discuss the very last stage of the gas envelope when it collapses onto the planet. Thus, the surface density of the gas will change from some accepted initial value down to zero, vanishing the gas. This configuration is favorable for the retrograde satellites. Our goal is to investigate how to get permanent gravitational capture for the retrograde satellites.

We will first discuss, in Section 2, the possible heliocentric orbital region that could make an asteroid to be captured by Jupiter. In Section 3 we will expose the gas model and show the results of our simulations. Finally, we discuss our results in Section 4.

2. Possible Origins for the Captured Asteroids

In order to study the asteroids initial conditions which lead to capture by the planet, we used the circular planar Sun-Jupiter system, and made a grid of initial conditions with semi-major axis and true longitude (λ). The others asteroid's initial conditions parameters were fixed in such a way that the asteroid had a circular planar orbit with respect the Sun. The semi-major axis (a) was varied from 4 to 7 Astronomical Units (AU) and the true longitude (λ) from 0° to 360°. For each trajectory in this grid, we measured the two-body energy of the asteroid relative to Jupiter from time $t = 0$ to time $t = 10^5$ years and the minimum value was stored. The two-body energy is not constant in the three-body system and will vary due to the Sun's perturbation. We were interested in the negatives values of the two-body energy, which configures the temporary gravitational capture [16], since negative value of this energy corresponds to an osculate elliptical orbit. Other possible option for this study would be the minimum distance between asteroid and Jupiter and a passage through the Hill sphere, but the energy was chosen because it takes the position and velocity of the particle with respect to the planet, and we can avoid a close passage with very high speed which is not interesting to our experiment.

In the search of proper initial conditions which leads to gravitational capture we did not consider the gas drag. The idea in this first moment is just to find the path for an encounter with Jupiter.

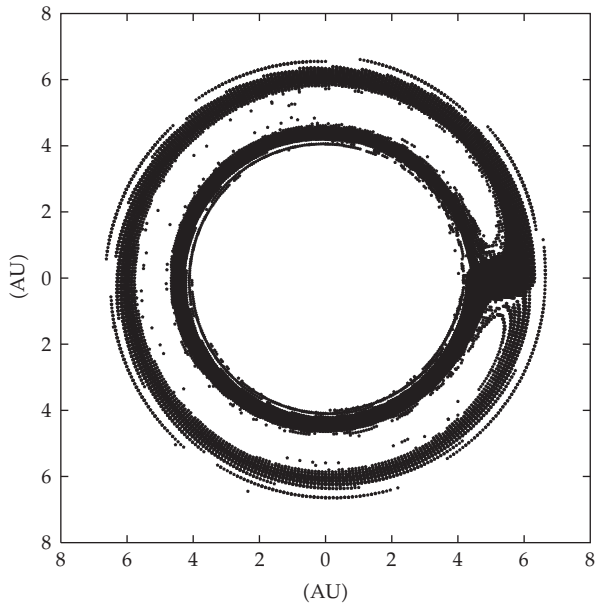


Figure 2: Asteroid initial conditions which lead to negative two-body energy w.r.t. Jupiter, taken from Figure 1, in polar coordinates.

The numerical integrations we made in this work used an integrator that uses Gauss-Radau spacings described by E. Everhart[17].

The result for this simulation can be seen in Figure 1, where we separated the initial conditions in the grid that lead to positive and negative two-body energy. The black points indicate the initial conditions of trajectories which achieved a minimal negative two-body energy at some point along the integration, while the other color dots correspond to those trajectories whose minimum energy was positive.

On Figure 2 we plotted only the initial conditions which lead to negative two-body energy of the asteroid with respect to Jupiter, in polar coordinates. From these two figures we can see the horseshoe libration zone [18] and the chaotically-unstable region around Jupiter's orbit, which is due to overlap of mean motion resonances [19].

In Figures 1 and 2 we can see the existence of two dense regions with initial conditions that lead to negative values of the two-body energy. These initial conditions are the ones that lead the asteroid to pass close enough to the planet and have a temporary gravitational capture.

3. Dynamics and Numerical Results

In order to turn a temporary capture into a permanent one, we used a drag force proportional to the velocity of the asteroid relative to the velocity of the gas, the cross section of a spherical asteroid, and the local density of the gas [20, 21]:

$$\overrightarrow{F}_d = -C_d \frac{\pi}{2} R^2 \rho(r) v_{\text{rel}} \overrightarrow{v}_{\text{rel}} , \tag{3.1}$$

where C_d is the drag coefficient, R is the asteroid radius, v_{rel} is the relative velocity of the asteroid with respect to the gas, and

$$\rho(r) = \rho_0 \left(\frac{r}{r_0} \right)^{\gamma}, \tag{3.2}$$

where $\rho(r)$ is the gas density at a distance r from the centre of the planet, and γ is the exponent that gives how the gas density decay with the distance from the planet. In this work we used $\gamma = -1$, as in Ćuk and Burns [14]. We also used

$$\rho_0 = \frac{\Sigma_0}{\sqrt{\pi} H_0} \tag{3.3}$$

as the gas density at r_0, with Σ_0 being the gas surface density, and H_0 the gas envelope height. We considered $r_0 = 100$ Jupiter Radius (JR), $H_0 = 0.05 r_0$, as in Ćuk and Burns [14]. The asteroidal radius was took as $R = 6$ km. We choose this small radius because we are interested in the retrograde satellites and this will reduce the decaying effect of the gas. Besides, this is a representative radius value for the retrograde satellites observed around Jupiter. The gas velocity rotation around the planet was chosen to be 90% of the Keplerian velocity.

In order to change the density of the gas envelope during the passage of the asteroid we implemented the following surface density time function:

$$\Sigma_0(t) = \begin{cases} \Sigma_0(t_0), & t < t_0 \\ \Sigma_0(t_0) + \alpha(t - t_0), & t_0 \leq t \leq t_1 \\ 0, & t > t_1, \end{cases} \tag{3.4}$$

where α is the time rate of the decreasing density. An exponential (or sigmoid) decay law will change only slightly our results.

The gravitational model used was the general three-body model, that is, the mass of the asteroid were considered. The density of the asteroid was considered 1.5 g/cm^3, and the gas density surface considered was $\Sigma_0(t_0) = 10^3$ g/cm^2, as in Ćuk and Burns [14].

The gas envelope is formed after the consumption of all the gas from the solar nebula. The lack of gas to fuel the planet makes the envelope collapses into the planet [20, 22]. It is believed that this collapse is quick and took only some hundreds of years to occur. That is, the surface density of the gas goes to zero in times of the order of 10^2 years.

3.1. Asteroid's Passage through the Gas Envelope

In order to study the passage of the asteroid through the gas envelope we use the results from Figure 1 to choose an adequate set of initial conditions. The initial conditions used are in the same plane of the gas envelope around Jupiter. We fixed a time interval of 100 years ($\Delta t = |t_1 - t_0|$) for the gas to vanish. The gas has its density changed according to (3.4), from $\Sigma_0(t_0)$ to zero in a time interval of $\Delta t = 100$ years, and after this period there is no more gas around the planet, $\Sigma_0(t) = \Sigma_0(t_1) = 0$.

The effect of the change of the surface gas density on the asteroid trajectory will be analysed by fixing the initial orbital configuration of the asteroid and Jupiter, but selecting different values of initial integration time from $t = -100$ years to $t = 0$ year, with steps of 0.1

year, while time variation of the gas envelope has ever the initial time $t_0 = -100$ years and final time $t_1 = 0$ year. That is, for each trajectory the asteroid and Jupiter have the same orbital elements with respect to the Sun, but the orbital integration of the system starts at different times, reaching the gas envelope at different surface density values Σ_0.

During the integration of the trajectory we monitored the relative two-body energy of the asteroid relative to Jupiter to verify its sign. Initially the asteroid is in heliocentric orbit and has a positive relative two-body energy with respect to Jupiter. We stoped the integration if one of the following happens: (i) the asteroid passes through the gas envelope; (ii) the asteroid collides with Jupiter; (iii) the integration time surpass $t = 50$ years.

In the first case, to know that the asteroid had passed through the gas envelope around Jupiter, we measured the relative two-body energy of the asteroid with respect to Jupiter during the encounter. The two-body energy turn into negative when the asteroid is close to Jupiter. Then, the relative two-body energy turn to positive again when the asteroid is distant from the planet.

For the second case, as we do not consider the Galilean satellites in the experiment, the collision with Jupiter is considered when the distance Jupiter-asteroid is smaller than the Jupiter's radius.

The third case happens if one of the following occurs: (a) the asteroid is permanently captured; (b) the asteroid is deflected by the gas envelope; (c) the asteroid is temporarily captured.

Case (a) is considered when the asteroid entered in the gas envelope, that is, its relative two-body energy was turned into negative, in a time that the surface density of the gas has a value between $\Sigma_0(t_0)$ e 0. Due to the gas, the asteroid suffers a drag and loose energy. But the gas surface density vanishes, the asteroid stop to loose its energy and, if it lost enough energy, it became a permanent satellite.

