Optimal Control for Halo Orbit Missions

Jerry Marsden

Martin Lo and Roby Wilson (JPL)
Wang-Sang Koon and Shane Ross (Caltech)
Linda Petzold and Radu Serban (Santa Barbara)

Control and Dynamical Systems and JPL
California Institute of Technology

CDS 270-1, Lecture 7A, May 8, 2000

marsden@cds.caltech.edu
http://www.cds.caltech.edu/~marsden/

Overview and Comments

- ***Specific Topic:*** Optimal control for halo orbit insertion. (Collaboration with ***Radu Serban, Linda Petzold and Roby Wilson***).
- Reference: Proceedings of the IFAC meeting, Princeton, March, 2000. Paper available at

  http://www.cds.caltech.edu/~marsden/

  (look under papers/papers published/2000).
- ***Related***: Application to energy efficient missions from the Earth to the moon and to the moons of Jupiter. (See Shane's talk).
- This talk gives one of many possible views on how to merge dynamics and optimal control. This is an ***exciting area*** for further exploration!

Optimal Insertion into a Halo Orbit

- ***Halo Orbit Insertion*** goes back to the early days of L_1 halo-orbit missions (eg, Farquhar et al [1980] for the ISEE-3 mission launched in 1978).
- Recall the nature of the halo orbit trajectory:

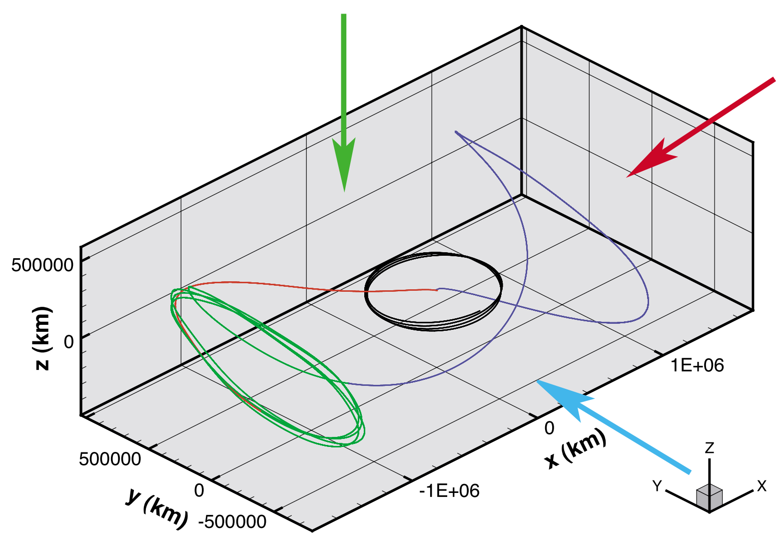


FIGURE 1: The Genesis Trajectory.

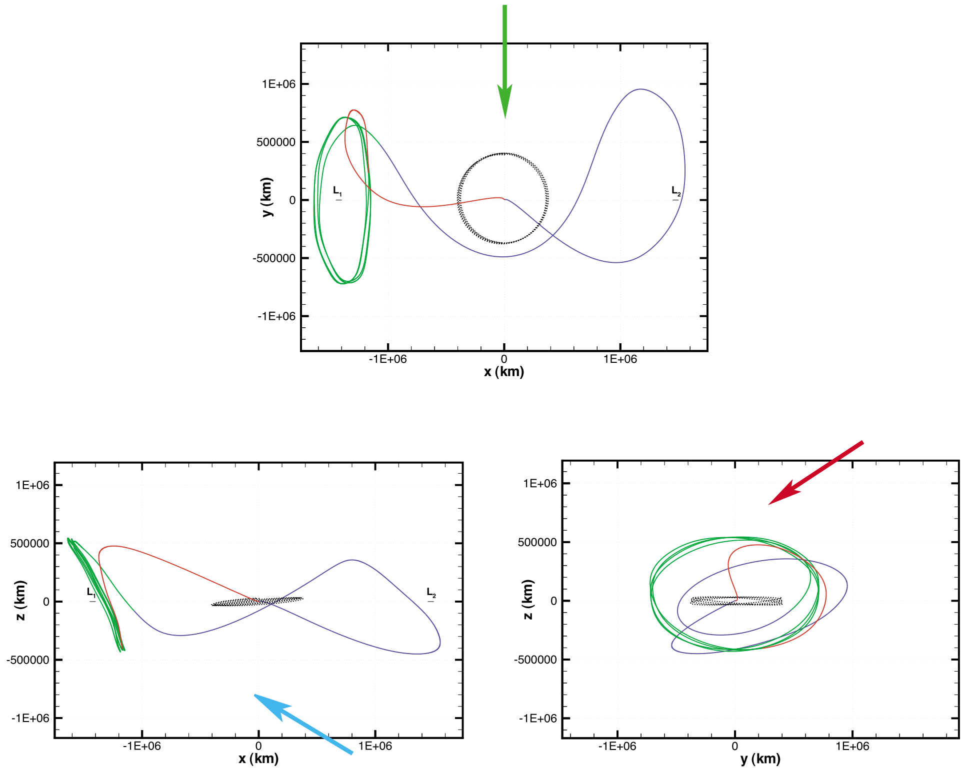


FIGURE 2: Projections of the Genesis Trajectory.

