2 Primer Vector Theory and Applications

John E. Prussing

Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, Urbana, IL

2.1 Introduction

In this chapter, the theory and a resulting indirect method of trajectory optimization are derived and illustrated. In an indirect method, an optimal trajectory is determined by satisfying a set of necessary conditions (NC), and sufficient conditions (SC) if available. By contrast, a direct method uses the cost itself to determine an optimal solution.

Even when a direct method is used, these conditions are useful to determine whether the solution satisfies the NC for an optimal solution. If it does not, it is not an optimal solution. As an example, the best two-impulse solution obtained by a direct method is not the optimal solution if the NC indicate that three impulses are required. Thus, post-processing a direct solution using the NC (and SC if available) is essential to verify optimality.

Optimal Control [1], a generalization of the calculus of variations, is used to derive a set of necessary conditions for an optimal trajectory. The primer vector is a term coined by D. F. Lawden [2] in his pioneering work in optimal trajectories. [This terminology is explained after Equation (2.24).] First-order necessary conditions for both impulsive and continuous-thrust trajectories can be expressed in terms of the primer vector. For impulsive trajectories, the primer vector determines the times and positions of the thrust impulses that minimize the propellant cost. For continuous-thrust trajectories, both the optimal thrust direction and the optimal thrust magnitude as functions of time are determined by the primer vector. As is standard practice, the word “optimal” is loosely used as shorthand for “satisfies the first-order NC.”

The most completely developed primer vector theory is for impulsive trajectories. Terminal coasting periods for fixed-time trajectories and the addition of midcourse impulses can sometimes lower the cost. The primer vector indicates when these modifications should be made. Gradients of the cost with respect to terminal impulse times and midcourse impulse times and positions were first derived by Lion and Handelsman [3]. These gradients were then implemented in a nonlinear

Figures 2.2 and 2.4–2.8 were generated using the MATLAB computer code written by Suzannah L. Sandrik [13].
2.2 First-Order Necessary Conditions

2.2.1 Optimal Constant-Specific-Impulse Trajectory

For a constant specific impulse (CSI) engine, the thrust is bounded by \( 0 \leq T \leq T_{\text{max}} \) (where \( T_{\text{max}} \) is a constant), corresponding to bounds on the mass flow rate: \( 0 \leq b \leq b_{\text{max}} \) (where \( b_{\text{max}} \) is a constant). Note that one can also prescribe bounds on the thrust acceleration (thrust per unit mass) \( \Gamma \equiv T/m \) as \( 0 \leq \Gamma \leq \Gamma_{\text{max}} \), where \( \Gamma_{\text{max}} \) is achieved by running the engine at \( T_{\text{max}} \). However, \( \Gamma_{\text{max}} \) is not constant but increases due to the decreasing mass. One must keep track of the changing mass in order to compute \( \Gamma \) for a given thrust level. This is easy to do, especially if the thrust is held constant, for example, at its maximum value. However, if the propellant mass required is a small fraction of the total mass because of being optimized, a constant \( \Gamma_{\text{max}} \) approximation can be made.

The cost functional representing minimum propellant consumption for the CSI case is

\[
J = \int_{t_0}^{t_f} \Gamma(t) dt. \tag{2.1}
\]

The state vector is defined as

\[
\mathbf{x}(t) = \begin{bmatrix} \mathbf{r}(t) \\ \mathbf{v}(t) \end{bmatrix} \tag{2.2}
\]

where \( \mathbf{r}(t) \) is the spacecraft position vector and \( \mathbf{v}(t) \) is its velocity vector. The mass \( m \) can be kept track of without defining it to be a state variable by noting that

\[
m(t) = m_0 e^{-F(t)/c} \tag{2.3}
\]

where \( c \) is the exhaust velocity and

\[
F(t) = \int_{t_0}^{t} \Gamma(\xi) d\xi. \tag{2.4}
\]

Note that from Equation (2.4), \( F(t_f) \) is equal to the cost \( J \). In the constant thrust case, \( \Gamma \) varies according to \( \dot{\Gamma} = \frac{1}{c} \Gamma^2 \), which is consistent with the mass decreasing linearly with time.

The equation of motion is

\[
\dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{r}} \\ \dot{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{v} \\ \mathbf{g}(\mathbf{r}) + \Gamma \mathbf{u} \end{bmatrix} \tag{2.5}
\]

with the initial state \( \mathbf{x}(t_0) \) specified.
In Equation (2.5), $\mathbf{g}(\mathbf{r})$ is the gravitational acceleration and $\mathbf{u}$ represents a unit vector in the thrust direction. An example gravitational field is the inverse-square field:

$$
\mathbf{g}(\mathbf{r}) = -\frac{\mu}{r^2} \mathbf{r} = -\frac{\mu}{r^3} \mathbf{r}.
$$

(2.6)

The first-order necessary conditions for an optimal CSI trajectory were first derived by Lawden [2] using classical calculus of variations. In the derivation that follows, an optimal control theory formulation is used, but the derivation is similar to that of Lawden. One difference is that the mass is not considered a state variable but is kept track of separately.

In order to minimize the cost in Equation (2.1), one forms the Hamiltonian using Equation (2.5) as

$$
H = \Gamma + \lambda_r^T \mathbf{v} + \lambda_v^T [\mathbf{g}(\mathbf{r}) + \Gamma \mathbf{u}].
$$

(2.7)

The adjoint equations are then

$$
\dot{\lambda}_r^T = -\frac{\partial H}{\partial \mathbf{r}} = -\lambda_v^T \mathbf{G}(\mathbf{r})
$$

(2.8)

$$
\dot{\lambda}_v^T = -\frac{\partial H}{\partial \mathbf{v}} = -\lambda_r^T
$$

(2.9)

where

$$
\mathbf{G}(\mathbf{r}) \equiv \frac{\partial \mathbf{g}(\mathbf{r})}{\partial \mathbf{r}}
$$

(2.10)

is the symmetric $3 \times 3$ gravity gradient matrix.