In case (b) the asteroid never reach a negative relative two-body energy with respect to Jupiter. In these kind of trajectories the asteroid passes in the outer layers of the gas envelope and it is deflected.

Finally, in case (c) the asteroid arrives at the planet when the surface density of the gas is already zero, or very low, and do not loose enough energy to be effectively captured.

We tested four initial conditions for semi-major axis at 6.0 AU, and other four initial conditions for semi-major axis at 4.5 AU. Each one with a different true longitude value (λ) for the asteroid. These initial conditions were obtained from Figure 1. In Figure 3 we can see the results of the trajectories of the asteroid which have initial semi-major axis of 6.0 AU. None of the trajectories was effectively captured by the planet, many trajectories entered the gas envelope and escaped quickly. Other trajectories never entered the gas envelope as shown in Figures 3(c) and 3(d), where there is a region in the interval $[-100, -40]$ years with no points. This occurs because the trajectories are deflected by the gas. In Figure 3(d), the lack of points for the region $[-7, 0]$ happens because the trajectories are with negative two-body energy, but the gas has low surface density and the trajectories will escape again due to the fact that the dissipative effect is too weak.

In Figure 4 we show the results for other four initial conditions, but now for $a = 4.5$ AU. The plots in the Figure show the time spent by the asteroid inside the gas envelope with negative two-body energy. Only for Figure 4(c) effective captures occur, as showed by the blue squares. In the other three cases there are only collisions and passages near the planet.

We also note in the plots the existence of upward spikes. These spikes occur due to the trajectories transition between trajectories which just pass through the gas envelope and

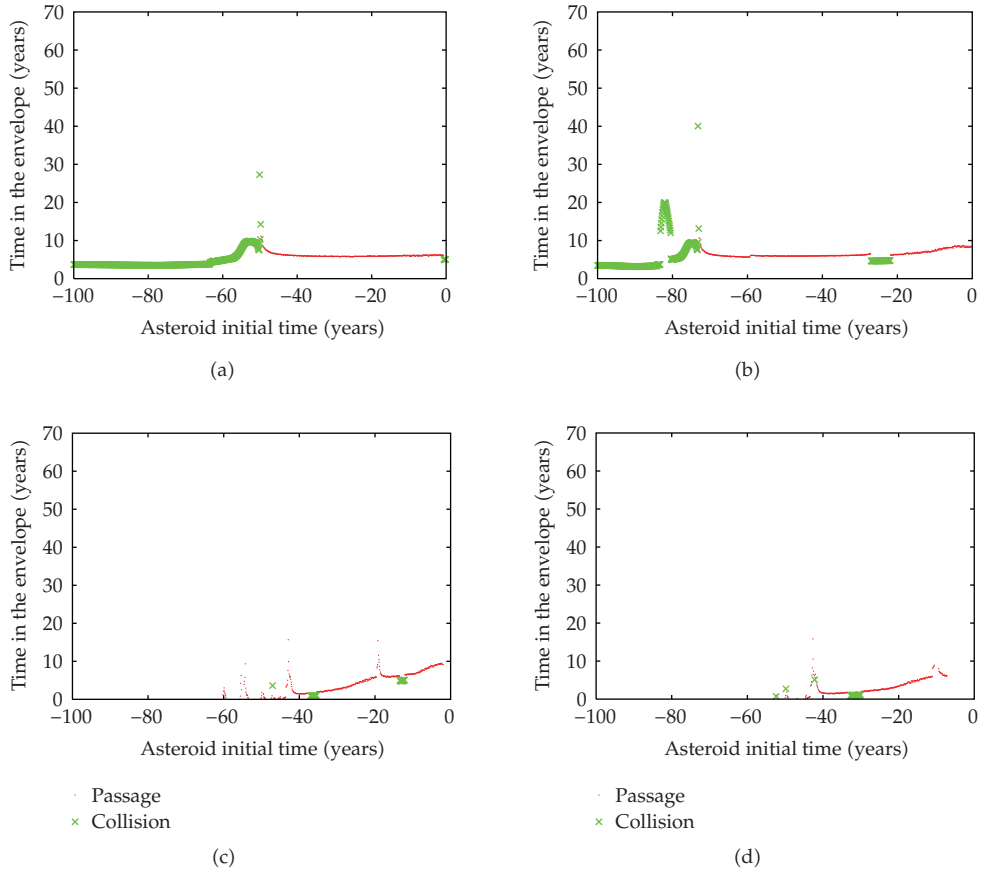


Figure 3: Length of time that the asteroid spent inside the gas envelope for each initial time simulation. Four initial conditions for semi-major axis of 6.0 AU: (a) $\lambda = 70°$; (b) $\lambda = 110°$; (c) $\lambda = 250°$; (d) $\lambda = 290°$.

trajectories that collide with the planet. These transition trajectories spend more time around the planet making more loops than the normal trajectories.

We studied some trajectories of Figure 4(c) and show them in Figure 5. The trajectories where condition (i) is satisfied, that is, a passage through the gas envelope, occurs when the asteroid initiated its trajectory when the gas density is close to zero. In this case the gas is so sparce that the asteroid almost does not feel it and escape (see Figure 5, trajectory d). Other possibility is when the asteroid initiated its trajectory just after the beginning of the gas density variation and the gas density is strong enough to make the asteroid to scatter in the extended atmosphere of the planet (see Figure 5, trajectory a).

Condition (ii), the collisions, mainly occurs for initial time interval [−40, −26.1] years of Figure 4(c). In this time interval the gas is not so dense and allows the asteroid to enter in the envelope, but it is a trap for the asteroid which has not enough energy to escape from the gas envelope, then it spirals and collides with the planet (see Figure 5, trajectory b).

Condition (iii), integration time greater than 50 years, happens in the interval [−26.0, −21.6] years of Figure 4(c). Figure 5, for trajectory c, shows a trajectory with initial time at $t = -25.0$ years. The asteroid enters the gas envelope and it is trapped by the gas. This trajectory has its semi-major axis reduced, the gas vanish and the trajectory is stabilized.

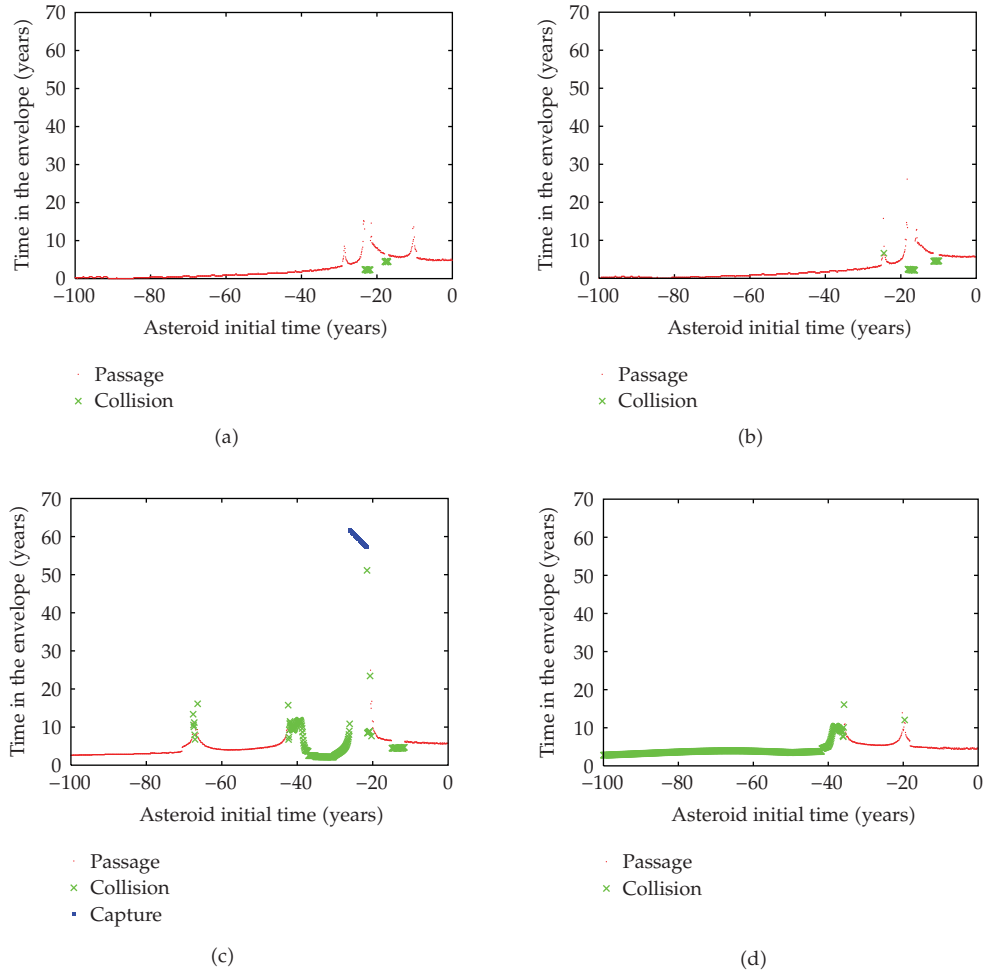


Figure 4: Length of time that the asteroid spent inside the gas envelope for each initial time simulation. Four initial conditions for semi-major axis of 4.5 AU: (a) $\lambda = 70°$; (b) $\lambda = 110°$; (c) $\lambda = 250°$; (d) $\lambda = 290°$.

