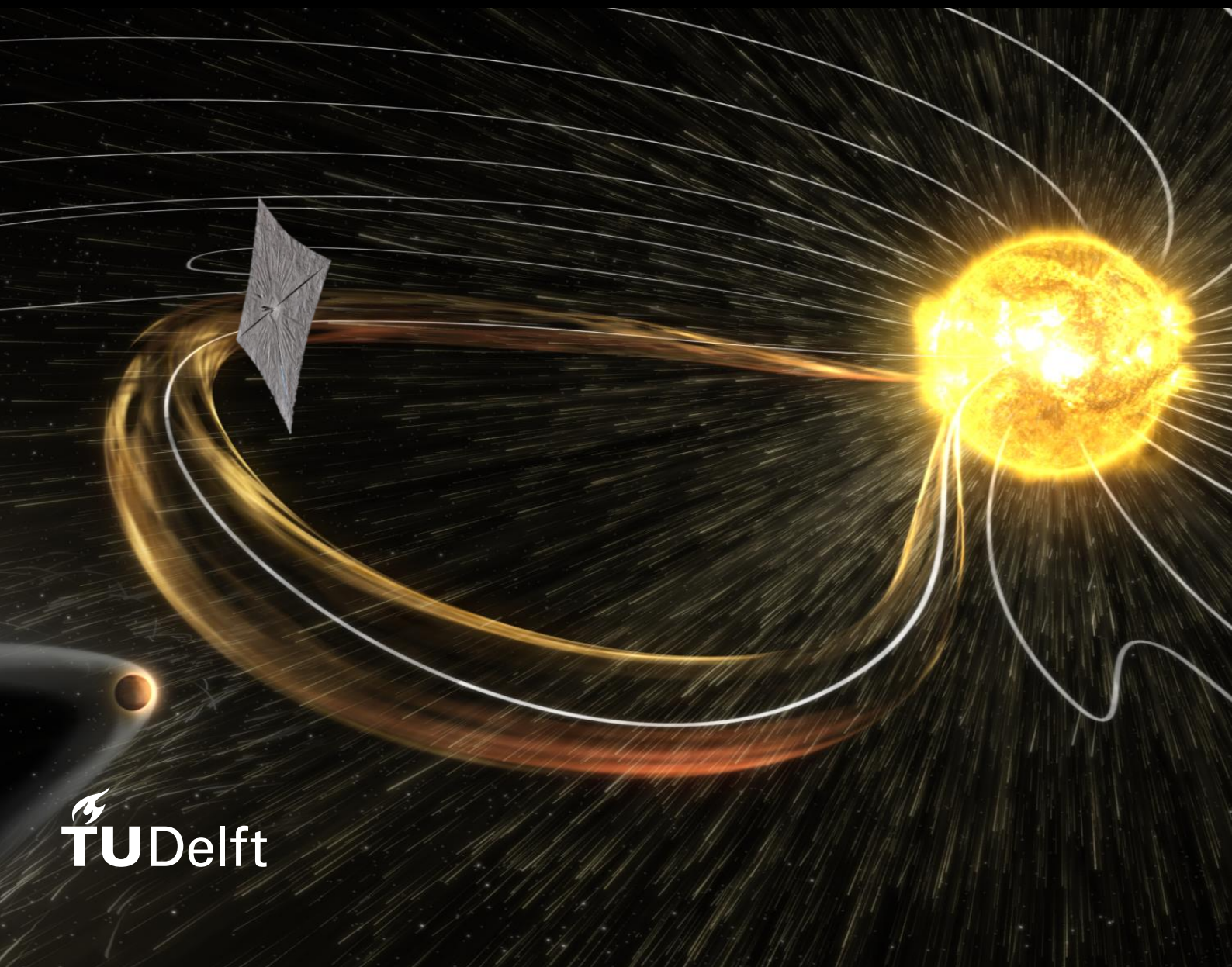
Solar-sail surfing along invariant manifolds to increase the warning time for solar storms

Master Thesis

Gonzalo Herrero Martínez

Technische Universiteit Delft



TU Delft

Solar-sail surfing along invariant manifolds to increase the warning time for solar storms

Master Thesis

by

Gonzalo Herrero Martínez

to obtain the degree of Master of Science in Aerospace Engineering
at the Delft University of Technology,
to be defended publicly on Wednesday, July 8th, 2020.

Supervisor:	Dr. M. Jeannette Heiligers,	Supervisor
Thesis committee:	Prof. Pieter Vieser,	Committee chair
	Dr. Ron Noomen,	Internal examiner
	Dr. Angelo Cervone,	External examiner

Credits cover NASA/GSFC,
https://svs.gsfc.nasa.gov/cgi-bin/details.cgi?aid=4370

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

I have always been amazed by Space. The books of Isaac Asimov were, in part, what helped me decide to choose the Space Exploration track from the Aerospace Engineering Master in TU Delft. I dream of the possibility of conquering space, becoming a multi-world civilization. Although this goal might not be feasible in my lifetime, I want to contribute to it. Solar sailing is a propellantless form of propulsion that harvests the energy of the Sun. Currently, It is considered for missions within our solar system and beyond. This thesis aims to design a near term mission using solar sails to protect the Earth for incoming coronal mass ejections from the Sun. Although this mission is still far from exploring other worlds, I consider it as a small step towards human expansion in the solar system.

I want to particularly thank Jeannette, my supervisor. Not only have you provided excellent technical advice and guidance during this thesis, but also you have given me all the flexibility I needed. I especially appreciate the opportunity you gave me to conduct my MSc thesis from Spain, alternating between Skype and presential meetings every two weeks. Finally, I thank you for your understanding when I started to work full-time at a company and needed to extend the project as a part-time thesis.

Gonzalo Herrero Martínez
Oslo during the quarantine, 2020

Abstract

Space-weather events have a large impact on Earth. In particular, Coronal Mass Ejections (CME) pose huge potential dangers to human technology both in orbit and on the surface of the planet, such as disruptions to power grids, increased radiation doses to astronauts and damage to sensitive components of satellites. Warnings for space-weather events are currently given by satellites in the vicinity of the L_1 point. When a CME passes by that point, the satellites emit a warning that reaches the Earth, on average, one hour before the CME. This thesis aims at using solar-sail technology to move a satellite closer to the Sun, detect the CME sooner, and thus increase the warning time.

Solar sails continuously generate thrust by reflecting solar photons off a large and highly reflective sail membrane. This continuous acceleration can be used to generate Artificial Equilibrium Points (AEPs) in the Circular Restricted Three-Body Problem (CRTBP) that are displaced away from the five classical Lagrange points. Like for the classical case, periodic orbits exist around these AEPs, enabling, for example, CME monitoring in a periodic orbit around an AEP that is located closer to the Sun than the Sun-Earth L_1 point from where current satellites detect CMEs. There have been some theoretical mission designs taking advantage of this possibility, but the increase in warning time is modest for any near-term sail performance.

This thesis investigates the use of solar-sail technology to travel upstream of the CME and significantly increase the warning time. The study considers the actual shape of CMEs as a constraint for the solar-sail trajectory that surfs along invariant manifold-like structures emanating from AEPs and Lyapunov orbits around sub-L_1 points, i.e., AEPs sunward of the classical L_1 point, to travel upstream of the CME.

As a preliminary solution, two strategies are evaluated. The first strategy considers a series of heteroclinic connections between different AEPs in the sub-L_1 region. The second strategy uses a homoclinic connection of a Lyapunov orbit around a sub-L_1 point. The trajectories aim to travel upstream along the path of the CME and back to the initial AEP or Lyapunov orbit to guarantee periodicity and, therefore, CME monitoring for as long as the sail remains operational. The homo- and heteroclinic connections are sought for by looking for connections between the unstable and stable manifolds emanating from the AEPs and Lyapunov orbits. To minimize the discontinuity in states at the linkage of the unstable and stable manifolds, a genetic algorithm approach is used to optimize the piece-wise constant attitude of the sail along the manifold trajectories and the location, i.e., which AEP or where along the Lyapunov orbit, from where the manifolds emanate.

Though the homo- and heteroclinic connections exhibit a discontinuity in the attitude of the sail at the connection of the unstable and stable manifolds, they provide a good initial guess for further optimization with a direct pseudospectral method, implemented in the software tool PSOPT. In the optimal control approach, the attitude of the sail is allowed to vary along the trajectory (instead of the piece-wise constant sail attitude in the genetic algorithm approach) such that the sail travels as far as possible upstream of the CME while staying as close as possible to the central axis of the CME.

The genetic algorithm results from both strategies show an improvement in warning time with respect to the warning time achieved by current satellites in the environment of the L_1 point. Under the assumptions taken in this research, the trajectories using the homoclinic connections from a Lyapunov orbit out-perform those that employ heteroclinic connections between AEPs. The best genetic algorithm solution offers an up to 10 times longer warning time than current satellites at L_1. This solution shows a small discontinuity in the states at the linkage of the unstable and stable manifold trajectories in the order of thousands of kilometers for the position and centimeters per second for the velocity. However, the discontinuity in the attitude of the sail of approximately 70 degrees renders the trajectory unfeasible before further optimization. Furthermore, some parts of the trajectory are too far from the axis of the CME to intercept CMEs approaching the Earth.

Finally, the trajectories obtained with PSOPT show that the sail remains within a defined distance to the axis of the CME while traveling upstream of the CME due to a control law that modifies the attitude of the sail at a rate achievable with state of the art technology. This strategy allows a 15 times longer average warning time compared to the warning time provided by current satellites at the L_1 point.

Contents

Nomenclature

Abbreviations

ACE	Advanced Composition Explorer
AEP	Artificial Equilibrium Point
AIAA	American Institute of Aeronautics and Astronautics
CME	Coronal Mass Ejections
CRTBP	Circular Restricted Three-Body Problem
DSCOVR	Deep Space Climate ObserVatoRy
EoM	Equations of Motion
GA	Genetic Algorithm
IKAROS	Interplanetary Kite-craft Accelerated by Radiation Of the Sun
JAXA	Japan Aerospace Exploration Agency
NASA	National Aeronautics and Space Administration
NEA	Near-Earth Asteroid
NOAA	National Oceanic and Atmospheric Administration
OKEANOS	Oversize Kite-craft for Exploration and Astronautics in the Outer Solar System
PCRTBP	Planar Circular Restricted Three-Body Problem
PSOPT	Pseudo-Spectral OPTimal (control solver)
SEP	Solar Energetic Particle
SOHO	SOlar and Heliospheric Observatory
STEREO	Solar TErrestrial RElations Observatory

Greek symbols

α	Cone angle
β	Lightness number
Δ	Variation
γ	Auxiliary angle for orbit parametrization
λ	Multiplier for cost function
μ	Dimensionless mass ratio
ω	Angular speed of the Earth and the Sun around their barycenter
Φ	State transition matrix
θ	Angle of line of stop condition
ε	Constant of small magnitude
φ	Angular half-width of CME

Latin symbols

a_s	Solar-sail acceleration
dir	Direction of the manifold
F	Cost function
F_B	Magnetic tension
F_G	Gravity force
F_H	Hydrodynamic streamlining force
H_X	Heaviside function along x-axis
H_Y	Heaviside function along y-axis
m_1	Mass of the main body (Sun)
m_2	Mass of the secondary body (Earth)
$\vec{n}$	Normal vector to the sail
$C(\hat{x}, \hat{y}, \hat{z})$	Sun-Earth synodic reference frame centered at the Sun-Earth barycenter
n	Front flattening of CME
$N_{eq_{points}}$	Number of equilibrium points
$N_{orbit_{points}}$	Number of points within a periodic orbit
$orbit$	Number of orbits within a family
p_1	Initial point

p_2	Final point
$\vec{R}$	Rotation matrix
$\vec{r}$	Position vector of the third body (sail)
$\vec{r}_1$	Sun-sail vector
$\vec{r}_2$	Earth-sail vector
R_t	Toroidal height of CME
r_{CME}	CME axis in polar coordinates
T	Period
t	Time
t_i	Time at node i
U	Gravitational plus centripetal potential
$u(t)$	Control variable
v	Velocity
V_c	Centripetal potential
V_g	Gravitational potential
W^{s+}, W^{s-}	Two branches of the stable manifold
W^{u+}, W^{u-}	Two branches of the unstable manifold
$\vec{x}(t)$	State vector
$\vec{x}_0$	Vector of initial conditions
$\vec{y^s}$	Eigenvector of the monodromy matrix associated to the stable manifold
$\vec{y^u}$	Eigenvector of the monodromy matrix associated to the unstable manifold
x, y, z	Position coordinates of the solar sail in $C(\hat{x}, \hat{y}, \hat{z})$

1 Introduction

Traveling through space without the need of carrying a limited amount of propellant that tightly constrains the possibilities of the mission sounds like science fiction. However, solar sails manage to achieve this as they generate a continuous thrust by reflecting solar photons off a large and highly reflective membrane. They can be used to modify the classical Circular Restricted Three-Body Problem (CRTBP) by including a continuous acceleration in the equations of motion. As such, solar sails allow new equilibrium points as well as broad possibilities for maneuvering throughout the solar system. Although solar-sail technology has been tested in space, these missions were only technology demonstration missions. One of the best candidates for the first scientific application of solar-sail technology is a space-weather warning mission, where a solar sail is placed between the Sun and the Earth to detect harmful space-weather events such as coronal mass ejections or solar winds before they reach the Earth [1]. The best example of such a mission was NASA's Sunjammer mission [2], which was scheduled for launch in January 2015 before it was canceled in October 2014. This thesis aims at designing a solar-sail mission to increase the warning time for solar storms even beyond the warning time that Sunjammer would have achieved with the hopes of a new space-weather mission that finally launches.

1.1. Solar sailing

The existence of solar radiation pressure was theoretically demonstrated by Maxwell in 1862 and its magnitude was first measured experimentally by Peter Levendew in 1900. Solar radiation pressure is extremely small, in the order of 9 N/km^2 at Earth's distance from the Sun [3]. Nevertheless, sufficiently large surfaces with a very low mass to area ratio - known as solar sails - can manage to generate a "free" acceleration to travel through the cosmos. In space, small accelerations take advantage of the absence of air resistance to build up large velocities for interplanetary missions. The acceleration acting on a solar sail is mostly directed normal to the surface of the sail regardless of its orientation with respect to the Sun, as illustrated by the force bubble in Figure 1.1 (left). However, the magnitude of the acceleration does depend on the orientation: it is maximum when the incident radiation is normal to the surface and zero when the radiation is parallel to the sail. Furthermore, it is not possible to generate a solar-sail acceleration with a component towards the Sun.

It took more than fifty years to transition from Levendew's experiment to actual mission designs using solar-sail technology. The first real advancements in solar sailing date back to the 1970s thanks to the appearance of the Space Shuttle with its capacity to take heavy payloads to space and the close passage of comet Halley predicted for the beginning of the 1980s. This particular comet, with a highly energetic retrograde orbit, had to be discarded for any possible rendezvous mission using conventional propulsion systems since the required ΔV was too large. Solar sailing, on the contrary, provided low but continuous thrust allowing the build-up of a considerable ΔV over a mission of some years and by exploiting close approaches to the Sun.

However, the proposed designs, a solar sail of 800 x 800 m^2 or a heliogyro with 12 blades of 7.5 km [3] were considered too risky and inappropriate for the first use of this technology. Still, the mission designs revived the interest in solar sailing and numerous papers were written [3, 4], enhancing the

marvelous opportunities of this technology. Furthermore, some campaigns such as the Moon race in the 1980s or the equivalent Mars race in the 1990s proposed a competition of solar-sail designs to reach these celestial bodies. More and more projects were envisaged [3, 5–7] and the concept of solar sailing was again extensively investigated. The problem was that none of these missions went much further than some theoretical designs and calculations. They were also too ambitious, proposing huge vehicles with long development times and unknown chances of success that no one dared to invest money in.

During the last decades, solar sailing has experienced significant development, reaching a readiness level that allowed four successful missions by the date of this publication. The first mission, IKAROS [8], was developed by JAXA, then NASA's NanoSail-D [9], and finally LightSail-1 and -2, shown in Figure 1.1 (right), from The Planetary Society [10]. There are also several missions planned for the near future, such as Near-Earth Asteroid (NEA) Scout from NASA [11] and the Oversize Kite-craft for Exploration and AstroNautics in the Outer Solar System (OKEANOS) by JAXA [12].

Figure 1.1: (Schematic drawing of solar-sail force bubble (left), edited from [13]. Lightsail-2 artist concept with Earth in the background (right) [14].

1.2. Space weather

Space weather relates to the activity of the Sun and its effect on Earth [15]. This activity is divided into Solar Energetic Particles (SEPs), solar flares, Coronal Mass Ejections (CMEs), and high-speed solar wind.

SEPs are electrically charged particles that travel at relativistic speeds along the magnetic field lines of the Sun and impact the Earth only if these magnetic field lines intersect those of the Earth. SEPs take from 20 minutes to several hours to cover the distance between the Sun and the Earth [16]. Solar flares release flashes of radiation from gamma-rays to radio waves, that will impact the Earth only if they are formed on the side of the Sun facing the Earth [17]. A. Isavnin defines CMEs as "large-scale explosive eruptions of magnetized plasma from the Sun into the heliosphere" [18]. CMEs can be ejected in any direction with varying speeds and cover the distance to the orbit of Earth within hours or days. CMEs will hit the Earth only under certain conditions and represent the main component of space weather. Finally, the solar wind is a mixture of ions and electrons that fills the space between the Sun and the planets. It escapes the Sun's outer atmosphere and travels up to the outer border of the heliosphere. High-speed solar wind originates in coronal holes and greatly influences the behavior of CMEs [15, 19].

Each space-weather event results in different effects on Earth. SEPs can be hazardous to satellite missions, damaging electronics and solar arrays, or blinding star-trackers, and, especially, they endanger astronauts by increasing their radiation doses to the maximum allowed for a lifetime within hours of the event [16]. Solar flares affect the ionosphere with negative effects upon radio navigation and communications. Furthermore, they can heat the atmosphere so it expands and drags satellites into lower orbits [17]. Solar winds and CMEs generate temporary disturbances on the Earth's magnetosphere when they impact the Earth, which is known as a geomagnetic or solar storm. These storms disrupt electric power grids and speed up the corrosion process of oil and gas pipelines. Some examples of these effects are: in September of 1859 a solar storm stopped all telegraph communications within the United States and Europe and in March 1989, a hydro-electrical power network in Quebec collapsed, leaving 6 million people in an energy blackout for more than 9 hours [20]. Nowadays, with the increased

technological dependency of humanity, a strong solar storm may have catastrophic consequences. In Reference [21], J. P. Eastwood estimates the cost of several CME impact events and the likelihood of the events: “For a 1-in-10-year sub-storm over western Europe, the direct cost is estimated to be €9340 million, with estimated international spillover costs in the range of €787−1108 billion”. For a 1-in-100-year event, such as the one in March 1989, the cost will reach the amount of trillions of Euros [21].

