An Architecture for a Generalized Spacecraft Trajectory Design and Optimization System

Cesar Ocampo
Assistant Professor,
Department of Aerospace Engineering and Engineering Mechanics,
The University of Texas at Austin, Austin, Texas, USA 78712

Abstract

The elements of a general high precision system for trajectory design and optimization for single or multiple spacecraft using one or more distinct propulsion systems, and operating in any gravitational environment within the solar system are discussed. The system architecture attempts to consolidate most all spacecraft trajectory design and optimization problems by using a single framework that requires solutions to either a system of nonlinear equations or a parameter optimization problem with general equality and/or inequality constraints. The use of multiple reference frames that generally translate, rotate, and pulsate between two arbitrary celestial bodies facilitates the analysis of multiple celestial body force field trajectories such as those associated with libration point missions, cycling trajectories between any set of celestial bodies, or any other type of trajectory or mission requiring the use of multiple celestial bodies. A basic trajectory building block, referred to as the basic segment, that can accommodate impulsive maneuvers, maneuver and non-maneuver based mass discontinuities, and finite burn or finite control acceleration maneuvers, is used to construct single or multiple spacecraft trajectories. The system architecture facilitates the modeling and optimization of a large range of problems ranging from single spacecraft trajectory design around a single celestial body to complex missions using multiple spacecraft, multiple propulsion systems, and operating in multiple celestial body force fields.

1 Introduction

Spacecraft trajectory optimization is a field that has received considerable attention over the last several decades. The field continues to evolve as a result of innovations in trajectory dynamics associated with spacecraft utilizing the simultaneous gravitational attraction of two or more celestial bodies where the contribution from any of the celestial bodies influences significantly the motion of the spacecraft (Ref.[1]). Further advancements have been made in both the analytical and numerical based solution methods required to solve these types of problems. The most recent developments in purely analytical solutions to low thrust orbit transfer problems have been given by Azimov and Bishop[2] and a recent numerical based solution method capable of solving complex low thrust multiple body gravity trajectories has recently been documented by Whiffen and Sims[3].

Fundamental results in trajectory optimization are based in part on original research due to Lawden[4]. Lawden introduced the Primer Vector and its use in the optimization of space trajectories is known as Primer Vector Theory. The Primer Vector and its applications form an integral part of the current system but the specific details associated with the optimization of both impulsive and finite burn maneuvers are not discussed in this article.
Associated with the theory of optimal space trajectories, are the numerical methods required to solve spacecraft trajectory problems in force fields where even closed form solutions of the uncontrolled trajectories are not available. A comprehensive survey on the computational issues applied specifically to the spacecraft trajectory optimization problem is given by Betts[5]. A clear exposition on the conversion of optimal control problems into sub-optimal parameter optimization problems whose solutions require nonlinear programming is given by Hull[6].

The focus of the current article is to describe a general approach that draws upon pertinent aspects of trajectory design and optimization theory. The effort is an attempt to present a single framework in which one or more spacecraft operating in a force field environment under the mutual attraction of one or more celestial bodies, and using one or more propulsion systems can be analyzed and solved efficiently. The framework for this architecture is currently implemented in a prototype trajectory design and optimization system called COPERNICUS that is under development at the University of Texas.

The term ‘architecture’ in the title of this article refers to the structure and methodology of the system. The architecture is defined by its basic components which include the force models, the coordinate frames, the numerical methods, and the method used to model the trajectories. The term ‘generalized’ implies that the system is designed to handle many classes of problems involving common gravitational force fields and acceleration models, different sets of boundary conditions, and various types of propulsion systems without resorting to specific procedures or algorithms to solve each type of problem separately with unique and distinct methods. Common in the spacecraft trajectory optimization literature is the description and implementation of specific methods and algorithms to solve specific problems. Though this is entirely valid, the current architecture attempts to unify the approach to these same problems under a single framework. The terms ‘trajectory design’ refer to the process of generating nominal and feasible solutions that satisfy a predetermined set of constraints and boundary conditions without considering the optimization of any aspect of such a solution. The term ‘optimization’ refers to the process of generating a trajectory design solution that extremizes some general scalar quantity of the solution, regardless of how complex it may be, provided that it can be determined or computed deterministically.

In a spacecraft trajectory optimization problem, the cost functions typically considered include, minimization of the total impulse required, minimization of the total transfer times, minimization of propellant used, or maximization of the final spacecraft mass. However, other allowable cost functions may include the minimization of the hyperbolic excess velocity relative to an arrival or flyby celestial body or the minimization of the value of the Jacobi constant of a spacecraft in a circular restricted three body system. Important here is that the cost function should be allowed to take on any value that is of interest to a mission, provided that it can be uniquely computed from the variables and system parameters used to model the problem.

The current prototype of the system uses explicit numerical integration for state propagation since the force fields encountered in these problems are in general nonlinear and non-autonomous. The solutions to the different classes of problems are obtained as solutions to either a system of nonlinear equa-
tions or to a nonlinear constrained parameter optimization problem. Efficient gradient based nonlinear root finding algorithms and sequential quadratic programming algorithms are used to obtain these solutions.

All spacecraft trajectory problems can be formulated as either multi-point boundary value problems or as nonlinear programming problems that explicitly extremize single or multi-objective cost functions. In the case of a multi-point boundary value problem formulation, two cases exist. First, feasible trajectory solutions with no implicit optimization satisfy only the boundary conditions associated with the kinematics or the physical variables of the problem such as the conditions related to the physical state of the spacecraft or the physical parameters of the force model. A simple example of this includes the central body, orbital two point boundary value problem that satisfies Lambert’s theorem (Ref.[7]). Here, a nonlinear search is required for the three components of the velocity vector that connects two distinct position vectors in a given flight time. However, there exist other types of targeting involving more complex boundary conditions that are general functions of the state variables. An example of one of these is to find the launch injection conditions from any given Earth centered parking orbit so that the trajectory terminates in a captured orbit about the Moon. This is the classic ballistic lunar capture trajectory first examined by Belbruno[8]. For this problem, a common parameterization of the launch conditions is given by the hyperbolic excess velocity magnitude, $v_\infty$, and the right ascension and declination of the outgoing asymptote, $\alpha_\infty$ and $\delta_\infty$, respectively. The flight time, $\Delta t$, is a free parameter, and one of the boundary conditions for a successful capture at the Moon is that the value of the Jacobi Constant of the spacecraft’s state measured with respect to the Earth-Moon system be between the Jacobi Constant values associated with the interior and exterior libration points that are near the Moon. The other boundary condition is that the radial distance between the spacecraft and the moon be less than the radius of the Hill sphere around the Moon at the time of the capture. The problem is further complicated by the fact the gravitational attraction of the Sun is required for this solution to exist. This problem has four unknowns ($v_\infty$, $\alpha_\infty$, $\delta_\infty$, $\Delta t$) but three boundary conditions given as inequalities. It is a targeting problem, but much more complex than the orbital two point boundary value problem, that can still be solved as a system of nonlinear equations.

The second case of a multi-point boundary value problem formulation results if optimal control theory is used to formulate the multi-point boundary value conditions associated with a specific problem. Here the solutions produced satisfy both the kinematic boundary conditions and the natural boundary conditions, commonly referred to as the transversality conditions, that are part of the first order necessary conditions for a solution of the optimal control problem. This case implicitly extremizes a scalar cost function given in either the Mayer, Lagrange, or Bolza forms and is referred to as an indirect method.

The nonlinear constrained parameter optimization problem directly extremizes a scalar cost function based on a variable parameter vector that can include parameters of the model and/or all other variables defining the states and dynamics of the spacecraft. A hybrid formulation is also possible where the search variables associated with the optimal control formulation of the multi-point boundary value problem augment the parameter vector of the pa-
parameter optimization problem. This eliminates the necessity of deriving and implementing the transversality conditions as constraint conditions associated with the optimal control problem.

Having stated and described a class of solution methods that can solve these problems, significant importance is placed on the modeling of the problems, the proper choice of coordinate frames for state definition and targetting, and the proper identification of the independent and dependent variables. Though there are multiple ways in which a problem can be modelled, there are some that have better convergence properties than others for a given choice of solution method. A well designed system should facilitate, via experimentation if needed, the modelling of many types of problems and the generation of the solution procedures required to solve them. A solution procedure entails identifying a single stage or multi-stage procedure that drives an initial estimate of a solution to convergence.

It is noted that the trajectory problems solved by this system, whether optimal in some sense or not, are referred to as ‘open loop’ solutions in contrast to the guidance or stabilization problems encountered in many control problems. These solutions will eventually have to be controlled with closed loop feedback systems because in a real operational implementation of the solutions, unmodelled perturbations, uncertainties in the state of the spacecraft, and maneuver execution errors will be present.

Fundamental to the general methodology presented here is the use of a basic element referred to as the basic segment that is composed of several basic entities. The basic segment is the basic building block from which all trajectories are constructed. Fundamentally, the basic segment is an arc that connects two endpoint nodes. The segment can accommodate velocity and mass discontinuities at either or both of the endpoint nodes. The velocity discontinuities result from either an impulsive maneuver produced by an engine or from an approximation to a gravity assist flyby about a given celestial body. However, whether the gravity assist is approximated by an impulse change in the velocity vector or whether it is a fully integrated trajectory arc about a celestial body is a choice allowed in the definition of the segment. The mass discontinuities at either node result from the mass depletion associated with an impulsive maneuver, or the mass changes that result when part of a spacecraft is discarded or added. This results when an engine stage is discarded or when a spacecraft is captured after a rendezvous. The arc connecting the end point nodes can be either a ballistic arc or a controlled arc if a finite burn engine or any other device (such as a solar sail) is used to produce an independent acceleration to the spacecraft. A ballistic arc is the same as a coast arc. Each segment can be altered to form several segment types. For example, a segment can be defined to be an impulse followed by a coast period, or a finite burn arc without the endpoint impulses, or simply a node point defined with an epoch, a position, and a velocity.

Inherent to the system is the capability to work with disconnected segments. A segment is disconnected to any other segment if any of the state variables (position, velocity, or mass) at either node of one segment is discontinuous with respect to any of the state variables at the nodes of the other segment. Common sets of disconnected segments are those that have position discontinuities. This allows multiple spacecraft missions to be studied because additional spacecraft are modelled as segments that may be dynamically
uncoupled from the segments used to model the trajectories of the other spacecraft. If some of the spacecraft are required to interact with other spacecraft, the node points of the segments that represent them are coupled via suitable boundary conditions to examine intercept, rendezvous, constrained formation flight, or any other type of problem that requires constraining the dynamic state of any spacecraft with respect to other spacecraft.

