3 Solving Two-Point Boundary Value Problems Using Generating Functions: Theory and Applications to Astrodynamics

VINCENT M. GUIBOUT* AND DANIEL J. SCHEERES†
* *MBDA France*
† *University of Michigan, Ann Arbor, Michigan*

Contents

3.1 Introduction

Two-point boundary value problems have a central place in the field of astrodynamics. In general, most of the hard problems in this field revolve around solving such problems. Examples include the targeting problem for mission design, the computation of periodic orbits for the analysis of systems, and the solution of optimal control problems.

In this chapter, a new methodology for solving two-point boundary value problems in phase space for Hamiltonian systems is presented. Using the Hamilton–Jacobi theory in conjunction with the canonical transformation induced by the phase flow, we show that the generating functions for this transformation solve any two-point boundary value problem in phase space. Properties of the generating functions are exposed, we especially emphasize multiple solutions, singularities, relations with the state transition matrix and symmetries. Next, we show that using Hamilton's principal function we are also able to solve two-point boundary value problems, nevertheless both methodologies have fundamental differences that we explore. Then we present and study an algorithm to compute the generating functions specialized to such two-point boundary value problems. This algorithm is able to compute the generating functions for a large class of practical two-point boundary value problems. Specifically, the algorithm naturally avoids singularities and allows one to specify the initial conditions as a function of a parameter. Finally, we

present applications of this method to two difficult problems of astrodynamics to show its generality, computation of periodic orbits and solution of an optimal control problem.

One of the most famous two-point boundary value problems in astrodynamics is Lambert's problem, which consists of finding a trajectory in the two-body problem which goes through two given points in a given lapse of time. Even though the two-body problem is integrable, no explicit solution to this problem exists. Many other two-point boundary value problems in astrodynamics can also be couched within a Hamiltonian formalism. These include all problems of orbital motion, excepting the effect of atmospheric drag on an orbiter, and also include all instances of optimal control problems. It is important to note that optimal control problems can all be recast into Hamiltonian systems via the necessary conditions. Thus, even non-conservative dynamical systems can be treated using the formalism we develop here when dealing with their optimal control. In the following, we do not make a distinction between whether a Hamiltonian dynamical system arises out of mechanics or out of optimal control, as the basic results of the Hamilton–Jacobi theory that we use apply to both.

For a general Hamiltonian dynamical system, a two-point boundary value problem is generally solved using iterative techniques such as shooting and relaxation methods. The shooting method [5, 29] consists of choosing values for all of the dependent variables at one boundary. These values must be consistent with any boundary conditions for that boundary, but otherwise are initially guessed "randomly". After integration of the differential equations, we in general find discrepancies between the desired boundary values at the other boundary. Then, we adjust the initial guess to reduce these discrepancies and reiterate this procedure again. The method provides a systematic approach to solving boundary value problems, but suffers several inherent limitations. As summarized by Bryson and Ho ([7] p. 214),

> The main difficulty with this method is getting started; i.e., finding a first estimate of the unspecified conditions at one end that produces a solution reasonably close to the specified conditions at the other end. The reason for this peculiar difficulty is that the extremal solutions are often very sensitive to small changes in the unspecified boundary conditions.

To get rid of the sensitivity to small changes in initial guesses, techniques such as the multiple shooting method [21] were developed. They consist of breaking the time domain into segments and solving a boundary value problem on each of these segments. In this manner, non-linear effects are limited over each segment, but on the other hand the size of the problem is increased considerably. However, the choice of the initial conditions still remains as the main hurdle to successfully apply shooting methods to general problems.

Relaxation methods [30] use a different approach. The differential equations are replaced by finite-difference equations on a mesh of points that covers the range of the integration. A trial solution consists of values for the dependent variables at each mesh point, not satisfying the desired finite-difference equations, nor necessarily even satisfying the required boundary conditions. The iteration, now called relaxation, consists of adjusting all the values on the mesh so as to bring them into successively closer agreement with the finite-difference equations and simultaneously with the boundary conditions. In general, relaxation works better than shooting when the boundary conditions are especially

delicate or subtle. However, if the solution is highly oscillatory then many grid points are required for accurate representation. Also, the number and positions of the required mesh points are not known a priori and must be adjusted manually for each problem. In addition, if solutions to the differential equations develop singularities, attempts to refine the mesh to improve accuracy may fail.

With the advent of computers, these two methods are able to solve most of the two-point boundary value problems. They may require substantial time to find an appropriate initial guess and/or computer memory to refine the mesh, but they generally succeed. However, there are problems for which these methods reach their limits. For instance, the design of space missions involving several spacecraft in formation requires one to solve a large number of boundary value problems for which the boundary conditions may in turn depend on parameters. Most research in this area to date has focused on the solution of such boundary value problems for linearized motion. Extension of these techniques to non-linear dynamics is much more difficult, with most progress being limited to non-linear expansions of the two-body problem with minimal perturbations added, if any [2, 24]. However, for precise control of formations over long periods of time or over large distances, it is crucial that non-linear solutions to these problems be available. For example, to reconfigure a formation of N spacecraft, there are $N!$ possibilities in general, i.e., $N!$ boundary value problems need to be solved [31]. Similarly, suppose that we plan to reconfigure a spacecraft formation to achieve a specific goal, such as for an interferometry mission where they may be required to be equally spaced on a circle perpendicular to the line of sight they observe. In that case, the final positions are specified in terms of the angle that indicates the position of the spacecraft on the circle. In order to find the value of the angle that minimizes fuel expenditure, infinitely many boundary value problems may need to be solved, if evaluated in a formal sense. As a result, the algorithms mentioned above are no longer appropriate as they require excessive computation and time for modeling non-linear situations. To address these complex problems arising in spacecraft formation design, Guibout and Scheeres [16–18] developed a novel approach for solving boundary value problems which outperforms traditional methods for spacecraft formation design. In the present contribution we generalize their method and study its properties. We first prove that it allows us to formally solve a non-linear two-point boundary value problem at a cost of a single function evaluation once generating functions are known.

In addition, properties of the generating functions are studied. In particular, for linear systems we show that generating functions and state transition matrices are closely related. The state transition matrix allows one to predict singularities of the generating functions whereas the generating functions provide information on the structure of the state transition matrix. This relationship also allows us to recover and extend some results on the perturbation matrices introduced by Battin in Ref. [4]. For non-linear systems, generating functions may also develop singularities (called caustics). Using the Legendre transformation, we propose a technique to study the geometry of these caustics. We illustrate our method with the study of the singularities of the F_1 generating function in the Hill three-body problem and relate the existence of singularities to the presence of multiple solutions to boundary value problems. Furthermore, we discuss Hamilton's principal function, a function *similar* to the generating functions that also solves two-point boundary value problems. We highlight the differences between Hamilton's function and

generating functions and justify our choice of focusing on generating functions. Next we outline and evaluate a method for constructing solutions for the generating functions. Finally, we present direct applications of this theory that have been identified in previous papers. These applications are presented to illustrate the application of our method to different problems in astrodynamics. We only present the main ideas and refer to previous papers for details.

3.2 Solving two-point boundary value problems

In this section, we review the principle of least action for Hamiltonian systems and derive the Hamilton–Jacobi equation. Local existence of generating functions is proved, but we underline that we do not study global properties. In general, we do not know a priori if the generating functions will be defined for all time, and in most of the cases we found that they develop singularities. We refer the reader to Refs. [1, 3, 9, 11, 14, 22, 23] for more details on local Hamilton–Jacobi theory, Refs. [1, 3, 23] for global theory and Refs. [1, 3, 8] and Section 3.2.4 of this chapter for the study of singularities.

3.2.1 The Hamilton–Jacobi theory

Let $(\mathcal{P}, \omega, X_H)$ be a Hamiltonian system with n degrees of freedom, and $H\colon \mathcal{P} \times \mathbb{R} \to \mathbb{R}$ the Hamiltonian function. We consider the symplectic charts whose existence is guaranteed by Darboux's theorem [6]. We denote the component functions (also called canonical coordinates) by (q_i, p_i) so that, in the symplectic chart, ω is locally written as:

$$\omega = \sum_{i=1}^{n} \mathrm{d}q_i \wedge \mathrm{d}p_i \,.$$

In the extended phase space $\mathcal{P} \times \mathbb{R}$, we consider an integral curve of the vector field X_H connecting the points (q_0, p_0, t_0) and (q_1, p_1, t_1). The principle of least action [23] reads:

Theorem 3.2.1. ***(The principle of least action in phase space)*** *Critical points of* $\int_0^1 p\mathrm{d}q - H\mathrm{d}t$ *in the class of curves* γ *whose ends lie in the n-dimensional subspaces* $(t = t_0, q = q_0)$ *and* $(t = t_1, q = q_1)$ *correspond to trajectories of the Hamiltonian system whose ends are* q_0 *at* t_0 *and* q_1 *at* t_1.

Proof. We proceed to the computation of the variation.

$$\begin{aligned}\delta \int_\gamma (p\dot{q} - H)\mathrm{d}t &= \int_\gamma \left(\dot{q}\delta p + p\delta\dot{q} - \frac{\partial H}{\partial q}\delta q - \frac{\partial H}{\partial p}\delta p \right) \mathrm{d}t \\ &= [p\delta q]_0^1 + \int_\gamma \left[\left(\dot{q} - \frac{\partial H}{\partial p} \right) \delta p - \left(\dot{p} + \frac{\partial H}{\partial q} \right) \delta q \right] \mathrm{d}t \end{aligned} \tag{3.1}$$

Therefore, since the variation vanishes at the end points, the integral curves of the Hamiltonian vector field are the only extrema. □

Now let $(\mathcal{P}_1, \omega_1)$ and $(\mathcal{P}_2, \omega_2)$ be symplectic manifolds,

Definition 3.2.1. A smooth map $f\colon \mathcal{P}_1 \times \mathbb{R} \to \mathcal{P}_2 \times \mathbb{R}$ is a canonical transformation from (q, p, t) to (Q, P, T) if and only if:
1. f is a diffeomorphism,
2. f preserves the time, i.e., there exists a function g_t such that $f(x, t) = (g_t(x), t)$ (from here we assume that $t = T$),
3. Critical points of $\int_{t_0}^{t_1}(\langle P, \dot{Q}\rangle - K(Q, P, t))\mathrm{d}t$ correspond to trajectories of the Hamiltonian system, where $K(Q, P, t) = H(q(Q, P, t), p(Q, P, t), t)$ is the Hamiltonian function expressed in the new set of coordinates.

Consider a canonical transformation between two sets of coordinates in the phase space $f\colon (q_i, p_i, t) \mapsto (Q_i, P_i, t)$ and let $H(q, p, t)$ and $K(Q, P, t)$ be the Hamiltonian functions of the same system expressed in different sets of coordinates. From Def. 3.2.1, trajectories correspond to critical points of $\int_{t_0}^{t_1}(\langle P, \dot{Q}\rangle - K(Q, P, t))\mathrm{d}t$. Therefore, they are integral of:

$$\begin{cases} \dot{Q}_i = \dfrac{\partial K}{\partial P_i}, \\ \dot{P}_i = -\dfrac{\partial K}{\partial Q_i}, \end{cases} \tag{3.2}$$

i.e., f preserves the canonical form of Hamilton's equations.

Conversely, suppose that f is a coordinate transformation that preserves the canonical form of Hamilton's equations and leaves the time invariant. Let $K(Q, P, t)$ be the Hamiltonian in the new set of coordinates, then from the modified Hamilton principle (Thm. 3.2.1), critical points of

$$\int_{t_0}^{t_1} \big(\langle P, \dot{Q}\rangle - K(Q, P, t)\big)\,\mathrm{d}t$$

correspond to trajectories of the system. Thus, f is a canonical map. These last two remarks are summarized in the following lemma:

Lemma 3.2.2. *The third item in Def. 3.2.1 is equivalent to:*
(4)—*f preserves the canonical form of Hamilton's equations and the new Hamiltonian function is $K(Q, P, t)$.*

Remark 3.2.1. The definition we give is different from the one given in many textbooks but in agreement with Arnold [3], Abraham and Marsden [1], and Marsden and Ratiu [23]. Often the third item reduces to:
(5)—f preserves the canonical form of Hamilton's equations.

We consider again a canonical transformation $f\colon (q_i, p_i, t) \mapsto (Q_i, P_i, t)$ and a Hamiltonian system defined by H. Along trajectories, we have by definition:

$$\delta \int_{t_0}^{t_1} \left(\sum_{i=1}^{n} p_i \dot{q}_i - H(q, p, t) \right) \mathrm{d}t = 0, \tag{3.3}$$

$$\delta \int_{t_0}^{t_1} \left(\sum_{i=1}^{n} P_i \dot{Q}_i - K(Q, P, t) \right) \mathrm{d}t = 0\,. \tag{3.4}$$

From Eqs. 3.3 to 3.4, we conclude that the integrands of the two integrals differ at most by a total time derivative of an arbitrary function F:

$$\sum_{i=1}^{n} p_i \mathrm{d}q_i - H\mathrm{d}t = \sum_{j=1}^{n} P_j \mathrm{d}Q_j - K\mathrm{d}t + \mathrm{d}F. \tag{3.5}$$

Such a function is called a generating function for the canonical transformation f. It is, a priori, a function of both the old and the new variables and time. The two sets of coordinates being connected by the $2n$ equations, namely, $f(q, p, t) = (Q, P, t)$, F can be reduced to a function of $2n+1$ variables among the $4n+1$. Hence, we can define 4^n generating functions that have n variables in $\mathcal{P}_1$ and n in $\mathcal{P}_2$. Among these are the four kinds defined by Goldstein [9]:

$$F_1(q_1, \ldots, q_n, Q_1, \ldots, Q_n, t), \quad F_2(q_1, \ldots, q_n, P_1, \ldots, P_n, t),$$
$$F_3(p_1, \ldots, p_n, Q_1, \ldots, Q_n, t), \quad F_4(p_1, \ldots, p_n, P_1, \ldots, P_n, t).$$

Let us first consider the generating function $F_1(q, Q, t)$. The total time derivative of F_1 reads:

$$\mathrm{d}F_1(q, Q, t) = \sum_{i=1}^{n} \frac{\partial F_1}{\partial q_i}\mathrm{d}q_i + \sum_{j=1}^{n} \frac{\partial F_1}{\partial Q_i}\mathrm{d}Q_i + \frac{\partial F_1}{\partial t}\mathrm{d}t. \tag{3.6}$$

Hence Eq. (3.5) yields:

$$\sum_{i=1}^{n} \left(p_i - \frac{\partial F_1}{\partial q_i}\right) \mathrm{d}q_i - H\mathrm{d}t = \sum_{j=1}^{n} \left(P_j + \frac{\partial F_1}{\partial Q_j}\right) \mathrm{d}Q_j - K\mathrm{d}t + \frac{\partial F_1}{\partial t}\mathrm{d}t. \tag{3.7}$$

Assume that (q, Q, t) is a set of independent variables. Then Eq. (3.7) is equivalent to:

$$p_i = \frac{\partial F_1}{\partial q_i}(q, Q, t), \tag{3.8}$$

$$P_i = -\frac{\partial F_1}{\partial Q_i}(q, Q, t), \tag{3.9}$$

$$K\left(Q, -\frac{\partial F_1}{\partial Q}, t\right) = H\left(q, \frac{\partial F_1}{\partial q}, t\right) + \frac{\partial F_1}{\partial t}. \tag{3.10}$$

If (q, Q) is not a set of independent variables, we say that F_1 is singular.

Let us consider more general generating functions. Let $(i_1, \ldots, i_s)(i_{s+1}, \ldots, i_n)$ and $(k_1, \ldots, k_r)\,(k_{r+1}, \ldots, k_n)$ be two partitions of the set $(1, \ldots, n)$ into two non-intersecting parts such that $i_1 < \cdots < i_s$, $i_{s+1} < \cdots < i_n$, $k_1 < \cdots < k_r$, and $k_{r+1} < \cdots < k_n$ and define $I_s = (i_1, \ldots, i_s)$, $\bar{I}_s = (i_{s+1}, \ldots, i_n)$, $K_r = (k_1, \ldots, k_r)$, and $\bar{K}_r = (k_{r+1}, \ldots, k_n)$. If

$$(q_{I_s}, p_{\bar{I}_s}, Q_{K_r}, P_{\bar{K}_r}) = (q_{i_1}, \ldots, q_{i_s}, p_{i_{s+1}}, \ldots, p_{i_n}, Q_{k_1}, \ldots, Q_{k_r}, P_{k_{r+1}}, \ldots, P_{k_n})$$

are independent variables, then we can define the generating function F_{I_s,K_r}:

$$F_{I_s,K_r}(q_{I_s}, p_{\bar{I}_s}, Q_{K_r}, P_{\bar{K}_r}, t) = F(q_{i_1}, \ldots, q_{i_s}, p_{i_{s+1}}, \ldots, p_{i_n}, Q_{k_1}, \ldots, Q_{k_r}, P_{k_{r+1}}, \ldots, P_{k_n}, t). \tag{3.11}$$

Expanding $\mathrm{d}F_{I_s,K_r}$ yields:

$$\mathrm{d}F_{I_s,K_r} = \sum_{a=1}^{p} \frac{\partial F_{I_s,K_r}}{\partial q_{i_a}} \mathrm{d}q_{i_a} + \sum_{a=p+1}^{n} \frac{\partial F_{I_s,K_r}}{\partial p_{i_a}} \mathrm{d}p_{i_a} + \sum_{a=1}^{r} \frac{\partial F_{I_s,K_r}}{\partial Q_{k_a}} \mathrm{d}Q_{k_a} + \sum_{a=r+1}^{n} \frac{\partial F_{I_s,K_r}}{\partial P_{k_a}} \mathrm{d}P_{k_a} + \frac{\partial F_{I_s,K_r}}{\partial t} \mathrm{d}t \tag{3.12}$$

and rewriting Eq. (3.5) as a function of the linearly independent variables leads to:

$$\sum_{a=1}^{p} p_{i_a} \mathrm{d}q_{i_a} - \sum_{a=p+1}^{n} q_{i_a} \mathrm{d}p_{i_a} - H\mathrm{d}t = \sum_{a=1}^{r} P_{k_a} \mathrm{d}Q_{k_a} - \sum_{a=r+1}^{n} Q_{k_a} \mathrm{d}P_{k_a} - K\mathrm{d}t + \mathrm{d}F_{I_s,K_r}, \tag{3.13}$$

where

$$F_{I_s,K_r} = F_1 + \sum_{a=r+1}^{n} Q_{k_a} P_{k_a} - \sum_{a=p+1}^{n} q_{i_a} p_{i_a}. \tag{3.14}$$

Eq. (3.14) is often referred to as the *Legendre transformation*, it allows one to transform one generating function into another.