The magnitude of the consequences of these events shows the need to react before they reach Earth. There are two possibilities: detection and prediction. Prediction works in a similar way as weather forecasting for Earth’s surface events, using information about future events that may impact the Earth provided by solar observations. Detection requires in-situ measurements made by satellites with dedicated payloads placed along the path of the event. Then, the satellites can send a warning for an event that will impact the Earth, so operators on the Earth will have some time to react and protect sensitive infrastructure: satellites can turn-off their most sensitive components, astronauts can start a safety procedure to enter a more shielded area of the station, and power stations can be prepared for unusual activity [20]. The warning time will be proportional to the distance from Earth at which the event is detected. Since SEPs and solar flares travel at speeds close to the speed of light, detecting these events will provide almost no reaction time; the only solution being the observation of the Sun to predict them [16, 17]. On the other hand, CMEs travel much slower than the speed of light, which offers the possibility of both observation and detection [18, 22].

1.3. Missions for space-weather forecasting

Numerous NOAA, ESA, and NASA satellites study both the Sun and the Earth to understand the origin and the effect of solar storms to reduce the consequences on vital space and ground infrastructure. Some of these satellites orbit the Sun-Earth L_1 point, such as the Deep Space Climate Observatory (DSCOVR) (NOAA/NASA, 2015), which gives real-time solar wind observations; WIND (NASA, 2004), that studies the interaction between the solar wind and the Earth’s iono- and magnetosphere; the Advanced Composition Explorer (ACE) (NASA, 1997) provides coverage for solar wind and measures the intensity of SEPs; and the Solar and Heliospheric Observatory (SOHO) (ESA/NASA, 1995), which has detected more than 7000 CMEs [23]. Other missions like the twin Solar Terrestrial Relations Observatories (STEREO A and B) (NASA, 2006) follow two heliocentric elliptical orbits, one closer to the Sun and one further from the Sun with respect to Earth’s orbit [24], to collect data from CMEs.

The majority of the current missions for space-weather forecasting orbit in the vicinity of the Sun-Earth L_1 point, approximately 1.5 million km closer to the Sun than the Earth. Due to the speed difference between CMEs and the warning message, the warning time that these missions can provide is one hour on average. Solar sails can be used to increase the distance from the Earth at which the CME is detected. When the continuous acceleration from a solar sail is added to the CRTBP, the well-known five classical Lagrange points [25, 26] become surfaces of artificial equilibria that span around these classical points [3]. The surface of equilibria in the environment of the L_1 point intersects the Sun-Earth line at the so-called sub-L_1 point, which lies sunwards of the classical point at a distance given by the efficiency of the solar sail. This sub-L_1 point has been the target of numerous solar-sail mission designs to increase the warning time for solar storms over the last two decades. In Reference [1], the author proposes injecting a sail into a sub-L_1 point and develops a series of trajectories to achieve periodic motion around the sub-L_1 point. Similarly, the Sunjammer mission [2] was a NASA project for increased solar-storm warning time that aimed at injecting a solar sail in a halo orbit around a sub-L_1 point. The launch date was planned for January 2015, but the mission was canceled in October 2014. Finally, in Reference [27], the authors also planned to inject a solar sail into a halo orbit around a sub-L_1 point but assumed that the vehicle is filled with SpaceChips [28], i.e., very small satellites with a high area to mass ratio offering a behavior similar to solar sails. These SpaceChips would then be deployed from the main spacecraft to follow the unstable manifold of the periodic orbit towards the Sun. On their journey along the manifold, they will perform observations of the Sun, greatly increasing the warning time for CMEs (between 4.5 and 9 hours). The main disadvantage of this mission was the finite quantity of these small satellites that can be carried and therefore limited mission lifetime.

1.4. Research goals and questions

The objective of this investigation is:
to increase the warning time for potential CMEs approaching the Earth with respect to current mission and mission designs by designing a solar-sail trajectory that travels upstream along the axis of the CME.
The current missions and mission designs detect the CMEs at different points along the Sun-Earth line, i.e., the classical L_1, the displaced sub-L_1 point, or closer to the Sun in the unstable manifold of an orbit around the sub-L_1 point. Using the approach proposed in this thesis, the sail will leave the Sun-Earth line to follow, as closely as possible, the path of CMEs directed towards the Earth. This novel approach allows the detection of the CME with longer average warning times for similar sail efficiencies than the previously discussed missions. For this, the invariant manifolds of different artificial equilibrium points in the environment of the L_1 point and Lyapunov orbits around a sub-L_1 point are exploited to generate hetero- and homoclinic connections that take the sail closer to the Sun along the path traveled by the CME.

To achieve this goal, the research should answer the following questions:

A. Is it possible to use heteroclinic connections between different solar-sail artificial equilibrium points in the sub-L_1 region to travel upstream of the CME using a piecewise-constant sail attitude?

B. Is it possible to use a homoclinic connection of a Lyapunov orbit around a sub-L_1 point to travel upstream of the CME using a piecewise-constant sail attitude?

C. Is it feasible to optimize the sail attitude from a piecewise-constant attitude to a continually variable attitude to further increase the warning time for solar storms?

It is expected that the answers to these questions will substantially increase the flexibility to design a solar-sail mission to adequately warn for Earth-approaching CMEs.

1.5. Report outline

The main content of the thesis is presented as a draft article in Chapter 2. The draft article style of the American Institute of Aeronautics and Astronautics (AIAA) is chosen due to the relation between the content of this thesis and their journal. The paper starts with an abstract and an introduction. Then Section II shows how to model the shape of CMEs. Section III explains the dynamics employed for the study and presents the concept of invariant manifolds. Section IV shows the preliminary results along invariant manifold-like structures employing a piece-wise constant attitude. Then, in section V, the best candidate from the options presented in Section IV is optimized with an optimal control solver (PSOPT) to further increase the warning time and fix the discontinuities from the piece-wise constant attitude trajectory. Finally, the results are discussed in Section VI and the conclusions presented in Section VII.

After the article draft, in Chapter 3 the research questions are answered in line with the results presented in the paper and the possibilities to expand the work are discussed. Finally, Appendix A includes the verification of the simulations used for the research.

2

Journal Article

Solar-Sail Surfing Along Invariant Manifolds to Increase the Warning Time for Solar Storms

Gonzalo Herrero Martinez*
Delft University of Technology, 2629 HS, Delft, The Netherlands

This paper proposes a new mission strategy to detect Coronal Mass Ejections (CMEs) and greatly increase the warning time for these CMEs. The warning time is proportional to the distance from the Earth at which the CMEs are detected, where the current warning time of one hour is achieved by satellites orbiting the Sun-Earth L_1 point. To date, several mission designs have taken advantage of solar-sail technology to displace the L_1 point sunward to a so-called sub-L_1 Artificial Equilibrium Point (AEP) and as such place the satellite closer to the Sun. However, the increase in warning time is modest for any near-term sail efficiency. We investigate the use of the invariant manifolds emanating from AEPs in the L_1 region and Lyapunov orbits around sub-L_1 points to generate a trajectory that travels upstream of the CME to increase the distance from the Earth at which the CME is detected and, consequently, increase the warning time. Initial results are obtained with a genetic-algorithm approach, where homoclinic connections of a solar-sail Lyapunov orbit provide the largest increase in warning time. These connections are then used as initial guesses for solving the optimal control problem with a direct pseudospectral method to reduce the distance to the axis of the CME and increase the warning time to an average of 15 hours.

I. Introduction

Coronal Mass Ejections (CMEs) are large-scale explosive eruptions of magnetized plasma from the Sun into the heliosphere [1] and they represent the main contribution to solar storms [2]. The solar wind is a mixture of ions and electrons that fills the space between the Sun and the planets and greatly influences the behavior of CMEs [2–4]. Solar winds and CMEs generate temporary disturbances on the Earth's magnetosphere when they impact the Earth, which is known as a solar storm. If undetected, these storms can result in disastrous consequences such as disruptions in electric power grids or accelerated corrosion of gas and oil pipelines. In Reference [5], the authors estimate the cost of a non-detected, 1-in-100-years CME to be in the order of trillions of Euros.

Detection requires in-situ measurements of the CME, for which satellites with dedicated payloads are placed along its path. Upon detection of the CME, the satellites send a warning to Earth, so that the operators on Earth have time to

*Graduate Student, Department of Astrodynamics and Space Missions, Faculty of Aerospace Engineering, G.J.HerreroMartinez@student.tudelft.nl

react and protect sensitive infrastructure. Numerous NOAA, ESA, and NASA satellites use the Sun-Earth L_1 point as a vantage point to detect CMEs and solar winds. Some examples include DSCOVR (NOAA/NASA, 2015), WIND (NASA, 2004), ACE (NASA, 1997), and SOHO (ESA/NASA, 1995) [6]. SOHO by itself has detected more than 7000 CMEs [7]. The warning time is proportional to the distance from Earth at which the CME is detected, so satellites at the L_1 point can achieve warning times of approximately one hour [2, 5].

Solar sails can be used to increase the distance from the Earth at which a CME is detected. When the continuous acceleration from a solar sail is added to the Circular Restricted Three-Body Problem (CRTBP), the well-known five classical Lagrange points [8, 9] become surfaces of equilibria that span around these classical points [10]. The surface of equilibria in the environment of the L_1 point intersects the Sun-Earth line at the so-called sub-L_1 point, which lies sunward from the classical point at a distance determined by the efficiency of the solar sail: the more efficient, the closer to the Sun.

During the last 30 years, solar sailing has experienced significant development, reaching a readiness level that allowed four successful missions by the date of this publication. The first mission, IKAROS [11], was developed by JAXA, then NASA's NanoSail-D [12], and finally LightSail-1 and -2 [13] from The Planetary Society. There are also several missions planned for the near future, such as Near-Earth Asteroid (NEA) Scout [14] from NASA (launch date to be determined between 2020-2021) and the Oversize Kite-craft for Exploration and AstroNautics in the Outer Solar System (OKEANOS) by JAXA (launch date is set for the year 2027) [15].

In the literature, several proposals can be found that use solar sails to design missions that increase the warning time for CMEs [16–19]. In Reference [16], the author designs a series of trajectories that keep the solar sail in the environment of a sub-L_1 point, providing an increased warning time with respect to the aforementioned satellites at the L_1 point. Similarly, NASA's Sunjammer mission [17] proposed a halo orbit around a sub-L_1 point to increase the warning time. In Reference [18], the authors take advantage of SpaceChips, i.e., very small satellites with a high area to mass ratio that behave like solar sails [19], by releasing them from a spacecraft in a halo orbit around a sub-L_1 point into its associated unstable manifold to travel sunwards along the Sun-Earth line for a limited amount of time. Finally, the authors of [20–22] study the dynamics of the Sun-Earth system with a solar sail to navigate between Artificial Equilibrium Points (AEPs). During the motion between these AEPs, the sail gets closer to the Sun than when it would remain at the AEP itself, again allowing a modest increase in the warning time for CMEs. These mission proposals always consider the CME traveling along the Sun-Earth line, which works for small distances from the Earth. However, to greatly increase the warning time, the actual path of CMEs as they travel through the Solar System [1] has to be considered, as it then becomes possible to design a solar-sail trajectory that truly follows the axis of the CME upstream of its interplanetary trajectory.

In this paper, we show the design of a periodic trajectory in which the sail travels upstream of the CME axis to increase the warning time for CMEs. We evaluate the feasibility of two strategies: a trajectory composed of two

heteroclinic connections between two AEPs belonging to the family of AEPs in the sub-L_1 region and a homoclinic connection of a Lyapunov orbit around the sub-L_1 point. Both strategies are optimized with a genetic algorithm to maximize the warning time and minimize the discontinuity in states at the linkage between the unstable and stable manifolds of the homo- and heteroclinic connections. Then, the best strategy is used as the initial guess for an optimal control solver that further optimizes the trajectory to eliminate the linkage error and increase the average warning time that can be obtained with the trajectory.

After this introduction, Section II shows how to model a CME in the CRTBP. Section III presents the dynamics of the solar sail in the planar CRTBP. It includes the computation of artificial equilibria, periodic orbits around the equilibria, and the invariant manifolds emanating from the equilibria and the periodic orbits. Section IV contains the search for optimal trajectories using a piecewise-constant attitude for the sail using a genetic-algorithm approach. In Section V, the best solution from Section IV is optimized with an optimal control solver to further increase the warning time. Then, Section VI discusses the results. Finally, the paper ends with the conclusions in Section VII.

II. Modelling Coronal Mass Ejections

The path or structure of a Coronal Mass Ejection (CME) can be modeled as a cylinder in equilibrium along the stream of solar wind radially outflowing from the Sun. The equilibrium is given by a balance between the forces of hydrodynamic streamlining, F_H, magnetic tension, F_B, and gravity, F_G, as $F_H = F_G + F_B$ [1]. The model is completed by including some deformations such as front flattening, pancaking, and skewing. Front flattening represents a compression along the Sun-Earth line, pancaking relates to the non-circular cross-section of the CME and skewing is caused by a rotation along an axis normal to the plane in which the Sun and the Earth orbit. To model the axis of the

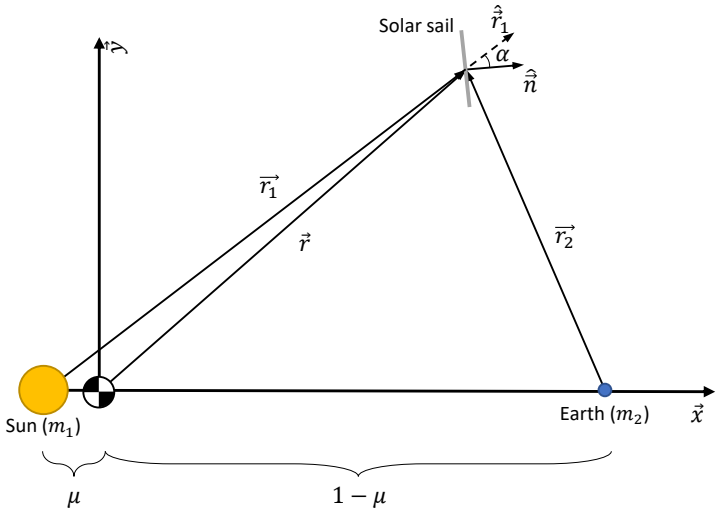


Fig. 1 Schematic drawing of a solar sail in the PCRTBP.

CME, a Sun-Earth synodic reference frame, $C(\hat{x}, \hat{y}, \hat{z})$ is used. The origin is set at the center of mass of the system, the x-axis lies along the line connecting the Sun and the Earth and points towards the Earth, the z-axis is directed perpendicular to the plane in which the Sun and the Earth orbit, and the y-axis completes the right-handed reference frame. This frame rotates at a constant angular velocity ω around its z-axis, $\vec{\omega} = \omega\hat{z}$. Figure 1 shows a projection of the reference frame employed, $C(\hat{x}, \hat{y}, \hat{z})$, onto the ecliptic plane. The unit of mass is defined as the sum of the masses of the system $m_1 + m_2 = 1$. Then, with the mass ratio $\mu = \frac{m_2}{m_1+m_2} = 3.0404 \cdot 10^{-6}$, the dimensionless masses of the massive bodies become $m_1 = 1 - \mu$ and $m_2 = \mu$. The unit of length is set as the distance between both massive bodies. Then, the distance from each massive body to the center of mass of the system is μ and $1 - \mu$ for the Sun and the Earth, respectively. Finally, the unit of time is chosen such that the orbital period of the Sun and the Earth around their barycenter is 2π, then the angular velocity of the reference frame becomes $\omega = 1$.

In this reference frame, the CME axis is defined in polar coordinates (r_{CME}, φ), with $r_{CME} = 0$ representing the location of the Sun, and where φ is measured in the ecliptic with $\varphi = 0$ at the Sun-Earth line [1]:

$$r_{CME}(\varphi) = R_t \cos^n(a\varphi) \tag{1}$$

where R_t is the toroidal height of the CME, set as the distance between the Sun and the Earth, $n = 0.5$ is the front flattening coefficient, and $a = (\pi/2)/\varphi_{hw}$, where $\varphi_{hw} = 30°$ is the angular half-width of the axis. Due to the dynamics of the Solar System, the CMEs evolve in an anti-clockwise manner, which limits the in-situ observation of incoming CMEs to the area between the Sun-Earth line and the quarter of the Earth's orbit immediately trailing the Earth's instantaneous position. Therefore, while the axis defined by Eq. (1) spans over the length of the CME, we are only interested in the fraction of the axis upstream of the Earth, see the dashed blue line in Figure 2.