- We study ***optimal control*** in the context of mission design with the aid of dynamical systems and invariant manifolds. In particular, we consider the problem of ***finding optimal burns*** for ***halo orbit insertion*** of ***Genesis***-type missions, although the methods are rather general.
- ***Low thrust*** and ***impulsive burn*** contexts are both important and the techniques can—presumably—handle either case.
- The optimization software COOPT (COntrol–OPTimization) is used to do an optimization of the ***cost function*** (minimizing ΔV) subject to the ***constraint of the equations of motion***. We vary the number of impulses and also consider the effect of delaying the first impulse.
- ***Aside:*** Another technique that is quite interesting involves substituting the controls from the equations of motion to get a ***higher***

order cost function rather than treating the equations of motion as constraints (most mechanics people shudder at treating the equations of motion as constraints); work of Heinzinger, Bloch, Crouch, Milam, etc.

- Of course, each of these methods has an associated Pontryagin theory and Hamilton-Jacobi-Bellman equations (and related methods such as receding horizon control).

- COOPT, which takes a ***brute force numerical approach***, is rather general and sophisticated software for ***optimal control and optimization of systems modeled by differential-algebraic equations*** (DAE), developed by the Computational Science and Engineering Group at University of California Santa Barbara. It has been designed to control and optimize a general class of DAE systems, which may be quite large. It uses multiple shooting and SQP techniques to do the optimization.

- Halo orbits are large three dimensional orbits shaped like the edges of a potato chip. The Y-amplitude of the ***Genesis*** halo orbit, which extends from the X-axis to the maximum Y-value of the orbit, is about 780,000 km. This is bigger than the radius of the Moon's orbit, which is about 380,000 km.
- The computation of halo orbits follows standard nonlinear trajectory computation algorithms based on parallel shooting. Due to the problem sensitivity and the ***instability*** of the halo orbit (albeit with a fairly long time constant in the Sun-Earth system), an ***accurate first guess*** is essential.
- This first guess is provided by a high order analytic expansion of minimum 3rd order using the Lindstedt-Poincaré method. For details of halo orbit computations and general algorithms, see Richard-

son [1980], Llibre, Martinez and Simó [1985], Howell and Pernicka [1988], and Parker and Chua [1989]. This was discussed in the preceding lectures.

- Simo et al, in conjunction with the SOHO mission in the 1980's were the first to study invariant manifolds of the halo orbit.
- ***Stable manifold of the halo orbit***–used to design the transfer trajectory which delivers the Genesis spacecraft from launch to insertion onto the halo orbit (HOI). ***Unstable manifold***–used to design the return trajectory which brings the spacecraft and its samples back to Earth via the ***heteroclinic connection***.
- Expected error due to launch is approximately 7 m/s for a boost of approximately 3200 m/s from a 200 km circular altitude Earth orbit. This error is then optimally corrected using impulsive thrusts. Halo orbit missions are ***very sensitive to launch errors***.

- ***Objective:*** *Find the maneuver times and sizes to minimize fuel consumption for a trajectory starting at Earth and ending on the specified halo orbit around the Lagrange point L_1 of the Sun-Earth system at a position and with a velocity consistent with the HOI time.*

■ **The Transfer to the Halo Orbit** The transfer trajectory is designed using the following procedure.

- A halo orbit $H(t)$ is first selected, where t represents time. The stable manifold of H, denoted W^s, consists of a family of asymptotic trajectories which take infinite time to wind onto H. These asymptotic solutions cannot be found numerically and are impractical for space missions where the transfer time needs to be just a few months.

- However, there is a family of trajectories that lie arbitrarily close to W^s that require just a few months to transfer between Earth and the halo orbit. These trajectories are said to ***shadow*** the stable manifold. It is these shadow trajectories that we can compute and that are extremely useful to the design of the *Genesis* transfer trajectory.

- Recall how to compute an approximation of the ***stable manifold of the halo orbit*** W^s. The basic idea is to linearize the equations of motion about the periodic orbit and then use the monodromy matrix provided by Floquet theory to generate a linear approximation of the stable manifold associated with the halo orbit. The linear approximation, in the form of a state vector, is integrated in the nonlinear equations of motion to produce the approximation of the stable manifold.