A similar method of using disconnected segments was used by Byrnes and Bright[9] to examine complex impulsive maneuver based multiple body flyby trajectories to connect and optimize initially disconnected segments similar in concept to those described here. Though their definition of a segment and how it is propagated is different than the one presented here, the concept is similar in that a solution is achieved by simultaneously minimizing the total sum of the impulsive maneuvers and achieving continuity in position, and optionally the velocity vectors, at initially disconnected node points that they referred to as ‘breakpoints’. Their method was able to robustly produce solutions to the Galileo tour of the Jovian system and the heliocentric multi-planetary flyby trajectory for the Cassini mission.

The principal goal of current the system and architecture is to solve complex problems in a standard way without the need to develop specific models and algorithms for specific problems. If designed and implemented correctly, a general trajectory design and optimization system should be the model of choice for any specific problem, regardless of its complexity. Such a system should be capable of solving practical trajectory design and optimization problems using multiple propulsion systems, multiple spacecraft, and multiple celestial bodies and with any set of measurable perturbing accelerations in the force field. Examples of the type and scope of the problems that can be solved by the system include:

1. Trajectories about a single celestial body for orbit transfers, rendezvous, intercepts, arrival and capture, departure and escape;
2. Transfer and return trajectories between any pair of celestial bodies that orbit each other or that are in orbit about another celestial body;
3. Trajectories associated with the libration points of any two celestial body system, including transfer trajectories to and between libration points or libration point orbits; or libration point orbit trajectory design;
4. Sample return missions including descent and ascent at the target celestial body including any necessary rendezvous maneuvers;
5. Ballistic, low energy, or low thrust cycler trajectories between any pair of celestial bodies;
6. Multiple body gravity assisted trajectories in the Solar System or any central body with one or more natural satellites using any combination of impulsive and/or low thrust maneuvers;
7. Ballistic or controlled low energy capture and escape from any celestial body using the direct influence of other celestial bodies.

A secondary goal of this system is to efficiently produce an accurate solution to a problem with minimal effort. Minimal effort is defined to be a level
of effort required to generate a solution based on a predefined procedure to model and solve a particular problem. Producing a predefined procedure for a given problem, i.e., determining how many segments should be used, what are the independent and dependent variables, etc., can be a lengthy process. However, once this process is shown to achieve convergence from a wide range of initial estimates, then it becomes an automated process facilitating the solution to similar problems with different parameters. Complex problems may require a multi-stage approach, where sub-problems are solved independently and then combined in an overall solution. A well designed system should at least generate solutions to the subproblems efficiently and attempt to solve the complex problem.

A general approach to the problem facilitates the solution and optimization of trajectories for many types of missions in any force field encountered in most spacecraft trajectory problems. Though it is desired that this architecture solve all spacecraft trajectory problems foreseen in the next several decades, problems may be posed that are not solvable with the current architecture without further analysis, development, and generalization.

Common to all of these problems is the requirement to have available an initial estimate of a solution that leads to convergence. This initial estimate becomes increasingly difficult to produce as the problem increases in complexity. This initial estimate has often been referred to as the ‘first guess’, however, this term will be avoided, because it should not be a ‘guess’. There is a lot of information available in a problem that can be used to construct a first estimate. Without going into details, the analytical solutions or approximations for simple force models and the dynamical systems based analysis currently being developed for more complex models (Ref.[1]) serve as a basis for the construction of these estimates. However, convergence to a solution from even a well founded initial estimate is not guaranteed, given the complexity and scope of the problems that can be considered.

Several unresolved issues remain that need to be addressed eventually if the system described here is to achieve some of the goals stated. First, it will be necessary to explore simplifications in the construction of initial estimates to most or all of the problems attempted. Secondly, if a comprehensive spacecraft dynamics system is to be produced, it will be necessary to incorporate a six degree of freedom spacecraft model to account for the attitude reorientation maneuvers required to properly align the spacecraft to perform the needed maneuvers. The spacecraft model considered in the current system is still restricted to be a three degree of freedom model. Thirdly, if a detailed spacecraft operations system is to be produced, it will be necessary to include the observability and navigation accuracy of the trajectory solutions as part of the cost function. And fourthly, for close proximity operations between multiple spacecraft such as formation flight, it will be necessary to add a general relative motion frame model to examine multi-spacecraft trajectory problems about arbitrary trajectories, flying in any force field environment.

**Notation:** All scalar quantities are typeset as italicized uppercase or lowercase, i.e., \( a \) and \( B \) are scalars. Vectors are typeset as bold lower case, i.e., \( \mathbf{a} \) and \( \mathbf{b} \) are vectors. Matrices are typeset as bold uppercase, i.e., \( \mathbf{A} \) and \( \mathbf{B} \) are matrices. The definition and dimensions of these variables or constants are context dependent and appear in the text where appropriate. Vectors are column vectors, so that if \( \mathbf{a} \) is an \( n \) vector and \( \mathbf{b} \) is an \( m \) vector and \( \mathbf{a} \) is a
function of $b$, then $\partial a/\partial b$ is an $n \times m$ matrix. Dots above any quantity represent differentiation with respect to time. Superscript and subscript symbols are used to further distinguish the meaning of a given quantity; the definition of these are given in the text where appropriate.

# Trajectory Design and Optimization Architecture

This section describes the basic elements of the system. This includes the formal definitions of the segments, trajectories, missions, force fields, and the coordinate systems.

## 2.1 Definition of the Basic Segment

The system makes use of what is termed the basic segment. It is a trajectory arc that connects two node points. The arc connecting the two node points is a solution to the equations of motion which are propagated by whatever means necessary. The force field models are generally smooth, highly nonlinear, and time dependent, thus requiring numerical based solutions of the equations of motion. A trajectory is composed of one or more segments. A complete mission is composed of one or more trajectories associated with one or more spacecraft. A spacecraft is any object in the model that is not a celestial body and does not influence the motion of any other object in the system such as another spacecraft or a celestial body. It is the object whose state at single or multiple times is being determined or controlled. A celestial body is any object that can influence the motion of a spacecraft via its gravitational potential or emitted radiation in the equations of motion. Celestial bodies also influence the motion of other celestial bodies in the system. This influence is directly available from a pre-computed ephemeris of the celestial bodies in the model. It is assumed that the motion of the celestial bodies is known, either from a realistic ephemeris or from an analytical approximation to a real ephemeris; i.e., the system only propagates the motion of one or all of the spacecraft in a mission.

The node points are tagged with an epoch that is referenced to a specified, but otherwise arbitrary, reference epoch, denoted as $t_{\text{epoch}}$. The epoch of the initial node point of any segment $i$ is $t_{i0}$ and the epoch of the final node point is $t_{if}$. The superscript $'i'$ denotes the segment number. A mission can have any number of segments so that $i = 1, ..., n$. There is no restriction on the values that $t_{i0}$ and $t_{if}$ can have, i.e.,

$$t_{i0} = t_{if} \quad \text{or} \quad t_{i0} > t_{if} \quad \text{or} \quad t_{i0} < t_{if}$$

However, the values of $t_{i0}$ and $t_{if}$ can be constrained in any way necessary during the solution process, i.e., both quantities can be independent if desired or functionally dependent with respect to each other or the time epochs of the node endpoints of any other segment. For example, the time of flight for a segment $i$ can be constrained to be less than or equal to a specified time of flight, or the time of flight of any other segment $k$; i.e., $dt^i = t_{if}^i - t_{i0}^i \leq dt^k$. Forward time or backward time propagations are handled in the same way and the temporal direction is only determined by the specific values of $t_{i0}^i$ and $t_{if}^i$. If $t_0 \neq t_f$, all segment propagations are from $t_0$ to $t_f$, regardless of their relative values. No propagation is made if $t_0 = t_f$. 

The state of the spacecraft at either node point is given by its position, \( r_{(0,f)} \), velocity \( \mathbf{v}_{(0,f)} \), and mass, \( m_{(0,f)} \). The subscript \((0,f)\) on each of these quantities denotes that the quantity is referenced to either the \( t_0 \) node or the \( t_f \) node. The superscript ‘−’ on \( v_{(0,f)} \) and \( m_{(0,f)} \) states that the value of the velocity and the mass, respectively, is given prior to any possible discontinuities in their values.

There are two types of velocity impulses allowed at either node. The first type is a maneuver based impulse provided by an engine and the second type is a gravity assist impulse. The gravity assist impulse is used to approximate the change in velocity relative to a fixed external reference frame not attached to the celestial body providing the gravity assist. It is useful only when solving a problem where the flyby celestial body is treated as a zero-point mass. The components of the gravity assist impulse are constrained to satisfy the conservation of energy across the flyby and optionally, a minimum flyby radius relative to the central body. A fully integrated flyby of a celestial body without any associated discontinuity in the velocity is also allowed, but this flyby is modelled as a numerically integrated segment with a non-zero time of flight duration. Thus two ways to model a gravity assist are available, with one of them being an approximation. The approximation can be used for broad searches that may include multiple flybys and accuracy is not critical. The integrated flybys are used for more accurate trajectories.

Both types of velocity impulses are treated the same way, except for the mandatory constraint imposed on the gravity assist impulse and the fact that the gravity assist impulse does not have an associated mass depletion. To model a ‘powered’ gravity assist flyby where in addition to the gravity assist impulse there is an additional maneuver based impulse, two segments are used. One of the segments can be a simple node with an impulse representing the gravity assist, and the other segment has an impulse representing the maneuver. If the epochs for both segments are the same, the order of the segments is not important.

The impulse at either node can have zero magnitude. After the impulse, the velocity vector is

\[
\mathbf{v}_{(0,f)}^{i+} = \mathbf{v}_{(0,f)}^{i-} + \Delta \mathbf{v}_{(0,f)}
\]  

where the superscript ‘\(+\)’ specifies the value after the impulse.

