INVESTIGATION INTO THE MITIGATION OF THE EFFECTS OF UNCERTAIN OPTICAL DEGRADATION ON AN INTERPLANETARY SOLAR SAIL MISSION USING A SINGLE MODEL UPDATE

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Aerospace Engineering

by

Jordan Trent Smiroldo

December 2013

COMMITTEE MEMBERSHIP

TITLE:	Investigation into the Mitigation of the Effects of Uncertain Optical Degradation on an Interplanetary Solar Sail Mission Using a Single Model Update
AUTHOR:	Jordan Trent Smiroldo
DATE SUBMITTED:	December 2013
COMMITTEE CHAIR:	Dr. Kira Abercromby, Assistant Professor of Aeronautics, California Polytechnic State University
COMMITTEE MEMBER:	Dr. Jordi Puig-Suari, Professor of Aeronautics, California Polytechnic State University
COMMITTEE MEMBER:	Dr. Eric Mehiel, Associate Professor of Aeronautics, California Polytechnic State University
COMMITTEE MEMBER:	Dr. Paul Choboter, Associate Professor of Mathematics, California Polytechnic State University

ABSTRACT

Investigation into the Mitigation of the Effects of Uncertain Optical Degradation on an Interplanetary Solar Sail Mission Using a Single Model Update

Jordan Trent Smiroldo

The renewed academic interest in using solar sails as a source of spacecraft propulsion has been accompanied by a recent fervor of investigations into non-ideal and off-nominal sail performance considerations. One of the most influential considerations, uncertain optical degradation, has been shown to present significant trajectory design difficulties. This paper investigates the potential of using a mid-course degradation model update to mitigate the risk of missing the target destination in a sample 300 day Earth-Venus trajectory. Using a range of potential degradation profiles, it is shown that correcting in the first half of the mission is highly likely to result in a trajectory that arrives sufficiently close to Venus at the end of the mission timeframe. Depending on the exact extent of the uncertainty, the data suggests that the latest a correction should take place ranges from 150 to 240 days into the mission. The influence of two different parameters, the extent and rate of degradation, are compared to show that the former of the two is more impactful on correcting timing than the latter.

ACKNOWLEDGMENTS

First, I would like to thank my family, friends and colleagues who allowed me to ignore them endlessly as I worked on this investigation. Second, I would like to recognize my thesis advisor Dr. Kira Abercromby, whose infinite patience was tested as I asked for extension after extension. Lastly, I would like to thank the developers of Google Scholar for making literature reviews convenient and easy.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background and General Information

1.1.1 Solar Sails

Solar sails are thin, reflective membranes used primarily as sources of spacecraft propulsion. By intercepting and reflecting radiation from the sun, a force acts upon the sail in a direction dictated by its orientation. With a pressure of about 4.5 $\frac{\mu N}{m^2}$ at a distance of one Astronomical Unit (AU) from the sun, the resulting force is typically several orders of magnitude smaller than those produce by traditional fuel-based propulsion systems. Yet through clever manipulation of the spacecraft's attitude this force has potential for use in applications ranging from minor orbital corrections to interplanetary travel. However, solar sails are still in their technological infancy; with little actual in-flight data their exact dynamics remain a mystery.

There are several theoretical advantages a solar sail has over more conventional forms of spacecraft propulsion, yet all pale in comparison to the lack of a propellant. Rocket powered propulsion requires significant quantities of fuel to be carried with the vehicle, often weighing as much, if not more than, the rest of the spacecraft combined. When this is coupled with weight constraints offered by current launch vehicles, it is easy to understand why such limited payloads have

been historically delivered to distant planets. Conversely, solar sails rely on the propellant made endlessly available from the sun: radiation. Only the sail itself and an adequate attitude control system are required to propel the craft, removing the need for fuel almost entirely. As a result, much more of a mass budget is available to the vehicle payload, potentially enhancing returns on an already expensive mission.

Likewise, these sails also have several theoretical disadvantages, the most influential of which is the lack of flight heritage. Without an extensive compilation of flight history, the sail's performance cannot be reliably predicted. Given the expense of spaceflight, such uncertainty is a risk few mission are willing to accept, perpetuating the problem.

It is precisely this uncertainty that this work attempts to help address. Although the best remedy to uncertainty is experimentation, (and thus removal of the uncertainty) the next best option is to quantify its potential impact. If it can be demonstrated that some unknown factor will have little impact in a meaningful way then it can be safely ignored. If not, the investigation could still stumble upon mitigation techniques or opportunities for a similar result. Even if neither is the case such investigations still contributes by helping evaluate the true risk presented by a solar sail.

This investigation targets the potential of an orbital correction to help mitigate optical property uncertainty. The reflectivity and absorptivity of the sail are the primary parameters responsible for allowing the sail to act as an effective source of propulsion, but they are subject to degrade under exposure to the radiation of the sun. To make matters worse this degradation has yet to be studied in the space environment so it is not precisely understood how big of an impact this degradation may have. Simulations, on the other hand, have suggested that failure to accurately predict optical degradation could cause significant positional error if left unchecked.

The remainder of this section one contains a short history of solar sails, an in-depth problem description and a brief literature review. Section 2 addresses the method and mathematics used simulate sail trajectories and corrections. Section 3 describes the employed solar sail model, as well as a couple solutions to problems it presents. Simulation results and discussion are presented in Section 4, leaving Section 5 for the conclusion.

1.1.2 A Short History of Solar Sails

The founding theory behind a solar sail is nothing new. James Clerk Maxwell theorized on the existence of light pressure in 1873 and physicists and science-fiction writers alike have since imagined using “tremendous mirrors” to navigate the solar system. The practical application of a solar sail was not significantly

considered by the engineering community until the early 1970's, when NASA's Jet Propulsion Laboratory led an investigation into the use of a sail as a means to rendezvous with comet Halley during it close approach of the sun. Due to the risk associated with the untested technology, a solar-electric propulsion system was chosen over the solar sail before the project was ultimately scrapped due to cost. [1]

General interest in solar sailing was rekindled in the late 2000's after the launch of several missions intending to employ a sail for more than test purposes. The four most notable missions, including their fate, are as follows:

1) Cosmos 1 (2005) was to be the first solar sail to use photonic pressure for trajectory change, but the launch vehicle it was launched upon failed during ascent. [2]

2) NanoSail-D (2008) met a similar fate as its predecessor when the Falcon1 rocket it launched upon experienced a failure. [3]

3) The Japanese Aerospace Exploration Agency successfully launched the IKAROS mission in 2008, traveling to Venus and demonstrating attitude control using solar pressure. [4]

4) NanoSail-D2 successfully completed its mission after initially failing to deploy from its mother ship in 2011. [5]

Both the IKAROS and NanoSail-D2 missions completed successfully yet a solar sail has yet to be used as a primary source of spacecraft propulsion. These

missions both deployed a solar sail, and demonstrated that the spacecraft experienced a photonic pressure that influenced its orbit, but the influence of solar radiation pressure is a well-known and planned for phenomenon. The value of these missions is instead of a technical demonstration for missions that will rely on a solar sail as the primary source of propulsion.

1.2 Literature Review and Related Work

To the knowledge of the author, there have been few publications that have investigated the impact of advanced optical degradation effects on solar sail trajectories. A generalized model for non-perfectly reflective solar sails was introduced by Rios-Reyes and Scheeres in 2005 [6] and has been used as the foundation for most contemporary non-ideal solar sail works. The same authors applied their work to demonstrate feasible control laws for orbiting artificial Lagrange points while under the influence of uncertain optical properties.[7] Other authors have used this model for purposes ranging from creating optimum solar sail control laws[8] to modeling the YORP effect on asteroids.[9][10]

Dachwald et. al. improved upon this model by allowing the solar sail optical properties to degrade exponentially as a function of radiation dosage.[11] This model has been used to gauge its influence on several different optimum interplanetary trajectories, demonstrating that the effects of optical degradation cannot be safely ignored.[12] These two publications are the primary works upon

which this thesis is based, and will be referred-to repeatedly in proceeding sections. Although the results of these two works are frequently cited, to the knowledge of the author the model has yet to be furthered in a considerable way.

As reported by in the previous two works, there has yet to be conclusive ground-based testing to define the sort of optical degradation that is expected to take place in orbit. One test reported measurable changes to absorptivity and emissivity of thin Kapton/Aluminum films when exposed to high-energy photons and electrons in a simulated space environment.[13] However, another work found that there was no measurable changes to the optical properties of similar films when exposed to electrons alone.[14]

1.3 Statement of Problem

Since the 1970's there has been considerable work creating solar sail models, many of which include different performance criteria. Likewise, there has also been considerable work in the field of ideal solar sail trajectory determination and optimization. However, to the knowledge of the author, little has been done to connect these two fields to determine the effect of performance criteria on trajectories, and even fewer investigations into mitigation of the impact of the performance effects. A more detailed examination of the historic work in the field can be found in the Literature Review section.

This lack of investigations is likely due to the history of the solar sail. Much of the foundational and ideal work was performed in the 1970's while NASA seriously considered a solar sail mission. When the mission was later abandoned, general interest dwindled until it was recently rekindled in the past 15 years. As a result, this particular field has had only a couple decades to mature. The impact of degradation on solar sail trajectories hasn't been studied in detail not because it is of little importance or interest, but instead because the field is still young.

As a continuation of the recent work connecting performance to trajectories, the general goal of this thesis is to further the investigation on the impact of optical degradation on interplanetary trajectories. Optical properties of the sail, such as the reflectivity and emissivity, are of critical importance because they are precisely the parameters responsible for converting radiation flux into photonic pressure. Without a firm knowledge of these parameters and how they change with time, it is impossible to calculate a realistic propulsive force, and thus orbits cannot be calculated with certainty.

To the knowledge of the author, there has only been one previous investigation into this particular subject. Dachwald et. al. implemented a simple exponential degradation model and showed that optimal interplanetary transits relied heavily on the extent and timing of optical depredation. As a first step into the field this work showed that this single performance aspect must be considered in potential

trajectory design. It also acknowledges that such data can only be obtained by analyzing actual on-flight data, due to the failure of ground-based testing to lead to conclusive results. [12]

This thesis intends to answer questions that logically extend from this previous work. Although it is clear that degradation impacts these trajectories, what is not clear is whether there are ways to mitigate this impact or whether knowledge of such degradation is even required before launch. Orbital corrections are carried out frequently with rocket-propelled missions, so it may be possible to gather degradation data mid-flight and correct during transit. If this is the case then it is worth investigating such options until detailed degradation profiles can be developed experimental data.

To be exact, the question this thesis attempts to answer could be phrased:

"Given a sample interplanetary mission, is there a time before which degradation data could be collected, and after which this data could be used to perform an orbital correction resulting in the successful approach of the target destination?"

Of course, this phrasing is quite broad, so it is of benefit to narrow a few words down before continuing. First, the "successful approach of the target destination" is taken to mean that the solar sail arrives within a threshold distance from the target planet at a pre-defined date. It is possible that a correction would not be

sufficient to make a particular date, but could be allow for a slightly late arrival. This possibility is not considered in this work in an effort to limit its scope. Second, an "orbital correction" is assumed to be more of a "trajectory correction" since solar sails the correction itself does not occur at only one time. Instead, this correction is taken to be a potentially new trajectory calculated after the degradation data has been collected.

1.3.1 Test Case

The sample mission assumed in this work is an interplanetary trajectory from Earth to Venus. As mentioned in another section, one of the biggest advantages of a solar sail is the ability to accumulate or remove large amounts of angular momentum over large period of time. For this reason, interplanetary trajectories are a natural choice for a candidate solar sail mission as the lengthy transfer makes the sail's strength most readily apparent.

Venus has been chosen as a destination because it orbits relatively close to sun. The force of a sail is inversely proportional to the square to the sail-sun distance, so decreasing distance to the sun is the fastest way to accumulate angular momentum. It is therefore desirable to initially orbit close to the sun in order to expedite the trajectory, even if the target sits in the outer solar system. It follows then that the results of this investigation are potentially relevant to all interplanetary solar sail missions.

An arbitrary launch date of August 28, 2003, and an assumed transfer time of 300 days define the assumed mission timeframe. This translates to potential trajectories that rotate less than one full revolution around the sun before arrival at Venus. It is also assumed the sail is solely responsible for the transfer between planets; the spacecraft's initial and final state is equal to that of Earth and Venus respectively. There is nothing special about these mission parameters except that they create a mission that takes place with overhead time to allow for a correction during the trajectory and require the solar sail to provide the momentum required for an interplanetary transfer. Although variations on these parameters could potentially improve the breadth of this investigation, they are considered out of the scope of this investigation. The remaining mission parameters can be found in appendix A.

1.3.2 Mathematical Description

It should be clear that a method of calculating spacecraft point-to-point trajectories is essential to this analysis. The following section offers a mathematical description of these trajectories to explicitly outline the goal of the algorithm described in a following section.

Generally speaking, the dynamics of a spacecraft moving through space is described by a set of six differential equations, typically referred to as its equations of motion. Since forces (which relate to the acceleration of the body)

define these equations of motion both the velocity and position are needed to fully describe the state of the spacecraft. Both position and velocity are three-dimension vectors so a total of six differential equations are required. These differential equations (Eq. 41-46) can be integrated to calculate the motion of the spacecraft as a function of time, i.e.:

$$S(t) = \int_{t_0}^{t} \dot{S}\, dt \qquad \mathbf{1}$$

where S is the state of the spacecraft in the classical mechanical sense, $\dot{S}$ are the equations of motion in differential form and t_0 is the time at the start of the simulation. Since the desired trajectories leave Earth at time t_0 and arrive at Venus at time t_f, these trajectories must also satisfy boundary conditions

$$S(t_0) = S_{Earth}(t_0) \qquad and \qquad S(t_f) = S_{Venus}(t_f) \qquad \mathbf{2}$$

where the subscripts indicate that the state S is also that of a planet. This implies that spacecraft is assumed to ideally arrive and depart with zero excess escape velocity.

Since analytical methods are currently insufficient to find solutions to Equation 1 that satisfy Equation 2, (given the complexity of advanced solar sail equations of motion) numerical estimations must be used. Estimation always leads to some degree of error, so it is accepted that the final boundary condition can never be satisfied completely. For this reason it is then convenient to define two different

types of trajectories: sufficient and insufficient trajectories. Sufficient trajectories are those that satisfy the final boundary condition to within some precision, and are classified as a "hit." Insufficient trajectories, on the other hand, fail to meet this precision and represent a "miss."

A combination of insight into the nature of trajectories and statistical analysis of numerical error is used to define this threshold, which can be found in section 4.3.1 Error Sources. The initial boundary condition can be guaranteed to always be true.

1.3.3 Independent Variables

This investigation focuses on the impact of two factors: the degradation factor d and degradation time constant λ which are discussed at length in section 2.2.1 Degrading Optical Parameters. Collectively these two variables define the extent and speed at which the sail is modeled to degrade, both of which are uncertain solar sail characteristics. In the previous work by Dachwald et. al.[12], only d was varied in order to study its impact on optimum solar sail trajectories; the influence of λ remain unclear. The influence of both of these variables on the possibility of orbital corrections will be investigated, including any synergistic effects. It should be noted that both of these variables relate to an exponential model, the only kind considered in this work.

1.4 General Notes

1.4.1 Assumptions

Several simplifying assumption are made about the nature of solar sails in the following analysis. It is assumed that:

1. The sail is perfectly flat and rigid.

2. The exact position and velocity of the sail is known without error.

3. Relativistic effects can be ignored.

4. The sail is always in thermal equilibrium

5. Sail degradation takes place as described by the model in section 2.2 Non-Ideal Solar Sail.

1.4.2 On Reflectivity and Ideal Sails

The general academic aerospace community considers an ideal solar sail to have the property of perfect reflectivity - a simplifying assumption used since the early days of sail analysis. However, this assumption is frequently removed in contemporary work now that computation and numerical methods can easily accommodate non-perfect reflectivity. Since this thesis addresses the impact of changing reflectivity, the phrase "ideal solar sail" is taken to imply a sail whose optical parameters remain constant with time, regardless of the value of the

reflectivity. In contrast, a "non-ideal sail" is reserved for those whose optical properties are expected to change, adding an additional level of complexity. This distinction is intended to help highlight behavior that is the result of optical degradation and not that of non-perfect reflectivity.

2. MODELS

2.1 Ideal Solar Sail

The simplest of solar sail models describe the force generated by a perfectly reflective, flat and time invariant solar sail. In such a case, the total force from solar radiation is composed by equal components resulting from impacting and reflected radiation. According to McInnes,[1] this force can be expressed given the equation:

$$\boldsymbol{F} = 2\,PA(\mathbf{r}\cdot\,\boldsymbol{n})^2\boldsymbol{n} \qquad \mathbf{3}$$

where $\boldsymbol{F}$ is the force due to solar radiation, P is solar pressure, and A is the area of the sail. When the sail is perpendicular to incoming radiation, and thus the unit vectors $\boldsymbol{n}$ (pointing in the sail normal direction) and $\boldsymbol{r}$ (pointing along the sun-sail line) are parallel, the force is equation to twice the product of the sail area and radiation pressure. The two in the equation is representative of the fact that there is a force that acts on the sail when from the interception of incident radiation, and another force that acts as the radiation is subsequently reflected. When the sail is parallel to incoming radiation the dot product $\mathbf{r}\cdot\,\boldsymbol{n}$ is equal to zero; no radiation is intercepted and there is no resulting force. For all intermediate cases it can be recognized that this dot product is equal to the cosine of the incident reflection angle α (referred to as the cone angle in later sections) allowing equation 3 can be re-written: [1]

$$\boldsymbol{F} = 2\,PA\cos^2\alpha\,\boldsymbol{n} \qquad \textbf{4}$$

Equation 4 includes the square of $\cos\ \alpha$: one factor is equal to the dot product $\mathbf{r} \cdot \boldsymbol{n}$ while the accounts for the cross section area of the sail as it is rotated.

This model has an appreciable quality that the non-ideal model will not; the resulting force will always act along the sail normal direction. This is due to the fact that the incident and reflection forces are equal in magnitude and symmetric about the sail normal direction. Components of these forces in the normal directions add together while the tangential components cancel out. Figure 26 in Appendix D can help illustrate the concept.

The ideal sail model allows a force to be created in any direction that the sail can be pointed. Equation 4 can be easily manipulated to determine the required incident/cone angle to create a force in any given direction. It should also be noted that this equation implies that the force-cone angle relationship is unique in the sense that an arbitrary force direction will only be the result of a single α within the domain $[-\frac{\pi}{2}, \frac{\pi}{2}]$. (These bounds are used to exclude the possibility of exposing the backside of the solar sail to radiation)

2.2 Non-Ideal Solar Sail

The ideal solar sail model can be improved by removing the perfect reflectivity assumption. Rios-Reyes and Scheeres[6] derive equations for the differential

force acting upon a different sail area as a function of various optical properties. First, the differential force in the sail normal direction is described by the equation

$$dF_{\perp} = P[a_1 \cos^2 \alpha + a_2 \cos \alpha]\, dA\, \boldsymbol{n} \quad \textbf{5}$$

where

$$a_1 = (1 + s\rho) \quad \textbf{6}$$

and

$$a_2 = \left[B_f(1-s)\rho + (1-\rho)\frac{\epsilon_f B_f - \epsilon_b B_b}{\epsilon_f + \epsilon_b}\right] \quad \textbf{7}$$

In the above expression ρ, s, ϵ and B are the reflectivity, specular reflection, emissivity and non-Lambertian property factor coefficients of the sail, respectively. The subscripts b and f refer to the back and the front of the solar sail, while $\perp$ indicates perpendicularity to sail face.

The coefficients ρ, s and B express, on a 0 to 1 scale, how "ideal" the property is. For instance, $\rho = 1$ indicates that all radiation is reflected and none is absorbed. Similarly, $s = 0$ and $s = 1$ indicates that the reflection takes place either perfectly diffusely or specularly, respectively. A coefficient B with a value of 1 indicates

that a surfaces is omni-directionally diffusely reflective, as opposed to a surface one that yeilds specular highlights.

The differential force in the sail traverse direction, yet still in the incident plane, can be expressed a similar fashion as the previous equation:

$$dF_{\parallel} = -Pa_3 \sin\alpha \cos\alpha \, dA \, \boldsymbol{t} \tag{8}$$

where

$$a_3 = (1 - s\rho) \tag{9}$$

and the unit vector $\boldsymbol{t}$ act perpendicular to $\boldsymbol{n}$ yet still in the orbital frame. Combined, these two components constitute the non-ideal solar sail force model.

There is a fairly significant implication to this model that should be addressed. It improves upon the ideal model by adding a term in the sail normal direction that is the result of the combined forces generated by re-emitted radiation from both the front and back of the sail as well as non-Lambertian reflection. At small cone angles, this term is small in magnitude compared to the reflection term and mainly acts to decrease the magnitude of the overall photonic force. At large cone angles, near $\pm\pi$, this term overtakes reflection term in magnitude, causing the sail normal force to act toward incoming incident rays.

This behavior adds a level of mathematical complexity to the model. Since the force is no longer solely due to the reflection of light, but instead the combination of several optical factors, it is difficult to invert the non-ideal model. This is to say that numerical methods are required to determine which set of cone and clock angles will produce a force in a pre-determine direction, whereas the reverse can be completed with simple substitution. The inversion of this model is the focus of section 2.2.3 Mapping Forces to Angles.