For terminal constraints of the form

$$
\psi[\mathbf{r}(t_f), \mathbf{v}(t_f), t_f] = 0,
$$

(2.11)

which may describe an orbital intercept, rendezvous, etc., the boundary conditions on Equations (2.8–2.9) are given in terms of

$$
\Phi \equiv \mathbf{v}^T \psi[\mathbf{r}(t_f), \mathbf{v}(t_f), t_f]
$$

(2.12)

as

$$
\lambda_r^T(t_f) = \frac{\partial \Phi}{\partial \mathbf{r}(t_f)} = \mathbf{v}^T \frac{\partial \psi}{\partial \mathbf{r}(t_f)}
$$

(2.13)

$$
\lambda_v^T(t_f) = \frac{\partial \Phi}{\partial \mathbf{v}(t_f)} = \mathbf{v}^T \frac{\partial \psi}{\partial \mathbf{v}(t_f)}.
$$

(2.14)

There are two control variables, the thrust direction $\mathbf{u}$ and the thrust acceleration magnitude $\Gamma$, that must be chosen to satisfy the minimum principle [1], that is, to minimize the instantaneous value of the Hamiltonian $H$. By inspection, the Hamiltonian of Equation (2.7) is minimized over the choice of thrust direction by aligning the unit
2.2 First-Order Necessary Conditions

Vector $\mathbf{u}(t)$ opposite to the adjoint vector $\lambda_v(t)$. Because of the significance of the vector $-\lambda_v(t)$, Lawden [2] termed it the primer vector $\mathbf{p}(t)$:

$$\mathbf{p}(t) \equiv -\lambda_v(t). \quad (2.15)$$

The optimal thrust unit vector is then in the direction of the primer vector, specifically

$$\mathbf{u}(t) = \frac{\mathbf{p}(t)}{p(t)} \quad (2.16)$$

and

$$\lambda_v^T \mathbf{u} = -\lambda_v = -p \quad (2.17)$$
in the Hamiltonian of Equation (2.7).

From Equations (2.9) and (2.15), it is evident that

$$\lambda_r(t) = \dot{\mathbf{p}}(t). \quad (2.18)$$

Equations (2.8), (2.9), (2.15), and (2.18) combine to yield the primer vector equation

$$\ddot{\mathbf{p}} = G(\mathbf{r}) \mathbf{p}. \quad (2.19)$$

The boundary conditions on the solution to Equation (2.19) are obtained from Equations (2.13) (2.14)

$$\mathbf{p}(t_f) = -v^T \frac{\partial \psi}{\partial \mathbf{v}}(t_f) \quad (2.20)$$

$$\dot{\mathbf{p}}(t_f) = v^T \frac{\partial \psi}{\partial \mathbf{r}}(t_f). \quad (2.21)$$

Note that in Equation (2.20), the final value of the primer vector for an optimal intercept is the zero vector, because the terminal constraint $\psi$ does not depend on $\mathbf{v}(t_f)$.

Using Equations (2.15)–(2.18), the Hamiltonian of Equation (2.7) can be rewritten as

$$H = -(p - 1)\Gamma + \dot{\mathbf{p}}^T \mathbf{v} - \mathbf{p}^T \mathbf{g}. \quad (2.22)$$

To minimize the Hamiltonian over the choice of the thrust acceleration magnitude $\Gamma$, one notes that the Hamiltonian is a linear function of $\Gamma$, and thus the minimizing value for $0 \leq \Gamma \leq \Gamma_{\text{max}}$ will depend on the algebraic sign of the coefficient of $\Gamma$ in Equation (2.22). It is convenient to define the switching function

$$S(t) \equiv p - 1. \quad (2.23)$$

The choice of the thrust acceleration magnitude $\Gamma$ that minimizes $H$ is then given by the “bang-bang” control law

$$\Gamma = \begin{cases} \Gamma_{\text{max}} & \text{for } S > 0 \ (p > 1) \\ 0 & \text{for } S < 0 \ (p < 1) \end{cases}. \quad (2.24)$$
That is, the thrust magnitude switches between its limiting values of 0 (an NT null-thrust arc) and $T_{\text{max}}$ (an MT maximum-thrust arc) each time $S(t)$ passes through 0 [$p(t)$ passes through 1] according to Equation (2.24). Figure 2.1 shows an example switching function for a three-burn trajectory.

The possibility also exists that $S(t) \equiv 0 \left[ p(t) \equiv 1 \right]$ on an interval of finite duration. From Equation (2.22), it is evident that in this case the thrust acceleration magnitude is not determined by the minimum principle and may take on intermediate values between 0 and $\Gamma_{\text{max}}$. This IT “intermediate thrust arc” [2] is referred to as a singular arc in optimal control [1].

Lawden explained the origin of the term primer vector in a personal letter in 1990: “In regard to the term ‘primer vector’, you are quite correct in your supposition. I served in the artillery during the war [World War II] and became familiar with the initiation of the burning of cordite by means of a primer charge. Thus, $p = 1$ is the signal for the rocket motor to be ignited.”