- In the case of ***quasiperiodic orbits*** that are not too far from periodic orbits, one approximates the orbit as periodic and the same algorithm is applied to compute approximations of W^s (see Howell, Barden and Lo [1997]; see also Gómez, Masdemont and Simó [1993]). For engineering purposes, at least for space missions, this seems to work well. Recently, a more refined approach based on reduction to the center manifold (or neutrally stable manifold) is provided by Jorba and Masdemont [1999].

- We will assume that the halo orbit, $H(t)$, and the stable manifold $M(t)$ are fixed and provided. Hence we will not dwell further on the theory of their computation which is well covered in the references (see Howell, Barden, and Lo [1997]). Instead, let us turn our attention to the trajectory correction maneuver (TCM) problem.

The TCM Problem

- ***Genesis*** will be launched from a Delta 7326 launch vehicle (L/V) using a Thiakol Star37 motor as the final upper stage. The ***most important error*** introduced by the ***inaccuracies of the launch vehicle*** is the velocity magnitude error. In this case, the expected error is 7 m/s (1 sigma value) relative to a boost of approximately 3200 m/s from a 200 km circular altitude Earth orbit.
- It is typical in space missions to use the magnitude of the ΔV as a measure of the spacecraft performance. The propellant mass is a much less stable quantity as a measure of spacecraft performance, since it is dependent on the spacecraft mass and various other parameters which change frequently as the spacecraft is being built.
- Although a 7 m/s error for a 3200 m/s maneuver may seem rather

small, it actually is considered quite large. Unfortunately, one of the characteristics of halo orbit missions is that, unlike interplanetary mission launches, they are extremely sensitive to launch errors.

- Typical interplanetary launches can correct launch vehicle errors 7 to 14 days after the launch. In contrast, halo orbit missions must generally correct the launch error within the first day after launch, due to energy concerns.
- This critical ***Trajectory Correction Maneuver*** is referred to as ***TCM1***, being the first TCM of any mission. Two clean up maneuvers, TCM2 and TCM3, generally follow TCM1 after a week or more, depending on the situation.
- From the equation for a conic orbit,

$$E = \frac{V^2}{2} - \frac{Gm}{R},$$

where E is Keplerian energy, V is velocity, Gm is the gravitational mass, and R is the position, it can be seen that

$$\delta V = \frac{\delta E}{V},$$

where δV and δE denote the variations in velocity and energy, respectively. In particular, for highly elliptical orbits, V decreases sharply as a function of time past perigee. Hence, the correction maneuver, ΔV, ***grows sharply in inverse proportion to the time from launch***.

- For a large launch vehicle error, which is possible in the case of *Genesis*, the correction maneuver TCM1 can quickly grow beyond the capability of the spacecraft's propulsion system.
- Because of necessary initial spacecraft checkout procedures after launch, which freqently requires up to a week, it is necessary to investigate the ***effect of delays in the first trajectory***