We then substitute Eq. (3.12) into Eq. (3.13):

$$\sum_{a=1}^{r} \left(P_{k_a} + \frac{\partial F_{I_s,K_r}}{\partial Q_{k_a}} \right) \mathrm{d}Q_{k_a} + \sum_{a=r+1}^{n} \left(\frac{\partial F_{I_s,K_r}}{\partial P_{k_a}} - Q_{k_a} \right) \mathrm{d}P_{k_a} - K\mathrm{d}t + \frac{\partial F_{I_s,K_r}}{\partial t} \mathrm{d}t = \sum_{a=1}^{p} \left(p_{i_a} - \frac{\partial F_{I_s,K_r}}{\partial q_{i_a}} \right) \mathrm{d}q_{i_a} - \sum_{a=p+1}^{n} \left(q_{i_a} + \frac{\partial F_{I_s,K_r}}{\partial p_{i_a}} \right) \mathrm{d}p_{i_a} - H\mathrm{d}t, \tag{3.15}$$

and obtain the set of equations that characterizes F_{I_s,K_r}:

$$p_{I_s} = \frac{\partial F_{I_s,K_r}}{\partial q_{I_s}}(q_{I_p}, p_{\bar{I}_p}, Q_{Kr}, P_{\bar{K}_r}, t), \tag{3.16}$$

$$q_{\bar{I}_s} = -\frac{\partial F_{I_s,K_r}}{\partial q_{\bar{I}_s}}(q_{I_p}, p_{\bar{I}_p}, Q_{Kr}, P_{\bar{K}_r}, t), \tag{3.17}$$

$$P_{K_r} = -\frac{\partial F_{I_s,K_r}}{\partial Q_{K_r}}(q_{I_p}, p_{\bar{I}_p}, Q_{Kr}, P_{\bar{K}_r}, t), \tag{3.18}$$

$$Q_{\bar{K}_r} = \frac{\partial F_{I_s,K_r}}{\partial P_{\bar{K}_r}}(q_{I_p}, p_{\bar{I}_p}, Q_{Kr}, P_{\bar{K}_r}, t), \tag{3.19}$$

$$K\left(Q_{K_r}, \frac{\partial F_{I_s,K_r}}{\partial P_{\bar{K}_r}}, -\frac{\partial F_{I_s,K_r}}{Q_{K_r}}, P_{\bar{K}_r}, t \right) = H\left(q_{I_s}, -\frac{\partial F_{I_s,K_r}}{\partial p_{\bar{I}_s}}, \frac{\partial F_{I_s,K_r}}{\partial q_{I_s}}, p_{\bar{I}_s}, t \right) + \frac{\partial F_{I_s,K_r}}{\partial t}. \tag{3.20}$$

For a generating function to be well-defined, we need to make the assumption that its variables are linearly independent. Later we see that this hypothesis is often not satisfied. The following property grants us that at least one of the generating function is well-defined at every instant (the proof of this result is given by Arnold [3]).

Proposition 3.2.3. *Let $f : \mathcal{P}_1 \times \mathbb{R} \to \mathcal{P}_2 \times \mathbb{R}$ be a canonical transformation. Using the above notations, there exist at least two partitions I_s and K_r such that $(q_{I_p}, p_{\bar{I}_p}, Q_{Kr}, P_{\bar{K}_r}, t)$ are linearly independent.*

3.2.2 The phase flow and its generating functions

The Hamilton–Jacobi theory has found many applications over the years but was first used to integrate the equations of motion of integrable Hamiltonian systems [10, 11]. This approach consists of finding a canonical map that transforms the system into an easily integrable one. Once the system is reduced to a trivial one, integration of the equations of motion are easily carried out. However, the search for such a map remains difficult and this aspect limits the use of the Hamilton–Jacobi theory in practice. Instead, in the present research we focus on a single transformation, the one induced by the phase flow that maps the system to its initial state. Under this transformation, the system is in equilibrium, every point in phase space is an equilibrium point. In general, we cannot compute this transformation (if we were able to find this transformation, it would mean that we could integrate the equations of motion) and so we focus on the generating functions that generate this transformation. In particular, we prove that they solve two-point boundary value problems.

Consider a Hamiltonian system and let Φ_t be its flow:

$$\begin{aligned} &\Phi_t : P \to P \\ &(q_0, p_0) \mapsto (\Phi_t^1(q_0, p_0) = q(q_0, p_0, t), \Phi_t^2(q_0, p_0) = p(q_0, p_0, t)) \,. \end{aligned} \tag{3.21}$$

Φ_t induces a transformation ϕ on $\mathcal{P} \times \mathbb{R}$ as follows:

$$\phi : (q_0, p_0, t) \mapsto (\Phi_t(q_0, p_0), t) \,. \tag{3.22}$$

ϕ^{-1} transforms the state of the system at time t to its state at the initial time while preserving the time. Let us now prove that ϕ, and a fortiori ϕ^{-1}, are canonical transformations.

Proposition 3.2.4. *The transformation ϕ induced by the phase flow is canonical.*

Proof. From the theory of differential equations[1], ϕ is an isomorphism. Moreover, the solution ϕ is by definition a diffeomorphism mapping from a symplectic space to a symplectic space, preserving the time, and trivially preserving the Hamiltonian. Thus, by Def. 3.2.1, ϕ is canonical. □

[1] Uniqueness of solutions of ordinary differential equations.

The inverse solution ϕ^{-1} maps the Hamiltonian system to equilibrium, i.e., its conditions at an initial epoch which are constant for all time along that trajectory. Therefore, the associated generating functions, F_{I_s,K_r}, verify the Hamilton–Jacobi equation (Eq. (3.20)). In addition, they must also verify Eqs. (3.16)–(3.19), where (Q, P) now denotes the initial state (q_0, p_0) and K is a constant that can be chosen to be 0:

$$p_{I_s} = \frac{\partial F_{I_s,K_r}}{\partial q_{I_s}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t), \tag{3.23}$$

$$q_{\bar{I}_s} = -\frac{\partial F_{I_s,K_r}}{\partial p_{\bar{I}_s}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t), \tag{3.24}$$

$$p_{0_{K_r}} = -\frac{\partial F_{I_s,K_r}}{\partial q_{0_{K_r}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t), \tag{3.25}$$

$$q_{0_{\bar{K}_r}} = \frac{\partial F_{I_s,K_r}}{\partial p_{0_{\bar{K}_r}}}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t), \tag{3.26}$$

$$0 = H\left(q_{I_s}, -\frac{\partial F_{I_s,K_r}}{\partial p_{\bar{I}_s}}, \frac{\partial F_{I_s,K_r}}{\partial q_{I_s}}, p_{\bar{I}_s}, t\right) + \frac{\partial F_{I_s,K_r}}{\partial t}. \tag{3.27}$$

The last equation is often referred to as the Hamilton–Jacobi equation. For the case where the partitions are $(1, \ldots, n)()$ and $(1, \ldots, n)()$ (i.e., $s = n$ and $r = n$), we recover the generating function F_1, which now verifies the following equations (note that the subscripts are suppressed in the following):

$$p = \frac{\partial F_1}{\partial q}(q, q_0, t), \tag{3.28}$$

$$p_0 = -\frac{\partial F_1}{\partial p_0}(q, q_0, t), \tag{3.29}$$

$$0 = H\left(q, \frac{\partial F_1}{\partial q}, t\right) + \frac{\partial F_1}{\partial t}. \tag{3.30}$$

The case $s = n$ and $r = 0$ corresponds to the generating function of the second kind:

$$p = \frac{\partial F_2}{\partial q}(q, p_0, t), \tag{3.31}$$

$$q_0 = \frac{\partial F_2}{\partial p_0}(q, p_0, t), \tag{3.32}$$

$$0 = H\left(q, \frac{\partial F_2}{\partial q}, t\right) + \frac{\partial F_2}{\partial t}. \tag{3.33}$$

If $s = 0$ and $r = n$, we recover the generating function of the third kind, F_3:

$$q = -\frac{\partial F_3}{\partial p}(p, q_0, t), \tag{3.34}$$

$$p_0 = -\frac{\partial F_3}{\partial q_0}(p, q_0, t), \tag{3.35}$$

$$0 = H\left(-\frac{\partial F_3}{\partial p}, p, t\right) + \frac{\partial F_3}{\partial t}. \tag{3.36}$$

Finally, if $s = 0$ and $r = 0$, we obtain F_4:

$$q = \frac{\partial F_4}{\partial p}(p, p_0, t), \tag{3.37}$$

$$q_0 = -\frac{\partial F_4}{\partial p_0}(p, p_0, t), \tag{3.38}$$

$$0 = H\left(\frac{\partial F_4}{\partial p}, p, t\right) + \frac{\partial F_4}{\partial t}. \tag{3.39}$$

To compute the generating functions, one needs boundary conditions to solve the Hamilton–Jacobi equation. At the initial time, the flow induces the identity transformation, and thus the generating functions should also do so. In other words, at the initial time,

$$q(t_0) = q_0, \;\; p(t_0) = p_0, \tag{3.40}$$

that is, $(q(t_0), p_0)$ and $(p(t_0), q_0)$ are the only sets of independent variables that contain n initial conditions and n components of the state vector at the initial time. As a consequence, all the generating functions save F_2 and F_3 are singular at the initial time, i.e., they are not defined as functions (they do not map a point into a single point). We will give deeper insight into this notion later, and we will especially show that singularities corresponds to multiple solutions from Eqs. (3.23) to (3.27).

Example 3.2.5. Let us look, for example, at the generating function of the first kind, $F_1(q, q_0, t)$. At the initial time, q is equal to q_0 whatever values the associated momenta p and p_0 take. We conclude that F_1 is singular.

We now focus on the boundary conditions for the F_2 and F_3 generating functions. At the initial time we must have:

$$\begin{cases} p_0 = \frac{\partial F_2}{\partial q}(q = q_0, p_0, t_0), \\ q_0 = \frac{\partial F_2}{\partial p_0}(q = q_0, p_0, t_0), \end{cases} \qquad \begin{cases} q_0 = -\frac{\partial F_3}{\partial p}(p = p_0, q_0, t_0), \\ p_0 = -\frac{\partial F_3}{\partial q_0}(p = p_0, q_0, t_0). \end{cases}$$

Due to the non-commutativity of the derivative operator and the operator that assigns the value t_0 at t, solutions to these equations are not unique. As a result, the boundary conditions verified by F_2 and F_3 may not be uniquely defined as well. For instance, they may be chosen to be:

$$F_2(q, p_0, t) = \langle q, p_0 \rangle, \;\; F_3(p, q_0, t) = -\langle p, q_0 \rangle, \tag{3.41}$$

or

$$F_2(q, p_0, t) = \frac{1}{t - t_0} e^{(t-t_0)\langle q, p_0 \rangle}, \quad F_3(p, q_0, t) = -\frac{1}{t - t_0} e^{(t-t_0)\langle p, q_0 \rangle}, \tag{3.42}$$

where $\langle , \rangle$ is the inner product. One can readily verify that Eqs. (3.41) and (3.42) generate the identity transformation (3.40) at the initial time $t = t_0$.

The singularity at the initial time of all but two generating functions is a major issue: it prevents us from initializing the integration, i.e., from solving the Hamilton–Jacobi equation for most generating functions. The algorithm we use circumvents this problem by specifying boundary value conditions for every generating functions at a later time (see Section 3.4).

Lemma 3.2.6. *Generating functions solve two-point boundary value problems.*

Consider two points in phase space, $X_0 = (q_0, p_0)$ *and* $X_1 = (q, p)$*, and two partitions of* $(1, \ldots, n)$ *into two non-intersecting parts,* $(i_1, \ldots, i_s)(i_{s+1}, \ldots, i_n)$ *and* $(k_1, \ldots, k_r)(k_{r+1}, \ldots, k_n)$*. A two-point boundary value problem is formulated as follows:*

Given $2n$ *coordinates* $(q_{i_1}, \ldots, q_{i_s}, p_{i_{s+1}}, \ldots, p_{i_n})$ *and* $(q_{0_{k_1}}, \ldots, q_{0_{k_r}}, p_{0_{k_{r+1}}}, \ldots, p_{0_{k_n}})$*, find the remaining* $2n$ *variables such that a particle starting at* X_0 *will reach* X_1 *in* T *units of time.*

From the relationship defined by Eqs. (3.23–3.26), we see that the generating function F_{I_s, K_r} *solves this problem. This remark is of prime importance since it provides us with a very general technique to solve any Hamiltonian boundary value problems.*

Example 3.2.7. Lambert's problem is a particular case of boundary value problem where the partitions of $(1, \ldots, n)$ are $(1, \ldots, n)()$ and $(1, \ldots, n)()$. Though, given two positions q_f and q_0 and a transfer time T, the corresponding momentum vectors are found from Eqs (3.23) to (3.26)

$$\begin{aligned} p_i &= \frac{\partial F_1}{\partial q_i}(q, q_0, T), \\ p_{0_i} &= -\frac{\partial F_1}{\partial q_{0_i}}(q, q_0, T). \end{aligned} \tag{3.43}$$

3.2.3 Linear systems theory

In this section, we particularize the theory developed above to linear systems. Specifically, we reduce the Hamilton–Jacobi equation to a set of four matrix ordinary differential equations. Then, we relate the state transition matrix and generating functions. We show that properties of one may be deduced from properties of the other. The theory we present has implications in the study of relative motion and in optimal control theory for instance [12, 26–28].

3.2.3.1 Hamilton–Jacobi equation

To study the relative motion of two particles, one often linearizes the dynamics about the trajectory (called the reference trajectory) of one of the particles. Then one uses this linear approximation to study the motion of the other particle relative to the reference trajectory (perturbed trajectory). Thus, the dynamics of relative motion reduces at first order to a time-dependent linear Hamiltonian system, i.e., a system with a quadratic Hamiltonian function without any linear terms:

$$\mathrm{H}^h = \frac{1}{2}\mathbf{X}^{h^T}\begin{pmatrix} H_{qq}(t) & H_{qp}(t) \\ H_{pq}(t) & H_{pp}(t) \end{pmatrix}\mathbf{X}^h, \tag{3.44}$$

where $\mathbf{X}^h = \binom{\Delta q}{\Delta p}$ is the relative state vector.

Lemma 3.2.8. *The generating functions associated with the phase flow transformation of the system defined by Eq. (3.44) are quadratic without linear terms.*

The proof of this lemma is trivial once we understand the link between the generating functions and the state transition matrix (see later in the section). From a heuristic perspective, we note that a linear term in the generating function would correspond to a non-homogenous term in the solution to the linear equation, which must equal zero for the dynamical system considered above.

From the above lemma, a general form for F_2 is:

$$F_2 = \frac{1}{2}\mathbf{Y}^T\begin{pmatrix} F_{11}^2(t) & F_{12}^2(t) \\ F_{21}^2(t) & F_{22}^2(t) \end{pmatrix}\mathbf{Y}, \tag{3.45}$$

where $\mathbf{Y} = \begin{pmatrix} \Delta q \\ \Delta p_0 \end{pmatrix}$ and $\begin{pmatrix} \Delta q_0 \\ \Delta p_0 \end{pmatrix}$ is the relative state vector at the initial time. We point out that both matrices defining H^h and F_2 are symmetric by definition. Then Eq. (3.31) reads:

$$\begin{aligned} \Delta p &= \frac{\partial F_2}{\partial \Delta q} \\ &= \left(F_{11}^2(t)\, F_{12}^2(t)\right)\mathbf{Y}, \end{aligned}$$

Substituting into Eq. (3.33) yields:

$$\boldsymbol{Y}^T\left\{\begin{pmatrix} \dot{F}_{11}^2(t) & \dot{F}_{12}^2(t) \\ \dot{F}_{12}^2(t)^T & \dot{F}_{22}^2(t) \end{pmatrix} + \begin{pmatrix} I & F_{11}^2(t)^T \\ 0 & F_{12}^2(t)^T \end{pmatrix}\begin{pmatrix} H_{qq}(t) & H_{qp}(t) \\ H_{pq}(t) & H_{pp}(t) \end{pmatrix}\begin{pmatrix} I & 0 \\ F_{11}^2(t) & F_{12}^2(t) \end{pmatrix}\right\}\boldsymbol{Y} = 0\,. \tag{3.46}$$

Though the above equation has been derived using F_2, it is also valid for F_1 (replacing $\mathbf{Y} = \begin{pmatrix} \Delta q \\ \Delta p_0 \end{pmatrix}$ by $\mathbf{Y} = \begin{pmatrix} \Delta q \\ \Delta q_0 \end{pmatrix}$) since F_1 and F_2 solve the same Hamilton–Jacobi equation (Eqs. (3.30) and (3.33)). Eq. (3.46) is equivalent to the following four matrix equations:

$$\begin{aligned}
&\dot{F}_{11}^{1,2}(t) + H_{qq}(t) + H_{qp}(t)F_{11}^{1,2}(t) + F_{11}^{1,2}(t)H_{pq}(t) + F_{11}^{1,2}(t)H_{pp}(t)F_{11}^{1,2}(t) = 0,\\
&\dot{F}_{12}^{1,2}(t) + H_{qp}(t)F_{12}^{1,2}(t) + F_{11}^{1,2}(t)H_{pp}(t)F_{12}^{1,2}(t) = 0,\\
&\dot{F}_{21}^{1,2}(t) + F_{21}^{1,2}(t)H_{pq}(t) + F_{21}^{1,2}(t)H_{pp}(t)F_{11}^{1,2}(t) = 0,\\
&\dot{F}_{22}^{1,2}(t) + F_{21}^{1,2}(t)H_{pp}(t)F_{12}^{1,2}(t) = 0,
\end{aligned} \tag{3.47}$$

where we replaced F_{ij}^2 by $F_{ij}^{1,2}$ to signify that these equations are valid for both F_1 and F_2. We also recall that $F_{21}^{1,2} = {F_{12}^{1,2}}^T$. A similar set of equations can be derived for any generating function F_{I_s,K_r}. However, in this section we only give the equations verified by F_3 and F_4:

$$\begin{aligned}
&\dot{F}_{11}^{3,4}(t) + H_{pp}(t) - H_{pq}(t)F_{11}^{3,4}(t) - F_{11}^{3,4}(t)H_{qp}(t) + F_{11}^{3,4}(t)H_{qq}(t)F_{11}^{3,4}(t) = 0,\\
&\dot{F}_{12}^{3,4}(t) - H_{pq}(t)F_{12}^{3,4}(t) + F_{11}^{3,4}(t)H_{qq}(t)F_{12}^{3,4}(t) = 0,\\
&\dot{F}_{21}^{3,4}(t) - F_{21}^{3,4}(t)H_{qp}(t) + F_{21}^{3,4}(t)H_{qq}(t)F_{11}^{3,4}(t) = 0,\\
&\dot{F}_{22}^{3,4}(t) + F_{21}^{3,4}(t)H_{qq}(t)F_{12}^{3,4}(t) = 0.
\end{aligned} \tag{3.48}$$

The first equations of Eqs. (3.47) and (3.48) are Riccati equations. The second and third are non-homogeneous, time varying, linear equations once the Riccati equations are solved and are equivalent to each other (i.e., transform into each other under transpose). The last are just a quadrature once the previous equations are solved.

