INDIRECT METHODS FOR SPACE TRAJECTORY OPTIMIZATION

Lorenzo Casalino

Outline

- introduction
- indirect methods: how to make them
  - optimization of discrete systems
  - optimization of continuous systems
    - calculus of variations and indirect optimization
    - solution of boundary value problems
  - indirect trajectory optimization
- indirect methods: how to make them work
  - general rules
  - hints and tips

Trajectory Design

- trajectory design usually requires maximization or minimization of a *performance index* (e.g., maximize final mass or payload, minimize, ΔV, propellant consumption or trip time) for given mission requirements
- typical missions
  - transfer between specified orbits
  - rendezvous mission (transfer from a specified orbit to a definite time-dependent position on a target orbit)
  - interplanetary missions (rendezvous missions with planets' gravity)
  - station keeping

Trajectory Model

- spacecraft state defined by state variables x_i, $i = 1, n$ (position, velocity, mass, angular position and velocity for attitude control)
- a propulsion system can produce thrust (or angular momentum) to modify the rate of change (derivative) of the spacecraft velocity
- thrust components are typically the control variables u_j, $j = 1, m$

Optimization Problem

- system of equations $\dot{x}_i = f_i[x_1(t), ..., x_n(t), u_1(t), ..., u_m(t), t]$, $i = 1, n$
- boundary conditions concerning state (and possibly control) variables at initial, final and possibly intermediate points
- find the control variables $u_j(t)$ $(j = 1, m)$ and the corresponding trajectory $x_i(t)$ $(i = 1, n)$ that maximize a given performance index J while satisfying differential equations and boundary conditions

Trajectory Assumptions

- preliminary design usually (but not necessarily) admits simplifications
  - two-body problem
  - patched-conic approximation for interplanetary transfers
- it is often convenient to split a complex trajectory into elementary legs that are then patched together

Solution Methods (1)

- direct methods
  - discretization of trajectory and controls
  - start from tentative solution
  - evaluate performance index and constraints and respective gradients
  - change tentative solution with the aim of improving performance index and constraint fulfillment

Solution Methods (2)

- indirect methods
  - continuous system described by differential equations with assigned boundary conditions
  - derive first-order necessary conditions for optimality and define a boundary value problem
  - start from tentative solution
  - evaluate error on boundary conditions and error gradient
  - change tentative solution with the aim of reducing the error

Solution Methods (3)

- evolutionary algorithms
  - more suitable to deal with a low number of parameters (impulsive missions)
  - solution described by finite set of variable values
  - random initialization of a population of different solutions
  - new solutions are created by combining old solutions (typically, using some random process) with the aim of creating new populations with improved performance index

Direct Methods

- pros
  - capability of treating complex problems
  - easier treatment of constraint “structure”
  - higher robustness
- cons
  - scarce accuracy (may require solution refinement)
  - high computational cost
  - solution may depend on tentative values and be suboptimal

Indirect Methods

- pros
  - high accuracy
  - low computational cost and time
  - theoretical insight
- cons
  - difficult treatment of complex equations or constraints and necessary preliminary assumption of constraint "structure" (an "a posteriori" analysis may suggest changes)
  - lower robustness
  - find stationary solutions, possibly suboptimal, and dependent on tentative values

Evolutionary Algorithms

- pros
  - no tentative solution
  - higher robustness
  - “global” optimization
- cons
  - (possibly) scarce accuracy (may require solution refinement)
  - difficult treatment of constraints
  - heuristic methods: no proof of finding actual optimum

Shooting Method

- both direct and indirect methods present unknown initial values and constraints or optimality conditions to be satisfied
- shooting method
  - choose initial tentative values for the unknowns
  - compute error on constraints
  - change tentative values
- errors on constraints may be very sensitive to the initial values, affecting convergence

Multiple Shooting

- split the trajectory into phases
- introduce the variable values at the start of each phase as additional unknowns
- enforce conditions for trajectory consistency (e.g., variable continuity) at the phase junctions
- reduced sensitivity, improved convergence
- higher computational cost ? not always

Discrete Systems

Summary

- notation and assumptions
- one-variable optimization
- unconstrained multivariable optimization
- constrained optimization
  - equality constraints
  - inequality constraints

Vectorial Notation (1)

- a column vector is indicated by a bold character $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix}$
- a row vector is written as the transpose of a column vector $\boldsymbol{a}^T = (a_1, a_2, \cdots, a_n)$
- the scalar product of two vectors (e.g., $\boldsymbol{a}$ and $\boldsymbol{b}$, same dimension) is written as $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T \boldsymbol{b} = \boldsymbol{b}^T \boldsymbol{a}$

Vectorial Notation (2)

- the derivative of a column vector (e.g., $\boldsymbol{a}$, n components) with respect to a scalar variable (e.g., t) is a column vector with components equal to the derivatives of the components of $\boldsymbol{a}$ with respect to t

$$\mathrm{d}\boldsymbol{a}/\mathrm{d}t = \dot{\boldsymbol{a}} = \begin{pmatrix} \mathrm{d}a_1/\mathrm{d}t \\ \mathrm{d}a_2/\mathrm{d}t \\ \dots \\ \mathrm{d}a_n/\mathrm{d}t \end{pmatrix}$$

- the derivative of a scalar quantity (e.g., ϕ) with respect to a column vector (e.g., $\boldsymbol{b}$, m components) is defined as a row vector with components equal to the derivatives of ϕ with respect to the components of $\boldsymbol{b}$

$$\mathrm{d}\phi/\mathrm{d}\boldsymbol{b} = (\mathrm{d}\phi/\mathrm{d}b_1, \mathrm{d}\phi/\mathrm{d}b_2, \cdots, \mathrm{d}\phi/\mathrm{d}b_m)$$

Vectorial Notation (3)

- by extension, the derivative of a column vector (e.g., $\boldsymbol{a}$) with respect to another column vector (e.g., $\boldsymbol{b}$) is a matrix
- the i-th row contains the derivatives of the i-th component of $\boldsymbol{a}$ with respect to $\boldsymbol{b}$
- the j-th column contains the derivatives of the constraints $\boldsymbol{a}$ with respect to the j-th component of $\boldsymbol{b}$

$$\mathrm{d}\boldsymbol{a}/\mathrm{d}\boldsymbol{b} = \begin{bmatrix} \mathrm{d}a_1/\mathrm{d}b_1 & \mathrm{d}a_1/\mathrm{d}b_2 & \cdots & \mathrm{d}a_1/\mathrm{d}b_m \\ \mathrm{d}a_2/\mathrm{d}b_1 & \mathrm{d}a_2/\mathrm{d}b_2 & \cdots & \mathrm{d}a_2/\mathrm{d}b_m \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{d}a_n/\mathrm{d}b_1 & \mathrm{d}a_n/\mathrm{d}b_2 & \cdots & \mathrm{d}a_n/\mathrm{d}b_m \end{bmatrix}$$

Assumptions

- performance index and boundary conditions are continuous and differentiable with respect to the variables
- existence of an optimal solution is usually assumed

One-Variable Unconstrained Optimization (1)

- $\phi(\hat{x})$ is maximum if $\phi(\hat{x}) > \phi(x)$ (strong) or $\phi(\hat{x}) \geq \phi(x)$ (weak) for any $x \neq \hat{x}$
- $\phi(\hat{x})$ is maximum (strong) if $\mathrm{d}\phi = \phi(\hat{x} + \mathrm{d}x) - \phi(\hat{x}) < 0$ for any choice of $\mathrm{d}x \neq 0$ (weak maximum if $\mathrm{d}\phi \leq 0$)
- if $\mathrm{d}x$ is small (local variation) use second-order Taylor's expansion of ϕ
  $\phi(x + \mathrm{d}x) = \phi(x) + g\mathrm{d}x + \frac{1}{2}H\mathrm{d}x^2$
- the first derivative of ϕ is $g = \dfrac{\partial \phi}{\partial x}$
- the second derivative of ϕ is $H = \dfrac{\partial^2 \phi}{\partial x^2}$

One-Variable Unconstrained Optimization (2)

- a stationary point is characterized by $g = 0$
- necessary conditions for a (local) maximum $g = 0$, $H \leq 0$
- sufficient conditions for a (local) maximum $g = 0$, $H < 0$
- $H = 0$ requires computation of additional derivatives to define the nature of the stationary point - if the first nonzero derivative is of order j
  - neither max nor min if j is odd
  - max if j is even and the derivative is negative
  - min if j is even and the derivative is positive
- the sign of H must be reversed for a minimum

Unconstrained Optimization (1)

- maximize $\phi(\boldsymbol{x})$
- $\boldsymbol{x}$ n-component vector of variables $\boldsymbol{x}^T = (x_1, x_2, ..., x_n)$
- ϕ is maximum if d$\phi \leq 0$ for any choice of d$\boldsymbol{x}$
- second-order Taylor's expansion of ϕ (small variations)
  $\phi(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) = \phi(\boldsymbol{x}) + \boldsymbol{g}^T \mathrm{d}\boldsymbol{x} + \frac{1}{2}\mathrm{d}\boldsymbol{x}^T \boldsymbol{H} \mathrm{d}\boldsymbol{x}$
- the *gradient* of ϕ is $\boldsymbol{g}$ (n-component vector)
  $\boldsymbol{g}^T = \frac{\partial \phi}{\partial \boldsymbol{x}} = (\partial\phi/\partial x_1, \partial\phi/\partial x_1, ..., \partial\phi/\partial x_n)$

Unconstrained Optimization (2)

- the *Hessian* of ϕ is $[\boldsymbol{H}]$ (nxn symmetric matrix)

$$[\boldsymbol{H}] = \frac{\partial \left(\frac{\partial \phi}{\partial \boldsymbol{x}}\right)^T}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial^2 \phi}{\partial x_1 \partial x_1} & \frac{\partial^2 \phi}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 \phi}{\partial x_n \partial x_1} \\ \frac{\partial^2 \phi}{\partial x_1 \partial x_2} & \frac{\partial^2 \phi}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 \phi}{\partial x_n \partial x_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \phi}{\partial x_1 \partial x_n} & \frac{\partial^2 \phi}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 \phi}{\partial x_n \partial x_n} \end{bmatrix}$$

Unconstrained Optimization (3)

- first-order expansion of ϕ provides $\mathrm{d}\phi = \frac{\partial \phi}{\partial \boldsymbol{x}}\mathrm{d}\boldsymbol{x} = \boldsymbol{g}^T \mathrm{d}\boldsymbol{x}$
- ϕ being a local maximum (or minimum) requires $\boldsymbol{g} = 0$ (necessary condition for a stationary point)
- an additional condition is required to determine the nature of the stationary point
- considering the second variation of ϕ at a stationary point ($\boldsymbol{g} = 0$), one has $\mathrm{d}\phi = \mathrm{d}\boldsymbol{x}^T \boldsymbol{H} \mathrm{d}\boldsymbol{x}$
- for a maximum $\mathrm{d}\boldsymbol{x}^T \boldsymbol{H} \mathrm{d}\boldsymbol{x} \leq 0$ for any $\mathrm{d}\boldsymbol{x}$, that is $\boldsymbol{H}$ negative semidefinite for a maximum (non-positive eigenvalues)
- sufficient conditions for a maximum $\boldsymbol{g} = 0$ and $\mathrm{d}\boldsymbol{x}^T \boldsymbol{H} \mathrm{d}\boldsymbol{x} < 0$ for any $\mathrm{d}\boldsymbol{x} \neq 0$, that is $\boldsymbol{H}$ negative definite for a maximum (negative eigenvalues)

Equality Constrained Optimization (1)

- maximize $\phi(\boldsymbol{x})$ ($\boldsymbol{x}$ is a n-component vector of variables) with $\boldsymbol{c}(\boldsymbol{x}) = 0$ ($\boldsymbol{c}$ is a m-component vector of constraints, $\boldsymbol{c}^T = [c_1, c_2, ..., c_m]$, $m < n$)
- feasible points are those which satisfy $\boldsymbol{c}(\boldsymbol{x}) = 0$
- at a feasible point ϕ is maximum if $\mathrm{d}\phi \leq 0$ for any *admissible* (i.e., that verifies $\boldsymbol{c}(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) = 0$) choice of $\mathrm{d}\boldsymbol{x}$
- at a feasible point, admissible variations require $\mathrm{d}\boldsymbol{c} = \boldsymbol{c}(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x}) = 0$

Equality Constrained Optimization (2)

- small variations are considered (local maximum)
- first order expansion of $\boldsymbol{c}$ gives $\boldsymbol{c}(\boldsymbol{x}+\mathrm{d}\boldsymbol{x})=\boldsymbol{c}(\boldsymbol{x})+[\boldsymbol{G}]\mathrm{d}\boldsymbol{x}$
- $[\boldsymbol{G}]$ is the *Jacobian* ($m \times n$ matrix)

$$[\boldsymbol{G}]=\left[\frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}\right]=\begin{bmatrix} \frac{\partial c_1}{\partial x_1} & \frac{\partial c_1}{\partial x_2} & \dots & \frac{\partial c_1}{\partial x_n} \\ \frac{\partial c_2}{\partial x_1} & \frac{\partial c_2}{\partial x_2} & \dots & \frac{\partial c_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial c_m}{\partial x_1} & \frac{\partial c_m}{\partial x_2} & \dots & \frac{\partial c_m}{\partial x_n} \end{bmatrix}$$

- the i-th row of $\boldsymbol{G}$ is the gradient of the i-th constraint c_i
- the j-th column of $\boldsymbol{G}$ contains the derivatives of the constraints with respect to the j-th variable x_j
- admissible (small) variations require $[\boldsymbol{G}]\mathrm{d}\boldsymbol{x}=0$ (m equations)

Equality Constrained Optimization (3)

- n variables, m equations allow for $n-m$ free parameters
- define $\boldsymbol{y}^T = [x_1, x_2, ..., x_m]$ as m state variables and $\boldsymbol{u}^T = [x_{m+1}, x_{m+2}, ..., x_n]$ as $n-m$ control (or decision) variables (the choice is usually not unique)
- obtain admissible variations d$\boldsymbol{y}$ as a function of d$\boldsymbol{u}$ from $[\boldsymbol{G}_y]\mathrm{d}\boldsymbol{y} + [\boldsymbol{G}_u]\mathrm{d}\boldsymbol{u} = 0$
- first order expansion $\phi(\boldsymbol{u} + \mathrm{d}\boldsymbol{u}) = \phi(\boldsymbol{u}) + \boldsymbol{g_y}^T\mathrm{d}\boldsymbol{y} + \boldsymbol{g_u}^T\mathrm{d}\boldsymbol{u}$
- $[\boldsymbol{G}_y] = \partial\boldsymbol{c}/\partial\boldsymbol{y}$, $[\boldsymbol{G}_u] = \partial\boldsymbol{c}/\partial\boldsymbol{u}$, $\boldsymbol{g}_y^T = \partial\phi/\partial\boldsymbol{y}$, $\boldsymbol{g}_u^T = \partial\phi/\partial\boldsymbol{u}$
- ϕ is maximum if $\mathrm{d}\phi \leq 0$ for any choice of d$\boldsymbol{u}$
- necessary condition for a stationary point $-\boldsymbol{g_y}^T[\boldsymbol{G}_y]^{-1}[\boldsymbol{G}_u] + \boldsymbol{g_u}^T = 0$

Equality Constrained Optimization (4)

- alternatively, maximize the augmented function $\phi^* = \phi(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{c}$ ($\boldsymbol{\lambda}$ is a m-component vector of adjoint parameters to be determined)
- ϕ and ϕ^* coincide for any choice of $\boldsymbol{\lambda}$ if the constraints are verified
- second-order Taylor's expansion of ϕ^* (local maxima are sought)
  $\phi^*(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) = \phi^*(\boldsymbol{x}) + (\boldsymbol{g}^T + \boldsymbol{\lambda}^T [\boldsymbol{G}])\mathrm{d}\boldsymbol{x} + \frac{1}{2}\mathrm{d}\boldsymbol{x}^T [\boldsymbol{H}^*]\mathrm{d}\boldsymbol{x}$
- augmented Hessian $[\boldsymbol{H}^*] = [\boldsymbol{H}] + \partial[\partial(\boldsymbol{\lambda}^T \boldsymbol{c}/\partial \boldsymbol{x})^T]\partial \boldsymbol{x}$, that is
  $H^*_{ij} = H_{ij} + \sum_{k=1}^{m} \lambda_k \frac{\partial^2 c_k}{\partial x_i \partial x_j}$

Equality Constrained Optimization (5)

- $\boldsymbol{\lambda}$ can be chosen to nullify the first variation of ϕ^* for *any* choice of d$\boldsymbol{x}$ (also those that do not satisfy $\boldsymbol{c} = 0$)
- first-order necessary condition for maximum ϕ is $\boldsymbol{g}^T + \boldsymbol{\lambda}^T[\boldsymbol{G}] = 0$, that is, $\boldsymbol{g} + [\boldsymbol{G}]^T\boldsymbol{\lambda} = 0$ (n conditions), and $\boldsymbol{c} = 0$ (m conditions) for $n + m$ unknowns $\boldsymbol{x}$ and $\boldsymbol{\lambda}$
- second-order necessary condition for a maximum d$\boldsymbol{x}^T[\boldsymbol{H}^*]d\boldsymbol{x} < 0$ for any *admissible* d$\boldsymbol{x}$ ($[\boldsymbol{H}^*]$ is not required to be negative semidefinite)
- sufficient conditions for a maximum $\boldsymbol{g}^T + \boldsymbol{\lambda}^T[\boldsymbol{G}] = 0$ and d$\boldsymbol{x}^T[\boldsymbol{H}^*]d\boldsymbol{x} < 0$ for any *admissible* d$\boldsymbol{x}$ ($[\boldsymbol{H}^*]$ is not required to be negative definite)

Equality Constrained Optimization (6)

- $\boldsymbol{g} + [\boldsymbol{G}]^T \boldsymbol{\lambda} = 0$ relates the gradient of ϕ to the gradients of the constraints
- with only one constraint ($m = 1$) $\frac{\partial \phi}{\partial \boldsymbol{x}} = -\lambda \frac{\partial c}{\partial \boldsymbol{x}}$, that is, the gradient of ϕ must be parallel to the gradient of c
- since the gradient of c is perpendicular the constraint $c = 0$, this means that the gradient of ϕ must be perpendicular to $c = 0$
- for more than one constraint ($m \geq 2$), the gradient of ϕ must be a linear combination of the constraints gradients