It follows then from Equation (2.3) that if $T = T_{\text{max}}$ and the engine is on for a total of $\Delta t$ time units,

$$\Gamma_{\text{max}}(t) = e^{F(t)/c} T_{\text{max}}/m_o = T_{\text{max}}/(m_o - b_{\text{max}}/\Delta t). \quad (2.25)$$

Other necessary conditions are that the variables $p$ and $\dot{p}$ must be continuous everywhere. Equation (2.23) then indicates that the switching function $S(t)$ is also continuous everywhere.

Even though the gravitational field is time-invariant, the Hamiltonian in this formulation does not provide a first integral (constant of the motion) on an MT arc, because $\Gamma$ is an explicit function of time as shown in Equation (2.25). From Equation (2.22)

$$H = -S \Gamma + \dot{p}^T \mathbf{v} - p^T \mathbf{g}. \quad (2.26)$$

Note that the Hamiltonian is continuous everywhere because $S = 0$ at the discontinuities in the thrust acceleration magnitude.
2.2.2 Optimal Impulsive Trajectory

For a high-thrust CSI engine the thrust durations are very small compared with the times between thrusts. Because of this, one can approximate each MT arc as an impulse (Dirac delta function) having unbounded magnitude \( \Gamma_{\text{max}} \to \infty \) and zero duration. The primer vector then determines both the optimal times and directions of the thrust impulses with \( p \leq 1 \) corresponding to \( S \leq 0 \). The impulses can occur only at those instants at which \( S = 0 \) \( (p = 1) \). These impulses are separated by NT arcs along which \( S < 0 \) \( (p < 1) \). At the impulse times the primer vector is then a unit vector in the optimal thrust direction.

The necessary conditions (NC) for an optimal impulsive trajectory, first derived by Lawden \([2]\), are shown in Table 2.1.

For a linear system, these NC are also sufficient conditions for an optimal trajectory \([5]\). Also in \([5]\), an upper bound on the number of impulses required for an optimal solution is given.

Figure 2.2 shows a trajectory (at top) and a primer vector magnitude (at bottom) for an optimal three-impulse solution. (In all of the trajectory plots in this chapter, the direction of orbital motion is counterclockwise.) Canonical units are used. The canonical time unit is the orbital period of the circular orbit that has a radius of one canonical distance unit. The initial orbit is a unit radius circular orbit, shown as the topmost orbit going counterclockwise from the symbol \( \oplus \) at \((1,0)\) to \((-1,0)\). The transfer time is 0.5 original (initial) orbit periods (OOP). The target is in a coplanar circular orbit of radius 2, with an initial lead angle (ila) of 270° and shown by the symbol \( \Box \) at \((0,-2)\). The spacecraft departs \( \bigcirc \) and intercepts \( \square \) at approximately \((1.8,-0.8)\) as shown. The + signs at the initial and final points indicate thrust impulses and the + sign on the transfer orbit very near \((0,0)\) indicates the location of the midcourse impulse. The magnitudes of the three \( \Delta V \)s are shown at the left, with the total \( \Delta V \) equal to 1.3681 in units of circular orbit speed in the initial orbit.

The examples shown in this chapter are coplanar, but the theory and applications apply to three-dimensional trajectories as well, for example, see Prussing and Chiu \([6]\).

The bottom graph in Figure 2.2 displays the time history of the primer vector magnitude. Note that it satisfies the necessary conditions of Table 2.1 for an optimal transfer.

Table 2.1. Impulsive necessary conditions

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The primer vector and its first derivative are continuous everywhere.</td>
</tr>
<tr>
<td>2.</td>
<td>The magnitude of the primer vector satisfies ( p(t) \leq 1 ) with the impulses occurring at those instants at which ( p = 1 ).</td>
</tr>
<tr>
<td>3.</td>
<td>At the impulse times the primer vector is a unit vector in the optimal thrust direction.</td>
</tr>
<tr>
<td>4.</td>
<td>As a consequence of the above conditions, ( dp/dt = \dot{p} = \hat{p}^T \hat{p} = 0 ) at an intermediate impulse (not at the initial or final time).</td>
</tr>
</tbody>
</table>
Primer Vector Theory and Applications

Note also that at a thrust impulse at time \( t_k \)

\[
\Gamma(t) = \Delta v_k \delta(t - t_k)
\]  

(2.27)

and from Equation (2.4)

\[
\Delta v_k = \int_{t_k^-}^{t_k^+} \Gamma(t)dt = F(t_k^+) - F(t_k^-)
\]  

(2.28)

where \( t_k^+ \) and \( t_k^- \) are times immediately after and before the impulse time, respectively. Equation (2.3) then becomes the familiar solution to the rocket equation:

\[
m(t_k^+) = m(t_k^-)e^{-\Delta v_k/c}.
\]  

(2.29)

### 2.2.3 Optimal Variable-Specific-Impulse Trajectory

A variable-specific-impulse (VSI) engine is also known as a power-limited (PL) engine, because the power source is separate from the engine itself, for example, solar panels, and radioisotope thermoelectric generator. The power delivered to the engine is bounded between 0 and a maximum value \( P_{\text{max}} \), with the optimal value being constant and equal to the maximum. The cost functional representing minimum propellant consumption for the VSI case is

\[
J = \frac{1}{2} \int_{t_0}^{t_f} \Gamma^2(t)dt.
\]  

(2.30)
2.3 Solution to the Primer Vector Equation

Writing $\Gamma^2$ as $\Gamma^T \Gamma$, the corresponding Hamiltonian function can be written as

$$H = \frac{1}{2} \Gamma^T \Gamma + \lambda_\Gamma^T \mathbf{v} + \lambda_v^T [\mathbf{g} r + \Gamma]. \quad (2.31)$$

For the VSI case, there is no need to consider the thrust acceleration magnitude and direction separately, so the vector $\Gamma$ is used in place of the term $\Gamma u$ that appears in Equation (2.7).