The evolution of the mass value across a node depends on three allowable and distinct mass discontinuities. The first mass discontinuity is a non-maneuver mass discontinuity that can result from either a mass drop off or a mass add on, so it can be positive or negative and is labeled as \( \Delta m_{(0,f)} \). In other words, a spacecraft component can be discarded prior to the impulse. Or, if the spacecraft has performed a rendezvous with another spacecraft and has captured it, a positive mass discontinuity represents that additional mass associated with this capture. The mass value after this first non-maneuver mass discontinuity is

\[
m_{(0,f)}^{i+} = m_{(0,f)}^{i-} + \Delta m_{(0,f)}
\]  

where \( \Delta m_{(0,f)} \) is the value of the mass change. Proceeding with the impulse, the mass value after the maneuver is

\[
m_{(0,f)}^{i+} = m_{(0,f)}^{i-} + \Delta m_{(0,f)}
\]
where $\Delta m_i^{(0,f)}$ is the mass change that results from the instantaneous depletion of propellant that results from a maneuver based impulse. It is zero for a gravity assist impulse. The maneuver based mass change is directly related to the magnitude of the maneuver, $\Delta v_i^{(0,f)}$, and the exhaust velocity of the engine used to provide this maneuver, $c_i^{(0,f)}$. The maneuver mass discontinuity is obtained from a form of the rocket equation,

$$\Delta m_i^{(0,f)} = m_i^{(0,f)} \left( e^{-\Delta v/c} - 1 \right)$$  \hspace{1cm} (4)

where the superscripts and subscripts on $\Delta v$ and $c$ have been omitted for notational simplicity. The exhaust velocity by definition is related to the specific impulse of the propellant used and the reference gravity acceleration value at the Earth’s surface, $c_i^{(0,f)} = \left( I_{sp_i}^{(0,f)} \right) \left( g_{earth} \right)$.

Following the impulse, another non-maneuver mass discontinuity with the same characteristics as the one prior to the impulsive maneuver is allowed. The mass value at the end of either node is then

$$m_i^{(0,f)} = m_i^{(0,f)} - \Delta m_i^{(0,f)} + \Delta m_i^{(0,f)}$$  \hspace{1cm} (5)

where $\Delta m_i^{(0,f)}$ is the post impulse, non-maneuver mass discontinuity. Typically this mass change results when the engine stage used to produce the impulsive maneuver is discarded. However, allowance is made so that a mass add on can occur at this point again for reasons associated with a spacecraft capture. The reason non-maneuver mass discontinuities are allowed on either side of the impulse at either node is because in the case of a spacecraft rendezvous and capture, the maneuver performing the rendezvous could occur at the initial node point of a given segment, or at the final node point of a previous segment to which the given segment is connected to.

Based on the evolution of the velocity and mass across a node point, the following distinct times labels that are equal in value but are used to distinguish the values of both the velocity and the mass are

- $t_{(0,f)}^{-}$: node initial time, and time prior to any velocity or mass discontinuities
- $t_{(0,f)}^{+}$: time after first non-maneuver mass discontinuity
- $t_{(0,f)}^{-}$: time after the velocity impulse
- $t_{(0,f)}^{+}$: node final time, and time after the second non-maneuver mass discontinuity

The arc that connects both node points can be either a ballistic arc with no independent control, or a controlled arc with thrust or acceleration controls. The control is provided by a thrust vector from an engine or a controlled acceleration that results from a non-mass depleting device such as a sail using radiation from a photon emitting celestial body such as the Sun or a star, or any other external momentum transfer device. Any system used to provide this control will be referred to as a propulsion system, though some may not use propellant such as in the case of a sail. The arc connecting the $t_0$ node to the $t_f$ node of a segment $i$ is defined by both the parameters of the propulsion system and the equations of motion. The control vector, $\Gamma(t)$, will in general have the following functional dependence,

$$\Gamma' = \Gamma'(c(t), T(t), P(t), \varepsilon, u(t), m(t), a_p, t)$$  \hspace{1cm} (6)
where \( c \) is the exhaust velocity, \( T \) is the thrust, \( P \) is the power, \( \varepsilon \) is the efficiency of the propulsion system, \( \mathbf{u} \) is the control direction unit vector, \( m \) is the instantaneous mass, and \( \mathbf{a}_p \) is a vector containing any other parameter that defines the propulsion system. For example, in the case of a sail, \( \mathbf{a}_p \) will contain parameters such as sail area, surface reflectivity, and other parameters that uniquely define it.

For any segment \( i \), the dynamic state of the spacecraft along the arc for times between \( t_{i0}^+ \) and \( t_{if}^- \) is defined as an augmented state vector comprised of its position, velocity, and mass \(( \mathbf{r}(t) \quad \mathbf{v}(t) \quad m(t) )^\top \) and satisfies the first order vector equation of motion,

\[
\begin{pmatrix}
\dot{\mathbf{r}}^i \\
\dot{\mathbf{v}}^i \\
\dot{m}^i
\end{pmatrix}^i = \begin{pmatrix}
\mathbf{v}(t) \\
\mathbf{g}(\mathbf{r}, \mathbf{v}, m, t, \mathbf{a}_g) + \Gamma(t) \\
-T(t)/c(t) + \dot{m}_p
\end{pmatrix}^i
\]  

(7)

Here, \( \mathbf{g}(\mathbf{r}, \mathbf{v}, m, t, \mathbf{a}_g) \) is the acceleration per unit mass resulting from control independent terms such as the acceleration due to the gravitational potential from any celestial body, radiation pressure from energy emitting celestial bodies, or nonconservative forces such as atmospheric drag. The symbol ‘\( g \)’ commonly has been used to represent the gravitational acceleration due to gravitating celestial bodies with a dependence only on position and time. Here it is generalized to also include a dependence on velocity, \( \mathbf{v} \), mass, \( m \), and an additional problem specific acceleration, \( \mathbf{a}_g \). The acceleration vector \( \mathbf{a}_g \) in the acceleration vector contains all of the constant or variable parameters of the force model such as the gravitational constant, the mass of the celestial bodies, the non-spherical gravitational potential field of the celestial body, the atmospheric parameters of the celestial body, the radiation parameters of any energy emitting celestial body, and any other term that may appear in the ballistic acceleration of the spacecraft. This acceleration may be time dependent. The time dependency of the force field results in part from the motion of the celestial bodies in the model or the rotation of a body about an internal axis if the non spherical gravitational potential of a celestial body is used. The mass rate, \( \dot{m}^i \), is a result of fuel consumed during an engine burn and is related to the thrust, \( T \), and exhaust velocity \( c \), or any other arbitrary mass depletion that results from intentional or unintentional continuous venting of liquids or gasses, or mass accumulation during flight through resisting a medium such as dust clouds or atmospheres. The non-thrust contribution to \( \dot{m}^i \) is given by the general term \( \dot{m}_p \). In summary, the spacecraft equations of motion between node points at \( t_{i0} \) and \( t_{if} \) are completely general and arbitrary, but known.

For any segment, the parameters for three propulsion systems need to be specified: one for the initial impulsive maneuver, one for the controlled arc, and one for the final maneuver. If the impulsive maneuvers are due to a gravity assist, then the propulsion system parameters for these are unimportant. The propulsion system parameters can all be defined for a single propulsion system, i.e., the exhaust velocity for the initial and final maneuver based impulses, and the controlled arc, assuming it is a constant exhaust velocity system, can be defined to be the same system by equating the exhaust velocity values of each segment node and the exhaust velocity of the controlled arc. But since these can be independent, multiple propulsion systems can be used in
one individual segment. If the segment is a pure coast or ballistic arc then all propulsion system parameters are set to zero. If a segment is initiated with a high thrust booster, whose maneuver can be approximated with an impulsive maneuver, and controlled arc of the segment uses an independent low thrust propulsion system, then the arc requires the definition of the initial node impulsive maneuver and the parameters defining the low thrust engine.

Another key entity of the basic segment is referred to as floating node point. It is similar to the two final endpoint nodes in that it has an associated time tag denoted as \( t_i \) that is required to lie between \( t_0 \) and \( t_f \),

\[
t_0 \leq t_i \leq t_f
\]  

(8)

The main restriction of the floating node point is that it cannot have any velocity or mass discontinuities. It is used as a position, velocity, and mass reference measured along the segment. Since it is a node, the state vector associated with it can be constrained, or it can serve as a constraint for other segments. For example, a rendezvous between a spacecraft on segment \( i \) and a spacecraft on segment \( k \) can occur at the floating node point of segment \( k \). The boundary conditions would be \( \mathbf{r}_i^f = \mathbf{r}_k^i \) and \( \mathbf{v}_i^f = \mathbf{v}_k^i \). Another example for the use of the floating endpoint is to find the location along a segment \( i \) that is at the periapsis point with respect to a celestial body. Here, \( t_i \) is a free parameter constrained by Eq. 8 and its value must be such that \( \dot{r}_i(t_i) = 0 \) and \( \ddot{r}_i(t_i) > 0 \), where \( \dot{r}_i(t_i) \) is the radial velocity with respect to the celestial body, and \( \ddot{r}_i(t_i) \) is the second time derivative of the position magnitude. Both of these conditions are sufficient for finding the location of periapsis. These types of constraints could be imposed without the need to define a floating node point because either endpoint nodes could be used to serve the same purpose. However, including the floating node point, simplifies the modelling process by removing the need to include an additional segment for some types of problems.

A sketch of the basic segment is given in Figure 1. Figure 2 shows the exploded time scale for the segment and the locations in time of the discontinuous and continuous states of the spacecraft. A subset of the parameters that define the basic segment are listed in Table 1. Within an individual segment, some of these quantities are independent and some are dependent. Outside of the segment, as in a trajectory or a mission, any of these quantities can be constrained and will thus be dependent in the solution process.

Figure 1: The basic segment Building Block. Velocity impulses can exist at either node and the arc connecting the nodes can be either a ballistic or an controlled arc with a time dependent variable control vector. Mass discontinuities can exist at either node that result from an impulsive maneuver or non-maneuver mass changes. A floating node point lies between the endpoint nodes.

A sketch of the basic segment is given in Figure 1. Figure 2 shows the exploded time scale for the segment and the locations in time of the discontinuous and continuous states of the spacecraft. A subset of the parameters that define the basic segment are listed in Table 1. Within an individual segment, some of these quantities are independent and some are dependent. Outside of the segment, as in a trajectory or a mission, any of these quantities can be constrained and will thus be dependent in the solution process.
Figure 2: Exploded View Representation of the basic segment. Mass discontinuities can exist at each node point. These are either stage drop offs, mass additions, or impulsive maneuver mass discontinuities. The non maneuver mass discontinuities occur on either side of the velocity impulses. The velocity impulses are due to either an impulsive maneuver or an approximated gravitational assist. The velocity before an impulsive is $v^-$, the velocity after an impulsive maneuver is $v^+$. The maneuver based impulses have an associated mass discontinuity, $\Delta m$. The non-maneuver mass discontinuities are $\Delta m^-$ if it occurs before the impulse, and $\Delta m^+$ if it occurs after the impulse. At either node point, the beginning mass value is $m^-$. The mass value prior to a velocity impulse is $m^- +$. The mass value after a velocity impulse is $m^+ -$. The mass value after the second non-maneuver mass discontinuity is $m^+$. The controlled accelerated are is between $t^+_0$ and $t^-_f$. All state variables vary continuously along this arc. The simplest type of segment is a node point at $t_0$ with all parameters set to zero except for the node state variables, $r_0, v_0, m_0$ which are required to be defined.