3.2. Captured Orbits

In order to understand what happens with the permanently captured orbits, we computed the average of the semi-major axis and eccentricity of the captured trajectories, presented in Figure 4(c). The average was made in the time interval between 0 and 50 years, that is, after the gas had vanished.

In Figure 6 we present the behavior of the averaged semi-major axis and eccentricity after the gas vanishes. For the first trajectories, the gas is denser at the time the asteroid arrives, which make its semi-major axis to decrease more, stabilizing the orbit with a semi-major axis close to 10 Jupiter's radius. With a less dense gas at the time of arrival of the asteroid, the average value of the semi-major axis is close to 320 Jupiter's radius.

The average value of the eccentricity is also shown in Figure 6. Most of the trajectories have eccentricities close to 0.95. The few ones which are less than 0.8 are the ones which are closer to the planet.

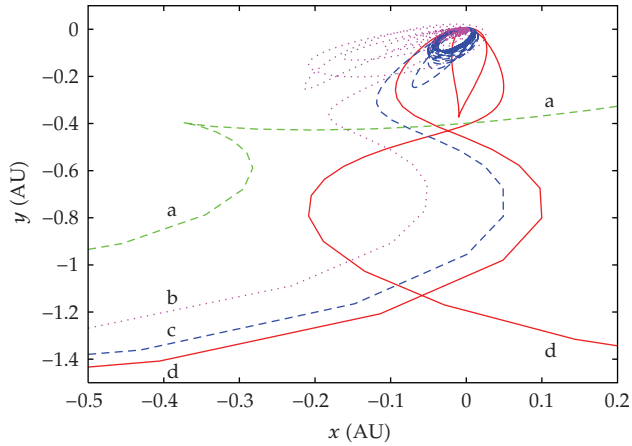


Figure 5: Four trajectories of Figure 4(c), all in a non rotating planetocentric reference frame with x pointing to the Vernal point. (a) Trajectory for gas density at $t = -80$ years. The asteroid scatters in the gas and goes away. (b) Trajectory for gas density at $t = -40$ years. The asteroid enters in the gas envelope, but the gas is strong and makes the asteroid to collide with the planet. (c) Trajectory for gas density at $t = -25$ years. The asteroid enters in the gas envelope in a time that the gas is just strong enough to make the asteroid to reduce its energy and not collide with the planet. After the gas vanishes the capture is permanent. (d) Trajectory for gas density at $t = -18$ years. The asteroid has a temporary gravitational capture but the gas is too sparce and the asteroid goes away.

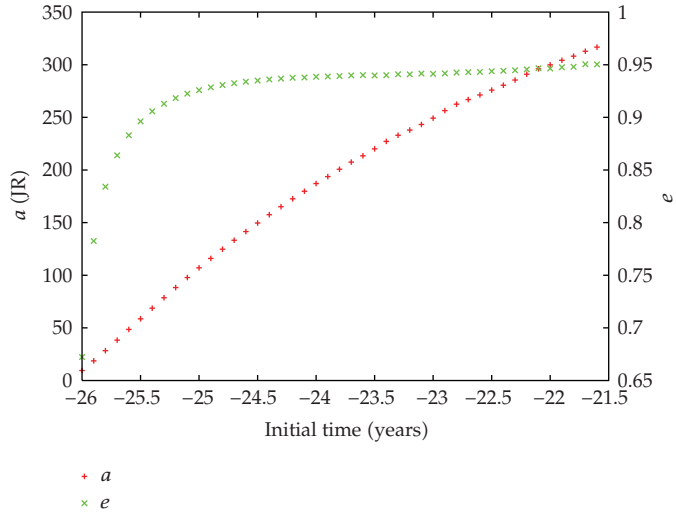


Figure 6: Averaged values of semi-major axis (left side) and eccentricity(right side) in the time interval between 0 and 50 years for the captures generated by gas drag.

We also pointed that the inclination of the asteroids during all the time after the gas was vanished for all captured trajectories were retrograde, that is, the inclinations never changes from $i = 180°$ with respect to Jupiter.

This process of capture is not suitable for the prograde capture. In the work of Ćuk and Burns [14] they show that the time needed to capture a Himalia satellite type is between 10^4 to 10^6 years. The process described here is rapid and the density of the gas envelope is

not enough to make a asteroid to lose orbital energy and turn it into a permanent prograde satellite.

If we compare this and the work of Ćuk and Burns [14], using the results of Pollack et al. [15], the capture of the prograde satellites might have occurred in the second phase of planetary formation, while the capture of the retrograde satellites might have occurred only in the third stage. In the work of Pollack et al. [15] they estimate a time between 10^6 years and 8×10^6 years in the second phase. Thus, the model of capture with gas drag show us that the capture of prograde satellites are much more probable than the retrograde ones. This contradicts the observations. There are more retrograde satellites observed than prograde ones. Therefore, the origin of the prograde and retrograde irregular satellites cannot be attributed to the gas drag capture mechanism alone.

4. Conclusions

We found a region of initial conditions, close to Jupiter, which produces temporary gravitational capture around Jupiter. We took eight sets of initial conditions from this region and varied the initial time for the integration in a dynamics which includes a gas envelope with decreasing density along the time to a moment that it vanishes. These simulations gave us information about what might be the outcomes of the evolution of such system. The possible outcomes are: the scatter of the asteroid, the temporary gravitational capture, the collision with Jupiter, and the permanent capture. Seven out of eight sets of initial conditions did not give effective capture. Only one set presented a few permanent captures due to the gas drag.

We have shown that the permanent capture of an asteroid in retrograde orbit around Jupiter due to a vanishing gas envelope is possible. We have the knowledge that this happens with low probability, but considering the supply of planetesimals available in the early time of the solar system formation, this kind of event could have happened. Although comparing with the probability of prograde capture it is less probable. The observations show more retrograde satellites than prograde ones. This is a problem for the gas drag capture mechanism as a whole.

It is necessary to look more regions of the space of semi-major axis, eccentricity, and inclination in order to have enough information about the gravitational capture with gas drag and also the time for the gas envelope to vanish needs to be tested. In order to implement this it is necessary to use some Monte Carlo method due to the huge range of free parameters.

The captures studied in this work, which were shown in Section 3.2, are close to Jupiter compared with the real irregular satellites. Some could cross the Galilean satellites orbits and collide with them. It is necessary to improve the gas envelope decaying mechanism to obtain more information of the process, but in this work we tried to proof this concept and further work is under development.

Acknowledgments

This work was funded by FAPESP and CNPq. These supports are gratefully acknowledged.

References

[1] J. B. Pollack, J. A. Burns, and M. E. Tauber, "Gas drag in primordial circumplanetary envelopes: a mechanism for satellite capture," *Icarus*, vol. 37, no. 3, pp. 587–611, 1979.