Because $H$ is a nonlinear function of $\Gamma$, the minimum principle is applied by setting

$$\frac{\partial H}{\partial \Gamma} = \Gamma^T + \lambda_\Gamma^T = 0^T \quad (2.32)$$

or

$$\Gamma(t) = -\lambda_v(t) = p(t) \quad (2.33)$$

using the definition of the primer vector in Equation (2.15). Thus for a VSI engine, the optimal thrust acceleration vector is equal to the primer vector: $\Gamma(t) = p(t)$.

Because of this, Equation (2.5), written as $\ddot{r} = \mathbf{g}(r) + \Gamma$, can be combined with Equation (2.19), as in [7] to yield a fourth-order differential equation in $r$:

$$r^{iv} - 2\dot{\mathbf{g}} \dot{r} + \mathbf{G}(\mathbf{g} - 2\dot{r}) = 0. \quad (2.34)$$

Every solution to Equation (2.34) is an optimal VSI trajectory through the gravity field $\mathbf{g}(r)$. But desired boundary conditions, such as specified position and velocity vectors at the initial and final times, must be satisfied.

Note also that from Equation (2.32)

$$\frac{\partial^2 H}{\partial \Gamma^2} = \frac{\partial}{\partial \Gamma} \left( \frac{\partial H}{\partial \Gamma} \right)^T = I_3 \quad (2.35)$$

where $I_3$ is the $3 \times 3$ identity matrix. Equation (2.35) shows that the (Hessian) matrix of second partial derivatives is positive definite, verifying that $H$ is minimized.

Because the VSI thrust acceleration of Equation (2.33) is continuous, a recently developed procedure [8] to test whether second-order NC and SC are satisfied can be applied. Equation (2.35) shows that an NC for minimum cost (Hessian matrix positive semidefinite) and part of the SC (Hessian matrix positive definite) are satisfied. The other condition that is both an NC and an SC is the Jacobi no-conjugate-point condition. Reference [8] details the recently developed test for that.

2.3 Solution to the Primer Vector Equation

The primer vector equation, Equation (2.19), can be written in first-order form as the linear system

$$\frac{d}{dt} \begin{bmatrix} \mathbf{p} \\ \dot{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_3 & I_3 \\ \mathbf{G} & \mathbf{O}_3 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \dot{\mathbf{p}} \end{bmatrix} \quad (2.36)$$

where $\mathbf{O}_3$ is the $3 \times 3$ zero matrix.
Equation (2.36) is of the form \( \dot{y} = A(t)y \), and its solution can be written in terms of a transition matrix \( \Phi(t, t_o) \) as

\[
y(t) = \Phi(t, t_o)y(t_o)
\]

for a specified initial condition \( y(t_o) \).

Glandorf [9] presents a form of the transition matrix for an inverse-square gravitational field. [In that Technical Note, the missing Equation (2.33) is \( \Phi(t, t_o) = P(t)P^{-1}(t_o) \).]

Note that on an NT (no-thrust or coast) arc, the variational (linearized) state equation is, from Equation (2.5),

\[
\delta \dot{x} = \begin{bmatrix} \delta \dot{r} \\ \delta \dot{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} O_3 & I_3 \\ G & O_3 \end{bmatrix} \begin{bmatrix} \delta r \\ \delta \mathbf{v} \end{bmatrix},
\]

which is the same as Equation (2.36). So the transition matrix in Equation (2.37) is also the transition matrix for the state variation, that is, the state transition matrix [10].

This state transition matrix has the usual properties from linear system theory and is also symplectic [10], which has the useful property that

\[
\Phi^{-1}(t, t_o) = -J\Phi^T(t, t_o)J
\]

where

\[
J = \begin{bmatrix} O_3 & I_3 \\ -I_3 & O_3 \end{bmatrix}.
\]

Note that \( J^2 = -I_6 \), indicating that \( J \) is a matrix analog of the imaginary number \( i \).

Equation (2.39) is useful when the state transition matrix is determined numerically because the inverse matrix \( \Phi^{-1}(t, t_o) = \Phi(t_o, t) \) can be computed without explicitly inverting a \( 6 \times 6 \) matrix.

### 2.4 Application of Primer Vector Theory to an Optimal Impulsive Trajectory

If the primer vector evaluated along an impulsive trajectory fails to satisfy the necessary conditions of Table 2.1 for an optimal solution, the way in which the NC are violated provides information that can lead to a solution that does satisfy the NC. This process was first derived by Lion and Handelsman [3]. For given boundary conditions and a fixed transfer time, an impulsive trajectory can be modified either by allowing a terminal coast or by adding a midcourse impulse. A terminal coast can be either an initial coast, in which the first impulse occurs after the initial time, or a final coast, in which the final impulse occurs before the final time. In the former case, the spacecraft coasts along the initial orbit after the initial time until the first impulse
Figure 2.3. A fixed-time impulsive rendezvous trajectory.

occurs. In the latter case, the rendezvous actually occurs before the final time, and the spacecraft coasts along the final orbit until the final time is reached.