The basic segment also includes a Lagrange multiplier vector, $\lambda$, adjoined to the physical state variables, $r, v, m$ at each time instant associated with each node and between $t^+_0$ and $t^-_f$. For $t^+_0 \rightarrow t^-_f$, the evolution of Lagrange multiplier vector is governed by the Euler-Lagrange differential equations associated with the optimal control problem. Other names given to this vector include the adjoint vector and the costate vector. This vector is composed of the position costate vector, $\lambda_r$, the velocity costate vector $\lambda_v$, and the mass costate scalar, $\lambda_m$.

Additionally, because of the basic segment definition, the segments can be all independent and initially disconnected. The dependency a segment may have with respect to another segment is specified in the constraint definitions. Because of this, the architecture allows then the modeling of multiple spacecraft problems. It is not necessary to specify how many spacecraft are in the model. All that is required is to model as many segments as needed and that they be constrained in any way necessary to represent the number of spacecraft in the model. For example, a spacecraft rendezvous between two
spacecraft requires a minimum of two segments. A possible constraint is that the final endpoint nodes of both spacecraft have the same position and velocity at the final solution. If both spacecraft have maneuvering capability, these maneuvers are then defined and allowed to be adjusted for each segment. If one spacecraft remains passive, then its segment definition has its propulsion parameters nulled out.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Variable Type/Name</th>
<th>Dim.</th>
<th>Dep.</th>
<th>Range or Constraint</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$</td>
<td>initial time</td>
<td>1</td>
<td>i</td>
<td>$-\infty &lt; t_0 &lt; \infty$</td>
<td>day</td>
</tr>
<tr>
<td>$r_0$</td>
<td>position</td>
<td>3</td>
<td>i</td>
<td>$0 \leq</td>
<td>r_0</td>
</tr>
<tr>
<td>$v_{-0}$</td>
<td>velocity</td>
<td>3</td>
<td>i</td>
<td>$0 \leq</td>
<td>v_{-0}</td>
</tr>
<tr>
<td>$m_{-0}$</td>
<td>mass</td>
<td>1</td>
<td>i</td>
<td>$0 \leq m_{-0} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\Delta m_{-0}$</td>
<td>mass discontinuity</td>
<td>1</td>
<td>i</td>
<td>$-\infty &lt; \Delta m_{-0} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\lambda_{-0}$</td>
<td>Lagrange multiplier</td>
<td>7</td>
<td>i</td>
<td>$0 &lt;</td>
<td>\lambda_{-0}</td>
</tr>
<tr>
<td>$\Delta v_{0}$</td>
<td>velocity impulse</td>
<td>3</td>
<td>i</td>
<td>$0 \leq</td>
<td>\Delta v_{0}</td>
</tr>
<tr>
<td>$I_{sp_{0}}$</td>
<td>specific impulse</td>
<td>1</td>
<td>i</td>
<td>$0 \leq I_{sp_{0}} &lt; \infty$</td>
<td>sec</td>
</tr>
<tr>
<td>$\Delta m_{0}$</td>
<td>maneuver mass discontinuity</td>
<td>1</td>
<td>d</td>
<td>$0 \leq \Delta m_0 &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\Delta m_{+0}$</td>
<td>mass discontinuity</td>
<td>1</td>
<td>i</td>
<td>$-\infty \leq \Delta m_{+0} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$v_{+0}$</td>
<td>velocity</td>
<td>3</td>
<td>d</td>
<td>$0 \leq</td>
<td>v_{+0}</td>
</tr>
<tr>
<td>$m_{+0}$</td>
<td>mass</td>
<td>1</td>
<td>d</td>
<td>$0 \leq m_{+0} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\lambda_{+0}$</td>
<td>Lagrange multiplier</td>
<td>7</td>
<td>d</td>
<td>$-\infty \leq</td>
<td>\lambda_{+0}</td>
</tr>
<tr>
<td>$t_i$</td>
<td>intermediate time</td>
<td>1</td>
<td>i</td>
<td>$t_0 \leq t_i \leq t_f$</td>
<td>day</td>
</tr>
<tr>
<td>$T$</td>
<td>thrust</td>
<td>1</td>
<td>d</td>
<td>$0 \leq T \leq T_{\text{max}} &lt; \infty$</td>
<td>Newtons</td>
</tr>
<tr>
<td>$I_{sp_i}$</td>
<td>specific impulse</td>
<td>1</td>
<td>i</td>
<td>$0 \leq I_{sp_i} \leq I_{sp_{\text{max}}} &lt; \infty$</td>
<td>sec</td>
</tr>
<tr>
<td>$u$</td>
<td>thrust direction unit vector</td>
<td>3</td>
<td>i</td>
<td>$</td>
<td>u</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>engine efficiency</td>
<td>1</td>
<td>i</td>
<td>$0 \leq \varepsilon \leq 1$</td>
<td>—</td>
</tr>
<tr>
<td>$t_f$</td>
<td>final time</td>
<td>1</td>
<td>i</td>
<td>$\infty &lt; t_f &lt; \infty$</td>
<td>day</td>
</tr>
<tr>
<td>$r_f$</td>
<td>position</td>
<td>3</td>
<td>d</td>
<td>$0 \leq</td>
<td>r_f</td>
</tr>
<tr>
<td>$v_{-f}$</td>
<td>velocity</td>
<td>3</td>
<td>d</td>
<td>$0 \leq</td>
<td>v_{-f}</td>
</tr>
<tr>
<td>$m_{-f}$</td>
<td>mass</td>
<td>1</td>
<td>d</td>
<td>$0 \leq m_{-f} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\Delta m_{-f}$</td>
<td>mass discontinuity</td>
<td>1</td>
<td>i</td>
<td>$-\infty \leq \Delta m_{-f} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\lambda_{-f}$</td>
<td>Lagrange multiplier</td>
<td>7</td>
<td>d</td>
<td>$-\infty \leq</td>
<td>\lambda_{-f}</td>
</tr>
<tr>
<td>$\Delta v_{f}$</td>
<td>velocity impulse</td>
<td>3</td>
<td>i</td>
<td>$0 \leq</td>
<td>\Delta v_{f}</td>
</tr>
<tr>
<td>$I_{sp_f}$</td>
<td>specific impulse</td>
<td>1</td>
<td>i</td>
<td>$0 \leq I_{sp_{f}} \leq \infty$</td>
<td>sec</td>
</tr>
<tr>
<td>$\Delta m_{f}$</td>
<td>maneuver mass discontinuity</td>
<td>1</td>
<td>d</td>
<td>$0 \leq \Delta m_f &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\Delta m_{+f}$</td>
<td>mass discontinuity</td>
<td>1</td>
<td>i</td>
<td>$-\infty \leq \Delta m_{+f} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$v_{+f}$</td>
<td>velocity</td>
<td>3</td>
<td>d</td>
<td>$0 \leq</td>
<td>v_{+f}</td>
</tr>
<tr>
<td>$m_{+f}$</td>
<td>mass</td>
<td>1</td>
<td>d</td>
<td>$0 \leq m_{+f} &lt; \infty$</td>
<td>kg</td>
</tr>
<tr>
<td>$\lambda_{+f}$</td>
<td>Lagrange multiplier</td>
<td>7</td>
<td>d</td>
<td>$-\infty \leq</td>
<td>\lambda_{+f}</td>
</tr>
</tbody>
</table>

Table 1: Basic set of Elements Defining the Basic Segment
Since the segments can be initially disconnected, the introduction of additional variables and possible constraints, though increases the number of variables, also decreases the sensitivity of the iterated solutions compared to strictly forward time or backward time propagated trajectories if the magnitude of the duration of a segment is longer than allowed for perturbations to be linearly valid in gradient based solution methods. However, if this is not an issue, any segment $j$ can be forced apriori to be connected sequentially to any other segment $i$ at any of the node points and distinct times of segment $i$. The idea of propagating disconnected segments that will eventually be connected or at least constrained in some way is a generalization of what is known as direct multiple shooting.

The definition of the basic segment described here can be altered to include more parameters. The working model described has been sufficient to examine a broad class of trajectory problems. The definition of the basic segment, however, is dynamic so that yet unforeseen trajectory problems can be accommodated. The simplest mission can be modelled with only one segment. More complex missions will require more than one segment.

2.1.1 Segment Types

The general definition of the segment allows for the modeling of different segment types. The type and number of segments used depends on the nature of the problem, the number of spacecraft in the model, the propulsion systems used in the model, and the complexity of the trajectory or mission. Based on the basic segment, the segment types that can be modelled, by properly specifying the values that define it, are listed in Table 2. Figure 3 shows the different segment types beginning with the most basic segment type and proceeding down to the simplest segment models which are simple node points or node points with impulsive maneuvers. Each segment type in Figure 3 is labeled with an integer identifier that corresponds to those listed in Table 2.

