ALGORITHMS FOR OPTIMAL CONTROL OF SYSTEMS DESCRIBED BY PARTIAL AND ORDINARY DIFFERENTIAL EQUATIONS

by

Theodore E. Baker

Memorandum No. UCB/ERL M88/45

21 June 1988

ALGORITHMS FOR OPTIMAL CONTROL OF SYSTEMS DESCRIBED BY PARTIAL AND ORDINARY DIFFERENTIAL EQUATIONS

by

Theodore E. Baker

Memorandum No. UCB/ERL M88/45

21 June 1988

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ALGORITHMS FOR OPTIMAL CONTROL OF SYSTEMS DESCRIBED BY PARTIAL AND ORDINARY DIFFERENTIAL EQUATIONS

by

Theodore E. Baker

Memorandum No. UCB/ERL M88/45

21 June 1988

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ALGORITHMS FOR OPTIMAL CONTROL OF SYSTEMS DESCRIBED BY PARTIAL AND ORDINARY DIFFERENTIAL EQUATIONS

Theodore E. Baker

Ph.D.

Dept. of Electrical Engineering
and Computer Science

Signature: ______________________
Committee Chairman

ABSTRACT

We examine optimal control problems for a class of systems described by partial and ordinary differential equations. After presenting a few illustrative optimal control problems, we formulate an abstract differential equation to describe system dynamics and a canonical optimal control problem with control constraints, terminal inequality constraints, and state-space constraints as an infinite dimensional nonlinear programming problem with inequality and box constraints. The abstract differential equation and the canonical problem are sufficiently general as to admit the as an infinite dimensional nonlinear programming problem with inequality previously described optimal control problems. We show that the costs and inequality constraints of the nonlinear program are Gateaux differentiable so that standard infinite dimensional nonlinear programming algorithms can be applied to problems with no control constraints. Equivalently, optimal control algorithms for problems with ODE's and control constraints are extended to PDE systems.

We extend the theory of relaxed controls as presented by Warga [War.1] and Williamson-Polak [Wil.1] to optimal control problems in which the dynamics are described by our abstract differential equation.

We introduce an extension of the Klessig-Polak [Kle.1] adaptive precision gradient method to perform a discretization of the PDE into a finite difference equation and show that for optimal control problems with control constraints the limits of the solutions to the discretized problems satisfy a necessary condition for optimality of the original problem.

We introduce an extension of the Pironneau-Polak [Pir.1] method of feasible directions with two new search direction finding subprocedures. Finally, using this algorithm we solve a collection of optimal control problems involving the rotation of a flexible beam with various objectives and constraints.

To my mother and father

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Professor Elijah Polak for supervising my work. We have spent many hours together, and his observations and suggestions have been invaluable to me.

I wish to thank Professor Robert Taylor for his time. I have enjoyed and benefited from the many discussions I have had with him.

I thank Professor Charles Desoer for serving on my dissertation committee and sharing with me some of his experiences.

I benefited from my discussions with Jerome Sackman, Tosio Kato and Shankar Sastry. I thank them for their time.

Many of my most productive discussions were with fellow students. I wish to thank Ywh-Pyng Harn and Joe Wiest for their comments. I am particularly indebted to Loc Vu, Joe Higgins, Harold Naparst, and Faryar Jabarri for their help.

I want to thank Vicki Kulik for her help and good humor, and Erika Buky for proofreading my thesis.

This research was supported in part by the National Science Foundation Grant ECS-8121149, Air Force Office of Scientific Research Grant 83-0361, Office of Naval Research Grants N00014-83-K-0295, N00014-86-K-0295, The State of California MICRO Program and the General Electric Company. I would like to acknowledge these organizations for their assistance.

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

The development of optimal control theory was motivated in the 1950s and 1960s by the need to extend the calculus of variations to solve the problem of controlling spacecraft trajectories. Determining a sequence of controls to maneuver the Apollo 11 command module to the Moon and back involved solving enormously complicated problems. Especially critical was the craft's return trajectory since incorrect trajectories could have resulted in the command module either burning up as it re-entered the Earth's atmosphere or skipping over the atmosphere and not returning to Earth. The trajectories of the Earth and Moon, the gravitational effects of the Sun, Moon, and Earth, and solar pressures, as well as the dynamics of the spacecraft had to be modeled. The resulting system was nonlinear and of high order. More general applications of open-loop optimal control theory have included calculating minimum time and minimum fuel trajectories for unmanned craft for both orbiting and interplanetary missions.

Optimal control theory has also been used to develop feedback laws to regulate finite-dimensional linear, time-invariant systems. An optimal control problem is formulated in which the cost is the integral of the sum of quadratic functions of the state error and the control. It has been shown (see [Ath.1] for example) that for the solution to this optimal control problem, the control at time t is a linear function of the state error at time t. The linear function is an easily found matrix multiplied by the solution to a Riccati equation.

However, many of the physical processes in the world cannot be described by ordinary differential equations. Two major classes of physical processes, vibration and diffusion, are usually described by partial differential equations. Most moving objects exhibit some degree of vibration such as the swaying of a building due to high winds or the vibration of a car moving down a road. Although some vibrational processes can be modeled effectively by ordinary differential equations[1], others exist for which ordinary differential equations models are inadequate. Many of the structures to be placed in space in the next decades will be hundreds of feet long, and very light. These structures will be flexible with low damping and so the appropriate models will be determined using partial differential equations.

Many physical processes can be modeled as diffusion processes. The best known diffusion process is the heat equation, which models the flow of heat in a homogeneous material. Variants of the heat equation can be used to model electron flow, and even the price of stock options [Bla.1]. More complicated equations can be used to model diffusion in membranes and fluid flows.

There are two major differences between ordinary and partial differential equations. First, most ODEs can be written in the form $\dot{z}(t) = f(z(t), u(t), t)$ where $f(\cdot, \cdot, \cdot)$ satisfies certain continuity assumptions. This formulation has allowed for a uniform development of algorithms to solve optimal control problems with ODE dynamics. There is no such canonical form for partial differential equations. Second, partial differential equations are often composed of unbounded operators such as $\frac{\partial}{\partial x}$. Although the meaning of the exponential of a bounded linear operator is well understood in the control literature, the properties of exponentials of unbounded operators, and the role they play in solving optimal control problems with PDEs needs further examination.

This dissertation provides a framework for solving optimal control problems arising in systems that can be modeled by partial differential equations. Central to this framework is a

[1] The vibration of the car is often modeled by a simple spring-mass-damper ordinary differential equation.

canonical abstract differential equation. Any problem whose dynamics fit the abstract differential equation and whose cost and constraints are continuously differentiable functions of the state can be solved. This abstract differential equation is general enough for many of the systems described above, as well as for any system whose dynamics can be described by first-order ordinary differential equations satisfying certain continuity requirements.

1.2 ALGORITHMS FOR ODES

In this section, we review earlier work in the field of optimal control of ODEs. The section is divided into three parts. We present several conceptual algorithms for solving optimal control problems with ODE's. These algorithms are primarily extensions of finite dimensional non-linear programming algorithms although some have no finite dimensional analog. The proofs of convergence for these optimal control algorithms follow their finite dimensional counterparts with one exception: for any bounded sequence in finite dimensional space, there exists at least one accumulation point. In infinite dimensional space, it is not necessary that a bounded sequence have an accumulation point. Since the proofs are based on showing various properties of accumulation points, when no accumulation points exists the proofs are null. We shall describe some previous work which has been done to close this gap. Finally, the above algorithms are *conceptual* in that they require the exact solution of ordinary differential equations and the calculation of an infinite dimensional design vector. We present a few of the previously developed methods for implementing these algorithms on a computer.

CONCEPTUAL ALGORITHMS:

We begin by stating a canonical optimal control problem for systems described by ordinary differential equations.

$$\mathbf{P}: \inf_{u,\tau} \{ g^0(u,\tau) \mid g^j(u,\tau) \le 0, j \in \{1,2,\cdots,m_1\},$$

$$g^j(u,\tau) = 0, j \in \{m_1+1,\cdots,m\}, \phi^k(u,\tau,t) \le 0, t \in [0,\tau], k \in \{1,2,\cdots,p\},$$

$$u \in G_\tau, \tau \in [\tau_{\min},\tau_{\max}], 0 \le \tau_{\min} \le \tau_{\max} \le \infty \}, \tag{1.1}$$

where $g^j(u,\tau) = h^j(x(\tau,u,\tau))$ for $j \in \{1,2,\cdots,m\}$, $\phi^k(u,\tau,t) = \bar{h}^k(x(t,u,\tau))$ for $k \in \{1,2,\cdots,p\}$, Ω is either a compact, convex subset of $\mathbb{R}^n$ or the entire space $\mathbb{R}^n$, $x(\cdot,u,\tau)$ is the solution to

$$\frac{d}{dt}x(t,u,\tau) = f(x(t,u,\tau),u(t)),\ t \in [0,\tau],\ x(0,u,\tau) = x_0, \tag{1.2}$$

$$G_\tau \triangleq \{ u \in L_2^m([0,\tau]) \cap L_\infty^m([0,\tau]) \mid u(t) \in U \}, t \in [0,\tau], \tag{1.3}$$

U is either a compact, convex subset of $\mathbb{R}^m$ or the entire space $\mathbb{R}^m$. When $m > 0$, we say that the problem has terminal constraints; when $p > 0$, the problem has state-space constraints. If $\tau_{\min} < \tau_{\max}$, **P** is called a free-time problem; otherwise it is called a fixed-time problem.

A standard method of transcription for converting free-time problems into fixed-time problems is described in Chapter 4. Its only drawback is that the resulting dynamics are non-linear even if the original dynamics were linear. With suitable assumptions on $f(\cdot,\cdot)$, $g^j(\cdot,\cdot)$ is differentiable in the control u. Consequently, it is at least theoretically possible to extend any algorithm designed to solve finite dimensional non-linear programming problems to an algorithm to solve an optimal control problem if there are no control constraints. A survey of such methods, including a discussion of infinite dimensional analogs appears in [Pol.4]. We also refer to Polak's book [Pol.1], which gives a detailed look at many of these algorithms, and Luenberger's book [Lue.1], which presents some additional general non-linear programming algorithms. Although the algorithms in [Lue.1] are not presented in the context of optimal control algorithms, the extensions are farily straightforward. In particular, we point out that the Newton method has been extended to solve unconstrained optimal control problems in which g^0 is twice continuously differentiable [Gol.1, Mit.1]. Furthermore, a Newton method can be used [Pol.1] to find a control which satisfies the necessary condition for optimality arising from the Maximum Principle.

Bertsekas [Ber.1] has extended the Goldstein-Levitin-Polyak gradient projection algorithm to problems with only control constraints. Mayne and Polak have developed a family of optimal control algorithms to solve problems with control constraints and terminal inequality constraints [Pol.2, Pol.3], control and terminal equality and inequality constraints [May.2], and control, state-space, and terminal equality and inequality constraints [May.3]. The algorithms in [Pol.2] and [Pol.3] are extensions of the Pironneau-Polak algorithm [Pir.1] to the infinite dimensional case with control constraints. [May.2] and [May.3] are extensions of exact penalty function algorithms to the infinite dimensional case with control constraints. Warga's [War.2] algorithm solves problems with state-space, control and terminal inequality constraints.

RELAXED CONTROLS:

Early in the study of optimal control theory it was realized that there may not be solutions to optimal control problems. For example, consider the problem:

$$\min\{\ g(u) \mid u \in G\ \}, \tag{1.4}$$

where

$$G \triangleq \{\ u \in L_\infty^m \mid u(t) \in U\,,\, t \in [0\,,1]\ \}, \tag{1.5}$$

U is a compact convex set of $\mathbb{R}^m$, and $g(u) = h(x(1\,,u))$ and $x(\cdot\,,u)$ is the solution to

$$\frac{d}{dt}x(t\,,u) = f(x(t\,,u)\,,u(t))\,,\ \ x(0\,,u) = x_0\,, \tag{1.6}$$

and $h(\cdot)$ is continuous. Consequently, $g(\cdot)$ is continuous. The set G is closed and bounded but not compact, and therefore, there is no guarantee that a solution to (2.1.4) exists. It was realized that for many infimizing sequences, the controls become increasingly chattering, and that the effects chattering controls on ODEs is the same as that exhibited by a modified ODE system:

$$\frac{d}{dt}x(t,u) \in \underset{u \in U}{\mathrm{co}}\, f(x(t,u),u),\ \ x(0,u) = x_0.^2 \tag{1.7}$$

This observation leads to a more formal theory [War.1] in which each member of the convex hull of $f(x(t,u),U)$ is expressed as an integral over the set U with a specific measure.

DISCRETIZATION:

The numerical solutions of optimal control problems require some form of discretization since all numerical integration techniques involve discretization. The most basic method [Ros.1] is to replace the differential equation modeling the dynamics with a finite difference equation obtained by discretizing the differential equation. By contrast, Canon, Cullum and Polak [Can.1] proposed using standard sampled data discretization of the control, which restricts the control to be piecewise constant with a finite number of discontinuities. These algorithms raise the question of the relation between the solutions to the discretized problems and the solution to the original problem. It was shown by Cullum [Cul.1] that only under reasonably restrictive assumptions can one be sure that the solutions of the discretized problems will converge in some sense to the solution of the original problem as the discretization is infinitely refined. Furthermore, she showed that if a free-time problem is discretized in an ad hoc fashion, then the solutions of the discretized problems would almost certainly not approximate a solution of the original problem. However, if the free-time problem is transcribed

[2] This means that $x(\cdot,u)$ is a function such that for all t, $\frac{d}{dt}x(t,u)$ is a member of the convex hull of $f(x(t,u),U)$.

into a fixed-time problem, then the solutions of the discretized problems will approximate a solution of the original problem. Similarly, she showed that state-space constraints must be transformed into affine constraints by the introduction of additional variables in order to make the problem stable with respect to discretization.

Klessig and Polak have developed a theory of adaptive precision discretization for implementation. In [Kle.1], they present an algorithm model for solving an unconstrained optimal control problem. Their theory is based on the ability to calculate the value of the cost and the gradient of the cost to an arbitrary accuracy. For a given value of the control and a given tolerance, an approximate cost and gradient of the cost are calculated within the tolerance. These approximations are substituted for the true cost and true gradient in a standard gradient algorithm. When a certain criterion is met, the required accuracy is increased (i.e., the tolerance is reduced). This approach has two main benefits: it is believed that accuracy of the calculation of the cost and its gradient is less critical farther from the optimal point. Since it is cheaper to calculate the cost and its gradient with lower accuracy, the adaptive technique could be more efficient in finding an optimal point. As for the convergence issue, Klessig and Polak have proven that accumulation points of sequences produced by such an adaptive strategy satisfy first order necessary optimality conditions for the original problem. They present a specific algorithm implementation for solving unconstrained optimal control problems. Their experimental data has shown their adaptive strategy to be much more efficient than the standard nonadaptive approach.

CHAPTER 2

OPTIMAL CONTROL WITH PDE DYNAMICS

In this chapter, we present two examples of problems in the optimal control of systems with PDE dynamics. Although the original motivation of our research was to solve the problem of optimal slewing of flexible structures, we discovered that this work is general enough to apply to a larger class of problems.

2.1 EXAMPLE - OPTIMAL SLEWING OF FLEXIBLE STRUCTURES:

Recent years have seen an increase in research in the optimal control of flexible structures. The primary motivation for this research is the control of flexible aerospace structures, which are becoming larger and more flexible while their performance requirements are becoming more stringent (see, e.g., [Tay.1, Nas.1]). For example, in tracking and other applications, satellites with large antennae, solar collectors, and other flexible components must perform fast slewing maneuvers while maintaining tight control over the vibrations of their flexible elements. Outside of the aerospace applications ,research on flexible structure control may have an impact on the control of mechanisms with flexible links.

There are two major reasons to study optimal control for slewing flexible structures. First, the equations describing the dynamics of flexible structures are usually pseudo-linear; there are linear partial differential equations (PDEs) which are coupled by non-linear ordinary differential equations (ODEs). It is therefore very difficult to control the structures effectively by feedback laws alone. Open loop optimal control can be used to bring a flexible structure close to the desired state and feedback control [1] can be used to ensure the final

[1] The feedback control law is based on linearized models of the system.

accuracy.[2] Second, the solution to the optimal control problem is by definition the best control given our performance criteria. We can use this as a benchmark with which to compare suboptimal (and possibly easier to compute) strategies.

We begin with four illustrative problems. These examples, although simple, are representative of optimal control problems involving slewing of flexible structures.

We consider the hollow aluminum tube depicted in figure 2.1. The tube is one meter long and has a uniform cross sectional radius of 1.0 cm and a thickness of 1.6 mm. Attached to one end of the tube is a mass of 1 kg, and attached to the other end is a shaft connected to a motor. For simplicity, we assume that the torque produced by the motor can be directly controlled. Our aim is to determine the torque necessary to rotate the tube and bring it to rest. The maximum torque produced by the motor is 5 Newton-meters.

[2] We also envision using a hybrid open-loop / closed-loop control in which the open-loop control is used to determine the setpoints for a series of closed loop controllers.

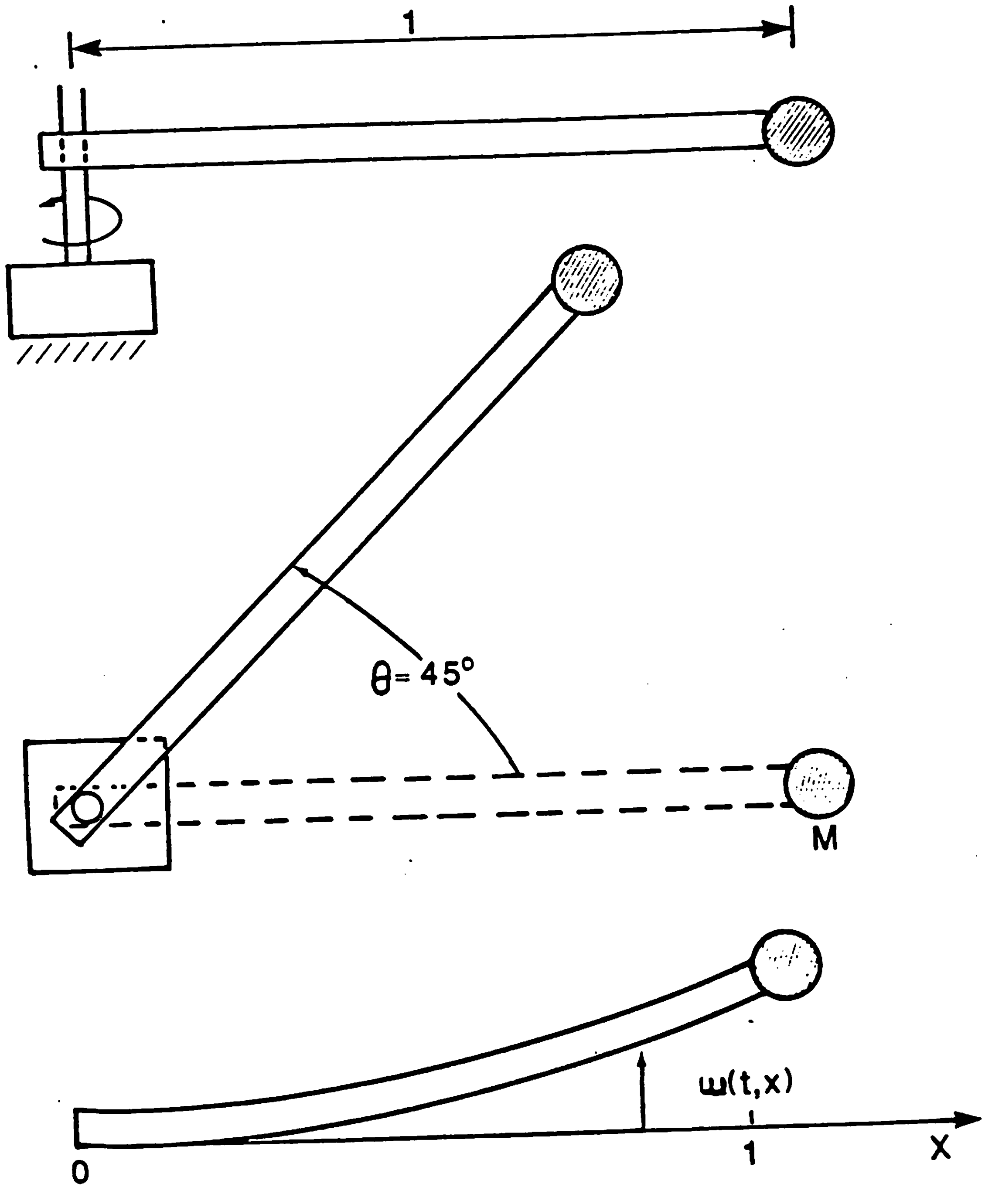

Figure 2.1 - Configuration of Slewing Experiment

the following relation between the stress in the bar, σ, and the strain, $\varepsilon = \frac{\partial u}{\partial x}$:

$$\sigma = E\varepsilon + \eta\frac{\partial \varepsilon}{\partial t}. \tag{2.1.2}$$

Since the partial derivative of stress with respect to distance, is equal to the mass per unit length times the acceleration,

$$\rho\frac{\partial^2 u}{\partial t^2} = \frac{\partial \sigma}{\partial x}, \tag{2.1.3}$$

$$\rho\frac{\partial^2 u}{\partial t^2} - \eta\frac{\partial^3 u}{\partial t \partial x^2} - E\frac{\partial^2 u}{\partial x^2} = 0. \tag{2.1.4}$$

The Maxwell assumptions lead to a different relation between stress and strain:

$$\frac{\partial \varepsilon}{\partial t} = \frac{1}{E}\frac{\partial \sigma}{\partial t} + \frac{\sigma}{\eta}. \tag{2.1.5}$$

Therefore, by taking the partial of (2.1.5) with respect to x, and using (2.1.3), we obtain

$$\frac{\partial \varepsilon}{\partial x \partial t} = \frac{\rho}{E}\frac{\partial^3 u}{\partial t^3} + \frac{\rho}{\eta}\frac{\partial^2 u}{\partial t^2}. \tag{2.1.6}$$

Integrating with respect to t, and setting the constant of intergration to zero, (2.1.6) becomes

$$\rho\frac{\partial^2 u}{\partial t^2} + \frac{\rho E}{\eta}\frac{\partial u}{\partial t} - E\frac{\partial^2 u}{\partial x^2} = 0. \tag{2.1.7}$$

In PDEs derived using the Kelvin-Voigt damping model, pulses travel infinitely fast through the beam, and therefore the Kelvin-Voigt assumption is accurate for only low frequency modeling. Conversely, Maxwell damping is not accurate for modeling low frequencies.

Returning to the beam, and applying Kelvin-Voigt and Maxwell damping to (2.1.1), we obtain (2.1.8) and (2.1.9) respectively.

Modeling

Several different PDEs can be obtained to model the dynamics of the tube. The most sophisticated PDEs, which arise from applying the Theory of Elasticity, are very accurate and can model vibrations that have wavelengths at the order of magnitude of the molecular level. However, calculation with this model is prohibitively expensive in all but the simplest cases. A simpler model is based on the Timoshenko assumption that the planar sections remain planar under deformation [Gra.1]. Using this model, we obtain a hyperbolic PDE which is fourth order in time and space. This model is considered to be fairly accurate for a wide range of wavelengths. Finally, a third model is based on the Euler-Bernoulli assumptions that the planar sections remain planar and that the planar sections remain perpendicular to the centroid axis under deformation [Gra.1]. This model gives rise to the classical beam equation:

$$\frac{\partial^2 u}{\partial t^2} + \gamma \frac{\partial^4 u}{\partial x^4} = q. \tag{2.1.1}$$

However, this model is limited in its ability to model high-frequency responses. For a uniform beam, this model becomes inaccurate when the wavelength of the vibration is less than about ten times the depth of the beam. It is easily shown that Equation (2.1.1) is parabolic, and therefore predicts that pulses will travel infinitely fast through the beam. This directly contradicts experimental evidence.

Several different models also exist for damping. For ease of exposition, these models are described in terms of longitudinal waves through a bar. The classical Kelvin-Voigt model assumes that each infinitesimally small section of the bar can be modeled as a spring in parallel with a dashpot. Its dual, Maxwell damping, assumes that this infinitesimally small section can be modeled as a spring in series with a dashpot. The Kelvin-Voigt assumptions lead to

$$\frac{\partial^2 u}{\partial t^2} + \alpha_k \frac{\partial^5 u}{\partial t \partial x^4} + \gamma \frac{\partial^4 u}{\partial x^4} = 0. \tag{2.1.8}$$

$$\frac{\partial^2 u}{\partial t^2} + \alpha_m \frac{\partial u}{\partial t} + \gamma \frac{\partial^4 u}{\partial x^4} = 0. \tag{2.1.9}$$

A third type of damping combines Kelvin-Voigt and Maxwell damping. A fourth type, proportional damping, in which the damping of a mode is proportional to its frequency, does not have a simple physical interpretation. However, recent experimental results on vibrating beams [Tay.2] indicate that proportional damping is the best model of the four. Applying proportional damping to (2.1.1), we obtain:

$$\frac{\partial^2 u}{\partial t^2} + \alpha_p \frac{\partial^3 u}{\partial t \partial x^2} + \gamma \frac{\partial^4 u}{\partial x^4} = 0. \tag{2.1.10}$$

Finally, non-linear models of deformation and vibration have been developed. See Vu-Quoc [Vuq.1]. These models are accurate under large deformations where both the Timoshenko and Euler-Bernoulli models break down due to inaccuracy in geometric modeling.

In summary, there are four models: the Theory of Elasticity, Timoshenko, Euler-Bernoulli, and the non-linear models. We have also discussed four types of damping: none, Kelvin-Voigt, Maxwell, and proportional. The abstract differential equation in Chapter 3 is general enough to admit the Euler-Bernoulli beam with each type of damping for fixed-time problems and Kelvin-Voigt and proportional damping for free-time problems.[3] For subsequent analysis and simulation (see Chapter 8), we have chosen to use the Euler-Bernoulli assumptions with Kelvin-Voigt damping. We have neglected coupling between the axial and flexural modes of the tube. This simplification may cause large modeling errors at high velocities [Sim.1].

---

[3] Models with Maxwell damping or no damping cannot be shown to generate the analytical semigroups necessary for free-time problems. See Chapter 4.

The appropriate equations of motion determined by application of the standard Euler-Bernoulli tube with Kelvin-Voigt visco-elastic damping are:

$$mw_{tt}(t,x) + CIw_{txxxx}(t,x) + EIw_{xxxx}(t,x) - m\Omega^2(t)w(t,x)$$
$$= \frac{m}{Ml^2 + \frac{1}{3}ml^3} u(t)x \, , x \in [0\, , l] \qquad (2.1.11)$$

with boundary conditions:

$$w(t,0) = 0\, , w_x(t,0) = 0\, , CIw_{txx}(t,1) + EIw_{xx}(t,1) = 0. \qquad (2.1.12)$$

$$M[\Omega^2(t)w(t,1) - w_{tt}(t,1) - u(t)l] + CIw_{txxx}(t,1) + EIw_{xxx}(t,1) = 0\, , \qquad (2.1.13)$$

and rigid body dynamics:

$$\frac{d}{dt}\Theta(t) = \Omega(t)\, , \quad \frac{d}{dt}\Omega(t) = \frac{1}{Ml^2 + \frac{1}{3}ml^3} u(t)\, , \qquad (2.1.14)$$

where $w(t,x)$ is the displacement of the tube from the *shadow tube* (which remains undeformed during the motion) due to bending as a function of time and distance along the tube; $u(t)$ is the torque applied by the motor, and $\Omega(t)$ is the resulting angular velocity (in radians per second). We shall denote by $\Theta(t)$ the angular displacement of the rigid body (in radians). The values for the parameters in (2.1.11) - (2.1.13) are: $l = 1.0$ m, $m = .257$ kg/m, $C = 6.30\times10^7$ Pascals/sec., $E = 6.30\times10^9$ Pascals, $I = 1.005\times10^{-8}m^4$, $M = 0.914$ kg. The tube is very lightly damped (0.1 percent).

We assume that the tube is initially at rest with no deformations, and so the following initial conditions hold:

$$w(0,x) = w_t(0,x) = 0\, , \quad x \in [0\, ,1]. \qquad (2.1.15a)$$

$$\Theta(0) = \Omega(0) = 0. \qquad (2.1.15b)$$

We consider four problems:

P_1: Minimize the time required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint.

P_2: Minimize the total energy required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint and the maneuver time not exceeding a given bound.

P_3: Minimize the time required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint and an upper bound on the potential energy due to deformation of the tube throughout the entire maneuver.

P_4: Minimize the total energy required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint, the maneuver time not exceeding a given bound, and an upper bound on the potential energy due to deformation of the tube throughout the entire maneuver.

ALGORITHMS FOR OPTIMAL SLEWING OF FLEXIBLE STRUCTURES:

There are two basic schools of thought concerning slewing of space structures. One advocates discretizing the PDE into an ODE using the Raleigh-Ritz or Finite Element Method. The resulting problem is then solved by replacing the PDE dynamics with the ODE dynamics using standard methods for optimal control of ordinary differential equations. Such an approach often involves an attempt to determine the spill-over effect of the ignored dynamics on the solution. The second school advocates solving the original problem with dynamics described by the PDE. The second approach is much more difficult, since it requires the invention of new methods. There is some debate as to whether it is necessary to use the second approach for slewing of flexible structures.[4]

---

[4] We need to distinguish here between slewing control and feedback control. By the former, we wish to slew a flexible structure close to the desired state. By the latter, we wish to use feedback to close the loop resulting in a stable system. We believe that spill-over effect is much more critical for feedback control because it can result in an unstable system.

We present one of the best known strategies of the first school for slewing a flexible structure. Junkins and Turner [Jun.1] have developed a method for solving the slewing problem of a rotating structure. They first consider the fixed-time problem in which the vibration at the final time is arrested and the final rotational velocity is fixed. The final rotation angle may be free or fixed. Using the Rayleigh-Ritz method, they determine a set of non-linear differential equations to approximate the dynamics of the system. Their objective is to minimize the performance index $J = \int_0^{t_f} [u^T(t)W_{uu}u(t) + x^T(t)W_{xx}x(t)]dt$ subject to the final state conditions where $u(\cdot)$ and $x(\cdot)$ are the input and state respectively, and W_{uu} and W_{xx} are positive definite matrices chosen by the designer. A linearized set of equations is generated and the resulting two-point boundary value problem (TPBVP) is solved. Finally, using a continuation technique [Sch.1], they determine the solution to the non-linear TPBVP. The continuation is iterative, and each iteration requires at least one solution of the non-linear state transition matrix. For problems in which a bifurcation point exists, the Chow-Yorke algorithm [Cho.1] may be used, but at considerable cost. By exploiting the sparse matrix properties of the problem, they are able to reduce computational effort by up to seventy-five percent. They have performed some frequency shaping on the control by augmenting the non-linear differential equation. For a simple problem, Junkins and Turner have produced numerical results. See [Jun.1]

For linear free-time problems, those in which t_f is not fixed, they solve the fixed-time problem for several different values of t_f. Then, using the minimum of these times such that J as described above is satisfactory as a starting point, an iterative algorithm determines the minimum time.

The limitations of this technique are: (1) it does not solve optimal control problems with hard control constraints (e.g. $u(t) \leq 1$ for $t \in [0, t_f]$); (2) it does not solve problems with state space constraints (e.g., bounding the potential energy resulting from the deformation of the tube throughout the entire maneuver); (3) The PDE dynamics are not used directly, and Junkins and Turner provide no analysis that shows the relation between their solution and the solution to the original problem. Additional work in this area includes [Bre.1, Chu.1, Flo.1]

A strategy from the second school for slewing a flexible structure is shown by Araya's [Ara.1] solution of the minimum-time optimal control problem:

$$\mathbf{P}_A : \min_{\tau, u}\{ \tau \mid \|x(\tau, u)\| \leq \delta, u \in G_\tau \}, \tag{2.1.16}$$

where $x(\cdot, u)$ is the mild solution of:

$$\frac{d}{dt}x(t) = Ax(t) + K(x(t)) + Bu(t), \quad x(0) = x_0, \tag{2.1.17}$$

A is a strongly continuous group generator, B is a linear finite dimensional operator, K is a C^∞ non-linear operator, δ is a small positive number, and G_τ is defined as in (2.1.9).

Araya chose the model general enough to be useful for many types of flexible structure slewing. In particular, he was interested in solving the NASA SCOLE design challenge, [Tay.1].

He creates a series of subproblems, $P_{\varepsilon, T}$:

$$P_{\varepsilon, T}\begin{cases} \min\int_0^T g^\varepsilon(x(s)) + H^\varepsilon(u(s))ds & H^\varepsilon(u) = \frac{1}{2\varepsilon}\sum[(|u_i| - c_i) +]^2 \\ \dot{x} = Ax + K(x) + Bu, \quad x(0) = x_0 & g^\varepsilon(u) = \Pi\left[\frac{\|x\|^2 - \delta^2}{2\varepsilon}\right] \end{cases} \tag{2.1.18}$$

where $\Pi \in C^\infty(0, 1)$, $\Pi'(s) \geq 0$, and

$$\Pi(s) = \begin{cases} 1 \text{ for } s \geq 2 \\ 0 \text{ for } s \leq 1 \end{cases}. \tag{2.1.19}$$

He derives maximum principles for P and $P_{\varepsilon,T}$. Finally he shows that as $\varepsilon \to 0$ and

$T \to \infty$, the controls derived from $P_{\varepsilon, T}$ approach the optimal control for P weakly in the $L_2[0, \tau]$ norm for every $\tau \geq 0$. It appears that state-space constraints can be added to his formulation without serious difficulty.