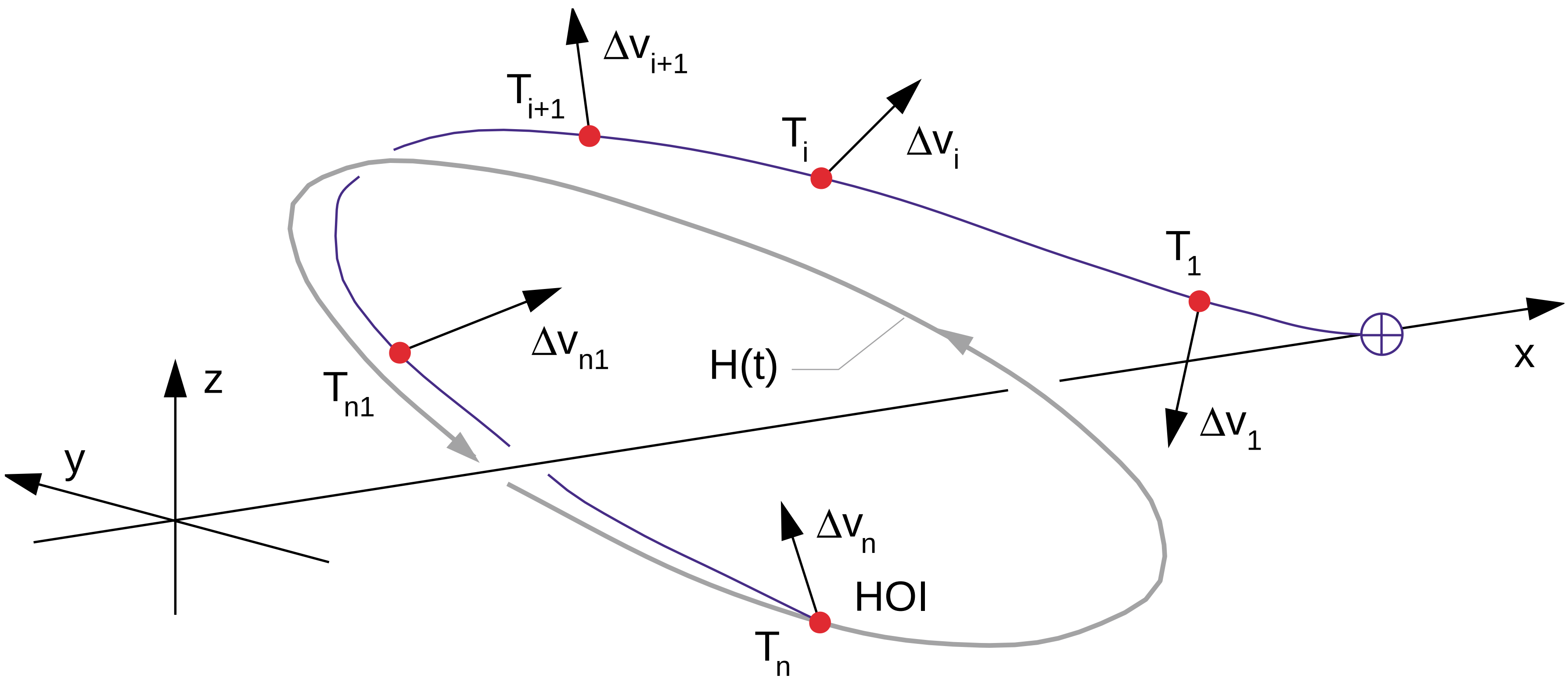
manouver. In fact, it is desirable to delay TCM1 by as long as possible, even at the expense of expenditure of the ΔV budget.

- The Genesis Project prefers that TCM1 be performed at 2 to 7 days after launch, or later if at all possible. The design of the current *Genesis* TCM1 retargets the state after launch back to the nominal HOI state (see Lo, Williams *et al* [1998]). This approach is based on linear analysis and is perfectly adequate if TCM1 is performed within 24 hours after launch. Beyond launch + 24 hours, the correction cost can become prohibitively high. See also Wilson, Howell, and Lo [1999] for another approach to targeting that may be applicable for *Genesis*.

- The desire to increase the time between launch and TCM1 suggests that one use a nonlinear approach, combining dynamical systems theory with optimal control techniques.

- We explore two similar but slightly different approaches and are able to obtain in both cases an optimal maneuver strategy that fits within the *Genesis* ΔV budget of 150 m/s for the transfer portion of the trajectory. These are:
  - ***HOI technique***: use optimal control techniques to retarget the halo orbit with the original nominal trajectory as the initial guess.
  - ***MOI technique***: target the stable manifold.

  Both methods are shown to yield good results.



Δv_{i+1}
T_{i+1}
T_i
Δv_i
T_1
Δv_{n1}
H(t)
x
z
T_{n1}
Δv_1
y
Δv_n
HOI
T_n

$\boxed{\texttt{Halo Insertion Movie}}$

TCM as a Trajectory Planning Problem

- Although different from a dynamical systems perspective, the HOI and MOI problems are very similar once cast as optimization problems. In the HOI problem, a final maneuver (jump in velocity) is allowed at $T_{\mathrm{HOI}} = t_{\max}$, while in the MOI problem, the final maneuver takes place on the stable manifold at $T_{\mathrm{MOI}} < t_{\max}$ and no maneuver is allowed at $T_{\mathrm{HOI}} = t_{\max}$. A halo orbit insertion trajectory design problem can be simply posed as:

  Find the maneuver times and sizes to minimize fuel consumption (ΔV) for a trajectory starting near Earth and ending on the specified halo orbit around the Lagrange point L_1 of the Sun-Earth system at a position and with a velocity consistent with the HOI time.

- The optimization problem as stated has two important features.
  - First, it involves discontinuous controls, since the impulsive maneuvers are represented by jumps in the velocity of the spacecraft. It can be readily reformulated to cast it into the framework required by continuous optimal control algorithms.
  - Secondly, the final halo orbit insertion time T_{HOI}, as well as all intermediate maneuver times, must be included among the optimization parameters. This too requires further reformulation of the dynamical model to capture the influence of these parameters on the solution at a given optimization iteration.
- Next, we discuss the reformulations required to solve the HOI discontinuous control problem; modifications of the following procedure required to solve the MOI problem are straightforward.
- Assume that the evolution of the spacecraft is described by a

generic set of six ODEs

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}),$$

where $\mathbf{x} = (\mathbf{x}^p; \mathbf{x}^v) \in \mathbb{R}^6$ contains both positions ($\mathbf{x}^p$) and velocities ($\mathbf{x}^v$). The dynamical model of the preceding equation can be either the CRTBP or a more complex model that incorporates the influence of the Moon and other planets. We use the CRTBP approximation.