<table>
<thead>
<tr>
<th>Segment Type</th>
<th>Segment Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$t_0$ node, $\Delta v_0$, controlled arc, $t_f$ node, $\Delta v_f$</td>
</tr>
<tr>
<td>2</td>
<td>$t_0$ node, $\Delta v_0$, controlled arc, $t_f$ node</td>
</tr>
<tr>
<td>3</td>
<td>$t_0$ node, controlled arc, $t_f$ node, $\Delta v_f$</td>
</tr>
<tr>
<td>4</td>
<td>$t_0$ node, controlled arc, $t_f$ node</td>
</tr>
<tr>
<td>5</td>
<td>$t_0$ node, $\Delta v_0$, ballistic arc, $t_f$ node, $\Delta v_f$</td>
</tr>
<tr>
<td>6</td>
<td>$t_0$ node, $\Delta v_0$, ballistic arc, $t_f$ node</td>
</tr>
<tr>
<td>7</td>
<td>$t_0$ node, ballistic arc, $t_f$ node, $\Delta v_f$</td>
</tr>
<tr>
<td>8</td>
<td>$t_0$ node, ballistic arc, $t_f$ node</td>
</tr>
<tr>
<td>9</td>
<td>$t_0$ node, $\Delta v_0$</td>
</tr>
<tr>
<td>10</td>
<td>$t_0$ node</td>
</tr>
<tr>
<td>11</td>
<td>$t_f$ node, $\Delta v_f$</td>
</tr>
<tr>
<td>12</td>
<td>$t_f$ node</td>
</tr>
</tbody>
</table>

Table 2: Segment Types as Sub Types of the Basic Segment

Recall that a trajectory for a single spacecraft is comprised of one or more segments. Therefore, it is possible to model a trajectory for a single spacecraft that is equipped with one or more propulsion systems. Further, mass discontinuities resulting from either impulsive maneuvers, mass drop offs, and mass
Figure 3: Segment Types Based on the basic segment. These are treated as building blocks to construct many types of trajectories for one or more spacecraft. The large vector arrows represent velocity impulses; the small vector arrows represent a finite acceleration control arc. Dynamically there is no difference between segment types 9 and 11 or 10 and 12 except that the time epochs are different since segment types 9 and 10 are the $t_0$ node and segment types 11 and 12 are the $t_f$ node, with and without a velocity impulse as shown.

Additions are accounted for by properly defining these quantities in the definition of the segment. A mission which is composed of one or more trajectories, which in turn is composed of one or more segments, will have as its basic building block any of the segment types shown in Figure 3 and tabulated in Table 2. The choice of which segments need to be connected to which other set of segments is problem dependent. In the case of multiple spacecraft problems, the connectivity of the segments is determined by the type of multiple-spacecraft mission being analyzed. For independent spacecraft that are not required to rendezvous or intercept, these may all be disconnected but constrained in any way necessary, such as maintaining their positions to satisfy some geometrical formation, for example. For intercept or rendezvous problems, clearly, some segments will need to be connected; i.e., for rendezvous problems, position and velocity are required to be continuous, or for intercept problems, only position needs to be continuous after the maneuver. A rendezvous between two spacecraft where one of the spacecraft assumes the additional mass of the other spacecraft will require the addition of a non-maneuver mass discontinuity which can be a fixed or variable quantity depending on whether mass of the captured spacecraft is fixed or variable.

Impulsive maneuvers, which are constrained to occur only at the endpoints, can be referenced to any allowable fixed-center frame or relative to the trajectory state associated with the respective node point allowing either velocity vector or radius vector referenced maneuvers. Finite acceleration maneuvers (finite engine burns, sails, etc.) can also be steered relative to the trajectory or remain fixed in any fixed-centered frame. In the case of finite burn maneuvers, both thrust constrained and power constrained thrust arcs can be modelled.
Algorithmically, no distinction is made between finite burn, high thrust or low thrust systems. Both are treated equally and the specified propulsion system parameters further defines the type of thrust arc used. The variation of the thrust magnitude, the exhaust velocity, and the power are constant or variable depends on how each of these are constrained.

2.2 A Conceptual Modelling Example

To illustrate the use of using the segment building blocks to construct a complex mission, consider the following hypothetical mission. It is desired to transfer a main spacecraft stationed at the vicinity of the interior libration point of the Earth-Moon system to the vicinity of the Jovian moon, Europa. A landing and ascent vehicle of fixed and known initial mass is attached to the main spacecraft. At Europa the lander separates from the main spacecraft, descends to the surface of Europa, performs the required objectives, and ascends to rendezvous with the main spacecraft. The main spacecraft returns to its starting point in the Earth-Moon system, but on its return route, it intercepts and flies by a comet. The main spacecraft is equipped with a nuclear powered continuous thrust, variable specific impulse engine. The lander/ascent vehicle is equipped with a high thrust constant specific impulse engine with fixed propellant mass. The mission performance objective is to minimize the initial mass of the main spacecraft while constraining the final mass (the dry mass) of the main spacecraft. The total mission duration is constrained to be no greater than a given value and the stay time for the lander at Europa is constrained to be no less than a given value. The mission can utilize any beneficial gravity assist maneuvers around celestial bodies that exist in the model. Clearly, there are many more details required to fully model this mission. However, the basic parameters stated are enough to conceptually illustrate the modeling of this type of mission. Figure 4 illustrates a possible set of segment blocks that can be used to construct the first iterate. The short dashed lines indicate the node points that need to be connected and required to be continuous at least in position and possibly velocity. In this sketch, the first assumption is that the mission can be modelled with ten segments. Other parts of the mission, such as the landing and ascent phases can be decomposed into further segments if necessary. To correctly design and optimize such a mission, it will be necessary to include the gravitational attraction from at least the principle celestial bodies involved which in this case include the Sun, Earth, Moon, Jupiter, and Europa. The comet can be assumed to be a non-gravitating celestial body and can be modelled as an independent segment if its state vector is known at some epoch, or it can be treated as a gravitating celestial body if its ephemeris is known. Further, it will be necessary to specify the initial state or orbit of the main spacecraft with respect to a reference frame in which the location of the interior libration point of the Earth-Moon system is known. At departure, the trajectory may take advantage of the complex force field that is dominated by the simultaneous attraction from the Sun, Earth, and Moon. At Jupiter, the trajectory may also exhibit a complex behavior since the force field will be dominated by a restricted four body model that includes the Sun, Jupiter, and Europa. Though a solution to this mission is not presented here, it is one example of a type of mission that can be modelled and solved with the system described.
Figure 4: Modelling of a Complex Mission. Different segment types are used to construct a round trip mission beginning at the interior libration point of the Earth-Moon system, to the Jupiter-Europa System, with an intermediate flyby of a comet, before returning to the starting point. Some segments will be low thrust arcs, others will be ballistic arcs with impulsive maneuvers at their nodes. All segments are shown initially disconnected. Some of these could have been defined to sequentially connected to other segments initially. The dashed lines between some of the nodes imply that these nodes are required to be connected at the final solution.

2.3 The Equations of Motion and the Propagation Reference Frame

For a general system, the choice of reference frame in which to model the spacecraft’s three degree of freedom point dynamics and the associated equations of motions is critical from several viewpoints. First, it is desired to have the simplest reference frame model allowable without compromising the validity of a solution and its accuracy. Second, in a comprehensive model, that may include multiple spacecraft, multiple propulsion systems, and multiple celestial bodies, it is desirable to work with a coordinate frame for state propagation in which transformations between the frames in which the states, targets, and maneuvers are referenced to are simple and computed efficiently. If the dynamics are modelled correctly, any reference frame used must yield the same result. On the other hand, some reference frames may include terms in the equations of motion which approach the limiting accuracy with which small terms are evaluated by the computer. Aware of the accuracy issues associated with the proper choice of reference frame for state propagation, the current system propagates the equations of motion in either a reference frame that is fixed (non-rotating) to any barycentered frame that can be defined or to any chosen celestial body. If it is attached to a specified celestial body, the frame translates but does not rotate. The celestial body selected to be the center of this frame depends on the nature of the problem being modelled.
Each segment in the mission can have a distinct reference frame for propagation if desired. The only restriction is that it be a non-rotating frame. For example, if interplanetary missions are being solved with respect to the Sun, and the other celestial bodies of interest are treated as non gravitating zero point masses, the reference frame should be centered at Sun or the solar system barycenter. For the same mission, if some of the segments are required to operate in the vicinity of a celestial body, such as in the case of a long duration spiral escape or capture from a given celestial body, then that segment should have as its propagation frame, a frame fixed to the celestial body of interest. In some problems, state definition and targeting can be specified in coordinates of a rotating frame defined between two celestial bodies. Though it is possible to integrate the equations in coordinates of the rotating frame, experimentation has so far determined that this does not provide any significant advantages. However, the use of rotating frame coordinates for state definition and targeting is a key component of the current architecture. The discussion of the rotating frames used for this purpose are reserved for a later section.

In a barycentered frame that is assumed to be inertial the second order equations of motion are

\[
\ddot{\mathbf{r}} = \sum_{j=1}^{n_{cb}} GM_j \frac{\mathbf{r} - \mathbf{r}_j}{|\mathbf{r} - \mathbf{r}_j|^3} + \mathbf{a}_{pert} + \mathbf{\Gamma}(t) \tag{9}
\]

where \( \mathbf{r} \) is the relative position vector of the spacecraft with respect to barycentered frame, \( \mathbf{r}_j \) is the relative position of celestial body \( j \) with respect to the barycentered frame, \( n_{cb} \) is the number of celestial bodies in the model, and \( GM_j \) is gravitational parameter for celestial body \( j \). All additional non-control related terms are contained in \( \mathbf{a}_{pert} \) and control related terms are contained in \( \mathbf{\Gamma} \).

If the propagation reference frame is centered at a specified celestial body, denoted by \( CB_f \), the equations of motion in this fixed (non-rotating) but translating frame is

\[
\ddot{\mathbf{r}} = -\frac{GM_{CB_f}}{r^3} \mathbf{r} - \sum_{j=1}^{n_{cb}} GM_j \left[ \frac{\mathbf{r} - \mathbf{r}_j}{|\mathbf{r} - \mathbf{r}_j|^3} + \frac{\mathbf{r}_j}{r_j^3} \right] + \mathbf{a}_{pert} + \mathbf{\Gamma}(t) \tag{10}
\]

where \( GM_{CB_f} \) is the gravitational parameter of the celestial body to which the frame is fixed, \( \mathbf{r}_j \) is the position vector of the other celestial bodies with respect to \( CB_f \), and \( n_{cb} \) is the number of celestial bodies treated as third bodies.

The vector terms associated with the gravitational acceleration due to third bodies include both the direct acceleration vector term and the indirect acceleration term needed to account for the fact the \( CB_f \) fixed centered frame is not an inertial frame. Eqs. 9 or 10 are used to propagate all of the segments from \( t_i^+ \rightarrow t_i^- \) (\( i = 1, \ldots, n \)) where \( n \) is the number of segments in the mission and noting that each segment can use either set of equations and have its own distinct choice for the center of the reference frame.