However, no effort has been made to implement this algorithm, and in fact implementation seems quite difficult since in its current form the algorithm requires the solution of an infinite-dimensional state transition matrix.

2.2 HEAT-EQUATION EXAMPLE

In this section we examine another example of an optimal control problem with PDE dynamics. In this example, the dynamics can be modeled as a diffusion process. Consider a cylinder (hot dog) of infinite length and radius 1. The cylinder has initial temperature $w(0, \cdot)$ zero. Starting at $t = 0$, we can control the time derivative of the temperature at the boundary $w(t, 1)$ of the cylinder, $u(t)$, such that $u(t) \in [-5, 5]$. The objective is to bring the entire cylinder to within a small tolerance of the temperature 5 in minimum time. We require the temperature of the cylinder $w(t, r)$ to be between 0 and 10 during the heating process.

By symmetry, we can consider the plane perpendicular to the axis of the cylinder. The heat equation in this plane in polar coordinates is:

$$\frac{\partial}{\partial t} w(t, r) = \frac{1}{r} w(t, r) + \frac{\partial^2}{\partial r^2} w(t, r), \quad w(0, r) = 0, r \in [0, 1], \tag{2.2.1}$$

with boundary conditions:

$$\frac{\partial}{\partial t} w(t, 1) = u(t), \quad \frac{\partial}{\partial r} w(t, 0) = 0. \tag{2.2.2}$$

We introduce an auxiliary variable $z(\cdot)$ so that

$$\frac{\partial}{\partial t}w(t,r) = \frac{1}{r}w(t,r) + \frac{\partial^2}{\partial r^2}w(t,r) \;\; w(0,r) = 0, r \in [0,1], \tag{2.2.3}$$

$$\frac{\partial}{\partial t}z(t) = u(t), z(0) = 0, \tag{2.2.4}$$

with boundary conditions:

$$w(t,1) = z(t), \quad \frac{\partial}{\partial r}w(t,0) = 0. \tag{2.2.5}$$

Therefore a mathematical formulation for the hot dog problem is:

$$\min_{u,\tau}\{ \tau \mid (\|w(t,\cdot)\|^2 - 5)^2 \le \varepsilon, \|\frac{\partial}{\partial x}w(t,\cdot)\|^2 \le \varepsilon, z(t) \in [0,10], u \in G_\tau \}, \tag{2.2.6}$$

where

$$G_\tau \triangleq \{ u \in L_\infty([0,\tau]) \mid u(t) \in [-5,5], t \in [0,\tau] \}. \tag{2.2.7}$$

2.3 THESIS SUMMARY AND CONTRIBUTIONS

Now that we have explained the problem and given some introductory notation, we are in the position to present an outline for the rest of this thesis.

Chapter 3: We introduce a canonical form for the dynamics:

$$\frac{d}{dt}x(t,u) = Ax(t,u) + F(x(t,u), u(t)), \; t \in [0,\tau], \; x(0,u) = x_0. \tag{2.3.1}$$

where $x(t,u)$ is a member of a Hilbert space X, $u \in G_\tau \triangleq \{ u \in L_\infty^m([0,1]) \mid u(t) \in U, [0,\tau] \}$, U is a convex, compact subset of $\mathbb{R}^m$, $A : D(A) \to X$ is the infinitesimal generator of a strongly continuous semigroup, and $F : X \times \mathbb{R}^m \to X$ is a nonlinear bounded operator satisfying certain assumptions (see Sections 3.2 and 3.3).

We establish Gateaux differentiability with respect to the control for fixed-time problems, i.e., for $u \in G_\tau$, $t \in [0,\tau]$, there exists $x_u(t,u) \in B(L_2^m[0,1] \cap L_\infty^m([0,1]); X)$ such that for any $h \in L_2^m([0,1]) \cap L_\infty^m([0,1])$,

$$\lim_{\lambda \to 0} \frac{1}{\lambda} \| x(t, u + \lambda h) - x(t, u) - \lambda x_u(t, u) h \| = 0. \tag{2.3.2}$$

Furthermore, we can show that $x_u(t, u)$ is continuous in u at all $u \in G_\tau$. Based on this differentiability, we construct an algorithm to solve optimal control problems with control constraints.

Chapter 4: We consider free-time problems and show that by using a standard transcription, the free-time problem becomes a fixed-time problem whose dynamics are similar to (2.3.1). We show Gateaux differentiability of the state for the transcribed free-time problem with respect to the control and the final time.

Chapter 5: We extend the theory of Chapter 3 to relaxed controls (i.e., G_τ is replaced by a new set with a new topology.)

Chapter 6: We present two implementable algorithms for solving the optimal control problem with control-constraints. These algorithms are extensions of the Klessig-Polak algorithms [Kle.1, Kle.2]. We briefly describe our changes and extensions. First, the Klessig-Polak algorithms consist of finding a sequence of approximations to the cost function and the gradient of the cost function. These algorithms find a sequence of approximations to the cost function and then find the exact gradient of each approximate cost function. The latter method is actually a subset of the former, but experimental evidence has shown that the latter method is far superior to general method of discretization described in [Kle.1] Second, we have extended the Klessig-Polak algorithms to the case with control constraints. Third, we have made less restrictive requirements on functions approximating the cost. Since integration methods for PDEs are less well understood, this extra leeway is a benefit. Fourth, we have extended the proofs of convergence to relaxed controls.

We also provide a simple example based on Fujii's [Fuj.1] analysis of Newmark's method on a vibrating string to show how these approximating cost functions can be constructed.

Chapter 7: We present our canonical optimal control problem:

$$\inf\{\ g^0(u,w) \mid g^j(u,w) \le 0, j \in \underline{p}, \phi(u,w,t) \le 0, t \in [0,1], u \in G, w \in C\ \} \tag{2.3.3}$$

where $g^j(u,w) = h^j(x(1,u,w))$, $j \in \underline{p} \triangleq \{\ 1,2,\cdots,p\ \}$ and $\phi(u,w,t) = \bar{h}(x(t,u,w)), t \in [0,1]$, and $x(\cdot,u,w)$ is the solution to an abstract differential equation with control u and initial state w.

An extended version of the Polak-Mayne [Pol.2] algorithm is presented with two different search direction finding subprocedures.

Chapter 8: We present numerical solutions to the problems $\mathbf{P}_1$ - $\mathbf{P}_4$ described above, obtained by using implementable versions of the algorithms in Chapter 7. We compare the two subprocedures presented in Chapter 7 and discuss the selection of parameters for these algorithms.

Chapter 9: Concluding Remarks.

2.4 EXTENSIONS

There are several directions in which to continue this research:

2.4.1. Examination of Different Models

For this thesis, we have been able to perform numerical tests only for a beam described by the Euler-Bernoulli equations with Kelvin-Voigt damping. Dynamics obtained from the Timoshenko model and proportional damping, as well as the non-linear rod model and computational procedure discussed in Vu-Quoc [Vuq.1], should be examined.

2.4.2 Extension of the Canonical Form of the Abstract Differential Equation (2.3.1)

We have done preliminary work to model a two-link flexible manipulator with revolute joints using the Euler-Bernoulli assumptions. This work indicates that equation (2.4.1) is not general enough to accommodate the resulting PDEs. We propose two extensions to (2.4.1). First for a two-link flexible manipulator in which there is no coupling from the flexible modes to the rigid body modes:

$$\frac{d}{dt}x(t,u) = C(z(t,u))Ax(t,u) + F_1(x(t,u), z(t,u), u(t)),\ t \in [0,\tau],\ x(0,u) = x_0. \quad (2.4.1)$$

$$\frac{d}{dt}z(t,u) = F_2(z(t,u), u(t)),\ z(0,u) = z_0. \quad (2.4.2)$$

where $x(t,u)$ is a member of a Hilbert space X, $z(t,u) \in \mathbb{R}^n$, u is the control, $A : D(A) \to X$ is the infinitesimal generator of a strongly continuous semigroup, $F_1 : X \times \mathbb{R}^n \times \mathbb{R}^m \to X$, and $F_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $C : \mathbb{R}^n \to \mathrm{B}(X,X)$ are non-linear bounded operators satisfying continuity assumptions similar to those in Sections 3.2 and 3.3.

To allow for coupling between rigid and flexible modes, we replace $f_2(z(t,u), u(t))$ by $f_3(x(t,u), z(t,u), u(t))$.

In (2.4.1), the operators on the control are bounded. In certain applications it may be desirable to make these operators unbounded. We suggest an additional canonical form to be examined:

$$\frac{d}{dt}x(t,u) = Ax(t,u) + F(x(t,u), z(t,u), u(t)) + Bu(t)\ t \in [0,\tau],\ x(0,u) = x_0. \quad (2.4.3)$$

where A and F are defined as in (2.4.1) and $B : \mathbb{R}^m \to X$ is an unbounded operator.

2.4.3. Development of Optimal Control Algorithms

The algorithms presented in Chapters 7 and 8 are based on first order nonlinear programming methods. Future work could examine other first-order and second order methods for nonlinear programming to see if they could be efficiently extended to optimal control problems (2.4.3). Several new first-order methods for nonlinear programming are being developed, with a particular view toward efficient scaling [Pol.6]; as these become available, they might also be extended.

CHAPTER 3

ABSTRACT FORMULATION FOR FIXED TIME PROBLEMS

3.1 INTRODUCTION

Extending optimal control algorithms for problems with ODEs, such as those described in Chapter 2, to optimal control problems with PDEs depends on how the PDE dynamics of the system are formulated into state space form. While it is still unclear which systems of partial differential equations are transcribable into the specific form (3.1.2), below, we know that system (2.2.5) can be transcribed (see Appendix 2). The central issue in validating a formulation is whether the resulting operator A, defined below, is the infinitesimal generator of a continuous semigroup. In this chapter, we shall consider the fixed time problem with only cost and hard control constraints. The minimum time problem with state-space and general constraints will be discussed in Chapters 4 and 7. We will use the following notation:

Let $\langle \cdot , \cdot \rangle$ denote the L_2 inner product. Let X denote a Hilbert Space with inner product $\langle \cdot , \cdot \rangle_X$ and dual space X^*. Since X is a Hilbert Space, X^* can be identified with X. Let $A : D(A) \to X$ be the infinitesimal generator of a continuous semigroup $\{ T(t) \}_{t \geq 0}$, let $F : X \times \mathbb{R}^m \to X$ be a nonlinear operator, and let G be the set of admissible controls defined by:

$$G \triangleq \{ u \in L_\infty^m([0,1]) \cap L_2^m([0,1]) \mid u(t) \in U , t \in [0,1] \} , \tag{3.1.1}$$

where U is a compact convex subset of $\mathbb{R}^m$. We choose $L_2^m([0,1]) \cap L_\infty^m([0,1])$ as the topology for our controls. This topology contains as open sets, sets which are open in both the $L_2^m([0,1])$ and $L_\infty^m([0,1])$ topologies. We need the boundedness associated with the L_∞ topology. However L_∞ is too fine a topology; the L_2 topology is sufficiently coarse for showing the differentiability properties that are needed for optimization. For notational convenience, we denote this space by $L_2 \cap L_\infty$.

PDE-FORM I:

We shall consider dynamics in a general and a specialized form. We now state the general form.

Let $x(t, u) \in X$, $\forall\ t \in [0, 1]$ denote the solution (if one exists) to:

$$\frac{d}{dt}x(t, u) = Ax(t, u) + F(x(t, u), u(t)), \quad x(0, u) = x_0 \in D(A). \tag{3.1.2}$$

The simplest canonical optimal control problem P that we will consider has the form:

$$P : \inf_{u \in G} \{ g(u) \} \tag{3.1.3}$$

where $g(u) = h(x(1, u))$ and h is a continuously differentiable function from X into $\mathbb{R}$.

3.2 EXISTENCE AND BOUNDEDNESS OF SOLUTIONS TO PDE

The first issue to consider is what type of solutions to (3.1.2) exist. There are two types of solutions of interest. Using standard definitions [Paz.1]:

Definition: A function $x(\cdot, u) \in C([0, 1], X)$ is a *mild solution* to (3.1.2) if

$$x(t, u) = T(t)x_0 + \int_0^t T(t - s)F(x(s, u), u(s))ds. \tag{3.2.1}$$

■

Definition: A function $x(\cdot, u) \in C([0, 1], X)$ is a *classical solution* to (3.1.2) if

(i) $x(\cdot, u)$ is continuously differentiable on $[0, 1]$;

(ii) $x(t, u) \in D(A)$, $t \in [0, 1]$;

(iii) (3.1.2) is satisfied. ■

To simplify the discussion, we define an open set O containing G:

$$O \triangleq \{ u \in L_2 \cap L_\infty \mid u(t) \in \tilde{U}, t \in [0, 1] \}, \tag{3.2.2}$$

where $\tilde{U}$ is an open set containing U.

Under our assumptions, if a mild solution exists, it is unique; and if a classical solution exists, then mild solution exists, and the two solutions are equal. We cannot guarantee the existence of either a mild or classical solution of (3.1.2) under our hypotheses. The following additional condition guarantees existence of a mild solution (Lemma A2.3.1):

Assumption 3.2.1: (Lipschitz Continuity of $F(\cdot,\cdot)$). There exists $K < \infty$ such that for all $\hat{x}, x \in X$ and all $\hat{u}, u \in \tilde{U}$, $\|F(\hat{x},\hat{u}) - F(x,u)\| < K[\|\hat{x} - x\| + \|\hat{u} - u\|]$. ■

This is too strict a condition for the dynamics in (1.8) Consequently the issue of whether a mild solution exists must be handled case by case.

We now introduce a more restrictive abstract form. Mild solutions exist for all systems which can be transcribed into this form. Furthermore, any system of PDEs transcribable into the second form (3.2.3a) - (3.2.3c) can also be transcribed into the first form, see Lemma A2.1.1

PDE-FORM II:

Let W denote a Hilbert Space with inner product $\langle \cdot,\cdot \rangle_W$ and dual space W^*. Since W is a Hilbert Space, W^* can be identified with W. Let $\Lambda_L(Y,Z)$ denote the space of Lipschitz continuous functions from Banach Spaces Y into Z, with Lipschitz constant L, let $\mathrm{B}(Y,Z)$ denote the the space of bounded linear functions from Y into Z, let $\hat{A} : D(\hat{A}) \to X$ be the infinitesimal generator of a continuous semigroup $\{\hat{T}(t)\}_{t \geq 0}$, and let $B : \mathbb{R}^n \to \mathrm{B}(W,W)$, $C : \mathbb{R}^n \to W$, and $E : \mathbb{R}^n \to W$ be twice continuously Frechet differentiable maps [1], $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be a twice continuously differentiable function.

For $u \in O$, let $w(t,u) \in W$, $\forall\, t \in [0,1]$, and $z(t,u) \in \mathbb{R}^n$, $\forall\, t \in [0,1]$, denote the solution to:

$$\frac{d}{dt}w(t,u) = (\hat{A} + B(z(t,u)))w(t,u) + C(z(t,u)) + E(u(t)), \tag{3.2.3a}$$

$$w(0,u) = w_0 \in D(\hat{A}), \tag{3.2.3b}$$

$$\frac{d}{dt}z(t,u) = f(z(t,u),u(t)), \quad z(0,u) = z_0. \tag{3.2.3c}$$

We make the additional assumptions:

Assumption 3.2.2:

(i) There exist $\tilde{M}$, $\omega \in \mathbb{R}_+$ such that $\|T(t)\| \le \tilde{M}e^{\omega t}$ and $\|\hat{T}(t)\| \le \tilde{M}e^{\omega t}$, for all $t \ge 0$. Let $M \triangleq \tilde{M}e^{\omega}$.

(ii) There exists $K < \infty$ such that for all $z \in \mathbb{R}^n$ and all $u \in \tilde{U}$, $\|f(z,u)\| \le K(\|z\| + 1)$.

(iii) For every bounded set $S \subset \mathbb{R}^n$, there exists K_S such that for all $z, \hat{z} \in S$ and all $u, \hat{u} \in \tilde{U}$.

(a) $\|f(\hat{z},\hat{u}) - f(z,u)\| \le K_S[\|\hat{z} - z\| + \|\hat{u} - u\|]$,

(b) $\|\frac{\partial f}{\partial z}(\hat{z},\hat{u}) - \frac{\partial f}{\partial z}(z,u)\| \le K_S[\|\hat{z} - z\| + \|\hat{u} - u\|]$,

(c) $\|\frac{\partial f}{\partial u}(\hat{z},\hat{u}) - \frac{\partial f}{\partial u}(z,u)\| \le K_S[\|\hat{z} - z\| + \|\hat{u} - u\|]$.

(iv) For every bounded set $S \subset \mathbb{R}^n$, there exists L_S such that

(a) $B \in \Lambda_{L_S}(S, \mathbf{B}(W,W))$,

[1] B is Frechet differentiable at z, i.e., for $z \in \mathbb{R}^n$, There exists $B_z(z) \in \mathbf{B}(W, \mathbf{B}(\mathbb{R}^n, W))$ such that For all $w \in W$, $\lim_{\|\delta z\| \to 0} \frac{\|B(z + \delta z)w - B(z)w - B_z(z)w\delta z\|}{\|\delta z\|} = 0.$

(b) $C \in \Lambda_{L_S}(S, W)$,

(c) $B_z \in \Lambda_{L_S}(S, \mathbf{B}(W, \mathbf{B}(\mathbb{R}^n, W)))$,

(d) $C_z \in \Lambda_{L_S}(S, \mathbf{B}(\mathbb{R}^n, W))$,

(e) $E \in \Lambda_{L_S}(\mathbb{R}^m, W)$. ■

Then we have the following theorems:

Lemma 3.2.3: For all $u \in O$, The ordinary differential equation (3.2.3c) has a unique solution $z(\cdot, u)$, and there exists $b_1 < \infty$ such that for all $u \in O$, $\|z(t, u)\| \le b_1$ for all $t \in [0, 1]$.

Proof: The existence and uniqueness of the solution is established in Chapter 1 of Hale[Hal.1]. Next,

$$z(t, u) = z_0 + \int_0^t f(z(s, u), u(s))ds, \tag{3.2.4}$$

$$\|z(t, u)\| \le \|z_0\| + K\int_0^t (\|z(s, u)\| + 1)ds. \tag{3.2.5}$$

Defining $y(s, u) \triangleq \|z(s, u)\| + 1$,

$$y(s, u) \le y(0, u) + K\int_0^t y(s, u)ds. \tag{3.2.6}$$

Therefore, by the Bellman-Gronwall Lemma, $y(t, u) \le y(0, u)e^K$, for $t \in [0, 1]$. Define $b_1 \triangleq e^K(\|z_0\| + 1)$. ■

Theorem 3.2.4: The system (3.2.3a)-(3.2.3c) has a mild solution for all $u \in L_2 \cap L_\infty$.

Proof: Lemma 3.2.3 shows that $z(t, u)$ exists and that there exists b_1 such that $\|z(t, u)\| \le b_1$ for all $t \in [0, 1]$ and all $u \in O$. Consequently by Assumption 3.2.2, (3.2.3a) satisfies all assumptions in Theorem A2.3.1, with $X = W$, and so (3.2.3a) has a mild solution.

Let $S \triangleq \{ z \mid \|z\| \leq b_1 \}$. Then, by Assumption 3.2.2(ii), there exists $L < \infty$ such that $B(\cdot)$, $C(\cdot)$ and $E(\cdot)$ are Lipschitz continuous on S with constant L.

Lemma 3.2.5: Let O be defined as in (3.2.2.). Then there exists $b_2 \in (b_1, \infty)$ such that for all $u \in O$, $\|w(t,u)\| \leq b_2$, for all $t \in [0,1]$.

Proof:

$$w(t,u) = T(t)w_0 + \int_0^t T(t-s)[B(z(s,u))w(s,u) + C(z(s,u)) + E(u(s))]ds. \tag{3.2.8}$$

Since B, C, and E are Lipschitz continuous, and $\|z(s,u)\| \leq b_1$, there exists b_3 such that $\|B(z(s,u))\| \leq b_3$, $\|C(z(s,u))\| \leq b_3$, $\|E(u(s))\| \leq b_3$ for all $s \in [0,1]$, $u \in O$. Hence

$$\|w(t,u)\| \leq M\|w_0\| + \int_0^t M(b_3\|w(s,u)\| + 2b_3)ds. \tag{3.2.9}$$

The constant b_2 is determined using the Bellman-Gronwall Lemma. ■

This abstraction is general enough for the slewing problems we have studied. Appendix 2, Section 2 shows how the problem developed in Chapter 2 can be transcribed into PDE-FORM II.

3.3 DIFFERENTIABILITY OF PDE SOLUTION WITH RESPECT TO CONTROL

We will show that the mild solution to (3.1.2) is Frechet differentiable with respect to the control, $u(\cdot)$, in the $L_2 \cap L_\infty$ topology. We make the following assumptions:

Assumption 3.3.1:

(i) For every bounded set $S \subset X$, there exists K_S such that for all $x, \hat{x} \in S$ and all $u, \hat{u} \in \tilde{U}$,

(a) $\|F(\hat{x},\hat{u}) - F(x,u)\| \leq K_S[\|\hat{x} - x\| + \|\hat{u} - u\|]$,

(b) $\left\| \frac{\partial F}{\partial x}(\hat{x},\hat{u}) - \frac{\partial F}{\partial x}(x,u) \right\| \le K_S[\|\hat{x}-x\| + \|\hat{u}-u\|]$,

(c) $\left\| \frac{\partial F}{\partial u}(\hat{x},\hat{u}) - \frac{\partial F}{\partial u}(x,u) \right\| \le K_S[\|\hat{x}-x\|\|\hat{u}-u\|]$.

(ii) For each $u \in O$, a mild solution to (3.1.2) exists, and there exists $b_4 < \infty$ such that for all $u \in O$ and $t \in [0,1]$, $\|x(t,u)\| \le b_4$. ■

For a system whose dynamics are of PDE-FORM II, Assumption 3.3.1 (ii) is satisfied by Lemmas 3.2.3 and 3.2.5. However, Assumption 3.3.1 (ii) must be verified on a case by case basis for all systems whose dynamics are of PDE-FORM I but not PDE-FORM II.

Lemma 3.3.2: (Lipschitz Continuity.) Let $x(\cdot,\cdot)$ be the mild solution of (3.1.2). Then, there exists b_5 such that for any $u,\hat{u} \in O$, $t \in [0,1]$

$$\|x(t,\hat{u}) - x(t,u)\| \le b_5\|\hat{u}-u\|_1 \le b_5\|\hat{u}-u\|_2, \tag{3.3.1}$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the L_1 and L_2 norms.

Proof: For any $u,\hat{u} \in O$ and $t \in [0,1]$,

$$x(t,\hat{u}) - x(t,u) = \int_0^t T(t-s)[F(x(s,\hat{u}),\hat{u}(s)) - F(x(s,u),u(s))]ds. \tag{3.3.2}$$

Since $\{ x(t,u) \mid t \in [0,1],\ u \in O \} \subset S \triangleq \{ x \in X \mid \|x\| \le b_4 \}$, by Assumption 3.3.1(ib), there exists a constant K_S, such that with $y(t) \triangleq \|x(t,\hat{u}) - x(t,u)\|$, $t \in [0,1]$,

$$y(t) \le MK_S\int_0^t [\|y(s)\|\|\hat{u}(s) - u(s)\|]ds. \tag{3.3.3}$$

Applying the Bellman-Gronwall Lemma, we obtain[2]

$$y(t) \le MK_Se^{MK_S}\|\hat{u}-u\|_1 \le b_5\|\hat{u}-u\|_2 \tag{3.3.4}$$

[2] For $u \in L_2^m([0,1])$, $\|u\|_1 \le \|u\|_2$ by the Schwartz Inequality.

where $b_5 \triangleq MK_S e^{MK_S}$. ■

We now define a linear approximation to $x(\cdot, u + \delta u)$ where δu is a perturbation. For $\delta u \in L_2 \cap L_\infty$, we define $\delta x(\cdot, u, \delta u) \in C([0,1], X)$ to be the solution to:

$$\delta x(t) = \int_0^t T(t-s)[\frac{\partial F}{\partial x}(x(s,u), u(s))\delta x(s) + \frac{\partial F}{\partial u}(x(s,u), u(s))\delta u(s)]ds, \quad t \in [0,1]. \quad (3.3.5)$$

It is easy to show that such a solution exists.

Lemma 3.3.3: (Differentiability) Let $x(\cdot,\cdot)$ be the mild solution to (3.1.2). Then, there exists $b_6 < \infty$ such that for any $u, \hat{u} \in O$, $t \in [0,1]$

$$\| x(t,\hat{u}) - x(t,u) - \delta x(t,u,\hat{u}-u) \| \leq b_6 \|\hat{u} - u\|_2^2. \quad (3.3.6)$$

To simplify notation, we make the following definitions:

$$\delta u(t) \triangleq \hat{u}(t) - u(t), \quad \Delta x(t,u,\delta u) \triangleq x(t,\hat{u}) - x(t,u), \quad (3.3.7a)$$

$$\tilde{A}(s) \triangleq \frac{\partial F}{\partial x}(x(s,u), u(s)), \quad \tilde{B}(s) \triangleq \frac{\partial F}{\partial u}(x(s,u), u(s)). \quad (3.3.7b)$$

Proof: By Assumption 3.3.1, there exists $b_7 < \infty$ such that for all $u, \hat{u} \in O$, all $s \in [0,1]$, the following are true: $\|\delta x(s,u,\delta u)\| \leq b_7 \|\delta u\|$, $\|x(s,u)\| \leq b_7$, $\|\frac{\partial F}{\partial x}(x(s,u), u(s))\| \leq b_7$, $\|\frac{\partial F}{\partial u}(x(s,u), u(s))\| \leq b_7$. Therefore,

$$\|\Delta x(t,u,\delta u) - \delta x(t,u,\hat{u})\| \leq M \| \int_0^t \{ F(x(s,\hat{u}), \hat{u}(s)) - F(x(s,u), u(s))$$

$$- \tilde{A}(s)\delta x(s,u,\delta u) - \tilde{B}(s)\delta u(s) \} ds \|$$

$$\leq M \int_0^t \Big\{ \|\tilde{A}(s)(\Delta x(s,u,\delta u) - \delta x(s,u,\delta u))\| +$$

$$M\int_0^1 \|\frac{\partial F}{\partial x}(x(s,u) + r\Delta x(s,u,\delta u), u(s) + r\delta u(s)) - \tilde{A}(s)\| dr \|\Delta x(s,u,\delta u)\| +$$

$$M\int_0^1 \|\frac{\partial F}{\partial u}(x(s,u) + r\Delta x(s,u,\delta u), u(s) + r\delta u(s)) - \tilde{B}(s)\| dr \|\delta u(s)\| \Big\} ds$$

$$\leq M\int_0^t \Big\{ \|\tilde{A}(s)(\Delta x(s,u,\delta u) - \delta x(s,u,\delta u))\|$$

$$+ \int_0^1 K_S r(\|\Delta x(s,u,\delta u)\| + \|\delta u(s)\|) dr \|\Delta x(s,u,\delta u)\|$$

$$+ \int_0^1 K_S r(\|\Delta x(s,u,\delta u)\| + \|\delta u(s)\|) dr \|\delta u(s)\| \Big\} ds$$

$$\leq M\int_0^t \{ b_7\|\Delta x(s,u,\delta u) - \delta x(s,u,\delta u)\|$$

$$+ MK_S[\|\Delta x(s,u,\delta u)\| + \|\delta u(s)\|]^2 \} ds. \tag{3.3.8}$$

Since by Lemma 3.3.4, $\|\Delta x(s,u,\delta u)\| \leq b_5\|\delta u\|_2$,

$$\|\Delta x(t,u,\delta u) - \delta x(t,u,\hat{u})\| \leq M\int_0^t \{ b_7\|\Delta x(s,u,\delta u) - \delta x(s,u,\delta u)\| \} ds$$

$$+ MK_S[b_5\|\delta u\|_2 + \|\delta u(s)\|]^2. \tag{3.3.9}$$

Applying the Bellman-Gronwall Lemma,

$$\|\Delta x(t,u,\delta u) - \delta x(t,u,\hat{u})\| \leq b_6\|\delta u\|_2^2, \tag{3.3.10}$$

where $b_6 \triangleq MK_S(1 + b_5)^2 e^{Mb_7}$. ■

The map $\delta u \to \delta x(t,u,\delta u)$ is linear in δu for each $t \in [0,1]$, $u \in O$. If we call this map $x_u(t,u)$, then we have the following theorem:

Theorem 3.3.4: For all $u \in O$ and $t \in [0,1]$, $x(t,u)$ admits a Gateaux derivative in the $L_2 \cap L_\infty$ topology. That is, there exists a linear operator $x_u(t,u) \in \mathbf{B}(L_2 \cap L_\infty, X)$ such that for all $\delta u \in L_2 \cap L_\infty$

$$\lim_{s \to 0} \frac{x(t,u+s\delta u) - x(t,u)}{s} = x_u(u,t)\delta u. \tag{3.3.11}$$
■

So far we have stated that for $u \in O$, $\delta u \in L_2 \cap L_\infty$, the directional derivative $x_u(t,u)\delta u$ exists. We now seek to find the differential $x_u(t,u) : L_2 \cap L_\infty \to X$ explicitly. We begin by defining $w(\cdot, u, \delta u) \in C([0,1],X)$, to be the mild solution to

$$\frac{d}{dt} w(t) = (A + \overline{B}(t,u))w(t) + \overline{C}(t,u)\delta u(t), \; w(0) = 0 \tag{3.3.12}$$

with $\overline{B}(t,u) \triangleq \frac{\partial F}{\partial x}(x(t,u),u(t))$, $\overline{C}(t,u) \triangleq \frac{\partial F}{\partial u}(x(t,u),u(t))$.

Since A is an infinitesimal generator,

$\overline{B}(\cdot,u) \in L_2^m([0,1],\mathbf{B}(X,X)) \cap L_\infty^m([0,1],\mathbf{B}(X,X))$ and , $\overline{C}(\cdot,u) \in L_2^m([0,1],X) \cap L_\infty^m([0,1],X)$, we can apply Lemma A1.3 (see Appendix I) to determine an evolution operator $U^u(t,s)$, $0 \le s \le t \le 1$ such that

$$w(t,u,\delta u) = \int_0^t U^u(t,s)\overline{C}(s,u)\delta u(s)ds \tag{3.3.13}$$

is the mild solution to (3.3.12). $w(t,u,\delta u)$ also satisfies:

$$w(t,u,\delta u) = \int_0^t T(t-s)[\overline{B}(s,u)w(s,u,\delta u) + \overline{C}(s,u)\delta u(s)]ds. \tag{3.3.14a}$$

Comparing with (3.3.8), $w(t,u,\delta u) = \delta x(t,u,\delta u)$. Therefore,

$$x_u(t,u)\delta u = \int_0^t U^u(t,s)\overline{C}(s,u)\delta u(s)ds. \tag{3.3.14b}$$

Next, we show that $x_u(t,u)$ is Lipschitz continuous in $u \in O$.

Lemma 3.3.5: (Lipschitz Continuity of $x_u(t,u)$ in u.)