To determine when a terminal coast will result in a trajectory that has a lower fuel cost, consider the two-impulse fixed-time rendezvous trajectory shown in Figure 2.3. In the two-body problem, if the terminal radii $r_o$ and $r_f$ are specified along with the transfer time $\tau \equiv t_f - t_o$, the solution to Lambert’s Problem [10] [11] provides the terminal velocity vectors $v_o^+$ (after the initial impulse) and $v_f^-$ (before the final impulse) on the transfer orbit. Because the velocity vectors are known on the initial orbit ($v_o^-$ before the first impulse) and on the final orbit ($v_f^+$ after the final impulse), the required velocity changes can be determined as

\[
\Delta v_o = v_o^+ - v_o^-
\]  

(2.41)

and

\[
\Delta v_f = v_f^+ - v_f^-.
\]  

(2.42)

Once the vector velocity changes are known, the primer vector can be evaluated along the trajectory to determine if the NC are satisfied. In order to satisfy the NC that on an optimal trajectory the primer vector at an impulse time is a unit vector in the direction of the impulse, one imposes the following boundary conditions on the primer vector

\[
p(t_o) \equiv p_o = \frac{\Delta v_o}{\Delta v_o}
\]  

(2.43)

\[
p(t_f) \equiv p_f = \frac{\Delta v_f}{\Delta v_f}
\]  

(2.44)
The primer vector can then be evaluated along the transfer orbit using the $6 \times 6$ transition matrix solution of Equation (2.37)

$$
\begin{bmatrix}
    p(t) \\
    \dot{p}(t)
\end{bmatrix} = \Phi(t, t_o) \begin{bmatrix}
    p(t_o) \\
    \dot{p}(t_o)
\end{bmatrix}
$$

(2.45)

where the $3 \times 3$ partitions of the $6 \times 6$ transition matrix are designated as

$$
\Phi(t, t_o) \equiv \begin{bmatrix}
    M(t, t_o) & N(t, t_o) \\
    S(t, t_o) & T(t, t_o)
\end{bmatrix}.
$$

(2.46)

Equation (2.45) can then be evaluated for the fixed terminal times $t_o$ and $t_f$ to yield

$$
p_f = M_{f, o}p_o + N_{f, o}\dot{p}_o
$$

(2.47)

and

$$
\dot{p}_f = S_{f, o}p_o + T_{f, o}\dot{p}_o
$$

(2.48)

where the abbreviated notation is used that $p_f \equiv p(t_f), M_{f, o} \equiv M(t_f, t_o)$, and so on. Equation (2.47) can be solved for the initial primer vector rate

$$
\dot{p}_o = N_{f, o}^{-1}[p_f - M_{f, o}p_o]
$$

(2.49)

where the inverse matrix $N_{f, o}^{-1}$ exists except for isolated values of $\tau = t_f - t_o$. With both the primer vector and the primer vector rate known at the initial time, the primer vector along the transfer orbit for $t_o \leq t \leq t_f$ can be calculated as using Equations (2.43–2.46, 2.49) as

$$
p(t) = N_{t, o}N_{f, o}^{-1}\Delta v_f + [M_{t, o} - N_{t, o}N_{f, o}^{-1}M_{f, o}]\Delta v_o.
$$

(2.50)

2.4.1 Criterion for a Terminal Coast

One of the options available to modify a two-impulse solution that does not satisfy the NC for an optimal transfer is to include a terminal coast period in the form of either an initial coast, a final coast, or both. To do this, one allows the possibility that the initial impulse occurs at time $t_o + dt_o$ due to a coast in the initial orbit of duration $dt_o > 0$ and that the final impulse occurs at a time $t_f + dt_f$. In the case of a final coast, $dt_f < 0$ in order that the final impulse occur prior to the nominal final time, allowing a coast in the final orbit until the nominal final time. A negative value of $dt_o$ or a positive value of $dt_f$ also has a physical interpretation as will be seen.

To determine whether a terminal coast will lower the cost of the trajectory, an expression for the difference in cost between the perturbed trajectory (with the terminal coasts) and the nominal trajectory (without the coasts) must be derived. The
discussion that follows summarizes and interprets results by Lion and Handelsman [3]. The cost on the nominal trajectory is simply

\[ J = \Delta v_o + \Delta v_f \]  

(2.51)

for the two-impulse solution. In order to determine the differential change in the cost due to the differential coast periods the concept of a noncontemporaneous, or “skew” variation is needed. This variation combines two effects: the variation due to being on a perturbed trajectory and the variation due to a difference in the time of the impulse. The variable \( d \) will be used to denote a noncontemporaneous variation in contrast to the variable \( \delta \) that represents a contemporaneous variation, as in Equation (2.38). The rule for relating the two types of variations is given by

\[ d\mathbf{x}(t_o) = \delta \mathbf{x}(t_o) + \mathbf{\dot{x}}_o^* dt_o \]  

(2.52)

where \( \mathbf{\dot{x}}_o^* \) is the derivative on the nominal (unperturbed) trajectory at the nominal final time and the variation in the initial state has been used as an example.