- To deal with the discontinuous nature of the impulsive control maneuvers, the equations of motion (e.o.m.) are solved simultaneously on each interval between two maneuvers. Let the maneuvers $M_1, M_2, ..., M_n$ take place at times T_i, $i = 1, 2, ..., n$ and let $\mathbf{x}_i(t)$, $t \in [T_{i-1}, T_i]$ be the solution of the dynamic equation on the interval $[T_{i-1}, T_i]$. Recall the figure for this:

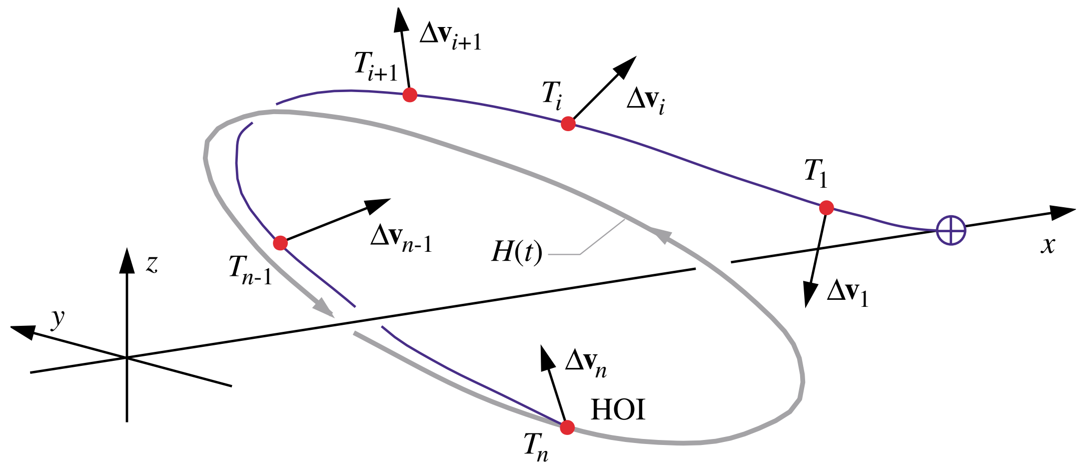


FIGURE 3: Transfer trajectory. Maneuvers take place at times $T_i, i = 1, 2, ..., n$. In the stable manifold insertion problem, there is no maneuver at T_n, i.e. $\Delta \mathbf{v}_n = \mathbf{0}$.

- To capture the influence of the maneuver times on the solution of the e.o.m. and to be able to solve the e.o.m. simultaneously, we scale the time in each interval by the duration $\Delta T_i = T_i - T_{i-1}$. As a consequence, all time derivatives in the e.o.m. are scaled by $1/\Delta T_i$. The dimension of the dynamical system is thus increased to $N_x = 6n$.

- Position continuity constraints are imposed at each maneuver, that is,

$$\mathbf{x}_i^p(T_i) = \mathbf{x}_{i+1}^p(T_i), \qquad i = 1, 2, ..., n-1.$$

- In addition, the final position is forced to lie on the halo orbit (or stable manifold), that is,

$$\mathbf{x}_n^p(T_n) = \mathbf{x}_H^p(T_n),$$

where the halo orbit is parameterized by the HOI time T_n.

- Additional constraints dictate that the first maneuver (TCM1) is delayed by at least a prescribed amount $TCM1_{\min}$, that is,

$$T_1 \geq TCM1_{\min},$$

and that the order of maneuvers is respected,

$$T_{i-1} < T_i < T_{i+1}, \qquad i = 1, 2, ..., n-1.$$

With a cost function defined as some measure of the velocity discontinuities

$$\Delta \mathbf{v}_i = \mathbf{x}_{i+1}^v(T_i) - \mathbf{x}_i^v(T_i),$$

$i = 1, 2, ..., n-1,$

$$\Delta \mathbf{v}_n = \mathbf{x}_H^v(T_n) - \mathbf{x}_n^v(T_n),$$

the optimization problem becomes

$$\min_{T_i, \mathbf{x}_i, \Delta \mathbf{v}_i} C(\Delta \mathbf{v}_i),$$

subject to the constraints given above.

Choice of Cost Function.