There exists $b_9 < \infty$ such that for all $t \in [0,1]$, $u \in O$, $\hat{u} \in O$

$$\sup_{t \in [0,1]} \sup_{\delta u \in L_2 \cap L_\infty} \frac{|x_u(t,\hat{u})\delta u - x_u(t,u)\delta u|}{|\delta u|_2} \le b_9 |\hat{u} - u|_2. \tag{3.3.15}$$

Proof: For $\hat{u} \in O$, $\delta x(t,\hat{u},\delta u) = x_u(t,\hat{u})\delta u$ satisfies

$$\delta x(t,\hat{u},\delta u) = \int_0^t T(t-s)\{ (\frac{\partial F}{\partial x}(x(s,\hat{u}),\hat{u}(s))\delta x(s,\hat{u},\delta u) + \frac{\partial F}{\partial u}(x(s,\hat{u}),\hat{u}(s))\delta u(s)) \}ds. \tag{3.3.16}$$

Therefore, $x_u(t,\hat{u})\delta u - x_u(t,u)\delta u = \delta x(t,\hat{u},\delta u) - \delta x(t,u,\delta u)$ and hence

$$|\delta x(t,\hat{u},\delta u) - \delta x(t,u,\delta u)| \le M|\int_0^t \{ \frac{\partial F}{\partial u}(x(s,\hat{u}),\hat{u}(s)) - \frac{\partial F}{\partial u}(x(s,u),u(s)) \}\delta u(s)ds$$

$$+ \int_0^t \{ \frac{\partial F}{\partial x}(x(s,\hat{u}),\hat{u})(\delta x(s,\hat{u},\delta u) - \delta x(s,u,\delta u))$$

$$+ (\frac{\partial F}{\partial x}(x(s,\hat{u}),\hat{u}(s)) - \frac{\partial F}{\partial x}(x(s,u),u(s)))\delta x(s,u,\delta u) \}ds|$$

$$\le M\int_0^t K_S[|x(s,\hat{u}) - x(s,u)| + |\hat{u}(s) - u(s)||][|\delta u(s)| + |\delta x(s,u,\delta u)ds|]ds$$

$$+ \int_0^t Mb_7|\delta x(s,\hat{u}) - \delta x(s,u)|ds$$

$$\le K_S[(1+b_5)|\hat{u} - u|_2(|\delta u|_2 + b_7|\delta u|_2)] + \int_0^t Mb_7|\delta x(s,\hat{u}) - \delta x(s,u)|ds. \tag{3.3.17}$$

Applying the Bellman-Gronwall Lemma,

$$|\delta x(t,\hat{u},\delta u) - \delta x(t,u,\delta u)| \le MK_S e^{Mb_7}(1+b_5)|\hat{u} - u|_2(1+b_7)|\delta u|_2, \tag{3.3.18}$$

and therefore,

$$\frac{|x_u(t,\hat{u})\delta u - x_u(t,u)\delta u|}{|\delta u|_2} \le b_9|\hat{u} - u|_2, \tag{3.3.19}$$

where $b_9 \triangleq MK_5 e^{Mb_7}(1 + b_7)(1 + b_5)$. ■

3.4 EXISTENCE OF GRADIENTS

Let O be defined as in (3.2.2) and consider the function $g : O \to \mathbb{R}$ defined by $g(u) = h(x(1,u))$ where $h : X \to \mathbb{R}$ is continuously differentiable and $x(\cdot, u)$ is the solution to (3.1.2). We denote the derivative of $h(\cdot)$ at x by $Dh(x)$. We will establish the existence of a gradient of $g(\cdot)$ at u for all $u \in O$.

Lemma 3.4.1:

(i) The function $g : O \to \mathbb{R}$ is continuous, i.e., For any sequence $\{ u_i \} \subset O$ which converges to $u \in O$ in the $L_2 \cap L_\infty$ topology, $\lim_{u_i \to u} g(u_i) = g(u)$.

(ii) The function $g(\cdot)$ admits a Gateaux derivative in the $L_2 \cap L_\infty$ topology at all $u \in O$. That is, there exists a linear operator $Dg(u) \in \mathbf{B}(L_2 \cap L_\infty, \mathbb{R})$ such that for all $\delta u \in L_2 \cap L_\infty$

$$\lim_{s \to 0} \frac{g(u + s\delta u) - g(u)}{s} = Dg(u)\delta u.$$

(iii) For every $u \in O$, there exists a unique $\nabla g(u) \in L_2 \cap L_\infty$ such that for all $\hat{u} \in O$,

$$\langle \nabla g(u), \hat{u} - u \rangle = Dg(u)(\hat{u} - u).$$

(iv) The function $\nabla g : L_2 \cap L_\infty \to L_2^m([0,1])$ is continuous.

(v) If the derivative of $h(\cdot)$, $Dh(\cdot)$, is Lipschitz continuous on bounded sets then there exists $b_{10} < \infty$ such that for all $u, v \in G$, $|\nabla g(u) - \nabla g(v)| \le b_{10}|u - v|$ and therefore $\nabla g(\cdot)$ is uniformly Lipschitz continuous.

Proof: (i) By Lemma 3.3.4, $x(1,\cdot)$ is continuous in u. Since h is continuous in x, $g(\cdot)$ is continuous in u.

(ii) For $u \in O$ and $\delta u \in L_2 \cap L_\infty$ define

$$Dg(u)\delta u \triangleq Dh(x(1,u))\delta x(1,u,\delta u). \tag{3.4.1}$$

For s sufficiently small, $u + s\delta u \in 0$ and

$$\frac{1}{s}|g(u+s\delta u) - g(u) - sDg(u)\delta u| = \frac{1}{s}|h(x(1,u+s\delta u)) - h(x(1,u)) - sDh(x(1,u))\delta x(1,u,\delta u)|$$

$$= \frac{1}{s}|\int_0^1 (Dh(x(1,u) + r(x(1,u+s\delta u) - x(1,u))) - Dh(x(1,u)))(x(1,u+s\delta u) - x(1,u))dr$$

$$+ Dh(x(1,u))(x(1,u+s\delta u) - x(1,u) - s\delta x(1,u,\delta u))|$$

$$\le \frac{1}{s}\int_0^1 \|Dh(x(1,u) + r(x(1,u+s\delta u) - x(1,u))) - Dh(x(1,u))\| dr b_7 \|\delta u\|_2$$

$$+ \|Dh(x(1,u))\| b_6 \|\delta u\|_2^2 s.$$

Therefore,

$$\lim_{s\to 0} \frac{1}{s}|g(u+s\delta u) - g(u) - sDg(u)\delta u| = 0. \tag{3.4.3}$$

(iii) Since X is a Hilbert space, we can identify X^* with X. Therefore, we denote the adjoint of $\overline{C}(s,u) \in B(\mathbb{R}^m, X)$ by $\overline{C}(s,u)^* \in B(X,\mathbb{R}^m)$; the adjoint of $U^u(t,s) \in B(X,X)$ by $U^u(t,s)^* \in B(X,X)$ and the adjoint of $Dh(x) \in B(X,\mathbb{R})$ by $Dh(x)^* \in B(\mathbb{R},X)$ which we identify with $\nabla h(x) \in X$. We define

$$\nabla g(u)(s) \triangleq \overline{C}(s,u)^* U^u(1,s)^* \nabla h(x(1,u)) \in \mathbb{R}^m. \tag{3.4.4}$$

First, we show that $\nabla g(\cdot) \in L_2 \cap L_\infty$. Since $\|\overline{C}(s,u)^*\| = \|\overline{C}(s,u)\| \le b_7$ and $\|U^u(1,s)^*\| = \|U^u(1,s)\| \le M$, for all $s \in [0,1], u \in O$, $\|\nabla g(u)(s)\| \le b_7 M \|Dh(x(1,u))\|$, so

$\nabla g(u)(\cdot)$ is bounded. Since it is also measurable, $\nabla g(u) \in L_2 \cap L_\infty$. Second,

$$\langle \nabla g(u), \hat{u} - u \rangle = \int_0^1 \langle \bar{C}(s,u)^* U^u(1,s)^* \nabla h(x(1,u)), \hat{u}(s) - u(s) \rangle ds$$

$$= \langle \nabla h(x(1,u)), \int_0^1 U^u(1,s)\bar{C}(s,u)(\hat{u}(s) - u(s))ds \rangle$$

$$= Dh(x(1,u))x_u(1,u)(\hat{u} - u) = Dg(u)(\hat{u} - u). \tag{3.4.5}$$

For $s \in [0,1]$, we define $p(s,u) \triangleq U^u(1,s)^* \nabla h(x(1,u))$. By Lemma A1.4, $p(\cdot,\cdot)$ is the mild solution to the problem:

$$\frac{d}{dt}p(t) = -(A^* + \bar{B}(t)^*)p(t), \quad p(1) = \nabla h(x(1,u)), \tag{3.4.6}$$

and

$$\nabla g(u)(s) = \bar{C}(s,u)^* p(s). \tag{3.4.7}$$

(iv) For any sequence $\{ u_i \} \subset O$ such that $u_i \to u \in O$,

$$\|\nabla g(u_i) - \nabla g(u)\| = \|Dg(u_i) - Dg(u)\| \tag{3.4.8}$$

$$\leq \|(Dh(x(1,u_i)) - Dh(x(1,u)))x_u(1,u_i)\| + \|Dh(1,u)(x_u(1,u_i) - x_u(1,u))\|$$

By Lemma 3.3.5 and the continuity of $Dh(\cdot)$,

$$\|\nabla g(u_i) - \nabla g(u)\| \to 0. \tag{3.4.9}$$

(v) Since $\|x(t,u)\| \leq b_4$ for all $t \in [0,1]$ and $u \in G$, there exists $b_{11} < \infty$ such that $\|Dh(x(1,v)) - Dh(x(1,u))\| b_{11} \|u - v\|$ for all $u, v \in G$. Therefore,

$$\|\nabla g(v) - \nabla g(u)\| \leq \|(Dh(x(1,v)) - Dh(x(1,u)))x_u(1,v)\| + Dh(x(1,u))(x_u(1,v) - x_u(1,u)\|$$

$$\leq [b_{11}\|x_u(1,v)\| + b_9\|Dh(x(1,u))\|]\|u - v\|, \tag{3.4.10}$$

Since there exists b_{12} such that $\|x_u(t,u)\| \leq b_{12}$ and $Dh(x(t,u))\| \leq b_{12}$ for $t \in [0,1]$ and $u \in G$,

$$\|\nabla g(v) - \nabla g(u)\| \leq b_{10}\|u - v\|, \tag{3.4.11}$$

where $b_{10} \triangleq (b_{11} + b_9)b_{12}$. ■

Lemma 3.4.2: (First Order Expansion) For any $\hat{u}, u \in O$,

$$g(\hat{u}) - g(u) = \int_0^1 \langle \nabla g(u + s(\hat{u} - u)), \hat{u} - u \rangle ds. \tag{3.4.12}$$

Proof: For $\hat{u}, u \in O$, we define $e(s) \triangleq g(u + s(\hat{u} - u))$. We now show that $e(\cdot)$ is differentiable at all $s \in (0, 1)$. Choose $s \in (0, 1)$, then

$$\frac{e(s + \Delta s) - e(s)}{\Delta s} = \frac{g(u + (s + \Delta s)(\hat{u} - u)) - g(u + s(\hat{u} - u))}{\Delta s}. \tag{3.4.13}$$

Set $z \triangleq u + s(\hat{u} - u) \in O$ and $\delta u \triangleq \hat{u} - u$. Then $\delta u \in L_2 \cap L_\infty$ and

$$\lim_{\Delta s \to 0} \frac{e(s + \Delta s) - e(s)}{\Delta s} = \lim_{\Delta s \to 0} \frac{g(z + \Delta s \delta u) - g(z)}{\Delta s} = Dg(z)\delta u \tag{3.4.14}$$

by Lemma 3.4.1(ii). Since $e(1) - e(0) = \int_0^1 e'(s)ds$,

$$g(\hat{u}) - g(u) = \int_0^1 \langle \nabla g(u + s(\hat{u} - u)), \hat{u} - u \rangle ds. \tag{3.4.15}$$

■

3.5 OPTIMIZATION ALGORITHMS I

In Section 3.4, we have shown that the gradient of $g(\cdot)$ exists and is continuous. We will show that this fact can be used to prove convergence of the following algorithm for solving problem **P**:

$$\mathbf{P} : \inf_{u \in G} \{ g(u) \} \tag{3.5.0}$$

Algorithm 3.5.1:

Data: $u_0 \in G$, $\alpha \in (0,1)$, $\beta \in (0,1)$.

Step 0: $i = 0$.

Step 1: Compute $\nabla g(u_i)$ using (3.4.6) and (3.4.7).

Step 2: Compute $\Theta(u_i)$ and v_i using

$$\Theta(u_i) \triangleq \min_{v \in G} \{ \frac{1}{2} \| v - u_i \|^2 + \langle \nabla g(u_i), v - u_i \rangle \}. \tag{3.5.1a}$$

$$v_i = v(u_i) \triangleq \arg\min_{v \in G} \{ \frac{1}{2} \| v - u_i \|^2 + \langle \nabla g(u_i), v - u_i \rangle \}. \tag{3.5.1b}$$

If $\Theta(u_i) = 0$, STOP.

Step 3: Compute the step size λ_i using

$$\lambda_i = \lambda(u_i) \triangleq \max\{ \lambda \in \{ 0, 1, \beta, \beta^2, \cdots \} \mid h(x(1, u_i + \lambda(v_i - u_i))) - h(x(1, u_i)) < \alpha\lambda\Theta(u_i) \}. \tag{3.5.2}$$

Step 4: Set $u_{i+1} = u_i + \lambda_i(v_i - u_i)$.

Step 5: Set $i = i + 1$; go to Step 1. ■

First, we show that $\Theta(\hat{u}) = 0$ is a necessary condition for $\hat{u}$ to be a minimizer of **P** and that Algorithm 3.5.1 is well-defined.

Lemma 3.5.2: If $\hat{u} \in O$ is such that $\Theta(\hat{u}) < 0$, then there exists $\hat{\lambda} \in (0,1)$ such that for all $\lambda \in (0, \hat{\lambda}]$,

$$g(\hat{u} + \lambda(v(\hat{u}) - \hat{u})) - g(\hat{u}) \le \lambda\alpha\Theta(\hat{u}), \tag{3.5.3a}$$

and therefore (i) $\hat{u}$ is not a minimizer of **P** and (ii) Algorithm 3.5.1 does not jam up in Step 3 at $\hat{u}$.

Proof: Let $\hat{v} \triangleq v(\hat{u})$. Then $\langle \nabla g(\hat{u}), \hat{v} - \hat{u} \rangle < \Theta(\hat{u}) = -\delta \le 0$. Therefore,

$$g(\hat{u} + \lambda(\hat{v} - \hat{u})) - g(\hat{u}) - \lambda\alpha\Theta(\hat{u})$$

$$= \lambda \langle \nabla g(\hat{u}), \hat{v} - \hat{u} \rangle - \alpha\lambda\Theta(\hat{u}) + \lambda\int_0^1 \langle \nabla g(\hat{u} + s\lambda(\hat{v} - \hat{u})) - \nabla g(\hat{u}), \hat{v} - \hat{u} \rangle ds$$

$$\le \lambda(1 - \alpha)\Theta(\hat{u}) + \int_0^1 \langle \nabla g(\hat{u} + s\lambda(\hat{v} - \hat{u})) - \nabla g(\hat{u}), \hat{v} - \hat{u} \rangle ds.$$

$$\le \lambda(- (1 - \alpha)\delta + \int_0^1 \langle \nabla g(\hat{u} + s\lambda(\hat{v} - \hat{u})) - \nabla g(\hat{u}), \hat{v} - \hat{u} \rangle ds. \tag{3.5.3b}$$

Since $\nabla g(\cdot)$ is continuous, there exists $\hat{\lambda} \in (0, 1)$ such that for all $\lambda \in (0, \hat{\lambda}]$, the right hand side of (3.5.3b) is less than zero, so that $g(\hat{u} + \hat{\lambda}(\hat{v} - \hat{u})) < g(\hat{u})$. Since $\hat{u} + \hat{\lambda}(\hat{v} - \hat{u}) \in G$, $\hat{u}$ is not a minimizer for **P**, and $\lambda(\hat{u}) \ge \beta\hat{\lambda}$ so that the algorithm does not jam up in Step 3 at $\hat{u}$. ∎

We shall prove that $v(\cdot)$ and $\Theta(\cdot)$ are continuous. We state a preliminary lemma.

Lemma 3.5.3: Let $C \subset \mathbb{R}^m$ be a compact, convex set, and x, y be arbitrary points in $\mathbb{R}^m$. If $\bar{x}$ and $\bar{y}$ are the closest points in C to x and y respectively, i.e.,

$$\bar{x} \triangleq \arg\min_{\xi \in C} \{ |\xi - x|^2 \}, \tag{3.5.4}$$

$$\bar{y} \triangleq \arg\min_{\xi \in C} \{ |\xi - y|^2 \}, \tag{3.5.5}$$

then

$$|\bar{x} - \bar{y}|^2 \le |x - y|^2. \tag{3.5.6}$$

Proof: Assume that for some $x, y \in \mathbb{R}^m$, $|\bar{x} - \bar{y}|^2 > |x - y|^2$. Define

$$\lambda_x \triangleq \arg\min \{ \lambda \in \mathbb{R} \mid |(1 - \lambda)\bar{x} + \lambda\bar{y} - x|^2 \}, \tag{3.5.7}$$

$$\lambda_y \triangleq \arg\min \{ \lambda \in \mathbb{R} \mid |(1 - \lambda)\bar{x} + \lambda\bar{y} - y|^2 \}, \tag{3.5.8}$$

$$\hat{x} \triangleq (1 - \lambda_x)\bar{x} + \lambda_x\bar{y}, \tag{3.5.9}$$

$$\hat{y} \triangleq (1-\lambda_y)\bar{x} + \lambda_y \bar{y}. \tag{3.5.10}$$

Then $\hat{x}$ is the point on the line through $\bar{x}$ and $\bar{y}$ nearest to x, and $\hat{y}$ is the point on the line through $\bar{x}$ and $\bar{y}$ nearest to y. Define a co-ordinate system for $\mathbb{R}^m$ with $\bar{x}$ as the origin, and chose an orthogonal basis $\{ e_i \}_{i \in \underline{m}}$ where $e_1 \triangleq \dfrac{\bar{y}-\bar{x}}{\|\bar{y}-\bar{x}\|}$. We can find $\alpha, \beta \in \mathbb{R}^m$ so that $x = \sum_{i=1}^{m} \alpha_i e_i$ and $y = \sum_{i=1}^{m} \beta_i e_i$. Then $\hat{x} = \alpha_1 e_1$ and $\hat{y} = \beta_1 e_1$, and therefore

$$\|\hat{x} - \hat{y}\|^2 = (\alpha_1 - \beta_1)^2 \le \sum_{i=1}^{m} (\alpha_i - \beta_i)^2 = \|x - y\|^2. \tag{3.5.11}$$

Now, we claim that $\lambda_y \ge 1$. Let us assume that $\lambda_y \in [0,1)$. Since C is convex, $\hat{y} \in C$, but this is not possible since $\|\hat{y} - y\| < \|\bar{y} - y\|^2$. Assume that $\lambda_y < 0$. Since $\|\cdot\|$ is strictly convex,

$$\|(1-\rho)\hat{y} + \rho\bar{y} - y\|^2 < (1-\rho)\|\hat{y} - y\|^2 + \rho\|\bar{y} - y\|^2 \le \|\bar{y} - y\|^2 \tag{3.5.12}$$

for $\rho \in (0,1)$, and there exists $\rho \in (0,1)$ such that $(1-\rho)\hat{y} + \rho\bar{y} = \bar{x}$ so that $\|\bar{x} - y\| < \|\bar{y} - y\|$ which is a contradiction. Similarly, $\lambda_x \le 0$, and therefore $\|\hat{x} - \hat{y}\| \ge \|\bar{x} - \bar{y}\| > \|x - y\|$ which contradicts (3.5.11). ■

Theorem 3.5.4: The function $v(\cdot)$ is continuous.

Proof: Let $\{ u_i \} \subset G$ be a sequence such that $u_i \to \hat{u} \in G$. For any $u \in G$,

$$v(u)(t) = \arg\min_{u \in U} \{ \tfrac{1}{2}\|v - (u(t) - \nabla g(u)(t))\|^2 \}, \tag{3.5.13a}$$

and so by Theorem 3.5.3,

$$\|v(u_j)(t) - v(u_i)(t)\|^2 \le \|(u_j(t) - \nabla g(u_j)(t)) - (u_i(t) - \nabla g(u_i)(t))\|^2. \tag{3.5.13b}$$

Since $(a+b)^2 \le 2[a^2 + b^2]$,

$$|v(u_j)(t) - v(u_i)(t)|^2 \le 2[|u_j(t) - u_i(t)|^2 + |\nabla g(u_j)(t) - \nabla g(u_i)(t)|^2]. \tag{3.5.13c}$$

Therefore,

$$\|v(u_j) - v(u_i)\|_2^2 \le 2\|u_j - u_i\|^2 + 2\|\nabla g(u_j) - \nabla g(u_i)\|^2. \tag{3.5.14}$$

Since $u_i \to \hat{u}$ and $\nabla g(u_i) \to \nabla g(\hat{u})$, $\{ u_i \}$ and $\{ \nabla g(u_i) \}$ are Cauchy sequences, and consequently $\{ v(u_i) \}$ is a Cauchy sequence, and since $L_2 \cap L_\infty$ is a Banach space, there exists $\hat{v} \in G$ such that $v(u_i) \to \hat{v}$. Suppose $\hat{v} \ne v(\hat{u})$. Then

$$\frac{1}{2}\|v(\hat{u}) - \hat{u}\|^2 + \langle \nabla g(\hat{u}), v(\hat{u}) - \hat{u} \rangle < \frac{1}{2}\|\hat{v} - \hat{u}\|^2 + \langle \nabla g(\hat{u}), \hat{v} - \hat{u} \rangle. \tag{3.5.15}$$

Since $u_i \to \hat{u}$ and $\nabla g(u_i) \to \nabla g(\hat{u})$, there exists i such that

$$\frac{1}{2}\|v(\hat{u}) - u_i\|^2 + \langle \nabla g(u_i), v(\hat{u}) - u_i \rangle < \frac{1}{2}\|v_i - u_i\|^2 + \langle \nabla g(u_i), v_i - u_i \rangle. \tag{3.5.16}$$

which is a contradiction. Therefore $v(u_i) \to v(\hat{u})$, and $v(\cdot)$ is continuous. ■

Corollary 3.5.5: The function $\Theta(\cdot)$ is continuous. ■

Theorem 3.5.6: If $\{ u_i \}$ is an infinite sequence generated by Algorithm 3.5.1, then any accumulation point $\hat{u} \in G$ satisfies the optimality condition $\Theta(\hat{u}) = 0$.

Proof: Assume that $\hat{u} \in G$ is an accumulation point of $\{ u_i \}$ such that $\Theta(\hat{u}) = -\delta < 0$. Let

$$\hat{k} = \min\{ k \in \{ 0, 1, 2, \cdots \} \mid g(\hat{u} + \beta^k(\hat{v} - \hat{u})) - g(\hat{u}) < \alpha\beta^k\Theta(\hat{u}) \}. \tag{3.5.17}$$

Such a $\hat{k} < \infty$ exists by Lemma 3.5.2. By the continuity of $g(\cdot)$, $\nabla g(\cdot)$, and $\hat{v}(\cdot)$, there exists i_0 such that for all $i > i_0$,

$$g(u_i + \beta^{\hat{k}}(v_i - u_i)) - g(u_i) < \alpha\beta^{\hat{k}}\Theta(u_i) < \frac{\alpha}{2}\beta^{\hat{k}}\Theta(\hat{u}). \tag{3.5.18}$$

Therefore $\lambda_i \ge \beta^{\hat{k}}$ and hence $g(u_{i+1}) - g(u_i) < \frac{\alpha}{2}\beta^{\hat{k}}\, \Theta(\hat{u})$, and $g(u_i) \to -\infty$ which is a contradiction since $g(\cdot)$ is continuous, and the theorem is proved. ■

CHAPTER 4

ABSTRACT FREE-TIME FORMULATION

In problem P (3.5.0), the final time is fixed at $\tau = 1$. Using suitable scaling, the final time can be fixed at any value. In this chapter, we consider free-time problems, and show that by using a particular transcription, a free-time problem can be cast into a form similar to that of the fixed-time problem. By extending the results of the previous chapter on the existence of continuous gradients, we can apply an algorithm similar to one used for the fixed-time problem.

4.1 ANALYTIC SEMIGROUP

We make one additional assumption on the operator A defined in Section 3.1:

Assumption 4.1.1: For some $\delta \in (0, \frac{\pi}{2})$, the semigroup generated by A, $\{ T(t) \}_{t \geq 0}$, can be extended to be analytic in a sector $\Delta_\delta \triangleq \{ z \mid |arg\ z| \leq \delta \}$, and $\|T(z)\|$ is uniformly bounded in every closed subsector $\overline{\Delta}_{\delta'}$, $\delta' \leq \delta$, of Δ_δ. The semigroup $\{ T(t) \}_{t \geq 0}$ is then called an *analytic semigroup*. ■

Showalter [Sho.1] has shown that for flexible structures with Kelvin-Voigt damping and Chen and Russel [Che.1] have shown that for flexible structures with proportional damping, the operator A generates an analytic semigroup. Appendix II (A2.2) shows that for the flexible beam ,(2.2.5)-(2.2.7), A generates an analytic semigroup.

We take the following theorem from Pazy [Paz.1]:

Theorem 4.1.2: The semigroup $\{ T(t) \}_{t \geq 0}$ generated by the operator A is analytic if and only if there exists a constant $\tilde{C} < \infty$ such that (i) $T(t)$ is differentiable in $t > 0$; (ii)

$\frac{d}{dt}T(t) = AT(t)$; and (iii) $\|AT(t)\| \le \frac{\tilde{C}}{t}$, for all $t > 0$. ■

Theorem 4.1.2 implies local Lipschitz continuity of $T(t)$. Furthermore, the following holds.

Lemma 4.1.3: For all $t \ge 0$ and $\tau \in [\tau_{min}, \tau_{max}]$, with $0 < \tau_{min} \le \tau_{max} < \infty$, the map $\tau \to T(\tau t)$ is Lipschitz continuous uniformly in τ, i.e., there exists $C < \infty$ such that for all $\hat{\tau}, \tau \in [\tau_{min}, \tau_{max}]$ and all $t \ge 0$, $\|T(\hat{\tau}t) - T(\tau t)\| \le C|\hat{\tau} - \tau|$.

Proof: Since $T(\cdot)$ is differentiable for $t > 0$,

$$\|T(\hat{\tau}t) - T(\tau t)\| = \|\int_{\tau t}^{\hat{\tau}t} AT(s)ds\| = \|\int_{\tau}^{\hat{\tau}} AT(ts)tds\| \le \int_{\tau}^{\hat{\tau}} \frac{\tilde{C}}{ts}tds = \tilde{C}\int_{\tau}^{\hat{\tau}} \frac{1}{s}ds \le \tilde{C}\int_{\tau}^{\hat{\tau}} \frac{1}{\tau_{min}}ds$$

$$\le C|\hat{\tau} - \tau|, \tag{4.1.1}$$

where $C \triangleq \tilde{C}/\tau_{min}$. ■

We next prove that $\frac{d}{dt}T(\cdot)$ is continuous. If A is bounded then since $\frac{d}{dt}T(t) = AT(t)$, $\frac{d}{dt}T(\cdot)$ is continuous. Since A is unbounded, the proof is more involved.

Lemma: 4.1.4 For $t > 0$, $\frac{d}{dt}T(t) = AT(t)$ is continuous.

Proof: Since A is analytic, $T'(t) \triangleq \frac{d}{dt}T(t)$ exists for all $t > 0$ and from Pazy [Paz.1, page 61],

$$T'(t) = \frac{1}{(2\pi i)}\int_0^{\infty} \rho e^{i\psi} e^{t\rho e^{i\psi}} R(\rho e^{i\psi}, A)d(\rho e^{i\psi}) + \frac{1}{(2\pi i)}\int_{-\infty}^{0} \rho e^{-i\psi} e^{t\rho e^{-i\psi}} R(\rho e^{-i\psi}, A)d(\rho e^{-i\psi}) \tag{4.1.3}$$

and $\|R(\rho e^{i\psi}, A)\| = \|R(\rho e^{-i\psi}, A)\| \le \frac{M}{\rho}$ where, $\psi \in (\pi/2, \pi/2 + \delta)$ and $R(\lambda, A) = (\lambda I - A)^{-1}$.

We assume without loss of generality that $t \to s$ and $t \geq s > 0$. Then

$$\|T(t) - T(s)\| \leq \frac{M}{\pi}\int_0^\infty |e^{t\rho e^{i\psi}} - e^{s\rho e^{i\psi}}|d\rho \tag{4.1.4}$$

$$\leq \frac{M}{\pi}\int_0^\infty e^{s\rho\cos\psi}|e^{(t-s)\rho\cos\psi + it\rho\sin\psi} - e^{is\rho\sin\psi}|d\rho. \tag{4.1.5}$$

Since $\cos\psi < 0$ and $\rho \in [0,\infty]$, $|e^{(t-s)\rho\cos\psi + it\rho\sin\psi}| \leq 1$ and $|e^{is\rho\sin\psi}| \leq 1$ for all $t \geq s > 0$, $\rho > 0$. Since the function e^x is Lipschitz continuous on bounded sets, there exists b such that

$$|e^{(t-s)\rho\cos\psi + it\rho\sin\psi} - e^{is\rho\sin\psi}| \leq |(t-s)\rho e^{i\psi}|b = \rho|t-s|b. \tag{4.1.6}$$

Therefore, integrating by parts:

$$\|T(t) - T(s)\| \leq |t-s|\frac{Mb}{\pi}\int_0^\infty \rho e^{s\rho\cos\psi}d\rho \leq \frac{Mb}{\pi(s\cos\psi)^2}|t-s|. \tag{4.1.7}$$

Hence, $T(t) \to T(s)$ as $t \to s$. ■

Theorem 4.1.5: For $\hat{\tau}$ and $\tau > \tau_{min}$, the following inequality holds:

$$(t-s)\|AT(\hat{\tau}(t-s)) - AT(\tau(t-s))\| \leq \frac{Mb|\hat{\tau}-\tau|}{\pi\tau_{min}^2\cos^2\psi} \quad \text{for } 1 \geq t \geq s \geq 0. \tag{4.1.8a}$$

Proof: Set $t = \hat{\tau}(t-s)$ and $s = \tau(t-s)$ in (4.1.7). (4.1.7) becomes:

$$\|AT(\hat{\tau}(t-s)) - AT(\tau(t-s))\| \leq \frac{Mb|(\hat{\tau}-\tau)(t-s)|}{\pi\tau^2\cos^2\psi|t-s|^2} = \frac{Mb}{\pi\tau^2\cos^2\psi}\frac{|\hat{\tau}-\tau||t-s|}{|t-s|^2}. \tag{4.1.8b}$$

Since $\tau \geq \tau_{min}$, (4.1.8a) is obtained from (4.1.8b). ■

With these additional properties, we can examine free-time problems. We state the canonical free-time, control constrained problem $\tilde{\mathbf{P}}_F$. We require τ to be in the interval $[\tau_{min}, \tau_{max}]$ where $\tau_{min} > 0$ and $\tau_{max} < \infty$.

$$\tilde{\mathbf{P}}_F : \inf \{ \tilde{g}(\tilde{u},\tau) \mid \tilde{u} \in G(\tau),\ \tau \in [\tau_{min},\tau_{max}] \}, \tag{4.1.9}$$

where $\tilde{g}(\tilde{u},\tau) = h(\tilde{x}(\tau,\tilde{u}))$, X is the Hilbert space defined in Section 3.1, $h : X \to \mathbb{R}$, is continuously differentiable, $G(\tau) \triangleq \{ u \in L_2^m([0,\tau]) \cap L_\infty^m([0,\tau]) \mid u(s) \in U,\ s \in [0,\tau] \}$, U is a compact, convex subset of $\mathbb{R}^m$, and $\tilde{x}(\cdot,\tilde{u})$ is the mild solution to the system:

$$\frac{d}{ds}\tilde{x}(s,\tilde{u}) = A\tilde{x}(s,\tilde{u}) + F(\tilde{x}(s,\tilde{u}),\tilde{u}(s)), \quad s \in [0,\tau], \quad \tilde{x}(0,\tilde{u}) = x_0 \in D(A). \tag{4.1.10}$$

4.2 TIME SCALING

We shall transcribe problem $\tilde{\mathbf{P}}_F$ into a fixed-time problem. First, by referring to (3.1.1), we see that $G = G(1)$. Next, with each $\tilde{u} \in G(\tau)$, we associate a $u \in G$ defined by $u(t) \triangleq \tilde{u}(\tau t)$ for $t \in [0,1]$. With each $\tilde{x} \in C([0,\tau],X)$, we associate $x \in C([0,1],X)$ defined by $x(t,u,\tau) \triangleq \tilde{x}(t\tau,\tilde{u})$ for all $t \in [0,1]$. Then,

$$\frac{d}{dt}x(t,u,\tau) = \frac{d}{dt}\tilde{x}(t\tau,\tilde{u}) = \tau[A\tilde{x}(t\tau,\tilde{u}) + F(\tilde{x}(t\tau),\tilde{u}(t\tau))] \tag{4.2.1}$$

$$= \tau[Ax(t,u,\tau) + F(x(t,u,\tau),u(t))].$$

Define $g(u,\tau) \triangleq h(x(1,u,\tau))$. Then $g(u,\tau) = h(\tilde{x}(\tau,\tilde{u})) = \tilde{g}(\tilde{u},\tau)$. Therefore $\tilde{\mathbf{P}}_F$ is equivalent to (4.2.2):

$$\mathbf{P}_F : \inf\{ g(u,\tau) \mid u \in G,\ \tau \in [\tau_{min},\tau_{max}] \}, \tag{4.2.2}$$

with the dynamics

$$\frac{d}{dt}x(t,u,\tau) = \tau[Ax(t,u,\tau) + F(x(t,u,\tau),u(t))], \quad x(0,u) = x_0 \in D(A), \ t \in [0,1]. \tag{4.2.3}$$

Thus, the form of $\mathbf{P}_F$ is similar to that of $\mathbf{P}$ (3.5.0). For any fixed value of τ, we can redefine A and $F(\cdot,\cdot)$ by multiplying each by τ, and apply the results of Chapter 3 to derive properties for the solutions $x(\cdot,u,\tau)$. We note that if the operator A generates the semigroup

$\{ T(t) \}_{t \geq 0}$, then τA generates the semigroup $\{ T(\tau t) \}_{t \geq 0}$

Theorem 4.2.1: Problems $\tilde{\mathbf{P}}_\mathbf{F}$ and $\mathbf{P}_\mathbf{F}$ are equivalent.