Equality Constrained Optimization - Example

- $\phi = -x_1^2 - x_2^2$, $\boldsymbol{c} = x_1 + x_2 - 1 = 0$ ($n = 2$, $m = 1$)
  $\phi^* = -x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1)$
  $\boldsymbol{g}^T = [-2x_1, -2x_2]$, $[\boldsymbol{G}] = (1, 1)$,
  $\boldsymbol{g}^T + \lambda^T[\boldsymbol{G}] = (-2x_1 + \lambda, -2x_2 + \lambda)$
  stationary point requires $-2x_1 + \lambda = 0$, $-2x_2 + \lambda = 0$,
  $x_1 + x_2 - 1 = 0$, that is $x_1 = x_2 = 1/2$ and $\lambda = 1$
- $[\boldsymbol{H}^*] = [\boldsymbol{H}] = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$, admissible variation is
  $(\mathrm{d}\boldsymbol{x})^T = (-a, a)$
  $\mathrm{d}\boldsymbol{x}^T[\boldsymbol{H}^*]\mathrm{d}\boldsymbol{x} = -4a^2 < 0$ therefore $\boldsymbol{x}^T = (1/2, 1/2)$ is a maximum (in this case, $[\boldsymbol{H}^*]$ is negative definite and $\mathrm{d}\boldsymbol{x}^T[\boldsymbol{H}^*]\mathrm{d}\boldsymbol{x} < 0$ for any $\mathrm{d}\boldsymbol{x}$)

Inequality Constrained Optimization (1)

- maximize $\phi(\boldsymbol{x})$ ($\boldsymbol{x}$ is a n-component vector of variables) with $\boldsymbol{c}(\boldsymbol{x}) \geq 0$ ($\boldsymbol{c}$ is a m-component vector of constraints, $\boldsymbol{c}^T = [c_1, c_2, ..., c_m]$, m may be larger than n)
- feasible points are those which satisfy $\boldsymbol{c}(\boldsymbol{x}) \geq 0$
- ϕ is maximum if d$\phi \leq 0$ for any *admissible* (i.e., that verifies $\boldsymbol{c}(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) \geq 0$) choice of d$\boldsymbol{x}$
- at a feasible point a constraint may be *inactive* ($c_j > 0$) and can be neglected or *active* ($c_j = 0$) and can be (at least, initially) treated as an equality constraint
- at a feasible point admissible variations require d$\boldsymbol{c}_a \geq 0$ ($\boldsymbol{c}_a$ only contains the active constraints, m_a components)

Inequality Constrained Optimization (2)

- main issue: determine the active set of constraints
- for a given active set, an equality constrained optimization problem must be solved
  - if the solution violates an inactive constraint, the constraint must be added to the active set and an augmented equality constrained problem must be solved
  - if the solution is feasible, the requirement of the active constraints must be checked and unnecessary constraints must be removed producing a reduced-dimension problem

Inequality Constrained Optimization (3)

- given the active set solve the equality constrained optimization
- maximize the augmented function $\phi^* = \phi(\boldsymbol{x}) + \boldsymbol{\lambda}_a^T \boldsymbol{c}_a$ (with $\boldsymbol{c}_a = 0$ active set of constraints and $\boldsymbol{\lambda}_a$ vector of adjoint parameters to be determined, m_a components)
- ϕ and ϕ^* coincide for any choice of $\boldsymbol{\lambda}_a$ if the active constraints are verified
- second-order Taylor's expansion of ϕ^* (local maxima are sought)
  $\phi^*(\boldsymbol{x} + \mathrm{d}\boldsymbol{x}) = \phi^*(\boldsymbol{x}) + (\boldsymbol{g}^T + \boldsymbol{\lambda}_a^T[\boldsymbol{G}_a])\mathrm{d}\boldsymbol{x} + \frac{1}{2}\mathrm{d}\boldsymbol{x}^T[\boldsymbol{H}^*]\mathrm{d}\boldsymbol{x}$
- augmented Hessian $[\boldsymbol{H}^*] = [\boldsymbol{H}] + \partial[\partial(\boldsymbol{\lambda}_a^T \boldsymbol{c}_a/\partial\boldsymbol{x})^T]\partial\boldsymbol{x}$, that is
  $H^*_{ij} = H_{ij} + \sum_{k=1}^{m} \lambda_{ak} \frac{\partial^2 c_{ak}}{\partial x_i \partial x_j}$

Inequality Constrained Optimization (4)

- for the equality constrained optimization given the active set
  - $\boldsymbol{\lambda}_a$ can be chosen to nullify the first variation of ϕ^* for *any* choice of d$\boldsymbol{x}$
  - first-order necessary condition $\boldsymbol{g}^T + \boldsymbol{\lambda}_a^T[\boldsymbol{G}_a] = 0$, that is, $\boldsymbol{g} + [\boldsymbol{G}_a]^T\boldsymbol{\lambda}_a = 0$ (n conditions) and $\boldsymbol{c}_a = 0$ (m_a conditions) for $n + m_a$ unknowns $\boldsymbol{x}$ and $\boldsymbol{\lambda}_a$
  - second-order necessary condition for a maximum d$\boldsymbol{x}^T[\boldsymbol{H}^*]d\boldsymbol{x} \leq 0$ for any *admissible* d$\boldsymbol{x}$ ($[\boldsymbol{H}^*]$ is not required to be negative semidefinite)
  - d$\boldsymbol{x}^T[\boldsymbol{H}^*]d\boldsymbol{x} < 0$ for any *admissible* d$\boldsymbol{x}$ is a sufficient condition

Inequality Constrained Optimization (5)

- at a feasible stationary point ($\mathrm{d}\phi^* = 0$) of the equality constrained problem ($\boldsymbol{c}_a = 0$), admissible variations of the inequality constrained problem require $\mathrm{d}\boldsymbol{c}_a \geq 0$
- $\mathrm{d}\phi = \mathrm{d}\phi^* - \boldsymbol{\lambda}_a^T \mathrm{d}\boldsymbol{c}_a$ and at a maximum ($\mathrm{d}\phi \leq 0$) $\boldsymbol{\lambda}_a^T \mathrm{d}\boldsymbol{c}_a \geq 0$ is required for any admissible $\mathrm{d}\boldsymbol{c}_a$
- this implies that all the components of $\boldsymbol{\lambda}_a$ must be positive
- the constraint $c_j = 0$ must be removed from the active set when $\lambda_{aj} < 0$
- generalizing, necessary conditions are $\boldsymbol{c} \geq 0$, $\boldsymbol{g} + [\boldsymbol{G}]^T \boldsymbol{\lambda} = 0$, $\lambda_j \geq 0$ when $c_j = 0$, $\lambda_j = 0$ when $c_j > 0$, $\mathrm{d}\boldsymbol{x}^T [\boldsymbol{H}^*] \mathrm{d}\boldsymbol{x} \leq 0$ for any admissible (with respect to the active constraints) $\mathrm{d}\boldsymbol{x}$
- if $\mathrm{d}\boldsymbol{x}^T [\boldsymbol{H}^*] \mathrm{d}\boldsymbol{x} < 0$ for any *admissible* $\mathrm{d}\boldsymbol{x}$ is enforced, a set of sufficient conditions is obtained

Inequality Constrained Optimization (6)

- $\boldsymbol{g} + [\boldsymbol{G}]^T \boldsymbol{\lambda} = 0$ relates the gradient of ϕ to the gradients of the constraints
- with only one constraint ($m = 1$) $\dfrac{\partial \phi}{\partial \boldsymbol{x}} = -\lambda \dfrac{\partial c}{\partial \boldsymbol{x}}$, that is, the gradient of ϕ must be parallel and *opposite* ($-\lambda \leq 0$) to the gradient of c
- since the gradient of c is perpendicular the constraint $c = 0$, this means that the gradient of ϕ must be perpendicular to $c = 0$ pointing toward constraint violation
- for more than one constraint ($m \geq 2$), the gradient of ϕ must be a *negative* linear combination of the constraints gradients
- - λ_i represents the change in ϕ for a unit change in c_i (assuming a linear approximation)

Inequality Constrained Optimization (7)

- a strict equality constraint $c_i = 0$ requires the simultaneous fulfillment of both $c_i \geq 0$ and $-c_i \geq 0$ and therefore, at a feasible point, an equality constraint is always active
- $\lambda_i > 0$ indicates that at a maximum (d$\phi < 0$) admissible dc_i must be positive, i.e., $c_i \geq 0$ is active (the index would be increased for $c_i < 0$)
- $\lambda_i < 0$ indicates that at a maximum (d$\phi < 0$) admissible dc_i must be negative, i.e., $c_i \leq 0$ is active (the index would be increased for $c_i > 0$)

Inequality Constrained Optimization

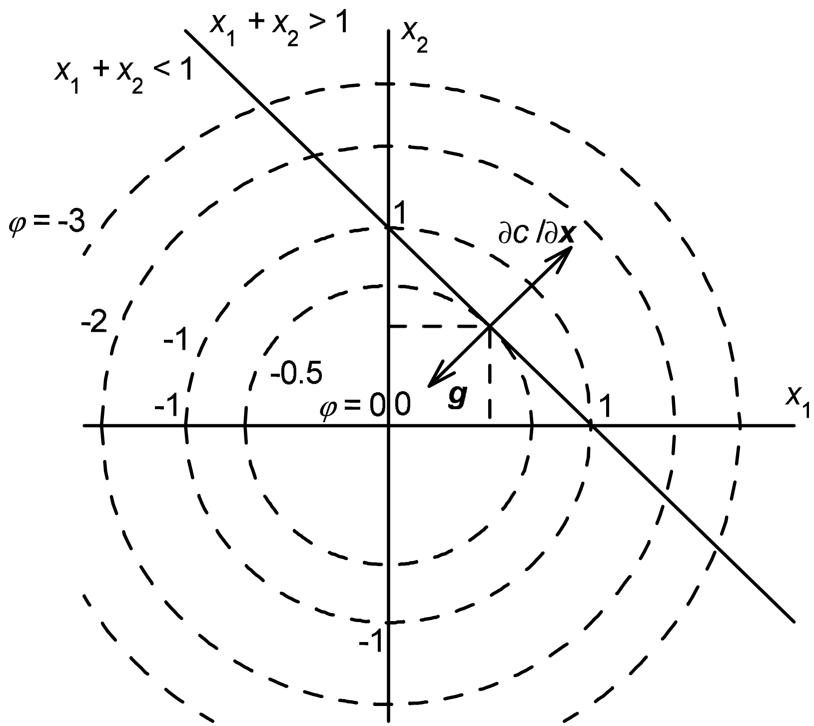
Examples (1)

- $\phi = -x_1^2 - x_2^2$, $\boldsymbol{c} = x_1 + x_2 - 1 \geq 0$
- assume c being active, define $\phi^* = -x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1)$
- stationary point requires $-2x_1 + \lambda = 0$, $-2x_2 + \lambda = 0$, $x_1 + x_2 - 1 = 0$, that is $x_1 = x_2 = 1/2$ and $\lambda = 1$
- $\boldsymbol{x}^T = (1/2, 1/2)$ is a maximum as $\lambda > 0$ and, for admissible variations $\mathrm{d}\boldsymbol{x}^T = (-a, a)$, $\mathrm{d}\boldsymbol{x}^T[\boldsymbol{H}]^*\mathrm{d}\boldsymbol{x} = -a^2 < 0$

Inequality Constrained Optimization Examples (2)

- $\phi = -x_1^2 - x_2^2$, $x_1 + x_2 - 1 \leq 0$, that is, $\boldsymbol{c} = -x_1 - x_2 + 1 \geq 0$
- assume c being active, define
  $\phi^* = -x_1^2 - x_2^2 + \lambda(-x_1 - x_2 + 1)$
- stationary point requires $-2x_1 - \lambda = 0$, $-2x_2 - \lambda = 0$,
  $x_1 + x_2 - 1 = 0$, that is $x_1 = x_2 = 1/2$ and $\lambda = -1$
- since $\lambda < 0$ the constraint is to be removed and the unconstrained optimization provides $\boldsymbol{x}^T = (0, 0)$ as the maximum

Example



$x_1 + x_2 > 1$
$x_1 + x_2 < 1$
x_2
$\varphi = -3$
-2
-1
-0.5
$\varphi = 0$
0
1
$\partial c / \partial \mathbf{x}$
$\mathbf{g}$
-1
1
x_1
-1

Continuous Systems

Summary

- necessary conditions for a stationary solution
- Hamilton-Jacobi-Bellman equation
- necessary and sufficient conditions for optimality

Continuous Systems

- t independent variable
- $\boldsymbol{x}$ state variables (n-component vector)
- $\boldsymbol{u}$ control variables (m-component vector)
- state equations $\dot{\boldsymbol{x}}(t) = \mathrm{d}\boldsymbol{x}/\mathrm{d}t = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t)$ (n-component vector of differential equations)
- boundary conditions $\boldsymbol{\psi}(\boldsymbol{x}_0, \boldsymbol{x}_f, t_0, t_f) = 0$ (q-component vector of algebraic equations, $q \leq n + 2$)

Bolza Problem

- determine the *extremal path* $\boldsymbol{x}(t)$ and the corresponding optimal control law $\boldsymbol{u}(t)$ satisfying
  - differential equations $\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t)$
  - boundary conditions $\boldsymbol{\psi}(\boldsymbol{x}_0, \boldsymbol{x}_f, t_0, t_f) = 0$

  to maximize (or minimize)

  $J = \phi(\boldsymbol{x}_0, \boldsymbol{x}_f, t_0, t_f) + \int_{t_0}^{t_f} \Phi(\boldsymbol{x}, \dot{\boldsymbol{x}}, t)\mathrm{d}t$
- Meyer's formulation $J = \phi(\boldsymbol{x}_0, \boldsymbol{x}_f, t_0, t_f)$ and $\Phi = 0$
- Lagrange's formulation $J = \int_{t_0}^{t_f} \Phi(\boldsymbol{x}, \dot{\boldsymbol{x}}, t)\mathrm{d}t$ and $\phi = 0$
- calculus of variations studies the behavior of functionals J
- optimal control theory applies CoV to the maximization or minimization of functionals

Equivalence of Formulations

- Lagrange's formulation can be reduced to Meyer's formulation by introducing the new state variable x' with $\dot{x}' = \Phi$, $x'_0 = 0$ and $\phi = x'_f$
- Meyer's formulation can be reduced to Lagrange's formulation if $\phi = \phi(\boldsymbol{x}_f, t_f)$ observing that $\phi(\boldsymbol{x}_f, t_f) = \int_{t_0}^{t_f} [(\partial\phi/\partial\boldsymbol{x}_f)\dot{\boldsymbol{x}} + \partial\phi/\partial t_f] \mathrm{d}t$ or by opportunely introducing additional variables in the general case
- minimization can be turned into maximization by changing the sign of ϕ and Φ
- the maximization problem is here considered

Adjoint Variables

- adjoint variables $\boldsymbol{\lambda}$ (n-component vector) and constants $\boldsymbol{\mu}$ (q-component vector)
- augmented index $J^* = \phi + \boldsymbol{\mu}^T \boldsymbol{\psi} + \int_{t_0}^{t_f} [\Phi + \boldsymbol{\lambda}^T (\boldsymbol{f} - \dot{\boldsymbol{x}})] \mathrm{d}t$
- $J^* = J$ for any choice of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ if state equations and boundary conditions are satisfied

Index Variation (1)

- using a first-order expansion of J^* one has

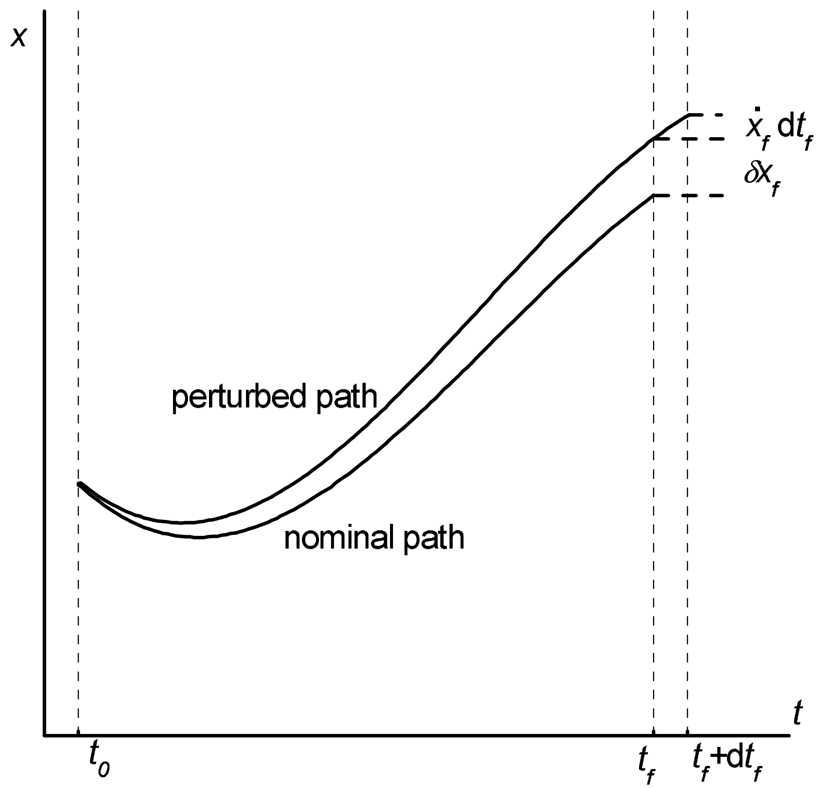
$$
\begin{aligned}
\mathrm{d}J^* \;=\; & \left(\frac{\partial\varphi}{\partial t_f} + \boldsymbol{\mu}^T \frac{\partial\boldsymbol{\psi}}{\partial t_f} + \Phi_f + \boldsymbol{\lambda}_f^T(\boldsymbol{f}_f - \dot{\boldsymbol{x}}_f)\right)\mathrm{d}t_f + \\
+ \; & \left(\frac{\partial\varphi}{\partial t_0} + \boldsymbol{\mu}^T \frac{\partial\boldsymbol{\psi}}{\partial t_0} - \Phi_0 - \boldsymbol{\lambda}_0^T(\boldsymbol{f}_0 - \dot{\boldsymbol{x}}_0)\right)\mathrm{d}t_0 + \\
+ \; & \left(\frac{\partial\varphi}{\partial \boldsymbol{x}_f} + \boldsymbol{\mu}^T \frac{\partial\boldsymbol{\psi}}{\partial \boldsymbol{x}_f}\right)\mathrm{d}\boldsymbol{x}_f + \left(\frac{\partial\varphi}{\partial \boldsymbol{x}_0} + \boldsymbol{\mu}^T \frac{\partial\boldsymbol{\psi}}{\partial \boldsymbol{x}_0}\right)\mathrm{d}\boldsymbol{x}_0 + \\
+ \; & \int_{t_0}^{t_f}\left[\frac{\partial(\Phi + \boldsymbol{\lambda}^T\boldsymbol{f})}{\partial\boldsymbol{x}}\delta\boldsymbol{x} + \frac{\partial(\Phi + \boldsymbol{\lambda}^T\boldsymbol{f})}{\partial\boldsymbol{u}}\delta\boldsymbol{u} - \boldsymbol{\lambda}^T\delta\dot{\boldsymbol{x}}\right]\mathrm{d}t
\end{aligned}
$$