[2] B. J. Gladman, P. D. Nicholson, J. A. Burns, et al., "Discovery of two distant irregular moons of Uranus," *Nature*, vol. 392, no. 6679, pp. 897–899, 1998.
[3] D. Jewitt and N. Haghighipour, "Irregular satellites of the planets: products of capture in the early solar system," *Annual Review of Astronomy and Astrophysics*, vol. 45, pp. 261–295, 2007.
[4] K. Tanikawa, "Impossibility of the capture of retrograde satellites in the restricted three-body problem," *Celestial Mechanics*, vol. 29, no. 4, pp. 367–402, 1983.
[5] E. Vieira Neto and O. C. Winter, "Time analysis for temporary gravitational capture: satellites of Uranus," *Astronomical Journal*, vol. 122, no. 1, pp. 440–448, 2001.
[6] D. S. de Oliveira, O. C. Winter, E. Vieira Neto, and G. de Felipe, "Irregular satellites of Jupiter: a study of the capture direction," *Earth, Moon and Planets*, vol. 100, no. 3-4, pp. 233–239, 2007.
[7] T. A. Heppenheimer and C. Porco, "New contributions to the problem of capture," *Icarus*, vol. 30, no. 2, pp. 385–401, 1977.
[8] E. Vieira Neto, O. C. Winter, and T. Yokoyama, "The effect of Jupiter's mass growth on satellite capture Retrograde case," *Astronomy and Astrophysics*, vol. 414, no. 2, pp. 727–734, 2004.
[9] E. Vieira Neto, O. C. Winter, and T. Yokoyama, "Effect of Jupiter's mass growth on satellite capture the prograde case," *Astronomy and Astrophysics*, vol. 452, no. 3, pp. 1091–1097, 2006.
[10] G. Colombo and F. A. Franklin, "On the formation of the outer satellite groups of Jupiter," *Icarus*, vol. 15, no. 2, pp. 186–189, 1971.
[11] C. B. Agnor and D. P. Hamilton, "Neptune's capture of its moon Triton in a binary-planet gravitational encounter," *Nature*, vol. 441, no. 7090, pp. 192–194, 2006.
[12] H. S. Gaspar, O. C. Winter, and E. Vieira Neto, "Irregular satellites of Jupiter: capturecongurations of binary asteroids," submitted to *Monthly Notices of the Royal Astronomical Society*.
[13] J. I. Lunine and D. J. Stevenson, "Formation of the Galiean satellites in a gaseous nebula," *Icarus*, vol. 52, pp. 14–39, 1982.
[14] M. Ćuk and J. A. Burns, "Gas-drag-assisted capture of Himaila's family," *Icarus*, vol. 167, pp. 369–381, 2004.
[15] J. B. Pollack, O. Hubickyj, P. Bodenheimer, J. J. Lissauer, M. Podolak, and Y. Greenzweig, "Formation of the giant planets by concurrent accreation of solid and gas," *Icarus*, vol. 124, pp. 62–85, 1996.
[16] E. Vieira Neto, O. C. Winter, and C. Melo, "The use of the two-body energy to study problems of escape/capture," in *Dynamics of Populations of Planetary Systems*, Z. Knezevic and A. Milani, Eds., Proceedings IAU Colloquium, no. 197, pp. 439–444, Cambridge University Press, Cambridge, UK, 2005.
[17] E. Everhart, "An efficient integrator that uses Gauss-Radau spacings," in *Dynamics of Comets: Their Origin and Evolution*, A. Carusi and G. B. Valsecchi, Eds., vol. 115 of *Proceedings of IAU Colloquium, no. 83*, pp. 185–202, Astrophysics and Space Science Library, Dordrecht, The Netherlands, 1985.
[18] C. D. Murray and S. F. Dermott, *Solar System Dynamics*, Cambridge University Press, Cambridge, UK, 1999.
[19] J. Wisdom, "The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body problem," *Astronomical Journal*, vol. 85, pp. 1122–1133, 1980.
[20] M. Podolak, W. B. Hubbard, and J. B. Pollack, "Gaseous acreation and the formation of giant planets," in *Protostars and Planets III*, E. H. Levy and J. I. Lunini, Eds., pp. 1109–1147, University of Arizona Press, Tucson, Ariz, USA, 1993.
[21] I. Adashi, C. Hayashi, and K. Nakazawa, "The gas drag effect on the elliptic motion of a solid body in the primordial nebula," *Progress in Theorical Physics*, vol. 56, no. 3, pp. 1756–1771, 1976.
[22] S. J. Weidenschilling, "Planetesimals from Stardust," in *From Stardust to Planetesimals*, Y. J. Pendleton and A. G. G. M. Tielens, Eds., vol. 122 of *ASP Conference Series*, p. 281, Astronomical Society of the Pacific, Santa Clara, Calif, USA, 1997.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 859678, 14 pages
doi:10.1155/2009/859678

Research Article

Atmospheric Reentry Dynamics of Conic Objects

J. P. Saldia,[1] A. Cimino,[1] W. Schulz,[1] S. Elaskar,[1,2] and A. Costa[1,2]

[1] *Departmento de Aeronáutica, Universidad Nacional de Córdoba, Vélez Sarsfield 1611, Córdoba 5000, Argentina*
[2] *CONICET, Rivadavia 1917, Argentina*

Correspondence should be addressed to W. Schulz, wschulz@efn.uncor.edu

Received 25 June 2009; Revised 27 September 2009; Accepted 4 November 2009

Recommended by Silvia Maria Giuliatti Winter

One of the key issues in a reentry risk analysis is the calculation of the aerodynamic coefficients. This paper presents a methodology to obtain these coefficients and couple it to a code that computes re-entry trajectories considering six degrees of freedom. To evaluate the different flight conditions encountered during the natural re-entry of conical objects, the Euler Equations for gasdynamics flows are used. A new scheme TVD (Total Variation Diminishing) is incorporated to a finite volume unstructured cell-centred formulation, for application to three-dimensional Euler flows. Finally, numerical results are obtained for a conical body at different attack angles and Mach. With these results, the calculation of the trajectories during atmospheric re-entry is completed.

1. Introduction

In the case of natural reentries (non-controlled), the orbital evolution of an object can only be monitored, with no or limited ability to control risks. The time window for reentry of a satellite is usually provided with a standard error of ±10% to ±20% of the remaining orbital lifetime. For the controlled reentries, it is required to simulate the different scenarios until the right window for the mission is found, being that the total or partial disintegration, or the landing on a safe place.

In this paper, the main objective is to conduct numerical simulations of the supersonic flow regime on a conical body, thereby using a code developed at the Department of Aeronautics of the National University of Córdoba, Argentina [1, 2]. This code uses the technique of finite volumes for solving Euler equations. The spatial discretization of the domain is done through a mesh of unstructured tetrahedral volumes. It has implemented a new technique for choosing the limiter functions that can reduce the artificial viscosity without the loss of strength robustness of the Total Variation Diminishing scheme—TVD [3–6].

The drag and lift coefficients, C_D and C_L, obtained by the numerical simulation of compressible flows are used in a code that allows to evaluate the trajectories considering six degrees of freedom [7]. As a result of this research, the trajectories of reentry into the Earth's atmosphere for conical objects having different initial flight conditions are presented.

2. Methodology

2.1. Description of the Numerical Scheme for Compressible Flow

The three-dimensional Euler equations can be written as

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F} = 0, \tag{2.1}$$

where $\mathbf{U}$ is the vector of conservative variables, and $\mathbf{F}$ is the 3D vectorial flow.

The temporal change of the conservative variables can be expressed as

$$\mathbf{U}^{n+1} = \mathbf{U}^n - \frac{\Delta t}{Vol} \sum_{l=1}^{l\text{faces}} \mathbf{F}_l^* \cdot \mathbf{n}_l A_l, \tag{2.2}$$

where the flux of the conservative variables $\mathbf{F}$ has been replaced by the numerical flux tensor $\mathbf{F}^*$. *Vol* indicates the volume where the integration is performed, $\mathbf{n}_l$ is the outward normal to the control surface (A_l).

Equation (2.2) allows the use of a locally aligned system of coordinates whose unit vector $\mathbf{i}$ coincides with the normal to the face l of the cell, and the unit vectors $\mathbf{j}$ and $\mathbf{k}$ are tangential directions. To achieve second-order accuracy, the numerical flux at the interface between cells l and l+1 in the direction normal to the face l is calculated by [8]

$$\mathbf{f}_{i+1/2}^* = \frac{\mathbf{f}_i + \mathbf{f}_{i+1}}{2} + \frac{1}{2} \sum_m \Phi_{i+1/2}^m \overline{\mathbf{K}}_{i+1/2}^m, \tag{2.3}$$

where $\mathbf{f}_i$ and $\mathbf{f}_{i+1}$ are the physical fluxes normal to the face in each cells, $\overline{\mathbf{K}}_{i+1/2}^m$ is the m-th right eigenvector, and $\Phi_{i+1/2}^m$ is, in the original Harten-Yee scheme [9–11], defined as

$$\Phi_{i+1/2}^m = g_i^m + g_{i+1}^m - \left|\lambda_{i+1/2}^m + \beta_{i+1/2}^m\right| \alpha_{i+1/2}^m, \tag{2.4}$$

being

$$g_i^m = \frac{S}{2} \max\left[0, \min\left(\left|\lambda_{i+1/2}^m \alpha_{i+1/2}^m\right|, S\left|\lambda_{i-1/2}^m\right| \alpha_{i-1/2}^m\right)\right], \tag{2.5}$$

$$S = \operatorname{sign}\left(\lambda_{i+1/2}^m\right), \tag{2.6}$$

$$\beta_{i+1/2}^m = \begin{cases} \dfrac{g_{i+1}^m - g_i^m}{\alpha_{i+1/2}^m} & \text{if } \alpha_{i+1/2}^m \neq 0 \\ 0 & \text{if } \alpha_{i+1/2}^m = 0, \end{cases} \tag{2.7}$$

where $\alpha^m_{i+1/2}$ is the jump of the conserved variables across the interfaces between cells i and i+1, $\lambda^m_{i-1/2}$ is the m-th eigenvalue of the Jacobian matrix, g^m_i is the limiter function, and S is the sign function of the corresponding eigenvalue [8–11]. Since the local Riemann problem is solved with rotated data, the eigensystem is calculated in the locally aligned coordinate frame.