Proof: Let $(\bar{\tau}, \bar{u}) \in [\tau_{min}, \tau_{max}] \times G(\bar{\tau})$ be optimal for $\tilde{\mathbf{P}}_\mathbf{F}$. Define $(\tau, u) \in [\tau_{min}, \tau_{max}] \times G$ where $\tau = \bar{\tau}$ and $u(s) = \bar{u}(\bar{\tau} s)$, $s \in [0, 1]$. Then $g(u, \tau) = g(\bar{u}, \bar{\tau})$ and so $\mathbf{P}_\mathbf{V} \leq \tilde{\mathbf{P}}_\mathbf{V}$. Similarly, we can show $\tilde{\mathbf{P}}_\mathbf{V} \leq \mathbf{P}_\mathbf{V}$ and so $\tilde{\mathbf{P}}_\mathbf{V} = \mathbf{P}_\mathbf{V}$. ■

4.3 DIFFERENTIABILITY OF SOLUTION OF THE PDE WITH RESPECT TO THE CONTROL AND FINAL TIME

We make the following assumptions:

Assumption 4.3.1:

(i) For all $\tau \in [\tau_{min}, \tau_{max}]$, a mild solution to (4.2.3) exists.

(ii) There exists $b_1 < \infty$ such that for all $\tau \in [\tau_{min}, \tau_{max}]$, $u \in O$ and $t \in [0, 1]$, $\|x(t, u, \tau)\| \leq b_1$. ■

The next few theorems are natural extensions of theorems in Chapter 3. We define $\overline{M} \triangleq \tilde{M} e^{\omega \tau_{max}}$.

Theorem 4.3.2: (Lipschitz Continuity of $x(t, u, \tau)$ in u.)

There exists $b_2 < \infty$ such that for all $\hat{u}, u \in O$, $t \in [0, 1]$, $\tau \in [\tau_{min}, \tau_{max}]$,

$$\|x(t, \hat{u}, \tau) - x(t, u, \tau)\| \leq b_2 \|\hat{u} - u\|. \tag{4.3.1}$$

■

For $u, \delta u \in L_2 \cap L_\infty$ such that $u \in O$, we define $\delta x(\cdot, u, \tau, \delta u) \in C([0, 1], X)$ to be the mild solution to

$$\delta x(t) = \int_0^t \left\{ T(\tau(t-s))\tau(\frac{\partial F}{\partial x}(x(s,u,\tau),u(s))\delta x(s) + \frac{\partial F}{\partial u}(x(s),u(s))\delta u(s)) \right\} ds. \qquad (4.3.2a)$$

Equation (4.3.2a) is an integral from of the first variation with respect to u of equation (4.2.1). Then we have:

Theorem 4.3.3: (Differentiability with Respect to the Control.)

There exists $b_3 < \infty$ such that for all $\hat{u}, u \in O$, $\tau \in [\tau_{min}, \tau_{max}]$

$$\|x(t,\hat{u},\tau) - x(t,u,\tau) - \delta x(t,u,\tau,\delta u)\| \leq b_3 \|\hat{u} - u\|_2^2, \qquad (4.3.2b)$$

where $\delta u = \hat{u} - u$. ∎

We will denote the linear map $\delta u \to \delta x(t,u,\tau,\delta u)$ by $x_u(t,u,\tau)$.

Theorem 4.3.4: (Continuity of the Differential in the Control)

For all $\tau \in [\tau_{min}, \tau_{max}]$, $\hat{u}, u \in G$ such that $u \to \hat{u}$ in the $L_2^m([0,1])$ norm

$$\lim_{u \to \hat{u}} x_u(t,u,\tau) \to x_u(t,\hat{u},\tau). \qquad (4.3.3)$$

∎

Next, we will show that x_u is continuous in τ. We present the following three preliminary lemmas.

Lemma 4.3.5: Let $\tilde{U}$ be defined as in (3.2.2). Then, there exists $b_4 \geq \max\{ b_1, b_2, b_3, \tau_{max}, 1 \}$, $b_4 < \infty$, such that for all $u, \hat{u} \in \tilde{U}$, all $\|x\| \leq b_1$, all $\|\hat{x}\| \leq b_1$, all $\tau, \hat{\tau} \in [\tau_{min}, \tau_{max}]$ and all $t \in [0,1]$:

(i) $\|\frac{\partial F}{\partial x}(x,u)\| \leq b_4$,

(ii) $\|\frac{\partial F}{\partial u}(x,u)\| \leq b_4$,

(iii) $\|\frac{\partial F}{\partial x}(\hat{x},\hat{u}) - \frac{\partial F}{\partial x}(x,u)\| \le b_4[\|\hat{x}-x\| + \|\hat{u}-u\|]$,

(iv) $\|\frac{\partial F}{\partial u}(\hat{x},\hat{u}) - \frac{\partial F}{\partial x}(x,u)\| \le b_4[\|\hat{x}-x\| + \|\hat{u}-u\|]$,

(v) $\|T(\tau t)\| \le b_4$,

(vi) $\|T(\hat{\tau}t) - T(\tau t)\| \le b_4|\hat{\tau} - \tau|$. ■

Lemma 4.3.6: Given a bounded set $S \subset X$, if f_i, $i \in \{1, 2, \cdots n\}$ is a finite sequence of Lipschitz Continuous functions from S to Y where Y is a Banach space, then $f \triangleq \prod_{i=1}^{n} f_i$ is Lipschitz Continuous on S.

Proof: Let the Lipschitz constant for each f_i be L_i. Furthermore, f_i must be bounded on S, and we denote this bound b_i. Then for $\hat{x}, x \in S$:

$$\|f(\hat{x}) - f(x)\| = \|\prod_{j=1}^{n} f_j(\hat{x}) - \prod_{j=1}^{n} f_j(x)\|$$

$$= \|\sum_{j=1}^{n}\left[\prod_{i=1}^{j-1} f_i(\hat{x})(f_j(\hat{x}) - f_j(x)) \prod_{i=j+1}^{n} f_i(x)\right]\|$$

$$\le \sum_{j=1}^{n}\left[\prod_{i=1}^{j} b_i\right] L_j\|\hat{x} - x\|\left[\prod_{i=j+1}^{n} b_i\right]. \tag{4.3.4}$$

Therefore, f has a Lipschitz constant $\left[\prod_{j=1}^{n} b_j\right]\left[\sum_{j=1}^{n}\frac{L_j}{b_j}\right] < \infty$. ■

Theorem 4.3.7: (Lipschitz Continuity of $x(t, u, \tau)$ in τ.)

There exists a constant $b_5 \in [b_4, \infty)$ such that for all $t \in [0, 1]$, $u \in G$, $\hat{\tau}, \tau \in [\tau_{\min}, \tau_{\max}]$

$$\|x(t, u, \hat{\tau}) - x(t, u, \tau)\| \le b_5|\hat{\tau} - \tau|. \tag{4.3.5}$$

Proof:

$$\|x(t, u, \hat{\tau}) - x(t, u, \tau)\| = \|\int_0^t T(\hat{\tau}(t-s))F(x(s, u, \hat{\tau}), u(s))ds - \int_0^t T(\tau(t-s))F(x(s, u, \tau), u(s))ds\|$$

$$\leq \int_0^t [\|[T(\hat\tau(t-s)) - T(\tau(t-s))]F(x(s,u,\hat\tau),u(s))\| +$$

$$\|T(\tau(t-s))(F(x(s,u,\hat\tau),u(s)) - F(x(s,u,\tau),u(s)))\|]ds$$

$$\leq Cb_3|\hat\tau - \tau| + \overline{M}\int_0^t K_S\|x(s,u,\hat\tau) - x(s,u,\tau)\|ds, \tag{4.3.6}$$

for some $K_S < \infty$ by Assumptions 3.3.1 (i) and 4.3.1 (ii). By applying the Bellman-Gronwall Lemma,

$$\|x(t,u,\hat\tau) - x(t,u,\tau)\| \leq Cb_3|\hat\tau - \tau|e^{\overline{M}K_s}. \tag{4.3.7}$$

■

Choose $b_5 \triangleq \max\{ b_4, Cb_3e^{\overline{M}K_s} \}$.

Theorem 4.3.8: (Lipschitz Continuity of $x_u(t,u,\tau)$ in τ.)

There exists $b_6 < \infty$ such that for all $u \in G$, all $\hat\tau, \tau \in [\tau_{\min}, \tau_{\max}]$ and all $t \in [0,1]$, $\|x_u(t,u,\hat\tau) - x_u(t,u,\tau)\| \leq b_6|\hat\tau - \tau|$.

Proof:

$$\|\delta x(t,u,\hat\tau,\delta u) - \delta x(t,u,\tau,\delta u)\| \tag{4.3.10}$$

$$= \Big\|\int_0^t \{ T(\hat\tau(t-s))\hat\tau(\frac{\partial F}{\partial x}(x(s,u,\hat\tau),u(s))\delta x(s,u,\hat\tau,\delta u) + \frac{\partial F}{\partial u}(x(s,u,\hat\tau),u(s))\delta u(s))$$

$$- T(\tau(t-s))\tau(\frac{\partial F}{\partial x}(x(s,u,\tau),u(s))\delta x(s,u,\tau,u) + \frac{\partial F}{\partial u}(x(s,u,\tau),u(s))\delta u(s)) \}ds\Big\|$$

$$\leq \Big\|\int_0^t \{ T(\hat\tau(t-s))\hat\tau(\frac{\partial F}{\partial x}(x(s,u,\hat\tau),u(s))(\delta x(s,u,\hat\tau,\delta u) - \delta x(s,u,\tau,\delta u))$$

$$+ (T(\hat\tau(t-s))\hat\tau\frac{\partial F}{\partial x}(x(s,u,\hat\tau),u(s)) - T(\tau(t-s))\tau\frac{\partial F}{\partial x}(x(s,u,\tau),u(s)))\delta x(s,u,\tau,\delta u)$$

$$+ \left(T(\hat{\tau}(t-s))\hat{\tau}\frac{\partial F}{\partial u}(x(s,u,\hat{\tau}),u(s)) - T(\tau(t-s))\tau\frac{\partial F}{\partial u}(x(s,u,\tau),u(s))\right)\delta u(s)) \} ds \|.$$

By Lemma 4.3.6, there exists $\bar{b}_6 < \infty$ such that

$$\|\delta x(t,u,\hat{\tau},\delta u) - \delta x(t,u,\tau,\delta u)\|$$

$$\leq \int_0^t \overline{M}\tau_{max} b_4 \|\delta x(s,u,\hat{\tau},\delta u) - \delta x(s,u,\tau,\delta u)\| ds + b_6 |\hat{\tau} - \tau| \|\delta u\|_2. \tag{4.3.11}$$

By applying the Bellman-Gronwall Lemma,

$$\|\delta x(t,u,\hat{\tau},\delta u) - \delta x(t,u,\tau,\delta u)\| \leq \bar{b}_6 e^{\overline{M}\tau_{max} b_4} |\hat{\tau} - \tau| \|\delta u\|. \tag{4.3.12}$$

Since

$$\|x_u(t,u,\hat{\tau}) - x_u(t,u,\hat{\tau})\| = \max_{\|\delta u\|} \|\delta x(t,u,\hat{\tau},\delta u) - \delta x(t,u,\tau,\delta u)\|, \tag{4.3.13}$$

Theorem 4.3.8 is proved with $b_6 \triangleq \bar{b}_6 e^{\overline{M}\tau_{max} b_4}$. ■

We now show differentiability of the PDE with respect to the final time, τ. We define $\delta x(t,u,\tau,\delta\tau)$ to be the solution to:

$$\delta x(t,u,\tau,\delta\tau) = \int_0^t T(\tau(t-s))\tau\frac{\partial F}{\partial x}(x(s,u,\tau),u(s))\delta x(s,u,\tau,\delta\tau)ds \tag{4.3.14}$$

$$+ \int_0^t (T(\tau(t-s)) + \tau(t-s)AT(\tau(t-s)))F(x(s,u,\tau),u(s))\delta\tau ds.$$

Equation (4.3.14) is an integral from of the first variation with respect to τ of equation (4.2.1). Because A generates an analytic semigroup, the map $\delta\tau \rightarrow \delta x(t,u,\tau,\delta\tau)$ is well-defined, linear and bounded, i.e., there exists b_7 such that $\|\delta x(t,u,\tau,\delta\tau)\| \leq b_7|\delta\tau|$.

Theorem 4.3.9: Differentiability of $x(t,u,\tau)$ in τ.

For all $u \in G$, all $t \in [0,1]$, $\tau, \hat{\tau} \in [\tau_{min}, \tau_{max}]$,

$$\lim_{\hat{\tau} \to \tau} \frac{\|x(t,u,\hat{\tau}) - x(t,u,\tau) - \delta x(t,u,\tau,\hat{\tau}-\tau)\|}{|\hat{\tau}-\tau|} = 0. \tag{4.3.15}$$

Proof: To simplify notation, we make the following definitions: $\delta\tau \triangleq \hat{\tau} - \tau$ and $\Delta x(t,u,\tau,\delta\tau) = x(t,u,\hat{\tau}) - x(t,u,\tau)$.

$$\|\Delta x(t,u,\tau,\delta\tau) - \delta x(t,u,\tau,\delta\tau)\| \le \int_0^t \|T(\hat{\tau}(t-s))\hat{\tau}F(x(s,u,\hat{\tau}),u(s))$$

$$- T(\tau(t-s))\tau F(x(s,u,\tau),u(s)) - T(\tau(t-s))\tau\frac{\partial F}{\partial x}(x(s,u,\tau),u(s))\delta x(s,u,\tau,\delta\tau)$$

$$- (I + \tau(t-s)A)T(\tau(t-s))F(x(s,u,\tau),u(s))\delta\tau\|ds$$

$$\le \int_0^t \{ \|(T(\hat{\tau}(t-s))\hat{\tau} - T(\tau(t-s))\tau - (I+\tau(t-s)A)T(\tau(t-s))\delta\tau)\|\|F(x(s,u,\tau),u(s))\|$$

$$+ \|T(\hat{\tau}(t-s))\hat{\tau}\|[\int_0^1 \|\frac{\partial F}{\partial x}(x(s,u,\tau) + r\Delta x(s,u,\tau,\delta\tau),u(s)) - \frac{\partial F}{\partial x}(s(s,u,\tau),u(s))\|dr$$

$$\|\Delta x(s,u,\tau,\delta\tau)\| + \|\frac{\partial F}{\partial x}(x(s,u,\tau),u(s))(\Delta x(s,u,\tau,\delta\tau) - \delta x(s,u,\tau,\delta\tau)\|]$$

$$+ \|T(\hat{\tau}(t-s))\hat{\tau} - T(\tau(t-s))\tau\|\|\frac{\partial F}{\partial x}(x(s,u,\tau),u(s))\delta x(s,u,\tau,\delta\tau)\| \}ds$$

$$\le |\hat{\tau}-\tau|\int_0^t \Big\{ \|T(\hat{\tau}(t-s)) - T(\tau(t-s))$$

$$+ \tau(t-s)\int_0^1 A(T(\tau + r(\hat{\tau}-\tau))(t-s)) - T(\tau(t-s)))dr\| \Big\}\|F(x(s,u,\tau),u(s))\|ds$$

$$+ \int_0^t \{ Me^{\omega\tau_{max}}\tau_{max}\|\Delta x(s,u,\tau,\delta\tau)\|^2 + b_6\|\Delta x(s,u,\tau,\delta\tau) - \delta x(s,u,\tau,\delta\tau)\| \}ds$$

$$+ \tilde{C}b_6\tau_{max}|\hat{\tau}-\tau|^2. \tag{4.3.16a}$$

By Lemma 4.1.4, $T(\cdot)$ and $AT(\cdot)$ are continuous, and by Theorem 4.3.7, $|\Delta x(s,u,\tau,\delta\tau)|^2 \le b_5^2|\hat{\tau}-\tau|^2$. Therefore,

$$|\Delta x(t,u,\tau,\delta\tau) - \delta x(t,u,\tau,\delta\tau)\| \le$$

$$|\hat{\tau}-\tau|O(\delta\tau) + b_6\int_0^t |\Delta x(s,u,\tau,\delta\tau) - \delta x(s,u,\tau,\delta\tau)|ds\,, \tag{4.3.16b}$$

where $\lim_{\delta\tau\to 0} O(\delta\tau) = 0$. To finish the theorem, apply Lemma 4.3.6 and the Bellman-Gronwall Lemma. ■

The map $\delta\tau \to \delta x(t,u,\tau,\delta\tau)$ is linear in $\delta\tau$ for each $t \in [0,1]$, $u \in G$, $\tau \in [\tau_{min},\tau_{max}]$. If we call this map $x_\tau(t,u,\tau)$, then we have the following theorem:

Theorem 4.3.10: For all $t \in [0,1]$, $u \in G$, $x(t,u,\tau)$ is Gateaux differentiable with respect to τ and its differential is given by $x_\tau(t,u,\tau)$. ■

Next, we show that $x_\tau(t,u,\tau)$ is continuous in $u \in G$, $\tau \in [\tau_{min},\tau_{max}]$.

Theorem 4.3.11: (Continuity of $x_\tau(t,u,\tau)$ in u,τ.) For $u,\hat{u} \in G$ and $\tau,\hat{\tau} \in [\tau_{min},\tau_{max}]$,

$$\lim_{\hat{u}\to u\ \hat{\tau}\to\tau} x_\tau(t,\hat{u},\hat{\tau}) \to x_\tau(t,u,\tau). \tag{4.3.17}$$

Proof:

$$|x_\tau(t,\hat{u},\hat{\tau})\delta\tau - x_\tau(t,u,\tau)\delta\tau| = |\delta x(t,\hat{u},\hat{\tau},\delta\tau) - \delta x(t,u,\tau,\delta\tau)|$$

$$\le \int_0^t [\|[T(\hat{\tau}(t-s)) - T(\tau(t-s))]\hat{\tau}\frac{\partial F}{\partial x}(x(s,\hat{u},\hat{\tau}),\hat{u}(s))\delta x(s,\hat{u},\hat{\tau},\delta\tau)|$$

$$+ |T(\tau(t-s))(\hat{\tau}-\tau)\frac{\partial F}{\partial x}(s,\hat{u},\hat{\tau}),\hat{u}(s))\delta x(s,\hat{u},\hat{\tau},\delta\tau)|$$

$$+ |T(\tau(t-s))\hat{\tau}|K_S[\|x(s,\hat{u},\hat{\tau}) - x(s,u,\tau)| + |\hat{u}(s) - u(s)|]|\delta x(s,\hat{u},\hat{\tau},\delta\tau)|$$

$$+ \| T(\tau(t-s))\tau \frac{\partial F}{\partial x}(x(s,u,\tau),u(s))(\delta x(s,\hat{u},\hat{\tau},\delta\tau) - \delta x(s,u,\tau,\delta\tau))\|$$

$$+ \| T(\hat{\tau}(t-s)) - T(\tau(t-s)) + (\hat{\tau}-\tau)(t-s)AT(\hat{\tau}(t-s))$$

$$+ \ + \tau(t-s)A(T(\hat{\tau}(t-s)) - T(\tau(t-s)))\|\|F(x(s,\hat{u},\hat{\tau}),\hat{u}(s))\delta\tau\|$$

$$+ \| T(\tau(t-s)) + \tau(t-s)AT(\tau(t-s))\| K_S[\| x(s,\hat{u},\hat{\tau}) - x(s,u,\tau)\| + \|\hat{u}(s) - u(s)\|]\|\delta\tau|]ds$$

$$\leq b_8[|\hat{\tau}-\tau| + \|\hat{u}-u\|_2]|\delta\tau| + b_8\int_0^t \|\delta x(s,\hat{u},\hat{\tau},\delta\tau) - \delta x(s,u,\tau,\delta\tau)\|ds \qquad (4.3.18)$$

for some $b_8 < \infty$. Applying the Bellman-Gronwall Lemma,

$$\| x_\tau(t,\hat{u},\hat{\tau})\delta\tau - x_\tau(t,u,\tau)\delta\tau\| \leq e^{b_8}b_8[|\hat{\tau}-\tau| + \|\hat{u}-u\|_2]|\delta\tau|. \qquad (4.3.19)$$

Therefore,

$$\| x_\tau(t,\hat{u},\hat{\tau}) - x_\tau(t,u,\tau)\| \leq e^{b_8}b_8[|\hat{\tau}-\tau| + \|\hat{u}-u\|_2]. \qquad (4.3.20)$$

We have shown existence and continuity of the partial derivatives $x_u(t,u,\tau)$ and $x_\tau(t,u,\tau)$. Since the partial derivatives are continuous, the function $x(t,u,\tau)$ is fully continuously differentiable in u,τ by Dieudonne[Die.1].

■

4.4 EXISTENCE OF GRADIENTS

Consider the function $g : G\times[\tau_{min},\tau_{max}] \to \mathbb{R}$, $g(u,\tau) = h(x(1,u,\tau))$ where h is continuously differentiable as in Section 3.4. Extending the results of Section 3.4, we obtain the following lemma:

Lemma 4.4.1:

(i) $g : G\times[\tau_{\min}, \tau_{\max}] \to \mathbb{R}$ is continuous in (u, τ).

(ii) $g : G\times[\tau_{\min}, \tau_{\max}] \to \mathbb{R}$ is differentiable in (u, τ), i.e., for all $u \in G$, $\tau \in [\tau_{\min}, \tau_{\max}]$, there exists a differential $Dg(u, \tau) : L_2\cap L_\infty\times\mathbb{R} \to \mathbb{R}$ such that for $u, \hat{u} \in G$, $\hat{\tau}, \tau \in [\tau_{\min}, \tau_{\max}]$:

$$\lim_{s \to 0} \frac{|g(u + s(\hat{u} - u), \tau + s(\hat{\tau} - \tau)) - g(u, \tau) - sDg(u, \tau)(\hat{u} - u, \hat{\tau} - \tau)|}{\|u_i - u\|_2 + |\tau_i - \tau|} = 0. \tag{4.4.1}$$

(iii) There exists $\nabla g(u, \tau) \in L_2\cap L_\infty\times\mathbb{R}$ such that $\langle \nabla g(u, \tau), (\hat{u} - u, \hat{\tau} - \tau) \rangle = Dg(u, \tau)(\hat{u} - u, \hat{\tau} - \tau)$ for all $\hat{u}, u \in G$, $\hat{\tau}, \tau \in [\tau_{\min}, \tau_{\max}]$.

(iv) $\nabla g : G\times[\tau_{\min}, \tau_{\max}] \to L_2\cap L_\infty\times\mathbb{R}$ is continuous.

Consequently, we can show that the following algorithm, which is an extension of Algorithm 3.5.1 is convergent, and that all accumulation points $(\hat{u}, \hat{\tau})$ satisfy the necessary condition for optimality, $\Theta(\hat{u}, \hat{\tau}) = 0$ where $\Theta(\cdot, \cdot)$ is defined by (4.4.2a).

Algorithm 4.4.2:

Data: $u_0 \in G$, $\tau_0 \in [\tau_{\min}, \tau_{\max}]$, $\alpha \in (0, 1)$, $\beta \in (0, 1)$.

Step 0: $i = 0$.

Step 1: Compute $\nabla g(u_i, \tau_i)$.

Step 2: Compute $\Theta(u_i, \tau_i)$ and $(v_i, \hat{\tau}_i)$ using

$$\Theta(u_i, \tau_i) \triangleq \min_{(v, \hat{\tau}) \in G\times[\tau_{\min}, \tau_{\max}]} \{ \frac{1}{2}\|v - u_i\|^2 + \frac{1}{2}|\hat{\tau} - \tau_i|^2 + \langle \nabla g(u_i, \tau_i), \begin{bmatrix} v - u_i \\ \hat{\tau} - \tau_i \end{bmatrix} \rangle \}. \tag{4.4.2a}$$

$$(v_i, \hat{\tau}_i) \triangleq \arg\min_{(v, \hat{\tau}) \in G\times[\tau_{\min}, \tau_{\max}]} \{ \frac{1}{2}\|v - u_i\|^2 + \frac{1}{2}|\hat{\tau} - \tau_i|^2 \tag{4.4.2b}$$

$$+ \langle \nabla g(u_i, \tau_i), \begin{bmatrix} v - u_i \\ \hat{\tau} - \tau_i \end{bmatrix} \rangle \}.$$

If $\Theta(u_i, \tau_i) = 0$, STOP.

Step 3: Compute the step size λ_i using

$$\lambda_i \triangleq \max\{ \lambda \in \{ 0, 1, \beta, \beta^2, \cdots \} \mid \tag{4.4.3}$$

$$g(u_i + \lambda(v_i - u_i), \tau_i + \lambda(\hat{\tau}_i - \tau_i)) - g(u_i, \tau_i) < \alpha\lambda\Theta(u_i, \tau_i) \}.$$

Step 4: Set $u_{i+1} = u_i + \lambda_i(v_i - u_i)$, $\tau_{i+1} = \tau_i + \lambda_i(\hat{\tau}_i - \tau_i)$.

Step 5: Set $i = i + 1$; go to Step 1. ■

Theorem 4.4.2: If $\{ (u_i, \tau_i) \}$ is an infinite sequence generated by Algorithm 4.4.2, then any accumulation point $(\hat{u}, \hat{\tau}) \in G \times [\tau_{min}, \tau_{max}]$ satisfies the optimality condition $\Theta(\hat{u}, \hat{\tau}) = 0$.■

CHAPTER 5

RELAXED CONTROLS

Theorem 3.5.4 states that any accumulation point of Algorithm 3.5.1, $\hat{u} \in G$, satisfies the optimality condition $\Theta(\hat{u}) = 0$. However, since the closed unit ball in $L_2 \cap L_\infty$ is not compact, there may be bounded sequences generated by the algorithm which have no accumulation points. In these cases, Theorem 3.5.4 gives no indication as to the performance of the algorithm. How then do we evaluate algorithm performance? An answer is either to change the set of admissible controls, G, or to change its topology. For problems with linear dynamics, results have been obtained by replacing the $L_2 \cap L_\infty$ topology by the $weak^*$–L_2 topology (see [Ara.1].) Since the unit ball and G are compact in the $weak^*$-topology (Banach-Alaoglu Theorem), and $g(\cdot)$ and $\nabla g(\cdot)$ are $weak^*$-continuous, the problem $\min\{ g(u) \mid u \in G \}$ defined in (3.5.0) has a solution. Furthermore, any bounded sequence generated by the algorithm admits an accumulation point satisfying the optimality condition $\Theta(\hat{u}) = 0$.

For problems with nonlinear dynamics, $g(\cdot)$ may not be $weak^*$-continuous. Warga [War.1] describes a method for analyzing problems in which the dynamics are described by a system of non-linear ordinary differential equations. The principle is to densely embed the set of admissible controls, G, into a larger set $\overline{G}$. The function $g : G \to \mathbb{R}$ is extended to $\overline{g} : \overline{G} \to \mathbb{R}$, and a topology is chosen so that $\overline{G}$ is compact and $\overline{g}(\cdot)$ is continuous on $\overline{G}$. Some additional continuity properties are also required. The set $\overline{G}$ is called the set of admissible relaxed controls. Although in this chapter we consider only fixed-time problems, the extension to free-time problems is simple.

5.1 DEFINITIONS

We give a brief description of the properties of relaxed controls. See Warga [War.1] for a more complete exposition. We recall that U is a compact, convex subset of $\mathbb{R}^m$. We define a *finite Radon measure* μ on the Borel sets of U, to be a measure such that $|\mu|(U) < \infty$. The set of finite Radon measures on U is denoted by *frm*(U). We define a *Radon probability measure* μ on U to be a positive measure such that $\mu(U) = 1$. The set of Radon probability measures is denoted *rpm*(U). A relaxed control, σ, is a measurable function $\sigma : [0,1] \to rpm(U)$. We define

$$G \triangleq \{ u \in L_2^m([0,1]) \cap L_\infty^m([0,1]) \mid u(t) \in U \text{ for almost all } t \in [0,1] \}, \tag{5.1.1}$$

$$\overline{G} \triangleq \{ \sigma : [0,1] \to rpm(U) \mid \sigma \text{ is measurable } \}, \tag{5.1.2}$$

$$\mathbf{R} \triangleq \{ \sigma : [0,1] \to frm(U) \mid \sigma \text{ is measurable } \}. \tag{5.1.3}$$

By the Dunford-Petis Theorem which is an extension of the Riesz-Representation Theorem, the set $\mathbf{R}$ is isomorphic to $L^1([0,1],C(U))^*$ where $C(U)$ is the set of continuous, real-valued functions on U. That is, for each $\sigma \in \mathbf{R}$, there exists $\psi \in L^1([0,1],C(U))^*$ such that

$$\|\sigma\|_{\mathbf{R}} \triangleq \operatorname*{ess\,sup}_{t \in [0,1]} |\sigma(t)|(U) = \|\psi\| \triangleq \sup_{\phi \in L^1([0,1],C(U))} \frac{(\phi,\psi)}{\|\phi\|}, \tag{5.1.4}$$

and

$$(\phi,\psi) = \int_0^1 \int_U \psi(t,u)\sigma(t)(du)dt, \tag{5.1.5}$$

where (u,v) denotes u being operated on by v which is a member the dual of the space in which u lies. The unit ball in $\mathbf{R}$, $B(\mathbf{R})$, is *weak**-compact since $\{ \psi \in L^1([0,1],C(U))^* \mid \|\psi\| \le 1 \}$ is *weak**-compact. We use the *weak**-topology on $L^1([0,1],C(U))^*$ to topologize $\overline{G}$. Consequently, $\{ \sigma_i \} \subset \overline{G}$ converges to $\sigma \in \overline{G}$ if and only if

$$\lim_{\sigma_i \to \sigma} \int_0^1 \int_U \phi(t,u)\sigma_i(t)(du)dt = \int_0^1 \int_U \phi(t,u)\sigma(t)(du)dt \text{ for all } \phi \in L^1([0,1],C(U)). \tag{5.1.6}$$

Since $\overline{G}$ is a closed subset of $B(\mathbf{R})$, it is compact. There is an injection of the ordinary controls into the relaxed controls. With each ordinary control, $u \in G$, we associate a relaxed control $\sigma \in \overline{G}$ such that $\sigma(t)(S) = \delta_{u(t)}(S)$ for all measurable sets $S \subset U$. $\delta_x(S) \triangleq 1$ if $x \in S$ otherwise 0. We note that $\phi(t,u(t)) = \int_U \phi(t,u)\delta_{u(t)}(du)$.

We now apply this theory to problems where the dynamics are described by nonlinear partial differential equations in the form of

$$\frac{d}{dt}x(t,u) = Ax(t,u) + F(x(t,u),u(t)), \quad x(t,u) = x_0 \in D(A), \tag{5.1.7a}$$

where the underlying Hilbert Space is called H. We extend the map $x : G \to C([0,1],H)$ to $\overline{G}$ by defining for each $\sigma \in \overline{G}$, $x(\cdot,\sigma)) \in C([0,1],H)$ to be the mild solution to

$$\frac{d}{dt}x(t,\sigma) = Ax(t,\sigma) + \int_U F(x(t,\sigma),u)\sigma(t)(du), \quad x(0,\sigma) = x_0. \tag{5.1.7b}$$

Such a solution can be shown to exist and to be unique.

Lemma 5.1.0: If $\sigma \in \overline{G}$ is an ordinary control, i.e., there exists $u \in G$ such that $\sigma(t)(S) = \delta_{u(t)}(S)$ for all measurable sets $S \subset U$ and almost all $t \in [0,1]$, then $x(t,\sigma) = x(t,u)$ for all $t \in [0,1]$ where $x(\cdot,\sigma)$ is the solution to (5.1.7b) and $x(\cdot,u)$ is the solution to (5.1.7a).

Proof: Substituting $\delta_{u(t)}(S)$ in (5.1.7b),

$$\frac{d}{dt}x(t,\sigma) = Ax(t,\sigma) + \int_U F(x(t,\sigma),u)\delta_{u(t)}(du)$$

$$= A(x(t,\sigma) + F(x(t,\sigma),u(t)). \tag{5.1.8}$$

Therefore $x(t,\sigma) = x(t,u)$ for all $t \in [0,1]$. ■

We shall show that the map $\sigma \to x(t,\sigma)$ is continuous in σ, uniformly in $t \in [0,1]$. We begin with a preliminary result.

Lemma 5.1.1: For every $\phi \in C([0,1]\times U;l_2)$ and any $\varepsilon > 0$, there exists $j_\varepsilon < \infty$ such that

$$\sum_{i=j_\varepsilon}^{\infty} \phi_i(t,u)^2 \leq \varepsilon \text{ for all } t \in [0,1],\ u \in U. \tag{5.1.9}$$

Proof: Suppose that the lemma is false. Then for some $\phi \in C([0,1]\times U;l_2)$, there exists $\varepsilon > 0$ such that for all $j \in \mathbb{Z}_+$, there exists $t_j \in [0,1]$ and $u_j \in U$ such that

$$\sum_{i=j}^{\infty} \phi_i(t_j,u_j)^2 > \varepsilon. \tag{5.1.10}$$

Since $[0,1]\times U$ is compact, there exists $(\hat{t},\hat{u}) \in [0,1]\times U$ and $K \subset \mathbb{Z}_+$ such that $t_k \xrightarrow{K} \hat{t}$ and $u_k \xrightarrow{K} \hat{u}$. There must exist $\hat{j} < \infty$ such that

$$\sum_{i=\hat{j}}^{\infty} \phi_i(\hat{t},\hat{u})^2 \leq \frac{\varepsilon}{2}. \tag{5.1.11}$$

Since ϕ is continuous, there exists k_0 such that for all $k \geq k_0$, $k \in K$, $\|\phi(t_k,u_k) - \phi(\hat{t},\hat{u})\|_{l_2}^2 \leq \frac{\varepsilon}{4}$.