Index Variation (2)

- at a given (constant) time the variable variation is $\delta \boldsymbol{x}$
- at the initial (or final) point the total variation $\mathrm{d}\boldsymbol{x} = \delta \boldsymbol{x} + \dot{\boldsymbol{x}}\mathrm{d}t$ is the sum of variation at constant t, i.e., $\delta \boldsymbol{x}$, and variation due to initial (or final) time variation $\dot{\boldsymbol{x}}\mathrm{d}t = \boldsymbol{f}\mathrm{d}t$
- the Hamiltonian is introduced $H = \Phi + \boldsymbol{\lambda}^T \boldsymbol{f}$
- term in $\delta \dot{\boldsymbol{x}}$ is integrated by parts

$$\int_{t_0}^{t_f} -\boldsymbol{\lambda}^T \delta \dot{\boldsymbol{x}} \mathrm{d}t = -\boldsymbol{\lambda}_f^T \delta \boldsymbol{x}_f + \boldsymbol{\lambda}_0^T \delta \boldsymbol{x}_0 + \int_{t_0}^{t_f} \dot{\boldsymbol{\lambda}}^T \delta \boldsymbol{x} \mathrm{d}t$$

Index Variation (3)



Index Variation (4)

$$
\begin{aligned}
\mathrm{d}J^* &= \left(\frac{\partial \varphi}{\partial t_f} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_f} + H_f\right) \mathrm{d}t_f + \\
&+ \left(\frac{\partial \varphi}{\partial t_0} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_0} - H_0\right) \mathrm{d}t_0 + \\
&+ \left(-\boldsymbol{\lambda}_f^T + \frac{\partial \varphi}{\partial \boldsymbol{x}_f} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_f}\right) \mathrm{d}\boldsymbol{x}_f + \\
&+ \left(\boldsymbol{\lambda}_0^T + \frac{\partial \varphi}{\partial \boldsymbol{x}_0} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_0}\right) \mathrm{d}\boldsymbol{x}_0 + \\
&+ \int_{t_0}^{t_f} \left[\left(\frac{\partial H}{\partial \boldsymbol{x}} + \dot{\boldsymbol{\lambda}}^T\right) \delta \boldsymbol{x} + \frac{\partial H}{\partial \boldsymbol{u}} \delta \boldsymbol{u}\right] \mathrm{d}t
\end{aligned}
$$

Stationary Point

- necessary condition for J being stationary is d$J = 0$ for any *admissible* choice of $\delta \boldsymbol{x}$, $\delta \boldsymbol{u}$, d$\boldsymbol{x}_f$, d$\boldsymbol{x}_0$, dt_f, dt_0 (that is, differential equations and boundary conditions must be satisfied)
- $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ can be *chosen* to nullify dJ^* for *any* choice of $\delta \boldsymbol{x}$, $\delta \boldsymbol{u}$, d$\boldsymbol{x}_f$, d$\boldsymbol{x}_0$, dt_f, dt_0 by nullifying their multiplying coefficients in dJ^*
- since J and J^* coincide when the constraints are satisfied, d$J^* = 0$ for any variation implies d$J = 0$ for any admissible variation

Equations for Adjoint and Control Variables

- Euler-Lagrange equations (n differential equations for the adjoint variables)
  $\frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t} = -\left(\frac{\partial H}{\partial \boldsymbol{x}}\right)^T$
- optimal control equations (m algebraic equations for the control variables)
  $\left(\frac{\partial H}{\partial \boldsymbol{u}}\right)^T = 0$
- the control equations do not formally depend on the performance index

Boundary Conditions for Optimality

- boundary conditions for optimality ($n + n$ algebraic equations at initial and final point)
$$\boldsymbol{\lambda}_0^T + \frac{\partial \varphi}{\partial \boldsymbol{x}_0} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_0} = 0 \qquad -\boldsymbol{\lambda}_f^T + \frac{\partial \varphi}{\partial \boldsymbol{x}_f} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_f} = 0$$
- transversality conditions ($1 + 1$ algebraic equations at initial and final time)
$$-H_0 + \frac{\partial \varphi}{\partial t_0} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_0} = 0 \qquad H_f + \frac{\partial \varphi}{\partial t_f} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_f} = 0$$
- with the imposed boundary conditions $\boldsymbol{\psi} = 0$ (q equations) one has $2n + q + 2$ equations which implicitly determine q adjoint constants ($\boldsymbol{\mu}$), 2 times (t_0 and t_f) and the initial values for $2n$ differential equations (for $\boldsymbol{x}$ and $\boldsymbol{\lambda}$)

Hamiltonian Time-Derivative

- $\frac{\mathrm{d}H}{\mathrm{dt}} = \frac{\partial H}{\partial t} + \frac{\partial H}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}} + \frac{\partial H}{\partial \boldsymbol{\lambda}}\dot{\boldsymbol{\lambda}} + \frac{\partial H}{\partial \boldsymbol{u}}\dot{\boldsymbol{u}}$
- one has $\left(\frac{\partial H}{\partial \boldsymbol{\lambda}}\right)^T = \boldsymbol{f} = \dot{\boldsymbol{x}}$, $\dot{\boldsymbol{\lambda}} = -\left(\frac{\partial H}{\partial \boldsymbol{x}}\right)^T$
- therefore $\frac{\mathrm{d}H}{\mathrm{dt}} = \frac{\partial H}{\partial t} + \frac{\partial H}{\partial \boldsymbol{u}}\dot{\boldsymbol{u}}$
- when H does not depend explicitly on time ($\frac{\partial H}{\partial t} = 0$) and the optimal control law is adopted ($\frac{\partial H}{\partial \boldsymbol{u}} = 0$), the Hamiltonian is constant

Simple Boundary Conditions on State Variables

- variable assigned at t_0, $(x_i)_0 - a_i = 0 \Rightarrow (\lambda_i)_0 + \mu = 0$, that is, the corresponding adjoint variable is free
- variable assigned at t_f, $(x_i)_f - b_i = 0 \Rightarrow -(\lambda_i)_f + \mu = 0$, that is, the corresponding adjoint variable is free
- one can consider $\mathrm{d}(x_i)_0 = 0$ (or $\mathrm{d}(x_i)_f = 0$) and drop the equation obtained by nullifying its coefficient
- variable free at t_0 and $\partial\phi/\partial(x_i)_0 = 0 \Rightarrow (\lambda_i)_0 = 0$, that is, the corresponding adjoint variable is null
- variable free at t_f and $\partial\phi/\partial(x_i)_f = 0 \Rightarrow (\lambda_i)_f = 0$, that is, the corresponding adjoint variable is null

Simple Boundary Conditions on Time

- initial time assigned $t_0 - t_a = 0 \Rightarrow -H_0 + \mu = 0$, that is, the corresponding Hamiltonian is free
- final time assigned $t_f - t_b = 0 \Rightarrow H_f + \mu = 0$, that is, the corresponding Hamiltonian is free
- one can consider $\mathrm{d}t_0 = 0$ (or $\mathrm{d}t_f = 0$) and drop the equation obtained by nullifying its coefficient
- initial time free and $\partial\phi/\partial t_0 = 0 \Rightarrow H_0 = 0$, that is, the corresponding Hamiltonian is null
- final time free and $\partial\phi/\partial t_f = 0 \Rightarrow H_f = 0$, that is, the corresponding Hamiltonian is null

Initial State Specified - No Terminal Constraints Fixed Terminal Time (1)

- $\boldsymbol{x}_0 = \boldsymbol{a}$, $t_0 = t_a$, $t_f = t_b$
- ϕ can be regarded as a function of $\boldsymbol{x}_f$ only, i.e., $\phi = \phi(\boldsymbol{x}_f)$
- H_0, H_f and $\boldsymbol{\lambda}_0$ are free as $\delta\boldsymbol{x}_0 = 0$, $\delta t_0 = 0$ and $\delta t_f = 0$
- optimality conditions provide $\boldsymbol{\lambda}_f = \left(\frac{\partial \phi}{\partial \boldsymbol{x}_f}\right)^T$ producing a two-point boundary value problem
- initial values $\boldsymbol{\lambda}_0$ are free and must be determined to satisfy the optimality conditions at the final point

Initial State Specified - No Terminal Constraints
Fixed Terminal Time (2)

- for Euler-Lagrange equations and optimality conditions satisfied (i.e., on an extremal path) non-admissible variations provide $\mathrm{d}J = \boldsymbol{\lambda}_0^T \mathrm{d}\boldsymbol{x}_0 - H\mathrm{d}t_0 + \int_{t_0}^{t_f} \frac{\partial H}{\partial \boldsymbol{u}} \delta \boldsymbol{u} \mathrm{d}t$
- $\boldsymbol{\lambda}_0$ is the gradient of J with respect to changes in $\boldsymbol{x}_0$ (influence functions) $\boldsymbol{\lambda}_0 = (\partial J / \partial \boldsymbol{x}_0)^T$
- $-H_0$ is the gradient of J with respect to changes in t_0 (time influence function) $-H_0 = \partial J / \partial t_0$
- $\frac{\partial H}{\partial \boldsymbol{u}}$ represents the change in J for a unit impulse $\delta \boldsymbol{u}$ (impulse response functions, must be null on an extremal path)

State Variables Specified at Fixed Final Time

State Variables Unspecified at Fixed Initial Time

- $(x_i)_f = b_i$ replaces $(\lambda_i)_f = \dfrac{\partial\phi}{\partial(x_i)_f}$
- similarly, if $(x_i)_0$ is free, $(\lambda_i)_0 = 0$ replaces $(x_i)_0 = a_i$ (ϕ is assumed to be a function of $\boldsymbol{x}_f$ only)
- the change in J with respect to a unit change in $(x_i)_0$, i.e., $(\lambda_i)_0$, must be null at the optimal initial point if $(x_i)_0$ is free
- if $\partial\phi/\partial(x_i)_0 \neq 0$ then $\mathrm{d}J/\mathrm{d}(x_i)_0 = \partial\phi/\partial(x_i)_0 + (\lambda_i)_0 = 0$ must be enforced

Unspecified Initial and Final Time

- $t_0 = t_a$ and $t_0 = t_b$ are dropped from $\boldsymbol{\psi}$
- the transversality conditions provide the initial and final time

$$-H_0 + \frac{\partial \varphi}{\partial t_0} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_0} = 0$$

$$H_f + \frac{\partial \varphi}{\partial t_f} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_f} = 0$$

Equality Path Constraints

- $\int_{t_0}^{t_f} N(\boldsymbol{x}, \boldsymbol{u}, t)\mathrm{d}t = k_b$
- introduce a new variable x_{n+1} with differential equation $\dot{x}_{n+1} = N$ and boundary conditions $(x_{n+1})_0 = 0$, $(x_{n+1})_f = k_b$

Equality Constraints on Control Variables

- $C(\boldsymbol{u}, t) = 0$ (only for $m \geq 2$)
- two options
  - eliminate one control variable u_j from $C(\boldsymbol{u}, t) = 0$ and consider the unconstrained problem with m reduced by one
  - use augmented Hamiltonian $H = \Phi + \boldsymbol{\lambda}^T \boldsymbol{f} + \lambda' C$ and derive E-L equations, controls and boundary conditions consequently

Equality Constraints on State and Control Variables

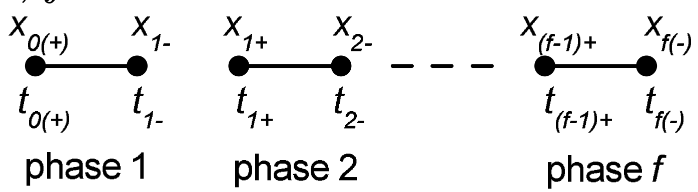
- $C(\boldsymbol{x}, \boldsymbol{u}, t) = 0$
- same approach as before, two options
  - eliminate one control variable u_j from $C(\boldsymbol{x}, \boldsymbol{u}, t) = 0$ and consider the unconstrained problem with m reduced by one
  - use augmented Hamiltonian $H = \Phi + \boldsymbol{\lambda}^T \boldsymbol{f} + \lambda' C$ and derive E-L equations, controls and boundary conditions consequently
- one of the state variables can alternatively be eliminated

Equality Constraints on Function of State Variables

- $S(\boldsymbol{x}, t) = 0$
- since S must be constant $\dot{S} = \frac{\partial S}{\partial \boldsymbol{x}} \dot{\boldsymbol{x}} + \frac{\partial S}{\partial t} = 0$
- if $\dot{S}$ depends on $\boldsymbol{u}$ eliminate one control variable u_j from $\dot{S} = 0$ and consider the unconstrained problem with m reduced by one
- if $\dot{S}$ is independent of $\boldsymbol{u}$ either
  - eliminate one state variable x_j from $\dot{S} = 0$ and consider the unconstrained problem with n reduced by one
  - compute subsequent time derivatives until $\boldsymbol{u}$ appears

Interior Point Constraints (1)

- functions of state variables specified or variables discontinuities at interior point (time may or may not be specified)
- split the integration into f intervals at the relevant $f-1$ intermediate points
- the j-th interval spans from $t_{(j-1)+}$ to $t_{(j)-}$, the variable values at the extremities are $\boldsymbol{x}_{(j-1)+}$ and $\boldsymbol{x}_{(j)-}$, respectively, $j = 1, f$



Interior Point Constraints (2)

- define the performance index J

$$J = \phi(\boldsymbol{x}_0, \boldsymbol{x}_{1\pm}, ..., \boldsymbol{x}_f, t_0, t_{1\pm}, ..., t_f) + \sum_{j=i}^{f} \int_{t_{(j-1)+}}^{t_{(j)-}} \Phi(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) \mathrm{d}t$$

- add conditions at intermediate points and build $\boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{x}_{(j-1)+}, \boldsymbol{x}_{(j)-}, t_{(j-1)+}, t_{(j)-})$, $j = 1, ...f$
- define the augmented performance index J^*

$$J^* = \phi + \boldsymbol{\mu}^T \boldsymbol{\psi} + \sum_{j=i}^{f} \int_{t_{(j-1)+}}^{t_{(j)-}} \Phi + \boldsymbol{\lambda}^T (\boldsymbol{f} - \dot{\boldsymbol{x}}) \mathrm{d}t$$

and derive Euler-Lagrange and control equations, and boundary conditions for optimality

Interior Point Constraints (3)

- boundary conditions for optimality ($j = 1, f$)

  $\boldsymbol{\lambda}^T_{(j-1)+} + \dfrac{\partial \varphi}{\partial \boldsymbol{x}_{(j-1)+}} + \boldsymbol{\mu}^T \dfrac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_{(j-1)+}} = 0$ (start of phase j)

  $-\boldsymbol{\lambda}^T_{(j)-} + \dfrac{\partial \varphi}{\partial \boldsymbol{x}_{(j)-}} + \boldsymbol{\mu}^T \dfrac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_{(j)-}} = 0$ (end of phase j)

- transversality conditions ($j = 1, f$)

  $-H_{(j-1)+} + \dfrac{\partial \varphi}{\partial t_{(j-1)+}} + \boldsymbol{\mu}^T \dfrac{\partial \boldsymbol{\psi}}{\partial t_{(j-1)+}} = 0$ (start of phase j)

  $H_{(j)-} + \dfrac{\partial \varphi}{\partial t_{(j)-}} + \boldsymbol{\mu}^T \dfrac{\partial \boldsymbol{\psi}}{\partial t_{(j)-}} = 0$ (end of phase j)

- a multipoint boundary value problem must be solved

Interior Point Constraints (4)

- at a generic point j (end of phase j, start of phase $j+1$)

$$\boldsymbol{\lambda}_{j+}^T + \frac{\partial \varphi}{\partial \boldsymbol{x}_{j+}} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_{j+}} = 0,\ j = 0, ..., f-1$$

$$-\boldsymbol{\lambda}_{j-}^T + \frac{\partial \varphi}{\partial \boldsymbol{x}_{j-}} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}_{j-}} = 0,\ j = 1, ..., f$$

$$-H_{j+} + \frac{\partial \varphi}{\partial t_{j+}} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_{j+}} = 0,\ j = 0, ..., f-1$$

$$H_{j-} + \frac{\partial \varphi}{\partial t_{j-}} + \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\psi}}{\partial t_{j-}} = 0,\ j = 1, ..., f$$

Interior Point Constraints (5)

- variable continuous and specified at intermediate time t_j, $(x_i)_{j+} - c_i = 0$, $(x_i)_{j-} - c_i = 0$, ϕ independent of $(x_i)_{j+}$ and $(x_i)_{j-} \Rightarrow (\lambda_i)_{j+} - (\lambda_i)_{j-} = -(\mu_1 + \mu_2)$, the corresponding adjoint variable has a free discontinuity
- variable continuous but unspecified at intermediate time t_j, $(x_i)_{j+} - (x_i)_{j-} = 0$, ϕ independent of $(x_i)_{j+}$ and $(x_i)_{j-} \Rightarrow$ $(\lambda_i)_{j+} + \mu = 0$ and $-(\lambda_i)_{j-} - \mu = 0$, the corresponding adjoint variable is continuous $(\lambda_i)_{j+} = (\lambda_i)_{j-}$

Interior Point Constraints (6)