III. Dynamics

This study is conducted within the well-known Sun-Earth Planar Circular Restricted Three-Body Problem (PCRTBP) dynamical framework [8, 9] augmented with a solar sail [10], that defines the motion of a massless particle (the solar sail) with respect to two massive bodies, the Sun (m_1) and the Earth (m_2), which orbit in a circular motion around their barycenter. In this framework, the motion of the solar sail in reference frame $C(\hat{x}, \hat{y}, \hat{z})$ is given by [9, 10]:

$$\ddot{\vec{r}} + 2\vec{\omega} \times \dot{\vec{r}} + \vec{\omega} \times (\vec{\omega} \times \vec{r}) = \vec{a}_s + \nabla V_g \tag{2}$$

where $\vec{r} = [x \quad y \quad z]^T$ is the position of the sail. The terms on the left-hand side of Eq. (2) are the kinematic, coriolis, and centripetal accelerations. The terms on the right-hand side are the sail acceleration and the gravitational acceleration exerted by the main bodies. The centripetal acceleration in Eq. (2) can be expressed as the gradient of a scalar potential

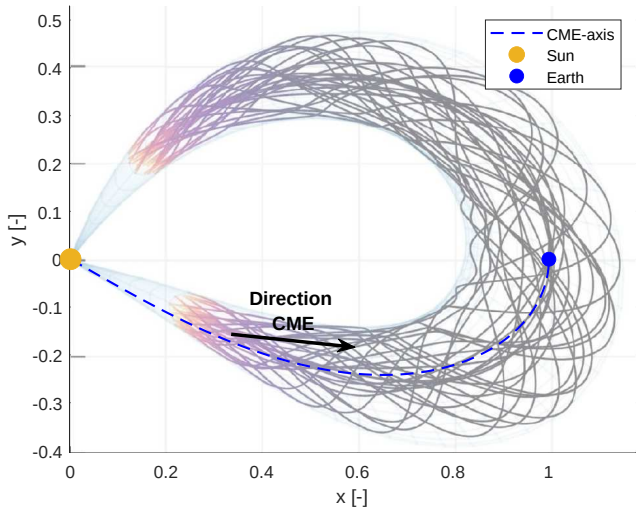


Fig. 2 Generic CME model projected onto the ecliptic. The dashed blue line represents the axis of the CME. Edited from [1].

function, $V_c = -\frac{1}{2}||\vec{\omega} \times \vec{r}||^2$. Then, the gravitational potential, V_g, can be combined with the centripetal potential into a new, effective potential, U [9]:

$$U = \frac{1}{2}(x^2 + y^2) + \frac{1-\mu}{r_1} + \frac{\mu}{r_2} \tag{3}$$

where $\vec{r}_1 = [(\mu + x) \quad y \quad z]^T$ is the Sun-sail vector and $\vec{r}_2 = [(x - (1-\mu)) \quad y \quad z]^T$ is the Earth-sail vector. As the research is conducted within the ecliptic plane, the vector components in the direction normal to this plane will always be zero, $z = 0$. Therefore, to ease the notation, the three-dimensional vectors will be expressed only with the two first coordinates $\vec{r} = [x \quad y]^T$. Finally, for the sail acceleration, we use an ideal-sail model, which assumes pure specular reflection of the incident photons. With this assumption, the solar-sail acceleration is defined as [10]:

$$\vec{a_s} = \beta \frac{1-\mu}{r_1^2} \cos^2 \alpha \, \hat{\vec{n}} = [a_{s_x}, \, a_{s_y}]^T \tag{4}$$

where β is the lightness number of the sail, α is the cone angle of the sail, and $\hat{\vec{n}}$ is the unit vector in the direction normal to the sail. The lightness number is defined as the ratio between the solar radiation pressure acceleration and the solar gravitational acceleration of the sail. In 2014, the Sunjammer mission was designed with a lightness number of $\beta = 0.0363$ [17]. In Reference [23], the authors suggest lightness numbers up to $\beta = 0.067$ as a near-term possibility. For the study presented in this paper, we have considered a modest lightness number of $\beta = 0.04$ as the reference value, which is a realistic value for present and near-term solar-sail missions. The cone angle defines the orientation of the sail and is measured clockwise as the angle between the normal to the sail $\hat{\vec{n}}$ and the direction of the incident radiation. It is defined as: $\cos \alpha = \hat{\vec{r}}_1 \cdot \hat{\vec{n}}$. The normal vector $\hat{\vec{n}}$ is obtained by a clockwise rotation of the unit vector in the Sun-sail

direction, $\hat{\vec{r}}_1$, around the z-axis over an angle α:

$$\hat{\vec{n}} = \vec{R}_\alpha \hat{\vec{r}}_1 = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \hat{\vec{r}}_1 \tag{5}$$

Note that several models represent the solar-sail acceleration with higher fidelity than the ideal model used. The non-ideal behavior of a sail is divided into three categories: attitude control, shape deformations, and optical imperfections [24]. The non-ideal behavior associated with attitude control mainly requires to reduce the maximum cone angle to, at least, 85 degrees. The non-ideal shape effects can be simplified with a small reduction in the sail performance. On the other hand, non-ideal optical effects have a larger impact on the sail behavior. As presented in Reference [24], a non-ideal model including numerous optical imperfections shows a mismatch in the thrust of up to 10% for cone angles below 60 degrees, while the mismatch increases to more than 100% as the cone angle approaches 90 degrees. The authors suggest a simplified non-ideal model where only 84% of the light is reflected specularly while the rest is absorbed with no re-emission. Since the current paper covers a preliminary mission design, the only non-ideal effect considered is a reduction in the maximum cone angle to 80 degrees. The remaining non-ideal effects are left for future research.

A. Equilibrium Points

The CRTBP, and equivalently the PCRTBP, exhibits five well-studied classical equilibrium points in the xy-plane at locations where the condition $\nabla U = \bar{0}$ is satisfied [8, 9]. The inclusion of the solar sail modifies the dynamics of the system such that the five classical points evolve into surfaces of equilibria [10, 25]. These surfaces are determined by the sail orientation and can be parametrized by the lightness number, β. Each point within these surfaces is known as an Artificial Equilibrium Point (AEP). As in the classical problem, the AEPs are located where no acceleration acts on the spacecraft when it has zero velocity. Substituting these conditions into Eq. (2) shows that the sail acceleration needs to equal the gradient of the potential at the AEP [25]:

$$\nabla U = \beta \frac{1-\mu}{r_1^2} \cos^2\alpha \, \hat{\vec{n}} \tag{6}$$

which can only be satisfied if $\hat{\vec{n}}$ is parallel to ∇U. Since $\hat{\vec{n}}$ is a unit vector, $\hat{\vec{n}} = \nabla U / |\nabla U|$ yields the appropriate orientation for the sail. Then, from Eq. (6) and the definition of the cone angle, we can derive an expression that defines the required lightness number to generate an AEP as a function of the sail position and orientation within the PCRTBP with

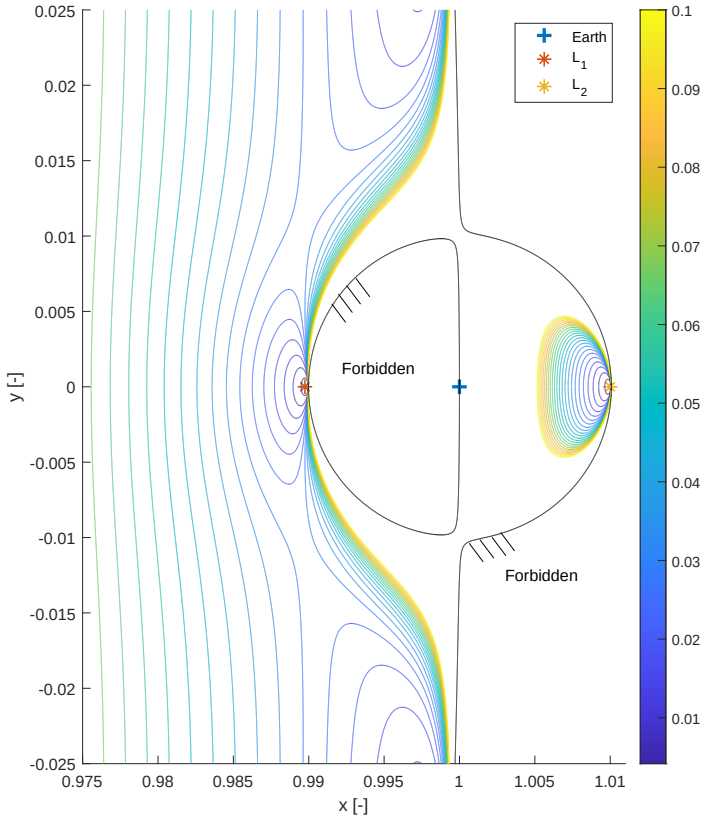


Fig. 3 Family of AEPs for different lightness numbers between 0 and 0.1 in the neighbourhood of the classical L_1 and L_2 points.

a solar sail as [25]:

$$\beta = \frac{r_1^2}{1-\mu} \frac{\nabla U \cdot \hat{\vec{n}}}{(\hat{\vec{r}}_1 \cdot \hat{\vec{n}})^2} \tag{7}$$

Figure 3 shows a contour plot of the required values for the lightness number close to the classical L_1 and L_2 points. In the case where the lightness number equals zero, the surfaces of AEPs are reduced to the classical five equilibrium points. As the lightness number increases, the surfaces expand around the classical points. We refer to the surfaces of equilibria in the vicinity of the L_1 point as the sub-L_1 region. Furthermore, the intersection of these surfaces of equilibria with the x-axis will be referred to as the sub-L_1 points, which are displaced closer to the Sun as the lightness number increases. The forbidden regions between L_1 and the Earth and farther away from L_2 appear due to the physical limitations of the solar sail, as it is not possible to direct the sail acceleration towards the Sun. Mathematically, this constraint can be expressed as $\hat{\vec{r}}_1 \cdot \hat{\vec{n}} \geq 0$ or equivalent limits of ± 90 degrees in the cone angle. The white areas outside the forbidden zones require larger lightness numbers than the range shown in Figure 3.

B. Periodic Orbits

The classical equilibrium points allow periodic motion around them that results in different types of orbits, such as Lyapunov, Lissajous, or halo orbits [8, 9]. The AEPs presented in Section III.A also allow the existence of these kinds of orbits [10, 25]. For example, displaced halo orbits such as the one considered as a destination for the previously proposed Sunjammer mission [17].

At any of the equilibrium points, we can find families of orbits that satisfy the conditions for periodic motion [21]. For a given sub-L_1 point (for a given lightness number), we can obtain the initial conditions of a periodic orbit from a catalog. In particular, we obtain the initial conditions of a planar Lyapunov orbit around the sub-L_1 point for $\beta = 0.03$ from [25].

To obtain periodic orbits from the initial condition in Reference [25], a differential correction scheme [8] is used, keeping the initial x-coordinate fixed, to slightly adjust the other initial conditions to ensure periodicity for our implementation. Then, the full family of Lyapunov orbits around the sub-L_1 point can be obtained with a continuation scheme [8]: one of the parameters, in our case, the initial value for the x-coordinate, is slightly increased (or decreased) by a small value $\varepsilon = 10^{-5}$, and via differential correction, the remaining initial coordinates are modified accordingly [8]. After several iterations, it is possible to obtain a family of orbits of smaller or larger amplitude.

Similarly, we can conduct a continuation in the lightness number by starting from the initial condition of any orbit within a family of orbits with constant lightness number and increase/decrease the lightness number by a small value $\varepsilon = 2 \cdot 10^{-5}$ and apply the differential corrector. It is important to highlight that the value for ε has to be small enough to allow convergence of the differential correction algorithm but not too small to require too many iterations. Figure 4 shows the results of these procedures. Figure 4 (left) shows six Lyapunov orbits for different lightness numbers between $\beta = 0.03$ and $\beta = 0.04$. The orbits shift towards the Sun as the lightness number increases. Instead, Figure 4 (right) shows a family of Lyapunov orbits around the sub-L_1 point that has been generated starting from the initial orbit with $\beta = 0.04$ from Figure 4 (left).

C. Invariant Manifolds

The classical L_1 point and the AEPs in the neighborhood of the sub-L_1 point are unstable [8, 20]. Similarly, AEPs and numerous periodic orbits in the CRTBP, as well as in the PCRTBP with a solar sail, are also unstable [8, 25]. A particle at rest at the equilibrium point or traveling along the unstable orbit that experiences a slight perturbation in the unstable direction will exponentially divert away from the equilibrium point or orbit. Likewise, a particle with the appropriate initial conditions will exponentially approach the equilibrium point or periodic orbit along the stable direction.

The manifolds of a periodic orbit can be constructed by propagating the dynamics defined in Eq. (2) from a state-vector along the stable or unstable direction. These directions are defined by the stable and unstable eigenvectors of

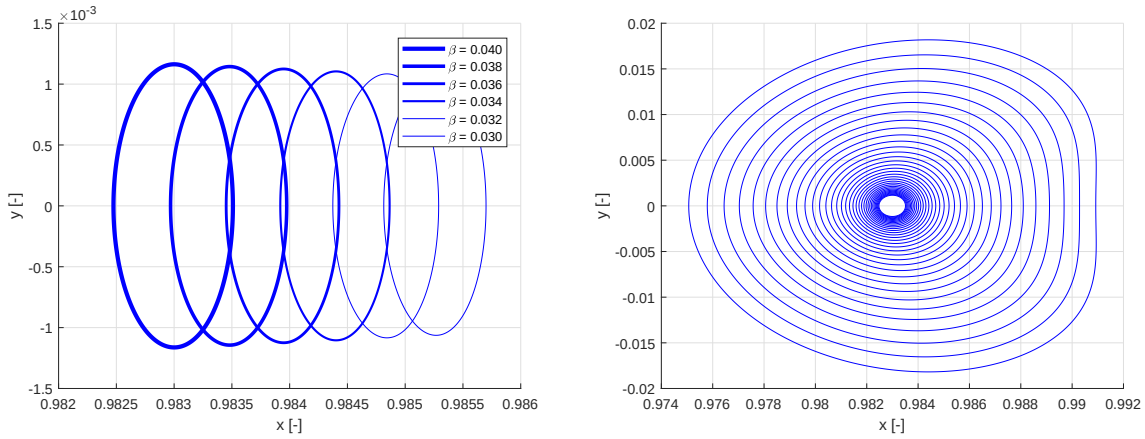


Fig. 4 Set of Lyapunov orbits with same initial velocity but different value for the lightness number (left). Family of Lyapunov orbits around the sub-L_1 point for $\beta = 0.04$ (right).

the monodromy matrix, which is the state transition matrix evaluated after one period of the orbit. The state transition matrix, $\Phi(t, t_0)$, approximates how a slight deviation in the state variables propagates along the trajectory, $\vec{r}_0$. It is propagated together with the equations of motion with the initial condition $\Phi(t_0, t_0) = I$ as [8]:

$$\dot{\Phi}(t, t_0) = \vec{A}\Phi(t, t_0) \tag{8}$$

where $\vec{A}$ is:

$$\vec{A} = \begin{bmatrix} 0 & \mathrm{I} \\ \left(\dfrac{\partial \nabla U}{\partial \vec{r}} + \dfrac{\partial \vec{a}_s}{\partial \vec{r}}\right)\Big|_{\vec{r}_0} & 2\Omega \end{bmatrix} \quad ; \quad \Omega = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{9}$$

In the PCRTBP, the monodromy matrix has four eigenvalues: $(\lambda, 1/\lambda, 1, 1)$. The stable eigenvector at the initial point of the orbit, $\vec{y^s}(\vec{x}_0)$, is the eigenvector associated to the eigenvalue smaller than one, while the unstable eigenvector at the initial point of the orbit, $\vec{y^u}(\vec{x}_0)$, is the eigenvector associated to the eigenvalue larger than one. Then, the state transition matrix can be used to propagate the stable and unstable eigenvectors to any point along the orbit as [8]:

$$\vec{y^s}(\vec{x}(t)) = \Phi(t, t_0)\, \vec{y^s}(\vec{x}_0) \quad ; \quad \vec{y^u}(\vec{x}(t)) = \Phi(t, t_0)\, \vec{y^u}(\vec{x}_0) \tag{10}$$

where $\vec{x}(t)$ is the state vector along the orbit as a function of time. Finally, the initial conditions of the trajectories along the manifolds can be described by [8]:

$$\vec{x^s}(\vec{x}) = \vec{x} \pm \varepsilon \vec{y^s}(\vec{x}) \quad ; \quad \vec{x^u}(\vec{x}) = \vec{x} \pm \varepsilon \vec{y^u}(\vec{x}) \tag{11}$$

where $\varepsilon = 10^{-6}$ is a constant that provides the magnitude of the perturbation. Larger perturbations require less time to

divert away from the periodic orbit but reduce the fidelity of the manifolds [8, 9]. Then, considering both the negative and positive signs from Eq. (11), two branches of the unstable and stable manifolds can be obtained. By integrating the initial conditions of the unstable manifold, $\vec{x^u}(\vec{x})$, forward in time, we generate trajectories shadowing the two branches of the unstable manifold $W^{u\pm}$. For the stable manifolds, $W^{s\pm}$, the same procedure can be applied, but the integration has to be performed backwards in time. In a similar manner, the invariant manifolds of any equilibrium point can be computed; only then, the monodromy matrix is substituted by the Jacobian evaluated at the equilibrium point ($\vec{A}$), i.e., $\vec{r}_0 = \vec{x}_{AEP}$, as defined in Eq. (9). Then, with Eq. (11), but using the state vector at the AEP ($\vec{x} = \vec{x}_{AEP}$), we obtain the initial conditions of the manifolds, which provide the unstable and stable branches of the manifolds when propagated forward or backward in time.

A change in the attitude of the solar sail will generate a different manifold-like structure, where the expression "manifold-like structure" is used because the term "invariant manifold" loses its meaning when the attitude is changed away from that used to maintain the orbit or the equilibrium point. However, to simplify the reading of the paper, we will refer to these manifold-like structures simply as manifolds. Figure 5 (left) shows the invariant manifolds of one of the Lyapunov orbits shown in Figure 4 (right), where the background is set as the shape of a CME as it propagates towards the Earth. The invariant manifolds for the same cone angle as in the Lyapunov orbit, $\alpha = 0$, appear in red (unstable) and blue (stable). In magenta (unstable) and cyan (stable), we also include the effect of changing the sail attitude in the propagation of the manifold trajectories. Note that, in order for the unstable and stable manifold trajectories to be symmetric with respect to the Sun-Earth line, the cone angle needs to be of opposite sign in the unstable and stable manifolds. Figure 5 (right) shows the invariant manifolds of the family of AEPs associated to the sub-L_1 point for $\beta = 0.04$. In red (unstable) and blue (stable), we show the unstable and stable manifolds for $\alpha = \pm 70°$, which bring the sail closer to the Sun. In magenta and cyan, we show the unstable and stable manifolds for $\alpha = \mp 35°$, which take the sail farther away from the Sun.

The intersection of the stable and unstable manifolds of a periodic orbit around an equilibrium point generates a family of homoclinic trajectories that connect the periodic orbit with itself, i.e., a particle starts at the periodic orbit, follows the unstable manifold until the intersection with the stable manifold, to then follow the stable manifold to arrive again to the periodic orbit. The intersection of the unstable manifold of a periodic orbit with the stable manifold of a different periodic orbit generates a family of heteroclinic trajectories that connects both periodic orbits [9]. Although the velocity components are missing in Figure 5, the direction of the velocity can be inferred as tangent to the trajectory. Therefore, if an unstable and a stable manifold trajectory intersect tangentially there is a great likelihood of the existence of a homo- or heteroclinic connection. In Figure 5 (left) we can see near-tangential intersections between the magenta and cyan manifolds at the locations $[x \approx 1.02, y \approx \pm 0.22]$. In Figure 5 (right) we can detect potential connections between the red and blue manifolds when they cross the x-axis at the point farthest from the family of AEPs. Similarly, the best connection (though less tangential than for the previously mentioned cases) between the cyan and magenta

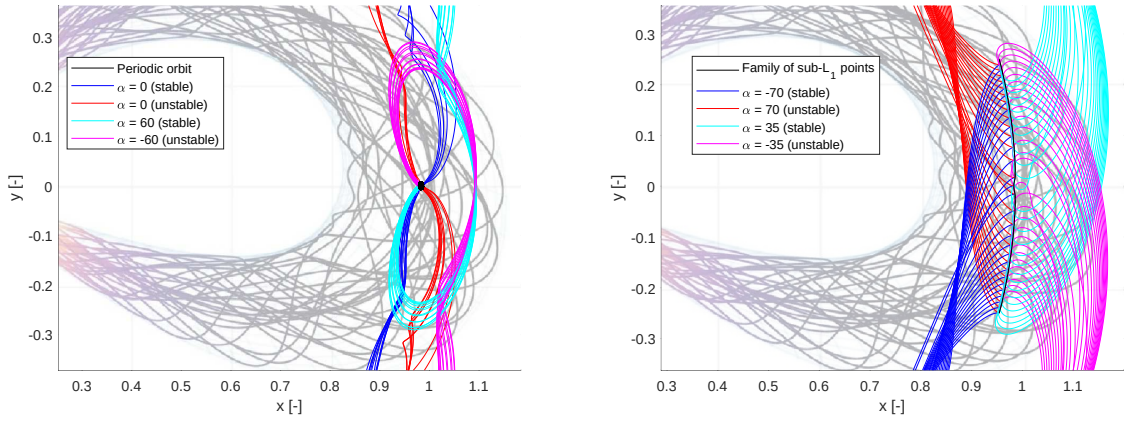


Fig. 5 Invariant manifolds of a Lyapunov orbit (left) and the sub-L_1 point (right) for $\beta = 0.04$. Red and magenta lines represent the unstable manifold branches while blue and cyan lines represent the stable ones.

manifolds in Figure 5 (right) can be found along the x-axis away from the Sun.