The limiter function given in (2.5) is known as *minmod* [3–6]. The minmod selects the minimum possible value, so that the scheme is TVD. The other end is the limiter function *superbee* that ponders the contribution of the high-order flux [3]. The only implementation of the superbee function leads to an excessively compressive scheme which it is not very robust for general practical aerospace applications [6].

In the numerical solution of the three-dimensional Euler equations, five wave families appear. If the five wave families are enumerated in correspondence with their speed, being one the slowest and five the fastest, it can be demonstrated that for waves of the families two to four, the characteristic velocities at both sides of the discontinuity are the same and equal to the velocity discontinuity [3, 5]. This property makes it very difficult to solve theses waves accurately. Generally they are solved diffusely because the numerical methods incorporate a large amount of artificial viscosity to track the contact discontinuity.

In this work, the possibility of implementing different limiter functions for different wave families is explored. The objective is to improve the numerical resolution of the discontinuities associated with the families two to four using compressive limiter functions (superbee), and without losing robustness mainly due to the use of diffusive limiter functions (minmod) for the wave families one and five. This technique implements the utilization of the superbee limiter function only in linear degenerate waves and the minmod function in nondegenerated nonlinear waves [1, 2].

To introduce in the numerical fluxes calculations the limiter function superbee, (2.5) is replaced by the following expression:

$$g^m_i = \begin{cases} 0 & \text{if } \alpha^m_{i+1/2}\alpha^m_{i-1/2} < 0 \\ \max[0, \min(2k, 1), \min(k, 2)]\dfrac{1}{2}\left|\lambda^m_{i-1/2}\right|\alpha^m_{i-1/2} & \text{if } \alpha^m_{i+1/2}\alpha^m_{i-1/2} \geq 0 \end{cases} \tag{2.8}$$

being

$$k = \frac{\left|\lambda^m_{i+1/2}\right|\alpha^m_{i+1/2}}{\left|\lambda^m_{i-1/2}\right|\alpha^m_{i-1/2}}. \tag{2.9}$$

To improve the overall scheme robustness, the implementation of different limiter functions is carried out only in those cells interfaces where the greater relative intensities of the discontinuities in central waves are registered, and using the conventional Harten-Yee TVD scheme in all other cases. Notice that the comparison among the intensity of the waves cannot be made using directly the coefficients of the spectral decomposition ($\alpha^m_{i+1/2}$) since these coefficients depend on the module assigned to each eigenvector.

In the local coordinate system adopted for computing the numerical fluxes across each face, the corresponding eigenvectors are given by [5]

$$\mathbf{K}_1 = \begin{bmatrix} 1 \\ u-c \\ v \\ w \\ H-uc \end{bmatrix}, \quad \mathbf{K}_2 = \begin{bmatrix} 1 \\ u \\ v \\ w \\ \dfrac{u^2+v^2+w^2}{2} \end{bmatrix}, \quad \mathbf{K}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ v \end{bmatrix}, \quad \mathbf{K}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ w \end{bmatrix}, \quad \mathbf{K}_5 = \begin{bmatrix} 1 \\ u+c \\ v \\ w \\ H+uc \end{bmatrix}, \tag{2.10}$$

where H is the stagnation enthalpy; u, v and w are the velocity vector components, and c is the sound velocity. It can be deduced from (2.10) that $\alpha^1_{i+1/2}$, $\alpha^2_{i+1/2}$ and $\alpha^5_{i+1/2}$ measure the density jump in the waves 1, 2, and 5, respectively, and that $\alpha^3_{i+1/2}$ and $\alpha^4_{i+1/2}$ measure the momentum jump in waves three and four. To compare these jumps it became necessary to select reference values for the density and velocity. Thus,

$$I_1 = \frac{\left|\alpha^1_{i+1/2}\right|}{\rho_{ref}}, \; I_2 = \frac{\left|\alpha^2_{i+1/2}\right|}{\rho_{ref}}, \; I_3 = \frac{\left|\alpha^3_{i+1/2}\right|}{\rho_{ref}u_{ref}}, \; I_4 = \frac{\left|\alpha^4_{i+1/2}\right|}{\rho_{ref}u_{ref}}, \; I_5 = \frac{\left|\alpha^5_{i+1/2}\right|}{\rho_{ref}}. \tag{2.11}$$

In this investigation, $\rho_{\text{ref}} = 0.5\ (\rho_i + \rho_{i+1})$ is taken as density reference, and as the velocity reference of the average of the sound velocities at the cells $u_{ref} = 0.5\ (c_i + c_{i+1})$, where c_i is the sound velocity. The parameters I_i permit to measure the wave intensities.

Finally, if the maximum of I_1, I_5 is higher than the maximum of I_2, I_3, I_4, the conventional Harten-Yee TVD scheme is used; otherwise, the values of g^2_i, g^3_i, g^4_i, are calculated with the limiter function superbee and g^1_i, g^5_i, with the limiter function minmod.

For the evaluation of g^m_i and g^m_{i+1} in (2.4), it is necessary to calculate the spectral decompositions of the conservative variables increments at the interfaces $i-1/2$, $i+1/2$, and $i+3/2$. In the context of three-dimensional not structured meshes of tetrahedrons, the identification of the cells i and $i+1$ is intuitive (they are two cells that share a face) but the determination of the points $i-1$ and $i+2$ is not direct. If two tetrahedrons that share a face are analyzed, the nodes not belonging to the common face can be used as representative points for $i-1$ and $i+2$. Then, these points can be used as imaginary cells. In this work these ideas have been implemented, being the nodal values calculated as a pondered average of the conservative variables between all cells that are in contact with the nodes $i-1$ and $i+2$. Such pondered average is given by

$$\mathbf{U}_{\text{node}\,k} = \frac{\sum_{i=1}^{n}\left(\mathbf{U}_{\text{cell}\,i}/\left(d_{GC_{\text{cell}\,i}-\text{node}\,k}\right)\right)}{\sum_{i=1}^{n}\left(1/\left(d_{GC_{\text{cell}\,i}-\text{node}\,k}\right)\right)}, \tag{2.12}$$

where $d_{GC_{\text{cell}\,i}-\text{node}\,k}$ is the distance that separates the gravity center of the cell i from the node k, and n is the cell number in contact with the node k.

The treatment of the boundary conditions is carried out through the imaginary cells technique [2–4]. Five different types of boundaries are considered (1) subsonic inlet; (2) supersonic inlet, (3) subsonic exit, (4) supersonic exit, (5) nonpenetration (solid boundary and symmetry).

2.2. Reentry Equations of Motion

The choice of a suitable set of coordinates and parameters of the trajectory to describe the movement of an object in atmospheric reentry is inherent to any investigation of guided spacecraft. To analyze a reentry trajectory it is appropriated to describe the motion of the center of mass using a set of elements known as Flight Coordinates [12].

Thus, the flight coordinates are described by the six orbital elements: magnitude of the position vector, r, longitude, θ, latitude, φ, magnitude of the velocity vector, v, flight-path angle, γ, and heading angle, ψ (azimuth of the velocity). At every moment this object is under the influence of a total force, $\overrightarrow{F}$, composed by the gravitational force, $\overrightarrow{F}_G$, the aerodynamic force, $\overrightarrow{A}$, and the force of propulsion, $\overrightarrow{T}$:

$$\overrightarrow{F} = \overrightarrow{T} + \overrightarrow{A} + \overrightarrow{F}_G. \tag{2.13}$$

The gravitational force is always present. For nonpowered flight, the propulsion force is zero, while for flights outside the atmosphere, the aerodynamic force vanishes.

To derive the equations of motion, we must use an Earth-fixed reference system. The kinematics equations of motion are [12]

$$\begin{aligned} \frac{dr}{dt} &= v \sin\gamma, \\ \frac{d\theta}{dt} &= \frac{v \cos\gamma \cos\psi}{r \cos\phi}, \\ \frac{d\phi}{dt} &= \frac{v \cos\gamma \sin\psi}{r}. \end{aligned} \tag{2.14}$$

It is desirable to separate the aerodynamic force into two components and define the tangential component of the nongravitational force, $\overrightarrow{F}_T$, along the velocity vector, and the normal component, $\overrightarrow{F}_N$, orthogonal to the velocity at the aerodynamic plane. When we have in plane flights, the normal vector $\overrightarrow{F}_N$ is in the plane $(\overrightarrow{r}, \overrightarrow{v})$, the vertical plane, and there is no lateral force. However, it is possible to create a lateral component of this force, which has the effect of changing the orbital plane. The non-gravitational force is then decomposed into a component on the vertical plane and orthogonal to the velocity vector, and a component orthogonal to this plane using the bank angle, σ.