- intermediate time fixed (and continuous), $t_{j+} - t_c = 0$, $t_{j-} - t_c = 0$, ϕ independent of t_{j+} and t_{j-} $\Rightarrow -H_{j+} + \mu_1 = 0$, $H_{j-} + \mu_2 = 0$, the Hamiltonian has a free discontinuity
- intermediate time free and continuous, $t_{j+} - t_{j-} = 0$, ϕ independent of t_{j+} and t_{j-} $\Rightarrow -H_{j+} + \mu = 0$, $H_{j-} - \mu = 0$, the Hamiltonian is continuous $H_{j+} = H_{j-}$

Inequality Constraints

- $C(\boldsymbol{x}, \boldsymbol{u}, t) \geq 0$
- assume the constraint being active $C(\boldsymbol{x}, \boldsymbol{u}, t) = 0$
- define augmented Hamiltonian $H = \Phi + \boldsymbol{\lambda}^T \boldsymbol{f} + \lambda' C$ and solve the equality constrained problem
  - $\lambda' \geq 0$ is required when $C = 0$ (active constraint)
  - $\lambda' = 0$ is required when $C > 0$ (inactive constraint)
- constraint to be removed if $\lambda' < 0$
- main issue: determine the constrained interval(s)

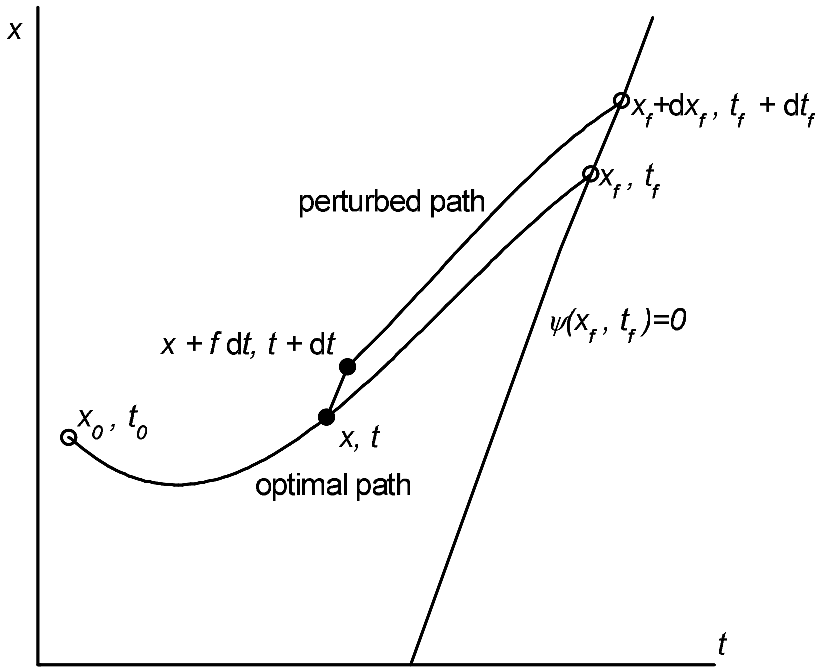
Hamilton-Jacobi-Bellman Equations (1)

- for a given initial point $\boldsymbol{x}_0$, t_0 and final constraints $\boldsymbol{\psi}(\boldsymbol{x}_f, t_f) = 0$ the optimal path provides the maximum index $J = J^o$ with control variables $\boldsymbol{u}^o(t)$ (unicity of the optimal path is here supposed)
- Hamilton-Jacobi theory provides differential equations concerning J^o and the optimal control $\boldsymbol{u}^o$, extended by Bellman to discrete systems (dynamic programming)
- any point x, t on the optimal path may be considered as the initial point
- starting from x, t on the optimal path the maximization of J provides the same control law $\boldsymbol{u}^o(t)$

Hamilton-Jacobi-Bellman Equations (2)

- for a generic starting point $\boldsymbol{x}$, t, one has $J^o(\boldsymbol{x},t) = \max\left\{\phi(\boldsymbol{x}_f,t_f) + \int_t^{t_f} \Phi(\boldsymbol{x},\boldsymbol{u},\tau)\mathrm{d}\tau\right\}$ with boundary condition $J^o(\boldsymbol{x}_f,t_f) = \phi(\boldsymbol{x}_f,t_f)$ on the terminal hypersurface $\boldsymbol{\psi}(\boldsymbol{x}_f,t_f) = 0$
- maximization with respect to $\boldsymbol{u}$
- assume J^o continuous with continuous first and second derivatives
- if a generic (possibly nonoptimal) control is used from $\boldsymbol{x}, t$ to $\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},t)\Delta t, t + \Delta t$, and then the optimal control is used up to the final point, the performance index will be $J'(\boldsymbol{x},t) = \Phi(\boldsymbol{x},\boldsymbol{u},t)\Delta t + J^o[\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},t)\Delta t, t + \Delta t]$

Hamilton-Jacobi-Bellman Equations (3)



Hamilton-Jacobi-Bellman Equations (4)

- $J' = J^o$ only if the optimal control is used from t to $t + \Delta t$, that is,
  $J^o(\boldsymbol{x}, t) = \max \{J^o[\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t)\Delta t, t + \Delta t] + \Phi(\boldsymbol{x}, \boldsymbol{u}, t)\Delta t\}$
- first order expansion provides $J^o(\boldsymbol{x}, t) =$
  $\max \left\{ J^o(\boldsymbol{x}, t) + \frac{\partial J^o}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t)\Delta t + \frac{\partial J^o}{\partial t}\Delta t + \Phi(\boldsymbol{x}, \boldsymbol{u}, t)\Delta t \right\}$
- for $\Delta t \to 0$ and considering that J^o and $\partial J^o / \partial t$ do not depend on u one has
  $-\frac{\partial J^o}{\partial t} = \max \left\{ \frac{\partial J^o}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t) + \Phi(\boldsymbol{x}, \boldsymbol{u}, t) \right\}$

Hamilton-Jacobi-Bellman Equations (5)

- according to the definition of influence functions
  $\mathrm{d}J^o = \boldsymbol{\lambda}^T \delta \boldsymbol{x} - H \delta t$
- therefore $\dfrac{\partial J^o}{\partial \boldsymbol{x}} = \boldsymbol{\lambda}^T \quad \dfrac{\partial J^o}{\partial t} = -H$ and
  $\Phi + \dfrac{\partial J^o}{\partial \boldsymbol{x}} \boldsymbol{f} = \Phi + \boldsymbol{\lambda}^T \boldsymbol{f} = H$
- comparison provides the Hamilton-Jacobi-Bellman equation
  $-\dfrac{\partial J^o}{\partial t} = H^o \left(\boldsymbol{x}, \dfrac{\partial J^o}{\partial \boldsymbol{x}}, t \right)$
  $H^o \left(\boldsymbol{x}, \dfrac{\partial J^o}{\partial \boldsymbol{x}}, t \right) = \max \left\{ H \left(\boldsymbol{x}, \dfrac{\partial J^o}{\partial \boldsymbol{x}}, \boldsymbol{u}, t \right) \right\} = \max \{ H (\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{u}, t) \}$
- maximization with respect to $\boldsymbol{u}$ given $\boldsymbol{x}$, $\boldsymbol{\lambda}$, and t

Hamilton-Jacobi-Bellman Equations (6)

- HJB equation state that the optimal controls must maximize the Hamiltonian over the whole set of admissible controls (global condition)
- the same results was obtained independently by Pontryagin and is known as Pontryagin's maximum principle (PMP)
- in the absence of bounds on state and control variables, HJB implies $H_u = 0$ and H_{uu} negative definite (local conditions)
- HJB also implies that Euler-Lagrange equations and the boundary conditions for optimality must be verified

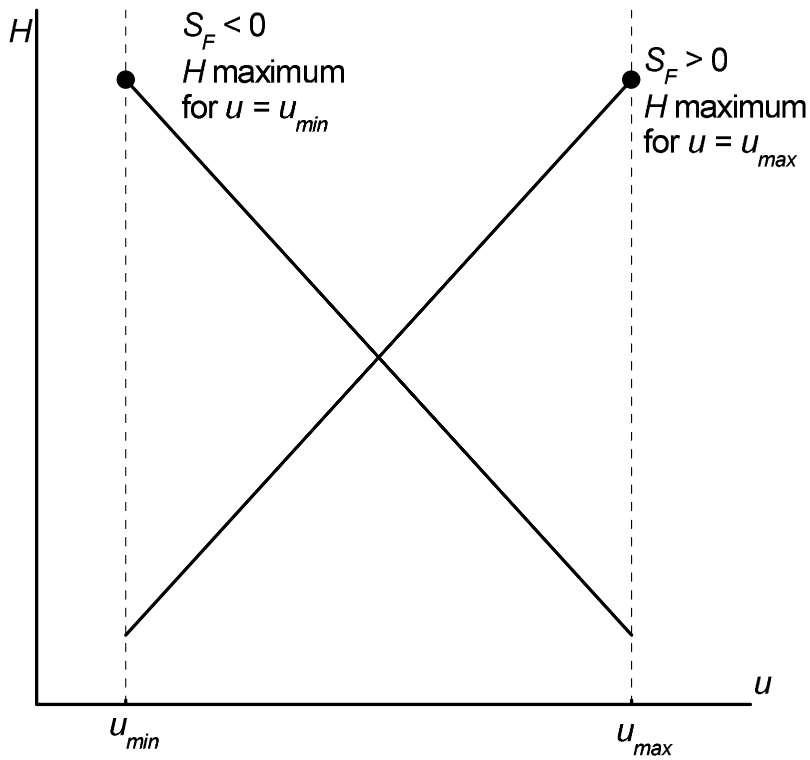
Hamilton-Jacobi-Bellman Equations (7)

- for any final point $\boldsymbol{x}_f$, t_f satisfying $\boldsymbol{\psi}(\boldsymbol{x}_f, t_f) = 0$ ($n+1$ variables, $q \leq n+1$ conditions), backward integration of HJB equation for any choice of $\boldsymbol{\lambda}_f$, $\boldsymbol{\mu}$ satisfying the boundary conditions for optimality and transversality condition at the final point ($n+q$ variables, $n+1$ conditions) provides all the possible optimal paths and the corresponding control laws (*field of extremals*)
- in general, only one optimal path will pass through a given point $\boldsymbol{x}, t$ and a unique optimal control $\boldsymbol{u}^o(\boldsymbol{x}, t)$ will be associated with each point (*optimal feedback control law*)
- main issue: *curse of dimensionality* - there are n free parameters and it is very difficult to find and store all the possible solutions

Bang-Bang Controls (1)

- if H is linear with respect to a control variable u_j then $\partial H/\partial u_j = 0$ does not contain u_j (indeterminate)
- HJB equation or PMP show that the control value which maximizes H must be adopted
- the problem has sense only if u_j is bounded $(u_j)_{min} \leq u_j \leq (u_j)_{max}$
- the switching function $S_F = \partial H/\partial u_j$ is introduced
  - $u_j = (u_j)_{max}$ when $S_F > 0$
  - $u_j = (u_j)_{min}$ when $S_F < 0$
  - singular arc when $S_F = 0$ over a finite interval (u_j obtained by nullifying the subsequent time derivatives of S_F)

Bang-Bang Controls (2)



Necessary Conditions for a Maximum (1)

- stationary feasible point: state and Euler-Lagrange equations verified for $t_0 \leq t \leq t_f$, imposed conditions and necessary conditions for optimality verified at initial, final, and intermediate points
- Weierstrass condition: H maximized by the optimal controls (global maximum) for $t_0 \leq t \leq t_f$
- for a local maximum Weierstrass condition is replaced by $\partial H/\partial \boldsymbol{u} = 0$ and Legendre-Clebsch condition H_{uu} negative semidefinite (usually, a minimization problem is considered and H_{uu} positive semidefinite is imposed, convexity condition)

Necessary Conditions for a Maximum (2)

- normality condition: $\widehat{Q}$ positive semidefinite, $\alpha \leq 0$ for $t_0 \leq t < t_f$
- Jacobi condition: $\widehat{S} - \widehat{R}\widehat{Q}^{-1}\widehat{R}^T$ finite for $t_0 < t < t_f$, that is, no conjugate points exist on the path (if the matrix becomes infinity some $\delta \boldsymbol{x}$ must be restricted and no maximization is possible as infinite equivalent solutions exist)
- Riccati equation $\dot{S} = -C - A^T S - SA + SBS =$
  $= -S\boldsymbol{f}_x - \boldsymbol{f}_x^T S - H_{xx} + (S\boldsymbol{f}_u + H_{ux}^T)H_{uu}^{-1}(H_{ux} + \boldsymbol{f}_u^T S)$
  $\dot{R} = -(A^T - SB)R$ and $\dot{Q} = R^T BR$
  $\dot{\boldsymbol{m}} = -(A^T - SB)\boldsymbol{m} \quad \dot{\boldsymbol{n}} = R^T B\boldsymbol{n} \quad \dot{\alpha} = \boldsymbol{m}^T B\boldsymbol{m}$
- with $A = \boldsymbol{f}_x - \boldsymbol{f}_u H_{uu}^{-1} H_{ux}$, $B = \boldsymbol{f}_u H_{uu}^{-1} \boldsymbol{f}_u^T$,
  $C = H_{xx} - H_{xu} H_{uu}^{-1} H_{ux}$

Sufficient Conditions for a Maximum (1)

- stationary feasible point: state and Euler-Lagrange equations verified for $t_0 \leq t \leq t_f$, imposed and necessary conditions for optimality verified at initial final and intermediate points
- Weierstrass condition: H maximized by the optimal controls (global maximum) for $t_0 \leq t \leq t_f$
- for a local maximum Weierstrass condition is replaced by $\partial H / \partial \boldsymbol{u} = 0$ and strengthened Legendre-Clebsch condition H_{uu} negative definite

Sufficient Conditions for a Maximum (2)

- normality condition: $\widehat{Q}$ positive definite, $\alpha < 0$ for $t_0 \le t < t_f$
- Jacobi condition: $\widehat{S} - \widehat{R}\widehat{Q}^{-1}\widehat{R}^T$ finite for $t_0 \le t < t_f$, that is, no conjugate points exist on the path (if the matrix becomes infinity some $\delta \boldsymbol{x}$ must be restricted and no maximization is possible as infinite equivalent solutions exist)

Practical Approach

- rely on intuition
- check results with perturbed conditions for optimality

Optimization Problem Formulation

- trajectory split into arcs
- homogeneous control law in each arc
- constraints and discontinuities at the arcs boundaries

Boundary Value Problem

- application of the theory of optimal control produces a multipoint boundary value problem (BVP)
- some constants (e.g., relevant times) and the initial values of some of the state and adjoint variables are unknown
- boundary conditions at the relevant points must be satisfied
- an iterative procedure based on Newton's method can be adopted

BVP Formulation (1)

- t independent variable
- $\boldsymbol{x}$ state variables $\mathrm{d}\boldsymbol{x}/\mathrm{d}t = \boldsymbol{f}$
- $\boldsymbol{\lambda}$ adjoint variables $\mathrm{d}\boldsymbol{\lambda}/\mathrm{d}t = -\left(\partial H/\partial \boldsymbol{x}\right)^T$
- $\boldsymbol{u}$ control variables obtained as functions of $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ by maximizing the Hamiltonian
- boundary conditions concerning $\boldsymbol{x}$, $\boldsymbol{\lambda}$ and t at initial, final and intermediate points

BVP Formulation (2)

- independent variable transformation: in the j-th arc
  $\varepsilon = j - 1 + \frac{t - t_{j-1}}{t_j - t_{j-1}} = j - 1 + \frac{t - t_{j-1}}{\tau_j}$
- in the j-th arc ε varies between $j - 1$ and j
- $\mathrm{d}\boldsymbol{x}/\mathrm{d}\epsilon = \tau_j \mathrm{d}\boldsymbol{x}/\mathrm{d}t$
- $\mathrm{d}\boldsymbol{\lambda}/\mathrm{d}\epsilon = \tau_j \mathrm{d}\boldsymbol{\lambda}/\mathrm{d}t$
- the relevant times or arc time-lengths are additional constant parameters
- constant parameters $\boldsymbol{y}$ with equation $\mathrm{d}\boldsymbol{y}/\mathrm{d}\epsilon = 0$

BVP Formulation (3)

- define vector of unknowns
  $\boldsymbol{z} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \\ \boldsymbol{y} \end{pmatrix}$
- differential equations $\partial \boldsymbol{z}/\partial \epsilon = \boldsymbol{g}(\boldsymbol{z}, \epsilon)$
- collect values at relevant boundaries
  $\boldsymbol{s} = (\boldsymbol{z}_{0_+}, \boldsymbol{z}_{1_\pm}, \ldots, \boldsymbol{z}_{(f-1)_\pm}, \boldsymbol{z}_{f_-})$
- boundary conditions in the form $\boldsymbol{\Psi}(\boldsymbol{s}) = 0$
- assume initial tentative values $\boldsymbol{z}_0 = \boldsymbol{p}$
- Newton's method to reduce the error on boundary conditions

BVP Solution (1)

- at the r-th iteration correct the initial values $\boldsymbol{p}^{r+1} = \boldsymbol{p}^r + \Delta\boldsymbol{p}$ with $\Delta\boldsymbol{p} = \boldsymbol{p}^{r+1} - \boldsymbol{p}^r = -\left[\partial\boldsymbol{\Psi}/\partial\boldsymbol{p}\right]^{-1}\boldsymbol{\Psi}^r$
- $[\partial\boldsymbol{\Psi}/\partial\boldsymbol{p}] = [\partial\boldsymbol{\Psi}/\partial\boldsymbol{s}]\,[\partial\boldsymbol{s}/\partial\boldsymbol{p}]$
- $[\partial\boldsymbol{\Psi}/\partial\boldsymbol{s}]$ by derivation
- $[\partial\boldsymbol{s}/\partial\boldsymbol{p}]$ collects values at the boundaries of $[\partial\boldsymbol{z}/\partial\boldsymbol{p}]$
- matrix $[\partial\boldsymbol{z}/\partial\boldsymbol{p}]$ by integration of the homogeneous differential system

$$\left[\dot{\frac{\partial\boldsymbol{z}}{\partial\boldsymbol{p}}}\right] = \left[\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{z}}\right]\left[\frac{\partial\boldsymbol{z}}{\partial\boldsymbol{p}}\right]$$

BVP Solution (2)