3.2.3.2 Initial conditions

Although F_1 and F_2 (or more generally F_{I_s,K_r} and $F_{I_s,K_{r'}}$ for all r and r') verify the same Hamilton–Jacobi partial differential equation, these generating functions are different. We noticed earlier that this difference is characterized by the boundary conditions. At the initial time, the flow induces the identity transformation, thus the generating functions should also do so. In other words, at the initial time,

$$\Delta q(t_0) = \Delta q_0, \quad \Delta p(t_0) = \Delta p_0\,.$$

In terms of generating functions this translates for F_2 to:

$$\frac{\partial F_2}{\partial \Delta q}(\Delta q_0, \Delta p_0, t_0) = \Delta p_0, \quad \frac{\partial F_2}{\partial \Delta p_0}(\Delta q_0, \Delta p_0, t_0) = \Delta q_0,$$

i.e.,

$$F_{11}^2 \Delta q + F_{12}^2 \Delta p_0 = \Delta p_0, \quad F_{21}^2 \Delta q + F_{22}^2 \Delta p_0 = \Delta q_0,$$

or equivalently:

$$F_{11}^2 = F_{22}^2 = 0, \quad F_{12}^2 = F_{21}^2 = Identity\,.$$

On the other hand, F_1 is ill-defined at the initial time. Indeed, at the initial time Eqs. (3.28) and (3.29) read:

$$F_{11}^1 \Delta q + F_{12}^1 \Delta q_0 = \Delta p_0, \quad F_{21}^1 \Delta q + F_{22}^1 \Delta q_0 = \Delta p_0\,.$$

These equations do not have any solutions. This was expected since we saw earlier that $(\Delta q, \Delta q_0)$ are not independent variables at the initial time $(\Delta q = \Delta q_0)$.

3.2.3.3 Perturbation matrices

Another approach to the study of relative motion at linear order relies on the state transition matrix. This method was developed by Battin in Ref. [4] for the case of a spacecraft moving in a point mass gravity field. Let Φ be the state transition matrix which describes the relative motion:

$$\begin{pmatrix} \Delta q \\ \Delta p \end{pmatrix} = \Phi \begin{pmatrix} \Delta q_0 \\ \Delta p_0 \end{pmatrix},$$

where $\Phi = \begin{pmatrix} \Phi_{qq} & \Phi_{qp} \\ \Phi_{pq} & \Phi_{pp} \end{pmatrix}$. Battin [4] defines the fundamental perturbation matrices C and $\tilde{C}$ as:

$$\tilde{C} = \Phi_{pq}\Phi_{qq}^{-1},$$
$$C = \Phi_{pp}\Phi_{qp}^{-1}\,.$$

That is, given $\Delta p_0 = 0$, $\tilde{C}\Delta q = \Delta p$ and given $\Delta q_0 = 0$, $C\Delta q = \Delta p$. He shows that for the relative motion of a spacecraft about a circular trajectory in a point mass gravity field the perturbation matrices verify a Riccati equation and are therefore symmetric. Using the generating functions for the canonical transformation induced by the phase flow, we immediately recover these properties. We also generalize these results to any linear Hamiltonian system.

Using the notations of Eq. (3.45), Eqs. (3.31) and (3.32) read:

$$\begin{aligned} \Delta p &= \frac{\partial F_2}{\partial \Delta q} \\ &= F_{11}^2 \Delta q + F_{12}^2 \Delta p_0, \\ \Delta q_0 &= \frac{\partial F_2}{\partial \Delta p_0} \\ &= F_{21}^2 \Delta q + F_{22}^2 \Delta p_0\,. \end{aligned}$$

We solve for $(\Delta q, \Delta p)$:

$$\Delta q = {F_{21}^2}^{-1} \Delta q_0 - {F_{21}^2}^{-1} F_{22}^2 \Delta p_0,$$
$$\Delta p = F_{11}^2 {F_{21}^2}^{-1} \Delta q_0 + (F_{12}^2 - F_{11}^2 {F_{21}^2}^{-1} F_{22}^2)\Delta p_0,$$

and identify the right-hand side with the state transition matrix:

$$\begin{cases} \Phi_{qp} = -{F_{21}^2}^{-1} F_{22}^2, \\ \Phi_{qq} = {F_{21}^2}^{-1}, \\ \Phi_{pp} = F_{12}^2 - F_{11}^2 {F_{21}^2}^{-1} F_{22}^2, \\ \Phi_{pq} = F_{11}^2 {F_{21}^2}^{-1}. \end{cases}$$

We conclude that

$$\tilde{C} = \Phi_{pq} \Phi_{qq}^{-1} = F_{11}^2 . \tag{3.49}$$

In the same manner, but using F_1, we can show that:

$$C = \Phi_{pp} \Phi_{qp}^{-1} = F_{11}^1 . \tag{3.50}$$

Thus, C and $\tilde{C}$ are symmetric by nature (as $F_{11}^{1,2}$ is symmetric by definition) and they verify the Riccati equation given in Eq. (3.47).

3.2.3.4 Singularities of generating functions and their relation to the state transition matrix

In the first part of this paper, we studied the local existence of generating functions. We proved that at least one of the generating functions is well-defined at every instant (Prop. 3.2.3). In general we can notice that each of them can become singular at some point, even for simple systems. As an example let us look at the harmonic oscillator.

Example 3.2.9. The Hamiltonian for the harmonic oscillator is given by:

$$H(q, p) = \frac{1}{2m} p^2 + \frac{k}{2} q^2,$$

The F_1 generating function for the phase flow canonical transformation can be found to be:

$$F_1(q, q_0, t) = \frac{1}{2} \sqrt{km} \csc(\omega t) \left[-2 q q_0 + (q^2 + q_0^2) \cos(\omega t) \right],$$

where $\omega = \sqrt{\frac{k}{m}}$. One can readily verify that F_1 is a solution of the Hamilton–Jacobi equation (Eq. (3.30)). Although it is well-defined most of the time, at $T = m\pi/\omega$, $m \in \mathbb{Z}$, F_1 becomes singular in that the values of the coefficients of the q's and q_0's increase without bound. To understand these singularities, recall the general solution to the equations of motion:

$$q(t) = q_0 \cos(\omega t) + p_0/\omega \sin(\omega t),$$
$$p(t) = -q_0 \omega \sin(\omega t) + p_0 \cos(\omega t) .$$

At $t = T$, $q(T) = q_0$, that is q and q_0 are not independent variables. Therefore the generating function F_1 is undefined at this instant. We say that it is singular at $t = T$.

However, F_1 may be defined in the limit: at $t = T$, $q = q_0$, and thus F_1 behaves as $m\frac{(q-q_0)^2}{2(t-T)}$ as $t \mapsto T$. Finally, at $t = T$, $q = q_0$ any values of p and p_0 are possible, i.e., singularities correspond to multiple solutions to the boundary value problem that consists of going from q_0 to $q = q_0$ in $T = 0$ unit of time.

The harmonic oscillator is a useful example. Since the flow is known analytically, we are able to explicitly illustrate the relationship between the generating functions and the flow ϕ. We can go a step further by noticing that both the state transition matrix and the generating functions generate the flow. Therefore, singularities of the generating functions should be related to properties of the state transition matrix:

$$\begin{aligned}\Delta p &= \frac{\partial F_2}{\partial \Delta p} \\ &= F_{11}^2 \Delta q + F_{12}^2 \Delta p_0, \\ &\text{but we also have} \\ \Delta p &= \Phi_{pq}\Phi_{qq}^{-1}\Delta q + (\Phi_{pp} - \Phi_{pq}\Phi_{qq}^{-1}\Phi_{qp})\Delta p_0 . \\ &\text{Similarly,} \qquad (3.51) \\ \Delta q_0 &= \frac{\partial F_2}{\partial \Delta p_0} \\ &= F_{21}^2 \Delta q + F_{22}^2 \Delta p_0, \\ &\text{but we also have} \\ \Delta q_0 &= \Phi_{qq}^{-1}\Delta q - \Phi_{qq}^{-1}\Phi_{qp}\Delta p_0 . \qquad (3.52)\end{aligned}$$

A direct identification yields:

$$F_{11}^2 = \Phi_{pq}\Phi_{qq}^{-1}, \qquad (3.53)$$

$$F_{12}^2 = \Phi_{pp} - \Phi_{pq}\Phi_{qq}^{-1}\Phi_{qp}, \qquad (3.54)$$

$$F_{21}^2 = \Phi_{qq}^{-1}, \qquad (3.55)$$

$$F_{22}^2 = \Phi_{qq}^{-1}\Phi_{qp} . \qquad (3.56)$$

Thus, F_2 is singular when and only when Φ_{qq} is not invertible. This relation between singularities of F_2 and invertibility of a sub-matrix of the state transition matrix readily generalizes to other kind of generating functions. For such linear systems in particular, we can show that

- F_1 is singular when Φ_{qp} is singular,
- F_2 is singular when Φ_{qq} is singular,
- F_3 is singular when Φ_{pp} is singular,
- F_4 is singular when Φ_{pq} is singular.

To extend these results to other generating functions, we must consider other block decompositions of the state transition matrix. Every $n \times n$ block of the state transition matrix is associated with a different generating function. Since the determinant of the

state transition matrix is 1, there exists at least one $n \times n$ sub-matrix that must have a non-zero determinant. The generating function associated with this block is non-singular, and we recover Prop. 3.2.3 for linear systems.

3.2.4 Non-linear systems theory

We have shown the local existence of generating functions and mentioned that they may not be globally defined. Using linear systems theory we are also able to predict where the singularities are and to interpret their meaning as multiple solutions to the two-point boundary value problem. In this section we generalize these results to singularities of non-linear systems.

The following proposition relates singularities of the generating functions to the invertibility of sub-matrices of the Jacobi matrix of the canonical transformation.

Proposition 3.2.10. *The generating function F_{I_s,K_r} for the canonical transformation ϕ is singular at time t if and only if*

$$\det\left(\frac{\partial \phi_i}{\partial z_j}\right)_{i\in I, j\in J} = 0, \tag{3.57}$$

where $I = \{i \in I_s\} \bigcup \{n+i, i \in \bar{I}_s\}$, $J = \{j \in \bar{K}_r\} \bigcup \{n+j, j \in K_r\}$, and $z = (q_0, p_0)$ is the state vector at the initial time.

Proof. For the sake of clarity, let us prove this property for F_1. In that case, $I = [1, n]$ and $J = [n+1, 2n]$. First we remark that

$$\left(\frac{\partial \phi_i}{\partial z_j}\right)_{i\in I, j\in J} = \left(\frac{\partial q_i}{\partial p_{0_j}}\right).$$

Thus, from the inversion theorem, if $\det\left(\frac{\partial \phi_i}{\partial z_j}\right)_{i\in I, j\in J} = 0$, there is no open set in which we can solve p_0 as a function of q and q_0.

On the other hand, suppose that F_1 is non-singular. Then, from Eq. (3.29), we have:

$$p_0 = -\frac{\partial F_1}{\partial q_0}(q, q_0, t), \tag{3.58}$$

that is, we can express p_0 as a function of (q, q_0). This is in contradiction with the result obtained from the local inversion theorem. Therefore, F_1 is singular. □

Example 3.2.11. From the above proposition, we conclude that the F_1 generating function associated with the phase flow of the harmonic oscillator is singular if and only if:

$$\det\left(\frac{\partial \phi_i}{\partial z_j}\right)_{i\in I, j\in J} = 0\,. \tag{3.59}$$

In this example, $I = 1$, $J = 2$, and $\phi = (q_0 \cos(\omega t) + p_0/\omega \sin(\omega t), -q_0 \omega \sin(\omega t) + p_0 \cos(\omega t))$. Therefore F_1 is singular if and only if $\sin(\omega t) = 0$, i.e., $t = 2\pi/\omega + 2k\pi$. We recover previous results obtained by direct computation of F_1.

Prop. 3.2.10 generalizes to non-linear systems the relation between singularities and non-uniqueness of the solutions to boundary value problems. Indeed, F_{I_s,K_r} is singular if and only if $z_J \mapsto \phi_I(t, z)$ is not an isomorphism. In other words, singularities arise when there exist multiple solutions to the boundary value problem.

To study the singularities of non-linear systems, we need to introduce the concept of Lagrangian submanifolds. The theory of Lagrangian submanifolds goes far beyond the results we present in this section: "Some believe that the Lagrangian submanifold approach will give deeper insight into quantum theories than does the Poisson algebra approach. In any case, it gives deeper insight into classical mechanics and classical field theories" (Abraham and Marsden [1]). We refer to Abraham and Marsden [1], Marsden [23] and Weinstein [32] and references given therein for further information on these subjects.

3.2.4.1 Lagrangian submanifolds and the study of caustics

Consider an arbitrary generating function F_{I_s,K_r}. Then the graph of $\mathrm{d}F_{I_s,K_r}$ defines a $2n$-dimensional submanifold called a canonical relation [32] of the $4n$-dimensional symplectic space $(\mathcal{P}_1 \times \mathcal{P}_2, \Omega = \pi_1^* \omega_1 - \pi_2^* \omega_2)$. On the other hand, since the variables (q_0, p_0) do not appear in the Hamilton–Jacobi equation (Eq. (3.27)), we may consider them as parameters. In that case the graph of $(q_{I_s}, p_{\bar{I}_s}) \mapsto \mathrm{d}F_{I_s,K_r}$ defines an n-dimensional submanifold of the symplectic space $(\mathcal{P}_1, \omega_1)$ called a Lagrangian submanifold [32]. The study of singularities can be achieved using either canonical relations [1] or Lagrangian submanifolds [3, 23]. In the following we assumed t fixed.

Theorem 3.2.12. *The generating function F_{I_s,K_r} is singular if and only if the local projection of the canonical relation $\mathcal{L}$ defined by the graph of dF_{I_s,K_r} onto $(q_{I_s}, p_{\bar{I}_s}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$ is not a local diffeomorphism.*

Definition 3.2.2. The projection of a singular point F_{I_s,K_r} onto $(q_{I_s}, p_{\bar{I}_s}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$ is called a caustic.

If one works with Lagrangian submanifolds then the previous theorem becomes:

Theorem 3.2.13. *The generating function[2] F_{I_s,K_r} is singular if the local projection of the Lagrangian submanifold defined by the graph of $(q_{I_s}, p_{\bar{I}_s}) \mapsto dF_{I_s,K_r}$ onto $(q_{I_s}, p_{\bar{I}_s})$ is not a local diffeomorphism.*

These theorems are the geometric formulation of Prop. 3.2.10. If the projection of the canonical relation defined by the graph of $\mathrm{d}F_{I_s,K_r}$ onto $(q_{I_s}, p_{\bar{I}_s}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$ is not

[2] We consider here that the generating function is a function of n variables only, and has n parameters.

a local diffeomorphism, then there exists multiple solutions to the problem of finding (q_0, p_0, q, p) knowing $(q_{I_s}, p_{\bar{I}_s}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$. From the local inversion theorem, this is equivalent to Prop. 3.2.10.

In the light of these theorems, we can give a geometrical interpretation to Thm. 3.2.3 on the existence of generating functions. Given a canonical relation $\mathcal{L}$ (or a Lagrangian submanifold) defined by a canonical transformation, there exists a $2n$-dimensional (or n-dimensional) submanifold $\mathcal{M}$ of $\mathcal{P}_1 \times \mathcal{P}_2$ (or $\mathcal{P}_1$) such that the local projection of $\mathcal{L}$ onto $\mathcal{M}$ is a local diffeomorphism.

To study caustics two approaches, at least, are possible depending on the problem. A good understanding of the physics may provide information very easily. For instance, consider the two-body problem in two dimensions, and the problem of going from a point A to a point B, symmetrically placed on a single line on either side of the body, in a certain lapse of time, T. For certain values of T, the trajectory that links A to B is an ellipse whose perigee and apogee are A and B. Therefore, there are two solutions to this problem depending upon which way the particle is going. In terms of generating functions, we deduce that F_3 is non-singular (there is a unique solution once the final momentum is given) but F_1 is singular (existence of two solutions).

Another method for studying caustics consists of using a known non-singular generating function to define the Lagrangian submanifold $\mathcal{L}$ and then study its projection. A very illustrative example is given by Ehlers and Newman [8]. Using the Hamilton–Jacobi equation they treat the evolution of an ensemble of free particles whose initial momentum distribution is $p = (1/(1+q^2))$. They identify a time t_1 at which F_1 is singular. Then, using a closed-form expression of F_3, they find the equations defining the Lagrangian submanifold at t_1. Its projection can be studied and they eventually find that the caustic is two folds. Nevertheless, such an analysis is not always possible as solutions to the Hamilton–Jacobi equation are usually found numerically, not analytically. To illustrate this method, let us consider the following example.

Example 3.2.14. (Motion about the Libration point L_2 in the Hill three-body problem.) Consider a spacecraft moving about and staying close to the Libration point L_2 in the normalized Hill three-body problem (see Appendix B for a description of the Hill three-body problem). The algorithm we develop in Section 3.4 computes the generating function for relative motion with respect to L_2 as a Taylor series expansion, of order N, of the exact generating function about L_2. For instance, F_2 reads:

$$\begin{aligned} F_2(q_x, q_y, p_{0_x}, p_{0_y}, t) = {} & f_{11}^2(t)q_x^2 + f_{12}^2(t)q_x q_y + f_{13}^2(t)q_x p_{0_x} + f_{14}^2(t)q_x p_{0_y} + f_{22}^2(t)q_y^2 \\ & + f_{23}^2(t)q_y p_{0_x}(t) + f_{24}^2(t)q_y p_{0_y} + f_{33}^2(t)p_{0_x}^2 + f_{34}^2(t)p_{0_x}p_{0_y} \\ & + f_{44}^2(t)p_{0_y}^2 + r(q_x, q_y, p_{0_x}, p_{0_y}, t), \end{aligned}$$

where (q, p, q_0, p_0) are relative position and momenta of the spacecraft with respect to L_2 at t and t_0, the initial time, and r is a polynomial of degree N in its spatial variables

with time-dependent coefficients and without any quadratic terms. At $T = 1.6822$, F_1 is singular but F_2 is not. Eqs. (3.31) and (3.32) reads:

$$p_x = 2f_{11}^2(T)q_x + f_{12}^2(T)q_y + f_{13}^2(T)p_{0_x} + f_{14}^2(T)p_{0_y} + D_1 r(q_x, q_y, p_{0_x}, p_{0_y}, T), \quad (3.60)$$

$$p_y = f_{12}^2(T)q_x + 2f_{22}^2(T)q_y + f_{23}^2(T)p_{0_x} + f_{24}^2(T)p_{0_y} + D_2 r(q_x, q_y, p_{0_x}, p_{0_y}, T), \quad (3.61)$$

$$q_{0_x} = f_{13}^2(T)q_x + f_{23}^2(T)q_y + 2f_{33}^2(T)p_{0_x} + f_{34}^2(T)p_{0_y} + D_3 r(q_x, q_y, p_{0_x}, p_{0_y}, T), \quad (3.62)$$

$$q_{0_y} = f_{14}^2(T)q_x + f_{24}^2(T)q_y + f_{34}^2(T)p_{0_x} + 2f_{44}^2(T)p_{0_y} + D_4 r(q_x, q_y, p_{0_x}, p_{0_y}, T), \quad (3.63)$$

where $D_i r$ represents the derivative of r with respect to its ith variable. Eqs. (3.60)–(3.63) define a canonical relation $\mathcal{L}$. By assumption F_1 is singular, therefore the projection of $\mathcal{L}$ onto (q, q_0) is not a local diffeomorphism and there exists a caustic.