IV. Genetic algorithm approach

From Figure 5, we inferred the possibility of homo- and heteroclinic connections between manifolds in the position domain. In this section, we perform an analysis to search for connections both in the position and velocity domain to confirm that there is indeed a match between unstable and stable manifold trajectories. In some cases, it is possible to reduce the dimension of the search space. For example, if one wants to find connections for a certain value of one of the coordinates, say along the x-axis, then one of the dimensions is already constrained as $y = 0$. Furthermore, if both the unstable and stable manifold trajectories are constructed for a zero cone angle, then the energy level will remain constant along the trajectory and one further dimension can be constrained through the energy conservation equation. In the problem considered in this paper, we do not want to keep the sail angle equal to zero, as this limits the search space too much, but we can define a surface of intersection with an auxiliary line as explained later in this section.

As a first approach to demonstrate the existence of connections in the four-dimensional space defined by position and velocity, we performed a grid search in which we compared the state vectors of the trajectories along the stable branch with those of the unstable branch for different integration times along both the stable and unstable manifolds. Then, the integration time that allows the smallest difference between a state vector from a trajectory from each manifold is defined as the intersection point. The grid search proved to be inefficient and slow, but confirmed the existence of connections and provided some insight into the optimal values for the variables of the optimization, such as the location of the intersection or the range of cone angles that will guarantee the existence of connections. However, the procedure to compute the invariant manifolds with sufficient resolution (amount of starting points along the Lyapunov orbits and the family of AEPs from which the manifolds emanate), integration times (number of points along the trajectories within

the manifold), and control (different cone angles) is very costly for a grid search. Therefore, a Genetic Algorithm (GA), implemented in the MATLAB® function *ga.m**, has been used to optimize the problem.

A. Methodology

The objective of the optimization is to generate a periodic trajectory that takes the sail upstream of the CME, increasing the warning time with respect to a sail at the sub-L_1 point. As previously mentioned, we consider two types of trajectories: a trajectory composed of two heteroclinic connections between two AEPs in the sub-L_1 region and a homoclinic connection of a Lyapunov orbit around the sub-L_1 point.

1. Heteroclinic trajectories between AEPs

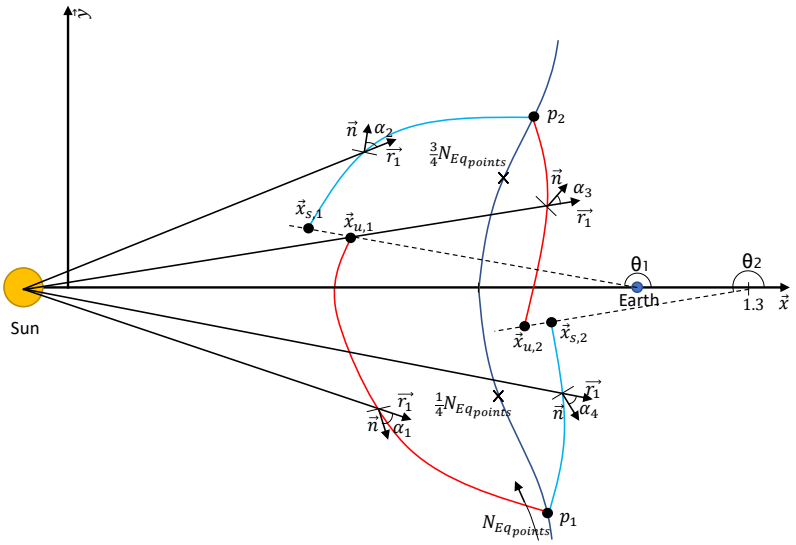


Fig. 6 Schematic of the heteroclinic trajectory showing the variables of the problem.

The trajectory to-be optimized, presented in Figure 6, starts at a certain AEP, p_1, within the family of AEPs located below the x-axis (negative y). The trajectory evolves along the unstable manifold associated to an initial cone angle, α_1, until the intersection with an auxiliary line defined by an angle θ_1 as:

$$y = \tan\theta_1 (x - 1) \tag{12}$$

The trajectory then follows the stable manifold associated to a different cone angle, α_2, which has been integrated

*ga.m - Find minimum of function using genetic algorithm. URL `https://www.mathworks.com/help/gads/ga.html` last accessed on 06/05/2020 using MATLAB® R2019b

backward in time starting from another AEP, p_2, in the family of AEPs located above the x-axis (positive y) until the intersection with that same line.

To return to the initial AEP, p_1, and construct a periodic trajectory that can be repeated over time, the second half of the trajectory to be optimized starts at p_2 and follows a trajectory along the unstable manifold associated to a cone angle, α_3, until the intersection with a second auxiliary line given by an angle θ_2:

$$y = \tan\theta_2(x - 1.3) \tag{13}$$

Then, for the solar sail to return to p_1, it follows a final trajectory segment along the stable manifold emanating from p_1 associated to the final cone angle, α_4, which has been integrated backwards in time until the intersection with that second auxiliary line. By defining the problem in this way, it can be represented by the following decision vector:

$$\begin{pmatrix} 1 \\ \frac{3}{4}N_{eq_{points}} \\ 0° \\ -30° \\ 135° \\ -30° \\ 0° \\ 135° \end{pmatrix} \leq \begin{pmatrix} p_1 \\ p_2 \\ \alpha_1 \\ \alpha_2 \\ \theta_1 \\ \alpha_3 \\ \alpha_4 \\ \theta_2 \end{pmatrix} \leq \begin{pmatrix} \frac{1}{4}N_{eq_{points}} \\ N_{eq_{points}} \\ 30° \\ 0° \\ 225° \\ 0° \\ 30° \\ 225° \end{pmatrix} \tag{14}$$

where $N_{eq_{points}}$ refers to the AEP number within the family of AEPs, which is defined by discretizing the continuous line of AEPs into 2850 points equally spaced along the y-axis between $y = 0.24$ and $y = -0.24$. Furthermore, due to the symmetry of the problem, and from the shape of the manifolds shown in Figure 5 (right) we have limited the search-space of this index to the upper and lower quarters of the family. The use of an integer (p_1, p_2) to select the starting and targeted locations allows a discrete representation of the family of AEPs without additional computational cost to obtain a state vector that satisfies the equilibrium conditions at each step of the GA.

The initial condition to integrate the invariant manifolds after a perturbation $\pm\varepsilon$ along the stable/unstable direction was defined in Eq. (11). This holds for the nominal manifolds, i.e., those in which the sail keeps the same cone angle as required to maintain the equilibrium. However, for the manifold-like structures of the family of AEPs, a change of angle provides a perturbation several orders of magnitude larger than that produced by ε. Therefore, instead of propagating the state vector at the AEP plus the perturbation given by Eq. (11), we simply integrate the state vector at the AEP with

a different attitude forwards and backward in time. Then, the only manifolds that travel towards the Earth are unstable manifolds associated to positive cone angles and stable manifolds associated with negative cone angles. Therefore, the cone angle theoretical limits of ±90 degrees, as well as the more realistic limit of ±80 degrees as defined in Section III, are too broad. Due to the mirrored symmetry of the problem, the stable manifolds directed towards the Sun can only be integrated for positive cone angles, while the unstable manifolds only emanate using negative cone angles. Finally, after trial and error, we have tightened the boundaries of the cone angle to reduce the computation time. All these considerations lead to the bounds as specified in Eq. (14).

The cost function for this problem, F_{hetero}, includes two main terms: $F_{forward}$ for the forwards (first) heteroclinic connection and $F_{backward}$ for the backwards (second) heteroclinic connection:

$$F_{hetero} = F_{forward} + F_{backward} = \Delta_{tot_1} + F_{penalty_1} + \Delta_{tot_2} + F_{penalty_2} \tag{15}$$

where Δ_{tot_1} and Δ_{tot_2} represent the quality of each heteroclinic connection, i.e., the error in position plus the error in velocity at the intersection, and are equally defined as Δ_{tot_i}:

$$\begin{gathered} \Delta_{tot_i} = \Delta_{pos_i} + \Delta_{vel_i} \\ \Delta_{pos_i} = \sqrt{(x_{u,i} - x_{s,i})^2 + (y_{u,i} - y_{s,i})^2} \quad ; \quad \Delta_{vel_i} = \sqrt{(v_{x_{u,i}} - v_{x_{s,i}})^2 + (v_{y_{u,i}} - v_{y_{s,i}})^2} \end{gathered} \tag{16}$$

where the sub-indexes refer to the unstable ($_{u,i}$) and stable ($_{s,i}$) branches of the manifold for each heteroclinic connection ($i = 1$ and $i = 2$). $F_{penalty_1}$ and $F_{penalty_2}$ are penalty functions to account for the location of the intersection. $F_{penalty_1}$ accounts for the location of the intersection of the first heteroclinic connection, giving a larger penalty to the solutions that do not travel upstream of the CME trajectory. It is computed as the distance between the location along the trajectory with minimum value for x, x_{min}, and a location upstream of the CME along the x-axis chosen by trial and error after several runs of the GA:

$$F_{penalty_1} = \lambda(x_{min} - 0.85)^2 \tag{17}$$

$F_{penalty_2}$ is used to limit the distance from the Earth along the x-axis traveled during the second heteroclinic connection:

$$F_{penalty_2} = \lambda(x_{max} - (1 - \mu))^2 \tag{18}$$

where x_{max} is the maximum value for x along the trajectory. Finally, λ is a multiplier that is used to determine the weight of the penalty functions within the cost function at the initial generations of the GA. As the GA progresses and the solutions are refined, a reduction in the weight allows the GA to focus on the quality of the connection rather than on the minimum value for x. The progressive weighting has been implemented after several runs of the GA, as it tended to

show bad quality connections with great variability between different runs. With this modification, the GA increased its repeatability and the quality of the connections. The multiplier takes the following values:

$$\lambda = \begin{cases} 0.1 & \text{if} \quad \Delta_{tot_i} \geq 10^{-2} \\ 0.001 & \text{if} \quad 10^{-2} \geq \Delta_{tot_i} \geq 10^{-6} \\ 0 & \text{if} \quad \Delta_{tot_i} \leq 10^{-6} \end{cases} \tag{19}$$

The problem defined above has been simulated in MATLAB® for eight different seeds, implemented using the MATLAB® function *rng.m*[†]: *rng*(1,2,3,...,8), where a maximum number of seed values of eight was consider sufficient due to the similarity in the shape of the trajectories found. Then, the trajectory with the smallest value for F_{hetero} from Eq. (15) is selected as the optimal one. For each run, the default settings of the MATLAB® *ga.m* function have been used, except for the function tolerance, which was set to 10^{-6}, and the population size, which was set to 150.

2. Homoclinic trajectories of a Lyapunov orbit

The process to obtain homoclinic solutions of a Lyapunov orbit follows a similar methodology to that defined for the heteroclinic solution. Figure 7 shows the trajectory to-be optimized, which starts at a location, p_1, of a certain Lyapunov orbit around the sub-L_1 point. The trajectory evolves along the unstable manifold associated with a cone angle, α_1, until the intersection with an auxiliary line defined by Eq. (12). The trajectory then continues along the stable manifold associated with another cone angle, α_2, which has been integrated backward in time from another location, p_2, of the same periodic orbit until the intersection with that same line. With this definition of the problem, it can be represented by the following decision vector:

$$\begin{pmatrix} 1 \\ 1 \\ -30^\circ \\ 0^\circ \\ 0 \\ 0 \\ 250^\circ \\ 1 \end{pmatrix} \leq \begin{pmatrix} p_1 \\ p_2 \\ \alpha_1 \\ \alpha_2 \\ dir_1 \\ dir_2 \\ \theta_1 \\ orbit \end{pmatrix} \leq \begin{pmatrix} N_{orbit_{points}} \\ N_{orbit_{points}} \\ 30^\circ \\ 80^\circ \\ 1 \\ 1 \\ 270^\circ \\ size_{family} \end{pmatrix} \tag{20}$$

[†]rng.m - Control random number generator. URL `https://www.mathworks.com/help/matlab/ref/rng.html` last accessed on 08/05/2020 using MATLAB® R2019b

where each Lyapunov orbit has been discretized with a resolution of $N_{orbit_{points}} = 10000$ points; $orbit$ is an integer that refers to the index of the orbit within the family of 100 Lyapunov orbits with different amplitudes as presented in Figure 4 (right), where only one in every five orbits is shown to ease readability; and dir_1 and dir_2 are logical variables that refer to the direction in which the manifold is propagated: a zero uses $-\varepsilon$ and a one uses $+\varepsilon$ in Eq. (11) to generate the trajectory along the manifold. While the change in cone angle is a sufficient perturbation to start the integration of the manifolds for the heteroclinic strategy in Section IV.A.1, the manifolds associated to small values of the cone angle emanating from a Lyapunov orbit are affected by the sign of ε. Finally, the boundaries for the cone angles come from the defined limits for a solar sail as defined in Section III, plus a reduction in the search-space based on trial and error.

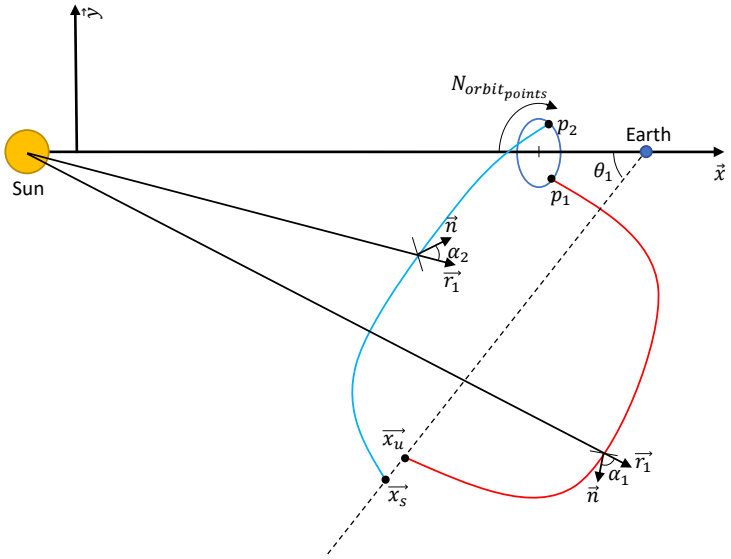


Fig. 7 Schematic of the homoclinic trajectory showing the variables of the problem.

The cost function used to guide the GA to the optimal solution is defined as:

$$F_{homo} = \Delta_{tot} + F_{penalty} \tag{21}$$

where the first term, Δ_{tot}, is the quality of the connection as defined in Eq. (16). The second term in the cost function accounts for the location and shape of the trajectory, which is composed of three terms:

$$F_{penalty} = \lambda \frac{1 - x_{min}}{|y_{min}|} (x_{min} - 0.7)^2 \tag{22}$$

where the multiplier λ follows the same three step definition as defined in Eq. (19), only using different values:

$$\lambda = \begin{cases} 5 & \text{if} \quad \Delta_{tot} \geq 10^{-2} \\ 0.05 & \text{if} \quad 10^{-2} \geq \Delta_{tot} \geq 10^{-5} \\ 0 & \text{if} \quad \Delta_{tot} \leq 10^{-5} \end{cases} \tag{23}$$

The second term represents a ratio between the maximum distance traveled along the x-axis between the sail and the Earth and the maximum displacement from the x-axis. This term is used to keep the solution close to the x-axis. Here, the location of the Earth along the x-axis is approximated by 1 instead of $1 - \mu$ due to the negligible impact of this correction. The third term refers to the distance traveled along the x-axis between the minimum value for x, x_{min}, and a point of reference upstream of the CME. This point has been chosen as $x = 0.7$ after several runs and analyses of the results of the GA. The objective of this third term is to guide the solution towards a location upstream of the CME along the x-axis without reducing the quality of the connections nor completely changing the shape of the solution. This factor is squared to increase the relative weight of this term and to avoid the possibility of flying away from the sub-L_1 point towards the L_2 region and beyond. Finally, there is a hard constraint that sets the cost function to $F = 100$ in case the trajectory does not intersect the line defined in Eq. (12).

The problem defined above has been simulated in MATLAB®, again for eight different seeds and with the same settings for the GA as for the heteroclinic solution. Like before, the trajectory with the smallest value for F_{homo} from Eq. (21) is selected as the optimal one.

B. Results

Figure 8 shows the results of the GA using a solar sail with a lightness number of $\beta = 0.04$ for the two discussed strategies: heteroclinic connections between AEPs (left) and homoclinic connections of a Lyapunov orbit (right). The background is set as the shape of a CME as it propagates towards the Earth to show the effectiveness of the solutions. The normal vector of the sail is shown as red arrows along the trajectory. Note that the change in attitude is assumed instantaneous at the location of the connections.

1. Heteroclinic connections

For the heteroclinic connections, the GA chooses the initial and targeted points very near to the maximum allowed amplitude for the family of AEPs. For this reason, we show the results of three runs of the GA for different values of the maximum amplitude to compare the effect of this parameter. This limitation in the maximum amplitude is implemented by constraining the limits on the variable $N_{eq_{points}}$: if we discard the first and last 425 AEPs in the family the amplitude is reduced to $y = \pm 0.17$ and discarding the first and last 700 points reduces the amplitude to $y = \pm 0.10$.

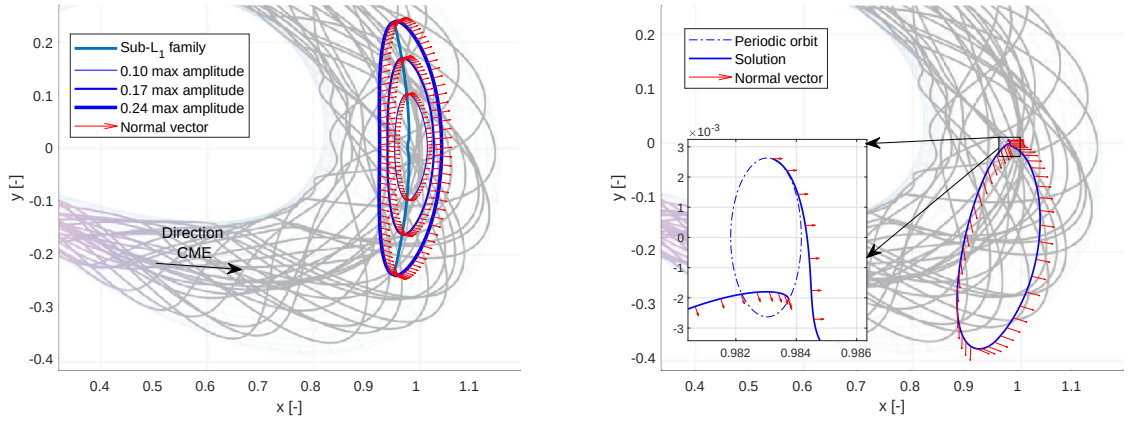


Fig. 8 GA optimized trajectory using heteroclinic connections between AEPs (left) and a homoclinic connection of a Lyapunov orbit (right) showing the normal vector of the sail along the trajectory.