- the matrix $[\partial\boldsymbol{\Psi}/\partial\boldsymbol{p}]$ can also be obtained numerically
- vary p_i by small quantity δp_i
- integrate equations and compute change of variable values at relevant points $\delta\boldsymbol{s}$
- compute change in error on boundary conditions $\delta\boldsymbol{\Psi}$
- approximate by linearization the i-th column of $[\partial\boldsymbol{\Psi}/\partial\boldsymbol{p}]$
  $[\partial\boldsymbol{\Psi}/\partial p_i] = \delta\boldsymbol{\Psi}/\delta p_i$

BVP Solution (3)

- linearization may introduce errors that can prevent convergence
  - perform check on error variation and reduce the parameter correction by posing $\boldsymbol{p}^{r+1} = \boldsymbol{p}^r + \Delta\boldsymbol{p}/2$ when $\boldsymbol{\Psi}^{r+1} > K_1 \boldsymbol{\Psi}^r$
  - use a reduced parameter correction $\Delta\boldsymbol{p} = -K_2 \left[\partial\boldsymbol{\Psi}/\partial\boldsymbol{p}\right]^{-1} \boldsymbol{\Psi}^r$ when the errors are large
  - $K_1 = 2$ usually provides good results
  - $0.01 \leq K_2 \leq 1$ usually provides good results (the larger values when the solution is close to the optimal one)

State Equations

- two-body problem with propulsion - vectorial formulation

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \boldsymbol{V}$$

$$\frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = \boldsymbol{g} + \frac{\boldsymbol{T}}{m}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{T}{c}$$

- Hamiltonian

$$H = \boldsymbol{\lambda}_r^T \boldsymbol{V} + \boldsymbol{\lambda}_V^T (\boldsymbol{g} + \boldsymbol{T}/m) - \lambda_m T/c$$

Euler-Lagrange Equations

- for thrust independent of state variables

$$\frac{\mathrm{d}\boldsymbol{\lambda}_r}{\mathrm{d}t} = -\left[\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{r}}\right]^T \boldsymbol{\lambda}_V$$

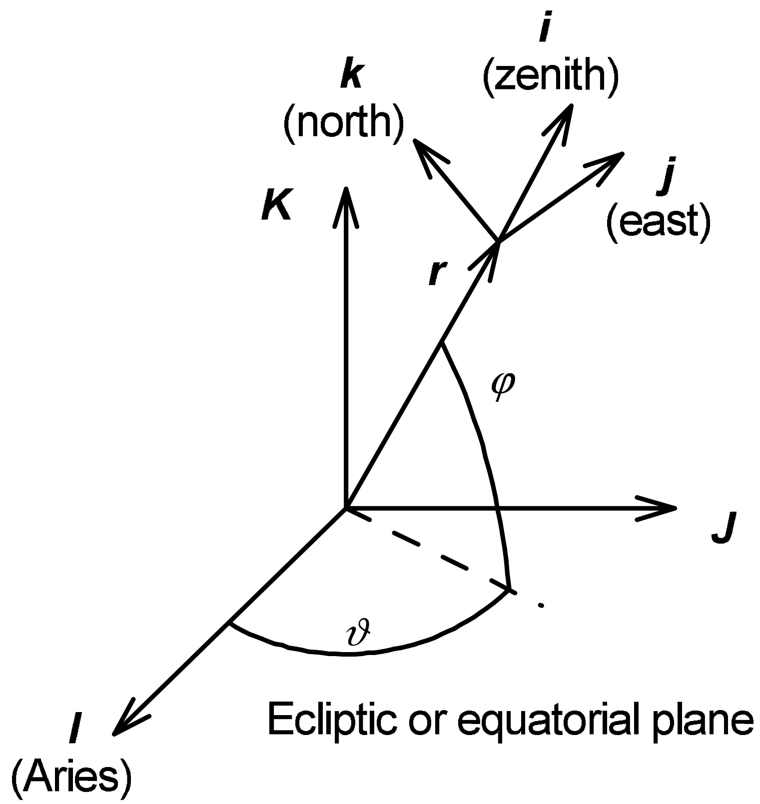
$$\frac{\mathrm{d}\boldsymbol{\lambda}_V}{\mathrm{d}t} = -\boldsymbol{\lambda}_r$$

$$\frac{\mathrm{d}\lambda_m}{\mathrm{d}t} = \frac{\boldsymbol{\lambda}_V \boldsymbol{T}}{m^2}$$

Optimal Thrust

- PMP states that the optimal controls must maximize H (given state and adjoint variables)
- thrust direction parallel to the velocity adjoint vector $\boldsymbol{\lambda}_V$ (also named *primer* vector), i.e., $\boldsymbol{T} = T\boldsymbol{\lambda}_V/\lambda_V$
- the Hamiltonian becomes $H = \boldsymbol{\lambda}_r^T \boldsymbol{V} + \boldsymbol{\lambda}_V^T \boldsymbol{g} + T(\lambda_V/m - \lambda_m/c)$
- maximum thrust T_{max} when the switching function $S_F = \lambda_V/m - \lambda_m/c$ is positive
- minimum thrust (typically, 0) when the switching function $S_F = \lambda_V/m - \lambda_m/c$ is negative
- $S_F = 0$ at the thrust switching points

Spherical Reference Frame



Scalar State Equations

- two-body problem with propulsion

$$\dot{r} = u \quad \dot{\vartheta} = v/(r\cos\phi) \quad \dot{\phi} = w/r$$

$$\dot{u} = -\mu/r^2 + (v^2 + w^2)/r + T\sin\gamma_T/m$$

$$\dot{v} = (-uv + vw\tan\phi)/r + T\cos\gamma_T\cos\psi_T/m$$

$$\dot{w} = (-uw - v^2\tan\phi)/r + T\cos\gamma_T\sin\psi_T/m$$

$$\dot{m} = -T/c$$

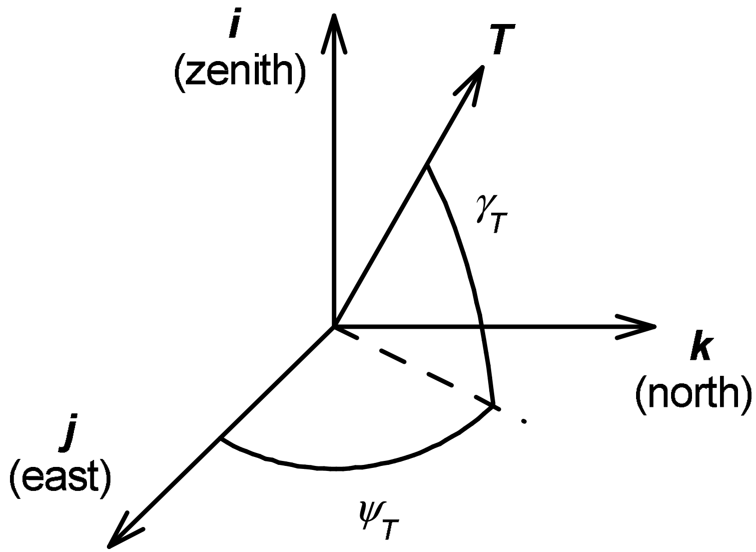
- $H = \lambda_r u + \lambda_\vartheta v/(r\cos\phi) + \lambda_\phi w/r$
  $+\lambda_u[-\mu/r^2 + (v^2 + w^2)/r + T\sin\gamma_T/m]$
  $+\lambda_v[(-uv + vw\tan\phi)/r + T\cos\gamma_T\cos\psi_T/m]$
  $+\lambda_w[(-uw - v^2\tan\phi)/r + T\cos\gamma_T\sin\psi_T/m] - \lambda_m T/c$

Scalar Euler-Lagrange Equations

- thrust independent of state variables (optimal thrust direction is assumed)

$$
\begin{aligned}
\dot{\lambda}_r &= \left[\lambda_\vartheta \frac{v}{\cos\phi} + \lambda_\phi w + \lambda_u \left(-\frac{2}{r} + v^2 + w^2\right) + \right. \\
&\quad \left. + \lambda_v \left(-uv + vw \tan\phi\right) + \lambda_w \left(-uw - v^2 \tan\phi\right)\right] / r^2 \\
\dot{\lambda}_\vartheta &= 0 \\
\dot{\lambda}_\phi &= \left(-\lambda_\vartheta v \sin\phi - \lambda_v v w + \lambda_w v^2\right) / (r \cos^2\phi) \\
\dot{\lambda}_u &= \left(-\lambda_r r + \lambda_v v + \lambda_w w\right) / r \\
\dot{\lambda}_v &= \left[-\lambda_\vartheta \frac{1}{\cos\phi} - 2\lambda_u v + \lambda_v \left(u - w \tan\phi\right) + 2\lambda_w v \tan\phi\right] / r \\
\dot{\lambda}_w &= \left(-\lambda_\phi - 2\lambda_u w - \lambda_v v \tan\phi + \lambda_w u\right) / r \\
\dot{\lambda}_m &= T\lambda_V / m^2
\end{aligned}
$$

Thrust Direction



Optimal Thrust Angles

- γ_T thrust elevation angle, ψ_T thrust heading angle
- optimal values by posing $\partial H/\partial\gamma_T = 0$ and $\partial H/\partial\psi_T = 0$
- one obtains
  $\lambda_u \cos\gamma_T - (\lambda_v \cos\psi_T + \lambda_w \sin\psi_T)\sin\gamma_T = 0$
  $-\lambda_v \sin\psi_T + \lambda_w \cos\psi_T = 0$
- these equations provide
  $\sin\gamma_T = \lambda_u/\lambda_V$
  $\cos\gamma_T \cos\psi_T = \lambda_v/\lambda_V$
  $\cos\gamma_T \sin\psi_T = \lambda_w/\lambda_V$
- primer vector magnitude $\lambda_V = \sqrt{\lambda_u^2 + \lambda_v^2 + \lambda_w^2}$

Thrust Management (1)

- the thrust magnitude is discontinuous
- the thrust level can be decided during integration according to the sign of S_F
- the thrust magnitude at a given instant and the result of the integration may depend on the integration step
- convergence difficulties may be experienced because of scarce accuracy in the evaluation of the error gradients
- variable-step integration strategies usually provide better results

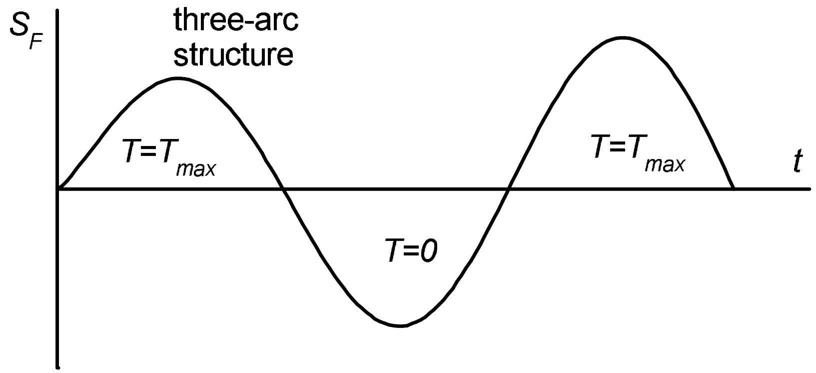
Thrust Management (2)

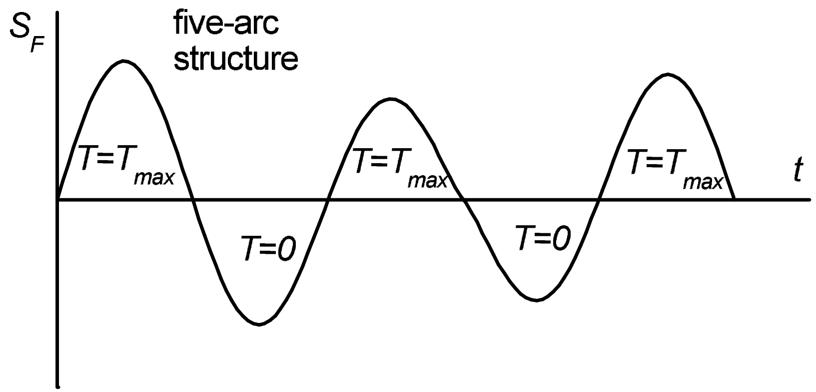
- an alternative approach consists in assuming "a priori" the *switching structure*, i.e., the succession of thrusting and coasting arcs, by splitting the trajectory into phases
- the thrust level during each phase is given and an accurate integration is allowed
- the switching times are additional unknown parameters of the BVP problem
- for free switching time, boundary conditions for optimality prescribe the Hamiltonian continuity $H_+ = H_-$
  $\boldsymbol{\lambda}_{r+}^T \boldsymbol{V}_+ + \boldsymbol{\lambda}_{V+}^T \boldsymbol{g}_+ + T_+ S_{F+} = \boldsymbol{\lambda}_{r-}^T \boldsymbol{V}_- + \boldsymbol{\lambda}_{V-}^T \boldsymbol{g}_- + T_- S_{F-}$
- for continuous state and adjoints variables (no constraints on the variables at the impulse) $S_F = 0$ is required

Thrust Management (3)

- the switching structure must be assigned a priori, based on experience and/or intuition
- the approach that decides the thrust magnitude during integration should be tried first
- if convergence difficulties are experienced, the switching function history may suggest the switching structure (thrusting arcs when $S_F > 0$, coasting arcs when $S_F < 0$)
- for a given switching structure, the solution must be checked in the light of PMP (verify that $S_F > 0$ during a thrusting phase and $S_F < 0$ during a coasting phase)
- thrusting or coasting arcs must be added or removed when PMP is violated

Switching Structure - Examples





Terminal Conditions (1)

- for maximum final mass $\lambda_{mf} = 1$
- for prescribed final position and velocity the final values of the corresponding adjoint variables are free
- for prescribed final position and free velocity the final values of the velocity adjoint variables are null and $\lambda_{Vf} = 0$ and, therefore, $S_{Ff} < 0$ (the thruster is off, since there is no use in thrusting if the velocity is free)

Terminal Conditions (2)

- minimum time, free final mass $\lambda_{mf} = 0$
- $\lambda_m \leq 0$ (in fact, its derivative is always positive) and $S_F > 0$ at any time (the thruster is always on)

Terminal Conditions (3)

- rendezvous requires $\boldsymbol{r}_f = \boldsymbol{r}_p(t_f)$, $\boldsymbol{V}_f = \boldsymbol{V}_p(t_f)$
- one has $\partial\boldsymbol{\psi}/\partial t_f = \partial\boldsymbol{\psi}/\partial(\boldsymbol{r}_p)_f\ (\dot{\boldsymbol{r}}_p)_f + \partial\boldsymbol{\psi}/\partial(\boldsymbol{V}_p)_f\ (\dot{\boldsymbol{V}}_p)_f = -\partial\boldsymbol{\psi}/\partial\boldsymbol{r}_f\ (\boldsymbol{V}_p)_f + \partial\boldsymbol{\psi}/\partial\boldsymbol{V}_f\ \boldsymbol{g}_f$ (with $\dot{\boldsymbol{r}}_p = \boldsymbol{V}_p$ and $\dot{\boldsymbol{V}}_p = \boldsymbol{g}$)
- boundary conditions for optimality and transversality condition combined provide $H - \boldsymbol{\lambda}_r^T \boldsymbol{V}_p - \boldsymbol{\lambda}_V^T \boldsymbol{g} = 0$ that is $TS_F = 0$, since $\boldsymbol{x} = \boldsymbol{x}_p$ at the rendezvous
- since the rendezvous is actually obtained when the thruster is switched on ($T > 0$), then $S_F = 0$ is usually imposed at the final point

Optimal Phasing

- assume optimal phasing between the relevant planets to define unconstrained optimum and find tentative solutions for the rendezvous problem
- "move" the target planet to the required position by evaluating position and velocity at the relevant time t as the values assumed at $t + t^*$, with t^* optimization parameter
- impose $\boldsymbol{r}_f = \boldsymbol{r}_p(t_f + t^*)$, $\boldsymbol{V}_f = \boldsymbol{V}_p(t_f + t^*)$
- boundary conditions for optimality require $\boldsymbol{\mu}^T(\partial\boldsymbol{\psi}/\partial t^*) = 0$
- since $\partial\boldsymbol{\psi}/\partial t^* = \partial\boldsymbol{\psi}/\partial t_f$, the transversality condition becomes $H_f = 0$, as in a time-free problem
- similar conditions are obtained if optimal phasing is assumed for an intermediate planet (flyby)

Impulsive Thrust (1)

- an impulse may me considered as a discontinuity in velocity and mass
- $\Delta \boldsymbol{V} = \boldsymbol{V}_+ - \boldsymbol{V}_-$, $m_+ = m_- \exp(-\Delta V/c)$
- boundary conditions for optimality require
  $-\boldsymbol{\lambda}_{V-} + (\mu m_-/c) \exp(-\Delta V/c)\, \Delta \boldsymbol{V}/\Delta V = 0$
  $\boldsymbol{\lambda}_{V+} - (\mu m_-/c) \exp(-\Delta V/c)\, \Delta \boldsymbol{V}/\Delta V = 0$
  $-\lambda_{m-} - \mu \exp(-\Delta V/c) = 0 \quad \lambda_{m+} + \mu = 0$
- this implies
  $\boldsymbol{\lambda}_{V-} = \boldsymbol{\lambda}_{V+} = \lambda_V \Delta \boldsymbol{V}/\Delta V$ (the impulse is parallel to the primer vector, which is continuous)
  $\lambda_{m+} = \lambda_{m-} \exp(+\Delta V/c)$ (the product $m\lambda_m$ is continuous)
  $\lambda_V = m\lambda_m/c$ (the switching function is null)

Impulsive Thrust (2)

- if the impulse time is free, transversality conditions require the continuity of the Hamiltonian through the impulse $H_+ = H_-$ (note that the mass and the corresponding adjoint variable do not appear in H, as no thrust is applied except at the impulse)
  $\boldsymbol{\lambda}_{r+}^T \boldsymbol{V}_+ + \boldsymbol{\lambda}_{V+}^T \boldsymbol{g}_+ = \boldsymbol{\lambda}_{r-}^T \boldsymbol{V}_- + \boldsymbol{\lambda}_{V-}^T \boldsymbol{g}_-$
- since position, gravity, and adjoint variable are continuous, one obtains $\boldsymbol{\lambda}_r^T(\boldsymbol{V}_+ - \boldsymbol{V}_-) = \boldsymbol{\lambda}_r^T \Delta \boldsymbol{V} = 0$
- this implies $\boldsymbol{\lambda}_r^T \boldsymbol{\lambda}_V = 0$, that is $\dot{\boldsymbol{\lambda}}_V^T \boldsymbol{\lambda}_V = \lambda_V \dot{\lambda}_V = 0$
- the derivative of the primer magnitude must be null, and, therefore, also $\dot{S}_F = 0$