- Typically in space missions, the spacecraft performance is measured in terms of the maneuver sizes $\Delta \mathbf{v}_i$. We consider the following two cost functions.

$$C_1(\Delta \mathbf{v}) = \sum_{i=1}^{n} \|\Delta \mathbf{v}_i\|^2$$

and

$$C_2(\Delta \mathbf{v}) = \sum_{i=1}^{n} \|\Delta \mathbf{v}_i\|.$$

- While the second of these may seem physically the most meaningful, as it measures the total sum of the maneuver sizes, such a cost function is ***nondifferentiable*** whenever one of the maneuvers vanishes. In our case, this problem occurs already at the first optimization iteration, as the initial guess transfer trajectory only has a single nonzero maneuver at halo insertion. The first cost function, on the other hand, is differentiable everywhere.

- Although the cost function C_1 is more appropriate for the optimizer, it raises two new problems. Not only is it not as physically meaningful as the cost function C_2, but, in some particular cases, decreasing C_1 may actually lead to increases in C_2.

  To resolve these issues, we use the following ***three-stage optimization sequence***:

  - Starting with the nominal transfer trajectory as initial guess, and allowing initially n maneuvers, we minimize C_1 to obtain a first optimal trajectory, $\mathcal{T}_1^*$.
  - Using $\mathcal{T}_1^*$ as initial guess, we minimize C_2 to obtain $\mathcal{T}_2^*$. It is possible that during this optimization stage some maneuvers can become very small. After each optimization iteration we moni-

tor the feasibility of the iterate and the sizes of all maneuvers. As soon as at least one maneuver decreases under a prescribed threshold (typically 0.1 m/s) at some feasible configuration, we stop the optimization algorithm.

- If necessary, a third optimization stage, using $\mathcal{T}_2^*$ as initial guess and C_2 as cost function is performed with a reduced number of maneuvers $\bar{n}$ (obtained by removing those maneuvers identified as "zero maneuvers" in step 2).

Sensitivity (Perturbation) Analysis

- Let f, h, and g be twice continuously differentiable on $\mathbb{R}^n$ and consider the problem ***P(u,v)***: *minimize* $f(x)$ *subject to equality constraints* $h(x) = u$ *and inequality constraints* $g(x) \leq v$.
- This problem is parameterized by the vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^r$.
- Assume that for $(u, v) = (0, 0)$ this problem has a local minimum x^*, which is regular and which together with its associated Lagrange multiplier vectors λ^* and μ^*, satisfies appropriate second order conditions in the calculus of variations.
- Then there is an open sphere S centered at $(u, v) = (0, 0)$ such that for every $(u, v) \in S$ there is an $x(u, v) \in \mathbb{R}^n$, $\lambda(u, v) \in \mathbb{R}^m$, and $\mu(u, v) \in \mathbb{R}^r$, which are a local minimum and associated Lagrange multipliers of ***P(u,v)***. (Not a big deal by using the implicit function theorem).

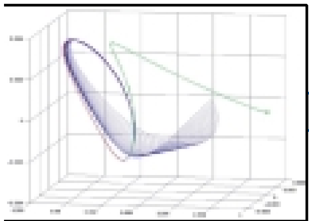
- More important—how to calculate: Let $p(u, v)$ be the ***optimal cost*** parameterized by (u, v):

$$p(u, v) = f(x(u, v)).$$

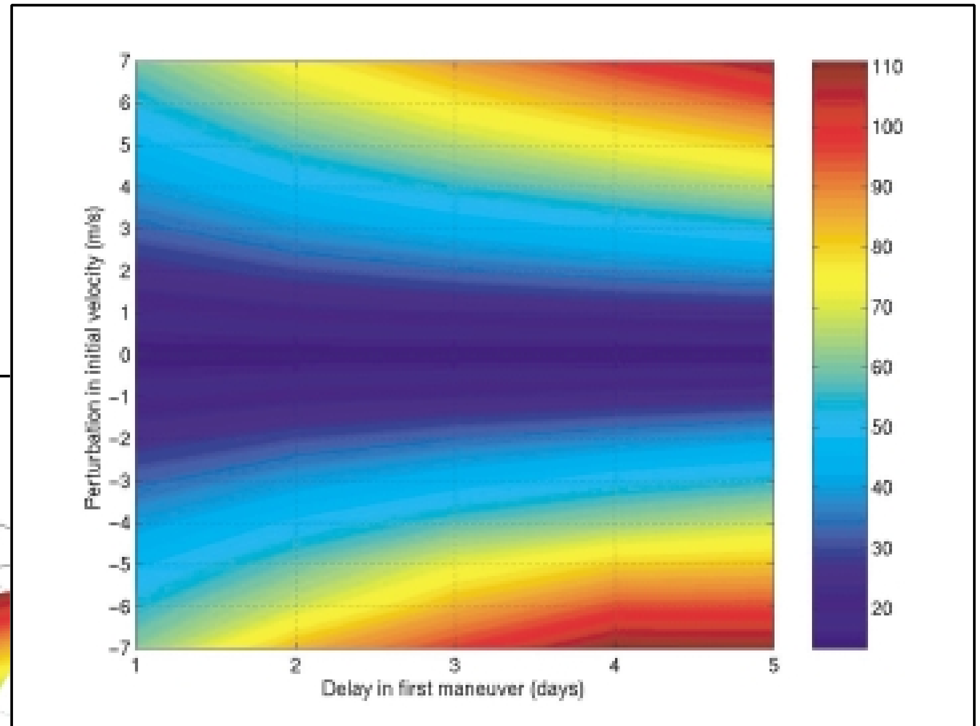
Here is how to calculate how p changes with (u, v):