Let us now study this caustic. Eqs. (3.60)–(3.63) provide p and q_0 as a function of (q, p_0), but to characterize the caustic we need to study the projection of the Lagrangian manifold on[3] (q, q_0). Hence, we must express p and p_0 as a function of (q, q_0). F_1 being singular, there are multiple solutions to the problem of finding p and p_0 as a function of (q, q_0), and one valuable piece of information is the number k of such solutions. To find p and p_0 as a function of (q, q_0) we first invert Eqs. (3.62) and (3.63) to express p_0 as a function of (q, q_0). Then we substitute this relation into Eqs. (3.60) and (3.61). The first step requires a series inversion that can be carried out using the technique developed by Moulton [25]. Let us rewrite Eqs. (3.62) and (3.63):

$$2f_{33}^2(T)p_{0_x} + f_{34}^2(T)p_{0_y} = q_{0_x} - f_{13}^2(T)q_x - f_{23}^2(T)q_y - D_3 r(q_x, q_y, p_{0_x}, p_{0_y}, T), \quad (3.64)$$

$$f_{34}^2(T)p_{0_x} + 2f_{44}^2(T)p_{0_y} = q_{0_y} - f_{14}^2(T)q_x - f_{24}^2(T)q_y - D_4 r(q_x, q_y, p_{0_x}, p_{0_y}, T). \quad (3.65)$$

The determinant of the coefficients of the linear terms on the left-hand side is zero (otherwise there is a unique solution to the series inversion) but each of the coefficients is non-zero, i.e., we can solve for p_{0_x} as a function of $(p_{0_y}, q_{0_x}, q_{0_y})$ using Eq. (3.64). Then we substitute this solution into Eq. (3.65) and we obtain an equation of the form

$$R(p_{0_y}, q_{0_x}, q_{0_y}) = 0, \quad (3.66)$$

that contains no terms in p_{0_y} alone of the first degree. In addition, R contains a non-zero term of the form $\alpha p_{0_y}^2$, where α is a real number. In this case, Weierstrass proved that there exist two solutions, $p_{0_y}^1$ and $p_{0_y}^2$, to Eq. (3.66).

[3] Since F_1 is a function of (q, q_0).

In the same way, we can study the singularity of F_1 at the initial time. At $t = 0$, F_2 generates the identity transformation, hence $f_{33}^2(0) = f_{34}^2(0) = f_{43}^2\ (0) = f_{44}^2(0) = 0$. This time there is no non-zero first minor, and we find that there exists infinitely many solutions to the series inversion. Another way to see this is to use the Legendre transformation:

$$F_1(q, q_0, t) = F_2(q, p_0, t) - q_0 p_0,$$

As t tends toward 0, (q, p) goes to (q_0, p_0) and F_2 converges toward the identity transformation $\lim_{t\to 0} F_2(q, p_0, t) = qp_0 \underset{t\to 0}{\longrightarrow} q_0 p_0$. Therefore, as t goes to 0, F_1 also goes to 0, i.e., the projection of $\mathcal{L}$ onto (q, q_0) reduces to a point.

The use of series inversion to quantify the number of solutions to the boundary value problem is a very efficient technique for systems with polynomial generating functions. From the series inversion theory we know that the uniqueness of the inversion is determined by the linear terms whereas the number of solutions (if many) depends on properties of non-linear terms (we illustrated this property in the above example). In addition, this technique allows us to study the projection of the canonical relation at the cost of a single matrix inversion only.

In the case where generating functions are (or can be approximated by a) polynomial, we can recover the phase flow (or its approximation) as a polynomial too. For instance, from

$$p_0 = \frac{\partial F_1}{\partial q_0}(q, q_0, t),$$

we can find $q(q_0, p_0)$ at the cost of a series inversion. Then, $q(q_0, p_0)$ together with $p = \frac{\partial F_1}{\partial q}(q, q_0, t)$ define the flow (or its polynomial approximation). On the other hand, generating functions are well-defined if and only if the transformation from the flow to the generating function has a unique solution (Prop. 3.2.10). From series inversion theory, we conclude that generating functions are well-defined if and only if the inversion of the linear approximation of the flow has a unique solution. Therefore, we have the following property:

Proposition 3.2.15. *Singularities of* ***polynomial*** *generating functions correspond to degeneracy of sub-matrices of the state transition matrix as in the linear case. In other words, using our previous notation,*

- F_1 *is singular when* $\det(\Phi_{qp}) = 0$,
- F_2 *is singular when* $\det(\Phi_{qq}) = 0$,
- F_3 *is singular when* $\det(\Phi_{pp}) = 0$,
- F_4 *is singular when* $\det(\Phi_{pq}) = 0$.

Using other block decompositions of the state transition matrix, these results can be extended to the generating function F_{I_s, K_r}.

Example 3.2.16. (Singularities of the generating functions in the Hill three-body problem.) To illustrate Prop. 3.2.11, let us determine the singularities of F_1 and F_2 in the normalized Hill three-body problem linearized about L_2.

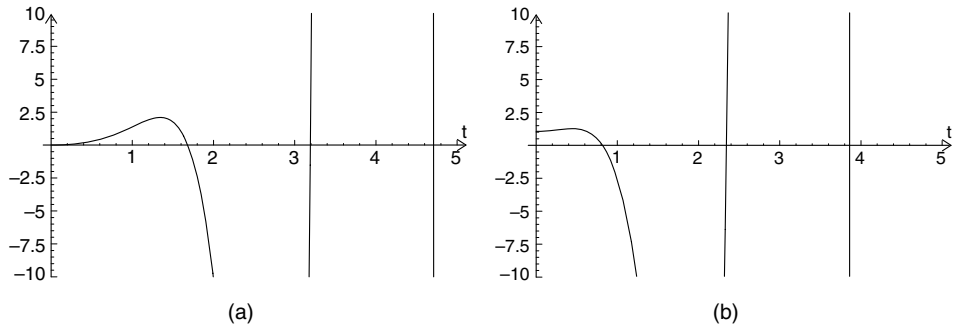


Fig. 3.1. Determinants of Φ_{qq} and Φ_{qp}.

The state transition matrix for this problem satisfies (Appendix B, Eq. (B.8)):

$$\dot{\phi}(t) = \begin{pmatrix} -8 & 0 & 0 & -1 \\ 0 & 4 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \phi(t), \quad \phi(0) = Identity\,.$$

We use the *Mathematica*© built in function *DSolve* to compute a symbolic expression of the state transition matrix. We plot in Figure 3.1 the determinant of Φ_{qq} and Φ_{qp} as a function of time. As noticed before F_1 is singular at the initial time and at $t = \{1.6821, 3.1938, 4.710\}$ and F_2 is singular at $t = \{0.809, 2.3443, 3.86\}$. The singularity at $t = 1.6821$ was studied above.

3.3 Hamilton's principal function

Though generating functions are used in the present research to solve boundary value problems, they were introduced by Jacobi, and mostly used thereafter, as fundamental functions which can solve the equations of motion by simple differentiations and eliminations, without integration. Nevertheless, it was Hamilton who first hit upon the idea of finding such a fundamental function. He first proved its existence in geometrical optics (i.e., for time-independent Hamiltonian systems) in 1834 and called it the characteristic function [19]. One year later he published a second essay [20] on systems of attracting and repelling points in which he showed that the evolution of dynamical systems is characterized by a single function called Hamilton's principal function:

> The former Essay contained a general method for reducing all the most important problems of dynamics to the study of one characteristic function, one central or radical relation. It was remarked at the close of that Essay, that many eliminations required by this method in its first conception, might be avoided by a general transformation, introducing the time explicitly into a part S of the whole characteristic function V; and it is now proposed to fix the attention chiefly on this part S, and to call it the Principal

> Function. (William R. Hamilton, in the introductory remarks of "Second essay on a General Method in Dynamics" [20]).

Although Hamilton's principal function has been introduced to derive solutions to the equations of motion, it may also be used to solve boundary value problems, similar to the generating functions. Therefore, in the next section we introduce Hamilton's principal function and prove that it solves two-point boundary value problems. Then we discuss how it compares to the generating functions.

3.3.1 Existence of the Hamilton principal function

Similarly to the generating functions, Hamilton's principal function may be derived using the calculus of variations. Consider the extended action integral:

$$A = \int_{\tau_0}^{\tau_1} (pq' + p_t t')\mathrm{d}\tau, \tag{3.67}$$

under the auxiliary condition $K(q, t, p, p_t) = 0$, where $q' = \mathrm{d}q/\mathrm{d}\tau$, p_t is the momentum associated with the generalized coordinates t and $K = p_t + H$.

Define a line element[4] $\mathrm{d}\sigma$ for the extended configuration space (q, t) by

$$\mathrm{d}\sigma = L\mathrm{d}t = Lt'\mathrm{d}\tau.$$

where L is the Lagrangian. Then, we can connect two points (q_0, t_0) and (q_1, t_1) of the extended configuration space by a shortest line γ and measure its length from:

$$A = \int_\gamma \mathrm{d}\sigma = \int_\gamma Lt'\mathrm{d}\tau.$$

The distance we obtain is a function of the coordinates of the end-points and, by definition, is given by the Hamilton principal function: $W(q_0, t_0, q_1, t_1)$.

From the calculus of variations (see e.g., Lanczos [22]) we know that the variation of the action A can be expressed as a function of the boundary terms if we vary the limits of the integral:

$$\delta A = p_1\delta q_1 + p_{t_1}\delta t_1 - p_0\delta q_0 - p_{t_0}\delta t_0\,.$$

On the other hand, we have:

$$\delta A = \delta W(q_0, t_0, q_1, t_1) = \frac{\partial W}{\partial q_0}\delta q_0 + \frac{\partial W}{\partial t_0}\delta t_0 + \frac{\partial W}{\partial q_1}\delta q_1 + \frac{\partial W}{\partial t_1}\delta t_1,$$

i.e.,

$$p_0 = -\frac{\partial W}{\partial q_0}(q_0, t_0, q_1, t_1), \tag{3.68}$$

$$p_1 = \frac{\partial W}{\partial q_1}(q_0, t_0, q_1, t_1), \tag{3.69}$$

[4] The geometry established by this line element is not Riemannian [22].

and

$$-\frac{\partial W}{\partial t_0}(q_0, t_0, q_1, t_1) + H\left(q_0, -\frac{\partial W}{\partial q_0}, t_0\right) = 0, \tag{3.70}$$

$$\frac{\partial W}{\partial t_1}(q_0, t_0, q_1, t_1) + H\left(q_1, \frac{\partial W}{\partial q_1}, t_1\right) = 0, \tag{3.71}$$

where K has been replaced by $p_t + H$. As with generating functions of the first kind, Hamilton's principal function solves boundary value problems of Lambert's type through Eqs. (3.68–3.69). To find W, however, we need to solve a system of two partial differential equations (Eqs. (3.70) and (3.71)).

3.3.2 Hamilton's principal function and generating functions

In this section, we highlight the main differences between generating functions associated with the phase flow and Hamilton's principal function. For sake of simplicity we compare $F_1(q, q_0, t)$ and $W(q, t, q_0, t_0)$.

3.3.2.1 Calculus of variations

Even if both functions are derived from the calculus of variations, there are fundamental differences between them. To derive generating functions the time t is considered as an independent variable in the variational principle. In contrast, we increase the dimensionality of the system by adding the time t to the generalized coordinates to derive Hamilton's principal function. As a consequence, generating functions generate a transformation between two points in the phase space, i.e., they act without passage of time. On the other hand, Hamilton's principal function generates a transformation between two points in the extended phase space, i.e., between two points in the phase space with different times. This difference may be viewed as follows: Generating functions allow us to characterize the phase flow given an initial time, t_0 (i.e., to characterize all trajectories whose initial conditions are specified at t_0), whereas Hamilton's principal function does not impose any constraint on the initial time. The counterpart being that Hamilton's principal function must satisfy two partial differential equations (Eq. (3.70) defines W as a function of t_0 and Eq. (3.71) defines W as a function of t_1) whereas generating functions satisfy only one.

Moreover, to derive the generating functions fixed endpoints are imposed, i.e., we impose the trajectory in both sets of variables to verify the principle of least action. On the other hand, the variation used to derive Hamilton's principal function involves moving endpoints and an energy constraint. This difference may be interpreted as follows: Hamilton's principal function generates a transformation which maps a point of a given energy surface to another point on the same energy surface and is not defined

for points that do not lie on this surface. As a consequence of the energy constraint, we have [22]:

$$\left| \frac{\partial^2 W}{\partial q_0 \partial q_1} \right| = 0 . \tag{3.72}$$

As noticed by Lanczos [22], "this is a characteristic property of the W-function which has no equivalent in Jacobi's theory". On the other hand, generating functions map any point of the phase space into another one, the only constraint is imposed through the variational principle (or equivalently by the definition of canonical transformation): we impose the system in both sets of coordinates to be Hamiltonian with Hamiltonian functions H and K, respectively.

3.3.2.2 Fixed initial time

In the derivation of Hamilton's principal function $\mathrm{d}t_0$ may be chosen to be zero, i.e., the initial time is imposed. Then Hamilton's principal function loses its dependence with respect to t_0. Eq. (3.70) is trivially verified and Eq. (3.72) does not hold anymore, meaning that W and F_1 become equivalent.

Finally, in Ref. [20] Hamilton also derives another principal function $Q(p_0, t_0, p_1, t_1)$ which compares to W as F_4 compares to F_1. The derivation being the same we will not go through it.

To conclude, Hamilton's principal function appears to be more general than the generating functions for the canonical transformation induced by the phase flow. On the other hand, the initial and final times are usually specified when solving two-point boundary problems and therefore, any of these functions will identically solve the problem. However, to find Hamilton's principal function we need to solve two partial differential equations whereas only one needs to be solved to find the generating functions. For these reasons, generating functions are more appropriate to address the problem of solving two-point boundary value problems.

3.4 Local solutions of the Hamilton–Jacobi equation

In this section, we provide a detailed discussion of a method we use to solve for the generating functions of the solution flow. This method only applies for systems with polynomial generating functions. This case obviously includes systems with polynomial Hamiltonian such as the double well potential. It also includes systems describing the relative motion of two particles moving in a Hamiltonian vector field and more generally, the motion of a particle in the vicinity of an equilibrium point or of a known trajectory. In the following we focus our discussion on the problem of relative motion between particles.

3.4.1 Direct solution for the generating function

Suppose we are interested in the relative motion of a particle whose coordinates are (q, p) with respect to another one on a known reference trajectory whose coordinates are (q^0, p^0), both moving in an Hamiltonian field. If both particles stay "close" to each other, we can expand (q, p) as a Taylor series about the reference trajectory. The dynamics of the relative motion is described by the Hamiltonian function H^h [16]:

$$\mathrm{H}^h(X^h, t) = \sum_{p=2}^{\infty} \sum_{\substack{i_1,\ldots,i_{2n}=0,\\ i_1+\cdots+i_{2n}=p}}^{p} \frac{1}{i_1!\cdots i_{2n}!} \frac{\partial^p H}{\partial q_1^{i_1}\cdots\partial q_n^{i_n}\partial p_1^{i_{n+1}}\cdots\partial p_n^{i_{2n}}}(q^0, p^0, t) X_1^{h\,i_1}\ldots X_{2n}^{h\;\;i_{2n}}, \tag{3.73}$$

where $\mathbf{X}^h = \begin{pmatrix} q \\ p \end{pmatrix}$ is the relative state vector. Since H^h has infinitely many terms, we are usually not able to solve the Hamilton–Jacobi equation but we can approximate the dynamics by truncating the series H^h in order to only keep finitely many terms. Suppose N terms are kept, then we say that we describe the relative motion using an approximation of order N. Clearly, the greater N is, the better our approximation is to the non-linear motion of a particle about the reference trajectory. When an approximation of order N is used, we look for a generating function F_{I_s,K_r} as a polynomial of order N in its spatial variables with time-dependent coefficients. It is important to note that even for a Hamiltonian with a finite expansion, the generating functions for that Hamiltonian will in general be analytic functions with infinite series expansions. Thus, truncation of a generating function at order N is not equivalent to a solution of the generating function of the order N Hamiltonian system, but is always only an approximation to it. Substituting the expansions into the Hamilton–Jacobi equation, it is reduced to a set of ordinary differential equations that can be integrated numerically. Once F_{I_s,K_r} is known, we find the other generating functions from the Legendre transformation, at the cost of a series inversion. If a generating function is singular, the inversion does not have a unique solution and the number of solutions characterizes the caustic.

Recall the Hamilton–Jacobi equation (Eq. (3.27)):

$$H\left(q_{I_s}, -\frac{\partial F_{I_s,K_r}}{\partial p_{\bar{I}_s}}, \frac{\partial F_{I_s,K_r}}{\partial q_{I_s}}, p_{\bar{I}_s}, t\right) + \frac{\partial F_{I_s,K_r}}{\partial t} = 0\,. \tag{3.74}$$

Since H is a Taylor series in its spatial variables, we look for a solution of the same form, that is, we assume that generating functions are Taylor series as well:

$$F_{I_s,K_r}(y, t) = \sum_{q=0}^{\infty} \sum_{\substack{i_1,\ldots,i_{2n}=0\\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}} f_{q,i_1,\ldots,i_{2n}}^{p,r}(t) y_1^{i_1}\cdots y_{2n}^{i_{2n}}, \tag{3.75}$$

where $y = (q_{I_s}, p_{\bar{I}_s}, q_{0_{K_r}}, p_{0_{\bar{K}_r}})$. We substitute this expression into Eq. (3.74). The resulting equation is an ordinary differential equation that has the following structure:

$$P\left(y, f_{q,i_1,\ldots,i_{2n}}^{p,r}(t), \dot{f}_{q,i_1,\ldots,i_{2n}}^{p,r}(t)\right) = 0, \tag{3.76}$$

where P is a series in y with time-dependent coefficients. An explicit expression of P up to order 3 is given in Appendix A. Equation (3.76) holds for all y if and only if all the coefficients of P are zero. In this manner, we transform the ordinary differential equation (Eq. (3.76)) into a set of ordinary differential equations whose solutions are the coefficients of the generating function F_{I_s,K_r}.

Now it remains to specify initial conditions for the integration. We have seen before that only F_2 and F_3 can generate the identity transformation, the other generating functions being singular. Let us look more closely at F_2 and F_3, and especially at the coefficients[5] $f^2_{q,i_1,\ldots,i_{2n}}(t_0)$ and $f^3_{q,i_1,\ldots,i_{2n}}(t_0)$. At the initial time we have:

$$\begin{aligned} p_0 &= p \\ &= \frac{\partial F_2}{\partial q}, \end{aligned}$$

and

$$\begin{aligned} q &= q_0 \\ &= \frac{\partial F_2}{\partial p_0}. \end{aligned}$$

Within the radius of convergence, the Taylor series defining the generating functions (Eq. (3.75)) converge normally. Therefore, we can invert the summation and the derivative operator. We obtain a unique set of initial conditions:

$$f^2_{q,i_1,\ldots,i_{2n}}(t_0) = \begin{cases} 1 & \text{if } q = 2,\ i_k = i_{k+n} = 1,\ i_{l\neq\{k,k+n\}} = 0,\ \forall (k,l) \in [1,n]\times[1,2n], \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we obtain for F_3:

$$f^3_{q,i_1,\ldots,i_{2n}}(t_0) = \begin{cases} -1 & \text{if } q = 2,\ i_k = i_{k+n} = 1,\ i_{l\neq\{k,k+n\}} = 0,\ \forall (k,l) \in [1,n]\times[1,2n], \\ 0 & \text{otherwise.} \end{cases}$$

These initial conditions allow one to integrate two generating functions among the 4^n, but what about the other ones? This issue on singular initial conditions is similar to the one on singularity avoidance during the integration. In the next section we propose a technique to handle these problems based on the Legendre transformation. But before going further, one remark needs to be made. After we proceed with the integration, one must always verify that the series converge and that they describe the true dynamics[6] in some open set. If these two conditions are verified we can identify the generating functions with their Taylor series.