Table 1 Heteroclinic connections: optimal values for the decision variables and error at the intersections between the unstable and stable manifolds

Amplitude [-]	p_1 [-]	p_2 [-]	α_1 [deg]	α_2 [deg]	θ_1 [deg]	α_3 [deg]	α_4 [deg]	θ_2 [deg]	Δ_{pos_1} [km]	Δ_{vel_1} [m/s]	Δ_{pos_2} [km]	Δ_{vel_2} [m/s]
0.10	12	1423	3.81	-3.79	179.48	-4.60	4.63	179.94	2.91	0.0017	0.32	0.0011
0.17	57	1978	5.05	-5.03	178.15	-5.39	5.42	179.62	2.62	0.012	1.53	0.0043
0.24	7	2846	6.71	-6.71	181.42	-6.84	6.84	180.43	0.54	0.007	19.16	0.25

Table 1 shows the optimal values for the decision variables for the three different amplitudes as well as the error in position and velocity for the two intersections of the stable and unstable manifolds. The three trajectories are very similar, using small values for the cone angle which gradually increases as the maximum allowed amplitude increases. Since the change in angle is very small, a realistic finite-time control law will not result in a significant change in the shape of the solution. The position error of up to 19 km represents a good performance considering that the distance traveled between the initial and targeted AEPs is in the order of 10^7 km. In a similar way, the error in velocity in the order of cm/s is sufficiently small compared to the magnitude of the velocity of the sail along the manifolds which is in the order of km/s.

Each of the three optimized trajectories shown in Figure 8 (left) is made up of four patched trajectories, two for the first half of the trajectory and another two for the second half. From the four trajectories that define the solution, the first one lies within the area of interest, the second and the third ones are situated downstream of the CME trajectory, and the fourth one lies in between the area of interest and an area that will detect CMEs after passing the Earth. Furthermore, the solution does not accurately follow the axis of the CME. In Section IV.B.3, we will analyze if these trajectories can still provide an increase in warning time.

Table 2 Homoclinic connections: optimal values for the decision variables and error at the intersection between the unstable and stable manifolds

p_1	p_2	α_1	α_2	dir_1	dir_2	θ_1	$orbit$	Δ_{pos}	Δ_{vel}
[-]	[-]	[deg]	[deg]	[-]	[-]	[deg]	[-]	[km]	[m/s]
2560	6344	-1.140	67.40	0	-1	267.77	56	3234.69	0.39

2. Homoclinic connections

For the solution using a homoclinic connection of a Lyapunov orbit, shown in Figure 8 (right), the sail remains upstream of the CME trajectory the majority of the time. A small fraction of the trajectory is located farther from the Sun than the L_1 providing no advantage in warning time, while another part of the trajectory goes beyond the expected path of the CME (see the part of the trajectory farthest from the y-axis). Furthermore, the solution does not accurately follow the axis of the CME.

Table 2 shows the optimal values for the decision variables. The control law for this trajectory is more demanding than that for the heteroclinic strategy as the instantaneous change in attitude at the connection of the unstable and stable manifold trajectories is much larger. The position error of 3234 km is still relatively small compared to the distance traveled, in the order of 10^7 km, although two orders of magnitude larger than that of the heteroclinic trajectories. The error in velocity for the homoclinic connection is again in the order of cm/s and therefore sufficiently small.

3. Warning time

The results in the previous two subsections showed that the sail is only located sunward of the L_1 point for parts of the homo- and heteroclinic connections. This section investigates if these connections can indeed increase the warning time with respect to the L_1 point. Figure 9 shows the ratio of increase in warning time for the sail with respect to a satellite at the classical L_1 point. The thick lines in magenta, green, and red show the instantaneous warning time for the heteroclinic solutions while the horizontal lines show their respective average; the blue lines provide the same information, but for the homoclinic solution, and the black solid line shows the warning time of a satellite at the sub-L_1 point. Figure 9 (left) uses the distance traveled along the x-axis to compute the warning time, while Figure 9 (right) uses the distance traveled along the axis of the CME. In both scenarios, negative ratios refer to a position of the sail that detects the CME after its impact on Earth; ratios smaller than one refer to worse warning times than those that can be achieved with current satellites at L_1; and ratios smaller than the black solid line show shorter warning time than that of a sail at the sub-L_1 point.

In Figure 9 (left), where the distance traveled along the x-axis is taken as reference to compute the increase in warning time, we see that the three solutions from the heteroclinic strategy and that of the homoclinic one provide a longer average warning time than a satellite at the L_1 point. Only the heteroclinic solution with the largest amplitude ($y = \pm 0.24$) offers a longer average warning time than the homoclinic solution. Instead, when taking the CME axis as

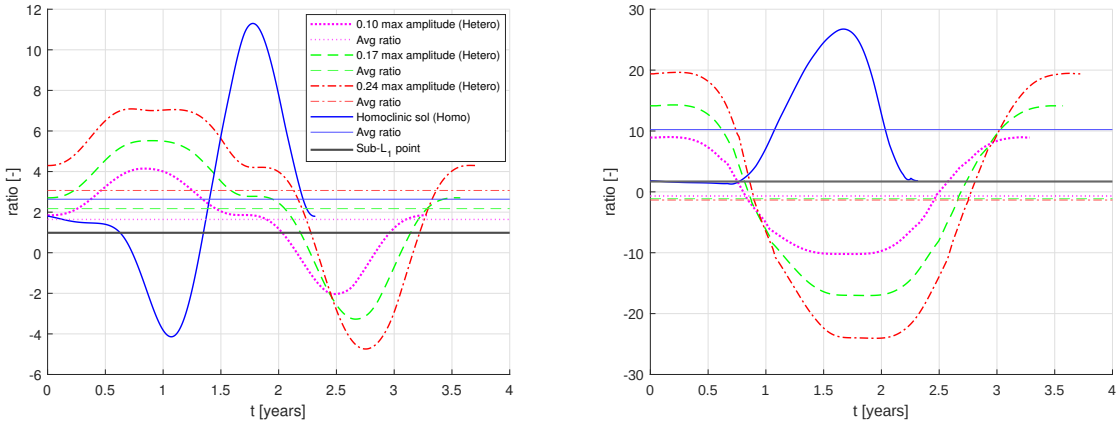


Fig. 9 Ratio of increase in warning time with respect to a satellite at the classical L_1 point using the distance travelled along the x-axis (left) or along the axis of the CME (right) to compute the increase.

reference, Figure 9 (right) shows negative values for the ratio of increase in average warning time for the three solutions of the heteroclinic strategy, with worse results for larger amplitudes. However, at the start and end of these heteroclinic connections, ratios of increase in warning of up to a factor of 20 (14 and nine for the smaller amplitude solutions) can be observed. This implies that, by using a constellation of two or three solar sails, it may be possible to greatly increase the warning time ratio to values near these maximum values. On the other hand, the solution from the homoclinic strategy shows a significant increase in the average warning time of approximately a factor of 10 by using only a single spacecraft. In addition, it guarantees a warning time never worse than that of a satellite at the sub-L_1 point. In this case, again, a constellation of two or three satellites may be used to further increase the average warning time and guarantee continuous coverage also during the fraction of the trajectory that is not located within the area of the CME as shown in Figure 8 (right).

4. Discussion of preliminary results

The results in the previous subsections show a different performance for each strategy. In order to choose the best strategy for further optimization, this section summarizes the key results for each strategy and draws conclusions on the trajectory that is selected as the initial guess for the trajectory with a continuously varying cone angle. The trajectories from the heteroclinic strategy present an error of up to 20 km in position and 0.25 m/s at the connections. In the homoclinic strategy, the errors are 3234 km for the position and 0.39 m/s for the velocity. Although the error for the homoclinic connection is larger than that of the heteroclinic connection, both strategies show sufficiently small relative errors: the distance traveled is approximately 10^7 km with velocities in the order of km/s at the transition between manifolds. Regarding the increase in warning time, the homoclinic strategy outperforms the heteroclinic strategy when the increase in warning time is computed with respect to the CME-axis as can be seen in Figure 9 (right).

The trajectories from the heteroclinic strategy need a constellation of at least two solar sails to provide a warning time longer than a satellite at L_1. Instead, in the homoclinic strategy, one sail guarantees an average warning time 10 times longer than a satellite at the sub-L_1 point, but at least two sails are needed to provide continuous detection of CMEs.

Therefore, we have chosen the trajectory from the homoclinic strategy as the best candidate to increase the warning time for CMEs. However, the warning time ratio can be increased further with the use of a continuous control law, as discussed in the following section.

V. Optimal control solver approach

In this section we discuss the methodology to further constrain and optimize the trajectory from Section IV in PSOPT and analyze its results.

A. Methodology

The limitations of the GA suggest the use of a different tool to improve the results of the study. We have chosen PSOPT, which is an open-source optimal control solver that uses direct collocation methods. PSOPT has been successfully used to generate optimal trajectories using homo- and heteroclinic connections within different CRTBPs such as the Earth-Moon [26] or the Sun-Earth [27, 28] system. To solve the optimal control problem, the time-dependent variables are approximated by Legendre polynomials at and in between the collocation points, which are spaced following Legendre-Gauss-Lobato nodes. Then, the finite non-linear programming problem can be solved with PSOPT default solver, IPOPT, which is an interior point optimizer. For the study, we have used PSOPT Release 2 version[‡], working with the Microsoft Visual 2010 compiler.

The optimal control problem to be solved can be described as follows. First, the objective function aims at maximizing the distance from the Earth at which the CMEs are detected. To this end, two different cost functions are defined:

$$F_1 = \sum_{i=1}^{N} x_i H_X H_Y \tag{24}$$

$$F_2 = \sum_{i=1}^{N} (x_i - x_{\text{sub-L}_1}) \frac{t_i - t_{i-1}}{t_f} \tag{25}$$

where x_i is the x-coordinate of the trajectory at node i, H_X and H_Y are heaviside functions that encourage the trajectory to be located in the part of the reference frame where the CME can be detected upstream of the sub-L_1 point, $x_{\text{sub-L}_1}$ is the x-coordinate of the sub-L_1 point, t_i and t_{i-1} are the times at the nodes i and $i-1$ and t_f is the final time of trajectory, so the factor $(t_i - t_{i-1})/t_f$ gives a weight to each node based on the length of that segment of the trajectory. Equation (24) is a direct minimization of the x-coordinate: the smaller the value for x, the closer to the Sun the CME is detected.

[‡]PSOPT Project Home. URL `http://www.psopt.org/` last accessed on 04/05/2020

On the other hand, Eq. (25) minimizes the distance to the sub-L_1 point, which is negative for x_i values smaller than $x_{\text{sub-L}_1}$ thus maximizing the distance to the sub-L_1 point.

The trajectory to-be optimized is represented by the following state vector:

$$\vec{x}(t) = [x \quad y \quad \dot{x} \quad \dot{y}]^T \tag{26}$$

where the initial and final states, $\vec{x}(t_0) = \vec{x}_0$ and $\vec{x}(t_f) = \vec{x}_f$, belong to the Lyapunov orbit around the sub-L_1 point, as will be explained in Section V.A.1. The boundary conditions for the state are:

$$[0.5 \quad -0.3 \quad -0.5 \quad -0.5]^T \leq \vec{x} \leq [1.02 \quad 0.05 \quad 0.5 \quad 0.5]^T \tag{27}$$

which are chosen after several trial runs of the optimization software. Initially, the boundaries were set equal to those of the GA trajectory with a 5% increase. Then, the margins were progressively reduced to guide the solution. Both cost functions, Eqs. (24,25), allow a smaller value for the upper bound of the x-coordinate. However, the ratio of increase in warning time becomes smaller and the integration errors grow larger. Furthermore, when the boundary is set below one, $x \leq 1$, convergence is not achieved.

The control of the to-be optimized trajectory is defined by the cone angle of the sail:

$$u(t) = \alpha \tag{28}$$

for which the following bounds were imposed:

$$-60^\circ \leq u \leq 60^\circ \tag{29}$$

The control variable, α, requires the physical constraint of $\pm 90^\circ$, or the more realistic bounds of $\pm 80^\circ$ as explained in Section III. However, the control profile of the sail for cone angles larger than 60 degrees showed unrealistic variations when no limit on the rate of change for the attitude was used. Furthermore, such loose limits on the control law did not allow convergence using realistic limits on the rate of change of the attitude of the sail.

There are two path constraints that the solution needs to abide by. The first one limits the rate of change of the cone angle:

$$-0.148 \text{ deg/day} \leq \dot{\alpha} \leq 0.148 \text{ deg/day} \tag{30}$$

where the limits are set by trial and error to obtain a smooth control profile. The second path constraint limits the distance of the trajectory to the axis of the CME, measured along the y-axis. The axis, defined in Eq. (1), has been

modeled as an eighth order polynomial in MATLAB® as $y_{CME}(x)$. Then, the second path constraint is defined as:

$$-0.17 \leq y_i - y_{CME}(x_i) \leq 0.15 \tag{31}$$

where the bounds have been progressively reduced up to the smallest values that still allow convergence.

The initial time is fixed at $t_0 = 0$ and the final time is set free within a broad range around the flight time of the trajectory optimized by the GA, 2.3 years as shown in Figure 9:

$$t_0 = 0 \quad ; \quad 1.9 \text{ years} \leq t_f \leq 3.5 \text{ years} \tag{32}$$

Finally, we use the solution from the GA as a first guess in PSOPT. The state vector and the cone angle of the GA solution are interpolated on a set of 50 Lagrange-Gauss-Lobatto nodes. The solution from PSOPT is also defined on 50 nodes, which could not be increased in a mesh refinement approach due to computational limitations. The convergence tolerance is set to 10^{-6} and the maximum number of iterations is 1000.

1. Initial and final conditions

The start- and endpoints of the trajectory are constrained to lie along the Lyapunov orbit chosen by the GA, see Figure 8. In MATLAB®, this orbit is represented by a discrete number of states, while PSOPT works better with continuous functions for the variables. We have therefore compared two methods to approximate the orbit with a continuous function. The first method uses a tenth order polynomial to fit three state variables, $[y(x), \dot{x}(x), \dot{y}(x)]^T$, to the remaining one, x, for the first half of the orbit as:

$$x_{orbit}(t) \approx x_{orbit_{pol}}(x) = [x \quad y(x) \quad \dot{x}(x) \quad \dot{y}(x)]^T \tag{33}$$

where $x_{orbit}(t)$ is the actual state in the orbit as a function of time and $x_{orbit_{pol}}$ is the approximation of the state using the tenth order polynomial as a function of x. Then, thanks to the symmetry with respect to the Sun-Earth line ($[x, y, \dot{x}, \dot{y}] = [x, -y, -\dot{x}, \dot{y}]$), the second half of the orbit is also parametrized. The second method uses an angular parameter $0 < \gamma < 2\pi$ to parametrize the orbit. Then, the four-state variables are approximated with an eighth order Fourier expansion as a function of γ:

$$x_{orbit}(t) \approx x_{orbit_{four}}(\gamma) = [x(\gamma) \quad y(\gamma) \quad \dot{x}(\gamma) \quad \dot{y}(\gamma)]^T \tag{34}$$

where $x_{orbit_{four}}$ is the approximated state of the orbit using the eight order Fourier series. The errors of the approximation in km (position) and m/s (velocity) for both methods are shown in Figure 10, where the error for the polynomial fitting

is only shown for the first half of the orbit. The error of the Fourier fitting is at least two orders of magnitude smaller, consequently, it is the chosen method. It is important to mention that the amplitude of the orbit is approximately $7 \cdot 10^5$ km, which means that the relative errors produced by this approximation (smaller than 10 kilometers) are reasonable. The average size of the error in dimensionless units is in the order of 10^{-8} for both the position and velocity. This value is used as the margin allowed by PSOPT to satisfy the constraints on the initial ($\vec{x} = \vec{x}_0$) and final ($\vec{x} = \vec{x}_f$) conditions as:

$$-10^{-8} \cdot [1 \quad 1 \quad 1 \quad 1]^T \leq \left(\vec{x} - \vec{x}_{orbit_{four}}(\gamma)\right) \leq 10^{-8} \cdot [1 \quad 1 \quad 1 \quad 1]^T \tag{35}$$

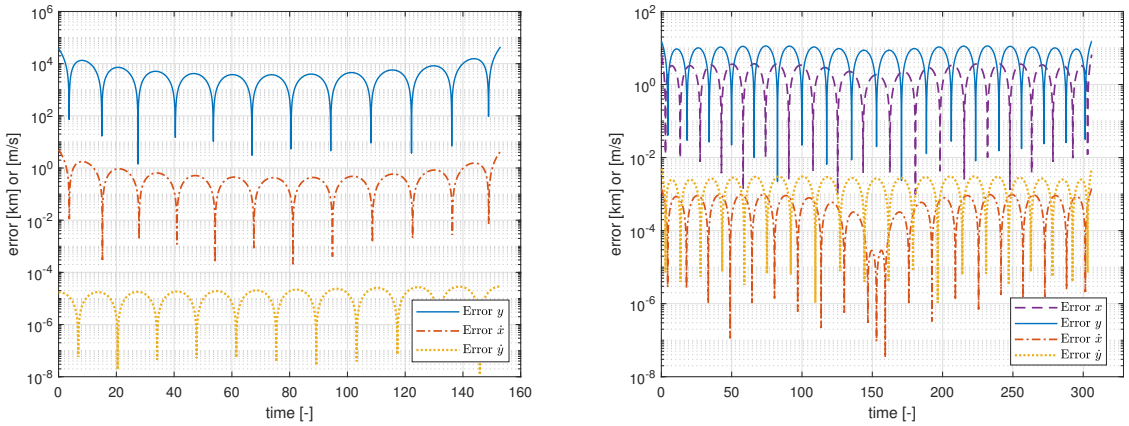


Fig. 10 Errors from approximating the Lyapunov orbit with polynomials (half orbit, left) or Fourier series (full orbit, right).

2. Re-integration in MATLAB®

The trajectories from PSOPT are represented by the value of the state, control, and time variables at the nodes. Even for a converged solution, PSOPT only guarantees that the dynamics, the boundaries for the variables, and other constraints are fulfilled at the nodes. To investigate the validity of the solution in between the nodes, these trajectories have to be re-integrated in MATLAB® using an interpolation of the control to show any discrepancies with respect to the result from PSOPT. For the interpolation of the control, we used the interpolating function $interp1.m$[§] with the interpolation method $pchip$. Other interpolation methods such as $linear$, $makima$, and $spline$ were tested, but these showed worse accuracy for the re-integration.

We consider two types of re-integration. The first re-integration takes the initial conditions from the first node of the solution and integrates the trajectory until the final time obtained in PSOPT. The difference between the state vector at the end of the re-integrated trajectory and the final state vector of PSOPT gives an indication of how accurate the

[§]interp.m - 1D data interpolation (table lookup). URL `https://www.mathworks.com/help/matlab/ref/interp1.html` last accessed on 06/04/2020 using MATLAB® R2019b

solution from PSOPT is, i.e., how well the trajectory in between the nodes abides by the constraints defined. The second method only integrates the trajectory from node to node. Then, the state error at each node is computed as an indication of the control effort required to fly the trajectory.