The force equations are [12]:

$$\begin{aligned}
\frac{dv}{dt} &= \frac{1}{m}F_T - \frac{1}{m}F_G \sin\gamma + \omega^2 r\cos\phi\left(\sin\gamma\cos\phi - \cos\gamma\sin\psi\sin\phi\right),\\
v\frac{d\gamma}{dt} &= \frac{1}{m}F_N\cos\sigma - \frac{1}{m}F_G\cos\gamma + \frac{v^2}{r}\cos\gamma + 2\omega v\cos\psi\cos\phi\\
&\quad + \omega^2 r\cos\phi\left(\cos\gamma\cos\phi + \sin\gamma\sin\psi\sin\phi\right),\\
v\frac{d\psi}{dt} &= \frac{1}{m}\frac{F_N\sin\sigma}{\cos\gamma} - \frac{v^2}{r}\cos\gamma\cos\psi\tan\phi + 2\omega v\left(\tan\gamma\sin\psi\cos\phi - \sin\phi\right)\\
&\quad - \frac{\omega^2 r}{\cos\gamma}\cos\psi\sin\phi\cos\phi.
\end{aligned} \tag{2.15}$$

Here ω is the rotation of the Earth that appears because we have to consider a reference system fixed on the planet.

2.3. Attitude Equations

Knowing the attitude of a space object means knowing the orientation of an axis system connected to the vehicle related to a vertical reference system. To specify the orientation of a rigid body in space, three independent parameters are needed. These parameters are commonly known as roll, pitch and yaw, the Euler angles.

However, the use of the Euler angles to compute the attitude evolution of a spacecraft is limited: the equations of motion in attitude have singularities for certain values of the pitch angle, namely $\pm\pi/2$. This limitation was solved with the substitution of the Euler angles with a set of variables known as quaternions [13].

The quaternions are denoted as $\mathbf{q} = (q_1, q_2, q_3, q_4)$. The components of $\mathbf{q}$ are defined in terms of the Euler angles using the convention xyz (or sequence 321) which is found in Goldstein [14].

Kinematics equations of motion have the following form [14]:

$$\begin{aligned}
\dot{q}_1 &= \frac{1}{2}\left(\omega_3 q_2 - \omega_2 q_3 + \omega_1 q_4\right),\\
\dot{q}_2 &= \frac{1}{2}\left(-\omega_3 q_1 + \omega_1 q_3 + \omega_2 q_4\right),\\
\dot{q}_3 &= \frac{1}{2}\left(\omega_2 q_1 - \omega_1 q_2 + \omega_3 q_4\right),\\
\dot{q}_4 &= \frac{1}{2}\left(-\omega_1 q_1 - \omega_2 q_2 - \omega_3 q_3\right).
\end{aligned} \tag{2.16}$$

The attitude dynamic equations of motion express the temporal dependence of the angular velocity related to the applied torques:

$$\mathbf{I}\frac{d\vec{\omega}}{dt} = \vec{N} - \vec{\omega} \times \mathbf{I}\vec{\omega}, \tag{2.17}$$

where $\mathbf{I}$ is the inertia matrix, and $\vec{N}$ the aerodynamic torque. Thus, with I_1, I_2, I_3 corresponding to the moments of inertia about the main axes of the vehicle, the dynamic equations of attitude are [13]:

$$\begin{aligned}
\dot{\omega}_1 &= \frac{N_1 + (I_2 - I_3)\omega_2\omega_3}{I_1}, \\
\dot{\omega}_2 &= \frac{N_2 + (I_3 - I_1)\omega_3\omega_1}{I_2}, \\
\dot{\omega}_3 &= \frac{N_3 + (I_1 - I_2)\omega_1\omega_2}{I_3}.
\end{aligned} \tag{2.18}$$

2.4. Aerodynamic Forces

If the vehicle under consideration operates on a symmetry condition, the velocity vector defines this plane of symmetry. So the attitude of the object is properly described by the attack angle, α, which is the angle between the velocity vector relative to the atmosphere and a vehicle's baseline, normally the longitudinal axis.

The aerodynamic force is decomposed into two components: the force opposite to the direction of motion, called Drag -and part of the non-gravitational force $\vec{F}_T$ in (2.15), and the orthogonal component, called Lift (part of $\vec{F}_N$):

$$\begin{aligned}
D &= \frac{1}{2}C_D\rho_{\text{atm}}S_{\text{sat}}V^2, \\
L &= \frac{1}{2}C_L\rho_{\text{atm}}S_{\text{sat}}V^2.
\end{aligned} \tag{2.19}$$

Here, C_D and C_L are called drag and lift coefficients, respectively, ρ_{atm} is the atmosphere's density, S_{sat} is the satellite's area and V is the relative velocity between the space object and the atmosphere. For each object it is necessary to calculate specific coefficients at every moment of the trajectory. The code developed in the Department of Aeronautics of the UNC [1] evaluates these parameters for the conical object into consideration.

The first phase of the calculation is to identify the object's surface areas. So, the nodes of the tetrahedrons, whose faces form the surface of the object, are identified in the code input file.

The forces acting on each face have a magnitude equal to the pressure divided by the area. The direction is the incoming normal to the surface, and the point of application corresponds to the geometric center of the face. In this way, the resultant force on the body is calculated as the vector sum of all forces acting on the discretized surface. Finally, the

resulting force components in directions parallel and perpendicular to the flow velocity vector are projected, in order to obtain the drag and lift coefficients, respectively.

3. Results

The numerical simulations were performed using as a core calculation code the one developed at UNC [1]. For delineation of the geometric bodies and meshing, we used the application GID 8.0 with temporary license issued by the manufacturer.

GID has been developed as an interface for geometric modelling, meshing, income data, and display results of all types of numerical simulation programs. The different menus can be modified according to specific user needs. The graphical interface adapted to the code is designed to allow the entrance of initial conditions, to mark the object's surface for the calculation of forces, to define the number of iterations, and other parameters inherent to the code. The ultimate objective of the use of GID is writing the data file that enters the UNC's code and the subsequent display of results, from reading the output file written by the code.

The limit of iterations in the code can be determined by the number of steps, or the limit time independently. When it comes to any of these, the program completes its implementation.

The motivation of the simulations performed is to obtain the main aerodynamic characteristics (drag and lift coefficients) of a conical body. They are used in the calculation of the trajectory during reentry into the atmosphere [15]. Due to the fact that the original code did not have the calculation of the pressure forces on a body, the necessary sentences to perform a simple summation of forces on the faces of the tetrahedrons lying on the object were written. For the identification of these faces, the graphical interface developed in GID is used, which has an option to mark the surface of the body.

The angle of the cone (10°) was chosen arbitrarily, in order to guarantee that this value is small. The cone's length (1 m) was chosen to facilitate the calculations.

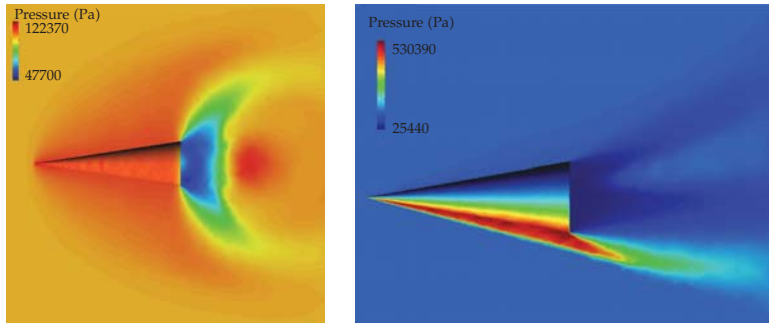
It is known that the aerodynamic characteristics of a body with the given geometry depend, in the case of nonviscous flow, only on the Mach number and the incidence angle of the free flow, without considering heat transfer or changes in the properties of air. That's the reason why these variables are used as independent ones.

Two sets of simulations were performed. The first case is for a cone-shaped object that enters in the atmosphere by sharp (narrow) side forward, while the second one was calculated in the assumption that the vehicle moves by wide (obtuse) side forward. Both calculations were made for several attack angles and Mach numbers.

It is important to note that the pressure distribution around the body depends only on the free flow Mach number. Then to obtain the C_D and C_L coefficients, it is necessary to modify only the free flow Mach number. For this reason the velocity is modified by keeping pressure and density constant and equal to those used in the remaining cases. The speed in the z direction is maintained equal to zero so there is flow symmetry around the xy plane.

3.1. Sharp Side Forward Case

The dimensions of the volume meshing are length in direction x = 2.2 m, length in direction y = 3 m, distance to surface point of impact in the axis x = 0.2 m, angle of revolution = 180°. The characteristics of the mesh are number of tetrahedrons around 130,000 and number of nodes around 29,000.



(a) (b)

Figure 1: Pressure distribution on the object: (a) Mach = 1, attack angle = 0°; (b) Mach = 4, a.a. = 15°.