2.2.1 Degrading Optical Parameters

Dachwald et. al.[12] expanded the above non-ideal solar sail model to allow for optical properties of the sail to degrade over time. Objects exposed to radiation for any length of time will experience some level of change in appearance, and solar sails are no different. The proposed model assumes that this optical degradation occurs exponentially as a function of both solar distance and impacting flux, thus requiring an integral over time. First, a dimensionless dosage Σ is defined as

$$\Sigma(\mathrm{t}) = \; r_0^2 \int_{t_0}^{t} \frac{\cos\alpha}{r^2} \quad per \quad year \tag{10}$$

where r and r_0 is the sun-sail distance at time t and t_0 respectively. This is accompanied by a degradation time constant λ based on a "half-life solar radiation dose" $\hat{\Sigma}$ as follows:

$$\lambda = \frac{\ln 2}{\hat{\Sigma}} \tag{11}$$

Each of the parameters τ in the non-ideal solar sail model are assumed to vary as follows:[12]

$$\begin{aligned} \frac{\tau(t)}{\tau_0} &= \frac{1 + de^{-\lambda\Sigma(t)}}{1 + d} && \tau \in \{\rho, s\} \\ \frac{\tau(t)}{\tau_0} &= 1 + d\left(1 - e^{\lambda\Sigma(t)}\right) && \tau = \epsilon_f \\ \frac{\tau(t)}{\tau_0} &= 1 && \tau \in \{\epsilon_b, B_f, B_b\} \end{aligned} \tag{12}$$

Equations 5-12 comprise the foundation of the model used in this thesis.

2.2.2 Rotation into the Orbital Frame

There are two frames of importance in this paper. The first is the Orbital Frame, which is defined by the orbit of the solar sail and used to express the equations of motion. The second frame is one fixed to the solar sail, called the body frame.

The orbital frame is composed by an orthogonal set of three orthogonal unit vectors and is illustrated in figure 1. The first unit vector, $\boldsymbol{r}$, acts along the line drawn between the sun and the solar sail. The second, $\boldsymbol{t}$, is normal to $\boldsymbol{r}$ and acts in the plane swept out by the orbit of the solar sail. The final direction, $\boldsymbol{h}$, completes the set, but is not included in the diagram because it would point toward the reader. The origin of the frame follows the solar sail, but it is not rigidly attached.

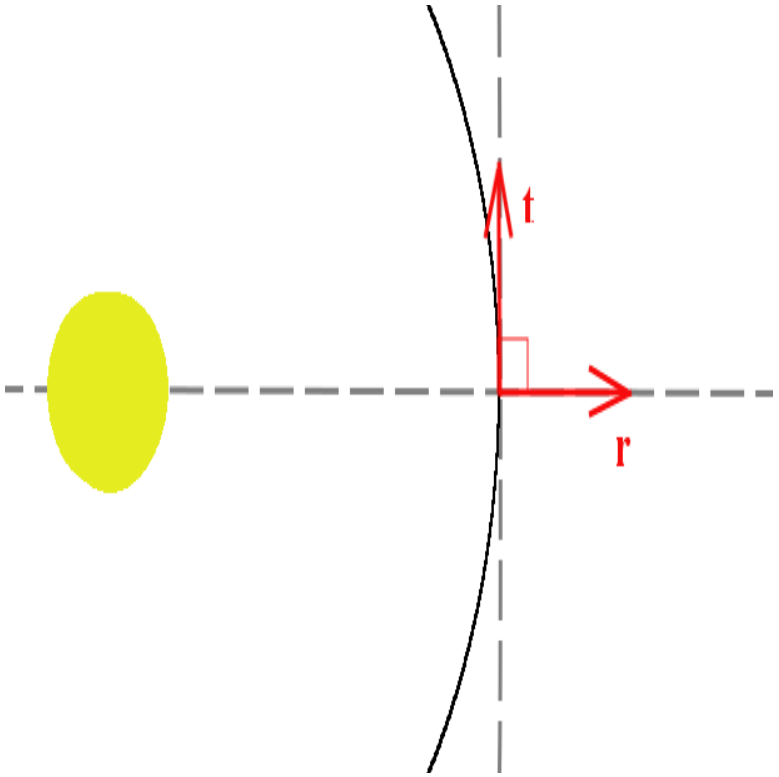

Figure 1 - The orbital frame

The second frame, called the body frame, is another orthogonal right-handed set of three unit vectors that are rigidly attached to the solar sail. Figure 2 is a diagram of this frame; the sail is the orange line and defines while α and δ are angles used to rotate into the Orbital Frame. The first unit vector is $\boldsymbol{n}$, which points in a direction perpendicular to the face of the sail. In figure 2 this vector is also given the notation $\perp$ to make the diagram correspond with equation 5. The second unit vector, $\parallel$, acts perpendicular to $\boldsymbol{n}$ yet also in the plane defined by the incident and secularly reflected radiation. The final unit vector, Ψ, completes the set and is omitted from the diagram because it would act toward the reader.

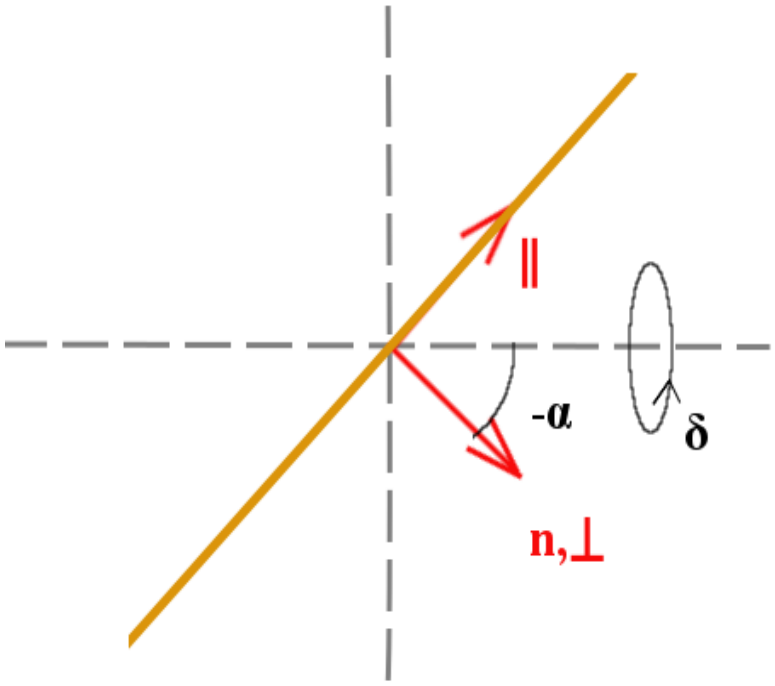

Figure 2 - The body frame

The angles α and δ serve as the control angles of the solar sail, and are often called the cone and clock angles. The cone angle, α, is the angle between the sail normal vector and the sun-sail line, as well as the angle of incident specular reflection. The clock angle, δ, is a rotation of the sail about $\boldsymbol{r}$, making the vector $\boldsymbol{n}$ sweep out the shape of a cone.

In order to rotate a force from the body frame to the orbital frame the following relation is used:

$$\begin{pmatrix} F_r \\ F_t \\ F_h \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\delta & \sin\delta \\ 0 & -\sin\delta & \cos\delta \end{bmatrix} \begin{bmatrix} \cos\alpha & sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} F_\perp \\ F_\parallel \\ F_\Psi \end{pmatrix} \quad \textbf{13}$$

where the subscripts correspond with the components of force in their respective frames.

It may be desired to calculate the cone and clock angle between these two frames. The cone angle can be found using the relation

$$\cos\alpha = \boldsymbol{r} \cdot \boldsymbol{n} \quad \textbf{14}$$

while the clock angle can be found with the equation[15]

$$\cos\delta = \frac{\boldsymbol{r}\times(\boldsymbol{n}\times\boldsymbol{r})}{|\boldsymbol{r}\times(\boldsymbol{n}\times\boldsymbol{r})|} \cdot \boldsymbol{t} \quad \textbf{15}$$

2.2.3 Mapping Forces to Angles

One of the functions performed by the algorithm, presented later in this thesis, determines the cone and clock angles necessary to create a force in a desired direction. Thus the non-ideal solar sail model is effectively inverted because the product of the model is used to find the corresponding inputs. While the clock angle is calculated analytically, the cone angle requires a numerical method to be found.

A three step strategy is used to calculate the cone angle. First, a new function is defined from the desired force direction, which is equal to zero when the cone angle has been found. Second, the derivative of this function with respect to the cone angle is calculated, set to zero and solved analytically. The result is a set of critical cone angles which mark the local bounds of the function. Finally, these angles are used as the starting points for a standard bisection method, which locates the zero crossing of the function, yielding the cone angles that produces the desired force.

2.2.3.1 Clock Angle

The clock angle is calculated by expanding equation 13 and observing that, after some careful factoring and reduction,

$$\frac{F_h}{F_t} = \tan\delta \qquad \textbf{16}$$

Since the ratio of F_h and F_t are given by the force direction vector, only a simple inverse tangent operation is required to find the desired angle.

2.2.3.2 Cone Angle

Unfortunately, a similar calculation for the cone angle was not found during the course of this work. This section presents the formulation of the numeric approach used to find the angle, as well some investigation into some of the implications of the model that complicates the calculation.

The formulation of the numerical method starts by mathematically defining the goal of the calculation. The desired cone angle is one that results in a realizable photonic force, F_{real}, that acts in the same direction as a desired force, F_{ideal}. This is calculated by normalizing each force and requiring each has an equal component in the $\boldsymbol{r}$ direction, i.e:

$$\frac{\mathrm{F_{r_{ideal}}}}{\|F_{ideal}\|} = \frac{F_{r_{real}}}{\|F_{real}\|} \quad \textbf{17}$$

In equation 17 the double bars indicate the use of the Euclidian norm. This equation requires that both components of force in the $\boldsymbol{r}$ direction are of equal magnitude and symmetric about $\boldsymbol{r}$. After some consideration, it can be concluded that this condition requires the forces to act in the same direction if the other two components of force share signs between both forces. To perform this calculate, the norm of F_{real} is rewritten using the Pythagorean theorem:

$$\|F_{ideal}\| = \frac{F_{real}}{\sqrt{F_{\perp}^2 + F_{\parallel}^2 + F_{\Psi}^2}} \quad \textbf{18}$$

Next, this equation is made specific to the non-ideal solar sail model by substituted in equations 5 and 8, yielding the equation:

$$\|F_{ideal}\| = \frac{1}{\sqrt{a_1^2 \cos^2\alpha + 2a_1 a_2 \cos\alpha + a_2^2 - a_3^2 \cos^2\alpha + a_3^2}} * \ldots$$

$$\begin{bmatrix} a_3 + a_2 \cos\alpha + a_1 \cos^2\alpha - a_3 \cos^2\alpha \\ \cos\alpha \cos\delta \sin\alpha\,(a_2 + a_1 \cos\alpha - a_3 \cos\alpha) \\ \cos\alpha \sin\delta \sin\alpha\,(a_2 + a_1 \cos\alpha - a_3 \cos\alpha) \end{bmatrix}$$

19

If cone and clock angles α and δ have been found that satisfy equation 19, the goal has then been achieved.

As is stated by equation 17, only the first component, in the $\boldsymbol{r}$ direction, is required. This is because δ can be calculated analytically, only α is unknown. The first component of the force, that in the sun-sail direction $\boldsymbol{r}$ is chosen because it only includes powers of $\cos\alpha$, a quality which is exploited later in this section. With this consideration, equation 17 can be simplified to the equation

$$F_{r_{norm}} - \frac{a_3 + a_2 \cos\alpha + a_1 \cos^2\alpha - a_3 \cos^2\alpha}{\sqrt{a_1^2 \cos^2\alpha + 2a_1 a_2 \cos\alpha + a_2^2 - a_3^2 \cos^2\alpha + a_3^2}} = 0 \qquad \textbf{20}$$

which takes the standard form to highlight that a zero crossing of this function yields the desired cone angle.

From here, a bisection method alone could be used to isolate zero crossings, but it is prudent to first discuss the solar sail behavior revealed by equation 20. Figure 3 shows the relationship between the constructed function and α, in blue, for a

perfectly reflective solar sail. The green lines mark artificial bounds of the domain at $\pm\frac{\pi}{2}$, which is put in place to prevent the sail from exposing its backside to the sun. The red lines identify a zero crossing of the function, which corresponds to cone angles that create the desired force.

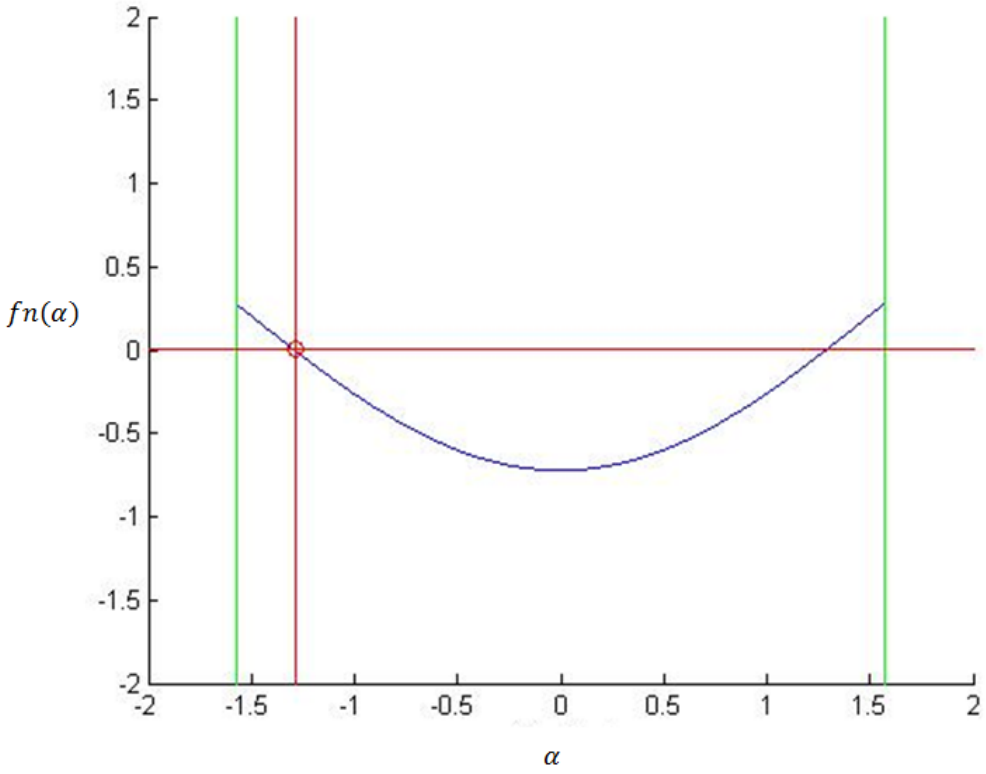

Figure 3 - The relationship between the constructed function (eq. 20) and cone angle α for an ideal solar sail.

As is evident in the figure, the force in the radial direction for a perfectly reflective solar sail displays simple, unimodal behavior. There are two zero

crossings to the function, but only one is calculated since the function is symmetric. Both cone angles result in a force with the appropriate component in the F_R direction, but a quadrant check is necessary to ensure that the other component have the correct sign.

Figure 4 is the same representation as the last figure for a non-perfectly reflective solar sail. It is immediately apparent that the function is no longer unimodal with respect to α. Instead, there are now potentially up to six different zeros crossings, three of which produce a force in the desired direction. This also implies that, as the sail is rotated in a single direction, the resulting force does not necessarily move in the same direction; an odd result when compared to an ideal sail. With up to three potential cone angles that result in a force in an arbitrary direction, care must now be taken to discriminate between zero crossing.

This effect is perhaps best visualized by examining a comparison of solar sail force bubbles in Figure 5. Here, the colored curves represent the range of possible force vectors that results from sweeping α from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ for both an ideal and non-ideal sail of arbitrary dimensions. The non-ideal sail force bubble, in red, is smaller in magnitude at all cone angles, but generally behaves in a similar manner.

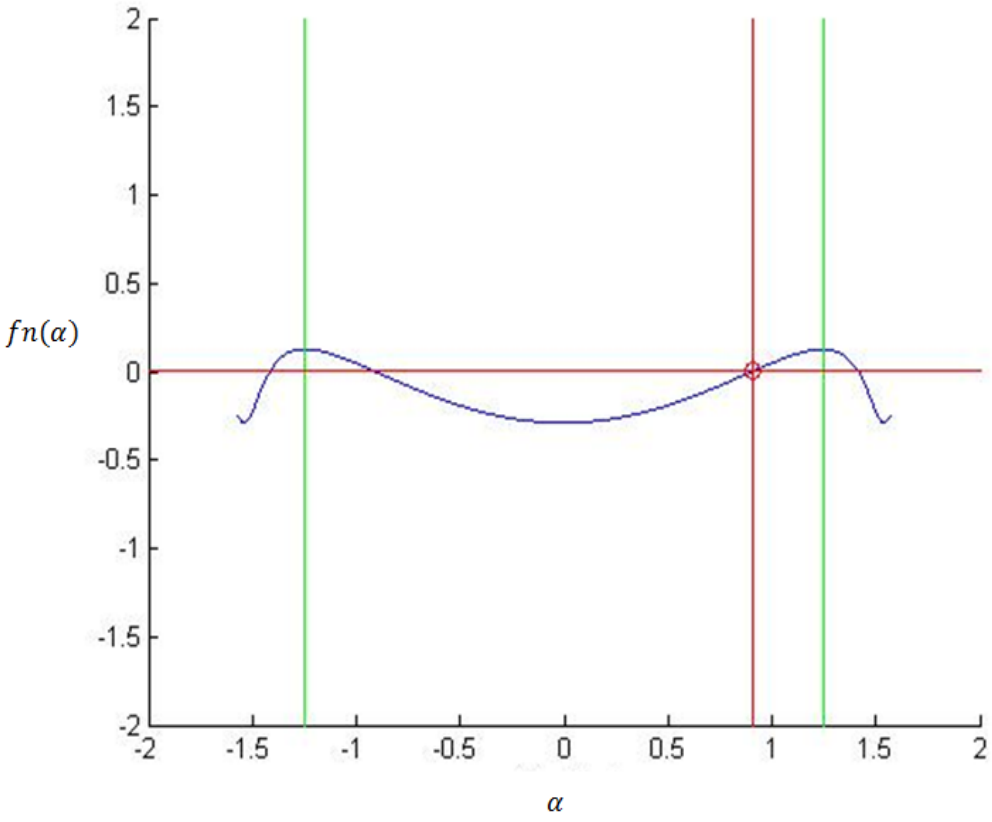

Figure 4 - The relationship between the constructed function (eq. 20) and cone angle α for a non-ideal solar sail.

The difference between the models, however, is most apparent when the force bubbles are examined at cone angles near $\pm\frac{\pi}{2}$, which can be viewed in Figure 6. The non-ideal solar sail model creates a "loop", where the forces due emission and reflection are nearly in balance. When viewed at this level it is clear how the model maps multiple cone angles to forces in a given direction. For instance, the intersection of the red curve with itself in Figure 6 corresponds to two different angles that yield forces purely in the F_r direction.

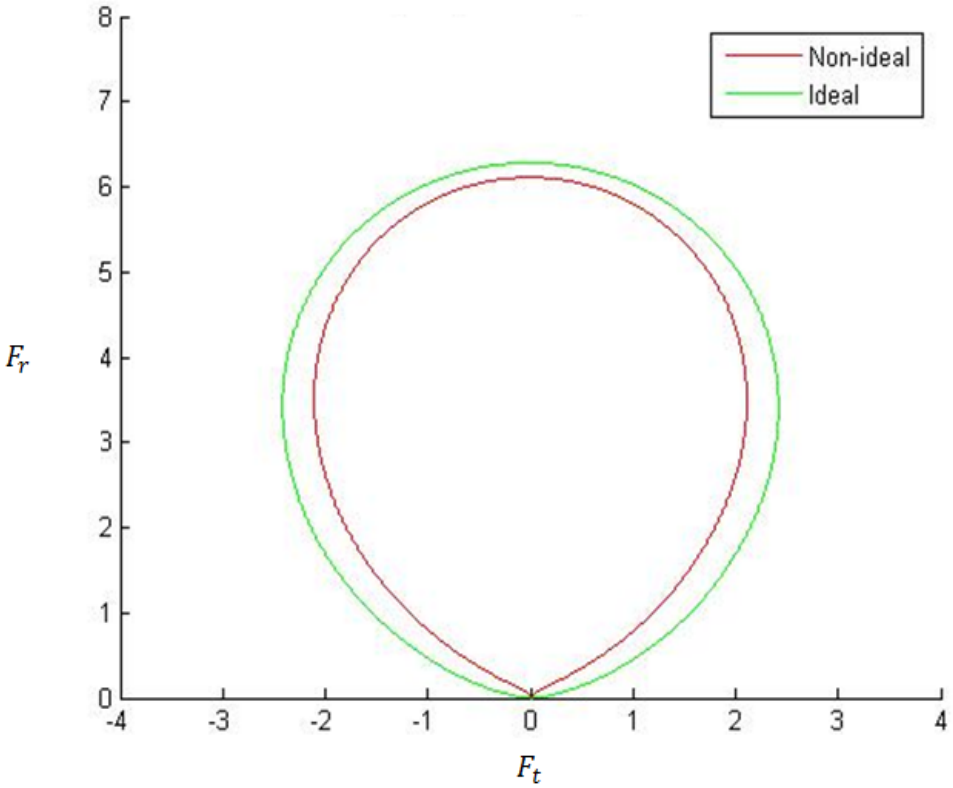

Figure 5 - A comparison between an ideal and non-ideal solar sail force bubble.

In order to avoid this non-uniqueness between angles and force directions, the "loop" is removed from consideration. The forces in this region are small; typically less than two orders of magnitude less than the rest of the bubble, so a result in this region is rounded to zero. By discounting these potential forces a small numeric error is potentially introduced to the algorithm, but it is considered small when compared to other unconsidered orbital perturbations.