B. Results

The trajectory that uses F_1 as defined in Eq. (24), which minimizes the x-coordinate, will be referred to as PSOPT-1, while the trajectory using F_2 from Eq. (25), which minimizes the (negative) distance to the sub-L_1 point, will be called PSOPT-2.

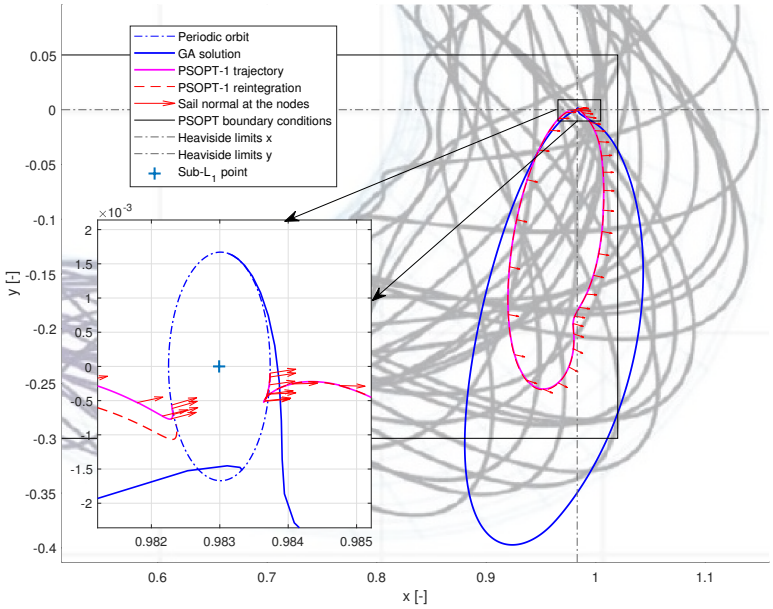


Fig. 11 PSOPT-1 trajectory plotted over the shape of the CME and compared with the GA trajectory.

Figure 11 shows the PSOPT-1 trajectory in magenta and Figure 12 the PSOPT-2 trajectory in cyan, the red arrows indicate the direction normal to the sail. The solid black lines represent the boundaries on the position coordinates as defined in Eq. (27). The two heaviside functions that guide the PSOPT-1 trajectory appear as thin dashed-dotted lines. The solution from the GA used as the initial guess is shown in blue; the red line shows the continuous re-integration of the trajectory in MATLAB®. At the bottom left of the figures, we show a detailed plot of the trajectories in close proximity of the Lyapunov orbit. There, we can analyze the departure and arrival locations along the orbit and the accuracy of the re-integration. It can be seen that the PSOPT-1 trajectory accumulates a smaller error during the re-integration. Both PSOPT-1 and PSOPT-2 trajectories remain within the path traveled by the CME for the whole trajectory, which can be compared to the GA trajectory that exits the area of interest and travels farther away from the

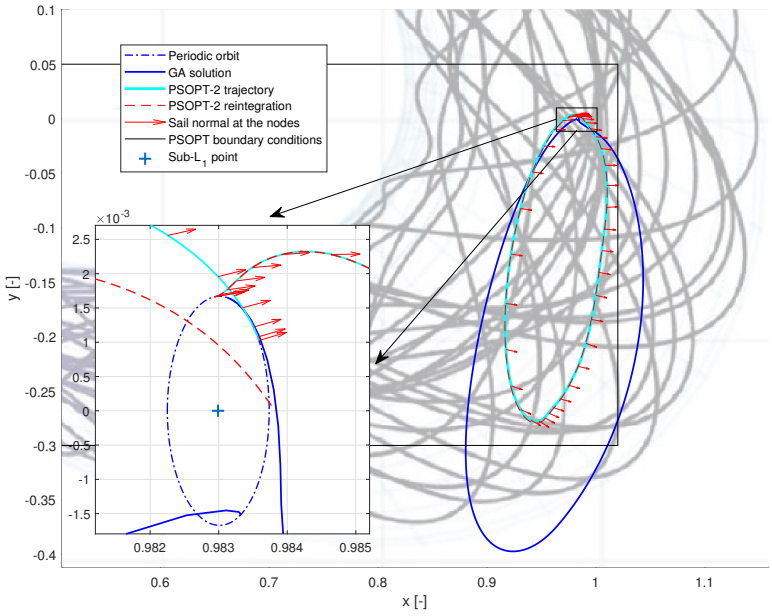


Fig. 12 PSOPT-2 trajectory plotted over the shape of the CME and compared with the GA trajectory

x-axis, providing a worse vantage point for CME monitoring.

The control law for the PSOPT-1 and PSOPT-2 trajectories is plotted in Figure 13. The attitude profile is similar for both trajectories, with maximum values of approximately +15 and -17 degrees. An analysis of the results in Figure 13 reveals a maximum rate of change for the cone angle smaller than 0.14 deg/day.

Figure 14 shows the ratio of increase in warning time for both trajectories, PSOPT-1 and PSOPT-2, with respect to a satellite at the classical L_1 point and compares it with that of the GA trajectory, following a similar approach as the one used in Figure 9. The magnitudes of the average warning time ratios are shown in the last two columns of Table 3 (the remaining columns of the table will be discussed later on in this section). In Figure 14 (left), we see that the ratio of increase in warning time by just considering the distance traveled along the x-axis is more uniform for PSOPT-1 and PSOPT-2 trajectories than for the GA trajectory. The PSOPT-1 trajectory shows a smaller average ratio than the GA trajectory, while the PSOPT-2 trajectory achieves a larger ratio. Figure 14 (right) shows the ratio of increase in warning time using the distance traveled along the CME-axis. The resulting trajectories after the optimization have a smaller peak for the ratio in warning time but a longer average warning time, guaranteeing better continuous detection of potential CMEs impacting the Earth.

In Table 3, we show the error in the re-integration of PSOPT-1 and PSOPT-2 trajectories with respect to the final node (first two columns) and the average ratio of increase in warning time for each trajectory (third and fourth columns). The re-integration error is four times smaller for the PSOPT-1 trajectory than for the PSOPT-2 trajectory. The velocity

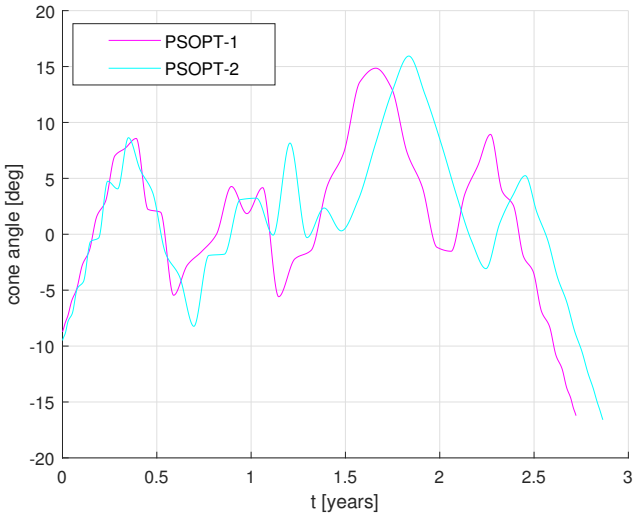


Fig. 13 Cone angle profile as a function of time for the PSOPT-1 and PSOPT-2 trajectories.

Table 3 Re-integration error and ratio of increase in warning time with respect to a satellite at L_1 point for the PSOPT-1 and PSOPT-2 trajectories.

	Position error [km]	Velocity error [m/s]	Warning time ratio x-axis [-]	Warning time ratio CME -axis [-]
PSOPT-1	44289.01	5.39	2.4	12.3
PSOPT-2	157743.56	19.19	2.85	15.1

of the sail along the Lyapunov orbit ranges between 35 and 65 m/s and its amplitude is $7 \cdot 10^5$ km, so the integration error is acceptable for the PSOPT-1 trajectory, but too large for the PSOPT-2 trajectory.

In Figure 15, we show the error of the re-integration from each node to the next one. The blue lines show the absolute error of the position in kilometers, with the yellow lines showing a relative measure of this error compared to the distance traveled. For the velocity, the red lines show the error in m/s and the purple lines, the relative error with respect to the velocity at the final point. The errors for the PSOPT-1 trajectory are given by the solid lines while the dashed lines show those errors for the PSOPT-2 trajectory. It can be seen that in this case, the errors are very similar for both solutions, and the error at arrival into the orbit is sufficiently small: in the order of tens of kilometers for the position and cm/s for the velocity. The errors along the trajectory reach some peaks of up to tens of thousands of kilometers and tens of m/s, but the distance traveled between nodes and the velocities at these nodes are in the order of millions of kilometers and several km/s, meaning very small relative errors as can be seen with the purple and yellow lines from Figure 15.

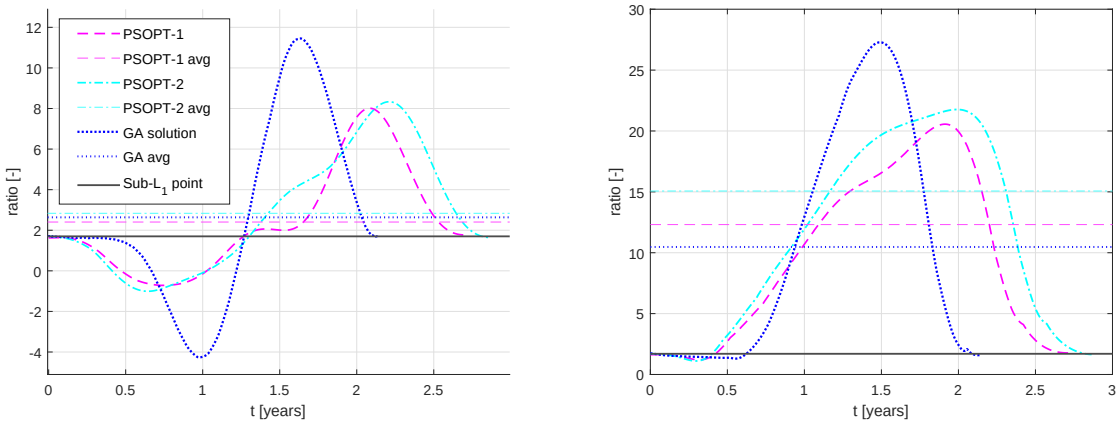


Fig. 14 Ratio of increase in warning time with respect to a satellite at the classical L_1 point using the distance travelled along the x-axis (left) and the distance travelled along the axis of the CME (right).

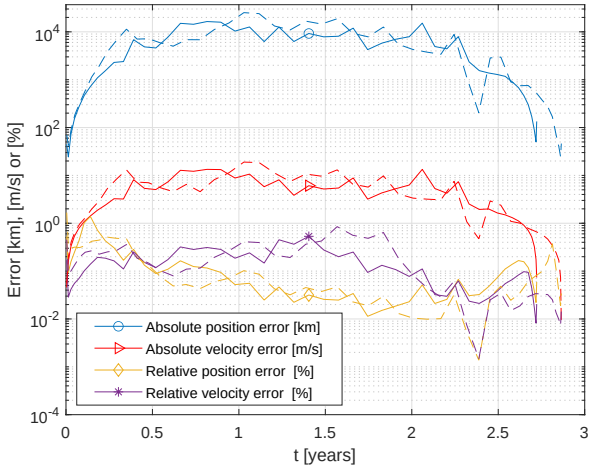


Fig. 15 Error between nodes for a node-to-node re-integration for the PSOPT-1 trajectory (solid lines) and the PSOPT-2 trajectory (dashed lines).

VI. Discussion of the results

The PSOPT-1 trajectory achieves a ratio of increase in warning time of 13 compared to a satellite at the classical L_1 point while the PSOPT-2 trajectory further increases this ratio to 15. However, the PSOPT-1 trajectory shows better fidelity to the dynamics of the system, achieving an accumulated error during the re-integration of one order of magnitude smaller. Both PSOPT trajectories provide effective warning for the full trajectory with a larger ratio than the trajectory optimized by the GA, which has a ratio of increase in warning time of 10 and showed some parts of the trajectory spent outside the path traveled by the CME.

Regarding the quality of the connection, the GA solution showed a position error of 3000 km and 0.4 m/s, see Table 2, at the connection between each half of the trajectory. The relative velocities and the vast distance traveled

translates this error into a very small relative error. In the case of the trajectories optimized by PSOPT, the trajectories are continuous until the re-insertion in the Lyapunov orbit, with an error smaller than the tolerance set, 10^{-6}. However, the trajectories from PSOPT show a re-integration error when the initial condition is propagated in MATLAB®. This re-integration error is approximately 45,000 km and 5 m/s for the PSOPT-1 trajectory and 160,000 km and 19 m/s for the PSOPT-2 trajectory. This re-integration error is relatively large for the PSOPT-2 trajectory, as the final state of the trajectory is the re-insertion into a Lyapunov orbit with an amplitude of 700,000 km and velocities ranging between 35 and 65 m/s. However, the re-integration error between nodes is similar for both trajectories, with small values between 0.001% and 1%.

The control profile of the trajectory optimized by the GA showed a large discontinuity at the intersection. Table 2 shows a required change in the cone angle of approximately 68° that needs to be performed instantaneously to guarantee the feasibility of the solution from the GA. On the contrary, the trajectories optimized by PSOPT have feasible control profiles which are shown in Figure 13. The small control effort from the solution and the relatively small re-integration errors in between nodes, see Figure 15, is expected to allow the development of some trajectory control that will correct for the small drift in between nodes. Therefore, PSOPT-2 is the best solution with the need for some sort of active trajectory control to adjust for the re-integration errors.

VII. Conclusions

This paper has shown the feasibility of using solar sails with a state-of-the-art lightness number, $\beta = 0.04$, to travel upstream of the path of Coronal Mass Ejections (CMEs) directed towards the Earth to increase the warning time for CMEs with respect to satellites at the L_1 point. For this purpose, two strategies have been evaluated and deemed feasible. The first strategy follows a series of two heteroclinic connections between two Artificial Equilibrium Points (AEPs) belonging to the family of AEPs in the sub-L_1 region, i.e., the region including AEPs sunward of the classical L_1 point. The AEPs used in this trajectory are nearly symmetrical with respect to the Sun-Earth line, which suggests the use of at least two solar sails to guarantee continuous coverage for CMEs directed towards the Earth, i.e., due to the rotation of the Sun-Earth system, the CMEs always travels towards the Earth in the same half of the ecliptic defined by the Sun-Earth line. The second strategy uses a homoclinic connection of a Lyapunov orbit around the sub-L_1 point to travel upstream of the CME. This second trajectory always remains in the half of the ecliptic in which the CME travels, but it also requires at least two solar sails for continuous coverage as parts of the trajectory are outside the path of the CME. Both strategies have been optimized with a genetic algorithm to increase the warning time for CMEs. The genetic algorithm selects the best starting AEPs along the family of AEPs or the location within the Lyapunov orbit and the piece-wise constant attitude of the sail along each manifold. The results of the optimization show better performance for the homoclinic strategy, increasing the warning time ratio compared to satellites at L_1 approximately by a factor of 10. Although the warning time is increased, the trajectory shows a discontinuity in the controls and the states at the

transition between the unstable and stable manifolds.

The trajectory following a homoclinic connection of a Lyapunov orbit around the sub-L_1 point is used as an initial guess in an optimal control problem to generate a continuous trajectory with a smaller average distance to the axis of the CME that further increases the warning time. The optimal control problem is solved for two slightly different cost functions. The first one aims to minimize the distance to the Sun along the Sun-Earth line, while the second one maximizes the distance between the sub-L_1 point and the sail also along the Sun-Earth line. The trajectory using the first cost function achieves an average ratio of increase in warning time of 12.3 with respect to satellites at L_1 with a sufficiently small re-integration error in the order of 40.000 km and 5 m/s. On the other hand, the trajectory using the second cost function achieves a larger average warning time ratio of 15.1 but develops a larger re-integration error in the order of 150.000 km and 19 m/s. Both trajectories use similar control laws where the attitude of the sail never exceeds ±20 deg with a rate of change smaller than 0.15 deg/day. The low-effort control law leaves a large margin in the sail controllability to design some sort of trajectory control to correct the re-integration error between nodes. Then, the trajectory using the second cost function can be used to provide continuous CME monitoring with a 15 times longer warning time than what is currently obtained with satellites in the environment of the L_1 point.

References

[1] Isavnin, A., "FRiED: a novel three-dimensional model of coronal mass ejections," *The Astrophysical Journal*, Vol. 833, No. 2, 2016, p. 267.

[2] Schwenn, R., "Space weather: the solar perspective," *Living Reviews in Solar Physics*, Vol. 3, No. 1, 2006, p. 2. https://doi.org/10.12942/lrsp-2006-2.

[3] "Modeling 3-D solar wind structure," *Advances in Space Research*, Vol. 32, No. 4, 2003, pp. 497 – 506. https://doi.org/https://doi.org/10.1016/S0273-1177(03)00332-6.

[4] Kechichian, J. A., Campbell, E. T., Werner, M. F., and Robinson, E. Y., "Solar surveillance zone population strategies with picosatellites using halo and distant retrograde orbits," *Acta Astronautica*, Vol. 56, No. 5, 2005, pp. 495–506.

[5] Eastwood, J. P., Hapgood, M. A., Biffis, E., Benedetti, D., Bisi, M. M., Green, L., Bentley, R. D., and Burnett, C., "Quantifying the economic value of space weather forecasting for power grids: an exploratory study," *Space Weather*, Vol. 16, No. 12, 2018, pp. 2052–2067. https://doi.org/10.1029/2018SW002003.

[6] NASA, "Solar storms and space weather - FAQ," *https://www.nasa.gov/mission_pages/sunearth/spaceweather/index.html#q5*, Accesed 30/06/2019.

[7] Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O., Plunkett, S., Rich, N., and Howard, R., "A catalog of white light coronal mass ejections observed by the SOHO spacecraft," *Journal of Geophysical Research: Space Physics*, Vol. 109, No. A7, 2004.

[8] Parker, J. S., and Anderson, R. L., *Low-energy lunar trajectory design*, John Wiley & Sons, 2014.

[9] Koon, W. S., Lo, M. W., Marsden, J. E., and Ross, S. D., *Dynamical systems, the three-body problem and space mission design*, Marsden Books, ISBN 978-0-615-24095-4., 2008.

[10] McInnes, C. R., *Solar sailing: technology, dynamics and mission applications*, Springer Science & Business Media, 2013.

[11] Tsuda, Y., Mori, O., Funase, R., Sawada, H., Yamamoto, T., Saiki, T., Endo, T., and Kawaguchi, J., "Flight status of IKAROS deep space solar sail demonstrator," *Acta Astronautica*, Vol. 69, No. 9-10, 2011, pp. 833–840.