Therefore, $\sum_{i=\hat{j}}^{\infty} \phi_i(t_k,u_k)^2 \leq \frac{3\varepsilon}{4}$. Contradiction. ■

Theorem 5.1.2: Let $\{ \sigma_i \} \subset \overline{G}$ be a sequence of relaxed controls such that $\sigma_i \to \sigma \in \overline{G}$. Then, for any $\phi \in C([0,1],C(U,l_2))$

$$\lim_{i\to\infty} \int_0^1 \int_U \phi(t,u)(\sigma_i(t) - \sigma(t))(du)dt = 0.$$

Proof: By Lemma 5.1.1, for any $\varepsilon > 0$, there exists $\hat{j}$ such that $\sum_{j=\hat{j}}^{\infty} \phi_j(t,u)^2 \leq \frac{\varepsilon}{8}$ for all $u \in U$, for all $t \in [0,1]$. Therefore,

$$\|\int_0^1\int_U \phi(t,u)(\sigma_i(t)-\sigma(t))(du)dt\|_{l_2}^2 = \sum_{j=1}^{\infty}\left(\int_0^1\int_U \phi_j(t,u)(\sigma_i(t)-\sigma(t))(du)dt\right)^2. \tag{5.1.12}$$

Now,

$$\sum_{j=\hat{j}}^{\infty}\left[\int_0^1\int_U \phi_j(t,u)(\sigma_i(t)-\sigma(t))(du)dt\right]^2 \le \sum_{j=\hat{j}}^{\infty}\left[\int_0^1\int_U |\phi_j(t,u)|(\sigma_i(t)+\sigma(t))(du)dt\right]^2. \tag{5.1.13}$$

Applying Schwartz inequality $((\int fg\,d\mu)^2 \le (\int f^2 d\mu)(\int g^2 d\mu))$ to (5.1.13), we obtain

$$\sum_{j=\hat{j}}^{\infty}\left[\int_0^1\int_U \phi_j(t,u)(\sigma_i(t)-\sigma(t))(du)dt\right]^2 \le \sum_{j=\hat{j}}^{\infty}\int_0^1\left(\int_U |\phi_j(t,u)|(\sigma_i(t)+\sigma(t))(du)\right)^2 dt. \tag{5.1.14}$$

Applying Schwartz inequality to (5.1.14), we obtain

$$\sum_{j=\hat{j}}^{\infty}\left[\int_0^1\int_U \phi_j(t,u)(\sigma_i(t)-\sigma(t))(du)dt\right]^2 \le$$

$$\sum_{j=\hat{j}}^{\infty}\int_0^1\left[\int_U \phi_j(t,u)^2(\sigma_i(t)+\sigma(t))(du)\int_U(\sigma_i(t)+\sigma(t))(du)\right]dt. \tag{5.1.15}$$

Applying Lesbegue's Monotone Convergence Theorem to (5.1.15), we obtain

$$\sum_{j=\hat{j}}^{\infty}\left[\int_0^1\int_U \phi_j(t,u)(\sigma_i(t)-\sigma(t))(du)dt\right]^2 \le 2\int_0^1\left(\int_U\left(\sum_{j=\hat{j}}^{\infty}\phi_j(t,u)^2\right)(\sigma_i(t)+\sigma(t))(du)\right)dt$$

$$\le \frac{\varepsilon}{2}. \tag{5.1.17}$$

Since $\sigma_i \to \sigma$, there exists i_ε such that for all $i \ge i_\varepsilon$,

$$\sum_{j=1}^{\hat{j}-1}\left|\int_0^1\int_U \phi_j(t,u)(\sigma_i(t)-\sigma(t))(du)dt\right|^2 \le \frac{\varepsilon}{2}. \tag{5.1.18}$$

Therefore, $\|\int_0^1\int_U \phi(t,u)\,(\sigma_i(t)-\sigma(t))(du)dt\|_{l_2}^2 \le \varepsilon$ for all $i \ge i_\varepsilon$, and the theorem is proved. ■

Corollary 5.1.3: Let Y be any separable Hilbert Space. Let $\{ \sigma_i \} \subset \overline{G}$ be a sequence of relaxed controls such that $\sigma_i \to \sigma \in \overline{G}$. Then, for any $\phi \in C([0,1],C(U,Y))$

$$\lim_{i \to \infty} \int_0^1 \int_U \phi(t,u)(\sigma_i(t) - \sigma(t))(du)dt = 0. \tag{5.1.19}$$

Proof: Any separable Hilbert space is isomorphic to l_2. ■

Corollary 5.1.4: If $\phi : [0,1] \to C(U,Y)$ is continuous except at a finite number of points $s \in [0,1]$, then the results of Corollary 5.1.3 hold. ■

Theorem 5.1.5: (Continuity of $x(\cdot,\sigma)$ in σ.) If the sequence $\{ \sigma_i \} \subset \overline{G}$ is such that $\sigma_i \to \sigma \in \overline{G}$, then $x(\cdot,\sigma_i) \to x(\cdot,\sigma)$.

Proof: Since for $\sigma \in \overline{G}$, $x(\cdot,\sigma)$ is the mild solution to (5.1.7b),

$$x(t,\sigma) = \int_0^t \int_U T(t-s)F(x(s,\sigma),u)\sigma(s)(du)ds. \tag{5.1.20}$$

Therefore,

$$\begin{aligned}
\|x(t,\sigma_i) - x(t,\sigma)\| &= \|\int_0^t [\int_U T(t-s)F(x(s,\sigma_i),u)\sigma_i(s)(du) - \\
&\qquad \int_U T(t-s)F(x(s,\sigma),u))\sigma(s)(du)]ds\| \\
&\le \|\int_0^t \int_U T(t-s)F(x(s,\sigma_i),u) - F(x(s,\sigma),u))\sigma_i(s)(du)ds\| \\
&\quad + \|\int_0^t \int_U T(t-s)F(x(s,\sigma),u)(\sigma_i(s) - \sigma(s))(du)ds\| \\
&\le MK_S\int_0^t \|x(s,\sigma_i) - x(s,\sigma)\|ds + \|\int_0^1 \int_U \phi^t(s,u)(\sigma_i(s) - \sigma(s))(du)ds\|,
\end{aligned} \tag{5.1.21}$$

where M is defined in Assumption 3.2.2 (i), S is a bounded set which contains $\{ x(t,\sigma) \}_{\{ t \in [0,1], \sigma \in \overline{G} \}}$, K_S exists and is finite by Assumption 3.1.1 (ib), and

$$\phi^t(s,u) = \begin{cases} T(t-s)F(x(s,\sigma),u) & 0 \le s \le t \\ 0 & t \le s \le 1 \end{cases}$$

By the Bellman-Gronwall Lemma,

$$\|x(t,\sigma_i) - x(t,\sigma)\| \le \| \int_0^1 \int_U \phi^t(s,u)(\sigma_i(s) - \sigma(s))(du)ds \| e^{MK_s}. \tag{5.1.22}$$

Therefore, by Corollary 5.1.4, for $t \in [0,1]$, $\lim_{i \to \infty} \|x(t,\sigma_i) - x(t,\sigma)\| = 0$. Since $\|\phi^t\|$ is bounded, the convergence is uniform.

5.2 EXISTENCE OF DIRECTIONAL DERIVATIVES

It turns out that $x(t,\sigma)$ is not Gateaux differentiable with respect to $\sigma \in \overline{G}$. However for the case in which the dynamics of the system can be described by ordinary differential equations, L. Williamson and E. Polak [Wil.1] have developed a directional derivative of $x(t,\sigma)$ in the $\delta\sigma$ direction for a specific set of directions. These derivatives are sufficient to generate a necessary optimality condition sufficiently strong that it is interesting. Here we extend their work to the case in which the dynamics are described by the partial differential equation (3.1.2).

For $\lambda \in [-1,1]$, $y \in C([0,1]\times U;\mathbb{R}^m)$, and $\sigma \in \overline{G}$, define $x(\cdot,\sigma,\lambda,y)$ and $\delta x(\cdot,\sigma,\lambda,y)$ to be the mild solutions to

$$\frac{d}{dt}x(t,\sigma,\lambda,y) = Ax(t,\sigma,\lambda,y) + \int_U F(x(t,\sigma,\lambda,y),u + \lambda y(t,u))\sigma(t)(du), \tag{5.2.1}$$

$$x(0,\sigma,\lambda,y) = x_0,$$

$$\frac{d}{dt}\delta x(t,\sigma,\lambda,y) = \int_U \{ (A + \frac{\partial F}{\partial x}(x(t,\sigma),u))\delta x(t,\sigma,\lambda,y) + \tag{5.2.2}$$

$$\frac{\partial F}{\partial u}(x(t,\sigma),u)\cdot\lambda y(t,u) \}\sigma(t)(du), \quad \delta x(0,\sigma,\lambda,y) = 0.$$

Equation (5.2.2) is the first variation of (5.2.1). By Lemma A1.3, there exists an evolution system, $U(t,s)$, $0 \le s \le t \le 1$ such that

$$\delta x(t,\sigma,\lambda,y) = \lambda \int_0^t U(t,s) \int_U \frac{\partial F}{\partial u}(x(s,\sigma),u) y(s,u) \sigma(s)(du). \tag{5.2.3}$$

Lemma 5.2.1: For all $y \in C([0,1]\times U;\mathbb{R}^m)$ there exists $L < \infty$ such that for any $\sigma \in \overline{G}$, $t \in [0,1]$ and λ sufficiently small, $\| x(t,\sigma,\lambda,y) - x(\sigma) \| \le L|\lambda|$.

Proof:

$$\| x(t,\sigma,\lambda,y) - x(t,\sigma) \| = \| \int_0^t \int_U T(t-s)[F(x(s,\sigma,\lambda,y),u+\lambda y(s,u)) - F(x(s,\sigma),u)]\sigma(s)(du)ds \|$$

$$\le M \int_0^t \{ K_S[\| x(s,\sigma,\lambda,y) - x(s,\sigma) \| + |\lambda| \| y(\cdot,\cdot) \|_\infty \} ds. \tag{5.2.5}$$

Applying the Bellman-Gronwall Lemma,

$$\| x(t,\sigma,\lambda,y) - x(t,\sigma) \| \le L|\lambda|, \tag{5.2.6}$$

where $L \triangleq M e^{MK_S} \| y(\cdot,\cdot) \|_\infty$ and K_S is defined in the proof of Lemma 3.3.4. ■

Lemma 5.2.2: For all $y \in C([0,1]\times U;\mathbb{R}^m)$, there exists $d_1 < \infty$ such that for all $\lambda \in [-1,1]$, $\sigma \in \overline{G}$,

$$\| x(t,\sigma,\lambda,y) - x(t,\sigma) - \delta x(t,\sigma,\lambda,y) \| \le d_1 |\lambda|^2, \quad t \in [0,1]. \tag{5.2.7}$$

Proof: We define $\Delta x(t,\sigma,\lambda,y) \triangleq x(t,\sigma,\lambda,y) - x(t,\sigma)$. Then,

$$\| \Delta x(t,\sigma,\lambda,y) - \delta x(t,\sigma,\lambda,y) \| \le \| \int_0^t \int_U T(t-s) \Big[F(x(s,\sigma,\lambda,y),u$$

$$+ \lambda y(s,u)) - F(x(s,\sigma),u)$$

$$- \frac{\partial F}{\partial x}(x(s,\sigma),u)\delta x(s,\sigma,\lambda,y) - \frac{\partial F}{\partial u}(x(s,\sigma),u)\cdot \lambda y(s,u) \Big] \sigma(s)(du)ds$$

$$\le M\int_0^t\int_U [\int_0^1 |\frac{\partial F}{\partial x}(x(s,\sigma,\lambda,y) + r\Delta x(s,\sigma,\lambda,y), u + r\lambda y(s,u)) - \frac{\partial F}{\partial x}(x(s,\sigma),u)|\,dr\,|\Delta x(s,\sigma,\lambda,y)||$$

$$+ \int_0^1 |\frac{\partial F}{\partial u}(x(s,\sigma,\lambda,y) + r\Delta x(s,\sigma,\lambda,y), u + r\lambda y(s,u)) - \frac{\partial F}{\partial u}(x(s,\sigma),u)|\,dr|\lambda||y(\cdot,\cdot)|_\infty$$

$$+ |\frac{\partial F}{\partial x}(x(s,\sigma),u)||\Delta x(s,\sigma,\lambda,y) - \delta x(s,\sigma,\lambda,y)|]\sigma(s)(du)ds$$

$$\le M\int_0^t\int_U [K_S(|\Delta x(s,\sigma,\lambda,y)| + |\lambda||y(\cdot,\cdot)|_\infty)|\Delta x(s,\sigma,\lambda,y)|$$

$$+ K_S(||\Delta x(s,\sigma,\lambda,y)| + |\lambda||y(\cdot,\cdot)|_\infty)|\lambda|y(\cdot,\cdot)|_\infty$$

$$+ b_7|\lambda||y(\cdot,\cdot)|_\infty|\Delta x(s,\sigma,\lambda,y) - \delta x(s,\sigma,\lambda,y)|]\sigma(s)(du)ds, \tag{5.2.10}$$

where b_7 is defined in Lemma (3.3.3). Since by Lemma 5.2.1 $|\Delta x(s,\sigma,\lambda,y)| \le L|\lambda|$, it follows from the Bellman-Gronwall Lemma that

$$|\Delta x(t,\sigma,\lambda,y) - \delta x(t,\sigma,\lambda,y)| \le MK_S e^{Mb_7}(|\Delta x(s,\sigma,\lambda,y)| + |\lambda||y(\cdot,\cdot)|_\infty)^2$$

$$\le d_1|\lambda|^2. \tag{5.2.11}$$

■

Consequently, $\lim_{\lambda\to 0}\frac{1}{\lambda}\delta x(t,\sigma,\lambda,y)$ is the directional derivative in the direction y. The natural question is: How can a member of $C([0,1]\times U;\mathbb{R}^m)$ be a direction for the set of relaxed controls? We define $\rho \triangleq \sigma\oplus y$ where $\sigma \in \overline{G}$ and $y \in C([0,1]\times U;\mathbb{R}^m)$,

$$\rho(t)(S) \triangleq \{\ \sigma(t)(R),\ R \triangleq \{\ u \in U \mid u + y(t,u) \in S\ \}\ \} \tag{5.2.12}$$

if $u + y(t,u) \in U$ for all $u \in U$ and almost all $t \in [0,1]$, otherwise ρ is undefined. It is easily seen that if ρ is well defined then $\rho \in \overline{G}$ and $x(1,\rho) = x(1,\sigma,1,y)$. It follows that $g(\sigma\oplus\lambda y) = h(x(1,\sigma,\lambda,y))$.

We now extend the function $g : G \to \mathbb{R}$ to $\bar{g} : \bar{G} \to \mathbb{R}$ by defining $\bar{g}(\sigma) \triangleq h(x(1,\sigma))$ for $\sigma \in \bar{G}$. Since the maps $\sigma \to x(1,\sigma)$ and $x(1,\sigma) \to h(x(1,\sigma))$ are continuous, it follows that $\bar{g}(\cdot)$ is continuous.

We shall call the "gradient" of $\bar{g}$ at σ the map $\nabla\bar{g}(\sigma):[0,1]\times U \to \mathbb{R}^m$ defined by

$$\nabla\bar{g}(\sigma)(t,u) \triangleq \frac{\partial F}{\partial u}(x(t,\sigma),u)^* p(t,\sigma); \tag{5.2.13}$$

where $p(t,\sigma)$ is the weak solution to

$$\frac{d}{dt}p(t,\sigma) = -\left[A + \int_U \frac{\partial F}{\partial x}(x(t,\sigma),u)\sigma(t)(du)\right]^* p(t,\sigma), \quad p(1,\sigma) = \nabla h(x(1,\sigma)). \tag{5.2.14}$$

The sense in which $\bar{g}$ is a gradient is made clear in Theorem 5.2.4. A mild solution to (5.2.14) exists and is unique by Theorem A1.4. Furthermore $p(t,\sigma) = U(1,t)^* p(1,\sigma)$ where $U(t,s)$, $0 \le s \le t \le 1$ is the evolution operator generated by $A + \int_U \frac{\partial F}{\partial x}(x(t,\sigma),u)\sigma(t)(du)$.

Theorem 5.2.3:

(i) For all $\sigma \in \bar{G}$, $\nabla\bar{g}(\sigma)(\cdot,\cdot)$ is uniformly Lipschitz continuous on $[0,1]\times U$, i.e., there exists $L < \infty$ such that for all $\sigma \in \bar{G}$ all $u,\hat{u} \in U$ and all $\hat{t},t \in [0,1]$,

$$|\nabla\bar{g}(\sigma)(\hat{t},\hat{u}) - \nabla\bar{g}(\sigma)(t,u)| \le L[|\hat{t} - t| + |\hat{u} - u|]. \tag{5.2.15}$$

(ii) The function $\nabla g(\cdot)$ is continuous, i.e., For every sequence $\{\sigma_i\} \subset \bar{G}$ such that $\sigma_i \to \sigma \in \bar{G}$, $\|\nabla\bar{g}(\sigma_i) - \nabla\bar{g}(\sigma)\|_\infty \to 0$.

Proof:

(i)

$$|\nabla\bar{g}(\sigma)(\hat{t},\hat{u}) - \nabla\bar{g}(\sigma)(t,u)| = \left|\frac{\partial F}{\partial u}(x(\hat{t},\sigma),\hat{u})p(\hat{t},\sigma) - \frac{\partial F}{\partial u}(x(t,\sigma),u)p(t,\sigma)\right|. \tag{5.2.16}$$

Since (a) $x(\cdot,\sigma)$ and $p(\cdot,\sigma)$ are mild solutions to partial differential equations, they are uniformly Lipschitz continuous on $[0,1]$ for all $\sigma \in \overline{G}$, (b) the set $\{ x(t,\sigma) \}_{\{ t \in [0,1], \sigma \in \overline{G} \}}$ is a subset of the bounded set S define in Lemma 3.3.2, and (c) $\frac{\partial F}{\partial u}(\cdot,\cdot)$ is Lipschitz continuous in x and u on with constant K_S, (i) follows.

(ii) Suppose that $\{ \sigma_i \} \subset \overline{G}$ such that $\sigma_i \to \sigma \in \overline{G}$. Then,

$$|\nabla \overline{g}(\sigma_i)(t,u) - \nabla \overline{g}(\sigma)(t,u)| = |\frac{\partial F}{\partial u}(x(t,\sigma_i),u)^* p(t,\sigma_i) - \frac{\partial F}{\partial u}(x(t,\sigma),u)^* p(t,\sigma)|. \tag{5.2.17}$$

Since $\frac{\partial F}{\partial u}(\cdot,\cdot) \in B(L_2;X)$ is Lipschitz continuous with constant K_S, its adjoint $\frac{\partial F}{\partial u}(\cdot,\cdot)^* \in B(X;L_2)$ is also Lipschitz continuous with constant K sub S. Therefore,

$$|\nabla \overline{g}(\sigma_i)(t,u) - \nabla \overline{g}(\sigma)(t,u)| \leq K_S |x(t,\sigma_i) - x(t,\sigma)||p(t,\sigma_i)|$$
$$+ |\frac{\partial F}{\partial u}(x(t,\sigma),u)^*||p(t,\sigma_i) - p(t,\sigma)|. \tag{5.2.18}$$

By Theorem 5.1.5 $x(\cdot,\sigma_i) \to x(\cdot,\sigma)$ in the L_∞ norm. Similarly, it can be shown $p(\cdot,\sigma_i) \to p(\cdot,\sigma)$ in the L_∞ norm. Since $\frac{\partial F}{\partial u}(\cdot,\cdot)^*$ is Lispschitz continuous, the set $\{ \frac{\partial F}{\partial u}(t,u) \}_{\{ t \in [0,1], u \in U \}}$ is bounded and $\nabla \overline{g}(\sigma_i)(\cdot,\cdot) \to \nabla \overline{g}(\sigma)(\cdot,\cdot)$ in the uniform topology on $[0,1] \times U$. ■

Finally, for each $\sigma \in \overline{G}$, we define an inner product $\langle \cdot,\cdot \rangle_\sigma : C([0,1]\times U;\mathbb{R}^m)\times C([0,1]\times U;\mathbb{R}^m) \to \mathbb{R}$ and a norm $|\cdot|_\sigma : C([0,1]\times U;\mathbb{R}^m) \to \mathbb{R}$ as

$$\langle y,z \rangle_\sigma \triangleq \int_0^1 \int_U \langle y(t,u),z(t,u) \rangle \sigma(t)(du)dt, \tag{5.2.22}$$

$$|y|_\sigma \triangleq \langle y,y \rangle_\sigma^{1/2}. \tag{5.2.23}$$

Theorem 5.2.4: For $y \in C([0,1]\times U;\mathbb{R}^m)$,

$$\lim_{\lambda \to 0} \frac{|h(x(1,\sigma,\lambda,y)) - h(x(1,\sigma)) - \lambda \langle \nabla \bar{g}(\sigma), y \rangle_\sigma|}{|\lambda|} = 0. \tag{5.2.24}$$

Proof:

$$|h(x(1,\sigma,\lambda,y)) - h(x(1,\sigma)) - \lambda \langle \nabla \bar{g}(\sigma), y \rangle_\sigma| \tag{5.2.25}$$

$$= | \langle \nabla h(x(1,\sigma)), \Delta x(1,\sigma,\lambda,y) - \delta x(1,\sigma,\lambda,y) \rangle + \langle \nabla h(x(1,\sigma), \delta x(1,\sigma,\lambda,\dot{y}) \rangle$$

$$+ \int_0^1 \langle \nabla h(x(1,\sigma) + r\Delta x(1,\sigma,\lambda,y)) - \nabla h(x(1,\sigma)), \Delta x(1,\sigma,\lambda,y) \rangle dr$$

$$- \lambda \int_0^1 \int_U \frac{\partial F}{\partial u}(x(t,\sigma),u)^* U(1,t)^* \nabla h(x(1,\sigma)) y(t,u) \sigma(t)(du) dt|$$

$$\leq K|\lambda|^2 + \int_0^1 \|\nabla h(x(1,\sigma) + r\Delta x(1,\sigma,\lambda,y)) - \nabla h(x(1,\sigma))\| dr M|\lambda|. \tag{5.2.26}$$

Since $\Delta x(1,\sigma,\lambda,y) < L|\lambda|$ and $\nabla h(\cdot)$ is continuous, the theorem is proven. ■

The map $y \to \langle \nabla \bar{g}(\sigma), y \rangle_\sigma$ is the directional derivative of $\bar{g}(\sigma)$ in the y direction.

5.3 OPTIMIZATION ALGORITHMS II: ALGORITHM WITH RELAXED CONTROLS

We shall present an algorithm for solving the problem

$$\bar{\mathbf{P}} : \min\{ \bar{g}(\sigma) \mid \sigma \in \bar{G} \}, \tag{5.3.1}$$

which is identical to Algorithm 3.5.1.

Algorithm 5.3.1: (This algorithm is identical to Algorithm 3.5.1)

Data: $u_0 \in G$, $\alpha \in (0,1)$, $\beta \in (0,1)$.

Step 0: $i = 0$.

Step 1: Compute $\nabla g(u_i)$ using (3.4.6) and (3.4.7).

Step 2: Compute $\Theta(u_i)$ and v_i using

$$\Theta(u_i) \triangleq \min_{v \in G} \{ \frac{1}{2}|v - u_i|^2 + \langle \nabla g(u_i), v - u_i \rangle \}. \tag{5.3.2a}$$

$$v_i = v(u_i) \triangleq \arg\min_{v \in G} \{ \frac{1}{2}|v - u_i|^2 + \langle \nabla g(u_i), v - u_i \rangle \}. \tag{5.3.2b}$$

If $\Theta(u_i) = 0$, STOP.

Step 3: Compute the step size λ_i using

$$\lambda_i \triangleq \max\{ \lambda \in \{ 0, 1, \beta, \beta^2, \cdots \} \mid h(x(1, u_i + \lambda(v_i - u_i))) - h(x(1, u_i)) < \alpha\lambda\Theta(u_i) \}. \tag{5.3.3}$$

Step 4: Set $u_{i+1} = u_i + \lambda_i(v_i - u_i)$.

Step 5: Set $i = i + 1$; go to Step 1. ■

Algorithm 5.3.1 produces a sequence of ordinary controls $\{ u_i \} \subset G$. Since there is an injection of the ordinary into the relaxed controls we can associate a relaxed control $\sigma_i \in \overline{G}$ with each $u_i \in G$. The sequence $\{ \sigma_i \} \subset \overline{G}$ must have accumulation points in $\overline{G}$ by the compactness property of $\overline{G}$. We shall show that any accumulation point of $\{ \sigma_i \}$ satisfies a first order necessary condition for optimality. We begin by stating the optimality condition.

Optimality Function 5.3.2: For $\sigma \in \overline{G}$, we define the optimality function

$$\overline{\Theta}(\sigma) \triangleq \frac{1}{2}|y(\sigma)|_\sigma^2 + \langle \nabla\overline{g}(\sigma), y(\sigma) \rangle_\sigma, \tag{5.3.4}$$

where

$$y(\sigma)(t,u) \triangleq \arg\min_{w \in U - u(t)} \{ \frac{1}{2}|w|^2 + \langle \nabla\overline{g}(\sigma)(t,u), w \rangle \}. \tag{5.3.5}$$

Lemma 5.3.3: If $\sigma \in \overline{G}$ is an ordinary control, i.e., there exists $u \in G$ such that $\sigma(t)(S) = \delta_{u(t)}(S)$ for all measurable sets $S \subset U$ and almost all $t \in [0, 1]$, then

(i) For $t \in [0,1]$, $\nabla \bar{g}(\sigma)(t,u(t)) = \nabla g(u)(t)$ where $\nabla g(u)(\cdot)$ is defined in (3.4.6), (3.4.7).

(ii) For $t \in [0,1]$, $v(u)(t) = u(t) + y(\sigma)(t,u(t))$ where $v(u)(\cdot)$ is defined in (3.5.1b).

(iii) $\Theta(u) = \bar{\Theta}(\sigma)$ where $\Theta(\cdot)$ is defined in (3.5.1a).

Proof:

(i) Since $x(t,\sigma) = x(t,u)$ by Lemma 5.1.0,

$$\nabla \bar{g}(\sigma)(t,u(t)) = \frac{\partial F}{\partial u}(x(t,u),u(t))p(t,\sigma), \tag{5.3.6}$$

where

$$\frac{d}{dt}p(t,\sigma) = -[A + \frac{\partial}{\partial x}F(x(t,u),u(t))]^* p(t,\sigma),\ p(1,\sigma) = \nabla h(x(1,u)),$$

and therefore $\nabla \bar{g}(\sigma)(t,u(t)) = \nabla g(u)(t)$.

(ii) By (5.3.5),

$$\begin{aligned} y(\sigma)(t,u(t)) &= \underset{w \in U - u(t)}{\arg\min} \{ \frac{1}{2}\|w\|^2 + \langle \nabla \bar{g}(\sigma)(t,u(t)), w \rangle \\ &= \underset{v \in U}{\arg\min} \{ \frac{1}{2}\|v - u(t)\|^2 + \langle \nabla g(u)(t), v - u(t) \rangle - u(t). \end{aligned} \tag{5.3.7}$$

Therefore, $u(t) + y(\sigma)(t,u(t)) = v(u)(t)$.

(iii)

$$\begin{aligned} \bar{\Theta}(\sigma) &= \frac{1}{2}\|y(\sigma)\|_\sigma^2 + \langle \nabla \bar{g}(\sigma), y(\sigma) \rangle_\sigma \\ &= \frac{1}{2}\int_0^1 \int_U \|y(\sigma)(t,u)\|^2 \delta_{u(t)}(du)dt + \int_0^1 \int_U \langle \nabla \bar{g}(\sigma)(t,u), y(\sigma)(t,u(t)) \rangle \delta_{u(t)}(du)(dt) \\ &= \frac{1}{2}\int_0^1 \|y(\sigma)(t,u(t))\|^2 dt + \int_0^1 \langle \nabla \bar{g}(\sigma)(t,u(t)), y(\sigma)(t,u(t)) \rangle dt \\ &= \frac{1}{2}\int_0^1 \|v(t)(t) - u(t)\|^2 dt + \int_0^1 \langle \nabla \bar{g}(u)(t), v(u)(t) - u(t) \rangle dt \end{aligned}$$

$$= \frac{1}{2}|v(u) - u|^2 + \langle \nabla g(u), v(u) - u \rangle = \Theta(u). \tag{5.3.8}$$

■

Lemma 5.3.4: Suppose that $\sigma \in \overline{G}$ is a local minimum for $\overline{\mathbf{P}}$. Then $\overline{\Theta}(\sigma) = 0$.

Proof: Assume that $\sigma \in \overline{G}$ and $\overline{\Theta}(\sigma) = -\delta < 0$. Then $\sigma \oplus \lambda y(\sigma) \in \overline{G}$ for $\lambda \in [0,1]$, and there exists $\hat{\lambda} \in (0,1]$ such that for all $\lambda \in [0,\hat{\lambda}]$, $\overline{g}(\sigma \oplus \lambda y(\sigma)) - \overline{g}(\sigma) - \lambda \langle \nabla \overline{g}(\sigma), y(\sigma) \rangle_\sigma \leq \frac{\lambda\delta}{2}$ by Theorem 5.2.4. Since $\langle \nabla \overline{g}(\sigma), y(\sigma) \rangle_\sigma < -\delta$, $\overline{g}(\sigma \oplus \lambda y(\sigma)) - \overline{g}(\sigma) \leq -\frac{\lambda\delta}{2}$ for all $\lambda \in [0,\hat{\lambda}]$ and hence σ is not a local minimizer. ■

Lemma 5.3.5: Let L be as defined in Theorem 5.2.3 (i). Then for all $\sigma \in \overline{G}$, $\hat{u}, u \in U$, and $\hat{t}, t \in [0,1]$,

$$|y(\sigma)(\hat{t},\hat{u}) - y(\sigma)(t,u)| \leq (L+2)[|\hat{t} - t| + |\hat{u} - u|] \tag{5.3.9}$$

Proof: Since

$$y(\sigma)(t,u) = \arg\min_{w \in U} \{ |w - (u - \nabla g(\sigma)(t,u))|^2 \} - u, \tag{5.3.10}$$

we can apply Lemma 3.5.3 and Theorem 5.2.3 (i) to obtain

$$|y(\sigma)(\hat{t},\hat{u}) - y(\sigma)(t,u)| \leq |\nabla g(\sigma)(\hat{t},\hat{u}) - \nabla g(\sigma)(t,u)| + 2|\hat{u} - u|$$

$$\leq (L+2)[|\hat{t} - t| + |\hat{u} - u|] \tag{5.3.11}$$

■

Lemma 5.3.6: The function $y(\cdot)$ is continuous, i.e., for a sequence $\{ \sigma_i \} \subset \overline{G}$ such that $\sigma_i \to \sigma \in \overline{G}$, $|y(\sigma_i) - y(\sigma)|_\infty \to 0$.

Proof: (By contradiction) Since $\{ y(\sigma_i) \}$ is equi-Lipschitz continuous (Lemma 5.3.5), there exists a $\hat{y} \in C([0,1] \times U; \mathbb{R}^m)$ and a subsequence $K \subset \mathbb{Z}_+$ such that $y(\sigma_i) \to \hat{y}$ on K. Suppose $\hat{y} \neq y(\sigma)$. Then,

$$\frac{1}{2}|y(\sigma)|_1^2 + \langle \nabla\bar{g}(\sigma), y(\sigma) \rangle_1 < \frac{1}{2}|\hat{y}|_1^2 + \langle \nabla\bar{g}(\sigma), \hat{y} \rangle_1, \tag{5.3.12}$$

where $\langle y,z \rangle_1 \triangleq \int_0^1 \int_U \langle y(t,u), z(t,u) \rangle du dt$ and $|y|_1 \triangleq \langle y,y \rangle_1^{1/2}$.

By the continuity of $\nabla\bar{g}(\cdot)$, there exists i_0 such that for all $i \geq i_0$, $i \in K$,

$$\frac{1}{2}|y(\sigma)|_1^2 + \langle \nabla\bar{g}(\sigma_i), y(\sigma) \rangle_1 < \frac{1}{2}|y(\sigma_i)|_1^2 + \langle \nabla\bar{g}(\sigma_i), y(\sigma_i) \rangle_1. \tag{5.3.13}$$

Hence, we obtain a contradiction. ■

Corollary 5.3.7: The function $\overline{\Theta}(\cdot)$ is continuous.