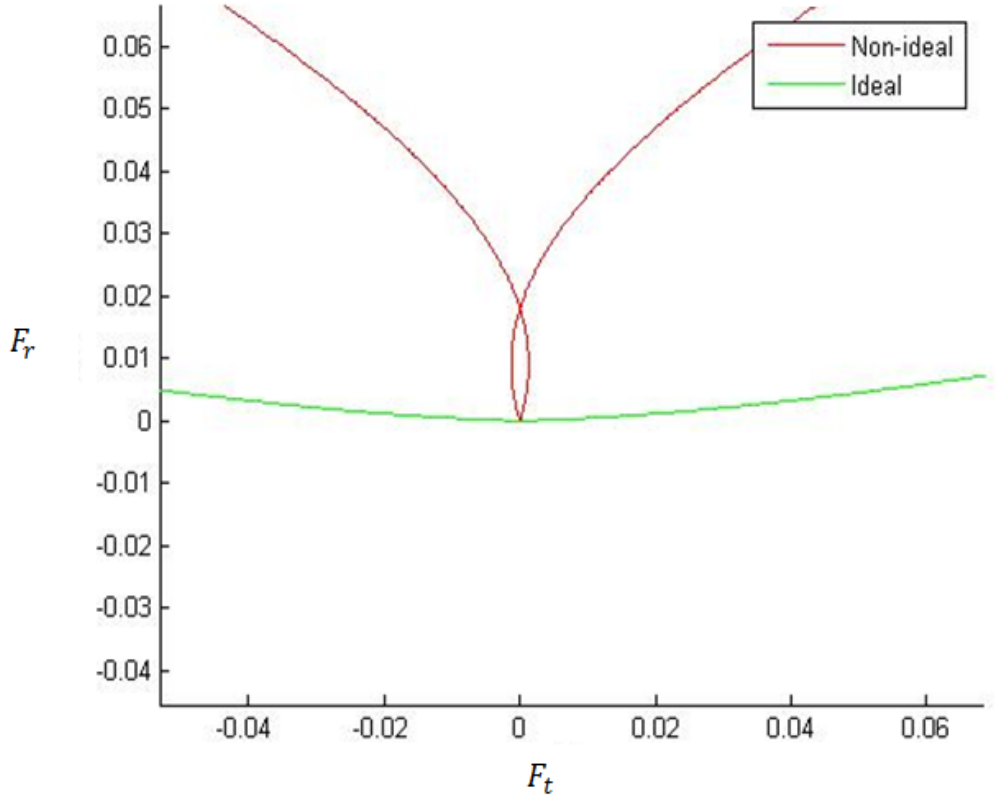

Figure 6 - A comparison between an ideal and non-ideal solar sail force bubble near the origin.

In order to isolate the zero crossings of equation 20, the local optima are found analytically. The method of bisection then uses these optima as starting points if they have opposite signs. By taking the derivative of equation 20 with respect to α, we find that the result takes the form

$$0 = \frac{n_1 n_2 n_3}{d_1} \qquad \textbf{21}$$

where

$$n_1 = \mathrm{A}^3\mathrm{P}^3 \cos^3 \alpha \sin \alpha \quad \textbf{22}$$

$$n_2 = \mathrm{a}_2 + \mathrm{a}_1 \cos \alpha - \mathrm{a}_3 \cos \alpha \quad \textbf{23}$$

$$\mathrm{n}_3 = \mathrm{a}_1^2 \cos^2 \alpha + 2\,\mathrm{a}_1\mathrm{a}_2 \cos \alpha - \mathrm{a}_1\mathrm{a}_3 + \mathrm{a}_2^2 - \mathrm{a}_2\mathrm{a}_3 \cos \alpha - \mathrm{a}_3^2 \cos^2 \alpha \ldots + \mathrm{a}_3^2 \quad \textbf{24}$$

and

$$d_1 = \left(\mathrm{A}^2\mathrm{P}^2 \cos^2 \alpha \left(\mathrm{a}_1^2 \cos^2 \alpha + 2\,\mathrm{a}_1\mathrm{a}_2 \cos \alpha + \mathrm{a}_2^2 - \mathrm{a}_3^2 \cos^2 \alpha + \mathrm{a}_3^2\right)\right)^{\frac{3}{2}} \quad \textbf{25}$$

Since the solution of equation 21 exist at the roots, an optima is found when n_1,n_2 or n_3 is equal to zero. By inspection, we can determine that n_1 only equals zero at when $\cos \alpha$ or when $\sin \alpha$ is equal to zero: all multiples of $\frac{\pi}{2}$. We can find the root due to n_2 by setting it equal to zero and solving for α, finding that

$$\alpha_{crit_1} = \operatorname{acos}\left(-\frac{a_2}{a_1 - a_3}\right) \quad \textbf{26}$$

The final root of the equation can be found by setting n_3 equal to zero and solving once more. Although it is more complex that the other two, we can observe that it is quadratic with respect to $\cos \alpha$, and thus the quadratic equation can be used to find that roots exist at:

$$\alpha_{\mathrm{crit}_{2,3}} = \operatorname{acos}(H) \quad \textbf{27}$$

where

$$H = \frac{a_2 a_3 \; - \; 2a_1 a_2 \; \pm \sqrt{(2a_1 a_2 \; - \; a_2 a_3)^2 \; - \; (4a_1^2 \; - \; 4\, a_3^2)(a_2^2 \; + \; a_3^2 \; - \; a_1 a_3)}}{2a_1^2 \; - \; 2a_3^2}$$

28

When all α_{crit} and corresponding optima of equation 21 has been found, the sign of the function at these points is used in a bisection method isolate zero crossing. If the function has multiple zero crossings, then only the optima relating to equation 23 are used to avoid the non-uniqueness of the problem. This corresponds with removing the "loop" identified in figure 6.

2.3 Equations of Motion

The equations of motion for a spacecraft can be expressed by a set of six differential equations. Known as Gauss' form of the variation equations, these equations are given by:[15]

$$\dot{a} = \quad \frac{2a^2}{\zeta}\left[e \sin\theta \, F_r + \frac{p}{r} F_t\right] \qquad \textbf{29}$$

$$\dot{e} = \quad \frac{1}{\zeta}[p \sin\theta \, F_r + \{(p + r)\cos\theta + re\}F_t] \qquad \textbf{30}$$

$$i = \quad \frac{r\cos(\omega + \theta)}{\zeta} F_h \qquad \textbf{31}$$

$$\dot{\Omega} = \quad \frac{r \sin(\omega + \theta)}{\zeta} F_h \qquad \textbf{32}$$

$$\dot{\omega} = \quad -\frac{r\sin(\omega+\theta)}{\zeta \tan i}F_r + \frac{1}{e\zeta}[-p\cos\theta F_r + \ (p+r)\sin\theta F_t] \qquad \textbf{33}$$

$$\dot{\theta} = \quad \frac{\zeta}{r^2} + \frac{1}{e\zeta}[p\cos\theta\, F_r - \ (p+r)\sin\theta F_t] \qquad \textbf{34}$$

where a, e, i, Ω, ω and θ are the six classical orbital elements and ζ is the specific relative angular momentum of the sail. F_r, F_t and F_h are the radial, tangential and out-of-plane perturbing forces in the orbital frame, which are a function of the sail's attitude and degradation. More information on how the force components are calculated can be found in section 2.2 Non-Ideal Solar Sail.

Of the classical orbital elements, the semi-major axis, a, and the eccentric, e, describe the size and shape of the ellipse that defines the orbit. The inclination, i, right ascension of the ascending node, Ω, and argument of periapsis , ω, are angles that define how the orbital plane is rotated with respect to the frame inertial fixed to the central body. The final classical orbital element is the true anomaly, θ, which defines where the body is in the orbital ellipse. A diagram of the classical orbital elements can be found in Appendix D. (Figure 27)

Although there are multiple possible representations for the system dynamics, this form is desirable because it includes the force of the body being orbited, and thus perturbing forces can be considered separately. This separation helps avoid

numerical difficulties associated with combining forces separated by several orders of magnitude.

However, even this form introduces difficulties. Several of these equations are singular when $e = 0$ or $i = 0 \pm 2n\pi$, so they can't be used to calculate trajectories that may pass through these regions. In order to avoid these singularities, the original orbital elements are translated into a different set of six, known as the equinoctial orbital elements, with relations given by Betts:[16]

$$\mathrm{p} = \mathrm{a}\,(1 - \mathrm{e}^2) \quad \textbf{35}$$

$$\mathrm{f} = \mathrm{e}\cos(\omega + \Omega) \quad \textbf{36}$$

$$\mathrm{g} = \mathrm{e}\sin(\omega + \Omega) \quad \textbf{37}$$

$$\mathrm{h} = \tan\left(\frac{i}{2}\right)\cos(\Omega) \quad \textbf{38}$$

$$\mathrm{k} = \tan\left(\frac{i}{2}\right)\sin(\Omega) \quad \textbf{39}$$

$$\mathrm{L} = \Omega + \omega + \theta \quad \textbf{40}$$

where $p, f, g, h\ k$ and L then act as the state of the spacecraft. After this translation, Equations 29 to 34 can be re-written as:[16]

$$\dot{p} = \frac{2p}{q}\sqrt{\frac{p}{\mu}}F_t \tag{41}$$

$$\dot{f} = \sqrt{\frac{p}{\mu}}\sin L\,F_r + \frac{1}{q}\sqrt{\frac{p}{\mu}}[(q+1)\cos L + f]F_t \ldots - \frac{g}{q}\sqrt{\frac{p}{\mu}}[\sin L - k\cos L]F_h \tag{42}$$

$$\dot{g} = -\sqrt{\frac{p}{\mu}}\cos L\,F_r + \frac{1}{q}\sqrt{\frac{p}{\mu}}[(q+1)\sin L + g]F_t \ldots + \frac{f}{q}\sqrt{\frac{p}{\mu}}[h\sin L - k\cos L]F_h \tag{43}$$

$$\dot{h} = \sqrt{\frac{p}{\mu}}\frac{s^2\cos L}{2q}F_h \tag{44}$$

$$\dot{k} = \sqrt{\frac{p}{\mu}}\frac{s^2\sin L}{2q}F_h \tag{45}$$

$$\dot{L} = \sqrt{\frac{p}{\mu}}\left(\frac{q}{p}\right)^2 + \sqrt{\frac{p}{\mu}}\frac{1}{q}[h\sin L - k\cos L]F_h \tag{46}$$

where

$$q = 1 + f\cos L + g\sin L \quad \textbf{47}$$

and

$$s^2 = 1 + h^2 + k^2 \quad \textbf{48}$$

In later parts of this section, references to the differential equation (DE) refer to Equations 42 to 46.

3. ALGORITHM

The following trajectories are calculated using a collocation method to transcribe a complex functional integration problem into a multivariable optimization problem. MATLAB's FSOLVE function is then used to optimize this new problem, resulting in a time history for a set of solar sail control angles. These angles are then sent to a multistep integration routine, which verifies that calculated control angles would result in an arrival at the intended target.

The algorithm proceeds through five cycles, each time improving the accuracy of the trajectory. In the first cycle a set of stochastically generated initial guesses are processed, after which any redundant resultant trajectory are deleted. At the beginning of every proceeding cycle the remaining trajectories are transformed into new initial guesses with a denser mesh and processed again. Every cycle increases the complexity of the problem, which in turn roughly doubles the time taken by the algorithm. In the interest of time the algorithm stops after the fifth cycle, but it certainly could be allowed to continue for improved accuracy.

This method was chosen over other potential candidate algorithms for several reasons. This first is that a collocation method can translate a complex set of differential equations into an even larger set of simpler equations. This is advantageous because there exists a wide array commercial, off-the-shelf equation solving algorithms that are suitable for estimating the solution of these less

complex problems, regardless of the number of equations. For instance, this collocation method allows the use of the FSOLVE routine, which is designed to minimize an arbitrarily large multivariable objective, to solve the non-linear boundary value problem.

The second major advantage of the collocation method is that it allows for "decisions" to be made considering the entire trajectory. Since a collocation method assumes a trajectory and then improves upon it, the entire trajectory is always available without integration. This means that changes can be made to the trajectory and that the subsequent effect can be easily measured. This is in opposition to a wide range of other methods, such as a shooter method, where the result of decisions made early in the trajectory is not clear until the trajectory has been integrated.

The proceeding section is arranged as a set of steps, which are analogous to those taken by the algorithm each cycle. Many portions of the algorithm is iterative, so some steps take place several times before a succeeding steps, but the general flow the of the algorithm should be sufficiently represented in the proceeding sections. The four step are followed by another section which introduces the algorithm in a more detailed break-down and introduces the associated mathematics.

3.1 Step 1: Form an Initial Guess

The first step of the algorithm is the formation of sets of initial guesses upon which the algorithm acts. In first of the five cycles this means creating stochastically generated trajectories, which satisfy the boundary conditions. Every subsequent cycle uses the results of the previous cycle as a base trajectory that it refines with a denser mesh before it is processed again.

The initial 1000 trajectories are produced by generating an equal number of randomly uniform points in a bounded five-dimensional space. These points are used as the parameterizing values for the states of the spacecraft at time $\frac{t_f - t_0}{2}$, excluding L. These values are then augmented with the initial and final values of each of the states, which takes the place of the parameterizing values at times $t = t_0$ and $\mathrm{t} = \mathrm{t_f}$ respectively. These three values per state are then used to form five cubic spline functions, describing an assumed time history of each state across the mission timeline.

L is excluded from the above process because it is the only of the six states whose dynamics are dominated by the gravitational force of the orbiting body. For instance, if a spacecraft has no propulsive force the first five of the orbital elements are equal to zero and the solar sail's dynamics are solely described by L. Even in the case where there is a non-gravitational force acting on the spacecraft, the behavior of L is still generally overwhelmed by the term describing

gravitational effects. If the time history of L was randomly determined in a manner similar to the other states, the result would be trajectories that required forces that rival gravity in magnitude. Of course, photonic forces is several orders of magnitude smaller than gravitation in this respect, so it is undesirable to use a random trajectory as an estimate of this state. Instead the intermediate parameterizing value is assumed to bisect the initial and final value of L, after which it is treated the same as the other states.

Table 1 below shows the initial value, final value, and bounds of each of the states used to create the initial guesses. With the exception of L, these bounds are intended to be wide enough to allow a diverse range of different trajectory possibilities, yet limited enough to omit obviously impossible trajectories. The bounds of p, f, g, h, and k were chosen such that they allowed the intermediate value to vary well outside of the interval defined by the initial and final values because these states are not linearly independent. Consequently, a change in one state often mandates a change in another; thus these bounds allow for the possibility that the state values deviate well away from boundary value interval to accommodate a change in other states. It should be noted that these bounds only serve to form initial trajectories, and that the algorithm is fully capable of ending at a result with states outside of these bounds. Both the upper and lower bounds of

L are set to 8.644 to force the initial guess to lie directly between the upper and lower bounds for the reason described in the previous paragraph.

The use of a cubic spline interpolant serves two major purposes. The first is obviously that of function parameterization, allowing the trajectories to be described as by a set of variables. Solving directly for trajectories can often create mathematically complex problems, while multivariable optimization is comparatively simpler. Furthermore, there exists a wide array commercial off-the-shelf multivariable optimizers available that are sufficient for the task. Thus by parameterizing the state functional, the problem is transcribed into a simpler form. The second purpose is to guarantee that the boundary conditions are always satisfied by any given candidate solution. Since a cubic spline can be defined by the points it passes through, the previously describe method of forming initial guesses will always define a trajectory that meets the boundary conditions.

Figure 7 below is an example to illustrate how a trajectory is formed. The red circles are the three values used to define the cubic spline interplant function, in blue. Notice that the function passes through each of the points that are used to define it. In the first cycle of algorithm only the second point allowed to vary, so the initial guesses only have a single degree of freedom. In later iterations, the mesh is refined introducing new ways to shape the interplant function. Figure 7 also shows how different interplant functions result from varying the second point

in the set. In all cases, the value of the function and its derivative at all points are defined by the values of the points that define it.

Table 1 - Boundary conditions and initial guess boundaries used to form the assumed, unrefined trajectories

	p (AU)	f	g	h	k	L (radians)
Upper Bound	1.5	0.5	0.5	0.05	0.05	8.644
Lower Bound	0.3	-0.5	-0.5	-0.05	-0.05	8.644
Initial Value	0.9997	-0.0037	0.0163	$-3.645 * 10^{-6}$	$-3.645 * 10^{-6}$	5.851
Final Value	0.7232	-0.0044	0.0051	0.0068	$.0288$	11.168

Initial guesses are no longer generated after the first cycle. As the mesh increases in density so does the number of values that parameterize the spacecraft's states, which makes it impractical to continue to try to span the entire dimensionality of the problem with randomly generated points. For the first cycle only the second value in each parameterization array is allowed to vary, so the initial guesses span

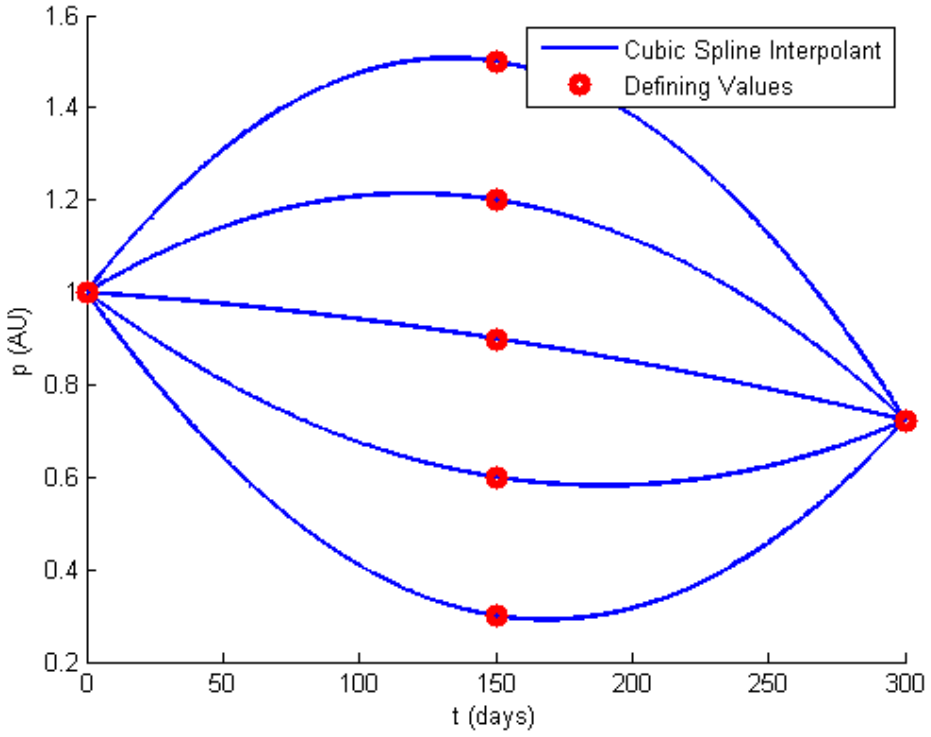

Figure 7 - Potential interpolant functions resulting from different parameterizing values

a 6 dimensional space. (one for each orbital element) In the second cycle the density of the mesh is doubled, meaning that the array population increases to five with two held constant to meet the boundary conditions. This translates to an 18 dimensional space, which would require more than 250,000 points to simply sample its corners. Consequently, every cycle after the first uses the results from the previous cycle its starting point.

3.2 Step 2: Evaluate Current Guess

The second step is the creation of an object that allows evaluation of the current guess. In practice this object is invoked repeatedly by the third step, but its function is distinct so it is assigned its own section. This evaluation consists of quantitating how well a trajectory "follows" the modeled system dynamics as measured by a multivariable optimization residual. When this residual is zero, or close to zero, the trajectory is physically realizable at all investigated times and therefore acts as a potential estimate to the boundary value problem. As the parameterization mesh becomes more dense, a near zero residual leads to an increasingly accurate answer, but this condition alone is not enough to guarantee that a trajectory is sufficiently accurate so verification is required.

The trajectory is evaluated using a collocation method in a manner similar to that used in the work of Fumenti et. al.[17] A "defect"Δ is defined as the difference between the time derivative of the assumed time history of the states, ψ, and the system dynamics in the form of the differential equation f evaluated at time t:

$$\Delta_t = \dot{\psi}(t) - f(\psi(t), \alpha, \delta, t) \qquad t \in [t_1, \dots, t_n] \tag{49}$$

When the defect is equal to zero at all points, the assumed trajectory has formed an estimate of the solution of the boundary value problem. In many ways, analyzing this defect is equivalent to inspecting the assumed trajectory at a set number of times across the mission timeframe and determining if the mission is

physically realizable. For instance, if there exists control angles α and δ that allows f to equal $\dot{\psi}$ for a given time, then the assumed state is changing in a way that is within the capability of the solar sail.

The algorithm makes a small departure from a traditional collocation method by using more investigation points than used to define the interpolant cubic spline. A cubic spline is usually composed of a number of linearly independent basis functions that is equal to number of conditions that the spline must meet. For instance, if three points define a cubic spline, then it is usually created from the sum of three independent basis functions to prevent the system form being over or under constrained. Similarly, collocation problems are often constructed such that there are an equal number of defects, (number of conditions) as points, which parameterize the assumed solution (number of basis functions). This means that the there are an equal number of unknowns as equations, and an estimate can be formed by setting all defects equal to zero.

However, this method provides little insight into the quality of the solution. When there are an equal number of equations and unknowns it is highly likely that there will always exist a solution that will set all defects to zero. This condition assures that the assumed solutions behaves in compliance with the system dynamics at the investigation points, but leaves no measure for the rest of the trajectory. It is entirely possible, and often the case, that this solution is unrealistic in between the

parameterization points and thus the value of the defects alone is insufficient to quantitatively describe the quality of the assumed trajectory.

In order to remedy this issue the number of inspection points is set greater than the number of parameterizing points. When this is the case the defects not only serves to define the assumed solution, they also act as a measure of quality for the solution by investigating between the parameterization points. This transforms the collocation method into a means to simultaneously create and evaluate an assumed solution.

The major downside to this modification is that this problem is over constrained. This, in turn, means that it is unlikely that a solution can be found that sets all defects equal to zero. Instead the focus is shifted to minimizing the defect, corresponding to the maximization of the trajectory quality. The absolute value of the defects are then treated as the objective of a traditional multivariable optimization problem, whose solution is found when this objective has been minimized.

3.3 Step 3: Improve Current Guess

The third general step of the algorithm is to improve the current guess by performing the optimization described in the previous step. MATLAB's FSOLVE function is used to minimize the defects of an assumed solution by adjusting its parameterizing values. The result is a trajectory that is an improvement over the

initial trajectory in the sense that it is more compliant with the equations of motion of a spacecraft.