$$\nabla_u p(u, v) = -\lambda(u, v),$$
$$\nabla_v p(u, v) = -\mu(u, v).$$

- The influence of delaying the first trajectory control maneuver TCM1 is computed from the Lagrange multiplier associated with the constraint of delaying the first maneuver. To evaluate sensitivities of the cost function with respect to perturbations in the launch velocity, we include this perturbation explicitly as an optimization parameter and fix it to some prescribed value through an equality constraint.

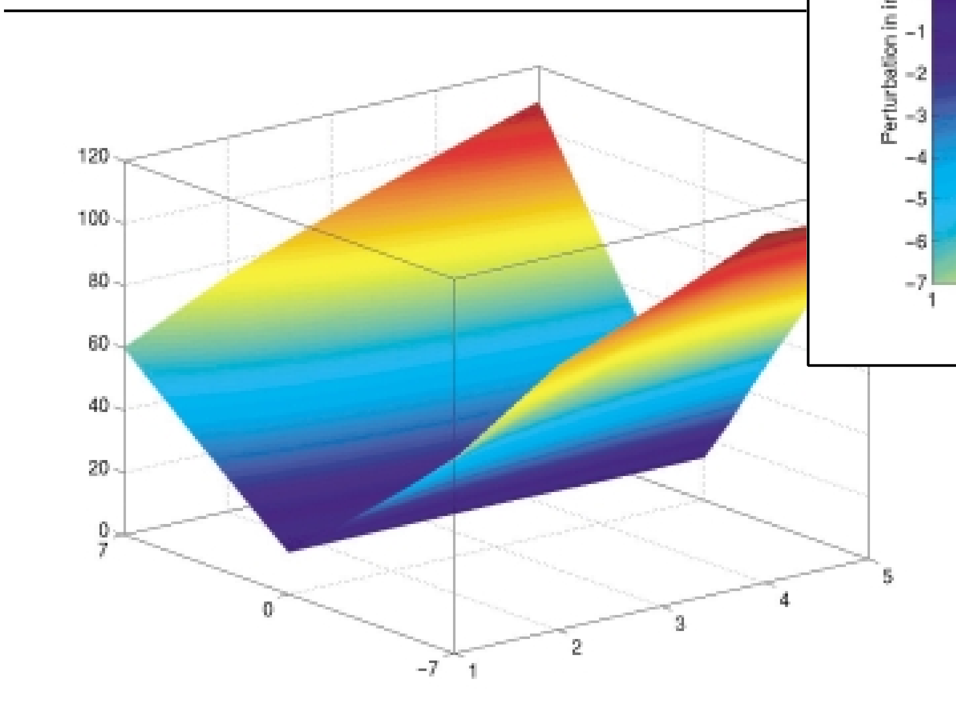


Parametric Study of the Optimal Solution



Number of maneuvers:

- **Unperturbed injection velocity: 1**
- **Perturbed injection velocity: 2**



Influence of:

- **Delay in TCM1**
- **Perturbation in launching velocity**

Optimal solutions found for all cases

More about the Software.

- COOPT is a software package for optimal control and optimization of systems modeled by differential-algebraic equations (DAE) (see Brenan, Campbell and Petzold [1995]), developed by the Computational Science and Engineering Group at the University of California, Santa Barbara.
- Designed to control and optimize a general class of ***DAE systems***, which may be quite large.
- Consider a DAE system:

$$\begin{aligned} \mathbf{F}(t, \mathbf{x}, \mathbf{x}', \mathbf{p}, \mathbf{u}(t)) &= \mathbf{0}, \\ \mathbf{x}(t_1, \mathbf{r}) &= \mathbf{x}_1(\mathbf{r}), \end{aligned} \tag{1}$$

- The ***control parameters*** $\mathbf{p}$ and $\mathbf{r}$ and the vector-valued control function $\mathbf{u}(t)$ must be determined such that an objective function of the form

$$\int_{t_1}^{t_{\max}} \Psi(t, \mathbf{x}(t), \mathbf{p}, \mathbf{u}(t))\, dt + \Theta(t_{\max}, \mathbf{x}(t_{\max}), \mathbf{p}, \mathbf{r}),$$

is minimized and some additional ***inequality constraints***

$$\mathbf{g}(t, \mathbf{x}(t), \mathbf{p}, \mathbf{u}(t)) \geq 0,$$

are satisfied. The optimal control function $\mathbf{u}^*(t)$ is assumed to be continuous.