The position of the other celestial bodies is obtained from an explicit time dependent ephemeris and hence the time dependence of the force model. The ephemeris provides the positions, and possibly the velocity, of any of the celestial bodies with respect to any other celestial body and can be a highly
accurate ephemeris, such as the Jet Propulsion Laboratory’s set of planetary ephemerides, or any user defined ephemeris. The user defined ephemeris is useful when studying the dynamics in simplified force models such as the circular or elliptic restricted three body problem; or user defined model solar systems with one or more arbitrary celestial bodies where the motion of the celestial bodies relative to the frame center can be simple Keplerian type orbits or based on a precomputed ephemeris.

For example, consider the analysis of cycling trajectories between two planets orbiting the Sun. Such a study should begin by assuming that the planets of interest are in circular orbits about the Sun. This model provides the basic properties associated with the construction of these trajectories. Once these properties have been understood, then the model is modified to account for the real motion of the planets. The assumption being that the circular model solutions carry over to the real planetary model with an expected variation. However, it is important to be aware that the model that uses the real planetary model may yield solutions that do not exist in the simplified circular model. If so, then a better approximation to the real planetary model should be used to identify the basic properties of cycling trajectories between these celestial bodies, such as an elliptic model for the motion of the planets of interest.

Current and future missions will require targeting relative to one or more solar system minor bodies such asteroids and comets. Assuming an ephemeris for these bodies is available, then these are treated as celestial bodies with an associated gravitational parameter. If only a state vector and epoch is known for any of these bodies then either an ephemeris is computed separately with the current system by treating it as a ballistic segment, or the body is treated as a ballistic segment in the solution process by assuming that its mass does not influence the motion of the other spacecraft segments.

2.4 Multi-Body Reference Frames

There are certain spacecraft trajectory design and optimization problems that are best solved in alternate reference frames. In the case where two or more celestial bodies are included in the force model and any two of these bodies are in bounded orbits relative to each other, the dynamics of a spacecraft is best analyzed and understood in a reference frame that can in general translate, rotate, and pulsate in such that the representation of the positions of the celestial bodies remain stationary. This frame is the well known rotating frame that has been used extensively in the study of trajectories in the Circular and Elliptic Restricted Three Body problems[10]. For example, consider the motion of a spacecraft in the Earth-Moon system. Analogous to the barycentered rotating frame of the restricted three body problem, a convenient frame in which both the Earth and Moon remain stationary is one that rotates with the instantaneous Earth-Moon line and pulsates with the varying distance between the Earth and Moon if the motion of the Moon about the Earth is based on a realistic ephemeris.

Another example of the use of a rotating frame is in the analysis of Earth-Mars trajectories where a suitable reference frame is one where both Earth and Mars remain stationary. Here, the fundamental axis is the instantaneous Earth-Mars line. The north pole axis of the Earth or the Sun can be used
to construct a normal vector to this line, from which the final right-handed coordinate frame is constructed. The advantage of this frame is clearly noted in the visualization and targeting of trajectories connecting both planets but more importantly, the capability then exists to specify initial state components and constraint functions directly in this frame greatly simplifying the targeting and optimization of trajectories relative to any set of moving bodies.

These two example reference frames can be examined under a single formulation. In the first example, two celestial bodies are in orbit about their common barycenter. In the other example two celestial bodies are orbiting a common reference center that is not their barycenter. The state, targeting, and maneuver definitions can be given in coordinates of this frame and appropriate transformations are needed to transform position, velocity, and acceleration between this frame and the segment propagation frame.

Consider the motion of two celestial bodies $CB_i$ and $CB_j$. The time varying position vectors for these bodies are known relative to some other fixed frame that can be an inertial frame or a frame centered at another body; i.e., $r_{CB_i}$ and $r_{CB_j}$ are known vector functions of time. Define a unit vector along the relative position vector of $CB_j$ with respect to $CB_i$

$$\hat{r} \equiv \frac{(r_{CB_j} - r_{CB_i})}{|r_{CB_j} - r_{CB_i}|} \quad (11)$$

Several options exist to define the remaining two basis vectors for the frame. If $CB_j$ is in a closed and bounded orbit about $CB_i$ such that its relative angular momentum vector with respect to $CB_i$ remains nearly constant, then the remaining basis unit vectors are defined as

$$\hat{s} \equiv \hat{t} \times \hat{r} \quad (12)$$

$$\hat{t} \equiv \frac{(r \times v)}{|r \times v|} \quad (13)$$

where the $\hat{t}$ unit vector is along the instantaneous specific angular momentum vector and $v$ is the relative velocity vector of $CB_j$ with respect to $CB_i$, $v = v_{CB_j} - v_{CB_i}$. This basis can be used, for example if the two bodies are a planet and the Sun, or a planet and its moon. Figures 5 and 6 illustrate the $CB_i$-$CB_j$ rotating frames in both a $CB_k$ fixed frame and a rotating frame centered at one of the celestial bodies.

Alternatively, if the relative angular momentum vector between the bodies changes in way that the motion alternates between being retrograde to prograde (such as the motion of two celestial bodies around a central body that are not in orbit about each other) then it is convenient to use a nearly constant vector, such as a reference north pole axis to define the remaining basis unit vector. Let this reference unit vector be $\hat{z}$. The remaining unit vectors then are

$$\hat{s} \equiv \frac{(\hat{z} \times \hat{r})}{|\hat{z} \times \hat{r}|} \quad (14)$$

$$\hat{t} \equiv \hat{r} \times \hat{s} \quad (15)$$

This set of basis vectors is ideally suited for a rotating-pulsating frame where the Earth is $CB_i$ and Mars is $CB_j$ (or vice-versa). Figures 7 and 8 illustrate the $CB_i$-$CB_j$ rotating-pulsating frames in both a $CB_k$ fixed frame and a rotating-pulsating frame centered at one of the celestial bodies.
The two celestial bodies, CB\textsubscript{i} and CB\textsubscript{j}, are in closed orbits about their common barycenter. In general, the distance between them varies.

The transformation of a position vector referenced in the rst frame to a fixed frame centered at CB\textsubscript{i} with basis unit vectors i, j, k (referred to as the ijk frame) and scaled with respect to the instantaneous distance between CB\textsubscript{j} and CB\textsubscript{i} is

$$r_{ijk} = \frac{r}{k}R_{rst}$$

where \(r = |r|\), \(k\) is a positive scaling constant with the same units as \(r\), and \(R\) is the transformation direction cosine matrix between the rst frame and the ijk frame,

$$R = \begin{pmatrix} i \cdot \hat{r} & i \cdot \hat{s} & i \cdot \hat{t} \\
 j \cdot \hat{r} & j \cdot \hat{s} & j \cdot \hat{t} \\
 k \cdot \hat{r} & k \cdot \hat{s} & k \cdot \hat{t} \end{pmatrix}$$

The fixed frame velocity and acceleration vectors are obtained by successive time differentiation of Eq. 16,

$$v_{ijk} = \frac{1}{k} \left[ (\dot{r}R + r\dot{R}) r_{rst} + rRv_{rst} \right]$$

$$a_{ijk} = \frac{1}{k} \left[ (\ddot{r}R + 2\dot{r}\dot{R} + r\ddot{R}) r_{rst} + \left(2\dot{r}\dot{R} + 2r\ddot{R}\right) v_{rst} + rRa_{rst} \right]$$

where \(\dot{r} = dr/dt\) and \(\ddot{r} = d^2r/dt^2\). The inverse transformation that provides expressions for \(r^{rst}, v^{rst}, \) and \(a^{rst}\) in terms of \(r^{ijk}, v^{ijk}, \) and \(a^{ijk}\) is readily available. Note that the higher order time derivatives for \(r\) and \(R\) will require up to a first order time derivative in the relative acceleration vector between CB\textsubscript{j} and CB\textsubscript{i}. If this information is not available from the ephemeris, then it will need to be estimated with any finite difference approximation provided at least the time dependent position vectors are available from the ephemeris.

In either the rotating or the rotating-pulsating frame, it becomes possible to define initial state components for any segment and to target and treat as constraints the state components or functions of them at either node of any
Figure 6: The $CB_i$-$CB_j$ Rotating Frame with respect to a $CB_i$ Fixed Centered Frame. $CB_j$ pulsates along the line connecting $CB_i$ and $CB_j$ if their absolute motion is non-circular with respect to each other.

segment. For example, finding an equilibrium point in a rotating-pulsating frame centered at the Earth with the Moon along the $\hat{r}$ direction requires a three dimensional search for the position components $r^{rst}$ with the conditions that $v^{rst} = a^{rst} = 0$, the definition of an equilibrium point. If there are no external accelerations beyond those from the gravitational acceleration due to the Earth and the Moon, then this equilibrium point will remain fixed in the rotating-pulsating frame. If external accelerations are present, such as the gravitational acceleration due to the Sun, then this equilibrium point can be only defined for a particular epoch.

2.5 State Transformations

Recall that all segment propagations are made in either a barycentered or celestial body fixed frame. It is required that a segment’s initial state vector and all other quantities that define it be referenced to the segment’s fixed reference frame used in its propagation. It is convenient for certain problems to express these quantities in any other reference frame or coordinate set. If this is the case, a suitable transformation to the segment’s propagation frame is necessary. Consider, for example, a segment $i$ in a gravity field composed of the celestial bodies, $CB_i$ and $CB_j$. Celestial body $CB_i$ can be defined to be $CB_f$, the center of the integration frame, and the motion of $CB_j$ is assumed known relative to $CB_f$. The initial state of the spacecraft can be referenced to any of the bodies. If it is referenced to $CB_j$, the required transformation is

$$
\begin{pmatrix}
(r_0)_{CB_i} \\
(v_0)_{CB_i}
\end{pmatrix}^i \rightarrow \begin{pmatrix}
(r_0)_{CB_f} \\
(v_0)_{CB_f}
\end{pmatrix}^i
$$

(20)

If there exists a more convenient vector or parametric representation of the initial state vector, say the classical Keplerian orbital elements of the spacecraft...
Figure 7: The $CB_i$-$CB_j$ Rotating Pulsating Frame with respect to a $CB_k$ Fixed Centered Frame. The two celestial bodies, $CB_i$ and $CB_j$, do not orbit in closed orbits about their common barycenter.

with respect to $CB_j$ at $t_{i0}$, then the required transformation is

$$
\begin{pmatrix}
a \\
e \\
i \\
\Omega \\
\omega \\
\nu
\end{pmatrix}^{i}_{CB_j} \rightarrow \begin{pmatrix}
(r_0)_{CB_j} \\
(v^{-}_0)_{CB_j}
\end{pmatrix}^{i} \rightarrow \begin{pmatrix}
(r_0)_{CB_f} \\
(v^{-}_0)_{CB_f}
\end{pmatrix}^{i}
$$

(21)

where $(a, e, i, \Omega, \omega, \nu)^{i}_{CB_j}$ is a vector whose components are the semi-major axis, eccentricity, inclination, right ascension of the ascending node, argument of perihelion, and true anomaly of the spacecraft at $t_{i0}$ relative to $CB_j$. Generally, it can be assumed that an initial state vector, $y(t_{i0})$, is comprised of any independent set of quantities that uniquely represent the initial state vector of the spacecraft. Ultimately, the transformation $y(t_{i0})_{CB_j} \rightarrow (r_i^i, v_i^i)_{CB_f}$ will be needed for any independent state vector representation with respect to any celestial body $CB_j$. There are many possible parameterizations for $y(t_{i0})_{CB_j}$ including those that contain departure, capture, or flyby parameters used in practice such as the hyperbolic excess velocity $v_\infty$ or the square of its value referred to as twice the hyperbolic energy, $C_3$, and the direction of the incoming or outgoing asymptote, $\alpha_v, \delta_v$.