The FSOLVE function is pre-programmed algorithm packaged with MATLAB. It uses a Levenberg-Marquardt method to attempt to optimize a non-square system by searching in a direction that falls somewhere between that of a traditional Gauss-Newton and steepest descent method.[18] It was chosen over other available pre-programmed MATLAB optimization algorithms because it is capable of minimizing a vector, (vector of defects) instead of a scalar, (sum of the defects) which ultimately led to improved performance.

In the actual implementation of this step FSOLVE acts upon the objective described in the second step so it is a bit misleading to separate this step from the other. FSOLVE approximates the derivative of the objective by perturbing the parameterization values to find the figurative downward direction and follows until it reaches a local minimum.

3.4 Step 4: Verify Current Guess

The final step of the algorithm is to verify the improved guess by extracting the control angles from the assumed trajectory and using them in conjunction with a time stepping algorithm. The time stepping algorithm integrates the equations of motions including the control angles to calculate how far the spacecraft is from the target at the end of the mission. The result of this verification step is taken to

be the "true solution" of the algorithm and used to determine if a sufficient trajectory has been found. The bulk of the verifier is a multistep predictor-corrector method called the Adams-Bashforth-Moulton method as presented by Mathews and Fink.[19]

The received radiation dosage of the sail is a parameter that receives special care during this step. As noted in the section Degrading Optical Parameters, the dosage is calculated by integrating the sun-sail distance and the sail control angles. As a result, it is itself changing with time as a function of the state and must be integrated alongside the state. A simple Riemann sum is used to estimate the dosage, which has a convenient property of allowing the verifier calculations to be explicit. Although there are certainly more accurate methods, it is assumed that this has little effect on the result due to the small step size.

Once all trajectories have passed through the algorithm five times, the remaining trajectory that is verified to have missed the target by the least is taken as the final result of the process. This overall process is then repeated 20 times, varying the correction timing.

3.5 Detailed Implementation.

This section describes the algorithm in greater detail by introducing a set of functions and describing the math upon which they are formed. Figure 8 is a simple flowchart that demonstrates the relationship between these functions,

noting the major inputs and outputs of each. This figure only captures part of the entire algorithm, excluding the verification block for brevity. When compiled together these functions perform the four steps described in the previous section

The algorithm begins by building a set of meshes that act as initial assumed trajectories. These meshes are created using a wide array of input parameters which generally fall into two categories: mission parameters and solver parameters. Mission parameters describe the problem that is being solved, and includes detail such as the arrival and departure destinations as well as the transit time. Solver parameters are variables that are solver specific, such as the density of the meshes and number of iterations to complete. These input parameters instruct the Mesh Generator block how to form the initial trajectories. The output of this block is $\boldsymbol{S}_{pi}$, which is a vector containing all values used to construct each of the six preliminary cubic spline.

This vector is passed to the FSOLVE function to serve as an initial starting point to the multivariable optimization. FSOLVE then acts upon the objective block by playing with these parameterization values to find the local minima of the residual, J, which is a vector constructed for all defects Δ from equation 49. When this minimum has been found to within a relative tolerance, in this case $1 * 10^{-12}$, the objective outputs the corresponding sail control angles to the Verifier block.

The impact of this tolerance is difficult to translate directly into something such as a distance or time. Each defect Δ corresponds with a different state at different times throughout the mission, which in turns effects the trajectory in different ways. Instead of trying to quantize the impact of the tolerance, its magnitude was chosen such that it was orders of magnitude smaller than required to have an appreciable impact on the trajectory.

The objective block is composed of five sequentially performed sub-functions: the Spline Interpolator, Ideal Force Calculator, Sail Model Inversion, System Dynamics and Residual Calculator routines. Combined, they compose the previously described collocation method and is used to evaluate the quality of the trajectory. The objective block accepts the current parameterization of the state, $\boldsymbol{S}_p$, and outputs both a set of solar sail control angles and a residual J, which acts again acts as a measure of trajectory quality.

The spline interpolator function bridges the parameterized states, $\boldsymbol{S}_p$, to their functional representation by using them to form six cubic spline functions. These functions are interpolated to find the value of the states, as well their derivatives, at the inspection points. These two new vectors are labeled ψ and $\dot{\psi}$ respectively in the flowchart and are used as inputs for almost all proceeding subroutines. The cubic spline is created and evaluated using MATLAB's SPLINE function, while the derivative is estimated using the finite difference approximation:

$$\dot{\psi}(t) = \frac{\psi(t+\epsilon) - \psi(t)}{\epsilon} \quad \textbf{50}$$

where ϵ is a perturbation that is taken to be 10^{-8}. Although this equation only gives a first-order approximation, it was chosen because it only required one additional evaluate of the cubic spline.

There are two solver parameters that are convenient to describe at this time. This first, n_{coes}, is an integer value which describes how many linearly spaced points parameterize each state within $\boldsymbol{S}_p$. The second value, n_{invest}, describes how linearly spaced points are used to investigate the trajectory with the residual. Although any ratio of n_{coes} to n_{invest} can be used, n_{invest} is chosen such that

$$n_{invest} = 2n_{coes} - 1 \quad \textbf{51}$$

for two reasons. First, this guarantees that n_{invest} is always greater than n_{coes} so that the defect can be used to measure the quality of the trajectory as described in the previous section. The second purpose is that this ratio forces roughly half of the evaluation points to line up exactly with the parameterization points, while the other half lines up exactly midway between these points. The greatest control over the shape of the assumed cubic spline is achieved at the parameterization points so it is desirable to ensure that the evaluation of the function does not only take place where the function is well controlled.

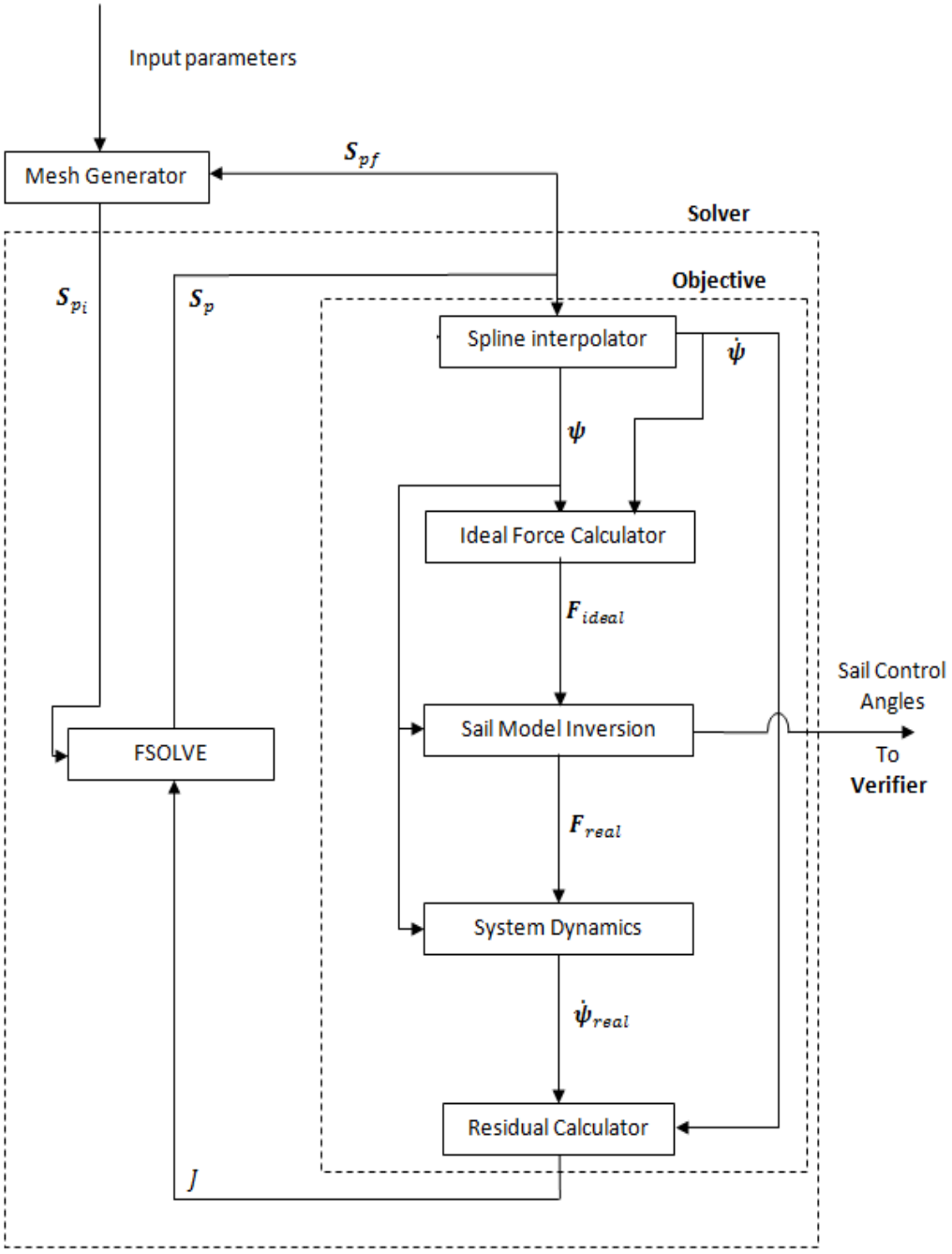

Figure 8 - A flowchart diagramming the flow of data between the functions that compose the first half of the algorithm.

The Ideal Force Calculation block takes both ψ and $\dot{\psi}$ and attempts to calculate the forces required to make the assumed trajectory physically realizable at the inspection points. To do this, it factors the equations of motion, (equations 41-48) such that they take the form

$$f = \boldsymbol{A}F + b \tag{52}$$

where $f = \left[\dot{p}\dot{f}\dot{g}\dot{h}k\ \dot{L}\right]^T$, $F = [F_r F_t F_h]^T$, $b = \left[0\ 0\ 0\ 0\ 0\ \sqrt{\frac{p}{\mu}}\left(\frac{q}{p}\right)^2\right]^T$ and $\boldsymbol{A}$ is the matrix that completes the set. When this equation is evaluated at a particular inspection point, then $f(t_i) = \dot{\psi}(t_i)$ and $\boldsymbol{A}$ is a function of ψ_i. It then follows that a force can be calculated using the equation

$$F_{ideal} = \boldsymbol{A_i}^+ * \left(\dot{\psi}_i - b\right) \tag{53}$$

where $\boldsymbol{A}^+$ is the psuedoinverse of the matrix $\boldsymbol{A}$. There is, however, no guarantee that the calculated force F_{ideal} is physically realizable by the sail; it is simply the least-squares solution to equation 53. This is the result of a mismatch between the number of states which must be controlled, six, and the number of controls available, two. Equation 53 is therefore over-determined and an F_{ideal} may not exist that allows the assumed ψ and $\dot{\psi}$ to coexist. For the mean time, this issue is ignored and later mitigated by the following blocks.

The Sail Model Inversion block takes an F_{ideal} and finds the closest force that the solar sail could actually produce at the assumed state ψ_i. This result is a new force, label F_{real}, and is calculated such that

$$\frac{\mathrm{F_{r_{ideal}}}}{\|F_{ideal}\|} = \frac{F_{r_{real}}}{\|F_{real}\|} \quad \textbf{54}$$

as described in the section 2.2.3 Mapping Force to Angles.

The System Dynamics function takes the assumed state ψ and applies the newly calculated forces F_{real} using the equations of motion. The result is a new state derivative vector, $\dot{\psi}_{real}$. If the solar sail is capable of creating the force required to follow the assumed trajectory at all inspection points, ψ and ψ_{real} are then equal. If these vectors are not equal then some combination of the following three conditions must be true:

1. The sail cannot create a force with the appropriate magnitude.
2. The sail cannot create a force in the necessary direction.
3. There does not exist a 3-dimensional F_{ideal} that would result in the desired six-dimensional $\dot{\psi}_i$.

However, these problems are mitigated when the difference in the two vectors is treated as the objective of the FSOLVE function. As it operates, the assumed trajectory becomes more "realistic" as the optimization routine minimizes the

difference in the two vector and therefore eliminate the three conditions above if the value of the minima is sufficiently small

The final function in the objective is referred to as the Residual Calculator and forms the vector of defects by subtracting ψ and ψ_{real}. The resultant vector, J, is then treated as the objective vector FSOLVE attempts to minimize as described above. Once this minimization has completed for all 1000 initial trajectories the corresponding states parameterizations $S_{\square_f}$ are sent back to the Mesh Generation function. It deletes any redundant trajectories, leaving one trajectory per minima. The remaining state parameterizations are then interpolated using another cubic spline in order to increase the density of the mesh. These new vectors once again become S_{pi}; the assumed trajectories in the following run of the algorithm.

At the same time, the solar sail control angles are extracted at the investigation points by the Sail Model Inversion block and Passed to the Verifier function.

The bulk of the verifier is a multistep predictor-corrector method called the Adams-Bashforth-Moulton method as presented by Mathews and Fink.[19] This method is a numerical integration technique that estimates a function value U at time t_{n_1} using the value of U at previous time steps. To perform the estimation, a Lagrange polynomial approximation p is calculated at t_{n_1} using the equation

$$p_{n+1} = U_n + \frac{k}{24}\left(-9f(U_{n-3}, t_{n-3}) + 37f(U_{n-2}, t_{n-2}) - 59f(U_{n-1}, t_{n-1}) + 55f(U_n, t_n)\right)$$

55

where k is the interval between time steps and f is the function's derivative with respect to time. This predictor, as it is called, is then used in a second Lagrange polynomial approximation to estimate the function at the proceeding time step

$$U_{n+1} = U_n + \frac{k}{24}\left(f(U_{n-2}, t_{n-2}) - 5f(U_{n-1}, t_{n-1}) + 19f(U_n, t_n) + 9f(p_{n+1}, t_{n+1})\right)$$

56

The difference between the prediction and correction is used to estimate the error over the time step as follows[19]

$$\text{Error} \approx -\frac{19}{270}(U_{n+1} - p_{n+1})$$

57

Although this error can be used to determine if time interval k is too large or too small for a desired relative error, it was found to be more time-efficient to set the step size to 10^{-5} and use equation 57 to verify that this step size was insignificant compared to other sources of error.

The Adams-Bashforth-Moulton method is not self-starting. In order to estimate the function at time t_{n+1}, function evaluations are required at the previous four time steps. These evaluations are unavailable for the initial time step, so a simple 4th order single-step Runge-Kutta scheme is used for the first three steps. The scheme is given by LeVeque and described by the following equations.[20]

$$Y_1 = U_n \quad \textbf{58}$$

$$Y_2 = U_n + \frac{1}{2}kf(Y_1, t_n) \quad \textbf{59}$$

$$Y_3 = U_n + \frac{1}{2}kf(Y_2, t_n + \frac{k}{2}) \quad \textbf{60}$$

$$Y_4 = U_n + kf(Y_2, t_n + \frac{k}{2}) \quad \textbf{61}$$

$$U_{n+1} = U_n + \frac{k}{6}\ldots * \left[f(Y_1, t_n) + 2f\left(Y_2, t_n + \frac{k}{2}\right) + 2f\left(Y_3, t_n + \frac{k}{2}\right)\ldots + f(Y_4, t_n + k)\right] \quad \textbf{62}$$

In the context of this work, f refers to the differential equations 41-46, making U the resulting estimation of its integral and state of the solar sail.

4. ANALYSIS

The models and algorithm described in the previous section were used to calculate point to point, non-ideal solar sail trajectories. In order to make sense of the resulting raw data, three forms of analysis took place: filtering, linear regression and statistical error analysis. Each of the these are addressed individually in the following section.

Before delving straight into the analysis, some time should be spent describing the raw data and how it was gathered. For any given set of degradation parameters, a base trajectory was first calculated from Earth to Venus using the "assumed" degradation profile. Control angles were extracted from this trajectory, and sent to the time stepping algorithm which calculated a new trajectory using these control angles along with the "true" degradation parameters. This new trajectory was considered the "true" trajectory the solar sail would follow using the calculated control angles. Needless to say, the true base trajectory always missed Venus to a degree largely dependent on the difference between the assumed and true degradation parameters.

Twenty more trajectories were then calculated, each of which started at points linearly spaced along the base trajectory. For this step, the algorithm was allowed knowledge of the true degradation parameters under the assumption that enough time had elapsed in the base trajectory to allow for the true degradation to be

determined. This set of trajectories simulated twenty different potential corrections, each of which attempted to reach Venus by the end the mission.

Recall that the trajectory-solving algorithm reports data from the trajectory which resulted in the smallest distance from Venus at the end of the simulation. This distance, henceforth called the Final Positional Error (FPE), acted as a figure of merit for the quality of correction trajectories. If the FPE of an attempted correction was less than .029 AU (the expected value of the error produced by the algorithm) then it was considered likely that an attempted correction would ensure arrival at Venus. On the other hand, if the FPE was greater than .059 AU (the expected value plus two standard deviations) it was considered highly unlikely that an attempted correction would be sufficient to ensure arrival. In comparison, .029 AU is about 7 times the sphere of influence of Venus.

For the purposes of this thesis, a run was defined as a set twenty trajectories that attempted to correct the same base trajectory. These trajectories only differed by the date at which the correction begins, so it is a natural grouping for analyzing the impact of correction timing. This grouping also enables filtering of the raw FPE data to ensure that each run is sequentially monotonic; an expected behavior which is used to help discern perceived trends from numerical error.

4.1 Monotonic Behavior

The FPEs of each run is expected to be monotonically increasing with respect to time. To help illustrate why, consider a correction trajectory starting on day 200 that corresponds to an FPE of .10 AU, and another starting at day 220 with an FPE of .03 AU. In this example a later correction results in less positional error than an earlier correction. It may not be apparent, but this indicates that the algorithm failed to find a potential trajectory in the first scenario: the trajectory of the second case. There always exists the option to delay a correction until a later date, such as waiting twenty days and performing the .03 AU FPE correction. Thus, it can be concluded that the smallest possible FPE of the first case is actually, at maximum, .03 AU. (that of the second case) This knowledge is then extended to an entire run; the FPE of any given trajectory must be, at maximum, equal to the lowest of all subsequent runs.

Unfortunately unprocessed runs are rarely monotonic. Although it is expected that the underlying function will follow this behavior, error from a number of different sources introduces noise into the results. This error is unavoidable, but non-monotonic points can be partially filtered from the results by disqualifying data that breaks this behavior. Fortunately, since it is known that the underlying behavior is monotonic, these points lend themselves to quantify numeric error introduced by the algorithm.

The filtering itself is very simple. Since the function must be monotonic, each FPE must be less than or equal to every FPE that occurs afterwards. If any point has a greater FPE than any subsequent point, the first point is re-assigned the lesser of the two values. This ensures that the post-filter data removes error and improves the quality of the data.

4.2 Linear Regression

As expected each run appears to be characterized two different regions. In the first region the data is characterized by what seems to be random error around a constant value that is within the predicted error threshold. This region is assumed to represent the range of correction timings that allow the spacecraft to arrive at Venus. This is also the region that displays the majority of the non-monotonic behavior, so it is used to estimate error.

The second region, on the other hand, is generally monotonic. It contains data that mostly exceeds the expected error, so this is the region in which a sufficient correction trajectory doesn't exist. This region is usually well summarized with a linear regression function, so a line is fit to this region. This line serves to represent an individual run when comparing multiple on the same plot, as well as estimating the earliest a correction could be required.

This two-region behavior is easily explained. For instance, one would expect that there are many different possible corrections that could be made early in the

mission that would allow for arrival at Venus. On the other hand, it also makes sense that there should exist a time after which there is no possible way a correction can take place in the remaining timeframe. These two cases are separated by a natural boundary, perhaps a "correction deadline," and it is this date that this work seeks to find.

To achieve this end, a linear regression is fit directly to the data to help estimate the earliest a correction could be required. This function is fit directly to the data, as opposed to the region bound by the error estimate, because the error only acts in the positive FPE direction. By fitting the line to the data, it estimates both the underlying behavior and the error, which serves to overestimate the urgency of a correction. It is expected that the x-intercept then creates a margin of time where a correction could potentially be successful, but unadvisable.

4.3 Solver Error Estimate

Since data is useless without an idea of its accuracy, a statistical analysis was performed to estimate the error introduced by the algorithm. Due to the complexity of the algorithm, it was not particularly clear how an analytic approach would manifest so simple statistics were used in its place. The resulting estimates were performed in such a way that it intentionally overestimated the potential error, adding further confidence in the conclusions made by this work.

The remainder of this section is broken into three parts. The first describes the sources of error and how they appear in the data. The second part outlines the strategy and reasoning behind the statistics. The third and final part presents the mathematics used to estimate the error.

4.3.1 Error Sources

There several different ways that error makes its way into the raw data, all of which fall into one of two categories. The first category is called solver error and stems from the methods used to calculate trajectories. The secondary category, verifier error, is numeric error that results from the time stepping algorithm.

These two categories are distinct in the way they impact the data. Solver error causes the FPE of a correction trajectory to be artificially large and acts in the positive direction. Verifier error, on the other hand, can act in either the positive or negative direction. This means that solver error increases the apparent urgency of a correction by shifting the linear regression function in the positive directive, while impact of verifier error is uncertain. Fortunately, it was computationally cheap to marginalize the magnitude of the verifier error so it had no impact on the conclusions of this work.

4.3.1.1 Solver Error

4.3.1.1.1 Global Minima

As described in the algorithm section, solar sail trajectories were calculated by minimizing an artificially constructed objective function. An advantage of this approach was that the objective was a relative measure of the quality of the trajectory - so minimizing the objective corresponded to a relatively maximized trajectory quality. A disadvantage of this approach was that best estimate to the boundary value problem corresponded to the objective's global minima, which is notoriously difficult to locate. Moreover, even if the global optima had been stumbled upon by luck alone it would be impossible to know with certainty that t had been found. Thus the objective served as a tool to improve the trajectory estimates, but could not be used to guarantee a perfect trajectory without unlimited resources.