Impulsive Thrust (3)

- the switching structure (i.e., number and sequence of impulses) must be assumed a priori
- impulses must be added when the switching function becomes positive, i.e., $\lambda_V > m\lambda_m/c$ (note that the larger is c, the smaller is the value that λ_V must assume at the impulse)
- impulse magnitude tends to vanish when the impulse is not required (signaling that the impulse should be removed from the switching structure)

Interplanetary Trajectories - Departure (1)

- the dimension of and the time spent inside the sphere of influence is neglected
- impulsive departure from parking orbit
- hyperbolic excess velocity $\boldsymbol{V}_\infty = \boldsymbol{V}_0 - \boldsymbol{V}_p$ with $\boldsymbol{V}_p$ planet heliocentric velocity and $\boldsymbol{V}_0$ spacecraft heliocentric velocity at escape
- impulse magnitude $\Delta V = \sqrt{V_\infty^2 + V_{esc}^2} - V_{orb}$, with V_{esc} and V_{orb} escape velocity and orbital velocity at departure, respectively (the departure point should be the periapsis of the parking orbit)
- escape mass $m_0 = m_{orb}\exp(-\Delta V/c)$
- position $\boldsymbol{r}_0 = \boldsymbol{r}_p$

Interplanetary Trajectories - Departure (2)

- boundary conditions for optimality require
  $\boldsymbol{\lambda}_{V0} - (\mu m_{orb}/c)\,\exp(-\Delta V/c)\,\boldsymbol{V}_\infty/\sqrt{V_\infty^2 + V_{esc}^2} = 0$
  $\lambda_{m0} + \mu = 0$
- this implies
  $\boldsymbol{\lambda}_{V0} = \lambda_{V0}\boldsymbol{V}_\infty/\sqrt{V_\infty^2 + V_{esc}^2}$ (the hyperbolic excess velocity is parallel to the primer vector)
  $\lambda_{V0} = (m_0\lambda_{m0}/c)V_\infty/\sqrt{V_\infty^2 + V_{esc}^2}$
- in comparison with a deep-space impulse, where $\lambda_V = m\lambda_m/c$, an impulse inside a sphere of influence occurs with a negative switching function, i.e., it is more convenient to use thrust close to a planet than outside its sphere of influence

Interplanetary Trajectories - Arrival

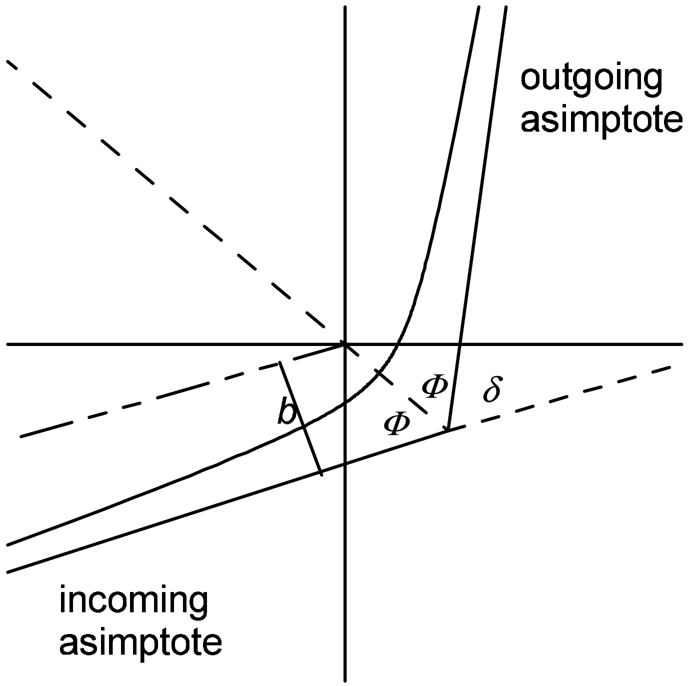
- specular maneuver compared to departure
- under the same assumptions $\boldsymbol{r}_f = \boldsymbol{r}_p$, $\boldsymbol{V}_\infty = \boldsymbol{V}_f - \boldsymbol{V}_p$, $\Delta V = \sqrt{V_\infty^2 + V_{esc}^2} - V_{orb}$
- mass delivered into the parking orbit $m_{orb} = m_f \exp(-\Delta V/c)$ is maximized
- boundary conditions are $\boldsymbol{\lambda}_{Vf} = -\lambda_{Vf} \boldsymbol{V}_\infty / \sqrt{V_\infty^2 + V_{esc}^2}$ (the hyperbolic excess velocity is parallel and opposite to the primer vector) $\lambda_{Vf} = (m_f \lambda_{mf}/c) V_\infty / \sqrt{V_\infty^2 + V_{esc}^2}$

Interplanetary Trajectories - Flyby (1)

- the dimension of and the time spent inside the sphere of influence is neglected
- planet intercept requires $\boldsymbol{r}_+ = \boldsymbol{r}_- = \boldsymbol{r}_p$
- planetocentric energy conservation requires conservation of hyperbolic excess velocity magnitude
  $\boldsymbol{V}^2_{\infty+} = (\boldsymbol{V}_+ - \boldsymbol{V}_p)^2 = \boldsymbol{V}^2_{\infty-} = (\boldsymbol{V}_- - \boldsymbol{V}_p)^2$
- the V_∞ turn depends on the height of the hyperbola periapsis R_p

Interplanetary Trajectories - Flyby (2)

- rotation angle
  $\delta = \pi - 2\Phi$
- $\cos\Phi = \sin(\delta/2) = 1/e$
- $\sin(\delta/2) = \dfrac{\mu_p/R_p}{V_\infty^2 + \mu_p/R_p}$
- $b = r_p \dfrac{\sqrt{V_\infty^2 + 2\mu_p/R_p}}{V_\infty}$



Free-Height Flyby

- imposed boundary conditions $\boldsymbol{r}_+ = \boldsymbol{r}_- = \boldsymbol{r}_p$, $\boldsymbol{V}^2_{\infty+} = \boldsymbol{V}^2_{\infty-}$, and $m_+ = m_-$
- boundary conditions for optimality
  $\boldsymbol{\lambda}_{r-} - \boldsymbol{\lambda}_{r+} = \boldsymbol{\mu}_1$ (the position adjoint vector presents a free discontinuity)
  $\lambda_{ms+} = \lambda_{ms-}$ (the mass adjoint variable is continuous)
  $\boldsymbol{\lambda}_{V-} = -2\mu_2 \boldsymbol{V}_{\infty-} \quad \boldsymbol{\lambda}_{V+} = -2\mu_2 \boldsymbol{V}_{\infty+}$
- this implies that the primer must be parallel to the hyperbolic excess velocity before and after the flyby and that its magnitude is continuous

Minimum-Height Flyby (1)

- additional boundary condition on the velocity turn $\boldsymbol{V}_{\infty+}^T \boldsymbol{V}_{\infty-} = -\cos 2\phi\, \boldsymbol{V}_{\infty-}^2$ with $\cos\phi = V_p^2/(\boldsymbol{V}_{\infty-}^2 + V_p^2)$ and $V_P = \sqrt{\mu_p/R_p}$ circular velocity at the minimum allowable distance from the planet
- boundary conditions for optimality state that the position adjoint vector presents a free discontinuity and the mass adjoint variable is continuous
- $\boldsymbol{\lambda}_{V-} = -2\mu_2 \boldsymbol{V}_{\infty-} + \mu_4 \boldsymbol{V}_{\infty+} + 2\mu_4 B\, \boldsymbol{V}_{\infty-}$
  $\boldsymbol{\lambda}_{V+} = -2\mu_2 \boldsymbol{V}_{\infty+} - \mu_4 \boldsymbol{V}_{\infty-}$ with
  $B = \cos 2\phi - A \sin 2\phi$ and $A = \dfrac{\mathrm{d}\phi}{\mathrm{d}\boldsymbol{V}_{\infty-}} \boldsymbol{V}_{\infty-} = \dfrac{2}{\tan\phi} \dfrac{V_{\infty-}^2}{V_{\infty-}^2 + V_p^2}$

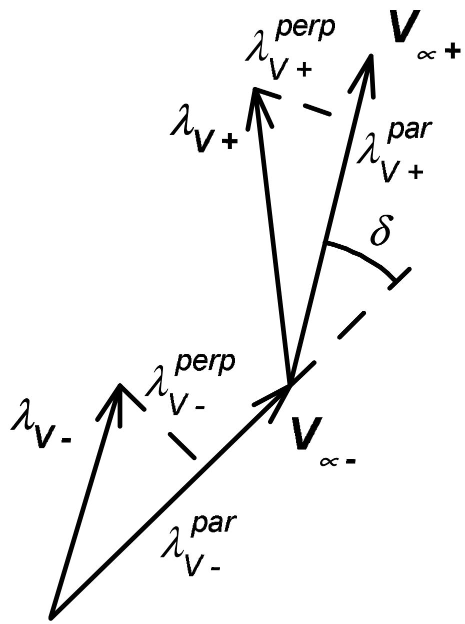
Minimum-Height Flyby (2)

- one has $\boldsymbol{\lambda}_{V-} \times \boldsymbol{V}_{\infty -} = -\mu_4 \boldsymbol{V}_{\infty +} \times \boldsymbol{V}_{\infty -}$ and $\boldsymbol{\lambda}_{V+} \times \boldsymbol{V}_{\infty +} = \mu_4 \boldsymbol{V}_{\infty -} \times \boldsymbol{V}_{\infty +} = \boldsymbol{\lambda}_{V-} \times \boldsymbol{V}_{\infty -}$ the primer component perpendicular to $\boldsymbol{V}_{\infty}$ is the same before and after the flyby ($\lambda_{V+}^{perp} = \lambda_{V-}^{perp}$)
- also $(\boldsymbol{\lambda}_{V-} \times \boldsymbol{V}_{\infty -}) \cdot \boldsymbol{V}_{\infty +} = 0$ and $(\boldsymbol{\lambda}_{V+} \times \boldsymbol{V}_{\infty +}) \cdot \boldsymbol{V}_{\infty -} = 0$ that is $\boldsymbol{\lambda}_{V} \cdot (\boldsymbol{V}_{\infty +} \times \boldsymbol{V}_{\infty -}) = 0$ the primer lies on the flyby plane before and after the flyby
- finally $\boldsymbol{\lambda}_{V-} \cdot \boldsymbol{V}_{\infty -} - 2\mu_2 V_{\infty -}^2 + \mu_4 \boldsymbol{V}_{\infty +} \cdot \boldsymbol{V}_{\infty -} + 2\mu_4 B\, V_{\infty -}^2 = 0$ and $-\boldsymbol{\lambda}_{V+} \cdot \boldsymbol{V}_{\infty +} + 2\mu_2 V_{\infty +}^2 + \mu_4 \boldsymbol{V}_{\infty -} \cdot \boldsymbol{V}_{\infty_+} = 0$ which provide the change in the primer component parallel to $\boldsymbol{V}_{\infty}$, that is $\lambda_{V_+}^{par} = \lambda_{V_-}^{par} + 2A\, \lambda_{V-}^{perp}$

Minimum-Height Flyby (3)

- the direction of the primer component perpendicular to the hyperbolic excess velocity must be checked to ensure the height constraint requirement
- the constraint is required when the $\boldsymbol{\lambda}_V$ perpendicular component is directed toward the planet, i.e., is in the direction of the velocity rotation
- the constraint must be removed when it is directed opposite to the velocity turn
- it is sufficient to check the sign of $(\boldsymbol{\lambda}_V \times \boldsymbol{V}_\infty) \cdot (\boldsymbol{V}_{\infty+} \times \boldsymbol{V}_{\infty-})$, which must be positive

Minimum-Height Flyby (4)



Flyby - Transversality Conditions (1)

- one has $t_+ = t_-$ and, as far as the condition at flyby on position and velocity are concerned (only the terms $\boldsymbol{r} - \boldsymbol{r}_p$ and $\boldsymbol{V} - \boldsymbol{V}_p$ appear), one has
  $\partial\boldsymbol{\psi}/\partial t = -(\partial\boldsymbol{\psi}/\partial\boldsymbol{r})\dot{\boldsymbol{r}}_p - (\partial\boldsymbol{\psi}/\partial\boldsymbol{V})\dot{\boldsymbol{V}}_p$
- the transversality condition becomes
  $H_- - \boldsymbol{\mu}^T(\partial\boldsymbol{\psi}/\partial\boldsymbol{r}_-\boldsymbol{V}_p + \partial\boldsymbol{\psi}/\partial\boldsymbol{V}_-\boldsymbol{g}) - \mu_\tau = 0$
  $-H_+ - \boldsymbol{\mu}^T(\partial\boldsymbol{\psi}/\partial\boldsymbol{r}_+\boldsymbol{V}_p + \partial\boldsymbol{\psi}/\partial\boldsymbol{V}_+\boldsymbol{g}) + \mu_\tau = 0$
- using the boundary conditions on the adjoint variables these equations become
  $H_- - \boldsymbol{\lambda}_{r-}^T\boldsymbol{V}_p - \boldsymbol{\lambda}_{V-}^T\boldsymbol{g} = H_+ - \boldsymbol{\lambda}_{r+}^T\boldsymbol{V}_p - \boldsymbol{\lambda}_{V+}^T\boldsymbol{g}$
- using $H = \boldsymbol{\lambda}_r^T\boldsymbol{V} + \boldsymbol{\lambda}_V^T\boldsymbol{g} + TS_F$ one has
  $\boldsymbol{\lambda}_{r-}^T\boldsymbol{V}_{\infty-} + T_-S_{F-} = \boldsymbol{\lambda}_{r+}^T\boldsymbol{V}_{\infty+} + T_+S_{F+}$

Flyby - Transversality Conditions (2)

- for a free-height flyby $\boldsymbol{\lambda}_V$ is parallel to $\boldsymbol{V}_\infty$ and $\lambda_{V+} = \lambda_{V-}$
- therefore, $S_{F+} = S_{F-}$ and $T_+ = T_-$ (in fact, the thrust magnitude is only determined by S_F)
- the transversality condition becomes $\dot{\lambda}_{V+} = \dot{\lambda}_{V-} = 0$
- on the contrary, for a minimum-height flyby S_F is discontinuous and the thrust may also be discontinuous

Propulsion System Optimization (1)

- when electric propulsion is employed, the available power and the values of thrust and specific impulse may be subject to optimization
- the propulsion system mass can be considered in the performance index by posing $\varphi = m_f - m_{ps}$
- assume m_{ps} proportional to the thrust power $m_{ps} = \beta T c/2$
- additional (constant) variables T and c with $\dot{T} = 0$ and $\dot{c} = 0$
- introduce adjoint variables λ_T and λ_c and derive Euler-Lagrange equations $\dot{\lambda}_T = -\partial H/\partial T$ and $\dot{\lambda}_c = -\partial H/\partial c$

Propulsion System Optimization (2)

- derive boundary conditions for optimality at the initial point $\lambda_{T0} = 0$, $\lambda_{c0} = 0$ and final point $\lambda_{mf} = 1$, $\lambda_{Tf} = -\beta c/2$, $\lambda_{cf} = -\beta T/2$
- solution provides the optimal values for T and c
- a constraint on the trip time is required to avoid $T \to 0$ and $t \to \infty$

Variable Specific Impulse (1)

- electric propulsion system with variable specific impulse and thrust at constant power (and efficiency) $P = Tc/(2\eta)$ and $T = 2\eta P/c$
- consider P and c as control variables with $0 \leq P \leq P_{max}$ with unbounded specific impulse (and thrust)
- the Hamiltonian for optimal thrust direction (parallel to the primer vector) is $H = \boldsymbol{\lambda}_r^T \boldsymbol{V} + \boldsymbol{\lambda}_V^T \boldsymbol{g} + \left(\frac{\lambda_V}{m} - \frac{\lambda_m}{c} \right) 2\eta \frac{P}{c}$
- optimal exhaust velocity from $\partial H / \partial c = 0$, that is $c = 2m\lambda_m/\lambda_V$, which gives $T = \eta P \lambda_V / (m\lambda_m)$
- the Hamiltonian becomes $H = \boldsymbol{\lambda}_r^T \boldsymbol{V} + \boldsymbol{\lambda}_V^T \boldsymbol{g} + \eta \frac{\lambda_V}{m} \frac{P}{c}$
- $\partial H / \partial P > 0$ and $P = P_{max}$ is always required

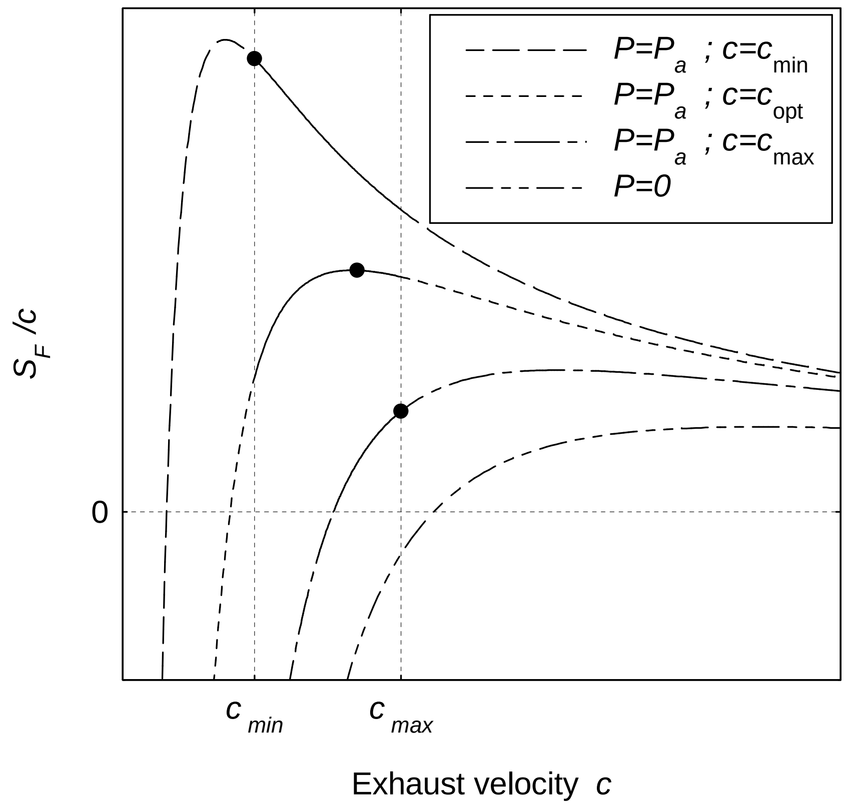
Variable Specific Impulse (2)