---

[5] We change our notation for convenience: f^2 stands for $f^{n,0}$, i.e., represents the coefficients of the Taylor series of F_2. We do the same for all four kinds of generating functions F_1, F_2, F_3 and F_4.

[6] Remember that even if a function is C^∞ and has a converging Taylor series, it may not equal its Taylor series. As an example take $f(x) = \exp(1/x^2)$ if $x \neq 0$, $f(0) = 0$, it is C^∞ and its Taylor series at $x = 0$ is 0, and therefore converges. However, f is not identically zero.

3.4.1.1 Singularity avoidance

We have seen that most of the generating functions are singular at the initial time. Moreover solutions to the Hamilton–Jacobi equations often develop caustics. These two issues prevent numerical integration. The goal of this section is to introduce a technique to overcome this difficulty.

We first need to recall the Legendre transformation, which allows one to derive one generating function from another (Eq. (3.14)). Suppose F_2 is known, then we can find F_1 from:

$$F_1(q, q_0, t) = F_2(q, p_0, t) - \langle q_0, p_0 \rangle, \tag{3.77}$$

where p_0 is viewed as a function of (q, q_0). Obviously, the difficulty in proceeding with a Legendre transformation lies in finding p_0 as a function of (q, q_0). To find such an expression we use Eq. (3.32):

$$q_0 = \frac{\partial F_2}{\partial p_0}(q, p_0, t), \tag{3.78}$$

and then solve for $p_0(q, q_0)$.

For the class of problems we consider, F_2 is a Taylor series. Therefore we need to perform a series inversion to eventually find p_0 as a Taylor series of (q, q_0). Series inversion is a classical problem, we adopt the procedure developed by Moulton [25]. We first suppose that there exists a series expansion of p_0 as a function of q and q_0. Then we insert this expression into Eq. (3.78) and balance terms of the same order. We obtain a set of linear equations, whose solution is found at the cost of a $n \times n$ matrix inversion (we recall that n is the dimension of the configuration space, it is small in general). If its rank is $n - p$, Weierstrass proved that the series inversion has $p + 1$ solutions (for instance, if $p = 1$, there are two solutions to the problem). This is the linear version of Prop. 3.2.10 for the F_1 generating function.

Let us return to the problem of singularity avoidance. So far, we were able to integrate generating functions of the second and third kinds since they have well-defined initial conditions. To integrate other generating functions, say F_{I_s,K_r}, we need to specify boundary conditions. Using the Legendre transformation, we can find the value of F_{I_s,K_r} at $t_1 > 0$ from the value of F_2 or F_3. This value can in turn be used to initialize the integration of the Hamilton–Jacobi equation for F_{I_s,K_r} which can be continued forward or backward in time until it encounters a singularity.

Now suppose F_2 is singular at t_2. Let us see how we can take advantage of the Legendre transformation to integrate F_2 for $t > t_2$. Proposition 3.2.3 tells us that at least one of the generating functions is non-singular at t_2. Without loss of generality, we assume that F_1 is non-singular at t_2. At $t_1 < t_2$ we carry out a Legendre transformation to find F_1 from F_2, then we integrate F_1 over $[t_1, t_3 > t_2]$ and carry out another Legendre transformation to recover F_2 at t_3. Once the value of F_2 is found at t_3, the integration of the Hamilton–Jacobi equation can be continued. Finally, we recall from Proposition 3.2.15 that we can predict the locus of the singularities using the state transition matrix.

We have described an algorithm to solve the Hamilton–Jacobi equation and developed techniques to continue the integration despite singularities. In the next section, we

introduce an indirect approach to compute the generating functions based on the initial value problem. This approach naturally avoids singularities but requires more computations (see Section 3.4.3).

3.4.2 An indirect approach

By definition, generating functions implicitly define the canonical transformation they are associated with. Hence, we may compute the generating functions from the canonical transformation, i.e., compute the generating functions associated with the flow from knowledge of the flow. In this section, we develop an algorithm based on these remarks.

Recall Hamilton's equations of motion:

$$\begin{pmatrix} q \\ p \end{pmatrix} = J\nabla H(q, p, t)\,. \tag{3.79}$$

Suppose that $q(q_0, p_0, t)$ and $p(q_0, p_0, t)$ can be expressed as series in the initial conditions (q_0, p_0) with time-dependent coefficients, truncate the series to order N and substitute these into Eq. (3.79). Hamilton's equations reduce to an ordinary differential equation of a form that is polynomial in (q_0, p_0). As before, we balance terms of the same order and transform Hamilton's equations into a set of ordinary differential equations whose variables are the time-dependent coefficients defining q and p as series of q_0 and p_0. Using $q(q_0, p_0, t_0) = q_0$ and $p(q_0, p_0, t_0) = p_0$ as initial conditions for the integration, we are able to compute an approximation of order N of the phase flow. Once the flow is known, we recover the generating functions by performing a series inversion.

Example 3.4.1. Suppose we want to compute F_1 at $t = T$. From $q = q(q_0, p_0, T)$ we carry out a series inversion to eventually find $p_0 = p_0(q, q_0, T)$. Then $p_0 = p_0(q, q_0, T)$ together with $p = p(q_0, p_0, T)$ defines the gradient of F_1:

$$\begin{aligned} \frac{\partial F_1}{\partial q}(q, q_0, T) &= p \\ &= p(q_0, p_0(q, q_0, T)), \end{aligned} \tag{3.80}$$

$$\begin{aligned} \frac{\partial F_1}{\partial q_0}(q, q_0, T) &= -p_0 \\ &= -p_0(q, q_0, T). \end{aligned} \tag{3.81}$$

We recover F_1 from its gradient by performing two quadratures over the polynomial terms. If one uses traditional numerical integrators to integrate the phase flow, Eqs. (3.80) and (3.81) are not integrable due to numerical round off

$$\left(\frac{\partial p(q_0, p_0(q, q_0, T))}{\partial q_0} \neq -\frac{\partial p_0(q, q_0, T)}{\partial q} \right).$$

Using symplectic algorithms to compute the approximate phase flow, we preserve the Hamiltonian structure of the flow and thus are assured that Eqs. (3.80) and (3.81) are integrable [13].

3.4.3 A comparison of the direct and indirect approach

We have introduced two algorithms that compute the generating functions associated with the phase flow. In this section, we highlight the advantages and drawbacks of each method. In addition, we show that by combining them we obtain a robust and powerful algorithm.

3.4.3.1 Method specifications

The direct approach

The direct approach provides us with a closed form approximation of the generating functions over a given time interval. However, there are inherent difficulties as generating functions may develop singularities which prevent the integration from going further in time. The technique we developed to bypass this problem results in additional computations. It requires us to first identify the times at which generating functions become singular, and then to find a non-singular generating function at each of these times. Over a long time simulation, this method reaches its limits as many singularities may need to be avoided.

The indirect approach

The main advantage of the indirect method is that it never encounters singularities, as the flow is always non-singular. On the other hand, this method requires us to solve many more equations than the direct approach (see below). Furthermore, a major drawback of the indirect approach is that it computes an expression for the generating functions at a given time only, the time at which the series inversion is performed. To generate solutions to a two-point boundary value problem over a range of times then requires that a series inversion be performed at each point in time.

The curse of dimensionality

In this paragraph, we point out a difficulty inherent to both methods, namely the "curse of dimensionality". As we solve the generating functions to higher and higher orders, the number of variables grows dramatically. This problem is the limiting factor for computation: typically on a $2GHz$ Linux computer with $1G$ RAM, we have trouble solving the generating functions to order 7 and up for a 6-dimensional Hamiltonian system.

Computation of the generating functions using the direct approach requires us to find all the coefficients of a $2n$-dimensional series with no linear terms. At order N, a $2n$-dimensional Taylor series has x terms, where

$$x = \binom{2n-1+N}{N} = \frac{(2n-1+N)!}{N!(2n-1)!}.$$

In the indirect approach we express the $2n$-dimensional state vector as Taylor series with respect to the $2n$ initial conditions. Therefore, we need to compute the coefficients of $2n$ $2n$-dimensional Taylor series.

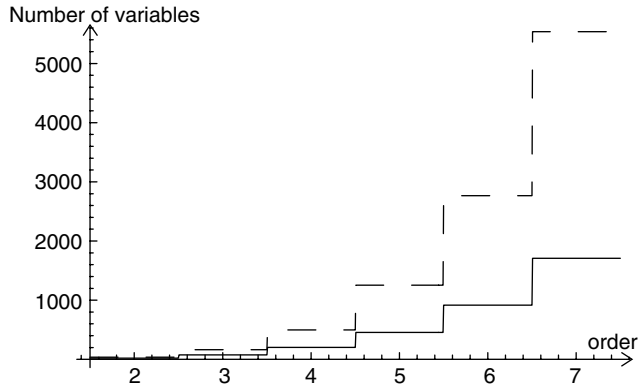


Fig. 3.2. Number of variables in the indirect (dashed) and direct (solid) methods.

To summarize, an approximation of order N of the generating functions is found by solving:

- $\sum_{k=2}^{N} \frac{(2n-1+k)!}{k!(2n-1)!}$ ordinary differential equations using the direct approach,
- $2n \sum_{k=1}^{N-1} \frac{(2n-1+k)!}{k!(2n-1)!}$ ordinary differential equations using the indirect approach[7].

In Figure 3.2, the solid line and dotted line indicate the numbers of equations that needs to be solved with the direct and indirect methods for a 6-dimensional Hamiltonian system.

3.4.3.2 A combined algorithm

In practice, to solve boundary value problems over a long time span it is most convenient to combine both methods. Typically, we first solve the initial value problem (indirect method) up to a time of interest, say T. Then we solve the Hamilton–Jacobi equation (direct approach) about T, with initial conditions equal to the values of the generating functions at T found using the indirect approach. This approach has been applied in [17, 18] to solve two-point boundary value problems over week-long time spans in low Earth orbit.

3.4.4 Convergence and existence of solutions

We now study the convergence properties of our algorithm. In particular, we provide a criterion to evaluate the domain in which the approximation of order N of the generating functions is valid. An example to illustrate this criterion is given.

[7] The summation goes from 1 to $N-1$ because the indirect approach computes the gradient of the generating functions.

3.4.4.1 Theoretical considerations

Recall the general form of a generating function (Eq. (3.75)):

$$F_{I_s,K_r}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t) = \sum_{q=0}^{\infty} \sum_{\substack{i_1,\ldots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}!} f^{p,r}_{q,i_1,\ldots,i_{2n}!}(t) y_1^{i_1}\cdots y_{2n}^{i_{2n}}.$$

Definition 3.4.1. (Radius of convergence.) The radius of convergence of the multi-variable series defining F_{I_s,K_r} at t is the real number R_t such that the sum:

$$\sum_{q=0}^{\infty} \left(\sum_{\substack{i_1,\ldots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}!} f^{p,r}_{q,i_1,\ldots,i_{2n}}(t) \right) \eta^q$$

converges absolutely $\forall \eta,\ 0 < \eta < R_t$ and diverges $\forall \eta > R_t$.

The following proposition, whose proof can be found in many textbooks, concerns the normal convergence of the series. Earlier, we used this result for finding the initial conditions to integrate the Hamilton–Jacobi equation.

Proposition 3.4.1. *Let R_t be the radius of convergence of the multi-variable series defining F_{I_s,K_r} at the time t. Then for all $\eta < R_t$ the series converges normally in $\{y \in \mathbb{R}^{2n} : \|y\| \leq \eta\}$ at t.*

The radius of convergence is not appropriate for studying series of functions as it is a function of time. To remove the time dependency, we define the domain of convergence, a domain $\mathcal{D}$ in $\mathbb{R} \times \mathbb{R}^{2n}$ in which the series converge uniformly.

Definition 3.4.2. (Domain of convergence.) The domain of convergence $\mathcal{D}$ is a region in $\mathbb{R} \times \mathbb{R}^{2n}$ in which the series

$$\sum_{q=0}^{\infty} \sum_{\substack{i_1,\ldots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}!} f^{p,r}_{q,i_1,\ldots,i_{2n}}(t) y_1^{i_1}\cdots y_{2n}^{i_{2n}}$$

converges uniformly.

In contrast with the radius of convergence, the domain of convergence is not uniquely defined. The spatial domain depends on the time interval and vice versa. For instance, $\sum_n t^n y^n$ converges if and only if $ty < 1$. $\mathcal{D} = \{(t, y) \in [0, 2] \times [0, 0.5]\}$ and $\mathcal{D} = \{(t, y) \in [0, 0.5] \times [0, 2]\}$ are two well-defined domains of convergence.

In Def. 3.4.2, the uniform convergence of the series is of prime importance. It allows one to bound the error between the true series and its truncation. Indeed, by definition we have:

$$\forall \epsilon > 0,\ \exists N > 0,\ \forall (t, y) \in \mathcal{D},$$
$$F_{I_s,K_r}(q_{I_p}, p_{\bar{I}_p}, q_{0_{K_r}}, p_{0_{\bar{K}_r}}, t) - \sum_{q=0}^{N} \sum_{\substack{i_1,\ldots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}} f^{p,r}_{q,i_1,\ldots,i_{2n}!}(t) y_1^{i_1} \cdots y_{2n}^{i_{2n}} < \epsilon\,. \tag{3.82}$$

In other words, given a domain of convergence and a precision goal ϵ, there exists a positive integer N such that the truncated Taylor series of order N approximates the true function within ϵ in the domain.

3.4.4.2 Practical considerations

In practice, for most of the problems we are interested in, we are only able to compute finitely many terms in the series. As a result, it is impossible to estimate a domain of convergence. Worse, we cannot theoretically guarantee that the generating functions can be expressed as Taylor series. In fact, we have seen earlier that even if the Taylor series of F_{I_s,K_r} converges on some open set and F_{I_s,K_r} is smooth, then F_{I_s,K_r} may not be equal to its Taylor series. One can readily verify that the function $f(x) = \exp(1/x^2)$ if $x \neq 0$, $f(0) = 0$ is smooth and has a converging Taylor series at 0. However, f is not equal to its Taylor series. In the following we make two realistic assumptions in order to develop a practical tool for estimating a domain of convergence.

We first assume that the flow may be expressed as a Taylor series in some open set. This is a very common assumption when studying dynamical systems. For example, we make this hypothesis when we approximate the flow by the state transition matrix at linear order. We noticed in the indirect approach that the generating functions may be computed from the flow at the cost of a series inversion. From the series inversion theory (see e.g., Moulton [25]), we conclude that the generating functions can also be expressed as Taylor series (when they are not singular). Thus, for almost every t, there exists a non-zero radius of convergence. In addition, the concept of domain of convergence is well-defined.

The second assumption we make is also reasonable. We assume that there exists a domain in which the first-order terms of the series defining F_{I_s,K_r} are dominant. In other words, we assume that there exists a domain in which the linear order is the largest, followed by the second order, third order, etc. This is again a very common assumption for dynamical systems. When approximating the flow with the state transition matrix, we implicitly assume that the linear term is dominant. However, in the present case, there is a subtlety due to the presence of singularities. We observe that this assumption no longer holds as we get closer to a singularity. Let us look at an example to illustrate this phenomenon.

Example 3.4.2. The Taylor series in x of $f(x,t) = (1-t)^x$ for $t \in (0,1)$ is

$$\sum_{r=0}^{\infty} a_n x^n, \text{ where } a_n = \frac{\log(1-t)^n}{n!}.$$

Its radius of convergence is $R_t = \infty$ for all $t \in (0,1)$ and it is singular at $t = 1$. In Figure 3.3, we plot the four first terms of the series as a function of x for different times. Clearly, as t gets closer to 1, the first-order terms are less and less dominant. Equivalently, the x-interval in which the first-order terms are dominant shrinks as t goes to 1. In Figure 3.4, we plot $(1-t)^x - \sum_{r=0}^{3} \frac{\log(1-t)^n}{n!} x^n$. One can readily verify that given a prescribed error margin, the domain in which the order 4 approximates f within this margin shrinks as t gets closer to 1. This is a very common behavior that motivates the need for a new criterion.

Suppose that the fourth-order approximation of f is to be used for solving a given problem where the time evolves from 0 to 0.6. We know that such an approximation is relevant if the first-order terms are dominant, i.e., $a_0 > a_1 > a_2 > a_3$. From Figure 3.3, we infer that this condition is satisfied if and only if $\|x\| \leq 1$. We call the domain $\mathcal{D}_u = \{[0,1],[0,0.6]\}$ the domain of use.

Let us formalize the concept of *domain of use*.

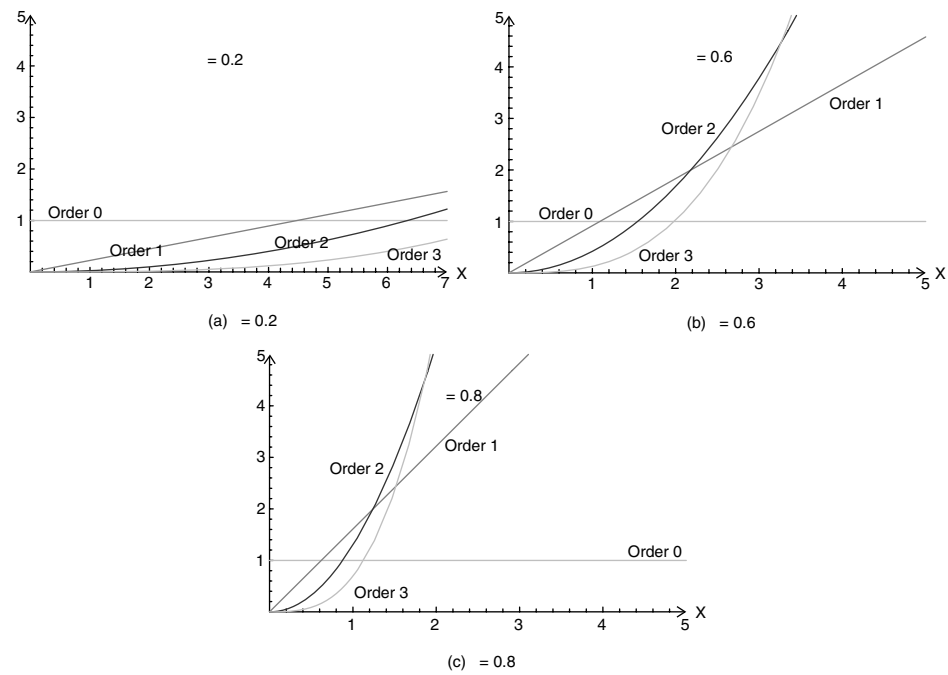


Fig. 3.3. Contribution of the first four terms in the Taylor series of $(1-t)^x$.