To work with this disadvantage the algorithm started with a large number of stochastically generated initial guesses. These guesses acted as a survey across a bounded domain of the objective, improving the odds that the global minima was found. However, this couldn't guarantee success in all cases, and when it failed its impact was fairly obvious. For instance, examine Figure 9; all data points from about 150 - 225 days have a value of about 0.03 AU. Except, of course, the third data point, which has relatively high value of 0.08 AU. This point is clearly in

error because the function must be monotonic, and is isolated to single point. This kind of error is called global minima error.

Global minima error has two defining characteristics. The first is that it usually appears randomly among the data. Since the initial guesses are stochastically generated, whether or not the global minima is found is a matter of probability. Thus it is not expected that this error appears on some data points, but only in the cases where the initial survey was insufficient. The second characteristic is that it is easy to identify by running the algorithm again. Every time the algorithm runs it re-generates a new set of initial guesses, decreasing the odds that global minima is missed again. If a data point has a different value after a second run, it can be concluded that the greater of the two is being influenced by global minima error.

Global minima error is very difficult to quantify by itself because it always acts in the positive direction: the same direction as discretization and interpolation error. Without an exact knowledge of one the other cannot be isolated. Instead, their impacts are quantified together to create bounding region for the underlying behavior.

4.3.1.1.2 Discretization and Interpolation Error

The second category of error is called discretization and interpolation error. It stems from approximations and conversions to and from discreet time space. These simplifications make the boundary value solvable but introduce a

systematic error into the result. The magnitude of this error is manageable by manipulating the step size of the simulation, but is impossible to remove altogether.

Below is a list of the most significant ways this error is generated. It is not comprehensive:

1. The discrete-time estimate of the trajectory must be interpolated into continuous-time to yield a trajectory. A cubic spline interpolant is used to serve this purpose, so any dynamics greater than third order are truncated.
2. The resulting control angles from the estimate are also based in discrete time. A zero-order hold is used to interpolate these angles. A zero order hold was chosen over other interpolant function because it proved to be more accurate when the control angles were near the bounds of their domains.
3. The process of converting an ideal force into a set of control angles uses a bisection method to solve for the cone angle. The accuracy of the method was set to 10^{-2} degrees for performance, so the angles are not exact.
4. The degradation must be integrated along-side the state of the trajectory. Error is added because Riemann sum is used in the place of an exact analytic integration.

5. The objective is numerically optimized. The tolerance of the optimizing function was set to 10^{-12} because the resulting trajectories proved to be very sensitive to the exact value of the objective function. Again, this non-perfect estimate contributes to the overall error in the trajectory.

All of these source, among others, combine to form what is referred to as discretization and interpolation error. This error is usually characterized by influencing every data point, shifting the raw data upward. The magnitude of this error grows with the size of the time step in the simulation, so the error is greater for longer trajectories. Consequently, this means that corrections that take place earlier in the mission are expected to experience greater error than those at the end.

4.3.1.2 Verifier Error

Verifier error is the error that comes from the verification step of the algorithm. Like all linear multistep numerical integrators, the Adams-Bashforth-Moulton method has an inherent local truncation error. This error can be easily shown to be on the order of h^5, where h is the step sized used for the integration. It was for this reason that a step size of 10^{-5} was used this error is henceforth considered to be negligible.

4.3.2 Solver Error Analysis

The error analysis began by isolating the data in the region of the results where it was presumed that the sailcraft could reach its destination after an orbital correction. The data in this region, labeled "Region 1" in Figure 9, would have an FPE of zero if the algorithm was perfect. Any non-zero FPE in this reason was therefore attributed as solver error and used to estimate the error in both region. Region 1 was generally characterized by a constant value with what appeared to be random noise. This was distinctly different than the other region, labeled "Region 2" in the figure, in which the data was characterized by a clear positively-sloped trend and monotonic data pre-processing.

The error was calculated in first region and was applied to both regions. Since the step size was smaller in the second region than the first, the error was likewise expected to be smaller in this region. This means that generated bounds are an over estimation of error in the area where the linear regression is applied. This, in effect, created a conservative approximation to the true nature of the of the underlying function by suggesting that a correction is more urgent than the raw data itself would suggest.

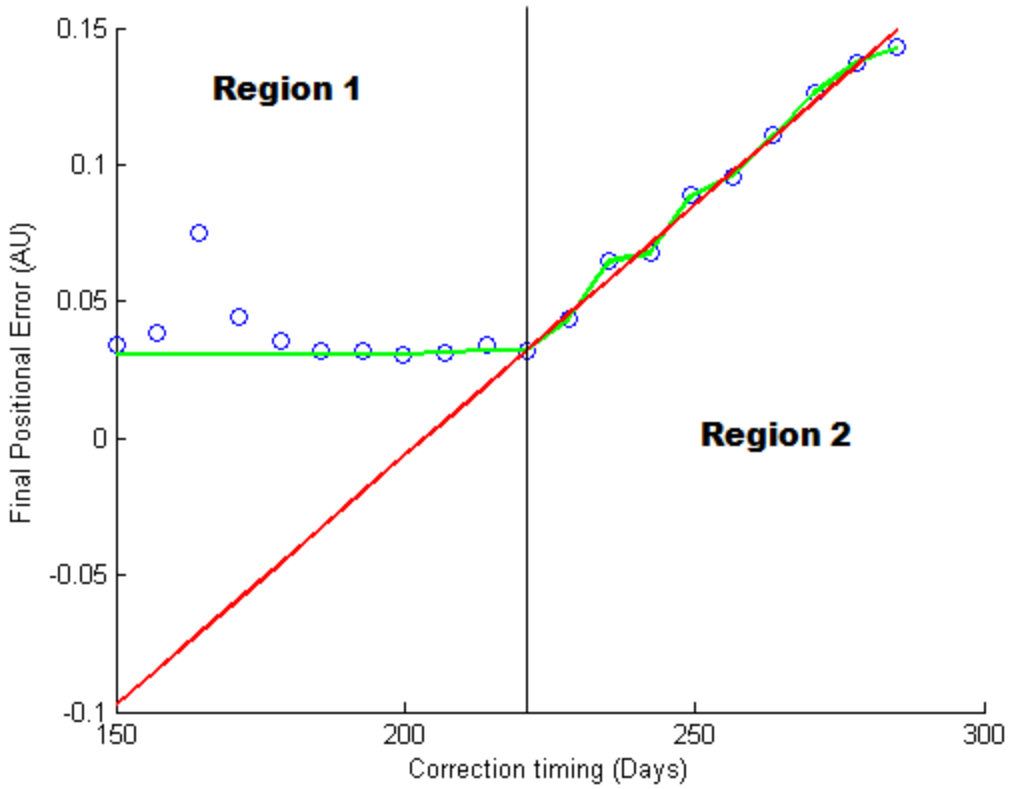

Figure 9 - A data set that exemplifies the two different observed behaviors

Once isolated, the error was assumed to be lognormally, randomly distributed. A lognormal distribution was a natural choice because both the population and the distribution are always positive. Figure 10 compares the two distributions; μ is the expected value of population while σ is the standard deviation of the normal or equivalent normal distributions, respectively.

Solver error was expected to always act in the positive direction. To understand why, imagine a situation where the solar sail has waited too long to correct and the closest it could approach Venus is with an FPE of 0.5 AU. It is impossible for the algorithm to find a set of angles, even in error, that would allow the solar sail to approach closer than 0.5 AU. Yet this is exactly what negative error implies. Just as negative distance does not make sense physically, neither does negative solver error. The algorithm can incorrectly calculate the angles necessary to lead the solar sail to with 0.5 AU FPE, making it miss by a greater distance, but never less. To incorporate this logic, error bars were added to the data in the negative FPE direction. Two bars were added to each data point; one that decreases the value of the data by the expected value μ, and another that decreases it by the expected value plus two standard deviations $\mu + 2\sigma$.

Table 2 below shows the calculated relevant statistical parameters for the lognormal distribution.

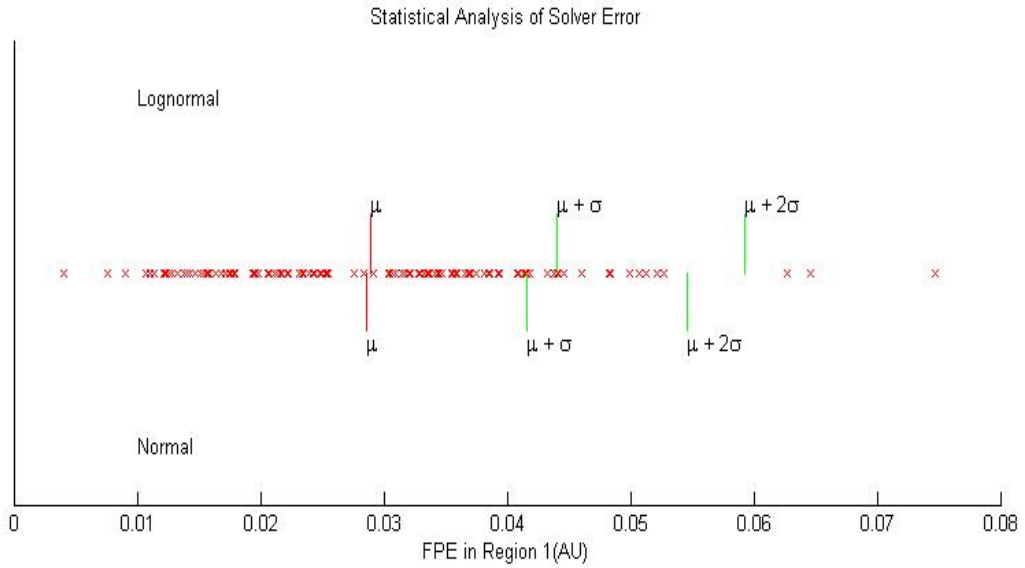

Figure 10 - Comparison of normal and lognormal parameter estimation for a population of solver error in region 1

The analysis was ran for the same trajectory three different time in an attempt to isolate non-stochastic error sources. These results are compared against the compilation of the three, which is comprised by the data points corresponding to the lowest FPE at each correction date. The results can be found in Table 3, and they suggest that the expected error decreases by almost half when the number of initial guesses is tripled.

Table 2 - Lognormal parameters for the population consisting of the FPEs in region 1

Population size - n	Expected Value - $\mu(AU)$	Variance - $\sigma^2(AU^2)$
144	.0289	0.00023

Table 3 - Comparison of the statistical estimations of the solver error from three different trials with the same parameters

Trial	Population size - n	Expected Value - $\mu(AU)$	Standard Deviation - $\sigma(AU)$
1	15	0.0194	0.0068
2	15	0.0262	0.0084
3	15	0.0276	0.0121
1+2+3	15	0.0167	0.0043

4.3.3 Equations

The mean and variance of the population is calculated as described by Shimizu.[21]

First, the population X is transformed by taking the natural logarithm, i.e.

$$Y = ln(X) \quad \textbf{63}$$

Since X is assumed to be lognormally distributed, Y is then a normal distribution with mean m and variance v, which are calculated as follows:

$$m = \frac{\sum_{i=1}^{n} Y_i}{n} \quad \textbf{64}$$

$$v = \sum_{i=1}^{n} (Y_i - m)^2 \quad \textbf{65}$$

The mean and variance of X then can be calculated as a function of m and v:

$$\mu = (1-\delta)e^{(m+\frac{v}{2})} \quad \textbf{66}$$

$$\sigma^2 = (1-\delta)\left(e^{v} - (1-\delta)\right)e^{2\mu+v} \quad \textbf{67}$$

where δ is taken to be in order to give the standard variance.

4.4 Processing Example

Figure 11 below is an example of a pre-processed run. The blue circles each represent the trajectory that passed through the verifier with smallest FPE among all trajectories tested for a given correction timing. There is a clear difference in behavior before and after the 250 day mark, which separates regions one and two. Whereas correcting before 250 days leads trajectories that end within an FPE of about .01 to .02 AU, correcting after leads to a noticeably larger FPE. The second region also displays the expected monotonic behavior.

Since the entire underlying function must be monotonic, any point that has an FPE value greater than any other point that succeeds it chronologically is filtered out from the result. This helps remove some numerical error generated by the solver, although it certainly cannot remove it completely. The resulting function, represented by the green line in Figure 12, acts as a better estimate for the underlying function than the pre-processed data.

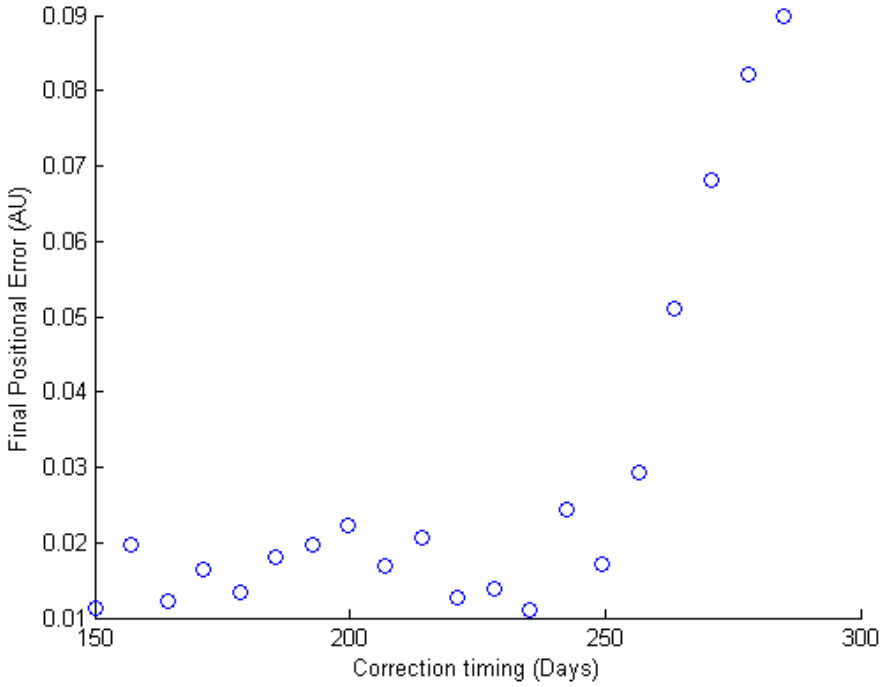

Figure 11 - A run to serve as an example run pre-processing. Notice that the 150-250 day region is non-monotonic and appears characteristic of noise.

The filtered data in Figure 12 (250 days onward) appears largely linear; a trend which that is observed in most of the trials. A line is fit to this region and is used in succeeding sections to compare results between runs. This line is generated by satisfying two conditions: the line must pass first pass through the last point that generally falls within the expected numerical error of the solver. The line also must have a slope that minimizes the least-squares error from points in region 2.

Finally, error bars are added to the plot, which can be observed in Figure 14. This helps define the end of the region where it is unclear if a correction is necessary

because of the precision of the solver. This particular example predicts that a sufficient correction can be made any time before 240 days into the mission. A correction after about 270 days is exceedingly likely to miss. The region between 240 and 270 days could be called a metaphorical "danger zone," where there is a significant risk of a failed correction.

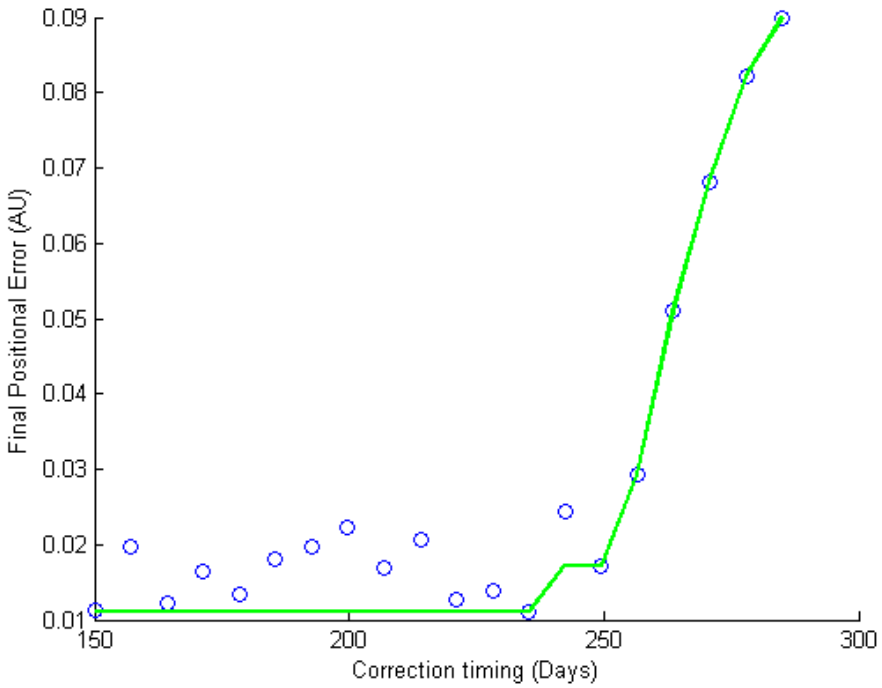

Figure 12 - The example run overlaid with the monotonically filtered results in green. This line represents a best guess at the underlying function.

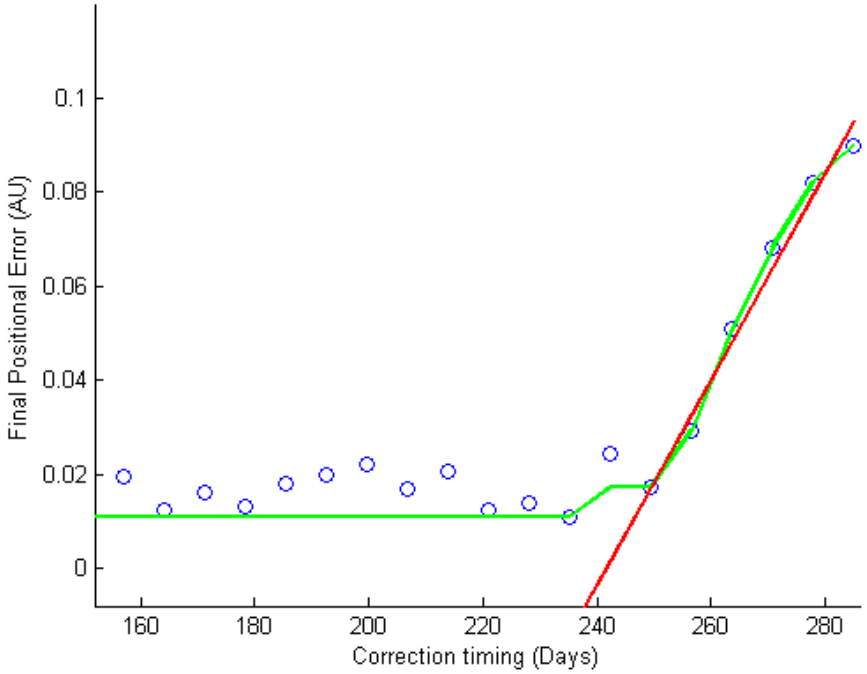

Figure 13 - A linear regression function is generated (red) for the 250+ day region and used as an estimate of this region when comparing results.

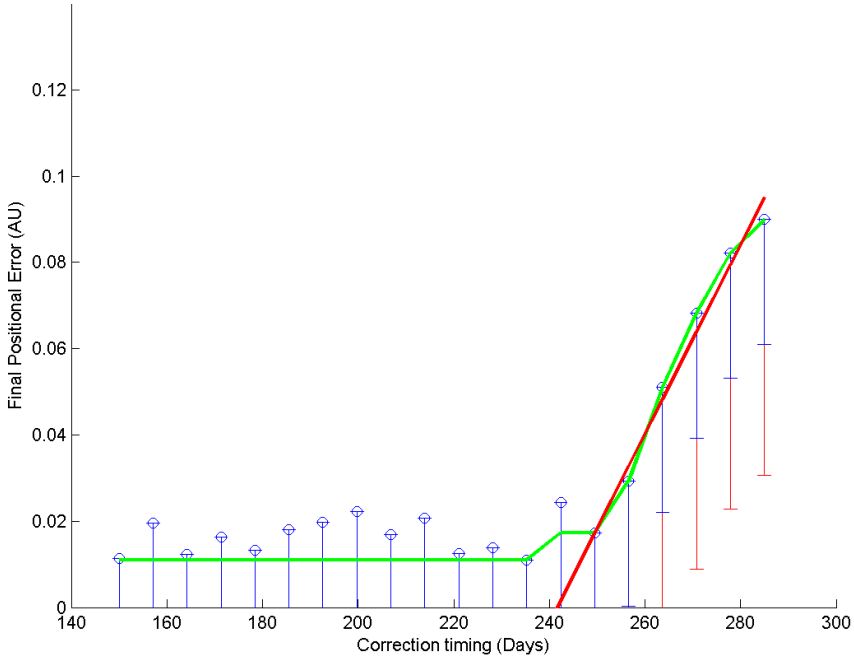

Figure 14 - The example run with error bars. The blue bar shows the expected numerical error, while the red adds two standard deviations.

5. RESULTS

The algorithms and models described in the previous sections have been used to investigate the potential of using a single model update to mitigate the influence of uncertain optical degradation in a sample Earth-Venus transfer. The simulation begins at Earth on August 27th, 2003 with an anticipated arrival at Venus 300 days later. It is assumed that the spacecraft starts with zero excess hyperbolic escape velocity and that a satisfactory trajectory likewise arrives at the appropriate planet with zero hyperbolic excess energy. A full list of mission parameters can be found in Appendix A.

This investigation focused on the relationship between three variables and their impact on the Final Positional Error (FPE) of the sailcraft. The FPE was defined as the distance between the target and the sail at the end of the 300 day transit and as a measure of quality for each trajectory found by the solver. If an FPE was within 0.03 AU of Venus, the expected numerical error from of the algorithm, it was assumed that the solar sail was capable of "hitting" the target. On the other hand, if an FPE was greater than this expected numerical error it indicated that the solver was incapable of finding a satisfactory trajectory. When this was the case for several sequential points, it was assumed that the sail was incapable performing a trajectory that arrived at the destination.