- a constraint on the total trip time is required to avoid solutions with $T \to 0$ and $c \to \infty$ for $t \to \infty$
- the quantity $m^2\lambda_m = b$ is constant, as $\frac{\mathrm{d}(m^2\lambda_m)}{\mathrm{d}t} = \lambda_V T - 2m\lambda_m\frac{T}{c} = 0$ for the optimal exhaust velocity
- one has $c = 2b/(m\lambda_V)$ and the acceleration $T/m = \eta P \lambda_V / b$ only depends on λ_V
- for a given trip time the acceleration history is fixed and does not depend on the mission parameters (power, efficiency)
- exhaust velocity history and payload, instead, depend on P and η

Variable Specific Impulse (3)

- for bounded specific impulse $c_{\text{min}} \leq c \leq c_{\text{max}}$ and thrust $T_{\text{min}} = 2\eta P/c_{\text{max}} \leq T \leq T_{\text{max}} = 2\eta P/c_{\text{min}}$, define the power switching function $S_P = \frac{\lambda_v}{m} - \frac{\lambda_m}{c_{\text{max}}} \geq \frac{\lambda_v}{m} - \frac{\lambda_m}{c}$
- when $S_P < 0$ the quantity $\partial H/\partial P$ is surely negative and $P = 0$, otherwise $P = P_{max}$ must be adopted
- $c = 2m\lambda_m/\lambda_V$ when $c_{min} \leq 2m\lambda_m/\lambda_V \leq c_{max}$, otherwise the closer extreme must be adopted (c_{min} when $c_{\text{opt}} \leq c_{\text{min}}$, c_{max} when $c_{\text{opt}} \geq c_{\text{max}}$)
- discrete values of the specific impulse could also be considered

Variable Specific Impulse (4)



Aerodynamic Forces

- equations of motion

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \boldsymbol{V}$$

$$\frac{\mathrm{d}\boldsymbol{V}}{\mathrm{d}t} = \boldsymbol{g} + \frac{\boldsymbol{T}}{m} + \frac{\boldsymbol{D}}{m} + \frac{\boldsymbol{L}}{m}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{T}{c}$$

- lift $\boldsymbol{L} = qSC_L\boldsymbol{L}/L$, drag $\boldsymbol{D} = qSC_D\boldsymbol{V}_r/V_r$, with dynamic pressure $q = \rho V_r^2/2$ and relative velocity $\boldsymbol{V}_r$
- assume parabolic polar $C_D = C_{D_0} + KC_L^2$

Optimal Aerodynamic Control

- Hamiltonian

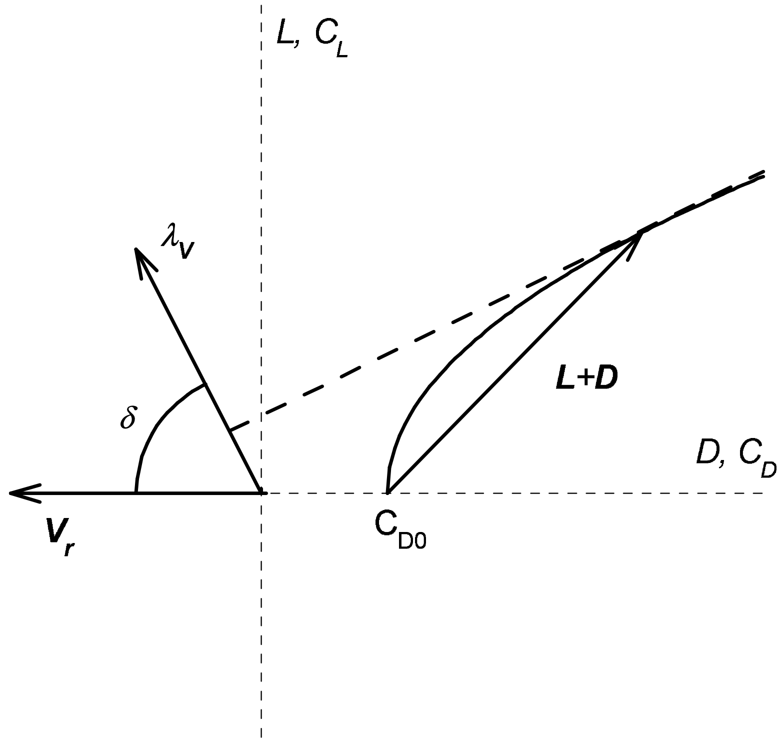
$$H = \boldsymbol{\lambda}_r^T \boldsymbol{V} + \boldsymbol{\lambda}_V^T \boldsymbol{g} + T S_F + \frac{qS}{m} A_F$$

- aerodynamic acceleration coefficient

$$A_F = -C_D \boldsymbol{\lambda}_V^T \boldsymbol{V}_r / V_r + C_L \boldsymbol{\lambda}_V^T \boldsymbol{L} / L$$

- optimal controls maximize the scalar projection of the aerodynamic force on $\boldsymbol{\lambda}_V$
- lift in the plane defined by relative velocity $\boldsymbol{V}_r$ and primer vector $\boldsymbol{\lambda}_V$
- lift coefficient $C_L = C_L = \tan\delta/2K$ (or $C_L = C_{L_{max}}$ if $\tan\delta/2K > C_{L_{max}}$) where δ is the angle between $\boldsymbol{V}_r$ and $\boldsymbol{\lambda}_V$

Optimal Aerodynamic Force



Geopotential Perturbation

- potential given by

$$\frac{V}{\mu/r} = -\left[1 + \sum_{n=2}^{N} \left(\frac{r_E}{r}\right)^n \sum_{m=0}^{n} \left(C_{nm} \cos m\lambda + S_{nm} \sin m\lambda\right) P_{nm}(\sin\varphi)\right]$$

- perturbing acceleration

$$(a_J)_u = \partial\Phi/\partial r$$

$$(a_J)_v = (\partial\Phi/\partial\vartheta)/(r\cos\varphi)$$

$$(a_J)_w = (\partial\Phi/\partial\varphi)/r$$

- derivarives of the adjoint variables are modified: simple calculations through a recursive scheme

Third-Body Gravitational Perturbation

- perturbing acceleration ($\boldsymbol{R} = \boldsymbol{r} - \boldsymbol{r}_p$ is the spacecraft position vector wrt third body)

$$\boldsymbol{a}_p = -(\mu_p/R^3)\boldsymbol{R} - (\mu_p/r_p^3)(\boldsymbol{r}_p)$$

- projection in the topocentric frame gives components of the perturbing accelerations as functions of position vector
- derivarives of the adjoint variables to position are modified: simple calculations (even though tedious)

Ascent: Maximum Dynamic Pressure

- an arc with constant dynamic pressure q at the maximum allowable value may be required
- thrust modulation to assure constant q
- q is a function of state variables through relative velocity and density $q = q[(\rho(r), V_{rel}] = q(r, u, v, w)$
- the equation

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial q}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}}$$

  provides the thrust magnitude
- $q = q_{max}$ at the start of the arc and $T = T_{max}$ at the arc end are the required boundary conditions
- Euler-Lagrange equations and boundary conditions for optimality are modified

Ascent: Maximum Heat Flux

- a constraint on the heat flux h may be required
- $h = h_{max}$ and possibly $\dot{h} = 0$ are the required boundary conditions
- h is a function of state variables through relative velocity and density $h = h[(\rho(r), V_{rel}] = h(r, u, v, w)$
- discontinuities in the adjoint variables arise
- good point for multiple shooting

Singular Arcs: When

- bang-bang control can be explained by the requirement of thrusting in the most favorable positions
  - larger thrust reduces velocity losses: use T_{max}
  - gravitational losses are smaller for larger velocity
- singular arcs arise when a reduced thrust is useful
  - aerodynamic drag: grows with velocity
  - three-body ?
- signal of requirement
  - chattering
  - irregular behavior of the switching function (it is not possible to satisfy PMP)

Singular Arcs in Atmospheric Flight

- Hamiltonian

$$H = \boldsymbol{\lambda}_r^T \boldsymbol{v} + \boldsymbol{\lambda}_v^T \boldsymbol{g} + \frac{T}{m} S_F + \frac{qS}{m} A_F$$

- thrust switching function and aerodynamic coefficient (δ: angle between velocity and primer vector)

$$S_F = \lambda_v - \lambda_m \, m/c$$

$$A_F = -C_D \lambda_v \cos\delta + C_L \lambda_v \sin\delta$$

$$C_L = \frac{\tan\delta}{2K}, \qquad C_D = C_{D0} + K C_L^2$$

Singular Arcs in Atmosphere: Derivation

- S_F and $\dot{S}_F$ vanish at arc extremities: boundary conditions to determine initial and final time, with

$$\dot{S}_F = \frac{\partial S_F}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}} + \frac{\partial S_F}{\partial \boldsymbol{\lambda}}\dot{\boldsymbol{\lambda}}$$

- $\dot{S}_F$ does not contain T

$$\dot{S}_F = -\frac{\boldsymbol{\lambda}_r^T \boldsymbol{\lambda}_v}{\lambda_v} - \frac{S}{m}\left[\frac{qA_F}{c} + \left(A_F \boldsymbol{Q}_v^T + q\boldsymbol{A}_v^T\right)\frac{\boldsymbol{\lambda}_v}{\lambda_v}\right]$$

- coefficients

$$\boldsymbol{Q}_v^T = \partial q / \partial \boldsymbol{v} \qquad \boldsymbol{A}_v^T = \partial A_F / \partial \boldsymbol{v}$$

Singular Arcs in Atmosphere: Thrust

- $\ddot{S}_F = 0$ provides thrust magnitude (first-order singular arc)

$$\ddot{S}_F = \frac{\partial \dot{S}_F}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}} + \frac{\partial \dot{S}_F}{\partial \boldsymbol{\lambda}}\dot{\boldsymbol{\lambda}}$$

- T appears in derivatives of velocity components and mass ($\boldsymbol{a}_0$: inertial + aerodynamic acceleration)

$$\ddot{S}_F = \Lambda - \frac{S}{m}\left(\boldsymbol{S}_r^T\dot{\boldsymbol{r}} + \boldsymbol{S}_v^T\boldsymbol{a}_0 + \boldsymbol{S}_\lambda^T\dot{\boldsymbol{\lambda}}_v + S_T T\right)$$

- setting $\ddot{S}_F = 0$ provides

$$T^* = \frac{m\Lambda/S - \boldsymbol{S}_r^T\boldsymbol{v} - \boldsymbol{S}_v^T\boldsymbol{a}_0 - \boldsymbol{S}_\lambda^T\dot{\boldsymbol{\lambda}}_v}{S_T}$$

- $S_T = 0$ vanishes in the absence of aerodynamic terms

Requirements for Convergence

- BVP solution is based on assumption of linear behavior
  - accurate evaluation of the error gradients
  - small corrections, i.e., tentative solution close to optimal solution

Error Gradient Evaluation

- accurate integration of the equations of motion
  - variable step (e.g., Adams Moulton) can handle control discontinuities
  - separation into phases can handle control discontinuities
- multiple shooting to reduce influence of single parameters
- normalization is usually mandatory (same order of magnitude for all the variables)

Tentative Solution

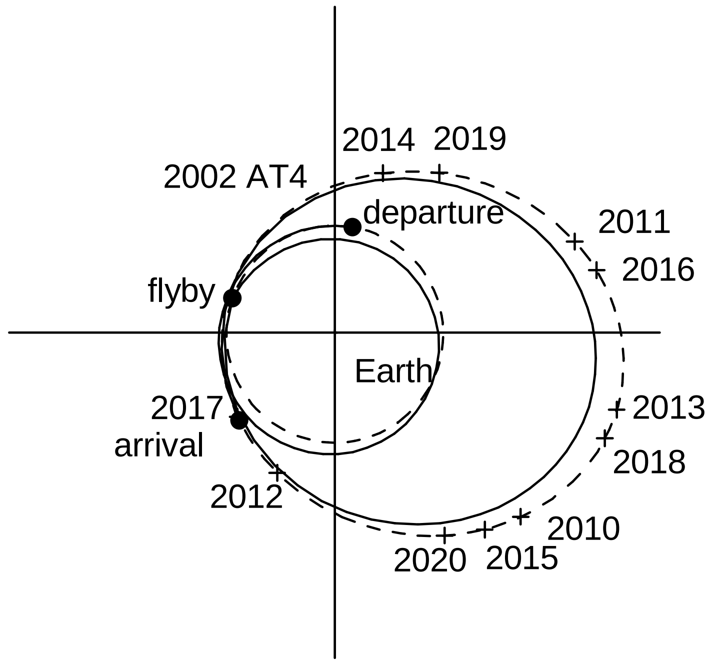
- must be sufficiently close to the optimal solution
  - switching structure/constraints assessment
  - avoid divergence
  - small values for the adjoint variables
- correction relaxation (i.e., use a fraction of the computed correction)

Optimal Phasing

- define mission opportunities in a large launch window (in particular when gravity assist is exploited)
  - evaluate optimal phasing transfer and its periodicity
  - check for target position on possible arrival dates
- find most favorable positions for gravity assists (plane/eccentricity changes)
- favorable opportunities when actual position close to desired one

Optimal Phasing - Asteroid Rendezvous

- optimal phasing transfers from earth repeat every year
- check target position on arrival day/month in available years
- one/two good opportunities every synodic cycle

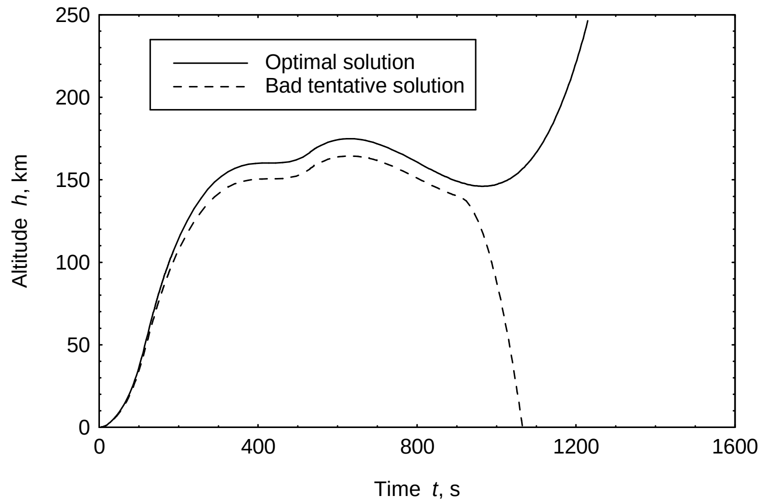


Multiple Shooting

- some cases are very sensitive to initial values
  - ascent: strong nonlinearity
  - multiple flybys: sudden discontinuities

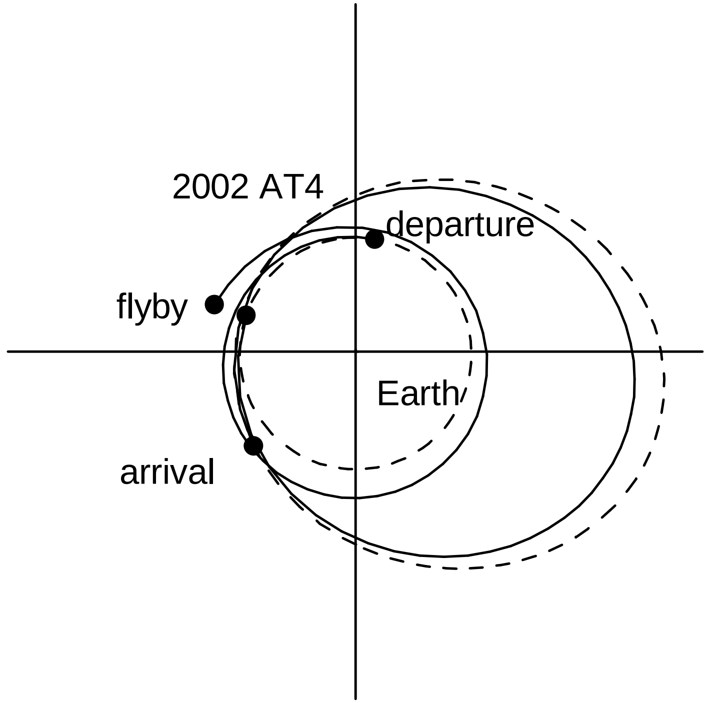
Multiple Shooting - Ascent

- uncorrect values: no integration possible
- split (e.g., at maximum heat flux)
- re-start with values that assure continuous ascent (e.g. positive radial velocity)



Multiple Shooting - Flybys

- use point immediately after flyby as split point
- accommodate for errors due to inaccurate estimations of the initial values
- position and velocity components suggested by physics of the problem
- try nonpropelled tentative solutions

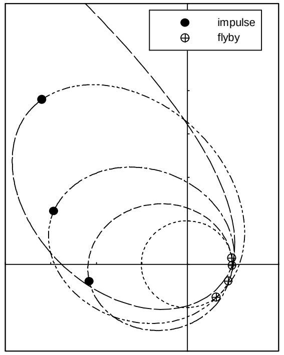


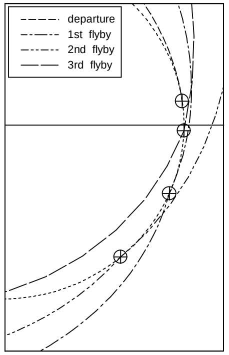
Trajectory Patching

- trajectories with multiple arcs
  - ΔV gravity-assist orbits
  - multiple-flyby trajectories
- solve simple one-arc problem with proper boundary/optimality conditions
- patch simple trajectories with a multiple-shooting approach