2.6 Open Loop Solution of a Trajectory or a Mission

Given the number of segments and their defining parameters, the open loop solution of the system can be determined. The open loop solution is the first estimate or iterate which will generally be infeasible since the constraints or boundary conditions are not necessarily satisfied. The segments are all propagated independently unless a particular segment is required to be connected.
sequentially to any other segment in which case the segments that are to be continued with other segments are propagated after the segments to which they are connected to.

If the segments are initially disconnected, i.e., discontinuous in any of the state variables and possibly the time, and these segments form a trajectory for a single spacecraft, the segments will have to be connected by the solution process. Some of the parameters defining these segments will be variables that can be adjusted, and constraint conditions are imposed as continuity conditions for values of all or either the position, velocity, mass, and time. For a multiple spacecraft mission where several of the spacecraft need to rendezvous with other spacecraft, some of the segments will have to be connected either in position or velocity, or both, at the appropriate nodes.

If a segment $k$ is forced to be connected sequentially to any other node of any other segment $i$ beforehand, the connectivity is enforced at any of the discrete node times, $t_{i-0}^{0-}, t_{i-0}^{0+}, t_{i-0}^{0-}, t_{i-0}^{0+}$, i.e., the continuity of the position, velocity, and mass of segment $i$ can be made prior to or after the initial or final impulsive maneuvers or any of the mass discontinuities, if they exist. For example, if a purely ballistic segment $k$ is required to be connected sequentially in all three state variables $(r,v,m)$ to any other segment $i$ in the model at the $t_f$ node after all, if any, of the velocity and mass discontinuities, the state vector of the $t_0$ node of segment $k$ is then

$$
\begin{align*}
\mathbf{r}^k_0 &= \mathbf{r}^i_{(0,f)} \\
\mathbf{v}^k_0^- &= \mathbf{v}^i_{(0,f)} \\
\mathbf{v}^k_0^+ &= \mathbf{v}^i_{(0,f)}
\end{align*}
$$

Here then, it is necessary that segment $i$ be evaluated entirely before segment $k$ can be evaluated. On the other hand, if the final solution requires that segments $i$ and $k$ be connected as in Eq. 22, these can be initially defined
independently and Eq. 22 is used as an equality constraint condition that must be satisfied by the final solution. Continuity is not allowed at any time between \( t^+ \) and \( t^- \) because only node points can be constrained. But this is not a restriction, because an additional segment is introduced such that one of its nodes can be used as the continuation to another segment.

This generality facilitates the modeling of complex trajectories that may be otherwise difficult to obtain as a single forward time or backward time propagated trajectory and in the modeling of multi-spacecraft problems where the other spacecraft are moving along different trajectories.

3 Solution Methods

For any given problem, that in general can be complex and involve multiple constraints, it is possible that no solution exists. A common reason for this is that the constraints are inconsistent meaning the problem definition itself precludes the possibility of a solution. Therefore, a necessary requirement is to formulate a consistent problem where the variables of the problem are independent and the constraints are at least linearly independent. If it is then assumed that a solution should exist, there are two possible solution methods. These methods provide solutions to the majority of the spacecraft trajectory design and optimization problems considered by the current architecture. This section describes these methods which are used to solve trajectory problems that are based on the set of basic segments.

Let \( n \) be the number of segments in the trajectory or the mission. Let the segment vector, \( s^i \), be a vector whose components are all of the independent and dependent quantities that uniquely define segment \( i \). Within an individual segment, some of these quantities will be either dependent or independent. For example, \( m^+ \), the vehicle mass at \( t^+ \), is a dependent quantity that depends on any impulsive maneuver at \( t_0 \) and/or \( t_f \), any mass depleting finite maneuver between \( t^0 \) and \( t^- \), and any non-maneuver mass discontinuity that may be present at either or both nodes. But \( m^- \) is an independent quantity (within the segment), though it may be constrained outside of the segment during the solution process. Depending on the target conditions required to be satisfied for a given mission, further information can be computed from \( s^i \) that can be used as constraints or as a cost function. For example, the state vector at any of the discrete times of the node point relative to any of the other celestial bodies that are present or any function of these can be computed provided the appropriate transformations are available. A common function is the two-body energy of a segment endpoint with respect to a celestial body.

Let \( x_p \) be a \( p \) vector containing all of the variable parameters that have been identified for any given problem. The parameter vector \( x_p \) can contain any independent element of any or all of the segments identified in \( s^i (i = 1, ..., n) \). It can also contain any parameter that defines the force model. Let \( c \) be a \( q \) vector containing all of the equality and inequality constraint functions. The constraint vector \( c \) is in general a nonlinear function of any of the elements in \( x_p \). Let \( J \) be a general nonlinear scalar function of any of the elements in \( x_p \). The value of \( J \) is to be extremized. A description of the two solution methods and their implementation into the current architecture follow.
3.0.1 Case 1: Solution of a System of Nonlinear Equations, $c(x_p)=0$.

If the number of variables in $x_p$ is equal to the number of constraint equations $c$ ($p = q$) and all of the elements in $c$ are equality constraints then the problem to solve is a targeting problem with no explicit optimization, i.e., no explicit scalar function is extremized. However, this does not exclude the possibility of implicitly extremizing a cost function if the equality constraint functions include necessary conditions based on optimal control theory.

The targeting is completely general and the solution for $x_p$ is a solution to a general nonlinear root finding problem for a system of nonlinear functions or equations with an equal number of unknown parameters. Functionally, the problem is to solve the set of equations $c(x_p) = 0$. It is assumed that the functions $c(x_p)$ be smooth, i.e., at least twice-continuously differentiable, in the independent variable vector $x_p$. Though in general, these equations are not available explicitly, the functions given in $c(x_p)$ are determined from a sequence of operations that depend on the specific model and problem. These operations will involve numerical propagations of the segments, transformations between reference frames, and function evaluations of the elements in the constraint vector, $c$.

The system can use any efficient and robust algorithm to solve the nonlinear root finding problem. The current prototype system uses a Newton-Raphson/Steepest descent correction to compute a search direction and is coupled with Broyden’s method for computing and updating the Jacobian matrix of the system, $\partial c/\partial x_p$. The analytical and numerical issues associated with this and similar methods can be found in Dennis and Schnabel[11], Gill et. al[12], and Nocedal and Wright[13].

The two-body orbital boundary value problem, which is known as the Lambert Problem if the gravitational force model includes only one celestial body with no other perturbations is a subset of this problem. The orbital boundary value problem, for any general gravitational force field which can include more celestial bodies, or a complex potential gravity field for one or more celestial bodies, or other perturbations, is a two point boundary value problem which is also a subset of this case. Other specific problems that can be solved as a nonlinear root finding problem in orbital mechanics include periodic orbit searches, Earth-to-Moon trajectories, including free return trajectories, multi-body gravity assisted trajectories in the Solar System, etc.

The functional optimization or optimal control problem for either impulsive maneuvers, finite engine burn maneuvers, or non-mass depleting acceleration controlled maneuvers using optimal control theory is also a subset of this case. Here the time varying Lagrange multipliers associated with each of the physical state variables augment the $s'$ vector for each segment, and the first order necessary conditions and the transversality conditions lead to a well defined and in general multi-point boundary value problem. If solved as a nonlinear root finding problem, the problem is an implicit optimization. An extremal solution that satisfies the first order necessary conditions from optimal control can be produced by solving the appropriate nonlinear system of functions. This is the well known indirect method for trajectory optimization and has been well documented in the literature. A comprehensive treatment of the optimal control problem and its solution methods is given by Hull[14].

Since the system of equations is in general nonlinear, multiple solutions
can be expected. If the functional optimization problem is solved indirectly, the solution provided by the nonlinear search will satisfy first order necessary conditions for an optimal solution. If the general targeting problem is solved without the use of the functional optimization conditions, then the solution is only a feasible solution.

It will always be necessary to provide a reasonable estimate the for the starting iterate given by \( x_p \). For simple problems, this starting guess can be estimated analytically. For more complex problems, it may be necessary to use results available from previous research in specific problems. For example, circular restricted three body trajectories are well understood though closed form solutions for these do not exist. For force models that are not too different from a simplified circular restricted three body force field model, the results from the simple circular restricted three body model can be used as starting iterates for more complex models.

If some of the elements in the constraint vector, \( c \), for a targeting problem are inequality constraints, these can be converted to equality constraints by adding one more variable to the parameter vector \( x_p \) known as a slack variable such that the final problem is an equality constrained problem and solved as such (Ref.[14]).

The underdetermined or overdetermined cases where there are more or less variables, respectively, than equality constraints is solved as a minimax problem (Ref. [12]) commonly used for nonlinear minimax data-fitting, where the constraint function with the maximum absolute value is minimized. The minimax solution to the system of nonlinear equations

\[
\mathbf{c}(x_p) = 0
\]

is the solution that minimizes the function, \( F \) defined by

\[
F(x_p) = \max |c_i(x_p)|
\]

where \( c_i \) is the \( i \)-th element of the constraint vector \( \mathbf{c} \).

This technique is robust and efficient for the class of nonlinear targeting problems considered here where there is not an explicit cost function and satisfaction of the constraint equations results in a feasible solution. In the underdetermined case, a solution to the minimax problem can be used as the starting estimate for the parameter optimization problem since all of the constraints are satisfied initially.