Figures 15 and 16 help illustrate the difference between trajectories that were considered correctable and those that were not. The green line represents the base trajectory, which fails to arrival at Venus due to the misestimated optical degradation. The blue and red lines represent calculated and simulated correction respectively. The black spheres have a radius equal to the anticipated solver error, used to distinguish between sufficient and insufficient trajectories. In Figure 15, the difference between the assumed and simulated degradation parameters was small and the correction took place relatively early mission - day 116. In this example, the simulated correction ended with an FPE of slightly less than .03 AU and was thus considered a "hit." In Figure 16, on the other hand, the difference between the assumed degradation parameters was much larger - and the attempted correction took place much later into the mission. The result was a simulated correction trajectory which varied imperceptibly from the base trajectory, ending well outside the FPE threshold. In this case it was concluded that the simulated correction was insufficient for arrival at Venus.

The main independent variable is the correction timing of the model update: the date on which the correction trajectory begins. A base trajectory is initially created using an assumed set of optical degradation parameters, which is incorrect to some degree. Until the correction date the sailcraft is controlled using the cone and clock angle history of the base trajectory, which causes it to veer off course.

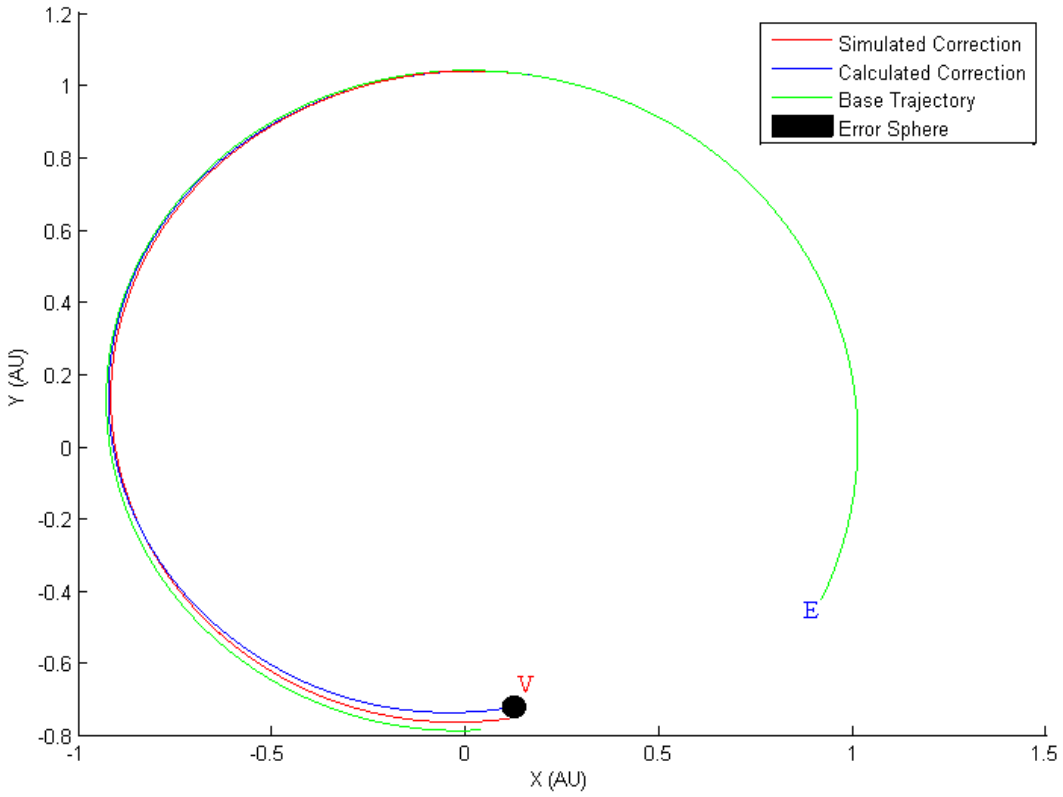

Figure 15 - A simulated correction that resulted in a Venus Arrival

After the correction date the solver attempts to create recovery trajectories with the correct set of optical degradation parameters. The correction timing ranges between 20 linearly spaced points between 50% and 95% of the total transit time.

The second and third independent variables are the end-of-life degradation factor, d, and half-life degradation rate, λ. As discussed in the introduction, these variables define the extent and rate of degradation, respectively. More technically, d is the percentage change in relevant optical properties while λ defines the radiation dosage required for half of the degradation to take place.

Each of these variables has two values: one assumed by the solver before the correction and one after. The difference between the two values simulates the uncertainty in the optical degradation. The verifier always uses the correct value for both parameters.

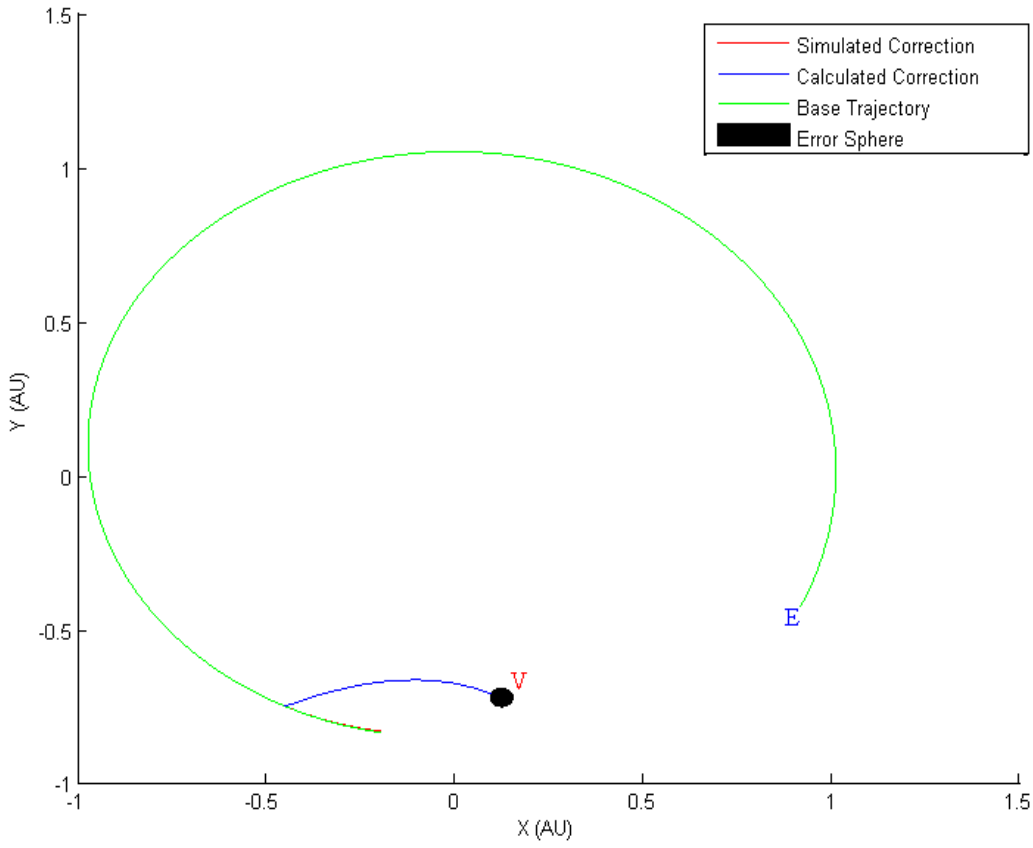

Figure 16 - A simulated correction that did not result in a Venus Arrival

The runs can be fit into three different groups. For the first six trajectories, d is the uncertain variable so λ is held constant. In the following two runs, λ is varied

while d is held constant. The final run is a worst-case scenario where both variables are assumed exceedingly poorly.

5.1 Processed Data

This section contains all 9 sets of data after processing. Each plot contains three lines, each of which is linear regression function described in the previous section. Each line is accompanied by a two-part annotation that indicates the assumed and correct values for the two degradation parameters. For instance, the annotation $d = .0 - .2 \quad \lambda = 1 - 1$ indicates that the initial, pre-corrected trajectory was calculated using the values: $d = 0$ and $\lambda = 1$. However, the true values of these parameters are $d = .2$ and $\lambda = 1$, indicating that λ was assumed correctly while d was not.

Each run can be viewed individually in Appendix B.

5.1.1 Degradation Limit - Part 1

The first set of runs focused on a range of different degradation factors with a degradation time constant of one. This constant was just below the 1.386 value used by Dachwald et. al.,[12] meaning that this particular simulation assumed that the sail degraded slightly slower than assumed in the other work. Three runs were generated, each assuming that the degradation factors was more severe in reality than initially assumed. The results are graphically summarized in Figure 17.

This case is representative of the situation where the degradation factor is uncertain but the degradation half-life radiation dosage is known. The earliest correction is required by most miss-predicted degradation factor: $d = .0 - .2$. This situation is representative what could happen if degradation factors were totally ignored in the trajectory planning stages and the true optical degradation happened to be the severe among the considered values. Even in this case it is anticipated that the entire first half of the mission could be used to gather orbital data and to re-plan the trajectory.

The other two runs, where d is assumed 10% lower than the actual value, require around half of the time for a correction compared to the first run. Between the two cases, the run with the higher d range is projected to require slightly more time to correct than the other, but this is smallest difference between the three cases.

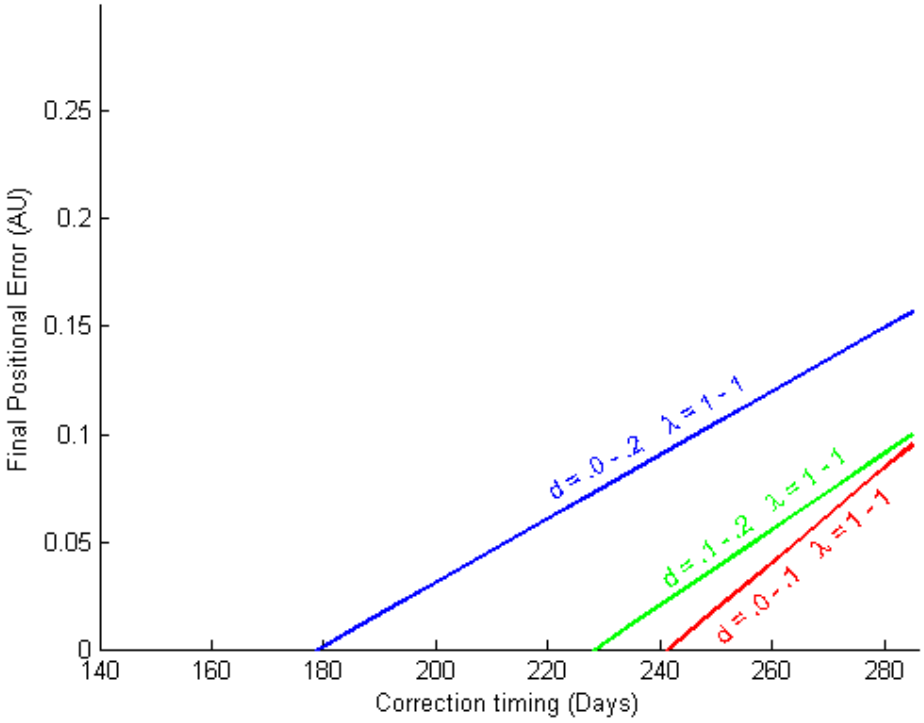

Figure 17 - The impact of an uncertain degradation factor (d) when $\lambda = 1$

5.1.2 Degradation Limit - Part 2

The second set of runs mirrored the first except λ was doubled. This is analogous to the case where half-life radiation dose was half of the previous three runs, which equates to a faster sail degradation than that considered by Dachwald et. al.[12] Even in this instance the worst case scenario allowed for a sufficient correction if the trajectory is corrected before 150 days. All three runs can be found in Figure 18.

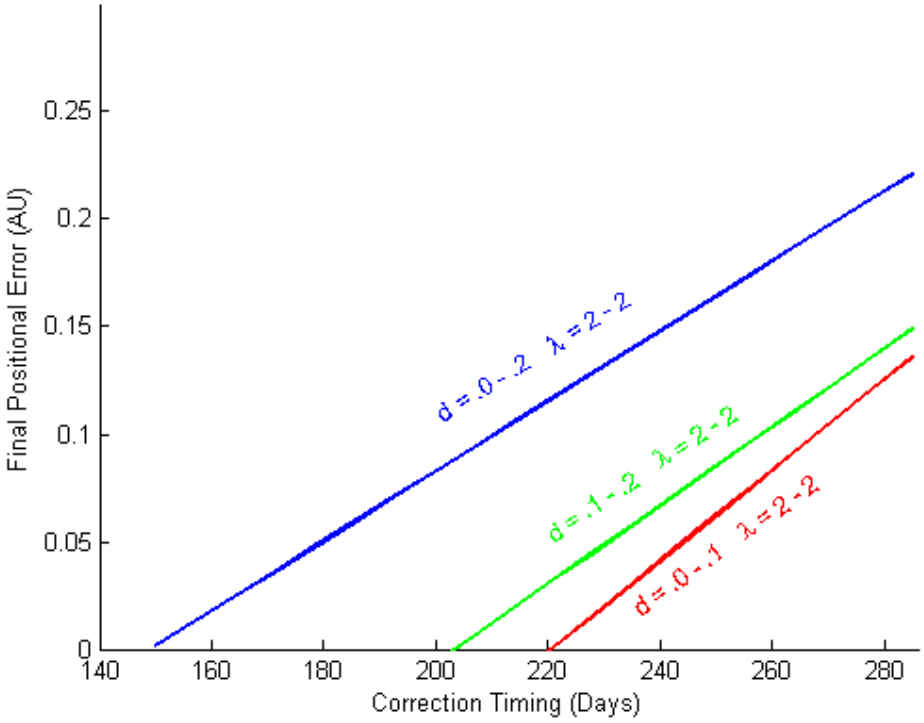

Figure 18 - The impact of an uncertain degradation factor (d) when λ= 2

As with the first three runs, it can be observed that assuming a degradation factor that is within 10% of the true factor affords additional time before a correction is necessary. Furthermore, a real interplanetary mission would be expected to be able to predict this factor with much greater accuracy than 10%, so it is conceivable that a model correction could be postponed well beyond 220 days into the mission without risking missing the target.

This result makes sense. The better the initial knowledge of sail's expected degradation, the more accurate the base trajectory is expected to be. With less error to correct for, the correction becomes less urgent because the sail requires less time to accumulate the necessary angular momentum.

5.1.3 Degradation Rate

The third set of runs held the degradation factor constant while lambda was allowed to vary. This is representative of the case where the dosage required to make the sail degrade is unknown. This acts as a simulation for the more feasible scenario where the required dosage is known, perhaps from laboratory experiments, but the exact flux from the sun is uncertain. Since solar events are usually unpredictable no amount of on-flight data will be able to fully mitigate the effects of this uncertainly.

Fortunately, these results seem to suggest that an uncertain $\boldsymbol{\lambda}$ is far less influential than an uncertain d. In both of the runs below, the blue and red runs in Figure 19, the final uncorrected trajectory yielded a FPE error that was within the solver error threshold. This means that it can't be certain that any correction is even necessary at all; these results fall within the precision of the solver. The behavior below is presented as a worst case-scenario, fitted in a manner similar to that described in a previous section. This provides an estimate of the earliest a correction could be necessary, if it all. Numerical error prevents making a solid

conclusion about the influence of an uncertain degradation rate, except that it is far less influential on the FPE than degradation factor uncertainty.

5.1.4 Worst Case Scenario

A single "worst case" scenario was run to determine the earliest that a correction could be necessary. In this scenario, the initial trajectory is created assuming that no degradation will take place. During the simulation, a degradation factor of $d = .2$ is used with a degradation time constant $\lambda = 1000$. Such a time constant causes the vast majority of the degradation to take place almost instantly. These settings are representative of the case where both the sail properties and expected degradation are grossly miss-estimated by the trajectory planners. It is hard to imagine an uncertainty of this extent occurring in any realized sail mission so it is solely meant to demonstrate the extrema of potential timing.

Curiously, the linear regression function from the worst case scenario predicts that the model correction can be postponed 15-20 days later than the much less severe $d = .0 - .2 \quad \lambda = 2 - 2$ case when the degradation rate is known. This is likely the combined result of the numerical error and the linear regression function itself.

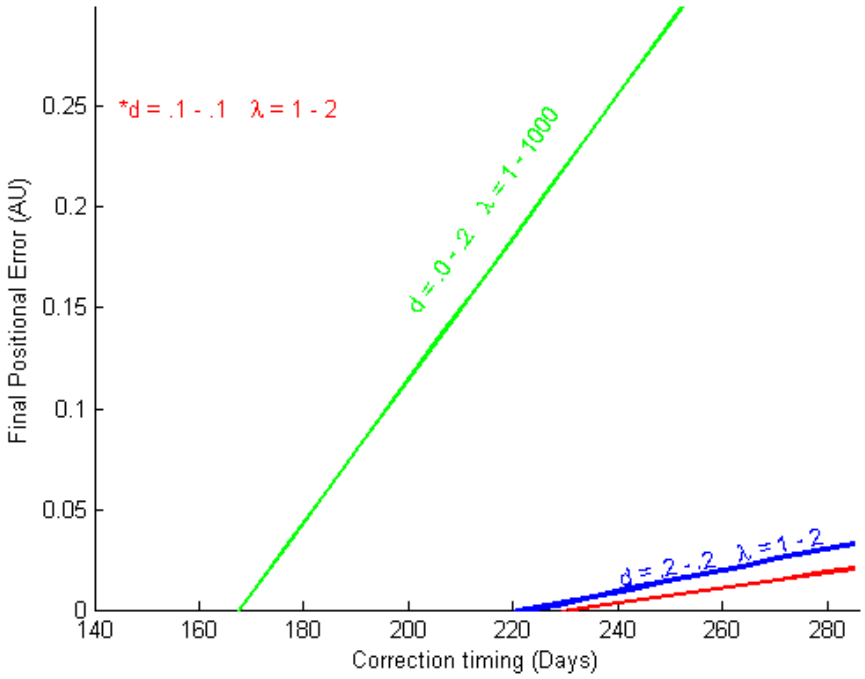

Figure 19 - Results for an uncertain lambda and the "Worst Case Scenario"

For instance, it becomes clear upon examination of Figure 25 and Figure 28 (in APPENDIX B - Unprocessed and Processed data) that the two function both transition from the region one to region two in the vicinity of the 180 day mark. However the FPE grows far more aggressively with a higher λ so the linear regression has a steeper slope, and as a result, also has a larger x-intercept.

5.2 Result analysis

These results strongly suggest that a miss-estimated trajectory can be recovered if the true degradation profile can be determined during the mission. In all cases, even where the assumed values were as incorrect as possible given the range considered, the FPE falls around the expected error before some point in the mission. Due to the variance in the required correction dates it cannot be stated that the uncertainty can be ignored completely. On the contrary, this simulation reaffirms that a correction will be required if the optical uncertainty is too great. However, this indicates that perfect knowledge of the sail is not necessarily required pre-launch.

When considered alone, the potential uncertainty range of degradation factors was far more influential than the degradation time constant. This means that understanding the extent of the degradation of the sail reflectivity, among other factors, is of far more valuable than understanding how quickly it will arrive at these values. This is convenient; ground test can be performed which expose the sail to gratuitous amounts of radiation and only the end-of-life value needs to be measured. Particular attention to the manner with which it achieves this value is therefore mostly irrelevant, as long as the exponential model is relatively accurate.

Although uncertainty in λ was the least influential of the two uncertainties, the value of λ itself does play a role in the predicted correction time limits. Given the same uncertainty in d, a higher lambda severed to shift the curves to the left, requiring an earlier model update.

6. CONCLUSION

An Earth -Venus solar sail transit with uncertain optical degradation is investigated. In the simulation the vehicle initially propels itself using the sail in a manner that would lead it Venus if assumed optical properties were accurate. After a set time, the sail model is updated and the true optical properties are realized, which are then used to attempt to create a correction trajectory that still arrives at Venus at the intended arrival date. The speed and extent of which the degradation takes place, for both the assumed and realized values, is varied to determine the influence of these two variables on the capability of a correction.

In all cases if was found that a correction could be performed at some point in the mission that yielded an FPE that was within threshold. This means that a correction trajectory was found within the precision of trajectory solving method even in the scenarios where the sail degradation is significantly underestimated. These results strongly suggest that the investigated Earth-Venus transit could be performed without prior knowledge of the sail degradation characteristics as long as they could be obtained during flight. However, these results also show that there is certainly a date after which a correction can be made. Ultimately, this means that while optical property measurements could be taken on-orbit, care must be taken to do so and correct before a theoretical deadline. The exact timing

of this deadline target varied with the exact degradation, so this risk should be well understood before the mission.

Between the two investigated parameters, it appears that the degradation factor d is more influential on correction timing than the degradation time constant λ. This is an important finding because it suggests that knowledge of the sail's end-of-life optical properties plays a larger role in corrections than how quickly it degrades; and while the former can be determined via laboratory experimentation and on-flight data, the latter is potentially dependent on solar weather and other unpredictable factors. Once more detailed analysis can create better estimates of the reaction of sail optical properties to solar emissions, the uncertainty in the degradation factor could potentially be removed, removing the urgency of a trajectory correction.

Although an Earth-Venus transit is hardly representative of all possible interplanetary missions, it can at least serve to demonstrate expected correction behavior in most cases. Travel to planets in the outer solar system usually includes an initial stage that involves decreasing perihelion to less than one AU where solar radiation is most intense. Since sail degradation is assumed to be proportional to received radiation dosage, this initial stage is also generally where the bulk of the optical degradation is expected to take place. This means that the

sample Earth-Venus trajectory could potentially share a considerable degree of similitude to the initial portions of most interplanetary solar sail missions.

6.1 Moving Forward

The findings of this thesis could be expanded in several ways. First, the analysis presented above did not consider the possibility of an arrival at the target at a date later than initially intended. A trajectory that "missed" was one that was not within a set tolerance at the end of the simulation. However, the solver attempts match both the position and velocity of the target. Even though the sail may not have reached its target adequately, it may well have a similar enough orbit that only a minor correction is necessary to bridge the distance.

These results are only directly applicable to the single Earth-Venus transit studied. Although the findings can perhaps be logically extrapolated to other cases, it cannot be 100% certain that the same behavior will be observed in trajectories to other planets until further analysis has been performed. It is therefore the recommendation of author that the same or similar analysis be performed for any specific mission parameters before it is assumed that a model update can ensure a timely arrival at the target. Although this was case for this particular mission, it cannot be safely guaranteed for all cases.