Proof: Suppose $\{ \sigma_i \} \subset \bar{G}$ is such that $\sigma_i \to \sigma \in \bar{G}$. Then

$$\begin{aligned}
\overline{\Theta}(\sigma_i) - \overline{\Theta}(\sigma) &= \frac{1}{2}|y(\sigma_i)|_{\sigma_i}^2 - \frac{1}{2}|y(\sigma)|_{\sigma}^2 + \langle \nabla g(\sigma_i), y(\sigma_i) \rangle_{\sigma_i} - \langle \nabla g(\sigma), y(\sigma) \rangle_{\sigma} \\
&= [\frac{1}{2}|y(\sigma_i)|_{\sigma_i}^2 - \frac{1}{2}|y(\sigma)|_{\sigma_i}^2] + [\frac{1}{2}|y(\sigma)|_{\sigma_i}^2 - \frac{1}{2}|y(\sigma)|_{\sigma}^2] \\
&\quad + [\langle \nabla g(\sigma_i), y(\sigma_i) \rangle_{\sigma_i} - \langle \nabla g(\sigma_i), y(\sigma) \rangle_{\sigma_i}] + [\langle \nabla g(\sigma_i), y(\sigma) \rangle_{\sigma_i} - \langle \nabla g(\sigma), y(\sigma) \rangle_{\sigma_i}] \\
&\quad + [\langle \nabla g(\sigma), y(\sigma) \rangle_{\sigma_i}] - \langle \nabla g(\sigma), y(\sigma) \rangle_{\sigma}].
\end{aligned} \tag{5.3.14}$$

Since $|y(\sigma_i) - y(\sigma)|_\infty \to 0$, $|\nabla g(\sigma_i) - \nabla g(\sigma)|_\infty \to 0$, $|y(\sigma)(\cdot,\cdot)|^2 \in C([0,1]\times U)$ and $\langle \nabla g(\sigma)(\cdot,\cdot), y(\sigma)(\cdot,\cdot) \rangle \in C([0,1]\times U)$, $|\Theta(\sigma_i) - \Theta(\sigma)| \to 0$. ■

Theorem 5.3.8: Suppose that $\{ u_i \} \subset G$ is a sequence generated by Algorithm 5.3.1. If the sequence is finite, then the last control, u_i, satisfies $\Theta(u_i) = 0$. Otherwise, there exists at least one accumulation point of the sequence $\{ \sigma_i \}$ in $\bar{G}$, and for any accumulation point $\hat{\sigma} \in \bar{G}$, $\overline{\Theta}(\hat{\sigma}) = 0$. Furthermore, $\lim_{i \to \infty} |g(u_{i+1}) - g(u_i)| = 0$.

Proof: If $\{ \sigma_i \}$ is an infinite sequence, then since $\bar{G}$ is compact, an accumulation point $\hat{\sigma}$ and a subsequence K such that $\sigma_i \to \hat{\sigma}$ on K exist. Assume $\overline{\Theta}(\hat{\sigma}) = -\delta < 0$. By Lemma

5.2.4, using the argument in Lemma 3.5.2, there exists $\hat{\lambda} \in (0,1)$ such that

$$\bar{g}(\hat{\sigma} \oplus \lambda y(\hat{\sigma})) - \bar{g}(\hat{\sigma}) < \alpha\lambda\bar{\Theta}(\hat{\sigma}), \quad \text{for all } \lambda \in [0,\hat{\lambda}]. \tag{5.3.15}$$

By continuity of $\bar{g}(\cdot)$, $y(\cdot)$, and $\bar{\Theta}(\cdot)$, there exists i_0 such that for all $i \geq i_0$, $i \in K$

$$\bar{g}(\sigma_i \oplus \lambda y(\sigma_i)) - \bar{g}(\sigma_i) < \alpha\lambda\bar{\Theta}(\sigma_i) < \frac{\alpha\lambda\bar{\Theta}(\hat{\sigma})}{2}. \tag{5.3.16}$$

Therefore, the a stepsize of at least $\lambda\beta$ is chosen by applying the armijo stepsize rule (Step 3), and so

$$\bar{g}(\sigma_{i+1}) - \bar{g}(\sigma_i) < \frac{\alpha\beta\hat{\lambda}\bar{\Theta}(\hat{v})}{2} = -\frac{\alpha\beta\hat{\lambda}\delta}{2} < 0. \tag{5.3.17}$$

Therefore $\bar{g}(\sigma_i) \to -\infty$ contradicting the continuity of $\bar{g}(\cdot)$. Finally, since $\bar{g}(\cdot)$ is continuous and $\sigma_i \to \hat{\sigma}$ on K, $g(\sigma_i) \to g(\hat{\sigma})$, and since $g(\sigma_i)$ is a non-increasing function of i, $\lim_{i \to \infty} |g(\sigma_{i+1}) - g(\sigma_i)| = 0$. ■

CHAPTER 6

DISCRETIZATION AND IMPLEMENTATION

6.1 PROBLEM STATEMENT

Section 3.5 presents Algorithm 3.5.1 to solve (3.1.3):

$$\mathbf{P} : \inf\{ g(u) \mid u \in G \}. \tag{6.1.1}$$

Theorem 3.5.4 states that any $L_2 \cap L_\infty$ accumulation point of a sequence generated by Algorithm 3.5.1 satisfies a necessary condition of optimality. We note in section 5.1 that such a sequence may not have an $L_2 \cap L_\infty$ accumulation point. However, by densely embedding G in the set $\overline{G}$ and using the topology of relaxed controls, we guarantee existence of an accumulation point and show (Theorem 3.5.6) that any accumulation point satisfies a necessary condition of optimality for the relaxed problem:

$$\overline{\mathbf{P}} : \min\{ g(u) \mid u \in \overline{G} \}. \tag{6.1.2}$$

However, none of the quantities used in Algorithm 3.5.1 or Algorithm 5.3.1 can be calculated exactly, since they require the exact solution of a partial differential equation. Consequently, Algorithm 3.5.1 is only a *conceptual algorithm*; it cannot be implemented directly on a computer.

In this chapter, we develop two *implementable algorithms* by performing a series of discretizations. We solve problem $\overline{\mathbf{P}}$ by introducing a series of discretized problems $\{ P_n \}$. We perform iterations to solve each P_n until a specific criterion is met, and then perform iterations to solve P_{n+1}, using the last iterate determined in solving P_n as an initial guess for P_{n+1}. This process produces an infinite sequence, and we shall show that there exists at least one accumulation point in the relaxed control topology of the sequence, $\hat{u}$, which satisfies

$\overline{\Theta}(\hat{u}) = 0.$

6.2 ABSTRACT IMPLEMENTATION

In this section, we introduce an abstract implementation scheme. We restate some of the important results from Chapter 3 and 5 as assumptions for Chapter 6.

Assumption 6.2.1:

(i) The function $\bar{g} : \bar{G} \to \mathbb{R}$ is continuous.

(ii) For each $u \in G$, there exists $\nabla g(u) \in L_2 \cap L_\infty$ such that for all $v \in G$,

$$\lim_{\lambda \to 0} \frac{1}{\lambda} | g(u + \lambda(v - u)) - g(u) - \lambda \langle \nabla g(u), v - u \rangle | = 0.$$

(iii) The function $\nabla g(\cdot)$ is uniformly continuous on G in the $L_2 \cap L_\infty$ topology. ■

We introduce a sequence of abstract discretized problems, $\{ P_n \}$. For each $n \in \mathbb{Z}_+$, we define G_n to be a compact, convex, finite-dimensional subset of G and $g_n(\cdot)$ to be a continuously differentiable function from G_n to $\mathbb{R}$ which approximates $g(\cdot)$ on G_n (Assumption 6.2.2.) For each $n \in \mathbb{Z}_+$, the discretized problem P_n is:

$$P_n: \ \min\{ g_n(u) \mid u \in G_n \}. \tag{6.2.1}$$

Since $g_n(\cdot)$ is continuous and G_n is compact, a minimum to (6.2.1) exists. A relaxed control $u \in \bar{G}$ may be isomorphic to an ordinary control in which case u either denotes a relaxed control or an ordinary control depending on the context. We make the following assumptions:

Assumption 6.2.2:

(i) For all $n \in \mathbb{Z}_+$, the functions $g_n : G_n \to \mathbb{R}$ are continuous, i.e., if a sequence $\{ u_i \} \subset G_n$ is such that $u_i \to u \in G_n$, then $g_n(u_i) \to g_n(u)$.

(ii) For all $n \in \mathbb{Z}_+$ and for each $u \in G_n$, there exists $\nabla g_n(u) \in L_2^m \cap L_\infty^m$ such that for all $v \in G_n$,

$$\lim_{\lambda \to 0} \frac{1}{\lambda} | g_n(u + \lambda(v - u)) - g_n(u) - \lambda \langle \nabla g_n(u), v - u \rangle | = 0.$$

(iii) For all $n \in \mathbb{Z}_+$, if a sequence $\{ u_i \} \subset G_n$ is such that $u_i \to u \in G_n$, then $\lim_{i \to \infty} \| \nabla g_n(u_i) - \nabla g_n(u) \| = 0$.

(iv) For all $n \in \mathbb{Z}_+$, $G_n \subset G_{n+1}$.

(v) The closure of $\bigcup_{n \in \mathbb{Z}_+} G_n$ is $\overline{G}$.

(vi) (Uniform Approximation Property.) For all $\varepsilon > 0$, there exists n_ε such that for all $n \geq n_\varepsilon$

(a) $|g(u) - g_n(u)| \leq \varepsilon$, for all $u \in G_n$,

(b) $\| \nabla g(u) - \nabla g_n(u) \| \leq \varepsilon$, for all $u \in G_n$. ■

For each $n \in \mathbb{Z}_+$, the function g_n is continuously differentiable and G_n is compact and so P_n can be solved using Algorithm 3.5.1.

6.3 IMPLEMENTABLE ALGORITHM

Using the above definitions and assumptions, we give an implementable algorithm to solve $\overline{\mathbf{P}}$ (6.1.2):

Algorithm 6.3.1

Data: $n \in \mathbb{Z}_+$, $u_0 \in G_n$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\varepsilon_0 > 0$.

Step 0: Set $i = 0$, $\varepsilon = \gamma^n \varepsilon_0$.

Step 1: Calculate $\nabla g_n(u_i)$.

Step 2: Calculate $\Theta_n(u_i)$ and v_i using

$$\Theta_n(u_i) \triangleq \min_{v \in G_n} \{ \frac{1}{2} \| v - u_i \|^2 + \langle \nabla g_n(u_i), v - u_i \rangle \}, \tag{6.3.1a}$$

$$v_i = v_n(u_i) \triangleq \arg\min_{v \in G_n} \{ \frac{1}{2} \| v - u_i \|^2 + \langle \nabla g_n(u_i), v - u_i \rangle \}. \tag{6.3.1b}$$

Step 3: If $\Theta_n(u_i) = 0$, set $\lambda_i = 0$, otherwise compute the step size λ_i using

$$\lambda_i \triangleq \max\{ \lambda \in \{ 0, 1, \beta, \beta^2, \cdots \} \mid \tag{6.3.2}$$
$$g_n(u_i + \lambda(v_i - u_i)) - g_n(u_i) < \alpha\lambda\Theta_n(u_i) \}.$$

Step 4: Set $u_{i+1} = u_i + \lambda_i(v_i - u_i)$, $n_{i+1} = n$.

Step 5: If $g_n(u_{i+1}) - g_n(u_i) > -\varepsilon$, { Set $n = n + 1$, $\varepsilon = \gamma\varepsilon$ }.

Step 6: Set $i = i + 1$; go to Step 1. ■

This algorithm performs a number of iterations on problem P_n (compare Steps 2, 3, 4 with Steps 2, 3, 4 of Algorithm 3.5.1.) until $g_n(u_{i+1}) - g_n(u_i) > -\varepsilon$ (The algorithm is making insufficient progress in decreasing $g_n(\cdot)$). Then, the discretization is refined (n is increased), and the algorithm begins to solve problem P_{n+1} using the final computed value for P_n as an initial guess for P_{n+1}. We shall show in Theorem 6.3.6 that there exists an accumulation point of the sequence $\{ u_i \}$, $\hat{u} \in \overline{G}$, which satisfies $\overline{\Theta}(\hat{u}) = 0$, where $\overline{\Theta}(\cdot)$ is defined in (5.3.4). We present some preliminary lemmas.

Lemma 6.3.2: For all $n \in \mathbb{Z}_+$, $\Theta_n(\cdot)$ and $v_n(\cdot)$ are continuous.

Proof: The proof follows the proof of Lemma 3.5.4 and Corollary 3.5.5. ■

We make one additional assumption. This assumption is satisfied when G_n is composed of piecewise constant functions and $\nabla g_n(\cdot)$ is piecewise constant. See Section 6.5.

Assumption 6.3.3: For $n \in \mathbb{Z}_+$, if $u \in G_n$ and

$v(u) = \arg\min_{v \in G}\{ \frac{1}{2}|v-u|^2 + \langle \nabla g_n(u), v-u \rangle \}$, then, $v(u) \in G_n$. ■

Lemma 6.3.4: For all $\delta > 0$ there exists $\rho > 0$ and n_0 such that if $\hat{u} \in \overline{G}$ is such that $\overline{\Theta}(\hat{u}) = -\delta < 0$, then $\Theta_n(u) \le -\frac{\delta}{2}$ for all $u \in G_n \cap B_R(\hat{u}, \rho)$ for all $n \ge n_0$, where $B_R(\hat{u},\rho) \triangleq \{ u \in \overline{G} \mid \|u - \hat{u}\|_R \le \rho \}$[1].

Proof:

(i) Recalling the definition of $\Theta_n(\cdot)$ and applying Assumption 6.3.3, for all $n \in \mathbb{Z}_+$ and all $u \in G_n$,

$$\Theta_n(u) = \min_{v \in G}\{ \frac{1}{2}|v-u|^2 + \langle \nabla g_n(u), v-u \rangle \}, \tag{6.3.4a}$$

$$\Theta(u) = \min_{v \in G}\{ \frac{1}{2}|v-u|^2 + \langle \nabla g(u), v-u \rangle \}. \tag{6.3.4b}$$

By the uniform approximation property, Assumption 6.2.2 (vi), for all $\varepsilon_1 > 0$, there exists n_0 such that for all $n > n_0$ and all $u \in G_n$, $v \in G_n$, $|\langle \nabla g_n(u) - \nabla g(u), v-u \rangle| \le |\nabla g_n(u) - \nabla g(u)||v-u| \le \varepsilon_1$. Let $\overline{v}$ be the arg minimizer of (6.3.4a). Then

$$\begin{aligned}\Theta(u) - \Theta_n(u) &\le [\frac{1}{2}|\overline{v}-u|^2 + \nabla g(u), \overline{v}-u \rangle] - [\frac{1}{2}|\overline{v}-u|^2 + \nabla g_n(u), \overline{v}-u \rangle] \\ &\le |\langle \nabla g(u) - \nabla g_n(u), \overline{v}-u \rangle| \le \varepsilon_1 \end{aligned} \tag{6.3.5}$$

Similarly by letting $\tilde{v}$ be the arg minimizer of (6.3.4b), $\Theta_n(u) - \Theta(u) \le \varepsilon_1$, and so for all $n \ge n_0$ and $u \in G_n$, $|\Theta_n(u) - \Theta(u)| \le \varepsilon_1$.

(ii) Since $\overline{\Theta}(\cdot)$ is continuous on the compact set $\overline{G}$, it is uniformly continuous. So for all $\varepsilon_2 > 0$, there exists $\rho > 0$ such that for all $u, v \in \overline{G}$ such that $|u - v|_R \le \rho$, $|\overline{\Theta}(u) - \overline{\Theta}(v)| \le \varepsilon_2$.

[1] The norm $|\cdot|_R$ is defined in (5.1.4).

Combining (i) and (ii) and setting $\varepsilon_1 = \varepsilon_2 = \dfrac{\delta}{4}$ and noting that $\overline{\Theta}(u) = \Theta(u)$ for all $u \in G$ (Lemma 5.3.3), we prove Theorem 6.3.4. ■

Next, we show that n increases to infinity.

Lemma 6.3.5: Let the sequence $\{ n_i \}$ be generated by Algorithm 6.3.1. Then, $\lim_{i \to \infty} n_i = \infty$.

Proof: Suppose $\{ n_i \}$ is bounded. There exists $\hat{n}, \hat{\imath} \in \mathbb{Z}_+$ such that $n_i = \hat{n}$ for all $i \geq \hat{\imath}$. Therefore, for $i \geq \hat{\imath}$ the steps carried out by Algorithm 6.3.1 are identical to Algorithms 3.5.1 and 5.3.1 with $g_{\hat{n}}$ and $\nabla g_{\hat{n}}$ replacing g and ∇g. Since $g_{\hat{n}}(\cdot)$ is continuously differentiable, by Theorem 5.3.8, $|g_{\hat{n}}(u_{i+1}) - g_{\hat{n}}(u_i)| \to 0$ thereby contradicting the Step 5 assumption that $g_{\hat{n}}(u_{i+1}) - g_{\hat{n}}(u_i) \leq -\varepsilon$ for all $i \geq \hat{\imath}$. ■

Lemma 6.3.6: Let $\{ u_i \} \subset G$ be the sequence generated by Algorithm 6.3.1. Let A be the set of all accumulation points of $\{ u_i \}$ in the relaxed control topology. Then $A \subset \overline{G}$ is closed and therefore compact.

Proof: The set A is closed and therefore compact by Theorem 3.5.2 in Munkres [Mun.1].■

Lemma 6.3.7: For any $\varepsilon > 0$, there exists $\hat{\imath}$ such that for all $i \geq \hat{\imath}$, $\min\{ \|u_i - u\| \mid u \in A \} \leq \varepsilon$.

Proof: Suppose there exists $\varepsilon > 0$ such that for all i, there exists $I(i) \geq i$ such that $\min\{ \|u_i - u\| \mid u \in A \} > \varepsilon$. We construct a sequence $\{ v_i \}$, $v_i \triangleq u_{I(i)}$, $i \in \mathbb{Z}_+$. The sequence $\{ v_i \}$ must have an accumulation point $\hat{v} \in A$. However, $\min\{ \|v - u\| \mid u \in A \} \geq \varepsilon$ which is a contradiction. ■

Theorem 6.3.8: Let $\{ u_i \} \subset G$ be the sequence generated by Algorithm 6.3.1. There exists $\hat{u} \in \overline{G}$ which is an accumulation point of $\{ u_i \}$, and $\overline{\Theta}(\hat{u}) = 0$, i.e., a necessary condition for optimality is satisfied.

Proof: Since $\overline{\Theta}(\cdot)$ is continuous and A is compact, if $\overline{\Theta}(u) < 0$ for all $u \in A$, there exists $\delta > 0$ such that $\overline{\Theta}(u) \le -\delta$ for all $u \in A$. We set ε in Lemma 6.3.7 equal to ρ in Lemma 6.3.4. By Lemmas 6.3.4, 6.3.5 and 6.3.7, there exists i_0 such that for all $i \ge i_0$, $\Theta_{n_i}(u_i) \le -\frac{\delta}{2}$.

By differentiability of $g_{n_i}(\cdot)$, Assumption 6.2.2 (ii),

$$g_{n_i}(u_i + \lambda(v_i - u_i)) - g_{n_i}(u_i) = \lambda \langle \nabla g_{n_i}(u_i), v_i - u_i \rangle + \lambda \int_0^1 \langle \nabla g_{n_i}(u_i + s\lambda(v_i - u_i)) - \nabla g_{n_i}(u_i), v_i - u_i \rangle ds. \tag{6.3.6}$$

Subtracting $\alpha\lambda\Theta_{n_i}(u_i)$ from each side of the equation and noting that $\langle \nabla g_{n_i}(u_i), v_i - u_i \rangle \le \Theta_{n_i}(u_i)$,

$$g_{n_i}(u_i + \lambda(v_i - u_i)) - g_{n_i}(u_i) - \alpha\lambda\Theta_{n_i}(u_i) \le (1 - \alpha)\lambda\Theta_{n_i}(u_i) + \lambda \int_0^1 \langle \nabla g_{n_i}(u_i + \lambda s(v_i - u_i)) - \nabla g_{n_i}(u_i), v_i - u_i \rangle ds. \tag{6.3.7}$$

We shall now show the existence of $i_1 \ge i_0$ and $\hat{\lambda} \in (0, 1]$ such that for all $i \ge i_1$, $\lambda \in [0, \hat{\lambda}]$ and $s \in [0, 1]$,

$$\langle \nabla g_{n_i}(u_i + \lambda s(v_i - u_i)) - \nabla g_{n_i}(u_i), v_i - u_i \rangle \le \frac{\delta(1 - \alpha)}{4}. \tag{6.3.8}$$

Since $\nabla g(\cdot)$ is uniformly continuous, Assumption 6.2.1, for all $\varepsilon > 0$, there exists $\mu > 0$ such that for all $u, v \in G$ such that $\| u - v \| \le \mu$, $\| \nabla g(u) - \nabla g(v) \| \le \varepsilon$. Therefore, there exists $\hat{\lambda} \in (0, 1]$ such that for all $i \in \mathbb{Z}_+$, $\lambda \in [0, \hat{\lambda}]$, $s \in [0, 1]$,

$$|\nabla g(u_i + \lambda s(v_i - u_i)) - \nabla g(u_i)| \leq \varepsilon. \tag{6.3.10}$$

By Assumption 6.2.2 (vi), there exists $i_1 \geq i_0$ such that for all $i \geq i_1$, $s \in [0,1]$,

$$|\nabla g_{n_i}(u_i + \lambda s(v_i - u_i)) - \nabla g(u_i + \lambda s(v_i - u_i))| \leq \varepsilon. \tag{6.3.11}$$

$$|\nabla g_{n_i}(u_i) - \nabla g(u_i)| \leq \varepsilon. \tag{6.3.12}$$

Furthermore, there exists $b < \infty$ such that for all $u, v \in G$, $|u - v| \leq b$. Choose $\varepsilon = \dfrac{\delta(1-\alpha)}{12b}$. Then for $i \geq i_1$, $\lambda \in [0, \hat{\lambda}]$, and $s \in [0,1]$,

$$\begin{aligned}
&\langle \nabla g_{n_i}(u_i + \lambda s(v_i - u_i)) - \nabla g_{n_i}(u_i), v_i - u_i \rangle \leq \\
&\qquad [|\nabla g_{n_i}(u_i + \lambda s(v_i - u_i)) - \nabla g(u_i + \lambda s(v_i - u_i))| \\
&\qquad\quad + |\nabla g(u_i + \lambda s(v_i - u_i)) - \nabla g(u_i)| + |\nabla g(u_i) - \nabla g_{n_i}(u_i)|]|v_i - u_i| \\
&\qquad \leq 3\left[\frac{\delta(1-\alpha)}{12b}\right]b = \frac{\delta(1-\alpha)}{4}.
\end{aligned} \tag{6.3.13}$$

By (6.3.13), we see that the Armijo step size rule (Step 3) of Algorithm 6.3.1 chooses $\lambda_i \geq \beta\hat{\lambda}$ for $i \geq i_1$ and so

$$g_{n_i}(u_{i+1}) - g_{n_i}(u_i) \leq \alpha\beta\hat{\lambda}\Theta_{n_i}(u_i) \leq -\frac{\alpha\beta\hat{\lambda}\delta}{2}. \tag{6.3.14}$$

By the uniform approximation property of $g(\cdot)$, Assumption 6.2.2 (vi), there exists $i_2 \geq i_1$ such that for all $i \geq i_2$

$$|g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_{i+1})| \leq \frac{\alpha\beta\hat{\lambda}\delta}{4}. \tag{6.3.15}$$

Therefore,

$$g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_i) \leq -\frac{\alpha\beta\hat{\lambda}\delta}{4}, \tag{6.3.16}$$

which contradicts the fact that $g_{n_i}(u_i) \to g(\hat{u})$. ■

The result in Theorem 6.3.8 is not strong. There may be many accumulation points of the sequence $\{ u_i \}$ such that at these accumulation points $\overline{\Theta} < 0$, i.e., the necessary condition for optimality is not satisfied. It may be difficult to find the accumulation point $\hat{u}$ such that $\overline{\Theta}(\hat{u}) = 0$. However, by replacing Assumption 6.2.2 (vi.a) by the stronger Assumption 6.3.9, we can show that if $\hat{u} \in \overline{G}$ is any accumulation point of $\{ u_i \}$ in the relaxed control topology, then $\overline{\Theta}(\hat{u}) = 0$. A similar algorithm was first proposed by Klessig and Polak [Kle.1] for the case with ODE dynamics and no hard-control constraints.

Assumption 6.3.9: There exists real numbers $\sigma \in (0,1)$ and $b < \infty$ such that for all $n \in \mathbb{Z}_+$ and $u \in G_n$:

$$|g(u) - g_n(u)| \le b\sigma^n. \tag{6.3.17}$$

Theorem 6.3.10: Let Assumption 6.2.2 (vi.a) be replaced by Assumption 6.3.10. Let $\{ u_i \} \subset G$ be the sequence generated by Algorithm 6.3.1. If $\hat{u} \in \overline{G}$ is an accumulation point of $\{ u_i \}$, then $\overline{\Theta}(\hat{u}) = 0$.

Proof: Assume that $u_i \to \hat{u}$ on the subsequence $K \subset \mathbb{Z}_+$ and $\overline{\Theta}(\hat{u}) = -\delta < 0$. By Theorem 6.3.8, there exists i_0 such that for all $i \ge i_0$, $i \in K$, $\Theta_{n_i}(u_i) \le -\frac{\delta}{2}$, and consequently there exists $i_1 \ge i_0$ and $\hat{\lambda} \in (0,1)$ such that $\lambda_i \ge \beta\hat{\lambda}$ for all $i \ge i_1$, $i \in K$, and so

$$g_{n_i}(u_{i+1}) - g_{n_i}(u_i) \le -\frac{\alpha\beta\hat{\lambda}\delta}{2} \tag{6.3.18}$$

for $i \ge i_1$ and $i \in K$.

For $i \in \mathbb{Z}_+$, $g_{n_i}(u_{i+1}) - g_{n_i}(u_i) \le 0$, and $n_{i+1} = n_i$ if $g_{n_i}(u_{i+1}) - g_{n_i}(u_i) \le -\varepsilon$, otherwise,

$n_{i+1} = n_i + 1$. Define $\overline{K} \triangleq \{\, i \in \mathbb{Z}_+ \mid n_{i+1} \neq n_i \,\}$. If $i \notin \overline{K}$, then since $g_{n_{i+1}}(u_{i+1}) = g_{n_i}(u_{i+1})$, $g_{n_{i+1}}(u_{i+1}) \le g_{n_i}(u_i)$. If $i \in \overline{K}$, then $g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_{i+1}) \le |g_{n_{i+1}}(u_{i+1}) - g(u_{i+1})| + |g(u_{i+1}) - g_{n_i}(u_{i+1})| \le \sigma^{n_{i+1}} + \sigma^{n_i}$. Consequently, $g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_i) \le \sigma^{n_{i+1}} + \sigma^{n_i}$.

For $i \ge i_1$, there are four cases:

(i) $i \in K_1 \triangleq K - \overline{K}$,

$$g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_i) \le -\frac{\alpha\beta\hat{\lambda}\delta}{2}, \tag{6.3.19}$$

(ii) $i \in K_2 \triangleq K \cap \overline{K}$,

$$g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_i) \le -\frac{\alpha\beta\hat{\lambda}\delta}{2} + \sigma^{n_{i+1}} + \sigma^{n_i}, \tag{6.3.20}$$

(iii) $i \in K_3 \triangleq \overline{K} - K$,

$$g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_i) \le \sigma^{n_{i+1}} + \sigma^{n_i}, \tag{6.3.21}$$

(iv) $i \notin \overline{K} \cup K$,

$$g_{n_{i+1}}(u_{i+1}) - g_{n_i}(u_i) \le 0. \tag{6.3.22}$$

Therefore,

$$\begin{aligned}
\lim_{i \to \infty} g_{n_i}(u_i) &= g_{i_1}(u_{i_1}) + \sum_{i = i_1}^{\infty} [g_{n_{i+1}}(u_{n_{i+1}}) - g_i(u_i)] \\
&\le g_{i_1}(u_{i_1}) + \sum_{i = i_1, i \in K_1}^{\infty} -\frac{\alpha\beta\hat{\lambda}\delta}{2} + \sum_{i = i_1, i \in K_2}^{\infty} -\frac{\alpha\beta\hat{\lambda}\delta}{2} + \sigma^{n_{i+1}} + \sigma^{n_i} + \sum_{i = i_1, i \in K_3}^{\infty} \sigma^{n_{i+1}} + \sigma^{n} \\
&\le g_{i_1}(u_{i_1}) + \sum_{i = i_1, i \in K_1 \cup K_2}^{\infty} -\frac{\alpha\beta\hat{\lambda}\delta}{2} + \sum_{i = i_1, i \in K_2 \cup K_3}^{\infty} \sigma^{n_{i+1}} + \sigma^{n_i} \\
&\le g_{i_1}(u_{i_1}) + \sum_{i = i_1, i \in K_1 \cup K_2}^{\infty} -\frac{\alpha\beta\hat{\lambda}\delta}{2} + \frac{2}{1 - \sigma}.
\end{aligned}$$

Therefore, $\lim_{i \to \infty} g_{n_i}(u_i) = -\infty$ which contradicts the fact that $g_{n_i}(u_i) \to g(\hat{u})$. ∎

We shall present an alternative algorithm for the case in which Assumption 6.2.2. (iv.a) is satisfied, but Assumption 6.3.9 may not be satisfied. This algorithm uses a different refinement criterion to produce a sequence $\{ u_i \}$ and a filter to produce a sequence $\{ w_n \} \subset \{ u_i \}$ such that any accumulation point $\hat{w} \in \overline{G}$ of $\{ w_n \}$ satisfies $\overline{\Theta}(\hat{w}) = 0$.[2] This algorithm is identical to Algorithm 6.3.1 except that the refinement criterion (Step 5) is replaced by Step 5':

Step 5': If $-\Theta_n(u_{i+1}) \le \varepsilon$, $\{$ Set $w_n = u_{i+1}$, $n = n + 1$, $\varepsilon = \gamma\varepsilon$ $\}$.

Algorithm 6.3.11:

Data: $u_0 \in G$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $n \in \mathbb{Z}_+$, $\varepsilon_0 > 0$.

Step 0: $i = 0$, $\varepsilon = \gamma^n \varepsilon_0$.

Step 1: Calculate $\nabla g_n(u_i)$.

Step 2: Calculate $\Theta_n(u_i)$ and v_i using (6.3.1a) and (6.3.1b).

Step 3: If $\Theta_n(u_i) = 0$, set $\lambda_i = 0$, otherwise compute the step size λ_i using (6.3.2).

Step 4: Set $u_{i+1} = u_i + \lambda_i(v_i - u_i)$, $n_{i+1} = n$.

Step 5': If $-\Theta_n(u_{i+1}) \le \varepsilon$, $\{$ Set $w_n = u_{i+1}$, $n = n + 1$, $\varepsilon = \gamma\varepsilon$ $\}$.

Step 6: Set $i = i + 1$; go to Step 1. ∎

Lemma 6.3.12: Let $\{ u_i \}$ and $\{ n_i \}$ be generated by Algorithm 6.3.11. Then $\lim_{i \to \infty} n_i = \infty$.

---

[2] The use of a filter or sieve to obtain a subsequence whose accumulation points satisfy a necessary condition for optimality was first proposed by Klessig and Polak [Kle.2].

Proof: Suppose $\{ n_i \}$ is bounded. There exists $\hat{n}, \hat{\imath} \in \mathbb{Z}_+$ such that $n_i = \hat{n}$ for all $i \geq \hat{\imath}$. Since $\{ u_i \} \subset G_{\hat{n}}$, and $G_{\hat{n}}$ is compact, there exists $\hat{u} \in G_{\hat{n}}$ and a subsequence $K \subset \mathbb{Z}_+$ such that $u_i \xrightarrow{K} \hat{u}$ and therefore by Theorem 3.5.6, $\Theta_{\hat{n}}(\hat{u}) = 0$ thereby contradicting the assumption that $-\Theta_{\hat{n}}(u_i) > \varepsilon$ for all $i \geq \hat{\imath}$ ■

Theorem 6.3.13: Let $\{ w_n \}$ be generated by Algorithm 6.3.11. If $\hat{w} \in \overline{G}$ is an accumulation point of $\{ w_n \}$ in the relaxed control topology, then $\overline{\Theta}(\hat{w}) = 0$.

Proof: Suppose that there exists $\hat{w} \in \overline{G}$ and a subsequence $K \subset \mathbb{Z}_+$ such that $w_n \xrightarrow{K} \hat{w}$, and $\overline{\Theta}(\hat{w}) = -\delta < 0$. By Lemma 6.3.4, there exists $\hat{n}$ such that for all $n \geq \hat{n}$, $n \in K$, $\Theta_n(w_n) \leq -\frac{\delta}{2}$. However, for all $n \in \mathbb{Z}_+$, $-\Theta_n(w_n) \leq \gamma' \varepsilon_0$ or $\Theta_n(w_n) \geq -\gamma' \varepsilon_0$ which is a contradiction. ■

6.4 EXAMPLE - PRELIMINARIES

In Section 6.2, we made a number of assumptions regarding $g_n(\cdot)$ and $\nabla g_n(\cdot)$. In the next two sections, we present a simple optimal control problem; we provide a discretization scheme; and we show that this scheme satisfies the assumptions of Section 6.2.