Next, the noncontemporaneous variation in the cost must be determined. Because the coast periods result in changes in the vector velocity changes, the variation in the cost can be expressed, from Equation (2.51) as

\[ dJ = \frac{\partial \Delta v_o}{\partial \Delta v_o} d\Delta v_o + \frac{\partial \Delta v_f}{\partial \Delta v_f} d\Delta v_f. \]  

(2.53)

Using the fact that for any vector \( \mathbf{a} \) having magnitude \( a \)

\[ \frac{\partial a}{\partial \mathbf{a}} = \frac{\mathbf{a}^T}{a} \]  

(2.54)

the variation in the cost in Equation (2.53) can be expressed as

\[ dJ = \frac{\Delta v_o^T}{\Delta v_o} d\Delta v_o + \frac{\Delta v_f^T}{\Delta v_f} d\Delta v_f. \]  

(2.55)

Finally, Equation (2.55) can be rewritten in terms of the initial and final primer vector using the conditions of Equations (2.43–2.44) as

\[ dJ = \mathbf{p}_o^T d\Delta v_o + \mathbf{p}_f^T d\Delta v_f. \]  

(2.56)

The analysis in [3] leads to the result that

\[ dJ = -\mathbf{p}_o^T \Delta v_0 dt_o - \mathbf{p}_f^T \Delta v_f dt_f \]  

(2.57)

The final form of the expression for the variation in cost is obtained by expressing the vector velocity changes in terms of the primer vector using Equations (2.43–2.44) as

\[ dJ = -\Delta v_0 \mathbf{p}_o^T dt_o - \Delta v_f \mathbf{p}_f^T dt_f. \]  

(2.58)
In Equation (2.58), one can identify the gradients of the cost with respect to the terminal impulse times \( t_o \) and \( t_f \) as

\[
\frac{\partial J}{\partial t_o} = -\Delta v_o \hat{p}_o^T \hat{p}_o
\]  

(2.59)

and

\[
\frac{\partial J}{\partial t_f} = -\Delta v_f \hat{p}_f^T \hat{p}_f.
\]  

(2.60)

One notes that the dot products in Equations (2.59–2.60) are simply the slopes of the primer magnitude time history at the terminal times, due to the fact that \( p^2 = \hat{p}^T \hat{p} \) and, after differentiation with respect to time, \( 2\hat{p}\ddot{p} = 2\dot{\hat{p}}^T \hat{p} \). Because \( \rho = 1 \) at the impulse times,

\[
\dot{\hat{p}}^T \hat{p} = \dot{\hat{p}}.
\]  

(2.61)

The criteria for adding an initial or final coast can now be summarized by examining the algebraic signs of the gradients in Equations (2.59–2.60):

If \( \dot{\hat{p}}_o > 0 \), an initial coast (represented by \( dt_o > 0 \)) will lower the cost. Similarly, if \( \dot{\hat{p}}_f < 0 \), a final coast (represented by \( dt_f < 0 \)) will lower the cost.

It is worth noting that, conversely, if \( \dot{\hat{p}}_o \leq 0 \), an initial coast will not lower the cost. This is consistent with the NC for an optimal solution and represents an alternate proof of the NC that \( p \leq 1 \) on an optimal solution. Similarly, if \( \dot{\hat{p}}_f \geq 0 \), a final coast will not lower the cost. However, one can interpret these results even further. If \( \dot{\hat{p}}_o < 0 \), a value of \( dt_o < 0 \) yields \( dJ < 0 \), indicating that an earlier initial impulse time would lower the cost. This is the opposite of an initial coast and simply means that the cost can be lowered by increasing the transfer time by starting the transfer earlier. Similarly, a value of \( \dot{\hat{p}}_f > 0 \) implies that a \( dt_f > 0 \) will yield \( dJ < 0 \). In this case, the cost can be lowered by increasing the transfer time by increasing the final time. From these observations, one can conclude that for a time-open optimal solution, such as the Hohmann transfer, the slopes of the primer magnitude time history must be zero at the terminal times, indicating that no improvement in the cost can be made by slightly increasing or decreasing the times of the terminal impulses. Figure 2.4 shows the primer time history for a Hohmann transfer rendezvous trajectory. An initial coast of 0.889 OOP is required to obtain the correct phase angle of the target body for the given ila and there is no final coast.

Figure 2.5 shows an example of a primer history that violates the NC in a manner indicating that an initial coast or final coast or both will lower the cost. The final radius is 1.6, the ila is 90°, and the transfer time is 0.9 OOP.

In this case, the choice is made to add an initial coast, and the gradient of Equation (2.59) is used in a nonlinear programming (NLP) algorithm to iterate on the time of the first impulse. This is a one-dimensional search in which small changes in the time of the first impulse are made using the gradient of Equation (2.59) until the gradient is driven to zero. On each iteration, new values for the terminal velocity
2.4 Application of Primer Vector Theory

Hohmann Trajectory: $\Delta V_H = 0.28446$

$t_H = 1.8077 \ \text{OOP}$
$t_{\text{wait}} = 0.88911 \ \text{OOP}$
$t_{\text{ellipse}} = 0.91856 \ \text{OOP}$

$\beta_0 = 270^\circ$ Canonical Distance

$\rho_0 = 270^\circ$ Canonical Distance

Figure 2.4. Hohmann transfer orbit and primer magnitude.

$t_f = 0.9 \ \text{OOP}, \ \Delta V = 0.37466$

 ila = 90

0.24379
0.13087

Figure 2.5. Primer magnitude indicating initial/final coast.
changes are calculated by re-solving Lambert’s Problem and a new primer vector solution is obtained. Note that once the iteration begins, the time of the first impulse is no longer $t_0$, but a later value denoted by $t_1$. In a similar way, if the final impulse time becomes an iteration variable, it is denoted by $t_n$ where the last impulse is considered to be the $n$th impulse. For a two-impulse trajectory, $n = 2$, but as will be seen shortly, optimal solutions can require more than two impulses. When the times of the first and last impulse become iteration variables, in all the formulas in the preceding analysis, the subscript $o$ is replaced by 1 everywhere and $f$ is replaced by $n$.

Figure 2.6 shows the converged result of an iteration on the time of the initial impulse.