The interface with GID requested that a new mesh was drawn every time a new calculation was made, because it needed a new data file for every case considered. For this reason, and because a nonstructured mesh was used, every case had a slightly different number of elements, due to arbitraries processes during meshing.

The criterion used for the meshing is based on

(i) using the symmetry of the flow on a plan to reduce the volume of mesh, and on two planes in the case of zero attack angle,

(ii) concentration of all elements near the surface of the cone, in particular areas considered critical for the calculation: impact point and base,

(iii) significant reduction of elements in remote areas of the body, to reduce computational cost.

Figures 1(a) and 1(b) display some examples of results obtained for the pressure field over a conical object considering two combinations of Mach number and attack angle.

The implemented numerical scheme has the capacity to simulate compressible flows in subsonic, transonic, and supersonic regimens. From Figure 1(a), it is possible to note, for Mach number equal 1, the concentration of sonic waves near of the body edge; however, the shock wave has not yet been formed. The shock wave can be seen clearly in Figure 1(b) when the free Mach flow is 4. Furthermore Figure 1 shows that the pressure distribution, including at the cone base, is different for transonic from that of supersonic flows , this explains the different behaviors of the C_D and C_L coefficients for different regimes.

Figure 2 shows the drag coefficients for the simulated cases, while Figure 3 shows the lift coefficients. In all cases it is considered the cone angle equal to 10°. The reference area used is the front section of the cone. It is important to note that the viscous effects are neglected to obtain theses figures.

Note from Figure 2 that drag coefficient is increasing accordingly as the attack angle of the cone is increasing too, this phenomenon occurs until 75°. From Figures 2 and 3, it is possible to observe that both coefficients, drag and lift, have lower variations for height Mach numbers and the greater variations occur at subsonic and transonic flows. For an attack angle of 45°, it is produced that the highest lift and the lower occur for an angle of 90°, for this test the lift is negative except for very reduced Mach number.

Figure 4 show some of the results of the reentry simulations. Figure 4(a) plots the variation of attack angle during the reentry trajectory for initial values of this angle

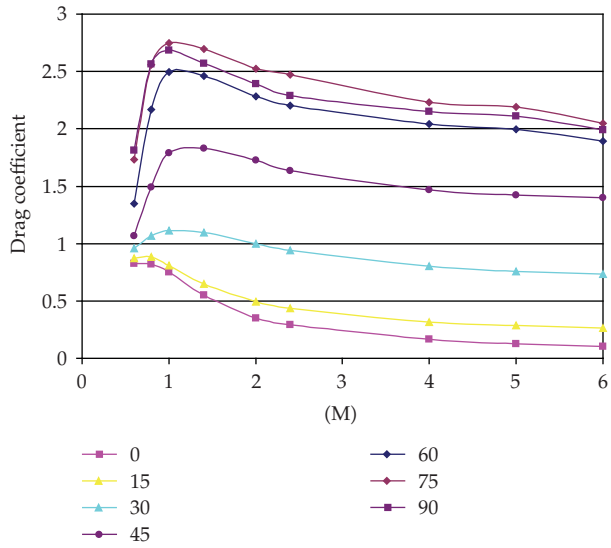


Figure 2: Sharp Side Forward: Drag coefficient as a function Mach for different angles of attack.

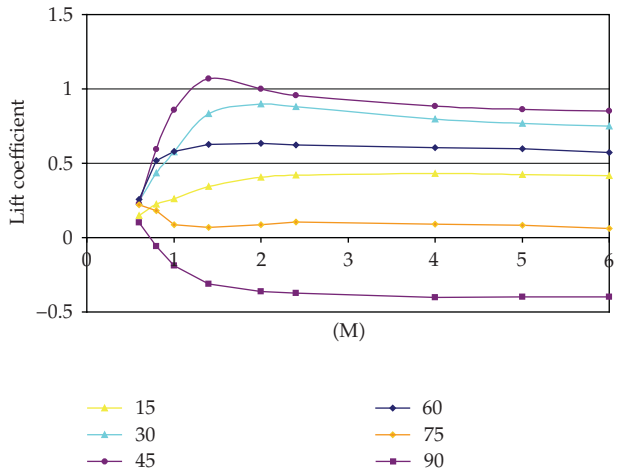


Figure 3: Sharp Side Forward: Lift coefficient as a function Mach for different angles of attack.

corresponding to 0°, 15°, and 30°, while Figure 4(b) presents the trajectory, meaning the altitude as a function of the time, for these angles.

We can notice the moment when the objects sense the presence of the atmosphere (around 1500 seconds) and begin to experience high variation of the attack angle during their descent. The first trend is to decrease the initial attack angle. However, although the atmosphere promotes an oscillatory movement around the zero attack angle condition, this is very unstable for objects reentering sharp side forward, and it is possible that for certain initial conditions, the atmosphere turns the object and reverses its attitude.

In what concerns the time evolution of the altitude, it could be seen that it is not affected until the height of 150 km. This is a limitation of the atmospheric model that has information on the atmospheric density just until this height. Since the aerodynamic forces

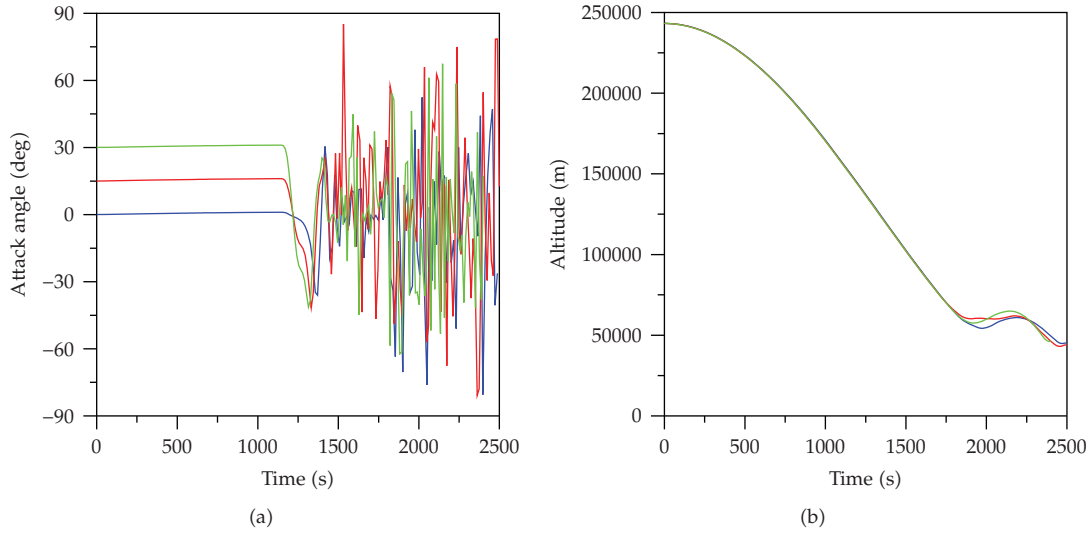


Figure 4: Sharp Side Forward: (a) Attack angles depending on the time; (b) Altitude. Initial attack angles: 0° (blue), 15° (red) and 30° (green).

depend directly on this parameter, it is assumed that the atmosphere is so faint over 150 km that it does not affect the trajectory. Anyway, under this limit, all trajectories show bouncing movements, indicating that the attitude variation affects the objects dynamics.

3.2. Wide Side Forward Case

In this case the characteristics of the mesh are number of tetrahedrons around 160,000 and number of nodes around 32,000. An additional criterion used for the meshing is based on the concentration of all elements over the cone's surface, in particular areas considered critical for the calculation: impact point and base. Also smaller elements were used in the zones nearby the cone base and along the axis of the cone in order to determine more precisely detached shock waves. Apart from that, the mesh was made thinner over the sides of the cone compared to the aerodynamic shadow in order to more accurately determine the expansion waves.

To simulate the atmospheric reentry of a cone shaped object by wide side forward, it is takes into account that this object has an attack angle around 180º. Some of the results found for this condition can be seeing on Figures 5 and 6.

Figure 5 shows the drag coefficients as a function of the Mach number for attack angles 105°, 120°, 135°, 150°, 165°, and 180°. It can be seen that this parameter presents higher values on the condition wide side forward compared with the same attitude angles on a sharp side forward situation for attack angles around the stability condition (180° in this case and 0° in the sharp side forward) as expected. For attack angles equal to or bigger than 135° (corresponding to 45° in the other case), the drag coefficient is very similar. On the other side, the lift coefficients tend to be smaller or maintain the same values (Figure 6).