We seek to solve the following problem:

$$P : \min \{ g(u) \mid u \in G \triangleq \{ u \in L_2^m([0,1]) \mid \|u(t)\|_\infty \le b, t \in [0,1] \} \}, \tag{6.4.0}$$

where $g(u)$ can be either the potential or kinetic energy of system (6.4.2) at the final time. That is

$$g(u) = P(u) = P(1,u),\ P(t,u) \triangleq \frac{1}{2}\|\frac{\partial}{\partial x}w(t,\cdot,u)\|^2, \tag{6.4.1a}$$

or

$$g(u) = K(u) = K(1,u),\ K(t,u) \triangleq \frac{1}{2}\|\frac{\partial}{\partial t}w(t,\cdot,u)\|^2, \tag{6.4.1b}$$

where $w(\cdot,\cdot,u)$ is the solution to system (6.4.2):

$$\frac{\partial^2}{\partial t^2}w(t,x,u) - \alpha\frac{\partial^3}{\partial t\partial x^2}w(t,x,u) - \frac{\partial^2}{\partial x^2}w(t,x,u) = f(x,u(t)), \quad t \in [0,1],\ x \in [0,1], \tag{6.4.2}$$

with boundary conditions:

$$w(t,0,u) = \frac{\partial}{\partial x}w(t,1,u) = 0,\ t \in [0,1], \tag{6.4.3}$$

and initial conditions:

$$w(0,x,u) = \gamma(x),\ \frac{\partial}{\partial t}w(0,x,u) = \beta(x),\ x \in [0,1], \tag{6.4.4}$$

with $\gamma(\cdot)$ and $\beta(\cdot)$ chosen to be smooth in x and to satisfy the boundary conditions; $\alpha \ge 0$; and $f(\cdot,\cdot)$ chosen such that for all $u \in \mathbb{R}^m$ such that $\|u\|_\infty \le b$, $f(x,u)$ is smooth in x and satisfies the boundary conditions (6.4.3). Finally $f(\cdot,u) \in C([0,1])$ is continuously differentiable with respect to u.

Let us assume we are interested in minimizing the potential energy at the final time, $g(u) = P(u)$. The case in which $g(u) = K(u)$ follows easily. By putting (6.4.2) into abstract form as was done for system (A2.2.1) and (A2.2.2) in Appendix 2, we follow the steps in Appendix 2 to show that the partial differential equation (6.4.2) gives rise to an infinitesimal generator of semigroup. The semigroup is analytic if $\alpha > 0$. Furthermore, we can see that $g(u) \triangleq \frac{1}{2}\|\frac{\partial w}{\partial x}(1,\cdot,u)\|^2$ is a continuously differentiable function in $u \in L_2 \cap L_\infty$ and therefore all requirements of Assumption 6.2.1 are satisfied.

We now derive the gradient $\nabla g(u)$ for $g(\cdot)$ defined by (6.4.1) for the case with no damping ($\alpha = 0$). The case with positive damping ($\alpha > 0$) follows easily.

Lemma 6.4.1: **Differentiability** Let $\delta w(\cdot,\cdot,u,\delta u)$ be the solution to:

$$\frac{\partial^2}{\partial t^2}\delta w(t,\cdot,u,\delta u) - \frac{\partial^2}{\partial x^2}\delta w(t,\cdot,u,\delta u) = \frac{\partial}{\partial u}f(\cdot,u(t))\delta u(t), \tag{6.4.5}$$

with boundary conditions:

$$\delta w(t,0,u) = \frac{\partial}{\partial x}\delta w(t,1,u) = 0,\ t \in [0,1], \tag{6.4.6}$$

and initial conditions:

$$\delta w(0,x,u) = \frac{\partial}{\partial t}\delta w(0,x,u) = 0,\ x \in [0,1]. \tag{6.4.7}$$

Then,

$$\lim_{h \to 0} \frac{1}{h}\|w(t,\cdot,u + h\delta u) - w(t,\cdot,u) - h\delta w(t,\cdot,u,\delta u)\|_X = 0, \tag{6.4.8}$$

where $\|\cdot\|_X$ is the energy norm, i.e.,

$$\|w(t,\cdot)\|_X^2 \triangleq \frac{1}{2}\|\frac{\partial}{\partial t}w(t,\cdot)\|^2 + \frac{1}{2}\|\frac{\partial}{\partial x}w(t,\cdot)\|^2. \tag{6.4.9}$$

Proof: Let $\eta(t,x,h,\delta u) \triangleq w(t,x,u+h\delta u) - w(t,x,u) - h\delta w(t,x,u,\delta u)$ and $\delta f(t,x,h,\delta u) \triangleq f(x,u(t)+h\delta u(t)) - f(x,u(t)) - h\frac{\partial}{\partial u}f(x,u(t))\delta u(t)$. Then η satisfies:

$$\frac{\partial^2}{\partial t^2}\eta(t,x,h,\delta u) - \frac{\partial^2}{\partial x^2}\eta(t,x,h,\delta u) = \delta f(t,x,h,\delta u), \tag{6.4.10}$$

with boundary conditions and initial conditions:

$$\eta(t,0,h,\delta u) = \frac{\partial}{\partial x}\eta(t,1,h,\delta u) = 0,\ t \in [0,1], \tag{6.4.11}$$

$$\eta(0,x,h,\delta u) = \frac{\partial}{\partial t}\eta(0,x,h,\delta u) = 0,\ x \in [0,1].$$

Multiplying (6.4.10) by $\frac{\partial}{\partial t}\eta(t,x,h,\delta u)$ and integrating from 0 to 1 in x yields

$$\frac{1}{2}\frac{\partial}{\partial t}\|\frac{\partial}{\partial t}\eta(t,\cdot,h,\delta u)\|^2 + \frac{1}{2}\frac{\partial}{\partial t}\|\frac{\partial}{\partial x}\eta(t,\cdot,h,\delta u)\|^2 = \int_0^1 \delta f(t,x,h,\delta u)\frac{\partial}{\partial t}\eta(t,x,h,\delta u)dx, \tag{6.4.12}$$

where we integrated the second term by parts and employed the boundary condition (6.4.11)[1],

Integrating in time from 0 to t and using the initial conditions,

$$\frac{1}{2}\|\frac{\partial}{\partial t}\eta(t,\cdot,h,\delta u)\|^2 + \frac{1}{2}\|\frac{\partial}{\partial x}\eta(t,\cdot,h,\delta u)\|^2 = \int_0^t\int_0^1 \delta f(s,x,h,\delta u)\frac{\partial}{\partial t}\eta(s,x,h,\delta u)dxds. \tag{6.4.13}$$

Noting that $\int_0^1 \delta f(s,x,h,\delta u)\frac{\partial}{\partial s}\eta(s,\cdot,h,\delta u)dx \le \frac{1}{2}\|\delta f(s,x,h,\delta u)\|^2 + \|\frac{\partial}{\partial s}\eta(s,\cdot,h,\delta u)\|^2$ and adding $\frac{1}{2}\|\frac{\partial}{\partial x}\eta(s,\cdot,h,\delta u)\|^2$ to the right hand side of (6.4.13), we obtain

$$\frac{1}{2}\|\frac{\partial}{\partial t}\eta(t,\cdot,h,\delta u)\|^2 + \frac{1}{2}\|\frac{\partial}{\partial x}\eta(t,\cdot,h,\delta u)\|^2 \tag{6.4.14}$$

[1] $\|\cdot\|$ denotes the L_2 norm unless otherwise stated.

$$\leq \int_0^t [\frac{1}{2}|\delta f(s,\cdot,h,\delta u)|^2 + \frac{1}{2}|\frac{\partial}{\partial s}\eta(s,\cdot,h,\delta u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}\eta(s,\cdot,h,\delta u)|^2]ds.$$

Applying the Bellman-Gronwall Lemma (see Section 6.6),

$$\frac{1}{2}|\frac{\partial}{\partial t}\eta(t,\cdot,h,\delta u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}\eta(t,\cdot,h,\delta u)|^2 \leq \frac{1}{2}\int_0^t |\delta f(s,\cdot,h,\delta u)|^2 ds\ e^{t/2}. \tag{6.4.15}$$

Equivalently,

$$\frac{1}{h^2}|\eta(t,\cdot,h,\delta u)|_X^2 \leq \frac{1}{2}\int_0^t \frac{1}{h^2}|\delta f(s,\cdot,h,\delta u)|^2 ds\ e^{t/2}. \tag{6.4.16}$$

Since $f(\cdot,u)$ is differentiable,

$$\lim_{h\to 0}\frac{1}{h}|\eta(t,\cdot,h,\delta u)|_X = 0, \tag{6.4.17}$$

and the theorem is proved. ■

Lemma 6.4.2: (Differentiability.) Let $g(u) = \frac{1}{2}|\frac{\partial}{\partial x}w(1,\cdot,u)|^2$. Then,

$$\lim_{h\to 0}\frac{1}{h}|g(u+h\delta u) - g(u) - h\langle \frac{\partial}{\partial x}w(1,\cdot,u), \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u)\rangle| = 0. \tag{6.4.18}$$

Proof:

$$\frac{1}{h}|g(u+h\delta u) - g(u) - h\langle \frac{\partial}{\partial x}w(1,\cdot,u), \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u)\rangle| = \tag{6.4.19}$$

$$|\frac{1}{2h}[\frac{\partial}{\partial x}\eta(1,\cdot,h,\delta u) + 2\frac{\partial}{\partial x}w(1,\cdot,u) + h\frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u), \frac{\partial}{\partial x}\eta(1,\cdot,h,u) + h\frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u)\ \}$$

$$-\ \langle \frac{\partial}{\partial x}w(1,\cdot,u), \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u)\rangle|$$

$$= \frac{1}{2h}\langle \frac{\partial}{\partial x}\eta(1,\cdot,h,\delta u), \frac{\partial}{\partial x}\eta(1,\cdot,h,\delta u)\rangle + \frac{1}{2}\langle \frac{\partial}{\partial x}\eta(1,\cdot,h,\delta u), \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u)\rangle$$

$$+ \frac{1}{h}\langle \frac{\partial}{\partial x}w(1,\cdot,u), \frac{\partial}{\partial x}\eta(1,\cdot,h,\delta u)\rangle$$

$$+ \frac{1}{2}\langle \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u), \frac{\partial}{\partial x}\eta(1,\cdot,h,\delta u)\rangle + h|\frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u)|^2.$$

The limit of the right hand side as h approaches zero is zero by (6.4.17). ■

Therefore the directional derivative of $g(\cdot)$, $dg(u;\delta u) = \langle \frac{\partial}{\partial x}w(1,\cdot,u), \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u) \rangle$. By Lemma 3.4.3, there exists $\nabla g(u) \in L_2 \cap L_\infty$ such that $\langle \nabla g(u), \delta u \rangle = dg(u;\delta u)$.

The weak form of (6.4.2) is:

For all $\phi \in H^* \triangleq \{ \psi \in H^1 \mid \psi(0) = 0 \}$,

$$\langle \frac{\partial^2}{\partial t^2}w(t,\cdot,u),\phi(\cdot) \rangle + \alpha \langle \frac{\partial^2}{\partial t \partial x}w(t,\cdot,u), \frac{\partial}{\partial x}\phi(\cdot) \rangle + \langle \frac{\partial}{\partial x}w(t,\cdot,u), \frac{\partial}{\partial x}\phi(\cdot) \rangle \quad (6.4.20)$$

$$= \langle f(\cdot,u(t)),\phi(\cdot) \rangle.$$

We now prove an inequality relating the energy of the system at any time to its initial energy and input energy. In Section 6.5, we shall propose a discretization scheme and use a discrete version of this inequality to show stability and convergence of our scheme.

Lemma 6.4.3 Energy Inequality: There exists $c < \infty$ such that for all initial states and inputs and times $t \in [0,1]$,

$$K(t,u) + P(t,u) \le c\{ K(0,u) + P(0,u) \} + c\int_0^t \| f(\cdot,u(s))\|^2 ds, \quad (6.4.21)$$

where $K(t,u) \triangleq \frac{1}{2}\|\frac{\partial}{\partial t}w(t,\cdot,u)\|^2$, $P(t,u) \triangleq \frac{1}{2}\|\frac{\partial}{\partial x}w(t,\cdot,u)\|^2$.

Proof: Set $\phi(\cdot) \triangleq \frac{\partial}{\partial t}w(t,\cdot,u)$ in the weak form (6.4.20):

$$\langle \frac{\partial^2}{\partial t^2}w(t,\cdot,u), \frac{\partial}{\partial t}w(t,\cdot,u) \rangle + \alpha \langle \frac{\partial^2}{\partial t \partial x}w(t,\cdot,u), \frac{\partial^2}{\partial t \partial x}w(t,\cdot,u) \rangle \quad (6.4.22)$$

$$+ \langle \frac{\partial}{\partial x}w(t,\cdot,u), \frac{\partial^2}{\partial t \partial x}w(t,\cdot,u) \rangle = \langle f(\cdot,u(t)), \frac{\partial}{\partial t}w(t,\cdot,u) \rangle.$$

Equivalently,

$H^1 \triangleq \{ u \in L_2([0,1]) \mid u$ is differentiable and $u' \in L_2([0,1]) \}$.

$$\frac{1}{2}\frac{\partial}{\partial t}|\frac{\partial}{\partial t}w(t,\cdot,u)|^2 + \frac{1}{2}\frac{\partial}{\partial t}|\frac{\partial}{\partial x}w(t,\cdot,u)|^2$$

$$= \langle f(\cdot,u(t)),\frac{\partial}{\partial t}w(t,\cdot,u)\rangle - \alpha\|\frac{\partial^2}{\partial t\partial x}w(t,\cdot,u)|^2. \tag{6.4.23}$$

Integrating with respect to time, we obtain:

$$\frac{1}{2}|\frac{\partial}{\partial t}w(t,\cdot,u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}w(t,\cdot,u)|^2 = \frac{1}{2}|\frac{\partial}{\partial t}w(0,\cdot,u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}w(0,\cdot,u)|^2 \tag{6.4.24}$$

$$+\int_0^t \langle f(\cdot,u(s)),\frac{\partial}{\partial t}w(s,\cdot,u)\rangle ds - \alpha\int_0^t |\frac{\partial^2}{\partial t\partial x}w(s,\cdot,u)|^2 ds.$$

Since $\langle f(\cdot,u(s)),\frac{\partial}{\partial t}w(s,\cdot,u)\rangle \le \frac{1}{2}|f(\cdot,u(s))|^2 + \frac{1}{2}|\frac{\partial}{\partial t}w(s,\cdot,u)|^2$ and

$-\alpha\int_0^t|\frac{\partial^2}{\partial t\partial x}w(s,\cdot,u)|^2 \le 0 \le \frac{1}{2}\int_0^t |\frac{\partial}{\partial x}w(s,\cdot,u)|^2 ds$, (6.4.24) becomes:

$$\frac{1}{2}|\frac{\partial}{\partial t}w(t,\cdot,u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}w(t,\cdot,u)|^2 \le \frac{1}{2}|\frac{\partial}{\partial t}w(0,\cdot,u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}w(0,\cdot,u)|^2$$

$$+\frac{1}{2}\int_0^t |f(\cdot,u(s))|^2 ds + \frac{1}{2}\int_0^t |\frac{\partial}{\partial s}w(s,\cdot,u)|^2 ds + \frac{1}{2}\int_0^t |\frac{\partial}{\partial x}w(s,\cdot,u)|^2 ds. \tag{6.4.25}$$

Finally, by applying the Bellman-Gronwall Lemma to the right hand side of (6.4.25), we obtain:

$$\frac{1}{2}|\frac{\partial}{\partial t}w(t,\cdot,u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}w(t,\cdot,u)|^2 \tag{6.4.26}$$

$$\le [\frac{1}{2}|\frac{\partial}{\partial t}w(0,\cdot,u)|^2 + \frac{1}{2}|\frac{\partial}{\partial x}w(0,\cdot,u)|^2 + \frac{1}{2}\int_0^t |f(\cdot,u(s))|^2 ds]e^{t/2}.$$

■

6.5 EXAMPLE - DISCRETIZATION

We use the Finite Element Method and Newmark's β integration scheme [New.1] with $\beta = 0$ to solve (6.4.2) - (6.4.4). The analysis of this integration scheme with $\beta \ge 0$ was originally done by Fujii [Fuj.1] for a class of vibrational problems. In this section we adapt his analysis to show that the Uniform Approximation Property, Assumption 6.2.2. (vi), is valid for optimal control problem (6.4.0). Oden and Fost [Ode.1] have extended Fujii's analysis to the non-linear wave equation $u_{tt} - F(u_x) = f(x,t)$ where $F(u_x)$ possesses properties generally encountered in non-linear elasticity.

Spatial Discretization (Finite Element Method)

Since our PDE is of second order, we use linear elements to approximate the solution. For a given q, there are $2^q + 1$ discretization points spaced equally over $[0,1]$. We choose basis functions $\{ \psi_q^i(\cdot) \}_{i=0}^{2^q}$. Let $k \triangleq 2^{-q}$.

$$\psi_q^i(x) = \begin{cases} (x-(i-1))1/k & x \in [(i-1)k, ik] \\ 1-(x-ik)1/k & x \in [ik,(i+1)k]. \\ 0 & \text{otherwise} \end{cases} \tag{6.5.0}$$

This basis generates the subspace H_q:

$$H_q \triangleq \{ \psi \in H^* \mid \psi = \sum_{i=0}^{2^q} \lambda_i \psi_q^i,\ \lambda \in \mathbb{R}^{2^q+1} \}, \tag{6.5.1}$$

where $2^q + 1$ is the number of grid points. It is clear that for $\psi = \sum_{i=0}^{2^q} \lambda_i \psi_q^i$ to be a member of H^* λ must be zero. We subsequently ignore the λ_0 term. We approximate $w(t,\cdot,u)$ by a member of H_q: $w(t,\cdot,u) = \sum_{i=1}^{2^q} \eta_q^i(t,u)\psi_q^i(\cdot)$ where $\eta_q : [0,1]\times G \to \mathbb{R}^{2^q}$. The weak form (1.9) becomes: For $\phi \in H_q$,

$$\langle \frac{\partial^2}{\partial t^2} \sum_{i=0}^{2^q} \eta_q^i(t,u)\psi_q^i(\cdot), \phi(\cdot) \rangle + \alpha \langle \frac{\partial^2}{\partial t \partial x} \sum_{i=1}^{2^q} \eta_q^i(t,u)\psi_q^i(\cdot), \frac{\partial}{\partial x}\phi(\cdot) \rangle + \tag{6.5.2}$$

$$\langle \frac{\partial}{\partial x} \sum_{i=1}^{2^q} \eta_q^i(t,u)\psi_q^i(\cdot), \frac{\partial}{\partial x}\phi(\cdot) \rangle = \langle f(\cdot,u(t)), \phi(\cdot) \rangle.$$

Temporal Discretization

For $n \in \mathbb{Z}_+$, we discretize time into a grid of 2^{2n} intervals of equal size and consider as inputs members of

$$G_n \triangleq R_n \cap G, \tag{6.5.3}$$

where

$$R_n \triangleq \{\, u \in L_2^m([0,1]) \mid \text{u is piece-wise constant, continuous from the right and changes only when } t = j2^n h \text{ for } j \in \{1,2,\ldots 2^n\}\,\}, \tag{6.5.4}$$

where $h \triangleq 2^{-2n}$. We approximate $\eta_q^i(jh,u)$ by $\xi_{n,q}^i(j,u)$, $j \in \{0,1,\cdots 2^{2n}\}$, and the time derivatives by finite differences. We use Newmark's β scheme to approximate the differential equation. We approximate the second order time derivative $\frac{\partial^2}{\partial t^2}$ by

$$\frac{\partial^2}{\partial t^2}\eta(jh,u) \approx D_t D_{\bar{t}}\xi(j,u) = \frac{1}{h^2}(\xi(j+1,u) - 2\xi(j,u) + \xi(j-1,u)) \tag{6.5.5}$$

and the first order time derivative $\frac{\partial}{\partial t}$ by

$$\frac{\partial}{\partial t}\eta(jh,u) \approx \frac{1}{2}(D_t + D_{\bar{t}})\xi(j,u) = \frac{1}{2h}(\xi(j+1,u) - \xi(j-1,u)). \tag{6.5.6}$$

Applying this discretization scheme (6.5.2) becomes:

$$\langle D_t D_{\bar{t}} \sum_{i=1}^{2^q} \xi_{n,q}^i(j,u)\psi_q^i(\cdot),\phi(\cdot)\rangle + \alpha\langle \frac{1}{2}(D_t + D_{\bar{t}})\sum_{i=1}^{2^q} \xi_{n,q}^i(j,u)\frac{\partial}{\partial x}\psi_q^i(\cdot),\frac{\partial}{\partial x}\phi(\cdot)\rangle \tag{6.5.7}$$

$$+ \langle \sum_{i=1}^{2^q} \xi_{n,q}^i(j,u)\frac{\partial}{\partial x}\psi_q^i(\cdot),\frac{\partial}{\partial x}\phi(\cdot)\rangle = \langle f(\cdot,u(jh)),\phi(\cdot)\rangle,$$

for all $\phi(\cdot) \in H_q$. By setting $\phi(\cdot) = \psi_q^i(\cdot) \in H_q$, $i \in 2^q$, (6.5.7) becomes

---

$D_t\xi(j,u) = \frac{1}{h}(\xi(j+1,u) - \xi(j,u))$; $D_{\bar{t}}\xi(j,u) = \frac{1}{h}(\xi(j,u) - \xi(j-1,u))$.

$$M_q D_t D_{\bar{t}} \xi_{n,q}(j,u) + \frac{\alpha}{2} K_q (D_t + D_{\bar{t}}) \xi_{n,q}(j,u) + K_q \xi_{n,q}(j,u) = f_{n,q}(j,u), \tag{6.5.8}$$

where

$$K_q \triangleq \begin{bmatrix} \langle \frac{\partial}{\partial x}\psi_q^1, \frac{\partial}{\partial x}\psi_q^1 \rangle & \cdots & \langle \frac{\partial}{\partial x}\psi_q^1, \frac{\partial}{\partial x}\psi_q^{2^q} \rangle \\ \vdots & & \vdots \\ \langle \frac{\partial}{\partial x}\psi_q^{2^q}, \frac{\partial}{\partial x}\psi_q^1 \rangle & \cdots & \langle \frac{\partial}{\partial x}\psi_q^{2^q}, \frac{\partial}{\partial x}\psi_q^{2^q} \rangle \end{bmatrix}, \tag{6.5.9a}$$

$$M_q \triangleq \begin{bmatrix} \langle \psi_q^1, \psi_q^1 \rangle & \cdots & \langle \psi_q^1, \psi_q^{2^q} \rangle \\ \vdots & & \vdots \\ \langle \psi_q^{2^q}, \psi_q^1 \rangle & \cdots & \langle \psi_q^{2^q}, \psi_q^{2^q} \rangle \end{bmatrix}, \quad f_{n,q} \triangleq \begin{bmatrix} \langle \psi_q^1, f(\cdot, u(jh)) \rangle \\ \vdots \\ \langle \psi_q^{2^q}, f(\cdot, u(jh)) \rangle \end{bmatrix}. \tag{6.5.9b}$$

M_q is positive definite, and K_q is positive semi-definite. If $\{ \xi_{n,q}(j,u) \}_{j=0}^{2^{2n}}$ is the solution to (6.5.8), then we hope that $\sum_{i=1}^{2^q} \xi_{n,q}^i(j,u)\psi_q^i(\cdot)$ is a good approximation to $w(jh,\cdot,u)$. We shall determine the validity of this approximation.

Stability

To show stability of this discretization scheme, we shall show the discrete analog of the energy inequality (6.4.21). We have an important lemma from Fujii[Fuj.1]:

Lemma 6.5.1 [Fujii]: For $q \in \mathbb{Z}_+$ and $\xi \in 2^q$,

$$\xi^T K_q \xi \leq \frac{\gamma_1^2}{k^2} \xi^T M_q \xi \text{ where } \gamma_1 = 2\sqrt{3} \text{ and } k = 2^{-q}.$$

■

Lemma 6.5.2: Discrete Energy Inequality.

For $q \in \mathbb{Z}_+, n \in \mathbb{Z}_+$ such that $2n - q \geq \log_2 3$, there exists $C < \infty$ such that for $u \in G_n$,

$$\| M_q^{1/2} D_{\bar{t}} \xi(r,u) \|^2 + \| K_q^{1/2} \xi(r,u) \|^2 \leq \tag{6.5.10}$$

$$C(\| M_q^{1/2} D_t \xi(0,u) \|^2 + \| K_q^{1/2} \xi(0,u) \|^2) + C \sum_{j=1}^{r-1} h \| f(\cdot, u(jh)) \|^2.$$

Proof: Set $\phi(\cdot) \triangleq \frac{1}{2}(D_t + D_{\bar{t}}) \sum_{k=1}^{2^q} \xi_{n,q}^k(j,u)\psi_q^k(\cdot)$ in equation (6.5.7):

$$\langle D_t D_{\bar{t}} \sum_{i=1}^{2^q} \xi_{n,q}^i(j,u)\psi_q^i(\cdot), \frac{1}{2}(D_t + D_{\bar{t}}) \sum_{k=1}^{2^q} \xi_{n,q}^k(j,u)\psi_q^k(\cdot) \rangle$$

$$+ \alpha \| \frac{1}{2}(D_t + D_{\bar{t}}) \sum_{k=1}^{2^q} \xi_{n,q}^k(j,u)\frac{\partial}{\partial x}\psi_q^k(\cdot)\|^2$$

$$+ \langle \sum_{i=1}^{2^q} \xi_{n,q}^i(j,u)\frac{\partial}{\partial x}\psi_q^i(\cdot), \frac{1}{2}(D_t + D_{\bar{t}}) \sum_{k=1}^{2^q} \xi_{n,q}^k(j,u)\frac{\partial}{\partial x}\psi_q^k(\cdot) \rangle$$

$$= \langle f(\cdot, u(jh)), \frac{1}{2}(D_t + D_{\bar{t}}) \sum_{k=1}^{2^q} \xi_{n,q}^k(j,u)\psi_q^k(\cdot) \rangle. \tag{6.5.11}$$

Define $M = M_q$ and $K = K_q$ as in (6.5.9). Then (6.5.11) becomes:

$$\frac{1}{2}(D_t D_{\bar{t}}\xi_{n,q}(j,u))^T M(D_t + D_{\bar{t}})\xi_{n,q}(j,u) + \frac{\alpha}{4}((D_t + D_{\bar{t}})\xi_{n,q}(j,u))^T K(D_t + D_{\bar{t}})\xi_{n,q}(j,u)$$

$$+ \frac{1}{2}\xi_{n,q}(j,u)^T K(D_t + D_{\bar{t}})\xi_{n,q}(j,u)$$

$$\leq \frac{1}{2}[\|f(\cdot, u(jh))\|^2 + \frac{1}{4}\|(D_t + D_{\bar{t}}) \sum_{k=1}^{2^q} \xi_{n,q}^k(j,u)\psi_q^k(\cdot)\|^2], \tag{6.5.12}$$

where we have used the fact that $\langle x,y \rangle \leq \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2$. Multiplying (6.5.12) by h and summing from $j = 1$ to $r - 1$,

$$\frac{1}{2h^2}\sum_{j=1}^{r-1}(\xi_{n,q}(j+1,u) - 2\xi_{n,q}(j,u) + \xi_{n,q}(j-1,u))^T M(\xi_{n,q}(j+1,u) - \xi_{n,q}(j-1,u))$$

$$+ \frac{1}{2}\sum_{j=1}^{r-1}\xi_{n,q}(j,u)^T K(\xi_{n,q}(j+1,u) - \xi_{n,q}(j-1,u))$$

$$= \frac{1}{2}\sum_{j=1}^{r-1} h\Bigg[\|f(\cdot, u(jh))\|^2 + \frac{1}{4}\|(D_t + D_{\bar{t}})M^{1/2}\xi_{n,q}(j,u)\|^2$$

$$- \frac{\alpha}{4}\|(D_t + D_{\bar{t}})K^{1/2}\xi_{n,q}(j,u)\|^2\Bigg]. \tag{6.5.13}$$

For simplicity, define $\xi_j \triangleq \xi_{n,q}(j,u)$. We can remove the α term so that (6.5.13) becomes:

$$\frac{1}{2h^2}[(\xi_r^TM\xi_r - 2\xi_r^TM\xi_{r-1} + \xi_{r-1}^TM\xi_{r-1}) - (\xi_1^TM\xi_1 - 2\xi_1^TM\xi_0 + \xi_0^TM\xi_0)] \tag{6.5.14}$$

$$+ \frac{1}{2}(\xi_{r-1}^TK\xi_r - \xi_1^TK\xi_0) \le \frac{1}{2}\sum_{j=1}^{r-1} h[\|f(\cdot,u(jh))\|^2 + \frac{1}{4}\|(D_t + D_{\bar{t}})M^{1/2}\xi_{n,q}(j,u)\|^2].$$

Regrouping terms, (6.5.14) becomes:

$$\frac{1}{2}(D_{\bar{t}}\xi_r)^TM(D_{\bar{t}}\xi_r) - \frac{1}{2}(D_t\xi_0)^TM(D_t\xi_0) + \frac{1}{2}[\xi_r^TK\xi_r + (\xi_{r-1} - \xi_r)^TK\xi_r \tag{6.5.15}$$

$$- \xi_0^TK\xi_0 - (\xi_1 - \xi_0)^TK\xi_0] \le \frac{1}{2}\sum_{j=1}^{r-1} h[\|f(\cdot,u(jh))\|^2 + \frac{1}{4}\|(D_t + D_{\bar{t}})M^{1/2}\xi_{n,q}(j,u)\|^2]$$

$$\frac{1}{2}(D_{\bar{t}}\xi_r)^TM(D_{\bar{t}}\xi_r) + \frac{1}{2}(\xi_r^TK\xi_r) \le \frac{1}{2}(D_t\xi_0)^TM(D_t\xi_0) + \frac{1}{2}\xi_0^TK\xi_0 \tag{6.5.16}$$

$$- \frac{1}{2}(hK^{1/2}D_{\bar{t}}\xi_r, K^{1/2}\xi_r) + \frac{1}{2}(hD_tK^{1/2}\xi_0, K^{1/2}\xi_0)$$

$$+ \frac{1}{2}\sum_{j=1}^{r-1} h[\|f(\cdot,u(jh))\|^2 + \frac{1}{4}\|(D_t + D_{\bar{t}})M^{1/2}\xi_{n,q}(j,u)\|^2].$$

Since $-\langle x,y\rangle \le \frac{1}{2\varepsilon}\|x\|^2 + \frac{\varepsilon}{2}\|y\|^2$ for $\varepsilon > 0$, Lemma 6.5.1 implies

$$\langle hK^{1/2}D_{\bar{t}}\xi_r, K^{1/2}\xi_r\rangle \le \frac{\varepsilon}{2}(\xi_r^TK\xi_r) + \frac{h^2}{2\varepsilon}(D_{\bar{t}}\xi_r^TKD_{\bar{t}}\xi_r)$$

$$\le \frac{\varepsilon}{2}(\xi_r^TK\xi_r) + \frac{\gamma_1^2}{2\varepsilon}\frac{h^2}{k^2}(D_{\bar{t}}\xi_r)^TMD_{\bar{t}}\xi_r. \tag{6.5.17}$$

And since

$$\sum_{i=1}^{r-1}\|\xi_{i+1} - \xi_{i-1}\|^2 \le 4\sum_{i=1}^{r-1}\|\xi_i - \xi_{i-1}\|^2 + 2\|\xi_r - \xi_{r-1}\|^2,$$

(6.5.16) becomes

$$\frac{1}{2}(1 - \frac{\gamma_1^2}{2\varepsilon}\frac{h^2}{k^2} - \frac{h}{2})(D_{\bar{t}}\xi_r)^TM(D_{\bar{t}}\xi_r) + (1 - \frac{\varepsilon}{2})(\xi_r^TK\xi_r) \le$$

$$\frac{1}{2}(1 + \frac{\gamma_1^2}{2\varepsilon}\frac{h^2}{k^2})(D_t\xi_0)^TM(D_t\xi_0) + (1 + \frac{\varepsilon'}{2})(\xi_0^TK\xi_0)$$

$$+ \frac{1}{2} \sum_{j=1}^{r-1} h[\|f(\cdot, u(jh))\|^2 + \|D_{\bar{t}} M^{1/2} \xi_j\|^2 + \xi_j^T K \xi_j]. \tag{6.5.18}$$

If there exists a positive constant η such that for some $\varepsilon > 0$,

$$1 - \frac{\varepsilon}{2} \geq \eta \text{ and } 1 - \frac{\gamma_1^2}{2\varepsilon} \frac{h^2}{k^2} - \frac{h}{2} \geq \eta, \tag{6.5.19}$$

then we can apply the discrete Bellman-Gronwall Lemma (see Section 6.6) to (6.5.18) to obtain (6.5.20) for some $C < \infty$. Therefore if $1 - \frac{3h^2}{k^2} - \frac{h}{2} > 0$, then there exists such a $\eta > 0$ such that (6.5.19) is true and

$$\|M^{1/2} D_{\bar{t}} \xi_r\|^2 + \|K^{1/2} \xi_r\|^2 \leq C[\|M^{1/2} D_t \xi_0\|^2 + \|K^{1/2} \xi_0\|^2] + C \sum_{j=1}^{r-1} h\|f(\cdot, u(jh))\|^2. \tag{6.5.20}$$

■

Convergence

The purpose of this section is to show that the finite element solution $\sum_{i=1}^{2^q} \xi_{n,q}^i(\cdot, u)\psi_q^i(\cdot)$ converges in the L_2 and the H^1 norms to the true solution $w(\cdot,\cdot,u)$ as n and q become large. This is done in two parts: we introduce the interpolation of $w(jh,\cdot,u)$ in the H_q space for $j \in \{0,1, \cdots 2^{2n}\}$. We determine the error between $w(jh,\cdot,u)$ and its interpolation. Second, the error between the interpolation and the finite element solution is found. The sum of these two errors bounds the error between the finite element solution and the true solution.