Note that the necessary condition $p \leq 1$ is satisfied and the gradient of the cost with respect to $t_1$, the time of the first impulse (at approximately $t_1 = 0.22$), is zero because $\dot{p}_1 = 0$, making the gradient of Equation (2.59) equal to zero. This simply means that a small change in $t_1$ will cause no change in the cost, that is, the cost has achieved a stationary value and satisfies the first-order necessary conditions. Comparing Figures 2.5 and 2.6, one notes that the cost has decreased significantly from 0.37466 to 0.21459, and that an initial coast is required but no final coast is required.

### 2.4.2 Criterion for Addition of a Midcourse Impulse

Besides terminal coasts, the addition of one or more midcourse impulses is another potential way of lowering the cost of an impulsive trajectory. The addition of an
impulse is more complicated than including terminal coasts because, in the general case, four new parameters are introduced: three components of the position of the impulse and the time of the impulse. One must first derive a criterion that indicates that the addition of an impulse will lower the cost and then determine where in space and when in time the impulse should occur. The where and when will be done in two steps. The first step is to determine initial values of position and time of the added impulse that will lower the cost. The second step is to iterate on the values of position and time using gradients that will be developed, until a minimum of the cost is achieved. Note that this procedure is more complicated than for terminal coasts, because the starting value of the coast time for the iteration was simply taken to be zero, that is, no coast.

When considering the addition of a midcourse impulse, let us assume $dt_o = dt_f = 0$, that is, there are no terminal coasts. Because we are doing a first-order perturbation analysis, superposition applies and we can combine the previous results for terminal coasts easily with our new results for a midcourse impulse. Also, we will discuss the case of adding a third impulse to a two-impulse trajectory, but the same theory applies to the case of adding a midcourse impulse to any two-impulse segment of an $n$-impulse trajectory. The cost on the nominal, two-impulse trajectory is given by Equation (2.50)

$$J = \Delta v_o + \Delta v_f.$$  

The variation in the cost due to adding an impulse is given by adding the midcourse velocity change magnitude $\Delta v_m$ to Equation (2.56)

$$dJ = p_o^T d\Delta v_o + \Delta v_m + p_f^T d\Delta v_f. \tag{2.62}$$

The analysis in [3] results in

$$dJ = \Delta v_m \left(1 - p_m \frac{\Delta v_m}{\Delta v_m} \right). \tag{2.63}$$

In Equation (2.63), the expression for $dJ$ involves a dot product between the primer vector and a unit vector. If the numerical value of this dot product is greater than one, $dJ < 0$ and the perturbed trajectory has a lower cost than the nominal trajectory. In order for the value of the dot product to be greater than one, it is necessary that $p_m > 1$. Here again we have an alternative derivation of the necessary condition that $p \leq 1$ on an optimal trajectory. We also have the criterion that tells us when the addition of a midcourse impulse will lower the cost.

If the value of $p(t)$ exceeds unity along the trajectory, the addition of a midcourse impulse at a time for which $p > 1$ will lower the cost.

Figure 2.7 shows a primer magnitude time history that indicates the need for a midcourse impulse (but not for a terminal coast). The final radius is 2, the ila is $270^\circ$, and the transfer time is relatively small, equal to 0.5 OOP.

The first step is to determine initial values for the position and time of the midcourse impulse. From Equation (2.63) it is evident that for a given $p_m$, the largest
decrease in the cost is obtained by maximizing the value of the dot product, that is, by choosing a position for the impulse that causes $\Delta v_m$ to be parallel to the vector $p_m$ and by choosing the time $t_m$ to be the time at which the primer magnitude has a maximum value. Choosing the position of the impulse so that the velocity change is in the direction of the primer vector sounds familiar because it is one of the necessary conditions derived previously, but how to determine this position is not at all obvious, and we will have to derive an expression for this. Choosing the time $t_m$ to be the time of maximum primer magnitude does not guarantee that the decrease in cost is maximized, because the value of $\Delta v_m$ in the expression for $dJ$ depends on the value of $t_m$. However, all we are doing is obtaining an initial position and time of the midcourse impulse to begin an iteration process. As long as our initial choice represents a decrease in the cost, we will opt for the simple device of choosing the time of maximum primer magnitude as our initial estimate of $t_m$. In Figure 2.7, $t_m$ is 0.1.

Having determined an initial value for $t_m$, the initial position of the impulse, namely the value $\delta r_m$ to be added to $r_m$, must also be determined. Obviously $\delta r_m$ must be nonzero, otherwise the midcourse impulse would have zero magnitude. The property that must be satisfied in determining $\delta r_m$ is that $\Delta v_m$ be parallel to $p_m$. The analysis of [3] results in

$$\Delta v_m = A \delta r_m \quad (2.64)$$
where the matrix $A$ is defined as

$$A \equiv -(M_{fm}^T N_{fm}^{-T} + T_{mo} N_{mo}^{-1}).$$

(2.65)

Next, in order to have $\Delta v_m$ parallel to $p_m$, it is necessary that $\Delta v_m = \varepsilon p_m$ with scalar $\varepsilon > 0$. Combining this fact with Equation (2.64) yields

$$A \delta r_m = \Delta v_m = \varepsilon p_m$$

(2.66)

which yields the solution for $\delta r_m$ as

$$\delta r_m = \varepsilon A^{-1} p_m$$

(2.67)

assuming $A$ is invertible.