These findings only studied the usage of a single trajectory correction to mitigate the influence of uncertain degradation. Since it is assumed that the perfect

knowledge of the degradation is acquired by the time of the correction, more than a single correction is redundant. However, if this knowledge turns out to be unobtainable then these findings are not applicable. The solution to this problem could come in the form of a similar investigation as the one presented in this paper, but instead using state feedback to create correction trajectories. If similar results were found in such a study, then it would significantly less risky for a mission to launch without knowledge of the exact degradation of the sail.

Many of these improvements could be performed by running the algorithm presented in this work inside a loop. Since the departure date, arrival date and correction timing(s) are input variables in the code, the could be easily manipulated by an external loop However, due to performance issues, it is recommended that the algorithm is made to run faster, as it already took between 2-3 weeks to gather the data presented above. Perhaps the easiest way to do this is to re-write some of the core functions, such as the cubic spline interpolant function, in another programming language and interfacing it with MATLAB. (Such as a MALAB executable file (MEX))

REFERENCES

1. **McInnes, Collin Robert.** *Solar Sailing: Technology, Dynamics, and Mission Applications*. London : Springer, 2004.

2. **BBC News Science/Nature.** Solar sail probe 'probably lost'. *BBC News*. [Online] June 23, 2005. [Cited: August 21, 2013.] http://newsvote.bbc.co.uk/mpapps/pagetools/print/news.bbc.co.uk/2/hi/science/nature/4110912.stm.

3. **Clark, Stephen.** Collision between rocket stages doomed Falcon 1. *Spaceflight Now*. [Online] August 6, 2008. [Cited: August 21, 2013.] http://www.spaceflightnow.com/falcon/003/update.html.

4. **Tsuda, Y, et al.** Flight status of IKAROS deep space solar sail demonstrator. *Acta Astronautica*. November-December 2011, Vol. 69, 9-10, pp. 833-840.

5. **Anderson, Janet L.** NASA's Nanosail-D 'Sails' Home -- Mission Complete. *NASA*. [Online] November 11, 2011. [Cited: August 21, 2013.] http://www.nasa.gov/mission_pages/smallsats/11-148.html.

6. *Generalized Model for Solar Sails*. **Rios-Reyes, L and Scheeres, D.** 2005, Journal of Spacecraft and Rockets, pp. 182-185.

7. *Robust Solar Sail Trajectory Control for Large*. **Rios-Reyes, L. and Scheeres, D. J.** San Francisco : American Institue of Aeronautics and Aatronautics, 2005.

8. **Mengali, Giovanni and Quarta, Alessandro A.** Optimal Control Laws for Axially Symmetric Solar Sails. *Journal of Spacecraft and Rockets*. 2005, Vol. 42, 6, pp. 1130-1133.

9. **Scheeres, D. J.** The dynamical evolution of uniformly rotating asteroids subject to YORP. *Icarus*. 2007, Vol. 188, 2, pp. 430-450.

10. *The effect of YORP on Itokawa*. **Scheeres, D J, et al.** 2, s.l. : Elsevier, Icarus, Vol. 189.

11. **Dachwald, Bernd, et al.** Parametric Model and Optimal Control of Solar Sails. *Journal of Guidance, Control, and Dynamics*. September-October 2006, Vol. 29, 5.

12. *Potential Effects of Optical Solar Sail Degredation on Trajectory Design*. **Dachwald, Bernd, et al.** Lake Tahoe : AAS Publications Office, 2005. ASS/AIAA Astrodynamics Specialists Conference. pp. 1-23.

13. *Investigations of foil samples in simulated radiation environment on KOBE facility*. **Glotov, A. I., et al.** Toulouse : European Space Agency, 2002. Proceedings of the European Space Components Conference, ESCCON 2002. ISBN 92-9092-817-4.

14. *Characterization of space environmental effects on candidate solar sail material*. **Edwards, David L., et al.** Seattle : Photonics for Space Environments VII, 2002. doi:10.1117/12.451455.

15. **Wie, Bong.** *Space Vehicles Dynamics and Control*. Reston : American Institute of Aeronautics & Astronautics, 2007. p. 809.

16. *Very low-thrust trajectory optimization using a direct SQP method*. **Betts, John T.** 2000, Journal of Computational and Applied Mathematics, pp. 27-40.

17. *Collocation points distributions for optimal spacecraft trajectories*. **Fumenti, Fredrico, Circi, Christian and Romagnoli, Daniele.** 2013, Commun Nonlinear Sci Numer Simulat, pp. 710-727.

18. **The Mathworks, Inc.** Documentation Center. *MathWorks*. [Online] 2013. [Cited: October 20, 2013.] http://www.mathworks.com/help/optim/ug/equation-solving-algorithms.html#brrx7_e-2.

19. **Mathews, John H. and Fink, Kurtis D.** *Numerical Methods Using Matlab*. Upper Saddle River : Prentice-Hall inc., 2004.

20. **LeVeque, Randall J.** *Finite Difference Methods for Ordinary and Partial Differential Equations*. Philadelphia : Society for Industrial and Applied Mathematics, 2007.

21. **Shimizu, Kunio.** Point Estimation. [book auth.] Kunio Shimizu and Edwin L. Crow. *Lognormal Distributions: Theory and Application*. New York : Marcel Dekker, Inc., 1988.

22. **Wright, Jerome.** *Sapce Sailing*. Philadelphia : Gordon and Breach, 1992. ISBN-10: 288124842X.

23. **Kreyzig, Erwin.** *Advanced Engineering Mathematics*. New York : John Wiley & Sons, 2006.

24. *Spline Collocation Approach for the Numerical Solution of a Generalized System of Second-Order Boundary-Value Problems*. **Khuri, S. A. and Sayfy, A.** 2009, Applied Mathematical Sciences, pp. 2227 - 2239.

25. **National Aeronautics and Space Administration.** Orbital Elements. *NASA*. [Online] 2012. [Cited: December 13, 2013.] http://spaceflight.nasa.gov/realdata/elements/graphs.html.

APPENDIX A

A.1 Mission Parameters

A.1.1 Optical parameters

Table 4 - Simulated initial optical properties.

Variable	Value	Description (Units)
Sail Initial Properties[22]		
ρ	.88	Initial reflectivity
s	.94	Initial specular coefficient
E_f	.05	Initial emissivity, front
E_b	.55	Initial emissivity, back
B_f	.79	Initial non-Lambertian reflectance coefficient, front
B_b	.55	Initial non-Lambertian reflectance coefficient, back

A.1.2 Physical parameters

Table 5 - Simulated physical characteristics of the sail,

Variable	Value	Description (Units)
P	$7.6953 * 10^{20}$	Solar radiation pressure at one AU (kg DU/TU)
$Area$	$4.4683 * 10^{-21}$	Area of sail (DU^2)
m	40	Mass of spacecraft (kg)

A.1 Solver Parameters

Table 6 - Relevant solver parameters

Parameter	Value	Description
$neqcoes$	$(3 * 2^{run-1}) - 1$	Number of points used to parameterize each assumed trajectory by the solver for a given run. Run-1: neqcoes = 3 Run-2: neqcoes = 5 Run-3: neqcoes = 9 Run-4: neqcoes = 17 Run-5: neqcoes = 33
$runmax$	5	The number of cycles the solver performs
num_start	1000	The number of initial assumed trajectories received by the solver.
$delete\_tol$	.0005	Defines the tolerance under which two points are considered redundant and one is deleted. e.g. if $\|\psi_1 - \psi_2\|$ $< delete\_tol$
uni_bound	$\begin{bmatrix} .3 & 15 \\ -.5 & .5 \\ -.5 & .5 \\ -.05 & .05 \\ -.05 & .05 \end{bmatrix}$	Defines the bounds of the first 5 equinoctial orbital elements for procedurally generated initial guesses.
$guess\_diff\_type$	'random'	Sting to define the method used to generate initial guesses. Although different strings are accept by the algorithm, only 'random' was used in the results presented.

n_invest	$2 * neqcoes - 1$	Number of collocation points used by the solver create the objective.