- Approximate $\mathbf{u}(t)$ using piecewise polynomials on $[t_1, t_{\max}]$, where their coefficients are determined by the optimization.
- For simplicity, assume that the vector $\mathbf{p}$ contains both the parameters and these $\mathbf{u}(t)$ coefficients. Also, suppose that the initial states are fixed and therefore discard the dependency of $\mathbf{x}_1$ on $\mathbf{r}$. Hence,

we consider

$$\mathbf{F}(t, \mathbf{x}, \mathbf{x}', \mathbf{p}) = 0, \qquad \mathbf{x}(t_1) = \mathbf{x}_1,$$

$$\text{minimize} \quad \int_{t_1}^{t_{\max}} \psi(t, \mathbf{x}(t), \mathbf{p})\, dt + \Theta(t_{\max}, \mathbf{x}(t_{\max}), \mathbf{p})$$

$$\mathbf{g}(t, \mathbf{x}(t), \mathbf{p}) \geq 0.$$

- COOPT implements the single shooting method and a modified version of the multiple shooting method, both of which allow the use of adaptive DAE software.
- In the multiple shooting method, the time interval $[t_1, t_{\max}]$ is divided into subintervals $[t_i, t_{i+1}]$ $(i = 1, \ldots, N)$, and the differential equations $\mathbf{F}(t, \mathbf{x}, \mathbf{x}', \mathbf{p}) = 0$ are solved over each subinterval, where additional intermediate variables $\mathbf{X}_i$ are introduced. On each subinterval we denote the solution at time t with initial value $\mathbf{X}_i$ at t_i by $\mathbf{x}(t, t_i, \mathbf{X}_i, \mathbf{p})$.

- Continuity between subintervals in the multiple shooting method is achieved via the continuity constraints

$$\mathbf{C}_1^i(\mathbf{X}_{i+1}, \mathbf{X}_i, \mathbf{p}) \equiv \mathbf{X}_{i+1} - \mathbf{x}(t_{i+1}, t_i, \mathbf{X}_i, \mathbf{p}) = \mathbf{0}.$$

- Additional constraints are required at the boundaries of the shooting intervals

$$\mathbf{C}_2^i(\mathbf{X}_i, \mathbf{p}) \equiv \mathbf{g}(t_i, \mathbf{X}_i, \mathbf{p}) \geq \mathbf{0}.$$

- Write

$$\Phi(t) = \int_{t_1}^{t} \psi(\tau, \mathbf{x}(\tau), \mathbf{p})\, d\tau,$$

so the discretized optimal control problem becomes

$$\min_{\mathbf{X}_2, \mathbf{X}_3, \dots, \mathbf{p}} \Phi(t_{\max}) + \Theta(t_{\max}),$$

subject to the constraints

$$\begin{aligned} \mathbf{C}_1^i(\mathbf{X}_{i+1}, \mathbf{X}_i, \mathbf{p}) &= \mathbf{0}, \\ \mathbf{C}_2^i(\mathbf{X}_i, \mathbf{p}) &\geq \mathbf{0}. \end{aligned}$$

- This problem can be solved by an optimization algorithm; we use SNOPT (Gill, Murray and Saunders [1997]), which incorporates a ***sequential quadratic programming (SQP)*** method (Gill, Murray and Wright [1981]).
- The SQP methods require a gradient and Jacobian matrix that are the derivatives of the objective function and constraints with respect to the optimization variables. One computes these derivatives via DAE sensitivity software DASPK3.0 (Li and Petzold [1999]).
- The sensitivity equations to be solved by DASPK3.0 are generated via the automatic differentiation software ADIFOR (Bischof, Carle, Corliss, Griewank and Hovland [1997]).

- The multiple-shooting strategy can work very well for small-to-moderate size ODE systems, and has an additional advantage that it is inherently parallel. However, for large-scale ODE and DAE systems there is a problem because the ***computational complexity*** grows rapidly with the dimension of the ODE system. COOPT implements a highly efficient modified multiple shooting method (Petzold, Rosen, Gill, Jay and Park [1992] and Serban [1999]) which reduces the computational complexity to that of single shooting for large-scale problems.
- We have found it sufficient to use single shooting for the trajectory design problems considered.

End

Internet site:

http://www.cds.caltech.edu/~marsden/

Email:

marsden@cds.caltech.edu