[12] Heaton, A. F., Faller, B. F., and Katan, C. K., "NanoSail: D orbital and attitude dynamics," *Advances in Solar Sailing*, Springer, 2014, pp. 95–113.

[13] Biddy, C., and Svitek, T., "LightSail-1 solar sail design and qualification," *Proceedings of the 41st Aerospace Mechanisms Symposium*, Jet Propulsion Lab., National Aeronautics and Space Administration Pasadena, CA, 2012, pp. 451–463.

[14] McNutt, L., Johnson, L., Kahn, P., Castillo-Rogez, J., and Frick, A., "Near-earth asteroid (NEA) scout," AIAA Space 2014 Conference and Exposition, 2014, p. 4435.

[15] Mori, O., Matsumoto, J., Chujo, T., Matsushita, M., Kato, H., Saiki, T., Tsuda, Y., Kawaguchi, J., Terui, F., Mimasu, Y., et al., "Solar power sail mission of OKEANOS," *Astrodynamics*, 2019, pp. 1–16.

[16] Yen, C.-w. L., "Solar sail geostorm warning mission design," *Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2004.*, 2004.

[17] Heiligers, J., Diedrich, B., Derbes, W., and McInnes, C., "Sunjammer: Preliminary end-to-end mission design," *AIAA/AAS Astrodynamics Specialist Conference*, 2014, p. 4127.

[18] Heiligers, J., and McInnes, C., "Novel solar sail mission concepts for space weather forecasting," *24th AAS/AIAA Space Flight Mechanics Meeting 2014*, 2014, pp. AAS–14.

[19] Barnhart, D. J., Vladimirova, T., and Sweeting, M. N., "Very-small-satellite design for distributed space missions," *Journal of Spacecraft and Rockets*, Vol. 44, No. 6, 2007, pp. 1294–1306.

[20] Farrés, A., and Jorba, A., "Solar sailing with invariant manifolds in the Earth-Sun system," *Proceedings of the 66the International Astronautical Congress, Jerusalem, Israel*, 2015.

[21] Farrés, A., and Jorba, À., "Solar sail surfing along families of equilibrium points," *Acta Astronautica*, Vol. 63, No. 1-4, 2008, pp. 249–257.

[22] Farrés, A., and Jorba, A., "Dynamical System Tools to Navigate in the Earth-Sun System," *Proceedings of the xxiv congress on differential equations and applications, Cádiz, Spain*, June 8-12, 2015.

[23] Gong, S., and Macdonald, M., "Review on solar sail technology," *Astrodynamics*, Vol. 3, No. 2, 2019, pp. 93–125.

[24] Spencer, H., and Carroll, K. A., "Real solar sails are not ideal, and yes it matters," *Advances in Solar Sailing*, Springer, 2014, pp. 921–940.

[25] McInnes, A. I., "Strategies for solar sail mission design in the circular restricted three-body problem," *MSE Thesis, School of Aeronautics and Astronautics, Purdue University*, 2000.

[26] Heiligers, J., "Homo- and Heteroclinic Connections in the Planar Solar-Sail Earth-Moon Three-Body Problem," *Frontiers in Applied Mathematics and Statistics*, Vol. 4, 2018, p. 42. https://doi.org/10.3389/fams.2018.00042.

[27] Farrés, A., Heiligers, J., and Miguel, N., "Road Map to L4/L5 with a solar sail," *Aerospace Science and Technology*, Vol. 95, 2019, p. 105458.

[28] Mingotti, G., Heiligers, J., and McInnes, C., "Optimal solar sail interplanetary heteroclinic transfers for novel space applications," *AIAA/AAS Astrodynamics Specialist Conference*, 2014, p. 4223.

3

Conclusions and Recommendations

This thesis aimed at increasing the warning time for incoming Coronal Mass Ejections (CMEs). Warnings for Earth-approaching CMEs are currently provided by in-situ observations with satellites at the Sun-Earth L_1 point. As the warning time is proportional to the distance to the Earth at which the CME is detected, this thesis used solar-sail propulsion to move away from the neighborhood of the L_1 point and travel upstream of the CMEs to detect CMEs sooner and thereby increase the warning time.

Several solar-sail trajectories that periodically travel upstream of the CME while remaining as close as possible to the axis of the CME were designed in this thesis. The trajectories are developed for a solar sail with a lightness number of $\beta = 0.04$ in the planar circular restricted three-body problem using two main strategies. The first strategy followed a series of two heteroclinic connections between two Artificial Equilibrium Points (AEPs) belonging to the family of AEPs in the sub-L_1 region, i.e., the region including AEPs sunward of the classical L_1 point. The second strategy used a homoclinic connection of a Lyapunov orbit around the sub-L_1 point to travel upstream of the CME. Both strategies have been optimized with a Genetic Algorithm (GA) to select the best starting AEP along the family of AEPs or the location within the Lyapunov orbit as well as the piece-wise constant attitude of the sail along the unstable and stable manifold to increase the warning time for CMEs. The results of the optimization showed better performance for the homoclinic strategy, increasing the warning time ratio compared to satellites at L_1 by approximately a factor of 10. Although the warning time is increased, the trajectory shows a discontinuity in the controls and the states at the transition between manifolds.

The trajectory optimized with the GA following a homoclinic connection of a Lyapunov orbit around the sub-L_1 point was used as an initial guess in an optimal control problem to generate a continuous trajectory with a smaller average distance to the axis of the CME that further increases the warning time. The optimal control problem is solved for two slightly different cost functions. The first one aims at minimizing the sail distance to the Sun along the Sun-Earth line, while the second one maximizes the distance between the sub-L_1 point and the sail also along the Sun-Earth line. Both trajectories use similar control laws where the attitude of the sail never exceeds $\pm$ 17 deg, with a rate of change smaller than 0.14 deg/day. The trajectory using the first cost function achieves an average ratio of increase in warning time of 12.3 with respect to satellites at L_1 while the trajectory using the second cost function further improves the warning time to a ratio of 15.1.

In Section 3.1, the research questions of Section 1.4 are directly answered in line with the results presented throughout the paper in Chapter 2. Then, in Section 3.2 the limitations of the research are explained to guide further investigations that could potentially improve the results.

3.1. Conclusions

The answers to the research questions formulated in Section 1.4 are presented based on the results obtained throughout this thesis work.

A. *Is it possible to use heteroclinic connections between different solar-sail artificial equilibrium points in the sub-L_1 region to travel upstream of the CME using a piecewise-constant sail attitude?*
The short answer to this question is “yes”. This thesis presents three similar trajectories that use two heteroclinic connections between two different AEPs of the family of AEPs in the sub-L_1 region

that, for some fractions of the trajectory, take the sail upstream of the CME along its path towards the Earth. The trajectories have been optimized to increase the average warning time for CMEs and to reduce the discontinuity in the states at the transition between the unstable and stable manifolds of each heteroclinic connection using a Genetic Algorithm (GA) approach implemented in MATLAB.

The GA uses as decision variables the initial and final AEPs within the family of AEPs in the sub-L_1 region, the piece-wise constant attitude of the sail along the unstable and stable manifolds, and two other angles that define the location for the transition between the unstable and stable manifolds. With this definition of the problem, the GA selected the AEPs with the largest separation from the Sun-Earth line along a line normal to it. Three different values for the maximum distance between the AEPs and the Sun-Earth line of 0.10, 0.17, and 0.24, dimensionless with the Sun-Earth distance, have been used to properly analyze these connections. The optimized trajectories did not improve the average warning time with respect to a satellite at L_1, as only one half of the trajectory develops upstream of the CME path towards the Earth. However, the instantaneous warning time ratio is nearly constant at the peak values (ratios up to 20, 14, and nine with respect to a satellite at the L_1 point depending on the maximum amplitude of the family of AEPs) to rapidly decay below one. These results suggest that a constellation of at least two solar sails may provide continuous coverage with better average warning times than the existing missions in the environment of the L_1 point.

Finally, regarding the feasibility of the trajectories, the discontinuities at the connections of the unstable and stable manifolds need to be addressed. The error in the states at the intersection between manifolds is sufficiently small to guarantee a smooth transition: the position error increases with the maximum amplitude of the family of AEPs allowed for the trajectory up to a value of 19 km which is sufficiently small compared to the 10^7 km traveled between the AEPs. The velocity error is in the order of cm/s while the velocity of the sail reaches velocities in the order of km/s. The discontinuity in the attitude of the sail is inherent due to the approach of using a piecewise-constant cone angle along each manifold. The optimal cone angles found by the genetic algorithm are nearly symmetrical along the Sun-Earth line with values that increase with the maximum amplitude allowed for the family of AEPs up to a maximum of 7 degrees. Since an instantaneous attitude change of nearly 14 degrees is impossible for a real sail, some sort of trajectory control will be required to smoothly transition from one manifold to the next one.

B. *Is it possible to use a homoclinic connection of a periodic orbit around a sub-L_1 point to travel upstream of the CME using a piecewise-constant sail attitude?*

The short answer is, again, “yes”. This thesis has demonstrated the existence of trajectories that follow homoclinic connections of a planar Lyapunov orbit around a sub-L_1 point that take the sail upstream of the CME while maintaining a relatively small distance to the axis of the CME for the majority of its duration. As for the heteroclinic strategy, a GA has been used to increase the average warning time attainable with this trajectory as well as to reduce the discontinuity at the transition between the unstable and stable manifolds. The decision variables are the starting and final points within the orbit, the amplitude of the Lyapunov orbit used, the piece-wise constant attitude of the sail used along the stable and unstable manifolds, and an auxiliary angle to define the location of the transition between the unstable and stable manifolds.

The optimized trajectory achieves an average warning time 10 times longer than existing satellites in the environment of the L_1 point. However, some parts of the trajectory are spent outside the path traveled by the CME. As for the trajectory using heteroclinic connections of AEPs, a constellation of at least two satellites may guarantee continuous coverage while further increasing the average warning time.

Regarding the feasibility of the trajectory, the discontinuities in the state and the controls are larger than those obtained for the heteroclinic approach discussed in the previous question. The discontinuity in the states for the position domain is two orders of magnitude larger than that of the heteroclinic connection, with an error of approximately 3000 km. This position error is still sufficiently small compared to the distance traveled by the sail in the order of 10^7 km. In the velocity domain, the error is very similar to that of the heteroclinic connection, in the order of cm/s compared to the velocity of the sail in the order of km/s. Finally, the attitude change required to use this connection is almost 70 degrees. To compensate for this unrealistic attitude change, a progressive transition of the sail angle will be needed, but this control law may affect the shape of the trajectory.

C. *Is it feasible to optimize the sail attitude from a piecewise-constant attitude to a continually variable attitude to further increase the warning time for solar storms?*
As a short answer, yes it is. The results from the two strategies using a piece-wise constant attitude for the sail demonstrate the existence of trajectories that periodically travel upstream of the path of CMEs towards the Earth. The trajectory following a homoclinic connection of a Lyapunov orbit around a sub-L_1 point showed the best results in terms of average warning time. However, the discontinuity in the states and especially in the cone angle at the transition between the unstable and stable manifold suggests the use of a continuously varying control law to improve the results. Using a pseudo-spectral optimizator called PSOPT, an optimal control problem with two slightly different definitions for the cost function has been developed to further increase the warning time. The first cost function aims at minimizing the average distance to the Sun along the Sun-Earth line, while the second cost function maximizes the distance to the sub-L_1 point also along the Sun-Earth line. Furthermore, a path constraint is used to limit the maximum distance between the trajectory and the axis of the CME.

The optimized trajectories for both cost functions showed similar control profiles, with maximum cone angles of ± 17 degrees and a maximum rate of change for the cone angle of 0.14 deg/day, much smaller than what can be attained with state-of-the-art technology. The trajectory maximizing the distance to the sub-L_1 point obtained the largest increase in warning time ratio with a factor of 15.1, while the minimization of the distance to the Sun provided a factor of 12. Finally, the solution from PSOPT was re-integrated in MATLAB using a continuous control law by interpolating the values given by PSOPT at the nodes. The trajectory obtained with the minimization of the distance to the Sun achieved the smallest re-integration error of approximately 44,000 km and 5 m/s, while the trajectory obtained with the maximization of the distance to the sub-L_1 point accumulated a re-integration error of approximately 160,000 km and 19 m/s. These errors are relatively large compared to the final state of the trajectory which lies along a Lyapunov orbit with an amplitude of 700,000 km and velocities ranging between 30 and 60 m/s. Nevertheless, the accumulated error between each node and the next ranges between 0.001% and 1% of the distance traveled or the velocity at the end of the node. This small relative error added to the relatively simple control law suggests the possibility of using some sort of trajectory control to correct for the small drift between nodes.

In conclusion, a trajectory that can be used to increase the warning time for incoming CMEs has been developed by using a solar sail to travel upstream of the axis of CMEs directed towards the Earth.

3.2. Recommendations

As in every piece of research, it is impossible to investigate everything. It is necessary to set a limit to make the study feasible with the available resources. In the case concerning this project, the main constraint is time. This limitation justifies certain decisions to simplify the problem that will be potential candidates to expand the research in the future for better results.

I. *Re-integration error*
One of the problems of the final trajectory presented in the paper is the error in the re-integration of the trajectory. This error is inevitable due to the nature of the optimization used to solve the optimal control problem. PSOPT uses direct collocation methods, which only satisfy the dynamics at the nodes and interpolates the time-dependent variables in between them. In contrast, the numerical integration implemented in the *ode45.m*[1] function within MATLAB® satisfies the dynamics along the trajectory. This discrepancy generates a certain error at the end of the trajectory that can be reduced in two different ways. The simplest possibility, in terms of human effort, is to use a more powerful computer to run the same optimization problem with a finer grid of nodes. In this way, there will be more nodes at which the dynamics are satisfied and the accumulated error along the trajectory will be smaller. The second possibility to deal with the re-integration error is the design of some sort of trajectory control to reduce the error between nodes. Although this may prove to be relatively simple, it was beyond the scope of the thesis.

II. *Extension to the three dimensional problem*

[1] ode45.m - Solve nonstiff differential equations - medium order method. URL `https://www.mathworks.com/help/matlab/ref/ode45.html` ® last accessed in 29/04/2020 using MATLAB R2019b

CMEs are three-dimensional torus-like structures with a near-circular cross-section around an axis that lies in the ecliptic plane. One of the advantages of solar sailing is that they generate the possibility of maintaining orbits that are displaced away from the ecliptic plane. In this thesis, the study is limited to the ecliptic plane. This reduction to a 2D problem limits the maneuverability of the sail to just the angle between the incident radiation and the normal to the sail, the cone angle (α). It is expected that an extension to the three-dimensional problem will allow a more optimal design of the trajectory that manages to further reduce the average distance to the CME axis by traveling outside of the ecliptic plane. However, the search for connections then becomes more complex as the search for connections between manifolds would require a match of six coordinates instead of four and the GA would include an extra variable for the second angle that defines the attitude of the sail.

III. *Non-ideal sail properties*
The study is conducted in the planar CRTBP with an ideal solar sail acceleration model. Apart from the extension to the 3D problem, several non-ideal modifications can be included in the sail model to represent its acceleration with higher fidelity. The non-ideal behavior of a sail is divided into three categories: attitude control, shape deformations, and optical imperfections, where the non-ideal-optical effects have the largest impact on the acceleration [29]. A further study should consider whether the trajectory is feasible just with the addition of some trajectory control or if the problem needs to be re-optimized to account for the change in the dynamics of the sail.

IV. *Increased fidelity for the dynamics*
This thesis has used the PCRTBP as the dynamical framework for the research, which can be expanded to the spatial CRTBP to improve the results as mentioned in a previous recommendation. Nevertheless, the dynamical system can be further expanded to a higher fidelity one using a deviation from the CRTBP such as the Elliptical Restricted Three-Body Problem (ERTBP) or even further, ephemeris coordinates for the main bodies.

The inclusion of higher fidelity dynamics, such as the eccentricity from the ERTBP, not only will affect the shape of the invariant manifolds and, thus, the shape of the developed trajectories, but it will make the system non-autonomous. The time needs to be included in the equations of motion through the true anomaly. Then, any periodic solution must have a period which is an integer multiple of the period of the perturbation, in this case, the period of the Sun-Earth orbit: one year.

V. *Different approach for heteroclinic connections strategy*
Additional runs of the GA algorithm with a different approach for the heteroclinic strategy have already shown promising results (not shown in the draft article). One of the disadvantages of the trajectories using the series of heteroclinic connections was the need to use at least two solar sails to guarantee continuous monitoring for CMEs, as half of the trajectory developed downstream of the path traveled by CMEs towards the Earth. Reducing the search-space for the AEPs to limit their maximum distance downstream of the CME was deemed feasible. Figure 3.1 shows the result of a preliminary run of the GA with an asymmetrical amplitude along the axis normal to the Sun-Earth line for the family of AEPs, where only a small fraction of the resulting trajectory lies downstream of the CME path, which is set as the background of the image.

Figure 3.1 is plotted in a synodic reference frame, centered at the barycenter of the Sun and the Earth with the x-axis along the Sun-Earth line and the y-axis normal to the x-axis within the ecliptic plane. The background is set as the path of a CME approaching the Earth to show the effectiveness of the resulting trajectory.

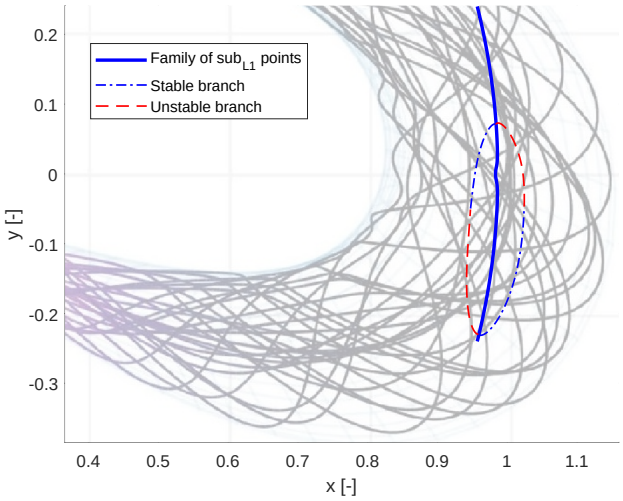


Figure 3.1: Preliminary results illustrating the direction of further research using heteroclinic connections with a modified approach.

A

Verification and Validation

This appendix includes the tests and comparisons used to gain trust in the methodology and implementation during the research. The similarities with results found in the literature, the small numerical errors, and the repeatability of the results are sufficient proof to verify and validate the procedures and the results.

A.1. Verification

A.1.1. Dynamical system model

The CRTBP has been extensively studied in the literature [25, 26], which is more than enough proof of its validity. Although the solar sail CRTBP is more recent and has less literature, it has been sufficiently studied to verify it [3, 30]. The dynamical system is modelled with Eq. 2 in Chapter 2. This second-order differential vectorial equation can be expanded to a system of four first-order differential equations as:

$$\begin{cases} \dot{x} = v_x \\ \dot{y} = v_y \\ \dot{v_x} = 2v_y - U_x + a_x \\ \dot{v_y} = 2v_x + U_y + a_y \end{cases} \tag{A.1}$$

The dynamical system can be verified by generating surfaces of equilibria. In Reference [31] the authors show the intersection of the 3D surfaces of equilibria with the ecliptic in the environment of the L_1 and L_2 points. Figure A.1 uses the surfaces of equilibria from the reference paper as the background image, then the same surfaces of equilibria computed in this thesis are shown in grey dotted lines. It is important to note the change in the x-axis with respect to Figure 3 from the paper in Chapter 2 as the authors of [31] define the axis as negative towards the secondary body, as opposed to the reference frame defined in this thesis. The accuracy of the result is enough to verify the dynamics.