Trajectory Patching - ΔV-GA Trajectories

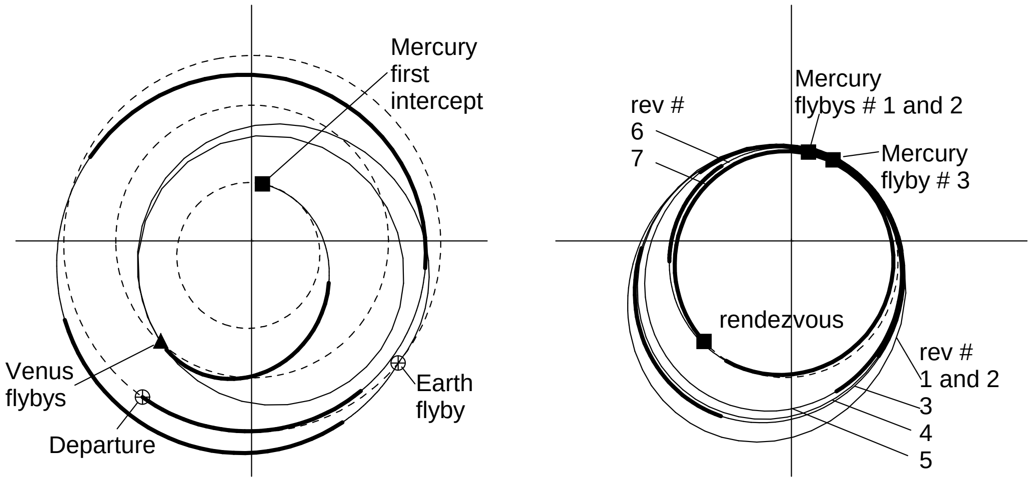
- maximize V_∞ gain for assigned propellant consumption (or dual problem)
- use nonpropelled resonant orbit as tentative solution





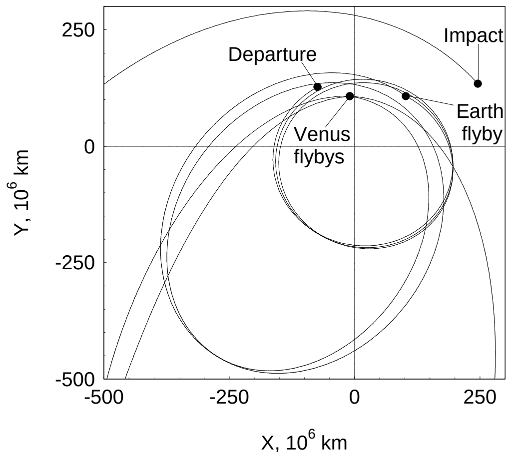
Trajectory Patching - Multiple Flybys

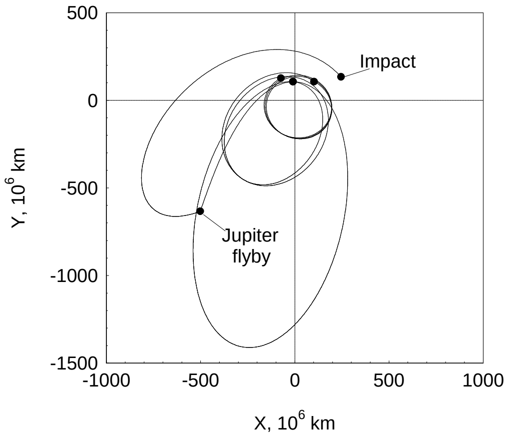
- connecting planets: Hohmann transfer
- resonant orbits: assign arrival V_∞ for following leg
- orientation suggested by geometry of orbits



Trajectory Patching - GTOC1

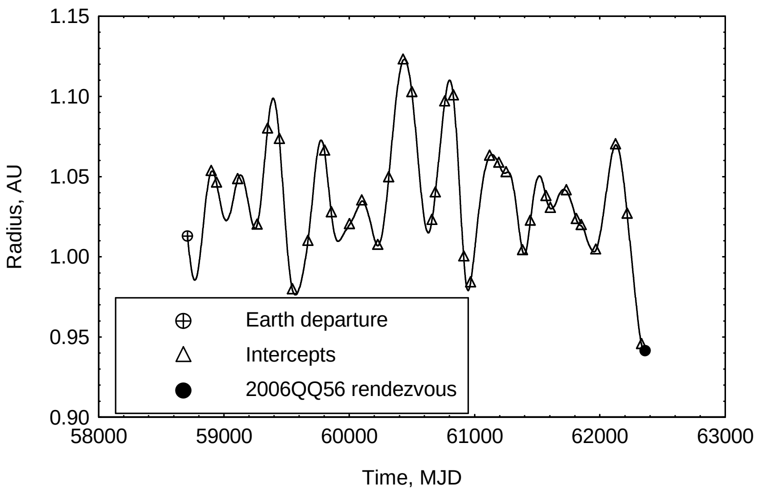
- EVVJ-GA transfer to deflect asteroid 2001 TW229
- first, each leg optimized separately
- final optimization of complete trajectory





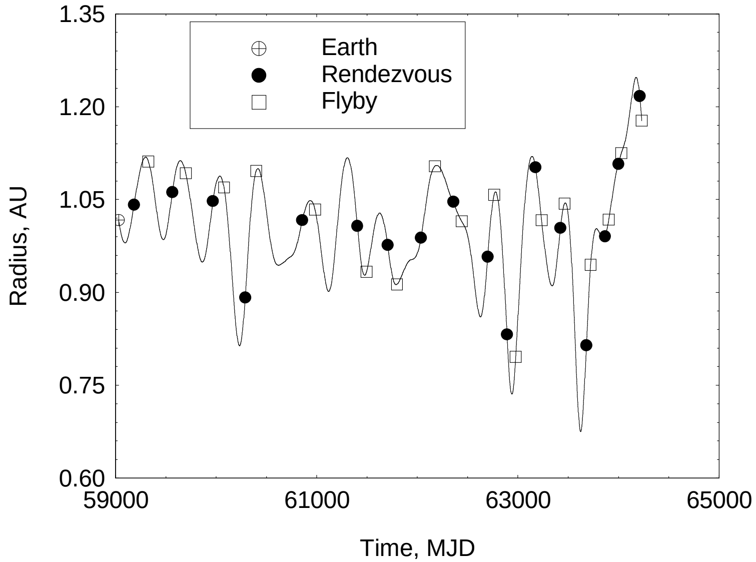
Trajectory Patching - GTOC4

- flyby of 36 asteroids
- legs encompassing up to 15 flybys are optimized
- start following leg with 3 or 4-asteroid overlap



Trajectory Patching - GTOC5

- rendezvous and flyby of 17 asteroids
- first, simple R(i)-F(i)-R(i+1) legs optimized separately, mixed time-mass performance index
- optimization of up to 5 joined legs follows

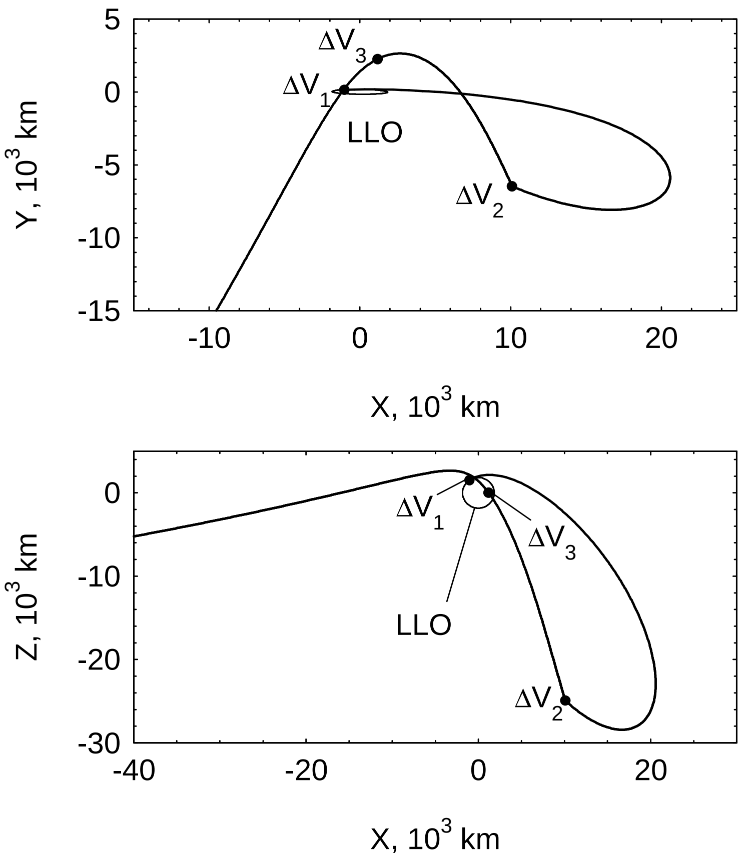


Homotopy

- complex trajectories/dynamics
  - three-body problem
- solve simple constrained trajectory
- solve transition problems towards optimal solution with blended boundary conditions between constrained/optimal problem

Homotopy - Moon Return

- fix proper times and impulses for starting solution
- linear combination of the boundary conditions with multipliers c and $1 - c$
- quite easy convergence

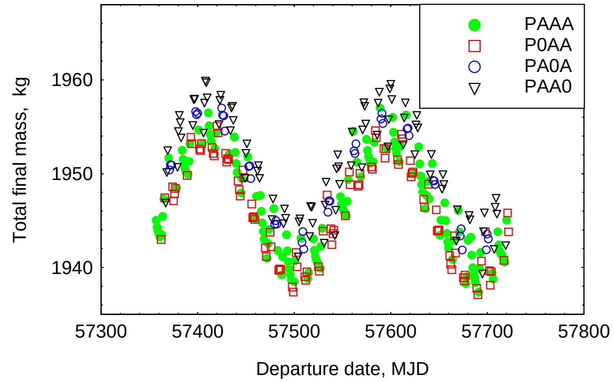


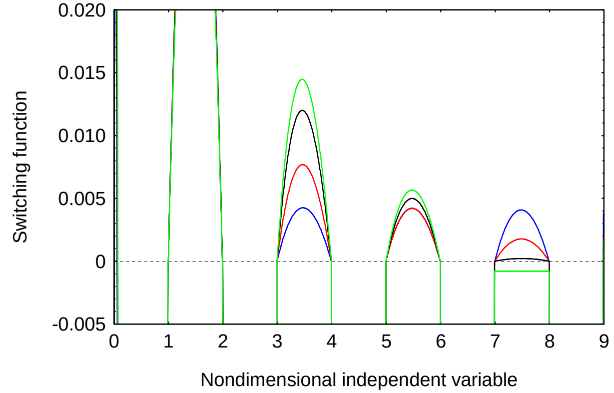
Continuation

- unknown switching structure:
  - lunisolar effect on Highly Elliptic Orbit deployment
- gradual introduction of disturbing parameter
- (automatic) check evolution of switching structure
- extremely more efficient than smoothing techniques

Continuation - HEO Deployment

- multiple revolutions
- energy modulated to have favorable Moon phasing at apogee passages
- (almost) equal burns at each apogee passage without the Moon
- apogee burns may vanish as Moon's gravity is introduced





Independent Variable Selection

- thrusting is usually preferred at specific positions, fixed in space
- quasi-periodicity with respect to longitude
- passage times may be unknown (depend on trajectory)
  - multiple revolution geocentric transfers
  - lunisolar effect on Highly Elliptic Orbit deployment
- use longitude as the independent variable

Independent Variable Selection - HEO Deployment

- burns at apsides
- orbit periods dictated by better use of lunisolar perturbation
- accurate integration reqiures splitting the trajectory into arcs
- switching times cannot be estimated, but switching angles can

Comparison

1/1/15-15/1/15

Time, hr	Long., rad
84.4-84.3	10.73-10.75
84.5-84.4	11.27-11.25
125.7-126.5	14.13-14.14
129.2-126.5	14.16-14.14
212.4-211.7	20.41-20.42
219.2-214.9	20.46-20.44
311.8-299.9	26.70-26.69
311.8-304.9	26.70-26.74
408.2-397.8	33.00-33.00

Independent Variable Selection - Multiple Revs.

- typical case: LeoGeo transfer
- continuation approach: gradually increase number of revs., final radius, inclination change
- small changes are accepted easily
- very difficult strict fulfillment of boundary conditions

Type	I_{sp}, s	t, d	Revs.	mass
Edelbaum	∞	191	1048	1
Ion thruster	3000	174	989	0.822
Hall thruster	1500	158	936	0.675
Arcjet	600	122	802	0.374

Smart Constraint Position

- proper constraint definition may simplify derivation and solution
- use alternative (equivalent or almost equivalent) simpler form of constraint
  - separation constraint at apogee in HEO: $|\boldsymbol{r}_1 - \boldsymbol{r}_2| = k$
  - can be turned into time constraint $t_{a2} - t_{a1} = k/v_a$

Nested Indirect Methods

- stepped approach to solve very complex problem
- in multiple-revolution transfers solve single revolution problem first
- patch single revolution solutions together
  - LeogGeo transfer with electric propulsion

Nested Indirect Methods: Single Revolution

- one revolution: almost circular orbit, small inclination change, constant mass (Edelbaum's approximation)
- apply OCT to define optimal controls (thrust angles, specific impulse if variable) and evaluate changes of a i and m in one revolution
- changes for unit radius orbit with unit mass
  - constant specific impulse $\Delta r = 8\pi P \cos\beta / c_N$, $\Delta i = 8P \sin\beta / c_N$, $\Delta m = -4\pi P / c_N^2$, control is out-of-plane angle β
  - variable specific impulse $\Delta r = 16\pi P K_1$, $\Delta i = 2\pi P K_3$, $\Delta m = -2\pi P (8K_1^2 + K_3^2)$, controls are K_1 and K_3

Nested Indirect Methods: Multiple Revolutions

- multiple revolutions: radius and mass changes are considered and elements changes are modified accordingly
- differential equations with r as independent variable
  - constant specific impulse

$$\mathrm{d}i/\mathrm{d}r = \tan\beta/(\pi r) \quad \mathrm{d}t/\mathrm{d}r = m/(2\sqrt{r^3}T\cos\beta)$$

$$\mathrm{d}m/\mathrm{d}r = -m/(2c\sqrt{r^3}\cos\beta)$$

  - variable specific impulse

$$\mathrm{d}i/\mathrm{d}r = K_3/(8rK_1) \quad \mathrm{d}t/\mathrm{d}r = m/(8P\sqrt{r^3}K_1)$$

$$\mathrm{d}m/\mathrm{d}r = -m(8K_1^2 + K_3^2)/(8\sqrt{r^3}K_1)\sqrt{r^3}T\cos\beta)$$

Nested Indirect Methods: Optimal Problem

- apply OCT to determine optimal values of the controls (either β or K_1 and K_2)
  - constant specific impulse

$$H = \lambda_i \tan\beta/(\pi r) + \lambda_t m/(2\sqrt{r^3}T\cos\beta) - \lambda_m m/(2c\sqrt{r^3}\cos\beta)$$

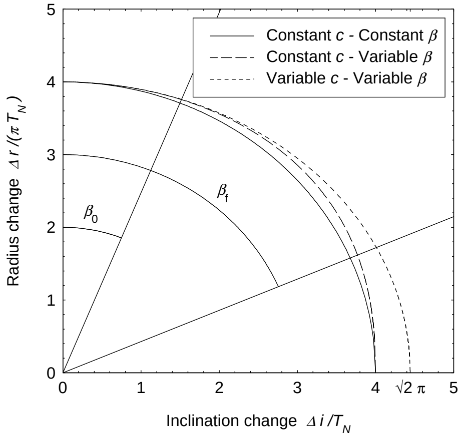
$$\sin\beta = \frac{2\lambda_i}{\pi m(\lambda_m/c - \lambda_t/T)}\sqrt{r}$$

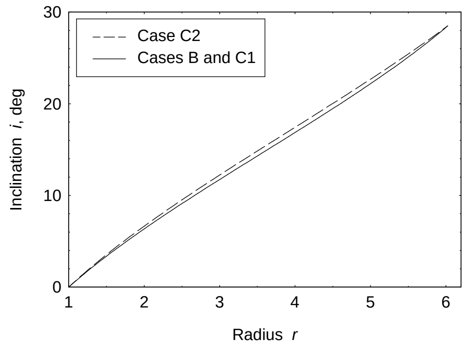
  - variable specific impulse

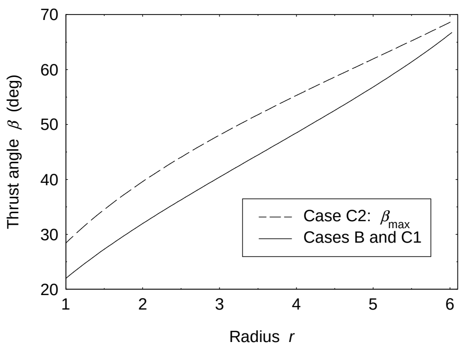
$$H = \lambda_i K_3/(8rK_1) + \lambda_t m/(8P\sqrt{r^3}K_1) - \lambda_m m(8K_1^2 + K_3^2)/(8\sqrt{r^3}K_1)$$

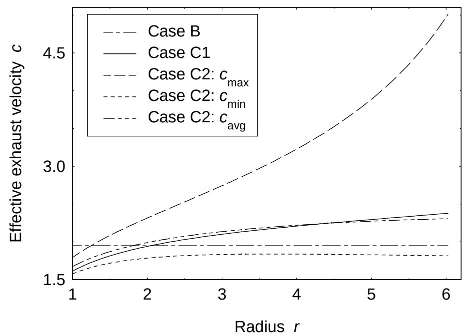
$$K_3 = \frac{\lambda_i\sqrt{r}}{2m\lambda_m} \qquad K_1 = \sqrt{\frac{1}{8}\left(\frac{1}{\lambda_m P} - K_3^2\right)}$$

Nested Indirect Methods: LeoGeo Results









Indirect Methods and Evolutionary Algorithms

- use an EA to determine the unknowns of the MPBVP from the application of OCT
  - solution may be inaccurate
  - usually not effective in complex problems
- use an EA to estimate the unknowns of the MPBVP (that is, to get a tentative solution)
  - requires solution of MPBVP with Newton's method

Future Developments

- applications suited for indirect methods are almost infinite
- possible interesting developments
  - nested optimization of geocentric transfers with eclipses and perturbations
  - developments of general (automatic) techniques to define and correct switching and constraint structure

Conclusions

- savvy application of indirect methods may allow for solution of very complex problems
- several techniques to permit convergence may be devised depending on the peculiarities of the problem
- fundamental relevance of problem formulation
- user's intuition necessary to guide the solution process