The question then arises how to select a value for the scalar $\varepsilon$. Obviously too large a value will violate the linearity assumptions of the perturbation analysis. This is not addressed in [3], but one can maintain a small change by specifying

$$\frac{\delta r_m}{r_m} = \beta$$

(2.68)

where $\beta$ is a specified small positive number such as 0.05. Equation (2.67) then yields a value for $\varepsilon$

$$\varepsilon \left| \frac{A^{-1} p_m}{r_m} \right| = \beta \Rightarrow \varepsilon = \frac{\beta r_m}{\left| A^{-1} p_m \right|}.$$  

(2.69)

If the resulting $dJ \geq 0$, then decrease $\varepsilon$ and repeat Equation (2.67). One should never accept a midcourse impulse position that does not decrease the cost, because a sufficiently small $\varepsilon$ will always provide a lower cost.

The initial values of midcourse impulse position and time are now determined. One adds the $\delta r_m$ of Equation (2.67) to the value of $r_m$ on the nominal trajectory at the time $t_m$ at which the primer magnitude achieves its maximum value (greater than one).

The primer history after the addition of the initial midcourse impulse is shown in Figure 2.8. Note that $p_m = 1$ but $\dot{p}_m$ is discontinuous and the primer magnitude exceeds unity, both of which violate the NC. However, the addition of the midcourse impulse has decreased the cost slightly, from 1.7555 to 1.7549.

### 2.4.3 Iteration on a Midcourse Impulse Position and Time

To determine how to efficiently iterate on the components of position of the midcourse impulse and its time, one needs to derive expressions for the gradients of the cost with respect to these variables. To do this, one must compare the three-impulse trajectory (or three-impulse segment of an $n$-impulse trajectory) that resulted from the addition of the midcourse impulse with a perturbed three-impulse trajectory.
Figure 2.8. Initial (nonoptimal) three-impulse primer magnitude.

Note that, unlike a terminal coast, the values of \( dr_m \) and \( dt_m \) are independent. (By contrast, on an initial coast \( dr_o = v_o^- dt_o \) and on a final coast \( dr_f = v_f^+ dt_f \).)

The cost on the nominal three-impulse trajectory is

\[
J = \Delta v_o + \Delta v_m + \Delta v_f
\]

and the variation in the cost due to perturbing the midcourse time and position is

\[
dJ = \frac{\partial \Delta v_o}{\partial \Delta v_o} d\Delta v_o + \frac{\partial \Delta v_m}{\partial \Delta v_m} d\Delta v_m + \frac{\partial \Delta v_f}{\partial \Delta v_f} d\Delta v_f
\]

which, analogous to Equation (2.56), can be written as

\[
dJ = p_o^T d\Delta v_o + p_m^T d\Delta v_m + p_f^T d\Delta v_f.
\]

The analysis of [3] leads to the result that

\[
dJ = (\dot{p}_m^+ - \dot{p}_m^-)^T dr_m - (p_m^T v_m^+ - p_m^T v_m^-) dt_m.
\]
In Equation (2.73), a discontinuity in $\dot{p}_m$ has been allowed because there is no guarantee that it will be continuous at the inserted midcourse impulse, as demonstrated in Figure 2.8.

Equation (2.73) can be written more simply in terms of the Hamiltonian function Equation (2.22) for $p_m = 1$: $H_m = \dot{p}_m^T v_m - \dot{p}_m^T g_m$ (for which the second term $\dot{p}_m^T g_m$ is continuous because $p_m = \Delta v_m/\Delta v_m$ and $g_m(r_m)$ are continuous).

$$dJ = (\dot{p}_m^+ - \dot{p}_m^-)^T dr_m - (H_m^+ - H_m^-) dt_m.$$  \hspace{1cm} (2.74)

Equation (2.74) provides the gradients of the cost with respect to the independent variations in the position and time of the midcourse impulse for use in a nonlinear programming algorithm:

$$\frac{\partial J}{\partial r_m} = (\dot{p}_m^+ - \dot{p}_m^-)$$  \hspace{1cm} (2.75)

and

$$\frac{\partial J}{\partial t_m} = -(H_m^+ - H_m^-).$$  \hspace{1cm} (2.76)

As a solution satisfying the NC is approached, the gradients tend to zero, in which case both the primer rate vector $\dot{p}_m$ and the Hamiltonian function $H_m$ become continuous at the midcourse impulse.

Note that when the NC are satisfied, the gradient with respect to $t_m$ in Equation (2.76) being zero implies that

$$H_m^+ - H_m^- = 0 = \dot{p}_m^T (v_m^+ - v_m^-) = \dot{p}_m^T \Delta v_m = \Delta v_m \dot{p}_m^T p_m = 0$$ \hspace{1cm} (2.77)

which, in turn, implies that $\dot{p}_m = 0$, indicating that the primer magnitude attains a local maximum value of unity. This is consistent with the NC that $p \leq 1$ and that $\dot{p}$ be continuous.

Figure 2.2 shows the converged, optimal three-impulse trajectory that results from improving the primer histories shown in Figures 2.7 and 2.8. Note that the final cost of 1.3681 is significantly less that the value of 1.7555 prior to adding the midcourse impulse. Also, the time of the midcourse impulse changed during the iteration from its initial value of 0.1 to a final value of approximately 0.17.

The absolute minimum cost solution for the final radius and ila value of Figure 2.2 is, of course, the Hohmann transfer shown in Figure 2.4. Its cost is significantly less at 0.28446, but the transfer time is nearly three times as long at 1.8077 OOP. Of this, 0.889 OOP is an initial coast to achieve the correct target phase angle for the Hohmann transfer. Depending on the specific application, the total time required may be unacceptably long.

(As a side note, a simple proof of the global optimality of the Hohmann transfer using ordinary calculus rather than primer vector theory is given in [12].)
REFERENCES