The result of these differences can be noticed in Figures 7(a) and 7(b). For a conic object that reenters the terrestrial atmosphere wide side forward, small angles of attack (here around 180°) tend to preserve their magnitude during the reentry trajectory as corresponding

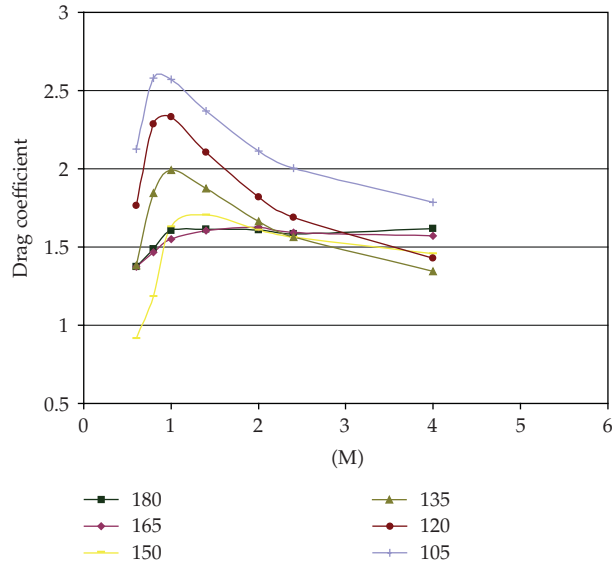


Figure 5: Wide Side Forward: Drag coefficient as a function Mach for different angles of attack.

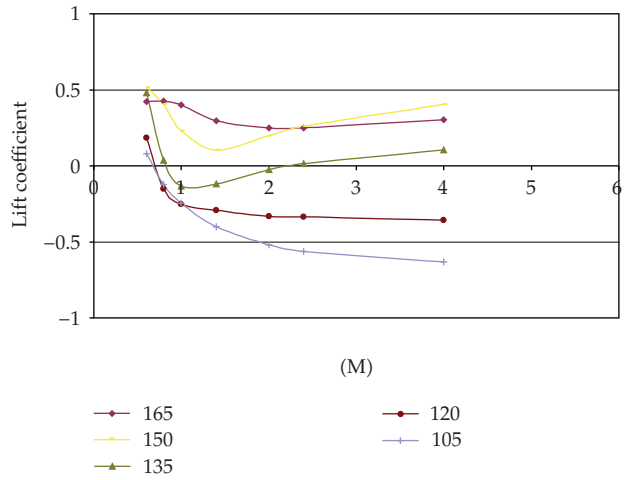


Figure 6: Wide Side Forward: Lift coefficient as a function Mach for different angles of attack.

to an equilibrium configuration. However, when the initial attitude is as far from 180° as 30°, equilibrium is lost during reentry and we found again the oscillations around the stability condition.

The most important consequence of the wide side forward reentry is over the altitude evolution. It can be seen in Figure 7(b) that all the cases present nearly ballistic paths. Although the far from the 180° initial attack angle condition, the less ballistic is the trajectory inside the dense part of the atmosphere, as can be seen by the dark green line that corresponds to an initial attack angle of 150°, the trajectories of initial attack angles equal to 180° and 175° are superimposed.

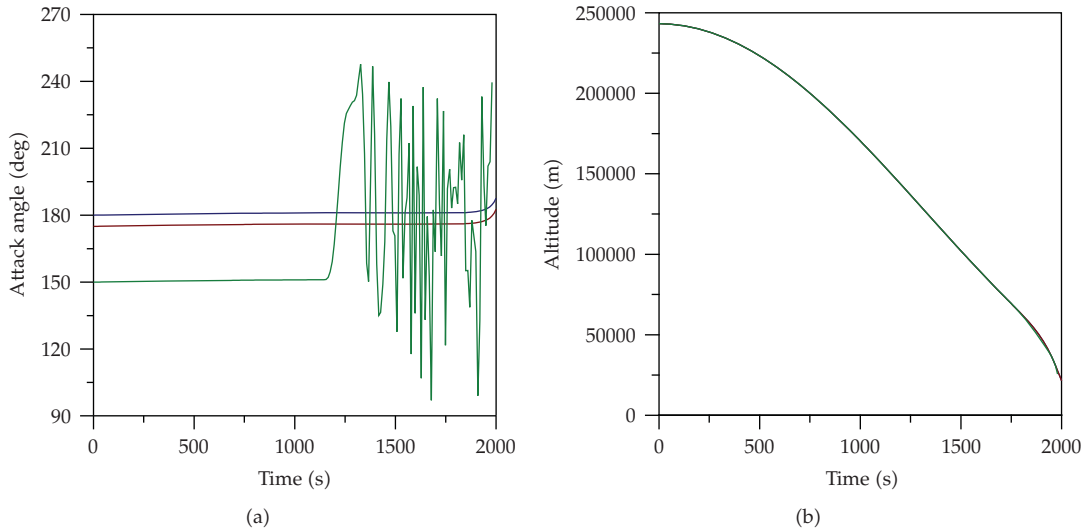


Figure 7: Wide Side Forward: (a) Attack angles depending on the time; (b) Altitude. Initial attack angles: 180° (dark blue), 175° (dark red) and 150° (dark green).

4. Conclusions

In this paper, a series of numerical simulations have been conducted using a code developed at UNC [1, 2], which solves the Euler equations by the method of finite volume, using unstructured 3D tetrahedral meshes. The code implements a new technique on the introduction of limiting functions, which aims to decrease the numerical viscosity, that is, increasing the contact discontinuities capture accuracy without loss of robustness regarding other TVD methods.

The goal was the calculation of aerodynamic characteristics of a cone under the effects of different flight conditions, to attach the results to other code that calculates the dynamic of atmospheric reentry [7].

It was possible to perform simulations of reentry trajectories for a specific cone under the influence of various aerodynamic effects for a variety of initial attack angles. Simulations show, as expected, that the trajectories are more affected when the object has initially a sharp side forward configuration.

From the obtained results, was compared the numerical slope of the normal force coefficient with analytical results. In none of the cases the errors exceed 6%. This good accuracy of the numerical results permits to induce that the error by not considering viscous effects in the calculation of the aerodynamic coefficients is low, and the obtained trajectories and attack angles evolution during the reentry are reliable.

Although the methodology implemented has been shown to be suitable for calculating the reentry trajectories inside the terrestrial atmosphere, it will be improved with the inclusion of viscous effects in the simulation of the aerodynamic flow in future works.

References

[1] O. A. Falcinelli, S. A. Elaskar, and J. P. Tamagno, "Reducing the numerical viscosity in non-structured three-dimensional finite volumes computations," *Journal of Spacecraft and Rockets*, vol. 45, no. 2, pp. 406–408, 2008.

[2] S. A. Elaskar, O. A. Falcinelli, J. P. Tamagno, and J. P. Saldía, "Further applications of scheme for reducing numerical viscosity: 3D hypersonic flow," *Journal of Physics: Conference Series*, vol. 166, Article ID 012018, 13 pages, 2009.
[3] C. Hirsch, *Numerical Computation of Internal and External Flows, Computational Methods for Inviscid and Viscous Flows*, vol. 2, John Wiley & Sons, New York, NY, USA, 1992.
[4] P. K. Sweeby, "High resolution schemes using flux limiters for hyperbolic conservation laws," *SIAM Journal on Numerical Analysis*, vol. 21, no. 5, pp. 995–1011, 1984.
[5] E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Springer, Berlin, Germany, 2nd edition, 1999.
[6] R. J. LeVeque, *Numerical Methods for Conservation Laws*, Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, Switzerland, 2nd edition, 1992.
[7] W. Schulz and P. Moraes Jr., "Reentry trajectory simulation of a small ballistic recoverable satellite," in *Advances in Space Dynamics 4: Celestial Mechanics and Astronautics*, H. K. Kuga, Ed., vol. 4, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil, 2004.
[8] B. Udrea, *An advanced implicit solver for MHD*, Ph.D. thesis, University of Washington, Seattle, Wash, USA, 1999.
[9] A. Harten, "On a class of high resolution total-variation-stable finite difference schemes," Tech. Rep., New York University, New York, NY, USA, 1982.
[10] A. Harten, "High resolution schemes for hyperbolic conservation laws," *Journal of Computational Physics*, vol. 49, no. 3, pp. 357–393, 1983.
[11] H. C. Yee, R. F. Warming, and A. Harten, "Implicit total variation diminishing (TVD) schemes for steady-state calculations," *Journal of Computational Physics*, vol. 57, no. 3, pp. 327–360, 1985.
[12] N. X. Vinh, *Optimal Trajectories in Atmospheric Flight*, Elsevier Scientific, Amsterdam, The Netherlands, 1981.
[13] J. R. Wertz, Ed., *Spacecraft Attitude Determination and Control*, D. Reidel, Dordrecht, The Netherlands, 1978.
[14] H. Goldstein, *Classical Mechanics*, Addison-Wesley Series in Physic, Addison-Wesley, Reading, Mass, USA, 2nd edition, 1980.
[15] J. P. Saldia, W. Schulz, and S. Elaskar, "Cálculo de la Dinámica de Reentrada de Objetos Espaciales," in *Proceedings of the 14th Colóquio Brasileiro de Dinâmica Orbital (CBDO '08)*, November 2008.