APPENDIX B - Unprocessed and Processed data

B.1 $d = 0-.1 \quad \lambda = 1-1$

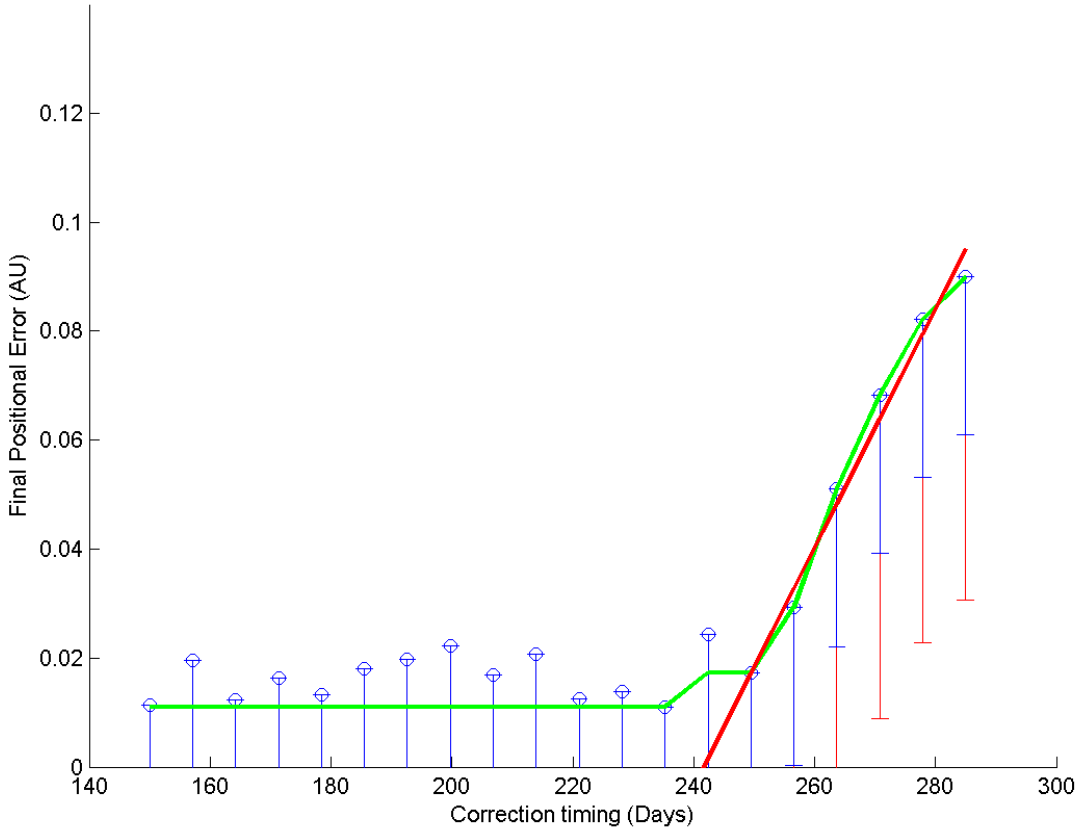

Figure 20 - d=0-.1 λ=1-1

B.2 $d = .1 - .2 \quad \lambda = 1 - 1$

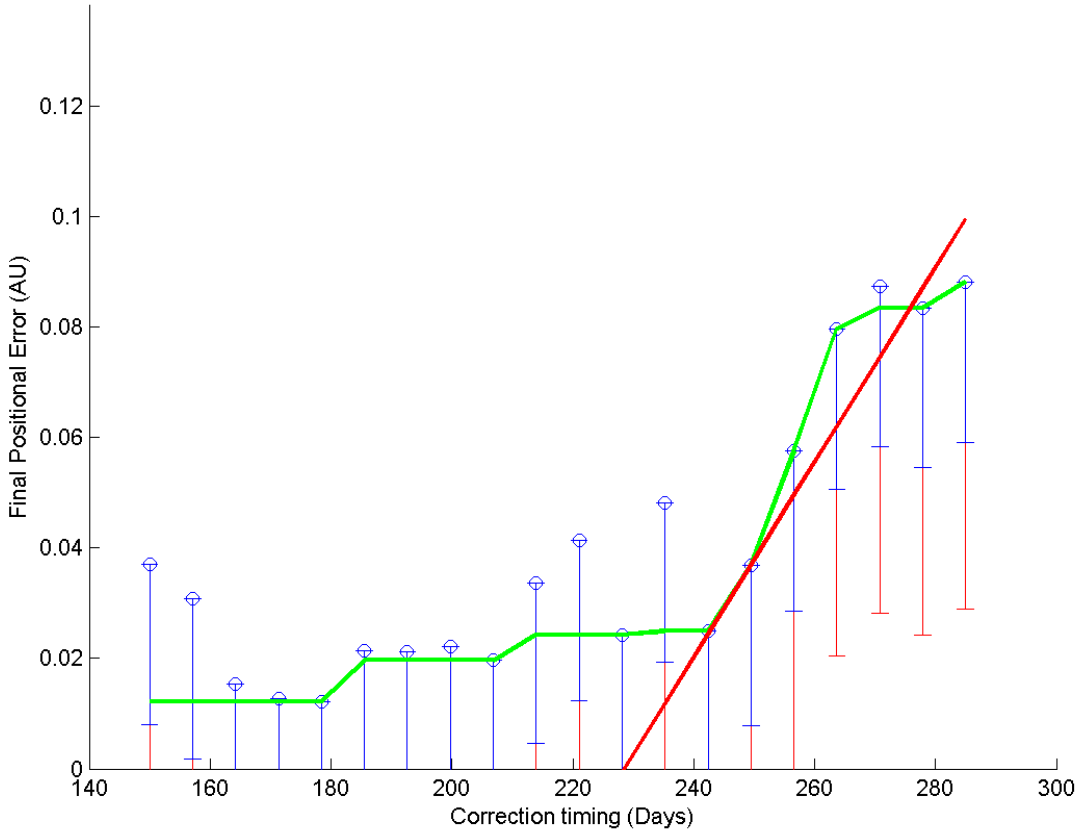

Figure 21 - d=.1 -.2 λ=1-1

B.3 $d = 0-.2 \quad \lambda = 1-1$

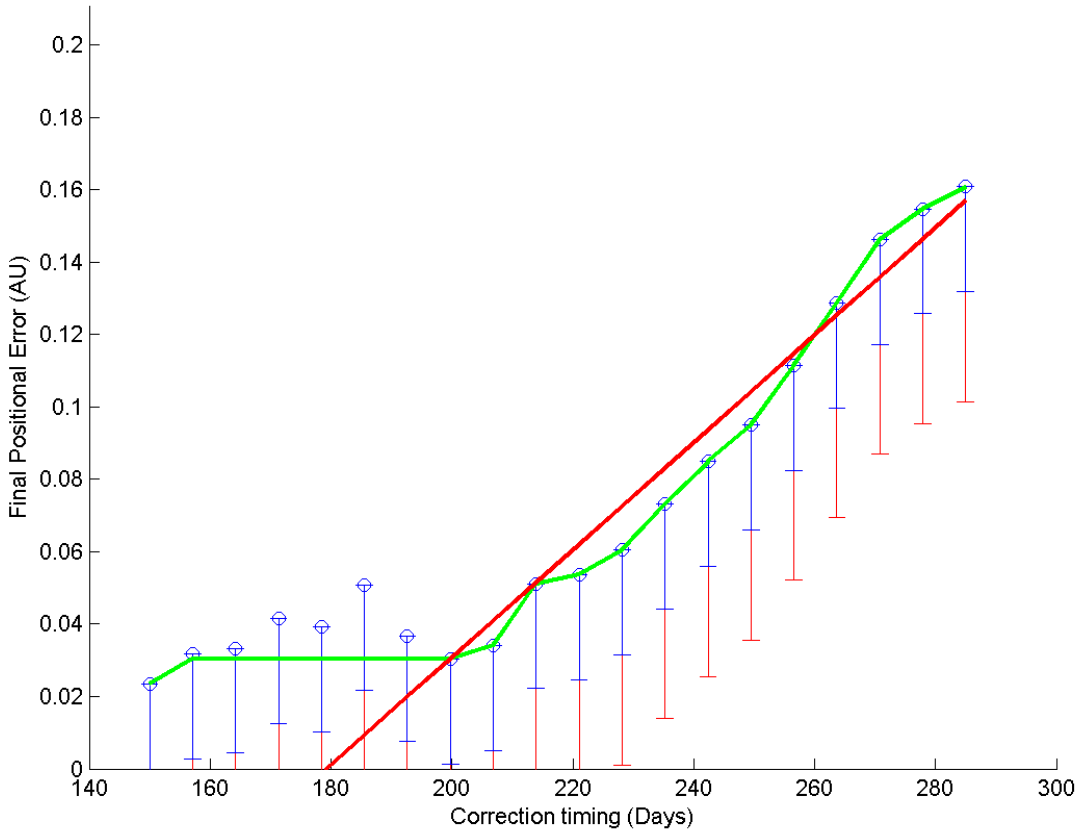

Figure 22 - d=0-.2 λ=1-1

B.4 $d = 0 - .1 \quad \lambda = 2 - 2$

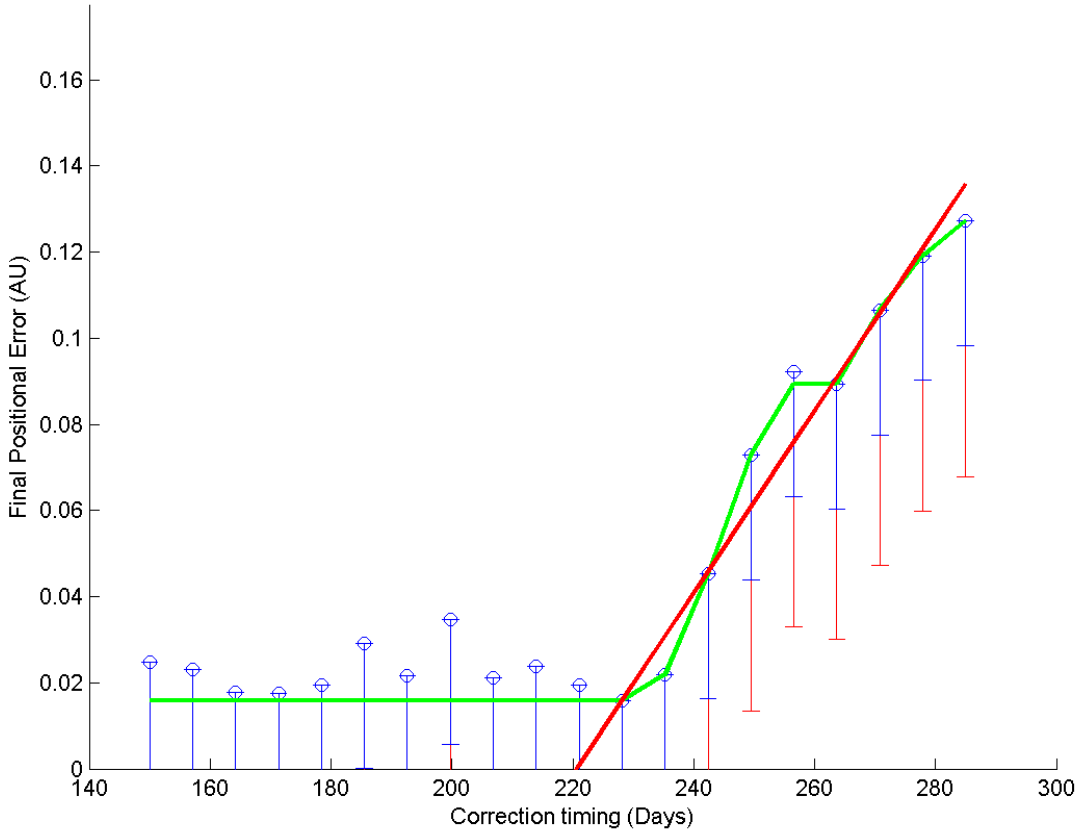

Figure 23 - d=0-.1 λ=2-2

B.5 $d = .1 - .2 \ \lambda = 2 - 2$

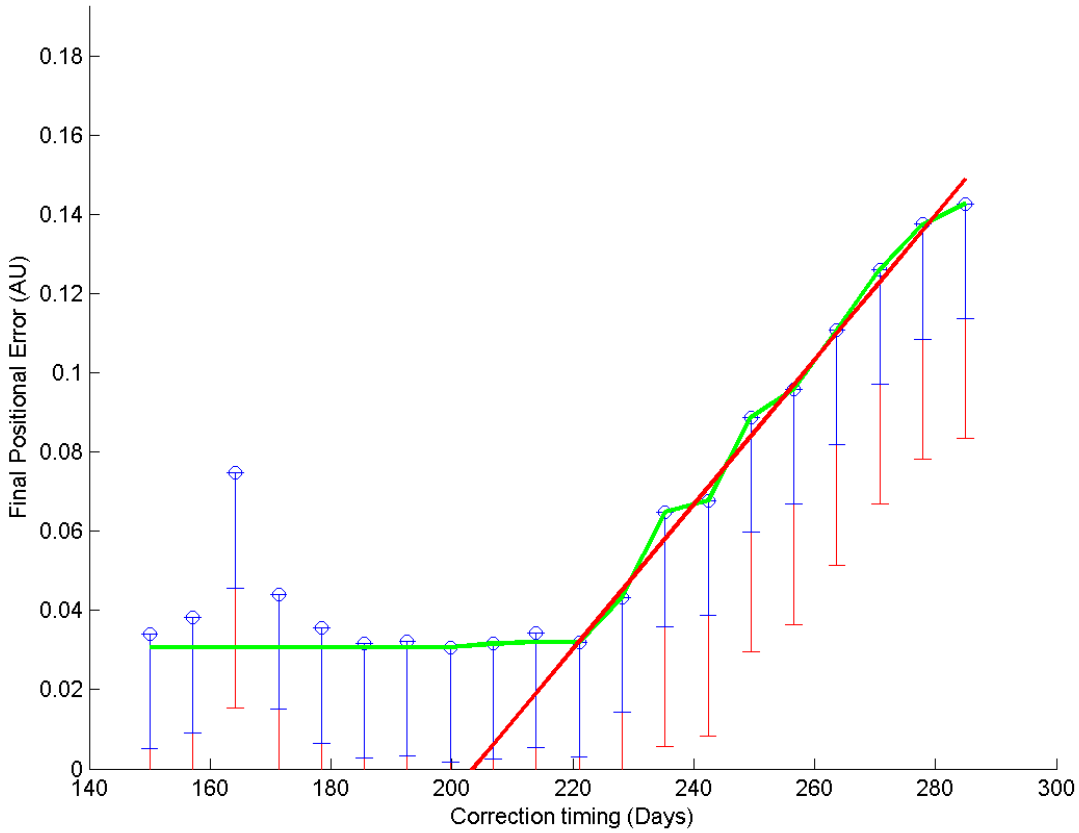

Figure 24 - d=.1 - .2 λ=2-2

B.6 $d = 0{-}.2 \quad \lambda = 2 - 2$

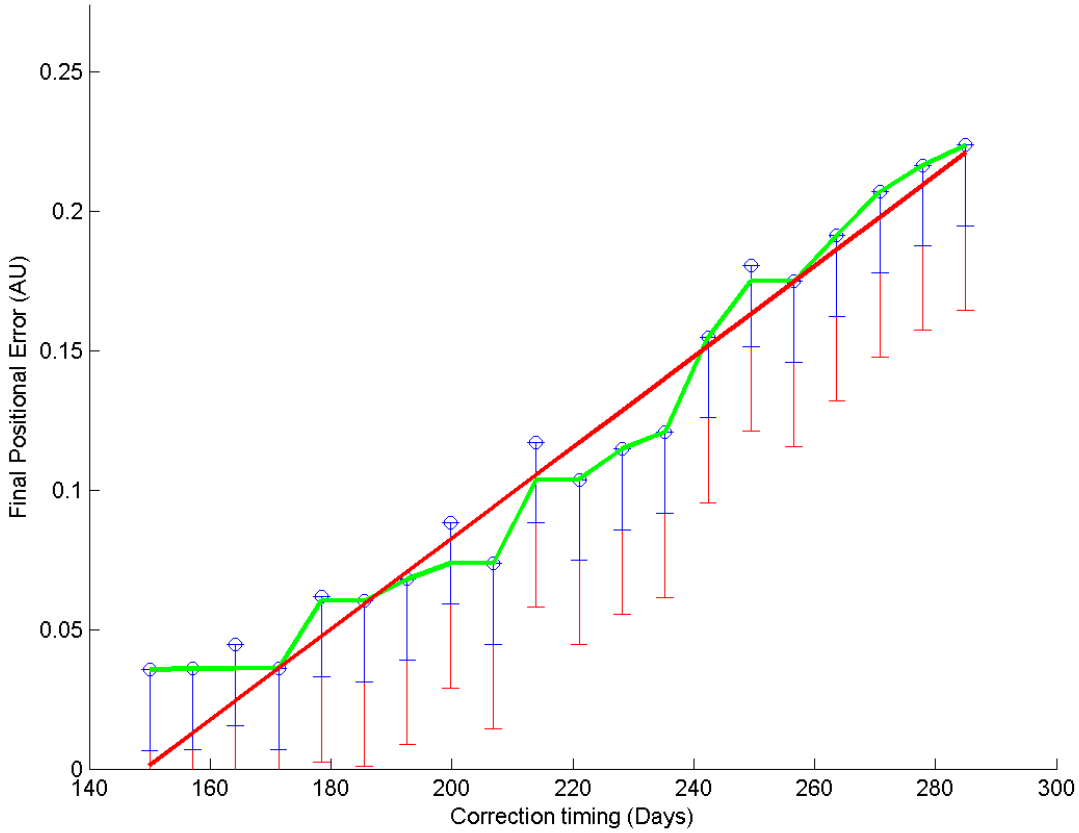

Figure 25 - d=0-.2 λ=2-2

B.7 $d = .1 - .1 \quad \lambda = 1 - 2$

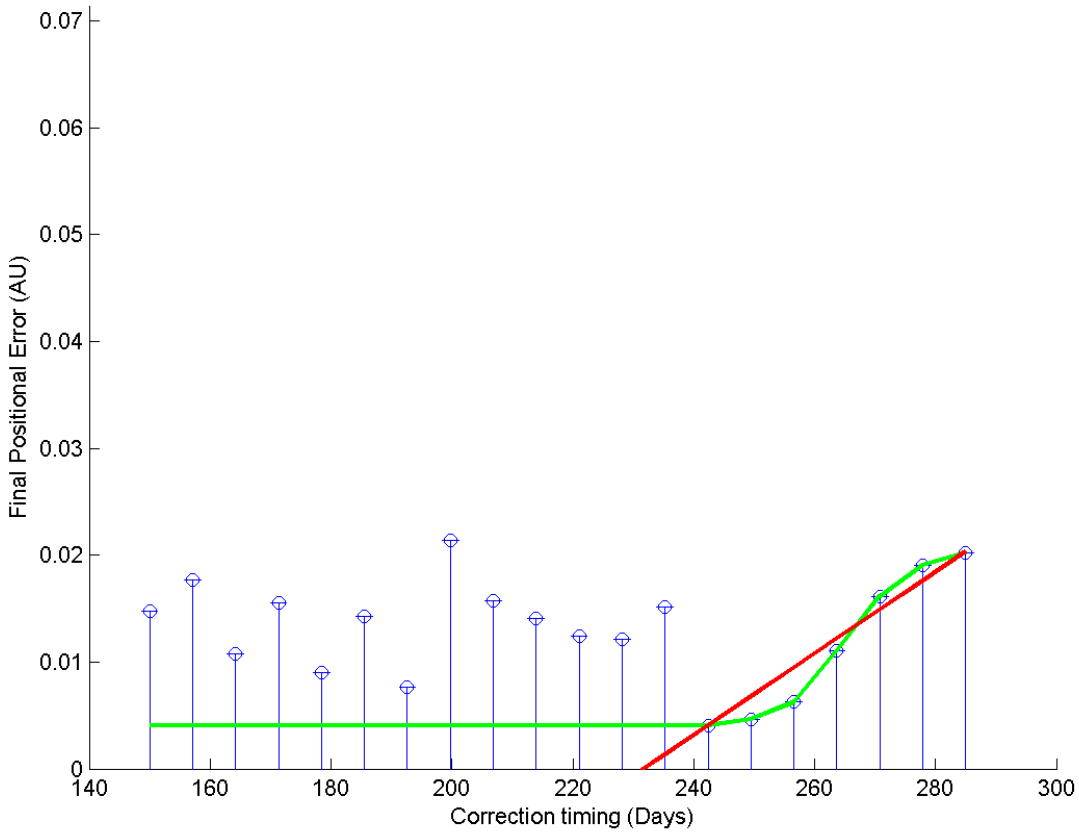

Figure 26 - d=0-.1 λ=1-2

B.8 $d = .2 - .2$ $\lambda = 1 - 2$

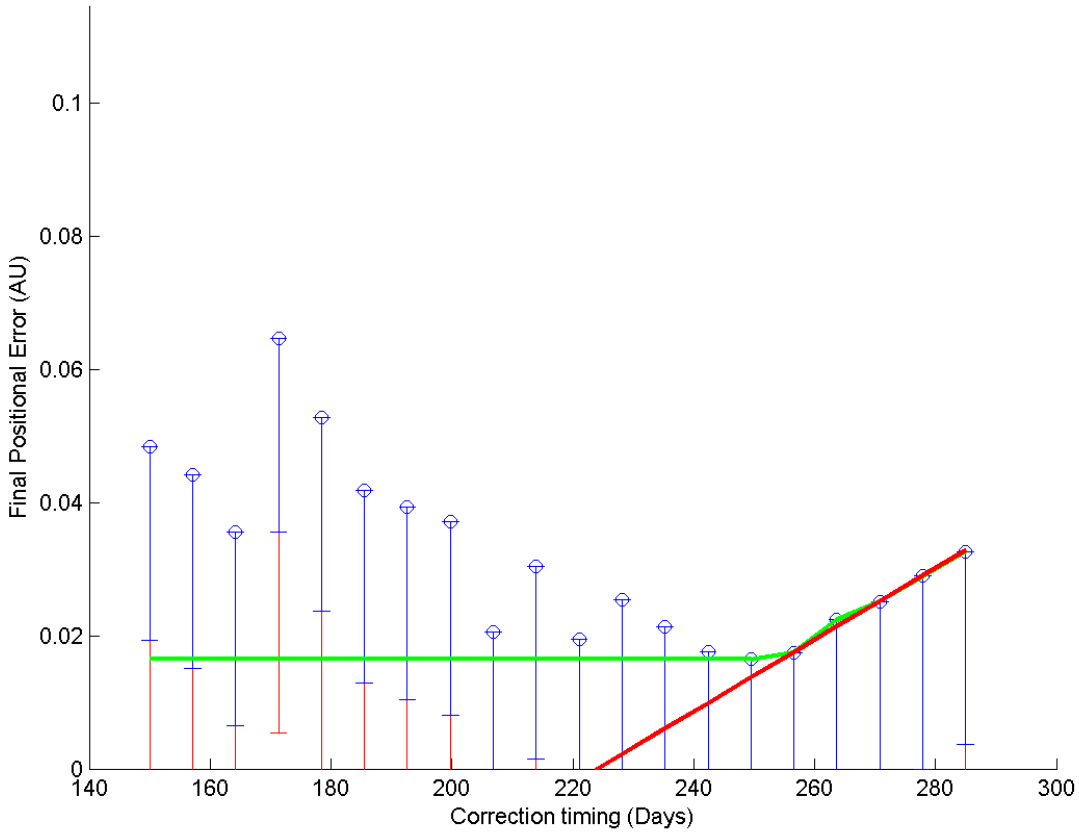

Figure 27 - d=.2-.2 λ=1-2

B.9 $d = 0 - .2 \quad \lambda = 1 - 1000$

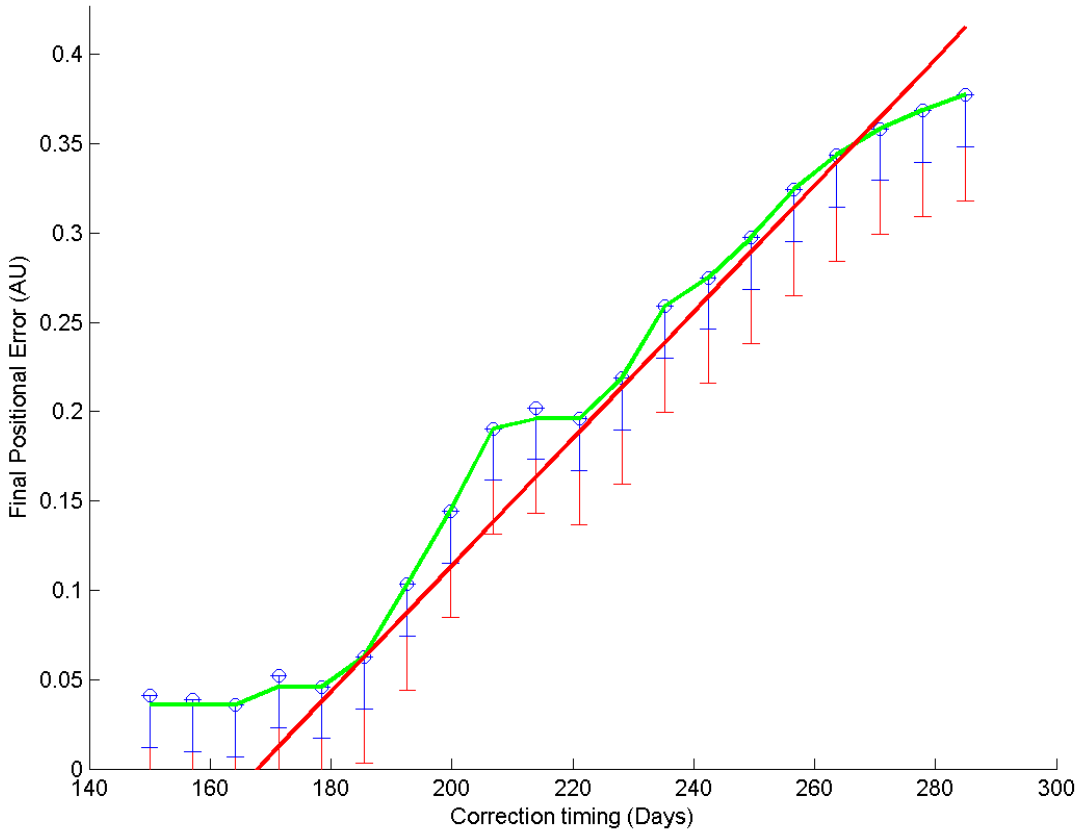

Figure 28 - d=0-.2 λ=1-1000

APPENDIX C - Function Parameterization

It is well known that functions are, in a sense, vectors by nature. For instance, through a Taylor series expansion, the sin function can be expressed in the following manner:

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \quad \textbf{68}$$

This equation can be expanded once more, rearranging the right hand into a simple vector multiplication:

$$\sin x = \begin{bmatrix} 1 & x & x^2 & x^3 & x^4 \ldots \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ -\frac{1}{3!} \\ 0 \\ \ldots \end{bmatrix} \quad \textbf{69}$$

Since only the values of the column are specified by the function, it can be thought of as a vector representation of the function. The row vector is then a mapping which, when multiplied by the function vector, translates any value within the domain of x into the corresponding range of $\sin x$.

In this particular example both vectors would need to continue on to infinity to represent the desired function exactly, as is the case for any infinite series equivalent to a function. It is therefore impossible to express many functions

exactly, so an approximation is used by truncating the length of these vectors. As one would expect, the longer the vector the better it can approximate said function, but it is often sufficient for most purposes to only use a few terms.

This principle is employed to avoid having to perform complex functional calculations in the course of finding the desired trajectories. It is assumed that, as was shown above, that the unknown functions $p(t)$, $f(t)$,$g(t)$,$h(t)$,$k(t)$ and $L(t)$ can be expressed as vectors. Each value in this vector is treated as independent variable – variables upon which a numerical solver can operate upon.

Instead of using a polynomial as the mapping function, the trajectories are parameterized by a cubic spline. One of the major downsides of high order polynomials is large fluctuations when $|x| \gg 0$, known as numerical instability.[23] Cubic splines avoid this instability, offering the local stability of the polynomial across the whole domain.

The employed cublic spine creates an interpellant function by using a linear combination of basis functions as described by Khuri and Sayfy. Each basis function is a continuous set of piecewise polynomials, described by the equation[24]

$$\psi_i(x) \qquad [x_i \quad , x_{i+1}]$$
$$[x_{i+1}, x_{i+2}]$$
$$= \frac{1}{h^3} \qquad [x_{i+2}, x_{i+3}]$$
$$[x_{i+3}, x_{i+4}]$$
$$otherwise \qquad \textbf{70}$$

$$(x - x_i)^3$$
$$h^3 + 3h^2(x - x_{i+1}) + 3h(x - x_{i+1})^2 - 3(x - x_{x+1})^3$$
$$* \; h^3 + 3h^2(x_{i+3} - x) + 3h(x_{i+3} - x)^2 - 3(x_{i+3} - x)^3$$
$$(x_{i+4} - x)^3$$
$$0$$

Where ψ_i is the basis function and x is the independent variable. The subscripts indicate the nodes of each piecewise function, which span the interval lh.

If this function is evaluated at a node, it only has 3 possible values: 0,1, or 4. It then follows that a cubic spline can be defined using n equally spaced values

$$\begin{bmatrix} 4 & 1 & 0 & 0 & \dots \\ 1 & 4 & 1 & 0 & \dots \\ 0 & 1 & 4 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{Bmatrix} O_1 \\ O_2 \\ O_3 \\ \vdots \end{Bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ \vdots \end{bmatrix} \qquad \textbf{71}$$

where $f(x_i)$ is the defining function values and O_i are values that parameterize the desired function. This equation is analogous to equation 69 in the sense that a set of variables once again represents a functional.

APPENDIX D - Additional Diagrams

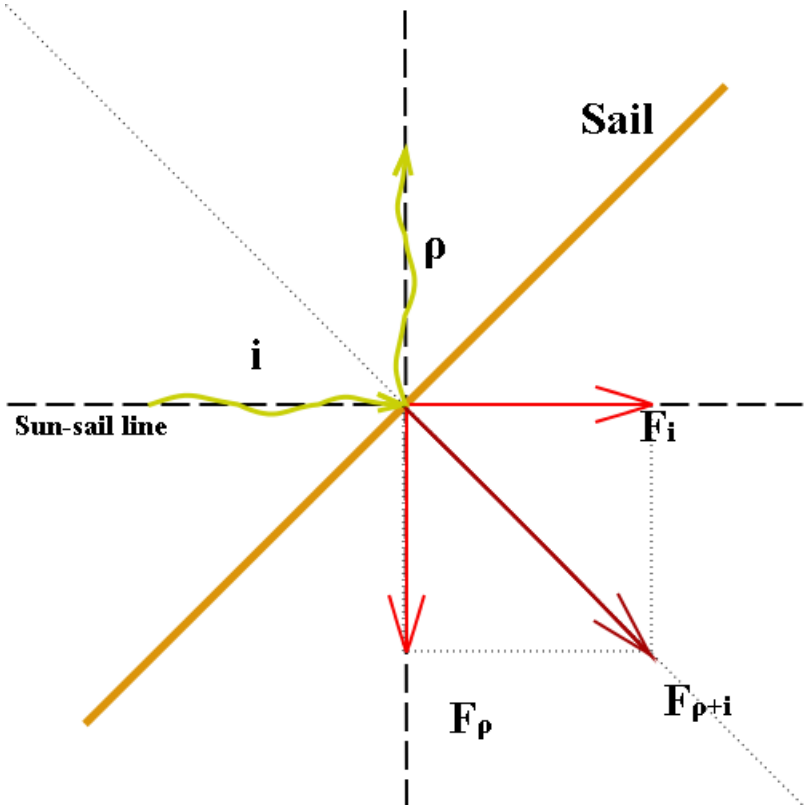

Figure 29 - The geometry of photonic forces for an ideal solar sail.

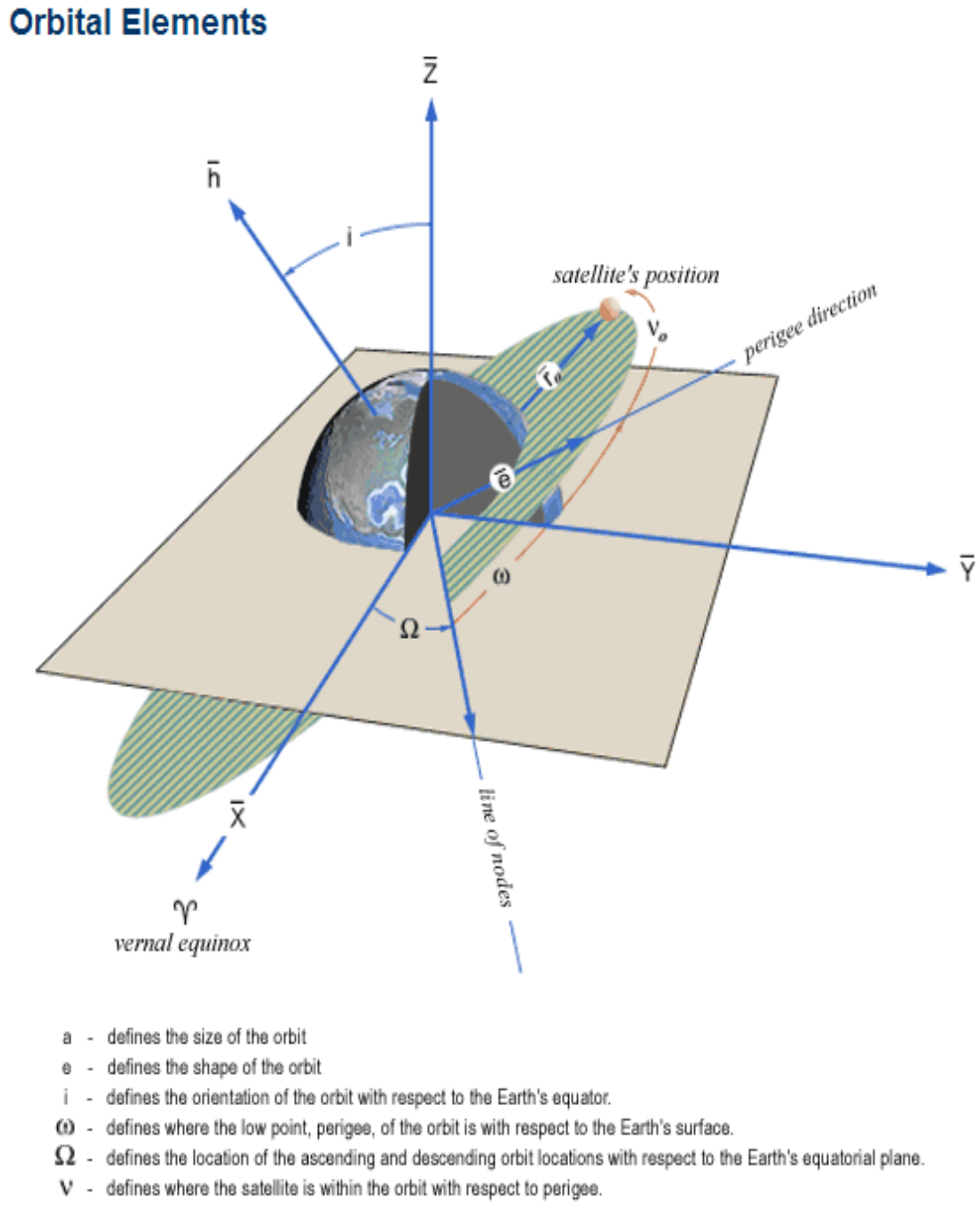

Figure 30 - An illustration of the classical orbital elements from the NASA website[25]