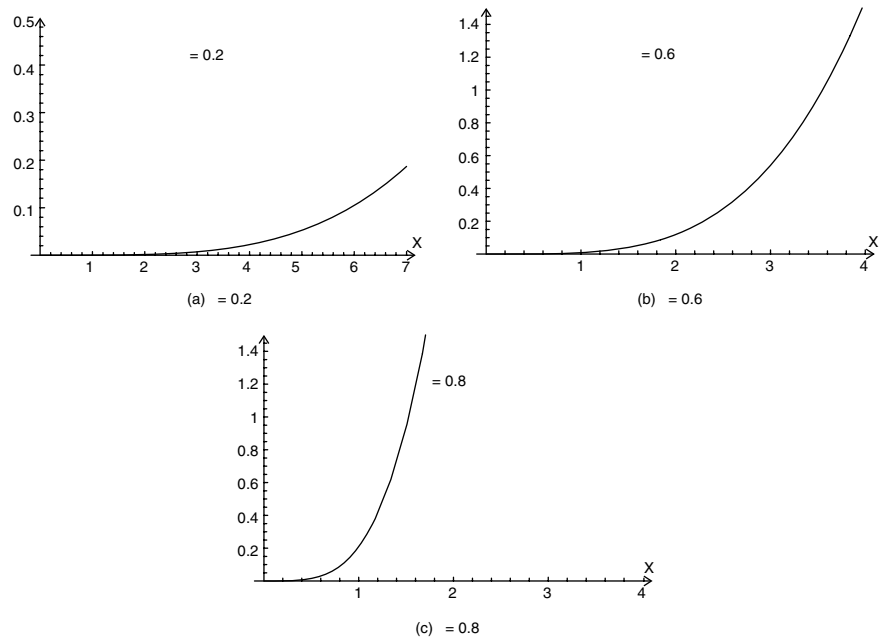


Fig. 3.4. $(1-t)^x - \sum_{r=0}^{3} \frac{\log(1-t)^n}{n!} x^n$.

Definition 3.4.3. (Domain of use.) The domain of use $\mathcal{D}_u$ is a domain in $\mathbb{R} \times \mathbb{R}^{2n}$ in which

$$\left(\sum_{\substack{i_1,\ldots,i_{2n}=0 \\ i_1+\cdots+i_{2n}=q}}^{q} \frac{1}{i_1!\cdots i_{2n}} f^{p,r}_{q,i_1,\ldots,i_{2n}}(t) y_1^{i_1} \cdots y_{2n}^{i_{2n}} \right)_q$$

is a decreasing sequence.

This definition is very conservative but very easy to work with. For a given problem, we identify a time interval (or a spatial domain) in which we want to use the generating functions. Then we compute the spatial domain (or the time interval) in which our solution is valid. Once we have identified the domain of use, one can safely work with the solution within this domain. Let us illustrate the use of the above tool with an example.

3.4.5 Examples

We consider the following fictional space mission: A formation of spacecraft is flying about the Libration point L_2 in the Hill three-body problem and we wish to use F_1 to solve the position to position boundary value problem in order to design a reconfiguration.

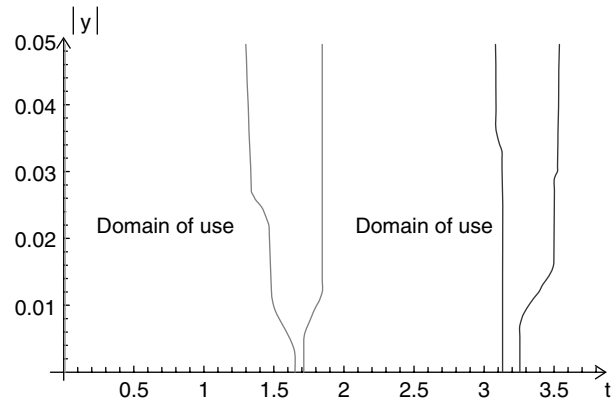


Fig. 3.5. Domain of use.

The mission specifications impose the spacecraft to stay within 0.05 units of length from the equilibrium point L_2 (i.e., 107 500 km in the Earth–Sun system). The normalized Hamiltonian describing the Hill's dynamics is (Appendix B):

$$H(q,p) = \frac{1}{2}(p_x^2 + p_y^2) + (q_y p_x - q_x p_y) - \frac{1}{\sqrt{q_x^2 + q_y^2}} + \frac{1}{2}(q_y^2 - 2q_x^2), \tag{3.83}$$

where $q_x = x$, $q_y = y$, $p_x = \dot{x} - y$, and $p_y = \dot{y} + x$. Using this coordinate system, L_2 is an equilibrium point with coordinates $(q_x, q_y) = (3^{-1/3}, 0)$. To use the above algorithm, H must be expressed as a Taylor series in the spatial variables. Hence, since we want to study the dynamics about L_2, we linearize H about L_2. Then, we use the algorithm to solve F_1 up to order 5 in the time interval $(0, 3.5)$. Using the direct approach, this is equivalent to solving 121 ordinary differential equations. We encounter a number of singularities for F_1 at $t = 0$, $t = 1.68$, and $t = 3.19$. In Figure 3.5, we plot the maximum value of $\|y\|$ so that the first five terms are in decreasing order[8]. We notice that as we get closer to the singularity, the maximum value of $\|y\|$ goes to 0. To find the domain of use, we only need to intersect this plot with $\|y\| = 0.05$.

3.4.5.1 Error in the approximation

We can verify a posteriori that the Taylor series expansion found for the generating function F_1 approximates the true dynamics within this domain. To do so, we again use the example from before and set $q(T) = q_1$ and q_0, and find $p(T) = p_1$ and p_0 from Eqs. (3.28) to (3.29). Then we integrate the trajectory whose initial condition is (q_0, p_0) to find $(q(T), p(T)) = (q_2, p_2)$. The error in the approximation is defined as the norm of

[8] Some terms may change sign and therefore may be very small. In that case we ignore these terms so that the decreasing condition can be satisfied (For instance if the order 2 term goes to 0, it will be smaller than any other terms and therefore must be ignored).

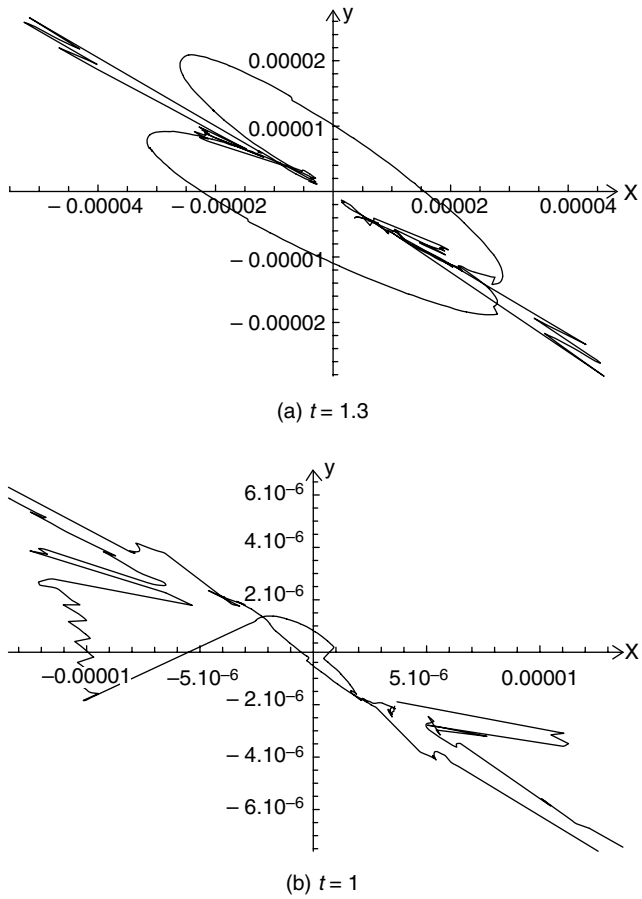


Fig. 3.6. Difference between the true and the approximate dynamics.

$(q_2 - q_1, p_2 - p_1)$. In Figure 3.6 we plot this error for $q_0 = 0$ and q_1 that takes values on the circle centered at L_2 of radius 0.05 for different values of t. The solution is checked at points along the circle, generating the error curves in the figures. We observe that the truncated series provide a good approximation of the true dynamics.

We also point out that since the series is converging and the magnitude of each order decreases in the domain of use, the accuracy must always increase if an additional order is taken into account. In Figure 3.7, we observe that the order two solution provides a poor approximation to the initial momentum because the error ranges up to 4.5×10^{-3} units of length (i.e., 9615 km in the Earth–Sun system). Order three and four give order of magnitude improvements, the error is less than 2.2×10^{-4} units of length (480 km) for order three and less than 3.5×10^{-5} units of length (77 km) for order four, over two orders of magnitude better than the order two solution.

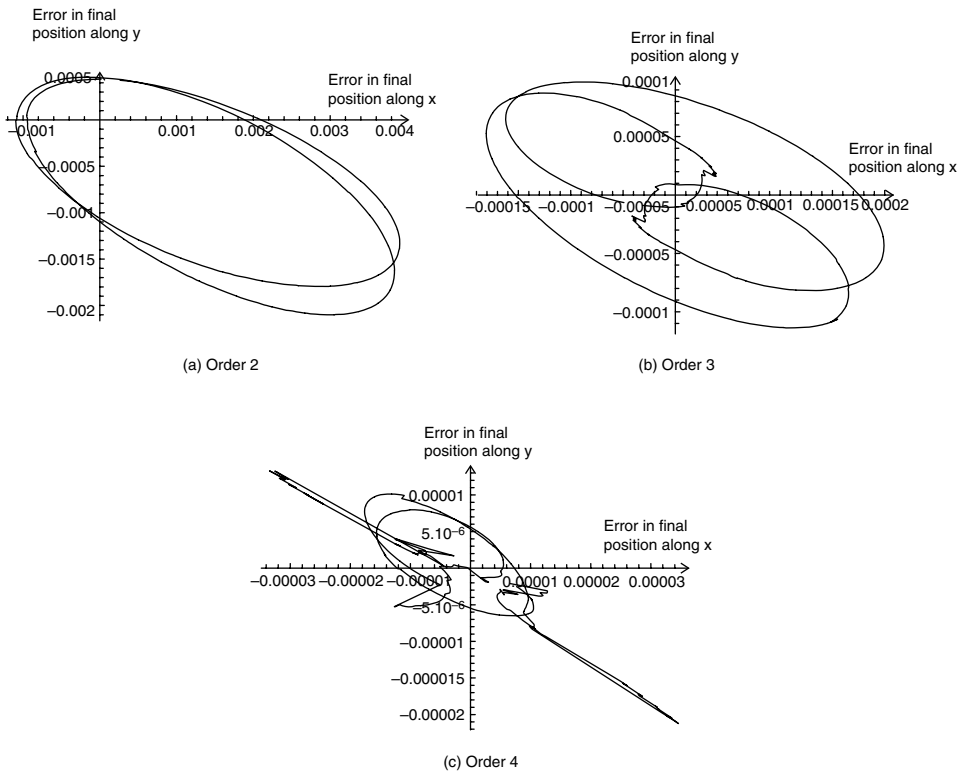


Fig. 3.7. Error in the normalized final position for $t = 0.9$.

3.5 Applications

We now illustrate the use of our novel approach to solving two-point boundary value problems. We consider the normalized Hill three-body problem (see Appendix B). Using our algorithm outlined previously we compute the Taylor series expansion of the generating function F_1 about the Libration point L_2 up to order N (N is determined by the accuracy we wish to achieve). In other words, F_1 can be expressed as a polynomial of order N with time-dependent coefficients whose values are known. The gradient of F_1 is then a polynomial of order $N-1$. Hence, solutions to any position to position boundary value problem (i.e., Lambert's type of boundary value problem) is solved by evaluating a polynomial of order $N-1$. Once F_1 is known, we can solve many problems. In this section, we choose to focus on the study of periodic orbits and on optimal control problems. The study of periodic orbits has been presented in Ref. [15] and the optimal control problems can be found in Ref. [12]. The method has also been applied to the formation flight of spacecraft, we refer the interested reader to Refs. [16–18].

3.5.1 The search for periodic orbits

3.5.1.1 Methodology

To find periodic orbits using the theory we developed above, we need to characterize them as solutions to two-point boundary value problems.

Periodic orbits in a $2n$-dimensional Hamiltonian dynamical system are characterized by the following equations:

$$q(T) = q_0, \tag{3.84}$$

$$p(T) = p_0, \tag{3.85}$$

where T is the period of the orbit, (q_0, p_0) are the initial conditions at time $t_0 = 0$ and $(q(t), p(t))$ verifies Hamilton's equations:

$$\dot{q} = \frac{\partial H}{\partial p}(q, p, t), \quad \dot{p} - \frac{\partial H}{\partial q}(q, p, t)\,. \tag{3.86}$$

In the most general case, the search for periodic orbits consists of solving the $2n$ equations (3.84) and (3.85) for the $2n+1$ unknowns (q_0, p_0, T). Simple methods that solve this problem take a set of initial conditions (q_0, p_0), and integrate Hamilton's equations. If there exists a time $t = T$ such that Eqs. (3.84) and (3.85) are verified, then a periodic orbit is found. Else, other initial conditions need to be guessed. In the approach we propose in this chapter, instead of looking at the initial conditions and the period as the only variables of the problem, we suppose that the period, n initial conditions as well as n components of the state vector at time T are unknowns. Then the search for periodic orbits reduces to solving the $2n$ equations (3.84) and (3.85) for these $2n+1$ unknowns.

For instance, if $(q(T), q_0, T)$ are taken to be the $2n+1$ unknowns, then the search for periodic orbits consists of solving the $2n$ equations (3.84) and (3.85) for $(q(T), q_0, T)$. Let us now find all periodic orbits of a given period. In other words, T is given and we need to find $(q(T), q_0)$ such that $q(T) = q_0$ and $p(T) = p_0$. This is a boundary value problem with constraints that can be solved with the generating function F_1. Combining Eqs. (3.28) and (3.29) and Eqs. (3.84) and (3.85) we obtain:

$$\begin{aligned} p(T) &= \frac{\partial F_1}{\partial q}(q, q_0, T), \qquad q(T) = q_0, \\ p_0 &= -\frac{\partial F_1}{\partial q_0}(q, q_0, T), \qquad p(T) = p_0, \end{aligned} \tag{3.87}$$

i.e.,

$$\frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T) = 0, \tag{3.88}$$

$$p = p_0 = \frac{\partial F_1}{\partial q}(q = q_0, q_0, T)\,. \tag{3.89}$$

Eqs. (3.88) and (3.89) define necessary and sufficient conditions for the existence of periodic orbits. Therefore, the search for all periodic orbits of a given period is reduced to solving n equations (3.88) for n variables, the q_0's, and then evaluate n equations (3.89)

to compute the corresponding momenta. Most importantly, once F_1 is known, solutions of these conditions are computed using algebraic manipulations, no integration is required.

Similarly, if we want to find all periodic orbits going through a given point in the position space, we set q_0 in Eq. (3.88) and solve for T. However, instead of solving the n equations defined by Eq. (3.88) for one variable, T, we may combine them in the following way:

$$\left\| \frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T) \right\| = 0, \tag{3.90}$$

where $\|\cdot\|$ is a norm. This equation can be easily solved numerically or even graphically.

3.5.1.2 Examples

In the following we use a truncated Taylor series of F_1 of order $N = 5$ to study periodic orbits about the Libration point L_2.

First, let us find all periodic orbits going through $q_0 = (0.01, 0)$. To solve this problem, we use the necessary and sufficient condition defined by Eq. (3.90). In Figure 3.8 we plot the left-hand side of Eq. (3.90) as a function of the normalized time. We observe that the norm vanishes only at $t = T = 3.03353$. Therefore, there exists only one periodic orbit going through q_0, and its period is T (there may be additional periodic orbits of period $T > 3.2$, but we cannot see them in this figure). Again, these results are in agreement with known results on periodic orbits about L_2. One can show that any point in the vicinity of L_2 belongs to a periodic orbit. The periods of these orbits increase as their distances from L_2 increase. In the limit, as the distance between periodic orbits and L_2 goes to zero, the period goes to $T = T_{linear} = 3.033019$.

Another problem is to find all periodic orbits of a given period $T = 3.0345$. To solve this problem, we use Eq. (3.88) which defines two equations with two unknowns that can be solved graphically. In Figure 3.9, we plot the solutions to each of these two equations and then superimpose them to find their intersection. The intersection corresponds to

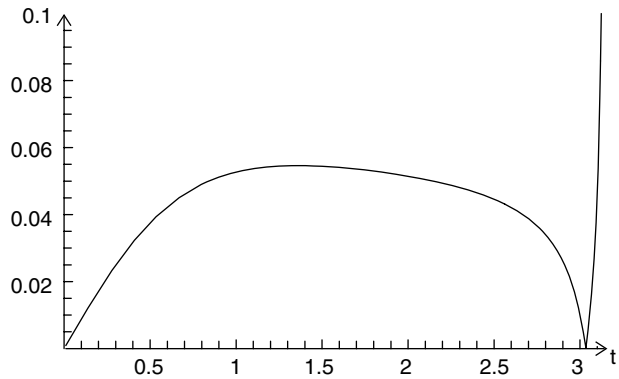
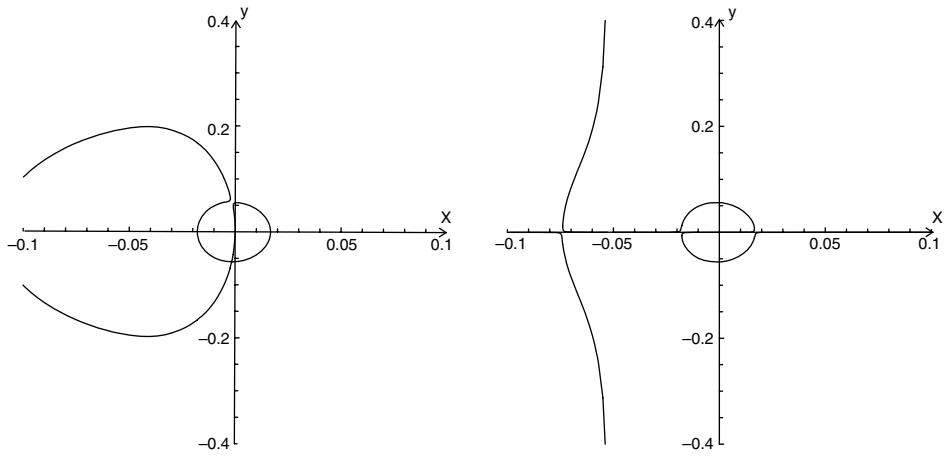
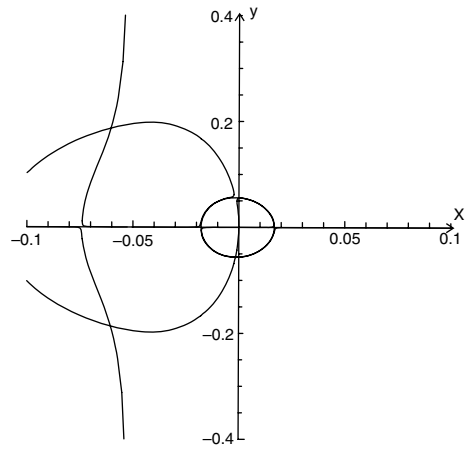


Fig. 3.8. Plot of $\| \frac{\partial F_1}{\partial q}(q = q_0, q_0, T) + \frac{\partial F_1}{\partial q_0}(q = q_0, q_0, T) \|$ where $q_0 = (0.01, 0)$.