### 3.0.2 Case 2: The Constrained Parameter Optimization Problem

This is the general problem of nonlinear functional optimization with nonlinear equality and inequality constraints and whose solution methods have been well documented in the literature (see for example Gill et al.[12] and Nocedal and Wright[13] ). The problem is to minimize or maximize the objective function

\[
J = J(x_p)
\]

subject to both equality and inequality constraints,

\[
\begin{align*}
\mathbf{c}_{eq}(x_p) &= 0 & \mathbf{c}_{eq} : q_{eq} \times 1 \\
\mathbf{c}_{ineq}(x_p) &\geq 0 & \mathbf{c}_{ineq} : q_{ineq} \times 1
\end{align*}
\]
where \( \mathbf{c}_{eq} \) is the \( q_{eq} \) vector of equality constraints and \( \mathbf{c}_{ineq} \) is the \( q_{ineq} \) vector of inequality constraints. The functions \( \mathbf{c}_{eq}(\mathbf{x}_p) \) and \( \mathbf{c}_{ineq}(\mathbf{x}_p) \) can represent lower and upper bounds on the individual elements in \( \mathbf{x}_p \), linear constraints on two or more of the elements in \( \mathbf{x}_p \), or more generally, nonlinear functions of the elements of \( \mathbf{x}_p \). The algorithm used in the current system is a sequential quadratic programming (SQP) variable metric method described in Ref.[12] and Ref.[13].

With the given architecture, the formulation of many parameter optimization problems is straightforward. The objective function to extremize (minimum or maximum) can be taken to be a single scalar variable in \( \mathbf{s}^i (i = 1, ..., n) \) or any scalar function of these elements. For example, a trajectory that begins in an orbit about \( CB_i \) and terminates in an orbit about \( CB_j \) can be such that the initial mass of the spacecraft is a minimum if the final mass has been prescribed. For this case \( J = \min(m_i^0) \) where \( i \) is first segment of the trajectory representing the spacecraft. The value of the initial mass required is a function of all the segments following it including all of the possible impulsive or finite burn maneuvers used to reach the final orbit, and any non maneuver mass discontinuities. The constraints associated with the initial orbit and the final orbit would be the nonlinear constraint functions in \( \mathbf{c} \), and \( \mathbf{x}_p \) would contain the parameters that can be estimated such as the time of the maneuvers and the parameters that define the maneuver and that are allowed to be adjusted.

4 General System Issues

Several main aspects important to a general trajectory design and optimization system are discussed in this section. This includes be benefits of using a system that is modular in all of its subcomponents, the importance of tuning the algorithms and automating the tuning process, and the use of integrated visualization as a key part of the solution process.

4.1 Modular Architecture

A general trajectory optimization system should be modular in the sense that the algorithms used to solve different parts of the overall trajectory problem are independent and can be readily modified or replaced. The basic components include algorithms for explicit numerical integration, solutions to systems of nonlinear equations, and solutions of the nonlinear constrained parameter optimization problem. The individual algorithms used for each of these functions should be the ‘best available’ and with the added flexibility of being updated or changed entirely as new and more robust and efficient algorithms are developed. Provisions should also be made to allow completely general force models to be used including the use of either realistic ephemeris models or user defined models. In summary, the assumption is to treat each algorithm to each sub-problem as ‘solved’. The continued development should be directed to the system architecture and all that it entails such as refinement of the definition of the basic segment, the maneuver models, and the coordinate frames for state definition, targeting, and maneuver parameterization.
4.2 Algorithm and System Tuning

Provided an estimate for a convergent solution is available for both the nonlinear equation solvers or the parameter optimization algorithms based on gradient information, there exists a sequence of perturbations step sizes for the parameter vector, $x_p$, that achieves a solution in a minimum number of function evaluations for the constraint vector, $c$, or both $c$ and the cost function, $J$. If the method of finite differences are being used to estimate gradients, typically the vector of perturbation step sizes, $\delta x_p$, is set as a constant during the iteration process. Regardless of the method used to compute the finite difference based gradients, determining the ‘best’ value for the perturbation vector is a process referred to as tuning and which is problem, algorithm, and processor dependent. A given value of $\delta x_p$ that provides accurate estimates for the gradients at the beginning of an iteration process may not be the best choice at other parts of the search space, therefore this vector should be allowed to change and recomputed periodically to produce accurate gradients at different points in the search space.

Any system should have in place an automatic tuning algorithm that can adjust the perturbation vector over the course of the iteration process to achieve convergence using the minimum amount of function and constraint vector evaluations. Included in the tuning process is the proper scaling of the parameter vector and the constraint functions if their values differ by large orders of magnitude. Also, a change to each of the elements in the parameter vector $x_p$ per iteration should be bounded to avoid evaluating subsequent iterates that cause the solution to diverge. This is common in trajectory problems where the endpoints of a trajectory operating in a complex force field are highly sensitive to changes in the parameters of earlier parts of the trajectory. But for well behaved solutions, if this upper bound is too small, many iterations will be needed for convergence. The proper choice of $\delta x_p$ per iteration, the scaling, and the maximum allowable changes to $x_p$ per iteration, are three factors that influence the convergence rate of a solution and have to be considered carefully in an automated tuning process.

A procedure that leads to accurate derivative estimates is based on the calculation of the state transition matrix (Ref. [7]) along the ballistic or controlled accelerated arcs for all segments between $t_f^+$ and $t_f^-$ and for the variables that are numerically integrated. The state transition matrix is the time varying fundamental matrix solution of the linearized variational equations evaluated along the trajectory arcs. If the state vector $x(t)$ is composed only of the physical state variables $r,v,m$ the state transition matrix at $t_f^-$ for a given statement provides the Jacobian matrix of the final state with respect to the initial state. It is a $7 \times 7$ matrix with the gradients $\partial x(t_f^-) / \partial x(t_f^+)$. If the Lagrange multiplier vector is part of the state vector, so that all of the variables $r,v,m,\lambda_r,\lambda_v,\lambda_m$ are numerically integrated along a trajectory arc, then the state transition matrix will be a $14 \times 14$ matrix. The state transition matrix is integrated along with the state vector thereby requiring the integration of $n + n^2$ first order equations where $n$ is the size of the state vector. The information available in the state transition matrix represents some of the terms necessary to compute the required gradients.

Given a constraint vector, $c$, or a scalar objective function $J$, that depends on an independent parameter vector, $x_p$, there exists analytical expressions
for both the gradient vector, $\partial J/\partial x_p$, and the Jacobian matrix, $\partial c/\partial x_p$, that are linearly valid near any nominal solution of a segment, trajectory, or mission. These expressions are based on state transition matrices along all of the segments, and the gradients across impulse points, discontinuous mass points, and any state and maneuver transformations used. Though numerical integration of the state vector and the state transition matrix is required to evaluate the quantities necessary to evaluate both $\partial J/\partial x_p$ and $\partial c/\partial x_p$, these provide accurate values at $x_p$ necessary for any gradient based solution method. The disadvantage of this process is that it is problem specific and requires the derivation of these expressions for each type of problem, a time consuming process if many classes of problems are to be considered. Nevertheless, these expressions can be evaluated and used as the actual gradients necessary in the solution methods. Alternatively, these expressions can be evaluated at discrete points in the iteration process, and used to tune the perturbation vector $\delta x_p$ at these points. The perturbation vector is then used to estimate the derivatives with a finite difference approximation.

4.3 Integrated Visualization

A general trajectory design and optimization system should include interactive visualization capability not only for presentation of intermediate and final results but as a key part of the solution process. The system currently provides the capability to visually display a three dimensional graphics representation of the trajectory design and optimization process in real processor time; i.e. immediate visual feedback during the targeting or optimization process is available. The reference frame used to visualize the dynamics and solution process can be independent from the reference frames used in the trajectory problems. The information shown in the graphics representation typically includes the positions of each point in the trajectory. However, other phase space variables can be visualized as well if the visualization of their evolution provides better information.

Though visualization has generally not been accepted as a necessary capability in the solution process, it is used here as a critical component in the design and optimization process for several reasons. First, it is immediately possible to determine if there is an error in the input information thereby allowing for the termination of the process and correction of the error. Secondly, it is possible to determine whether the solution is converging, and if so, the rate of convergence. If convergence is slow, the speed of convergence can be increased by re-tuning the algorithms. Thirdly, an intuitive understanding of the dynamics associated with complex trajectories in multi-body gravity fields or rotating frames can be obtained, thereby facilitating future investigation of more difficult problems. All of these benefits can be obtained from a non-graphics based system in a post execution manner by examining the output data generated and plotting any desired information. But clearly, the time saved by not having to perform these post execution tasks is a reason that supports the use of integrated graphic visualization in the solution process.
5 Conclusions

The key aspects associated with a general trajectory optimization system have been presented. The continuing effort is an attempt to construct a system that can be used to analyze a large range of problems that are currently required in trajectory design and optimization. Future spacecraft missions will use innovative trajectory dynamics that take advantage of the natural and complex dynamics associated with multi-body gravity fields, multiple and hybrid propulsion systems, and multiple spacecraft. A main objective is the development of a system that can facilitate the solution to these problems using a general framework that is applicable to all of the sub-problems required for the overall solution. Such a system should produce results that can be used in research topics involving both trajectory optimization theory and numerical methods for the solution to these problems. The system should also be able to produce results that can be used in actual spacecraft operations by accounting for actual mission operation constraints; detailed propulsion system models; attitude control, constraints, and requirements; and some level of information regarding the observability of the solutions. A deterministically optimal solution based on propellant or mass performance only may not be the best solution given a measurement model to observe the state of the spacecraft. The capability to produce an integrated solution that offers an acceptable compromise between performance and the ability to accurately navigate it is desirable.

Caution must be exercised in not having a system so general that it solves many problems only superficially and with many restricting assumptions. There will always be a balance between how general a system is and how detailed and complex the solutions produced by such a system are for any given problem. A measure of the usefulness of such a system is the scope of the problems it can solve while solving them to the level of detail needed to be valid enough. A useful system should produce solutions of practical interest to spacecraft missions that will require only minor adjustments when and if the mission is actually flown.

No claim has been made concerning the superiority of the described architecture over existing architectures and systems. The only claim made is that the given architecture facilitates the modeling and optimization of many classes of trajectories. The system remains evolutionary, meaning that it can be changed and enhanced as needed to address problems whose requirements can not currently be met with the current system architecture. If this is the case, then it is postulated that the majority of the changes needed will be directed to the modification and enhancement of definition of the basic segment; and not so much in the solution methods.

References