A.1.2. Periodic orbits

The periodic Lyapunov orbits used as initial conditions for the homoclinic connections are presented in Figure 4 in Chapter 2. As the dynamics of the system have already been verified, demonstrating the periodicity of the orbits is enough to verify them. Figure A.2 shows the norm of the error between the first and the last point of the orbit. The size of the error is always smaller than 10^{-11} and the smallest amplitude of the orbits is 10^{-3}. Therefore, the periodicity of the orbits is considered verified.

A.1.3. Invariant manifolds

The invariant manifolds associated with the family of equilibrium points can be computed using the eigenvalues and associated eigenvectors of the linearized dynamics at the equilibrium point. These linearized dynamics are given by the Jacobian of the flow field computed as [26, 32]:

$$\vec{A} = \begin{bmatrix} 0 & \mathrm{I} \\ \left(\dfrac{\partial \nabla U}{\partial \vec{r}} + \dfrac{\partial a_s}{\partial \vec{r}} \right)\Big|_{\vec{r_0}} & 2\Omega \end{bmatrix} \quad \text{where} \quad \Omega = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{A.2}$$

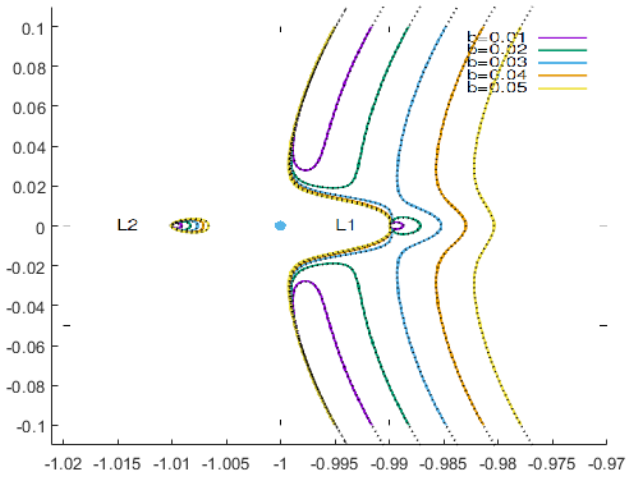


Figure A.1: Intersection of the surfaces of equilibria generated by adding a solar sail to the CRTBP and the ecliptic plane for different lightness numbers (β in this thesis, b in this plot as used by the original authors). Edited from [31] to verify the dynamics of this thesis.

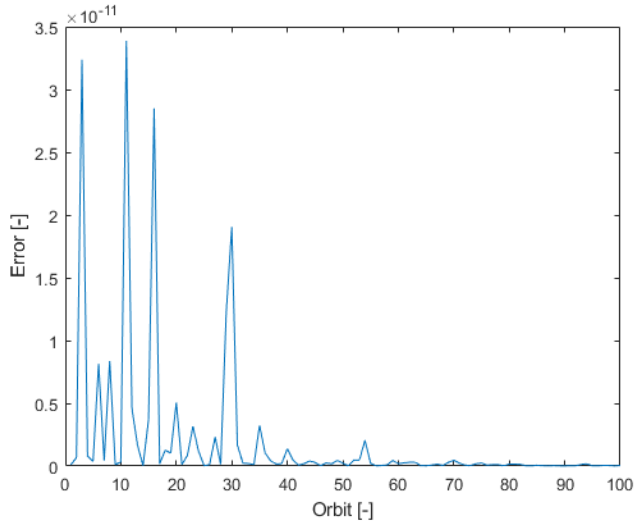


Figure A.2: Error in periodicity for the planar orbits used in the thesis, measured as the difference in the initial and final state vectors.

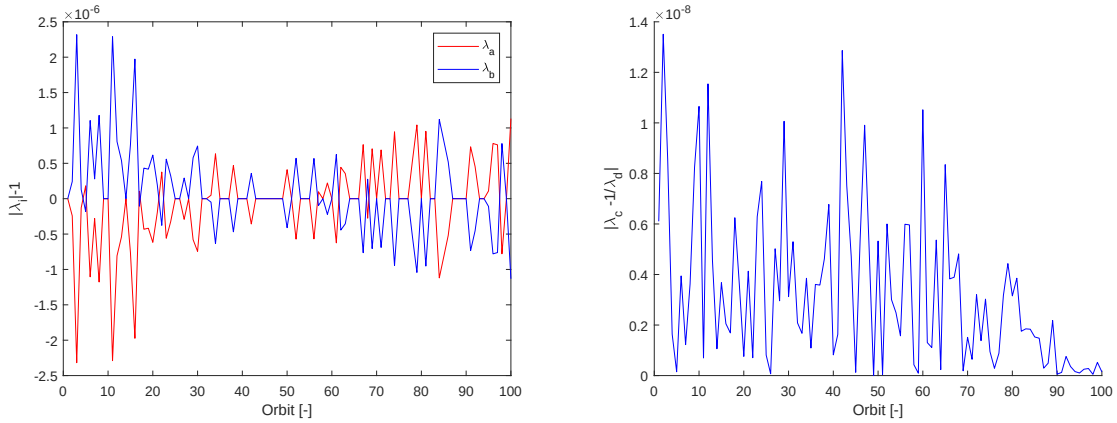


Figure A.3: Error in the eigenvalues from the Monodromy matrix. Norm of the unity eigenvalues minus one (left) and difference between a non-unity eigenvalue and the inverse of its reciprocal one (right).

Since the dynamics are already verified, it is only necessary to confirm that the derivatives of the potential, U, and the solar sail acceleration, a_s, are properly computed. This process is performed with the help of Maple® [1], computing the symbolic derivatives of both the potential and the sail acceleration, which are then written into MATLAB®. The eigenvectors and eigenvalues are computed using the MATLAB® function *eig.m*[2]. Since these programs are extensively verified, it is possible to confirm that the manifolds associated with the equilibrium points are also verified.

For the invariant manifolds of the periodic orbit, the eigenvalues and eigenvectors of the monodromy matrix are used. The monodromy matrix is the state transition matrix evaluated after one period of the orbit. The monodromy matrix of periodic orbits in the planar CRTBP has four eigenvalues: (λ_1, $1/\lambda_1$, 1, 1) [25]. Then, the family of periodic orbits around the sub-L_1 for $\beta = 0.04$ as presented in Figure 4 in Chapter 2 also needs to satisfy this condition. From the four eigenvalues of each monodromy matrix, the two that are closest to one are selected, λ_a and λ_b. Since some of them show small complex components,the difference between the complex magnitude and one is plotted ($|\lambda_i| - 1$, with $i = a, b$) in Figure A.3 (left). The two remaining eigenvalues have to be reciprocal. To confirm it, $|\lambda_c - 1/\lambda_d|$ is plotted in Figure A.3 (right) The size of the errors in the unit eigenvalues in the order of 10^{-6} and in the reciprocity of the two other eigenvalues of 10^{-8} verify this results.

A.1.4. Genetic Algorithm

The optimization with constant cone angle along the manifolds is performed using the MATLAB® function *ga.m*[3]. This function has been extensively used and is, therefore, verified. However, the optimization needs to be properly designed. To verify it, the GA is run at least eight times for the same configuration with different seeds using the function *rng.m*[4]: rng(1,2,3,...,8). Then, the results of each run are compared to see if they converge to a similar solution. Figure A.4 shows the eight runs for each of the three amplitudes in the search for heteroclinic connections (left) and the eight runs for the search of homoclinic connections (right). For the heteroclinic connections, it can be seen that the runs are very similar to each other, except for one run in between the middle and smallest amplitude. However, the shape of the solution is constant and the optimization can be verified. In the case of the homoclinic connections, again, seven of the connections are very similar to each other with one of them showing some variations. It is important to mention that the most optimal solution (the minimum value for the cost function as defined in the paper) is not the odd case. It can be seen that each solution uses differ-

[1] Maple® - interactive Computer Algebra System, version 2018. `https://www.maplesoft.com/products/maple/students/` last accessed 09/03/2020

[2] eig.m - Eigenvalues and eigenvectors. URL `https://www.mathworks.com/help/matlab/ref/eig.html` last accessed 24/04/2020 using MATLAB® R2019b

[3] ga.m - Find minimum of function using genetic algorithm. URL `https://www.mathworks.com/help/gads/ga.html` last accessed 06/05/2020 using MATLAB® R2019b

[4] rng.m - Control random number generator. URL `https://www.mathworks.com/help/matlab/ref/rng.html` last accessed 08/05/2020 using MATLAB® R2019b

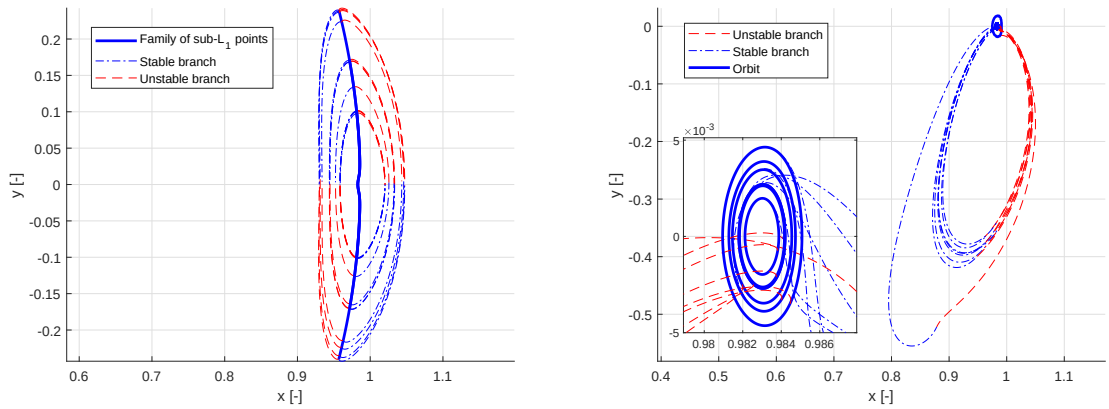


Figure A.4: Set of solution for different amplitudes for heteroclinic connections between equilibria (left) and homoclinic connections from a periodic orbit (right) for eight different seeds.

ent initial orbits, but the position within them is very similar along the solutions, which means that the code converges to the same solution but the initial orbit is not so relevant to the shape of the trajectory. Results can be considered verified.

A.1.5. Pseudo-Spectral Optimization

The optimization in MATLAB® is complemented with further optimization using PSOPT[5]. This tool is written in C++ and the dynamics have to be translated there. Although the process has been carefully performed, to verify the results, the potential function U is evaluated at the nodes. The difference between the potential evaluated at the nodes in MATLAB® and in PSOPT is smaller than the numerical errors, which guarantees an appropriate translation of the dynamics.

Then, to verify the results of PSOPT, the error at re-integration in MATLAB® is considered. This error is presented in Figure 15 and Table 3 in the paper. Although the error is not negligible, it is sufficiently small considering the different natures of the two programs and the scarce number of nodes for the optimization.

A.2. Validation

Finally, it is important to address how good is the local minimum obtained in the optimization. This procedure is, in general, a complex task. Regarding the optimization with the genetic algorithm, several cost functions have been studied and showed worse results than those presented in the paper. Although the solution cannot be guaranteed as the global minimum, it is the lowest minimum among numerous tests. Furthermore, as the genetic algorithm is used as an initial guess for further optimization with a considerably different approach, it is possible to say that a good local minimum, i.e., a trajectory that already provides a better warning time than those used a reference, is a sufficiently good result.

For the final solution obtained with PSOPT, the paper only presents the best two trajectories among numerous tested candidates. Again, this can only show that the solution is the best local minimum obtained so far, but cannot guarantee that the solution is the global minimum. The presented solution offered an average warning time up to 15 times longer than what satellites at the L_1 point can currently achieve. The target of this thesis was to increase the warning time beyond the existing mission proposals. In the literature, the are no missions that can increase the average warning by more than a factor of two using near-term sail performances as in this thesis. Therefore, the designed trajectory can be considered validated, as it fulfills the expected outcome of the thesis.

[5]PSOPT Project Home. URL `http://www.psopt.org/` last accessed 04/05/2020

Bibliography

[1] Yen, C. W. L., "Solar sail geostorm warning mission design," *Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2004.*, 2004.

[2] Heiligers, J., Diedrich, B., Derbes, W., and McInnes, C., "Sunjammer: Preliminary end-to-end mission design," *AIAA/AAS Astrodynamics Specialist Conference*, 2014, p. 4127.

[3] McInnes, C. R., *Solar sailing: technology, dynamics and mission applications*, Springer Science & Business Media, 2013.

[4] Jacobson, R. A., and Thornton, C. L., "Elements of solar sail navigation with application to a Halley's comet rendezvous," *Journal of Guidance and Control*, Vol. 1, No. 5, 1978, pp. 365–371.

[5] Friedman, L., "Interplanetary travel: A puff of wind for the solar sail?" *Nature*, Vol. 302, No. 5904, 1983, pp. 104–104.

[6] Staehle, R. L., Graham, J. M., and Champa, J., "Solar sail expedition to the moon and Mars-Mission update," *Spaceflight*, Vol. 34, 1992, pp. 256–258.

[7] Horvath, I., Carroll, K. A., and Williams, A., "The Canadian solar sail project," *In Canadian Space Agency, Proceedings of the Sixth CASI Conference on Astronautics 6 p*, Vol. 1990, 1990.

[8] Tsuda, Y., Mori, O., Funase, R., Sawada, H., Yamamoto, T., Saiki, T., Endo, T., and Kawaguchi, J., "Flight status of IKAROS deep space solar sail demonstrator," *Acta Astronautica*, Vol. 69, No. 9-10, 2011, pp. 833–840.

[9] Alhorn, D., Casas, J., Agasid, E., Adams, C., Laue, G., Kitts, C., and O'Brien, S., "Nanosail-d: The small satellite that could!" 2011.

[10] Biddy, C., and Svitek, T., "LightSail-1 solar sail design and qualification," *Proceedings of the 41st Aerospace Mechanisms Symposium*, Jet Propulsion Lab., National Aeronautics and Space Administration Pasadena, CA, 2012, pp. 451–463.

[11] McNutt, L., Johnson, L., Kahn, P., Castillo-Rogez, J., and Frick, A., "Near-earth asteroid (NEA) scout," AIAA Space 2014 Conference and Exposition, 2014, p. 4435.

[12] Mori, O., Matsumoto, J., Chujo, T., Matsushita, M., Kato, H., Saiki, T., Tsuda, Y., Kawaguchi, J., Terui, F., Mimasu, Y., et al., "Solar power sail mission of OKEANOS," *Astrodynamics*, 2019, pp. 1–16.

[13] Dachwald, B., "Solar sail dynamics and control," *Encyclopedia of Aerospace Engineering*, 2010.

[14] Society, T. P., "LightSail 2 artist concept with Earth behind," *https://www.planetary.org/multimedia/space-images/spacecraft/ls2-earth.html*, Accesed 27/04/2020.

[15] Schwenn, R., "Space weather: the solar perspective," *Living Reviews in Solar Physics*, Vol. 3, No. 1, 2006, p. 2. doi:10.12942/lrsp-2006-2.

[16] Klein, K.-L., and Dalla, S., "Acceleration and propagation of solar energetic particles," *Space Science Reviews*, Vol. 212, No. 3-4, 2017, pp. 1107–1136.

[17] Benz, A. O., "Flare observations," *Living reviews in solar physics*, Vol. 14, No. 1, 2017, p. 2.

[18] Isavnin, A., "FRiED: a novel three-dimensional model of coronal mass ejections," *The Astrophysical Journal*, Vol. 833, No. 2, 2016, p. 267.

[19] "Modeling 3-D solar wind structure," *Advances in Space Research*, Vol. 32, No. 4, 2003, pp. 497 – 506. doi:https://doi.org/10.1016/S0273-1177(03)00332-6.

[20] NASA, "Solar storms and space weather - FAQ," *https://www.nasa.gov/mission_pages/sunearth/spaceweather/index.html#q5*, Accesed 30/06/2019.

[21] Eastwood, J. P., Hapgood, M. A., Biffis, E., Benedetti, D., Bisi, M. M., Green, L., Bentley, R. D., and Burnett, C., "Quantifying the economic value of space weather forecasting for power grids: an exploratory study," *Space Weather*, Vol. 16, No. 12, 2018, pp. 2052–2067. doi:https://doi.org/10.1029/2018SW002003.

[22] Harrison, R. A., Davies, J. A., Biesecker, D., and Gibbs, M., "The application of heliospheric imaging to space weather operations: Lessons learned from published studies," *Space Weather*, Vol. 15, No. 8, 2017, pp. 985–1003.

[23] Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O., Plunkett, S., Rich, N., and Howard, R., "A catalog of white light coronal mass ejections observed by the SOHO spacecraft," *Journal of Geophysical Research: Space Physics*, Vol. 109, No. A7, 2004.

[24] Kaiser, M. L., Kucera, T., Davila, J., Cyr, O. S., Guhathakurta, M., and Christian, E., "The STEREO mission: An introduction," *Space Science Reviews*, Vol. 136, No. 1-4, 2008, pp. 5–16.

[25] Parker, J. S., and Anderson, R. L., *Low-energy lunar trajectory design*, John Wiley & Sons, 2014.

[26] Koon, W. S., Lo, M. W., Marsden, J. E., and Ross, S. D., *Dynamical systems, the three-body problem and space mission design*, Marsden Books, ISBN 978-0-615-24095-4., 2008.

[27] Heiligers, J., and McInnes, C., "Novel solar sail mission concepts for space weather forecasting," *24th AAS/AIAA Space Flight Mechanics Meeting 2014*, 2014, pp. AAS–14.

[28] Barnhart, D. J., Vladimirova, T., and Sweeting, M. N., "Very-small-satellite design for distributed space missions," *Journal of Spacecraft and Rockets*, Vol. 44, No. 6, 2007, pp. 1294–1306.

[29] Spencer, H., and Carroll, K. A., "Real solar sails are not ideal, and yes it matters," *Advances in Solar Sailing*, Springer, 2014, pp. 921–940.

[30] Macdonald, M., *Advances in solar sailing*, Springer Science & Business Media, 2014.

[31] Farrés, A., and Jorba, A., "Solar sailing with invariant manifolds in the Earth-Sun system," *Proceedings of the 66the International Astronautical Congress, Jerusalem, Israel*, 2015.

[32] McInnes, A. I., "Strategies for solar sail mission design in the circular restricted three-body problem," *MSE Thesis, School of Aeronautics and Astronautics, Purdue University*, 2000.