(a) Plot of the solution to the first equation defined by Eq. (3.88)

(b) Plot of the solution to the second equation defined by Eq. (3.88)

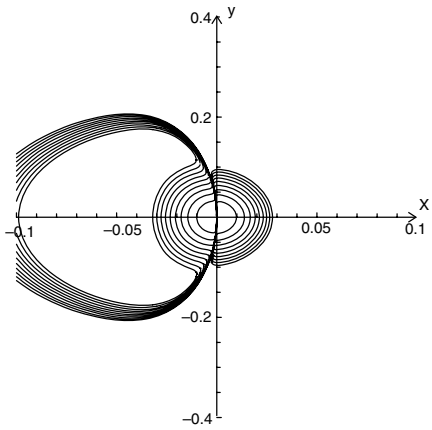


(c) Superposition of the two sets of solutions

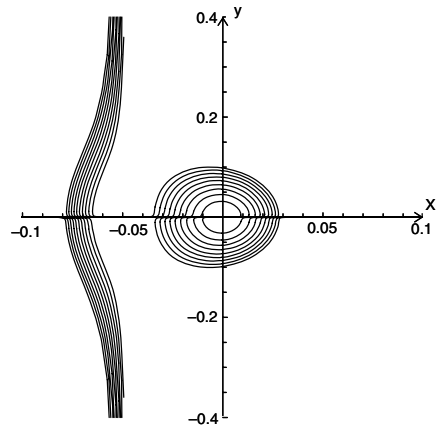
Fig. 3.9. Periodic orbits for the non-linear motion about a Libration point.

solutions to Eq. (3.90), i.e., to the set of points that belongs to periodic orbits of period T. We observe that the intersection is composed of a circle and two points whose coordinates are $(q_x, q_y) = (-0.0603795, \pm 0.187281)$. The circle is obviously a periodic orbit but the two points are not equilibrium points, and rather correspond to out-of-plane periodic orbits[9].

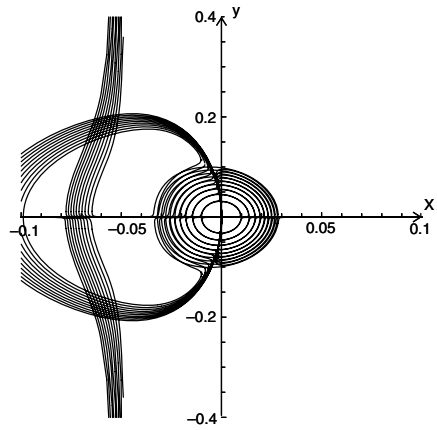
[9] In the Hill three-body problem these out-of-plane orbits do not exist. At that distance of L_2, our approximation of the dynamics of order 5 is no more valid, therefore these two points do not have any physical meaning. In practice, we can evaluate a domain in which an approximation of order N of the dynamics is valid. We refer to Section 3.4.4 and Ref. [12] for more details.



(a) Plot of the solution to the first equation defined by Eq. (3.88) for = 3.033+0.0005 ∈ {1...10}



(b) Plot of the solution to the second equation defined by Eq. (3.88) for = 3.033+0.0005 ∈ {1...10}



(c) Superposition of the two sets of solutions for = 3.033 + 0.0005 ∈ {1...10}

Fig. 3.10. Periodic orbits for the non-linear motion about a Libration point.

By plotting the intersection for different periods T, we generate a family of periodic orbits around the Libration point. In Figure 3.10 we represent the solutions to Eq. (3.88) for $t = 3.033 + 0.0005n,\ n \in \{1 \cdots 10\}$. For $t = 3.033$ (which is less that the periodic orbit period of the linearized system), the intersection only contains the origin, which is why there are only 9 periodic orbits shown around the origin. We note that at larger values of $x^2 + y^2$ the curves do not overlay precisely, indicating that higher order terms are needed.

The method we propose allows us to search for periodic orbits anywhere in the phase space or in the time domain without requiring any initial guess or knowledge of a periodic orbit that belongs to the family. This is a major advantage compared to traditional methods.

Most important, we reduce the search for periodic orbits to solving a non-linear system of equations. Once the generating functions are known, no integration is required to find periodic orbits of different periods and/or going through different points in the phase space. This is a fundamental property of the generating functions; once the generating functions are known, any two-point boundary value problem can be solved at the cost of a single function evaluation.

3.5.2 Optimal control and mission planning

To illustrate the use of the generating functions to solve non-linear optimal control problems we now consider a targeting problem in the two-dimensional Hill three-body problem (Appendix B). We consider a spacecraft away from the Libration point L_2 and want to find the control sequence that moves the spacecraft at the equilibrium point L_2 while minimizing the fuel consumption. Specifically, this optimal control problem formulates as follows:

We want to minimize the cost function $J = \frac{1}{2}\int_{t=0}^{t=t_f} \left(u_x^2 + u_y^2\right) \mathrm{d}t$ subject to the constraints

$$\begin{cases} \dot{x_1} = x_3, \\ \dot{x_2} = x_4, \\ \dot{x_3} = 2x_4 - \frac{x_1}{(x_1^2+x_2^2)^{3/2}} + 3x_1 + u_x, \\ \dot{x_4} = -2x_3 - \frac{x_2}{(x_1^2+x_2^2)^{3/2}} + u_y, \end{cases} \tag{3.91}$$

and the boundary conditions:

$$X(t=0) = X_0, \quad X(t=t_f) = X_{L_2} = (3^{-1/3}, 0, 0, 0), \tag{3.92}$$

where $r^2 = x^2 + y^2$ and $X = (x_1, x_2, x_3, x_4) = (x, y, \dot{x}, \dot{y})$. Define the Hamiltonian:

$$\begin{aligned} H(X, P, U) =& p_1 x_3 + p_2 x_4 + p_3 \left(2x_4 - \frac{x_1}{(x_1^2+x_2^2)^{3/2}} + 3x_1 + u_x\right) \\ &+ p_4 \left(-2x_3 - \frac{x_2}{(x_1^2+x_2^2)^{3/2}} + u_y\right) + \frac{1}{2}u_x^2 + \frac{1}{2}u_y^2, \end{aligned}$$

where $P = (p_1, p_2, p_3, p_4)$ and $U = (u_x, u_y)$. Then, from $\frac{\partial H}{\partial U} = 0$, we find the optimal control feedback law:

$$u_x = -p_3, \quad u_y = -p_4\,.$$

Substituting $U = (u_x, u_y)$ into H yields:

$$\begin{aligned} \bar{H}(X, P) =& p_1 x_1 + p_2 x_2 + p_3 \left(2x_4 - \frac{x_1}{(x_1^2+x^2)^{3/2}} + 3x_1 - p_3\right) \\ &+ p_4 \left(-2x_3 - \frac{x_2}{(x_1^2+x^2)^{3/2}} - p_4\right) + \frac{1}{2}p_3^2 + \frac{1}{2}p_4^2\,. \end{aligned} \tag{3.93}$$

We deduce the necessary conditions for optimality:

$$\dot{X} = \frac{\partial \bar{H}}{\partial P}, \tag{3.94}$$

$$\dot{P} = -\frac{\partial \bar{H}}{\partial X}, \tag{3.95}$$

$$X(t=0) = X_0, \quad X(t=t_f) = (0,0,0,0).$$

This is a position to position boundary value problem that can be solved using F_1. In this example, we compute F_1 at order 4 and use this approximation together with Eqs. (3.28) and (3.29) to find the value of the co-state P at the initial and final times. Then, the optimal trajectory is found by integrating Hamilton's equations (Eqs. (3.94) and (3.95)).

In Figure 3.11, we plot the trajectories for different final times. As t_f increases, the trajectory tends to wrap around the Libration point so that the spacecraft takes advantage of the geometry of the Libration point (Appendix B). On the other hand, if the transfer time is small, the trajectory is almost a straight line, it completely ignores the dynamics. In Figure 3.12 the associated control laws are represented. As expected, the longer the transfer time is, the smaller the magnitude of the control. We emphasize that we only need to evaluate the gradient of F_1 (which is a polynomial of order 3) seven times and integrate Eqs. (3.94) and (3.95) seven times to obtain the seven curves in Figure 3.11. Similarly, in Figure 3.13, at the cost of sixteen evaluations of the gradient of F_1, we are able to represent the optimal trajectories of spacecraft starting at $X_0 = (r\cos(\theta), r\sin(\theta))$ where $r = 10\,700$ km and $\theta = k\pi/8$, and ending at L_2 in 145 days. In Figure 3.14 the corresponding optimal control law is represented.

Further, if different types of boundary conditions are imposed (for instance, the terminal state is not fully specified) then we need to perform a Legendre transform to find the generating function that solves this new boundary value problem. There is no need

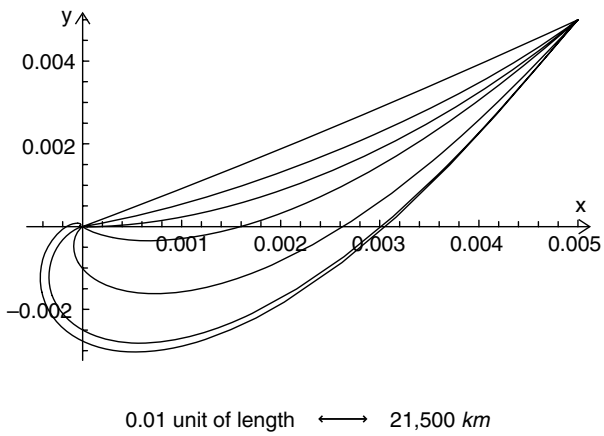
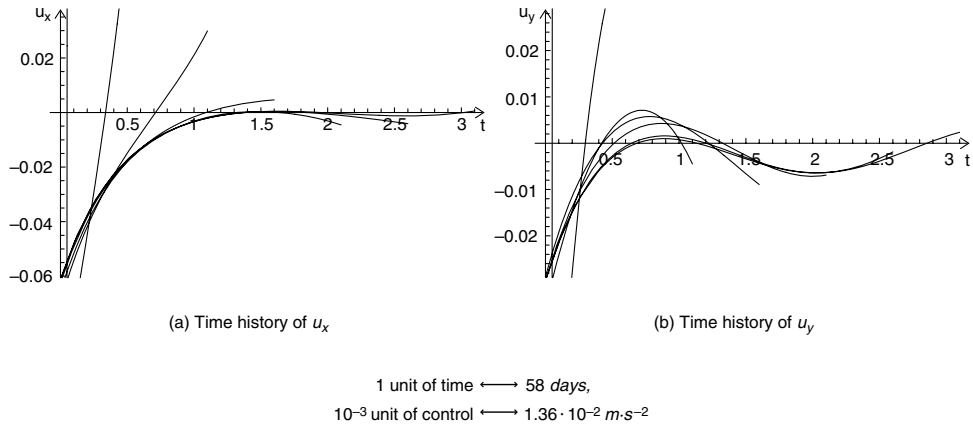


Fig. 3.11. Optimal trajectories of the spacecraft for different transfer times.

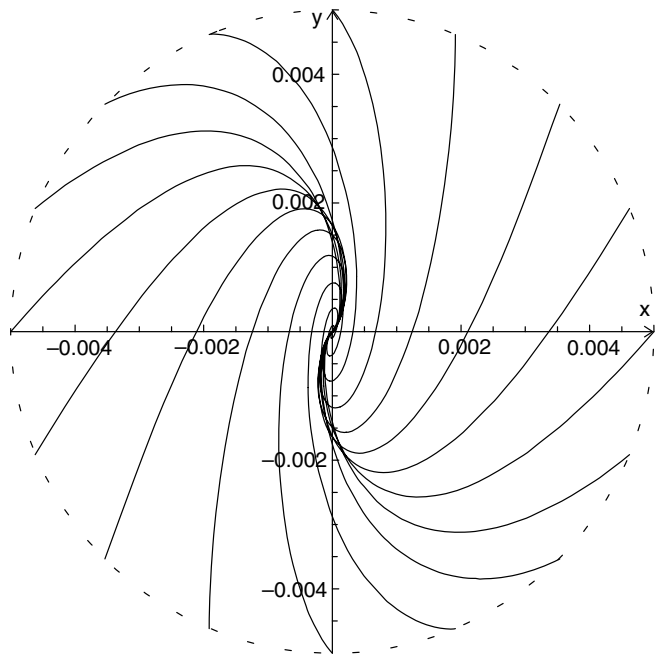


(a) Time history of u_x

(b) Time history of u_y

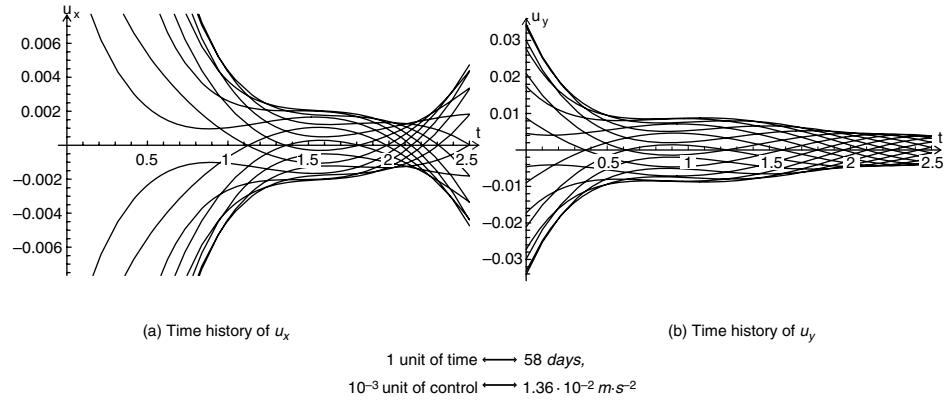
1 unit of time $\longleftrightarrow$ 58 *days*,
10^{-3} unit of control $\longleftrightarrow$ $1.36 \cdot 10^{-2}$ $m{\cdot}s^{-2}$

Fig. 3.12. Time history of the control laws.



0.01 unit of length $\leftrightarrow$ 21,500 *km*

Fig. 3.13. Optimal trajectories of the spacecraft as a function of the initial position.



(a) Time history of u_x (b) Time history of u_y

1 unit of time ⟷ 58 *days*,

10^{-3} unit of control ⟷ $1.36 \cdot 10^{-2}$ $m \cdot s^{-2}$

Fig. 3.14. Time history of the control laws.

to resolve the optimal control problem. This is an intrinsic property of the generating functions that opens the doors to truly reconfigurable optimal control. In [27, 28] the application of generating functions to solve optimal control problems is studied in detail, and a number of strong results are discovered for the necessary and sufficient conditions for optimal control.

Finally, we point out that this method is very efficient for mission planning. By varying the transfer time for instance, we can find times for which the optimal transfer requires less fuel expenditure. For more details on this topic we refer to Ref. [16]. Another major application of the theory presented in this chapter is spacecraft formation design. We noticed in the introduction that the reconfiguration of a formation of N spacecraft requires us to solve $N!$ boundary value problems. Using the generating functions this task can be achieved at the cost of $N!$ function evaluations. This problem and others relating to formation flight are solved in Refs. [17, 18].

3.6 Conclusions

The method we develop in this chapter is based on the Hamilton–Jacobi theory. We have observed that the generating functions associated with the phase flow readily solve any Hamiltonian two-point boundary value problems. This observation has many consequences that we now re-state. Above all, it provides a very general methodology for solving boundary value problems for Hamiltonian systems. Whereas traditional methods solve boundary value problems about an initial guess only, our approach gives a "full picture". In particular, traditional methods completely ignore the number of solutions to the boundary value problem. Our approach, however, indicates the presence of multiple solutions as singularities of generating functions. In turn, we proved and illustrated that these singularities can be studied and the number of solutions may be determined.

In linear systems theory, it is well-known that perturbation matrices solve boundary value problems. These matrices have distinctive properties that are studied in the literature.

Using generating functions we have recovered and extended some of these properties. Most importantly, we have proved that they correspond to coefficients of the generating functions. As a result, our approach naturally contains the theory of perturbation matrices. The relation between perturbation matrices and generating functions may also be investigated using the state transition matrix. In this respect, we have shown that state transition matrix and generating functions are closely related. One of the main consequences, is that we can predict singularities of the generating functions using the state transition matrix. This result broadens to some extent to non-linear systems.

In non-linear systems theory, there is no equivalent of the perturbation matrices. Thus, the approach we have proposed is the first to define functions, namely the generating functions, that directly relate boundary conditions. Obviously, no results as general as the ones derived for linear systems may be gleaned in this case. However, for polynomial generating functions we have established that singularities of the generating functions may still be predicted from the state transition matrix. As a result, the existence of multiple solutions to two-point boundary value problems is fully predicted by the linear dynamics. The number of solutions, however, depends on the non-linear dynamics.

The solution of the Hamilton–Jacobi equation for a generating function is extremely difficult in general. Thus, to make our observations and theory realizable it is essential that we be able to construct solutions for some general class of problems. In this chapter we detail an explicit set of algorithms that allows us to do so, found by performing a Taylor series expansion of the generating function and the Hamiltonian function about a nominal solution to the dynamical system. This provides a solution for the generating function which is analytic in its spatial variables, and can incorporate high-order non-linearities, where the coefficients of the expansion are found from numerically integrating a series of ordinary differential equations in time. Through some specific examples we demonstrate some of the convergence properties of this algorithm and establish that it provides an accurate representation of the generating function in the neighborhood of a solution.

Finally, the approach we have presented to solve two-point boundary value problems applies to any Hamiltonian systems. It is therefore not surprising that it has implications in several fields. In particular, it allows us to develop new methods to study the phase space structure, solve optimal control problems and design spacecraft formations.

Appendix A. The Hamilton–Jacobi equation at higher orders

In this appendix, we give an explicit expression of P as defined by Eq. (3.76). We assume a $2n$-dimensional Hamiltonian system with polynomial Hamiltonian function and polynomial generating functions. We have seen that the Hamilton–Jacobi partial differential equation reduces to an ordinary differential equation of the form

$$P(y, f^{p,r}_{i_1,\ldots,i_{2n}}(t), \dot{f}^{p,r}_{i_1,\ldots,i_{2n}}(t)) = 0. \tag{A.1}$$

In the following we use tensor notation in order to derive an explicit expression of P. In tensor notation, a Taylor series expansion writes as:

$$f(x,t) = f^0(t) + f^1(t)\cdot\vec{x} + (f^2(t)\cdot\vec{x})\cdot\vec{x} + ((f^3(t)\cdot\vec{x})\cdot\vec{x})\cdot\vec{x} + \cdots. \tag{A.2}$$

Applying this formula to $H(\vec{x}, t)$ and $F_2 = F(\vec{y}, t)$ yields:

$$H(\vec{x}) = h_{i,j}(t)x_i x_j + h_{i,j,k}(t)x_i x_j x_k + \cdots, \tag{A.3}$$

$$F(\vec{y}) = f_{i,j}(t)y_i y_j + f_{i,j,k}(t)y_i y_j y_k + \cdots, \tag{A.4}$$

where we assume the summation convention. Let us now express $\vec{x} = (\Delta q, \Delta p)$ as a function of $\vec{y} = (\Delta q, \Delta p_0)$ (we drop the time dependence in the notation, i.e., we shall write $h_{i,j}$ instead of $h_{i,j}(t)$). For all $a \leq n$ and $j = n + a$

$$x_a = y_a, \tag{A.5}$$

$$x_j = \frac{\partial F}{\partial y_a} \tag{A.6}$$

$$= f_{a,k}y_k + f_{k,a}y_k + f_{a,k,l}y_k y_l + f_{k,a,l}y_k y_l + f_{k,l,a}y_k y_l + \cdots, \tag{A.7}$$

where n is the dimension of the configuration space. The Hamilton–Jacobi equation becomes:

$$\dot{f}_{i,j}y_i y_j + \dot{f}_{i,j,k}y_i y_j y_k + \cdots + h_{i,j}x_i x_j + h_{i,j,k}x_i x_j x_k + \cdots = 0\,. \tag{A.8}$$

Replacing $\vec{x}$ by $\vec{y}$ in Eq. (A.8) using Eq. (A.7), and keeping only terms of order less than 3 yields:

$$\begin{aligned}
0 = {} & \dot{f}_{i,j}y_i y_j + \dot{f}_{i,j,k}y_i y_j y_k + h_{a,b}y_a y_b + h_{a,b,c}y_a y_b y_c \\
& + (h_{a,n+b} + h_{n+b,a})y_a(f_{b,k}y_k + f_{k,b}y_k + f_{b,k,l}y_k y_l + f_{l,b,k}y_k y_l + f_{k,l,b}y_k y_l) \\
& + h_{n+a,n+b}(f_{a,k}y_k + f_{k,a}y_k + f_{a,k,l}y_k y_l + f_{l,a,k}y_k y_l + f_{k,l,a}y_k y_l) \\
& \times (f_{b,m}y_m + f_{m,b}y_m + f_{b,m,p}y_m y_p + f_{p,b,m}y_m y_p + f_{m,p,b}y_m y_p) \\
& + (h_{n+a,b,c} + h_{c,n+a,b} + h_{b,c,n+a})y_b y_c(f_{a,k}y_k + f_{k,a}y_k) \\
& + (h_{n+a,n+b,c} + h_{n+b,c,n+a} + h_{c,n+a,n+b})y_c(f_{a,k}y_k + f_{k,a}y_k)(f_{b,l}y_l + f_{l,b}y_l) \\
& + h_{n+a,n+b,n+c}(f_{a,k}y_k + f_{k,a}y_k)(f_{b,l}y_l + f_{l,b}y_l)(f_{c,m}y_m + f_{m,c}y_m).
\end{aligned} \tag{A.9}$$

Eq. (A.9) is the expression of P up to order 3 as defined by Eq. (3.76). It is a polynomial equation in the y_i variables with time-dependent coefficients and holds if every coefficient is zero. The equations of order 3 reads:

$$\begin{aligned}
& \dot{f}_{i,j,k}y_i y_j y_k + (A_{i,j,k} + B_{i,j,k} + C_{i,j,k})y_i y_j y_k + (D_{a,i,j} + E_{a,i,j})y_a y_i y_j \\
& \quad + G_{a,b,i}y_a y_b y_i + h_{a,b,c}y_a y_b y_c = 0,
\end{aligned} \tag{A.10}$$

where

$$\begin{aligned}
A_{i,j,k} &= h_{n+a,n+b,n+c}(f_{a,i} + f_{i,a})(f_{b,j} + f_{j,b})(f_{c,k} + f_{k,c}), \\
B_{i,j,k} &= h_{n+a,n+b}(f_{a,i} + f_{i,a})(f_{b,j,k} + f_{j,k,b} + f_{k,b,j}), \\
C_{i,j,k} &= h_{n+a,n+b}(f_{b,i} + f_{i,b})(f_{a,j,k} + f_{j,k,a} + f_{k,a,j}),
\end{aligned}$$

$$D_{a,i,j} = (h_{a,n+b,n+c} + h_{n+c,a,n+b} + h_{n+b,n+c,a})(f_{b,i} + f_{i,b})(f_{c,j} + f_{j,c}),$$
$$E_{a,i,j} = (h_{a,n+b} + h_{n+b,a})(f_{b,i,j} + f_{j,b,i} + f_{i,j,b}),$$
$$G_{a,b,i} = (h_{a,b,n+c} + h_{b,n+c,a} + h_{n+c,a,b})(f_{c,i} + f_{i,c})\,. \tag{A.11}$$

We deduce the coefficients of $y_i y_j y_k$:

- Coefficients of $y_{i\le n}^3$

$$A_{i,i,i} + B_{i,i,i} + C_{i,i,i} + D_{i,i,i} + E_{i,i,i} + \dot{f}_{i,i,i} + G_{i,i,i} + h_{i,i,i} = 0\,. \tag{A.12}$$

- Coefficients of $y_{i>n}^3$

$$A_{i,i,i} + B_{i,i,i} + C_{i,i,i} + \dot{f}_{i,i,i} = 0\,. \tag{A.13}$$

- Coefficients of $y_{i\le n}^2 y_{j\le n}$

$$(A + B + C + D + E + \dot{f} + G + h)_{\tau(i,i,j)} = 0\,, \tag{A.14}$$

where $\tau(i, j, k)$ represents all the distinct permutations of (i, j, k), that is

$$A_{\tau(i,j,k),l} = A_{i,j,k,l} + A_{i,k,j,l} + A_{k,i,j,l} + A_{k,j,i,l} + A_{j,k,i,l} + A_{j,i,k,l}$$

but

$$A_{\tau(i,i,j),l} = A_{i,i,j,l} + A_{i,j,i,l} + A_{j,i,i,l}\,.$$

- Coefficients of $y_{i\le n}^2 y_{j>n}$

$$(A + B + C + \dot{f})_{\tau(i,i,j)} + (D + E)_{i,\tau(i,j)} + G_{i,i,j} = 0\,. \tag{A.15}$$

- Coefficients of $y_{i\le n} y_{j\le n} y_{k\le n}$:

$$(A + B + C + D + E + \dot{f} + G + h)_{\tau(i,j,k)} = 0\,. \tag{A.16}$$

- Coefficients of $y_{i\le n} y_{j\le n} y_{k>n}$

$$(A + B + C + \dot{f})_{\tau(i,j,k)} + (D + E)_{i,\tau(j,k)} + (D + E)_{j,\tau(i,k)} + G_{\tau(i,j),k} = 0\,. \tag{A.17}$$

- Coefficients of $y_{i>n}^2 y_{j\le n}$

$$(A + B + C + \dot{f})_{\tau(i,i,j)} + (E + D)_{j,i,i} = 0\,. \tag{A.18}$$

- Coefficients of $y_{i>n}^2 y_{j>n}$

$$(A + B + C + \dot{f})_{\tau(i,i,j)} = 0\,. \tag{A.19}$$

- Coefficients of $y_{i\le n} y_{j>n} y_{k>n}$

$$(A + B + C + \dot{f})_{\tau(i,j,k)} + (D + E)_{i,\tau(j,k)} = 0\,. \tag{A.20}$$

- Coefficients of $y_{i>n} y_{j>n} y_{k>n}$

$$(A + B + C + \dot{f})_{\tau(i,j,k)} = 0\,. \tag{A.21}$$

Eqs. (A.12)–(A.21) allow us to solve for F_2 (and F_1 since they both verify the same Hamilton-Jacobi equation, only the initial conditions being different). The process of deriving equations for the generating functions can be continued to arbitrarily high order using a symbolic manipulation program (we have implemented and solved the expansion to order 8 using *Mathematica*©).

Appendix B: The Hill three-body problem

The three-body problem describes the motion of three-point mass particles under their mutual gravitational interactions. This is a classical problem that covers a large range of situations in astrodynamics. An instance of such situations is the motion of the Moon about the Earth under the influence of the Sun. However, this problem does not have a general solution and thus we usually consider simplified formulations justified by physical reasoning. In this chapter, we consider three simplifications. We assume that:

1. One of the three bodies has negligible mass compared to the other two bodies (for instance a spacecraft under the influence of the Sun and the Earth).
2. One of the two massive bodies is in circular orbit about the other one.
3. One of the two massive bodies has larger mass than the other one (the Sun compared to the Earth for instance).

Under these three assumptions, the Hamiltonian for this system reads:

$$H(q,p) = \frac{1}{2}(p_x^2 + p_y^2) + (q_y p_x - q_x p_y) - \frac{1}{\sqrt{q_x^2 + q_y^2}} + \frac{1}{2}(q_y^2 - 2q_x^2), \tag{B.1}$$

and the equations of motion become:

$$\begin{cases} \dot{q}_x = p_x + q_y, \\ \dot{q}_y = p_y - q_x, \\ \dot{p}_x = p_y + 2q_x - \frac{q_x}{(q_x^2+q_y^2)^{3/2}}, \\ \dot{p}_y = -p_x - q_y - \frac{q_y}{(q_x^2+q_y^2)^{3/2}}, \end{cases} \tag{B.2}$$

where $q_x = x$, $q_y = y$, $p_x = \dot{x} - y$ and $p_y = \dot{y} + x$.

This problem has two equilibrium points, L_1 and L_2 whose coordinates are

$$L_1\left(-\left(\frac{1}{3}\right)^{1/3}, 0\right) \quad \text{and} \quad L_2\left(\left(\frac{1}{3}\right)^{1/3}, 0\right).$$

Using linear systems theory, one can prove that the libration points have a stable, an unstable and two center manifolds (Figure 3.15).

To study the relative motion of a spacecraft about L_2, we use Eq. (3.73) to compute H^h, the Hamiltonian function describing the relative motion dynamics.

$$H^h = \frac{1}{2} X^{h^T} \begin{pmatrix} H_{qq}(t) & H_{qp}(t) \\ H_{pq}(t) & H_{pp}(t) \end{pmatrix} X^h + \cdots, \tag{B.3}$$

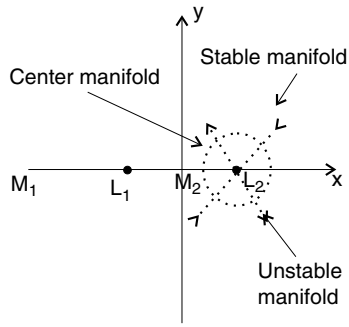


Fig. 3.15. The Libration points in the Hill three-body problem.

where $X^h = \begin{pmatrix} q - q_0 \\ p - p_0 \end{pmatrix} = \begin{pmatrix} \Delta q_x \\ \Delta q_y \\ \Delta p_x \\ \Delta p_y \end{pmatrix}$, $(q_0, p_0) = \left(\left(\frac{1}{3}\right)^{1/3}, 0, 0, \left(\frac{1}{3}\right)^{1/3} \right)$ refers to the state at the equilibrium point L_2 and,

$$H_{qq}(t) = \begin{pmatrix} \frac{1}{(q_{0x}^2+q_{0y}^2)^{3/2}} - \frac{3q_{0x}^2}{(q_{0x}^2+q_{0y}^2)^{5/2}} - 2 & -\frac{3q_{0x}q_{0y}}{(q_{0x}^2+q_{0y}^2)^{5/2}} \\ -\frac{3q_{0x}q_{0y}}{(q_{0x}^2+q_{0y}^2)^{5/2}} & \frac{1}{(q_{0x}^2+q_{0y}^2)^{3/2}} - \frac{3q_{0y}^2}{(q_{0x}^2+q_{0y}^2)^{5/2}} + 1 \end{pmatrix}, \tag{B.4}$$

$$H_{qp}(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{B.5}$$

$$H_{pq}(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{B.6}$$

$$H_{pp}(t) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{B.7}$$

Substituting (q_0, p_0) by its value yields the expression of H^h at second order:

$$H^h = \frac{1}{2} \begin{pmatrix} \Delta q_x & \Delta q_y & \Delta p_x & \Delta p_y \end{pmatrix} \begin{pmatrix} -8 & 0 & 0 & -1 \\ 0 & 4 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta q_x \\ \Delta q_y \\ \Delta p_x \\ \Delta p_y \end{pmatrix}. \tag{B.8}$$

At higher order, we find:

$$H^h = \frac{1}{2} \begin{pmatrix} \Delta q_x & \Delta q_y & \Delta p_x & \Delta p_y \end{pmatrix} \begin{pmatrix} -8 & 0 & 0 & -1 \\ 0 & 4 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta q_x \\ \Delta q_y \\ \Delta p_x \\ \Delta p_y \end{pmatrix}$$
$$+ 3^{4/3}\Delta q_x^3 - \frac{3^{7/3}}{2}\Delta q_x \Delta q_y^2 - 3^{5/3}\Delta q_x^4 + 3^{8/3}\Delta q_x^2 \Delta q_y^2 - \frac{3^{8/3}}{8}\Delta q_y^4 \cdots \tag{B.9}$$

We point out that H^h is time-independent.

Finally, we give in the following table the values of the normalized variables for the Earth–Sun system.

Normalized units		Earth–Sun system
0.01 unit of length	$\longleftrightarrow$	21,500 km
1 unit of time	$\longleftrightarrow$	58 days 2 hours
1 unit of velocity	$\longleftrightarrow$	428 m/s
1 unit of acceleration	$\longleftrightarrow$	$1.38 \cdot 10^{-5}\,\mathrm{m/s^2}$

References

1. Abraham Ralph, and Marsden, Jerrold E. (1978). Foundations of Mechanics (W.A. Benjamin, ed.) 2nd edition, 1978.
2. Alfriend, K.T., Yan, Hui and Valadi, S.R. (2002). Nonlinear considerations in satellite formation flying. In *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit*. AIAA.
3. Arnold, Vladimir I. (1988). Mathematical Methods of Classical Mechanics. Springer-Verlag, 2nd edition.
4. Battin, Richard H. (1999). An Introduction to the Mathematics and Methods of Astrodynamics. American Institute of Aeronautics and Astronautics, revised edition.
5. Betts, John T. (1998). Survey of numerical methods for trajectory optimization. *Journal of Control, Guidance, and Dynamics*, **21**(2), pp. 193–207.
6. Bloch, A.M., Baillieul, J., Crouch, P.E. and Marsden, J.E. (2003). Nonholonomic Mechanics and Control. Springer.
7. Bryson, Arthur E. and Ho, Yu-Chi (1975). Applied Optimal Control: Optimization, estimation, and control. Halsted Press, revised edition.
8. Ehlers, Juergen and Newman, Ezra T. (2000). The theory of caustics and wavefront singularities with physical applications. *Journal of Mathematical Physics A*, **41**(6), pp. 3344–3378.
9. Goldstein, Herbert (1965). Classical Mechanics. Addison-Wesley.
10. Goldstein, Herbert (1980). Classical Mechanics. Addison-Wesley, 2nd edition.
11. Greenwood, Donald T. (1977). Classical Dynamics. Prentice-Hall.
12. Guibout, Vincent M. (2004). The Hamilton–Jacobi theory for solving two-point boundary value problems: Theory and Numerics with application to spacecraft formation flight, optimal control and the study of phase space structure. PhD thesis, University of Michigan.
13. Guibout, Vincent M. and Bloch, Anthony M. (2004). Discrete variational principles and Hamilton-Jacobi theory for mechanical systems and optimal control problems. *Physica D*, submitted.
14. Guibout, Vincent M. and Scheeres, Daniel J. (2002). Formation flight with generating functions: Solving the relative boundary value problem. In *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Monterey, California. Paper AIAA 2002-4639*. AIAA.
15. Guibout, Vincent M. and Scheeres, Daniel J. (2003). Periodic orbits from generating functions. In *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference and Exhibit, Big Sky, Montana. Paper AAS 03-566*. AAS.
16. Guibout, Vincent M. and Scheeres, Daniel J. (2003). Solving relative two-point boundary value problems: Spacecraft formation flight transfers application. *AIAA, Journal of Control, Guidance and Dynamics*, **27**(4), 693–704.
17. Guibout, Vincent M. and Scheeres, Daniel J. (2004). Spacecraft formation dynamics and design. In *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Providence, Rhode Island.*
18. Guibout, V.M. and Scheeres, D.J. (2006). Spacecraft formation dynamics and design. *Journal of Guidance, Control, and Dynamics*, **29**(1), 121–133.

19. Hamilton, William Rowan (1834). On a general method in dynamics. *Philosophical Transactions of the Royal Society, Part II*, pp. 247–308.
20. Hamilton, William Rowan (1835). Second essay on a general method in dynamics. *Philosophical Transactions of the Royal Society, Part I*, pp. 95–144.
21. Keller, H.B. (1968). Numerical Methods for Two-point Boundary Value Problems. Blaisdell.
22. Lanczos, Cornelius (1977). The variational principles of mechanics. University of Toronto Press, 4th edition.
23. Marsden, Jerrold E. and Ratiu, Tudor S. (1998). Introduction to Mechanics and Symmetry: A basic exposition of classical mechanical systems. Springer-Verlag, 2nd edition.
24. Melton, Robert G. (2000). Time-explicit representation of relative motion between elliptical orbits. *Journal of Guidance, Control, and Dynamics*, **23**, 604–610.
25. Moulton, Forest R. (1930). Differential equations. The Macmillan company.
26. Park, Chandeok, Guibout, Vincent M. and Scheeres, Daniel J. (2006). Solving optimal continuous thrust rendezvous problems with generating functions. *Journal of Guidance, Control, and Dynamics*, **29**(2), 321–331.
27. Park, Chandeok and Scheeres, Daniel J. (2003). Indirect solutions of the optimal feedback control using Hamiltonian dynamics and generating functions. In *Proceedings of the 2003 IEEE conference on Decision and Control, 2003. Maui, Hawaii*. IEEE.
28. Park, Chandeok and Scheeres, Daniel J. (2004). Solutions of optimal feedback control problems with general boundary conditions using Hamiltonian dynamics and generating functions. In *Proceedings of the American Control Conference, Boston, Massachusetts, June 2004. Paper WeM02.1*.
29. Powers, David L. (1987). Boundary Value Problems. San Diego, Harcourt Brace Jovanovich.
30. Press, William H., Teukolsky, Saul A., Vetterling, William T. and Flannery, Brian P. (1992). Numerical Recipes in C, the Art of Scientific Computing. Cambridge University Press, 2nd edition.
31. Wang, P.K.C. and Hadaegh, F.Y. (1999). Minimum-fuel formation reconfiguration of multiple free-flying spacecraft. *The Journal of the Astronautical Sciences*, **47**(1–2), 77–102.
32. Weinstein, Alain (1977). Lectures on symplectic manifolds. *Regional Conference Series in Mathematics*, 29.