Suppose $u \in G_n$ for $n \in \mathbb{Z}_+$. Let $w(\cdot,\cdot,u)$ denote the solution to (6.4.2) which has a nodal vector $\bar{w}_{n,q}(\cdot,u) \in \mathbb{R}^{(2^{2n}+1) \times 2^q}$ (i.e., $w(jh, i2^{-q}, u) = \bar{w}_{n,q}^i(j,u)$ for $i \in \{1,2,\dots 2^q\}$ and $j \in \{0,1,\dots 2^{2n}\}$). Again $k = 2^{-q}$ and $h = 2^{-2n}$.

To obtain the first part, we invoke the following lemma from Strang [Str.1]:

Lemma 6.5.3 [Strang]: There exists $c < \infty$ such that for all $u \in G$ such that if $w(\cdot,\cdot,u)$ is the classical solution to (6.4.2),

$$\| \sum_{i=1}^{2^q} \overline{w}_{n,q}(jh,u)\psi_q^i(\cdot) - w(jh,\cdot,u) \| \le ck^2, \tag{6.5.21}$$

$$\| \frac{\partial}{\partial x} \sum_{i=1}^{2^q} \overline{w}_{n,q}(jh,u)\psi_q^i(\cdot) - \frac{\partial}{\partial x} w(jh,\cdot,u) \| \le ck. \tag{6.5.22}$$

■

Since all $u \in G_n$, where G_n is defined in (6.5.3), are piecewise continuous, all $u \in G_n$ admit classical solutions to (6.4.2), see Pazy [Paz.1]. Since $u \in G \cap R_n$ is piecewise constant, a classical solution exits to (6.4.2). Temporal discretization has been done in the following way. For any $n \in \mathbb{Z}_+$, the time horizon $[0,1]$ is divided into 2^{2n} equal segments for the purpose of integration. However, for determining controls, the time horizon is divided into only 2^n equal segments. This means the control can change only after 2^n integration steps have been performed. We define $J_n \triangleq \{ 0,1, \cdots 2^{2n} \}$ and $K_n \triangleq \{ 0, 2^n, 2 \times 2^n, \cdots 2^{2n} \}$. For $j \in J_n$,

$$w((j+1)h,\cdot,u) = w(jh,\cdot,u) + h\dot{w}_+(jh,\cdot,u) + \frac{h^2}{2}\ddot{w}_+(jh,\cdot,u) + \frac{h^3}{6}\dddot{w}_+(jh,\cdot,u) \tag{6.5.23}$$
$$+ \frac{h^4}{24}\ddddot{w}_+((j+\theta_j)h,\cdot,u),$$

$$w((j-1)h,\cdot,u) = w(jh,\cdot,u) - h\dot{w}_-(jh,\cdot,u) + \frac{h^2}{2}\ddot{w}_-(jh,\cdot,u) - \frac{h^3}{6}\dddot{w}_-(jh,\cdot,u) \tag{6.5.24}$$
$$+ \frac{h^4}{24}\ddddot{w}_-((j-\overline{\theta}_j)h,\cdot,u),$$

with $\theta_j, \overline{\theta}_j \in [0,1]$. For $j \in J_n/K_n$, $\dot{w}_+(jh,\cdot,u) = \dot{w}_-(jh,\cdot,u)$, $\ddot{w}_+(jh,\cdot,u) = \ddot{w}_-(jh,\cdot,u)$, $\dddot{w}_+(jh,\cdot,u) = \dddot{w}_-(jh,\cdot,u)$, and $\ddddot{w}_+(jh,\cdot,u) = \ddddot{w}_-(jh,\cdot,u)$. Summing (6.5.23) and (6.5.24),

$$D_t D_{\bar{t}} w(jh,\cdot,u) = \ddot{w}(jh,\cdot,u) + \frac{h^2}{24}\rho(jh,\cdot,u), \tag{6.5.25}$$

where $\rho(jh,\cdot,u) \triangleq (\ddddot{w}((j+\theta_j)h,\cdot,u) - \ddddot{w}((j-\overline{\theta}_j)h,\cdot,u)$. Substituting (6.5.23), (6.5.24) into

(6.4.20) and neglecting damping ($\alpha = 0$)[2],

$$\langle D_t D_{\bar{t}} w(jh,\cdot,u),\phi(\cdot)\rangle + \langle \frac{\partial}{\partial x} w(jh,\cdot,u), \frac{\partial}{\partial x}\phi(\cdot)\rangle + \frac{h^2}{24}\langle \rho(jh,\cdot,u),\phi(\cdot)\rangle \tag{6.5.26}$$

$$= \langle f(\cdot,u(jh)),\phi(\cdot)\rangle$$

for all $\phi(\cdot) \in H_q$ and $j \in J_n/K_n$.

We state the identity:

$$\langle D_t D_{\bar{t}} \sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\psi^i_q(\cdot),\phi(\cdot)\rangle + \langle \frac{\partial}{\partial x}\sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\psi^i_q(\cdot),\phi(\cdot)\rangle \tag{6.5.27}$$

$$= \langle D_t D_{\bar{t}} \sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\psi^i_q(\cdot),\phi(\cdot)\rangle + \langle \frac{\partial}{\partial x}\sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\psi^i_q(\cdot),\phi(\cdot)\rangle.$$

Adding equations (6.5.27) and (6.5.26) and subtracting (6.5.7), we obtain:

$$\langle D_t D_{\bar{t}} \sum_{i=1}^{2^q} (\xi^i_{n,q}(j,u) - \overline{w}^i_{n,q}(j,u))\psi^i_q(\cdot),\phi(\cdot)\rangle + \langle \sum_{i=1}^{2^q} (\xi^i_{n,q}(j,u) - \overline{w}^i_{n,q}(j,u))\frac{\partial}{\partial x}\psi^i_q(\cdot), \frac{\partial}{\partial x}\phi(\cdot)\rangle$$

$$= \langle D_t D_{\bar{t}} (\sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\psi^i_q(\cdot) - w(jh,\cdot,u)),\phi(\cdot)\rangle$$

$$+ \langle \frac{\partial}{\partial x}\sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\psi^i_q(\cdot) - \frac{\partial}{\partial x} w(jh,\cdot,u), \frac{\partial}{\partial x}\phi(\cdot)\rangle$$

$$+ \frac{h^2}{24}\langle \rho(jh,\cdot,u),\phi(\cdot)\rangle \text{ for all } \phi(\cdot) \in H_q. \tag{6.5.28}$$

For $j \in K_n$, (6.5.26) does not hold. We subtract (6.5.7) from (6.5.27) yielding

$$\langle D_t D_{\bar{t}} \sum_{i=1}^{2^q} (\xi^i_{n,q}(j,u) - \overline{w}^i_{n,q}(j,u))\psi^i_q(\cdot),\phi(\cdot)\rangle + \langle \sum_{i=1}^{2^q} (\xi^i_{n,q}(j,u) - \overline{w}^i_{n,q}(j,u))\frac{\partial}{\partial x}\psi^i_q(\cdot), \frac{\partial}{\partial x}\phi(\cdot)\rangle$$

$$= \langle D_t D_{\bar{t}} \sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\psi^i_q(\cdot),\phi(\cdot)\rangle + \langle \sum_{i=1}^{2^q} \overline{w}^i_{n,q}(j,u)\frac{\partial}{\partial x}\psi^i_q(\cdot), \frac{\partial}{\partial x}\phi(\cdot)\rangle$$

[2]The case in which $\alpha > 0$ can be similarly handled.

$$- \langle f(\cdot, u(jh)), \phi(\cdot) \rangle \text{ for all } \phi(\cdot) \in H_q. \tag{6.5.29}$$

Define $e_j = e_{n,q}(j,u) \triangleq \overline{w}_{n,q}(j,u) - \xi_{n,q}(j,u)$, and set $\phi(\cdot) \triangleq \frac{1}{2}(D_t + D_{\bar{t}}) \sum_{i=1}^{2^q} e^i_{n,q}(j,u)\psi^i_q(\cdot)$ in equations (6.5.28) and (6.5.29). Multiplying (6.5.28) and (6.5.29) by h and summing from $j = 1$ to $r - 1$, and applying the Discrete Bellman-Gronwall Lemma ($c_i, i \in \{1, 2, \cdots\}$ are real, positive numbers less that infinity.),

$$\frac{1}{2}[1 - \frac{\gamma_1^2}{2\varepsilon}\frac{h^2}{k^2}](D_{\bar{t}}e_r)^T M D_{\bar{t}} e_r + (1 - \frac{\varepsilon}{2})e_r^T K e_r \le \tag{6.5.30}$$

$$\frac{1}{2}[1 + \frac{\gamma_1^2}{2\varepsilon}\frac{h^2}{k^2}](D_t e_0)^T M D_t e_0 + (1 - \frac{\varepsilon}{2})e_0^T K e_0$$

$$+ h\left[\sum_{i=1, i \neq j2^n, j \in \mathbf{Z}_+}^{r-1} c_1 k^4 + c_2 k^2 + c_3 h^4\right] + h\left[\sum_{i=1, i=j2^n, j \in \mathbf{Z}_+} c_4\right]$$

$$\le \frac{1}{2}[1 + \frac{\gamma_1^2}{2\varepsilon}\frac{h^2}{k^2}](D_t e_0)^T M D_t e_0 + (1 - \frac{\varepsilon}{2})e_0^T K e_0 + c_5(2^{-2q} + 2^{-n}).$$

Let $e_0 = \overline{w}_{n,q}(0,u) - \xi_{n,q}(0,u) = 0$. Since the forcing function is bounded, $|e_1| \le c_6 h$. Therefore, when the stability requirement (6.5.19) is met, (6.5.30) becomes

$$(D_{\bar{t}}e_r)^T M D_{\bar{t}} e_r + e_r^T K e_r \le c_7(2^{-2q} + 2^{-n}). \tag{6.5.31}$$

Therefore,

$$\left\| \sum_{i=1}^{2^q} e^i_{n,q}(j,u)\frac{\partial}{\partial x}\psi^i_q(\cdot)\right\|^2 = e_j^T K e_j \le c_7(2^{-2q} + 2^{-n}). \tag{6.5.32}$$

By Friedrich's inequality, $\|\phi(\cdot)\| \le c_8 \|\frac{\partial}{\partial x}\phi(\cdot)\|$ for all $\phi(\cdot) \in H^*$,

$$\left\| \sum_{i=1}^{2^q} e^i_{n,q}(j,u)\psi^i_q(\cdot)\right\| = \le c_9(2^{-q} + 2^{-n/2}). \tag{6.5.33}$$

We define $g_{n,q}(u)$ to be an approximation to the potential energy at the final time, $g_{n,q}(u) \triangleq \frac{1}{2}\| \sum_{i=1}^{2^q} \xi^i_{n,q}(2^{2n},u)\frac{\partial}{\partial x}\psi^i_q(\cdot)\|^2$. Applying Lemma 6.5.3, we obtain the following theorem:

Theorem 6.5.4: There exits $c < \infty$ such that for all $n \in \mathbb{Z}_+$ and all $u \in G_n$,

$$(1)\quad |g(u) - g_{n,q}(u)| \le c(h^{1/2} + k^2) = c(2^{-n} + 2^{-2q}), \tag{6.5.34}$$

$$(2)\quad \|\frac{\partial}{\partial x}w(jh,\cdot,u) - \sum_{i=1}^{2^q} \xi^i_{n,q}(j,u)\frac{\partial}{\partial x}\psi^i_q(\cdot)\| \le c(h^{1/4} + k) = c(2^{-n/2} + 2^{-2q}). \tag{6.5.35}$$

Corollary 6.5.5: Theorem 6.5.4 is also true for any damping $\alpha \ge 0$. ■

We define an approximation to $\delta w(jh,\cdot,u,\delta u)$ to be $\sum_{i=1}^{2^q} \delta\xi_{n,q}(j,u,\delta u)\psi^i_q(\cdot)$ where $\delta\xi_{n,q}(\cdot,u,\delta u)$ is the solution to

$$M_qD_lD_{\bar{l}}\delta\xi_{n,q}(j,u,\delta u) + \frac{\alpha}{2}K_q(D_l + D_{\bar{l}})\delta\xi_{n,q}(j,u,\delta u) + K_q\delta\xi_{n,q}(j,u,\delta u) \tag{6.5.36}$$

$$= \frac{\partial}{\partial u}f_{n,q}(j,u)\delta u.$$

We define an approximation to the directional derivative of g(u) to be

$$dg_{n,q}(u;\delta u) = \langle \sum_{i=1}^{2^q} \xi^i_{n,q}(2^{2n},u)\frac{\partial}{\partial x}\psi^i_q(\cdot), \sum_{i=1}^{2^q} \delta\xi^i_{n,q}(2^{2n},u,\delta u)\frac{\partial}{\partial x}\psi^i_q(\cdot) \rangle. \tag{6.5.37}$$

Since $w(\cdot,\cdot,u)$ and $\delta w(\cdot,\cdot,u,\delta u)$ solve similar partial differential equations, and $\xi_{n,q}(\cdot,u)$ and $\delta\xi_{n,q}(\cdot,u,\delta u)$ solve similar finite difference equations, a result similar to Theorem 6.5.4 applies.

Theorem 6.5.6: There exists $c' \in [c,\infty)$ such that for all $n \in \mathbb{Z}_+$, $u \in G_n$ and $\delta u \in G_n - u$,

$$(1)\quad \|\frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u) - \sum_{i=1}^{2^q} \delta\xi^i_{n,q}(2^{2n},u,\delta u)\frac{\partial}{\partial x}\psi^i_q(\cdot)\| \le c'(2^{-n/2} + 2^{-q}), \tag{6.5.38}$$

(2) $|dg(u;\delta u) - dg_{n,q}(u;\delta u)| < c'(2^{-n/2} + 2^{-q}).$ (6.5.39)

Proof: The proof of (1) is similar to that of Theorem 6.5.4, part 2. We must show how $dg_{n,q}(u;\delta u)$ approximates $dg(u;\delta u)$:

$$|dg(u;\delta u) - dg_{n,q}(u;\delta u)| = \langle \frac{\partial}{\partial x}w(1,\cdot,u), \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u) \rangle$$

$$- \langle \sum_{i=1}^{2^q} \xi^i_{n,q}(2^{2n},u)\frac{\partial}{\partial x}\psi^i_q(\cdot), \sum_{i=1}^{2^q} \delta\xi^i_{n,q}(2^{2n},u,\delta u)\frac{\partial}{\partial x}\psi^i_q(\cdot) \rangle$$

$$\leq \| \frac{\partial}{\partial x}w(1,\cdot,u) - \sum_{i=1}^{2^q} \xi^i_{n,q}(2^{2n},u)\frac{\partial}{\partial x}\psi^i_q(\cdot)\| \| \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u)\|$$

$$+ \| \sum_{i=1}^{2^q} \xi^i_{n,q}(2^{2n},u)\frac{\partial}{\partial x}\psi^i_q(\cdot)\| \| \frac{\partial}{\partial x}\delta w(1,\cdot,u,\delta u) - \sum_{i=1}^{2^q} \delta\xi^i_{n,q}(1,u,\delta u)\frac{\partial}{\partial x}\psi^i_q(\cdot)\|$$

$$\leq c'(2^{-n/2} + 2^{-q}). \tag{6.5.40}$$

■

If we choose $q = n$ in the above equations, then all assumptions in Section 6.2 are satisfied and Theorem 6.3.6 is true.

Summary

We have presented an example problem to illustrate how the assumptions of Section 6.2 can be satisfied. In this example, we chose the spacial and temporal discretization to be the same and the control to be piecewise constant with $\sqrt{n}$ discontinuities where n is the number of time intervals. Fujii suggests that the restriction between space and time discretizations (6.5.19) can be removed if Newmark's β integration scheme with $\beta \geq \frac{1}{4}$ is used. Furthermore, from experimental results (see Chapter 8) it appears that the piecewise constant controls with n discontinuities can be used without problems.

6.6 APPENDIX FOR CHAPTER 6

Bellman-Gronwall Lemma:

Suppose $c \geq 0$, $r(\cdot) \geq 0$ and $k(\cdot) \geq 0$.

If $r(t) \leq c + \int_0^t k(s)r(s)ds$ for all $t \in [0,T]$, then $r(t) \leq ce^{[\int_0^t k(s)ds]}$ for all $t \in [0,T]$.

Discrete Bellman-Gronwall Lemma: Suppose $c \geq 0$ and $k_i \geq 0$.

If $r_N \leq c + \sum_{i=1}^{N-1} k_i r_i$, then $r_N \leq \prod_{i=1}^{N-1} (1 + k_i)c$.

Proof: Let $s_N \triangleq c + \sum_{i=1}^{N-1} k_i r_i$. Then $r_N \leq s_N$ and $s_{N+1} - s_N = k_N r_N$.

$$s_{N+1} = k_N r_N + s_N \leq (1 + k_N)s_N.$$

$$r_N \leq s_N \prod_{i=1}^{N-1} (1 + k_i)s_1 = \prod_{i=1}^{N-1} (1 + k_i)c.$$

■

CHAPTER 7

ADVANCED ALGORITHMS

In chapter 3, we presented a conceptual algorithm to solve fixed-time problems with control constraints. In Chapter 4, we showed that by using a specific time scaling, free-time problems can be transcribed into fixed-time problems. In Chapter 6, this conceptual algorithm was transcribed into an implementable algorithm by introducing: (1) a scheme to discretize the PDE spatially into an ODE; (2) a scheme consistent with (1) to discretize the ODE into a finite difference equation; (3) a rule to determine when and how to refine the spatial and temporal discretizations.

Chapter 7 provides an extension of the Polak-Trahan-Mayne algorithm to solve both fixed-time and free-time problems with control, terminal state and state-space constraints. In Section 7.1, we derive a formulation for these problems. Section 7.2 presents an algorithm to solve the resulting problems. With the PDE discretized to a finite difference equation, the algorithm can be implemented. Section 7.3 presents two subprocedures to find a search direction for the algorithm in Section 7.2.

7.1 FORMULATION

In Chapter 3, we discussed a *conceptual algorithm* to solve the fixed-time problem with hard control constraints:

$$\mathbf{P}: \ \inf \{ \ g(u) \mid u \in G \ \}. \tag{7.1.1}$$

In Chapter 4, this was extended to the free-time problem

$$\inf_{u,\tau}\{ \ \bar{g}(\bar{u},\tau) \mid \bar{u} \in G(\tau), \tau \in [\tau_{\min}, \tau_{\max}] \ \}. \tag{7.1.2}$$

The problems encountered in Chapter 2 are more complicated:

$$\inf_{u,\tau}\{\ \bar{g}(\bar{u},\tau)\mid \bar{g}^j(\bar{u},\tau)\le 0, j\in \underline{p-1}, \bar{\phi}(\bar{u},\tau,t)\le 0, t\in[0,\tau], \quad (7.1.3)$$
$$\bar{u}\in G(\tau), \tau\in[\tau_{\min},\tau_{\max}]\ \},$$

where $\bar{g}^j(\bar{u},\tau) = h^j(\tau,\bar{x}(1,\bar{u}))$, $\bar{\phi}(\bar{u},\tau,t) = h^p(\tau,\bar{x}(t,\bar{u}))$, for all $t\in[0,\tau]$ and $h^j : \mathbb{R}\times X\to\mathbb{R}$ are continuously differentiable for all $j\in\{\ 0,\ \cdots,p\ \}$, and $\bar{x}(\cdot,\bar{u})$ is the solution to (4.1.10).

By performing the time scaling as in Section 4.2, we obtain the problem:

$$\inf_{u,\tau}\{\ g(u,\tau)\mid g^j(u,\tau)\le 0, j\in \underline{p-1}, \phi(u,\tau,t)\le 0, t\in[0,\tau], \quad (7.1.4)$$
$$u\in G, \tau\in[\tau_{\min},\tau_{\max}]\ \},$$

where $g^j(u,\tau) = h^j(\tau,\hat{x}(1,u,\tau))$, $j\in\{\ 0,\ \cdots,p-1\ \}$, $\phi(u,\tau,t) = h^p(\tau,\hat{x}(t,u,\tau))$, with the dynamics from Section 4.2:

$$\frac{d}{dt}\hat{x}(t,u,\tau) = \tau[A\hat{x}(t,u,\tau) + F(\hat{x}(t,u,\tau),u(t))],\ t\in[0,1],\ \hat{x}(0,u,\tau) = \hat{x}_0. \quad (7.1.5)$$

The functions g^j are dependent on both the state and a parameter τ. To simplify exposition, we define an additional state whose value is τ for all time. We can therefore transform (7.1.4) to a new problem in which the function g^j is dependent only on the state.

We define $x^0(\cdot,u,\tau)\in C([0,1])$ so that $x^0(t,u,\tau) = \tau$ for $t\in[0,1]$. Then,

$$\frac{d}{dt}\begin{bmatrix}\hat{x}(t,u,\tau)\\ x^0(t,u,\tau)\end{bmatrix} = \begin{bmatrix}x^0(t,u,\tau)\hat{f}(\hat{x}(t,u,\tau),u(t))\\ 0\end{bmatrix},\ t\in[0,1],\ \begin{bmatrix}\hat{x}(0,u,\tau)\\ x^0(0,u,\tau)\end{bmatrix} = \begin{bmatrix}\hat{x}_0\\ \tau\end{bmatrix}. \quad (7.1.7)$$

Define $x(t,u,\tau) \triangleq \begin{bmatrix}\hat{x}(t,u,\tau)\\ x^0(t,u,\tau)\end{bmatrix}\in X\times\mathbb{R}$.

By defining $f(x(t,u,\tau),u(t)) = \begin{bmatrix}x^0(t,u,\tau)\hat{f}(\hat{x}(t,u,\tau),u(t))\\ 0\end{bmatrix}$, (7.1.7) becomes:

$$\frac{d}{dt}x(t,u,\tau) = f(x(t,u,\tau),u(t)), \quad x(0,u,\tau) = \begin{bmatrix} \hat{x}_0 \\ \tau \end{bmatrix}. \tag{7.1.8}$$

Hence (7.1.4) becomes

$$\inf_{u,w}\{\, g^0(u,w) \mid g^j(u,w) \le 0, j \in \underline{p-1}, \phi(u,w,t) \le 0,\ t \in [0,1],\ u \in G,\ w \in C \,\}, \tag{7.1.9}$$

where $g^j(u,w) = h^j(x(1,u,w)),\ j \in \{\, 0,1,...p-1 \,\}$ and $\phi(u,w,t) = h^p(x(t,u,w)),\ t \in [0,1]$, with $h^j : X \times \mathbb{R} \to \mathbb{R}$ being continuously differentiable for $j \in \{\, 0,1,\cdots,p \,\}$. $C \triangleq \hat{x}_0 \times [\tau_{\min}, \tau_{\max}] \subset X \times \mathbb{R}$ and $x(\cdot,u,w) \in C([0,1]; X \times \mathbb{R})$ satisfies

$$\frac{d}{dt}x(t,u,w) = f(x(t,u,w),u(t)), \ x(0,u,w) = w. \tag{7.1.10}$$

In what follows, we will generalize to the case in which C is any compact, convex subset of $X \times \mathbb{R}$. The functions $h^j(\cdot)$ in (7.1.9) are different than those in (7.1.3).

We close this section with a discussion of state-space constraints ($\phi(u,w,t) \le 0, t \in [0,1]$). The algorithm presented in Section 7.2 does not solve problems with state-space constraints explicitly. However, if we define

$$g^p(u,w) \triangleq \int_0^1 [\max\{\, \phi(u,w,t), 0 \,\}]^2 dt. \tag{7.1.11}$$

Then $g^p(u,w) = 0$ if and only if $\phi(u,w,t) \le 0$ for all $t \in [0,1]$. If we append a new state, $\bar{x}(\cdot,u,w) \in C([0,1])$ such that

$$\frac{d}{dt}\bar{x}(t,w,u) = \max\{\, h^p(y(t,u,w)), 0 \,\}^2, \quad \bar{x}(0,u,w) = 0, \tag{7.1.12}$$

we obtain a new system:

$$\frac{d}{dt}\begin{bmatrix} x(t,u,w) \\ \bar{x}(t,u,w) \end{bmatrix} = \begin{bmatrix} f(x(t,u,w),u(t)) \\ \max\{ h^p(x(t,u,w)),0 \}^2 \end{bmatrix}, \quad \begin{bmatrix} x(0,u,w) \\ \bar{x}(0,u,w) \end{bmatrix} = \begin{bmatrix} w \\ 0 \end{bmatrix}. \tag{7.1.13}$$

With this notation, (7.1.9) is transcribed into (7.1.14) with no explicit state-space constraints.

$$\mathbf{P}_{\mathrm{II}} : \inf_{u,w}\{ g^0(u,w) \mid g^j(u,w) \le 0, j \in \underline{p}, u \in G, w \in C \}, \tag{7.1.14}$$

where $g^j(u,w) = h^j(x(1,u,w))$, $h^j \in \{ 1,2,...p-1 \}$, $g^p(u,w) = \bar{x}(1,u,w)$.

Note of Warning: There may be inherent poor conditioning associated with this transcription. This shall be discussed in Section 7.2

In Sections 7.2 and 7.3 we present algorithms to solve (7.1.14). It is clear that any such algorithms are *conceptual* ; they still require exact solution of an partial differential equation. However, by applying a discretization scheme with a consistent refinement rule, we obtain an implementation.

7.2 POLAK-TRAHAN-MAYNE SEARCH DIRECTION

In 1979, Polak, Trahan, and Mayne [Pol.5] proposed a method based on earlier work of Polak and Pirroneau [Pir.1, Pir.2] to solve the problem:

$$\min_{x \in \mathbb{R}^n} \{ g^0(x) \mid g^j(x) \le 0, j \in \underline{p} \}, \tag{7.2.1}$$

with $g^j : \mathbb{R}^n \to \mathbb{R}$ being continuously differentiable. We define

$$\psi(x) \triangleq \max\{ g^j(x) \mid j \in \underline{p} \} \tag{7.2.2}$$

and

$$\psi_+ \triangleq \max\{ \psi(x), 0 \}. \tag{7.2.3}$$

The optimality function and search direction are:

$$\tilde{\Theta}(x) = \min_h \{ \frac{1}{2}\|h\|^2 + \tag{7.2.4}$$

$$\max_{j \in \underline{p}} \{ -\rho\psi_+(x) + \langle \nabla g^0(x), h \rangle ; g^j(x) + \langle \nabla g^j(x), h \rangle - \psi_+(x) \} \}.$$

$$\tilde{h}(x) = \arg\min_h \{ \frac{1}{2}\|h\|^2 + \tag{7.2.5}$$

$$\max_{j \in \underline{p}} \{ -\rho\psi_+(x) + \langle \nabla g^0(x), h \rangle ; g^j(x) + \langle \nabla g^j(x), h \rangle - \psi_+(x) \} \}.$$

The Polak-Trahan-Mayne Algorithm consists of using (7.2.5) to obtain the search direction, $\tilde{h}(x)$, and using the Armijo rule to determine a step-size. By taking $h = 0$ in (7.2.4), it is easily seen that $\tilde{\Theta}(x) \le 0$ and that $\tilde{\Theta}(x) = 0$ is a necessary condition of optimality (see Theorem 7.2.2). The positive number ρ is a scaling factor. For x infeasible, $\psi(x) > 0$, a larger value of ρ makes the term $-\rho\psi_+(x) + \langle \nabla g^0(x), h \rangle$ in (7.2.4) more negative for any h. Consequently, the minimizing h in (7.2.4) is "less concerned" with $\langle \nabla g^0(x), h \rangle$ being a large negative number than with smaller ρ. In the case $\rho = \infty$ and x is infeasible, (7.2.4) reduces to finding a search direction based on minimizing a convex approximation to $\psi(x + h)$.

We present an algorithm which is an extension of the Polak-Trahan-Mayne algorithm and the Polak-Mayne [Pol.2] algorithm for optimal control. We consider the case in which the explicit state-space constraints have been removed by way of the transcription in Section 7.1. We define an optimality function:

$$\Theta(u, w) \triangleq \min\{ \phi(u, w, v, z) \mid v \in G, z \in C \}, \tag{7.2.6}$$

where

$$\phi(u, w, v, z) \triangleq \frac{1}{2}\|v - u\|^2 + \frac{1}{2}\|z - w\|^2 + \max_{j \in \underline{p}} \{ -\rho\psi_+(u, w) + \langle \nabla g^0(u, w), \begin{bmatrix} v - u \\ z - w \end{bmatrix} \rangle ;$$

$$g^j(u, w) + \langle \nabla g^j(u, w), \begin{bmatrix} v - u \\ z - w \end{bmatrix} \rangle - \psi_+(u, w) \} , \tag{7.2.7}$$

where

$$\psi(u,w) \triangleq \max_{j \in \underline{p}}\{ g^j(u,w) \}, \quad \psi_+(u,w) \triangleq \max\{ 0, \psi(u,w) \}, \tag{7.2.8}$$

$$\begin{bmatrix} v(u,w) \\ z(u,w) \end{bmatrix} \triangleq \arg \Theta(u,w), \tag{7.2.9}$$

i.e., $\begin{bmatrix} v(u,w) \\ z(u,w) \end{bmatrix}$ is the unique minimizer of (7.2.6).

We present the extended Algorithm.

Algorithm 7.2.1:

Data: $u_0 \in G$, $w_0 \in C$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1]$, $\rho > 0$.

Step 0: $i = 0$.

Step 1: Calculate $\nabla g^j(u_i)$ for $j \in \{ 0, 1, \cdots, p \}$.

Step 2: Calculate $v_i \in G$ and $z_i \in C$ such that $\phi(u_i, w_i, v_i, z_i) \leq \gamma\Theta(u_i, w_i)$.

Step 3: If $\Theta(u_i, w_i) = 0$, $\lambda_i \triangleq 0$.

Else if $\psi(u_i, w_i) > 0$ (Phase I),

$\lambda_i \triangleq \max\{ \lambda \in \{ 1, \beta, \beta^2, \cdots \} \mid$

$$\psi(u_i + \lambda(v_i - u_i),\ w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) < \alpha\lambda\phi(u_i, w_i, v_i, z_i)$$

Else ($\psi(u_i, w_i) \leq 0$, (Phase II)

$\lambda_i \triangleq \max\{ \lambda \in \{ 1, \beta, \beta^2, \cdots \} \mid$

$g^0(u_i + \lambda(v_i - u_i),\quad w_i + \lambda(z_i - w_i)) - g^0(u_i, w_i) \quad < \alpha\lambda\phi(u_i, w_i, v_i, z_i)$ and

$\psi(u_i + \lambda(v_i - u_i),\ w_i + \lambda(z_i - w_i)) \leq 0$.

Step 4: Set $u_{i+1} = u_i + \lambda_i(v_i - u_i)$, $w_{i+1} = w_i + \lambda_i(z_i - w_i)$.

Step 5: Set $i = i + 1$; go to Step 1. ■

We now show that with $\Theta(u,w) = 0$ is a necessary condition for optimality for P_{II} (7.1.14).

Theorem 7.2.2: For $u \in C$ and $w \in G$, if $\Theta(u,w) = -\delta < 0$, then (u,w) is not a minimizer of (7.1.9) and Algorithm 7.2.1 will not jam up at (u,w).

Proof: Assume that $(u,w) \in G \times C$ is such that $\Theta(u,w) = -\delta < 0$. Then there exists $(v,z) \in G \times C$ such that $\phi(u,w,v,z) \le -\gamma\delta$ where $\gamma \in (0,1]$ is datum in Algorithm 7.2.1. We consider two cases:

Case I: $\psi(u,w) > 0$:

Since $\phi(u,w,v,z) \le -\delta\gamma$,

$$g^j(u,w) + \langle \nabla g^j(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle - \psi(u,w) < \phi(v,z,u,w) \le -\delta\gamma,\ j \in \underline{p}. \tag{7.2.10}$$

Let

$$J(u,w) \triangleq \{\ j \in \underline{p} \mid g^j(u,w) = \psi(u,w)\ \} \tag{7.2.11}$$

be the set of active constraints. For $j \in J(u,w)$, $\langle \nabla g^j(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle < -\delta\gamma$. Since $g^j(\cdot,\cdot)$ is continuously differentiable, it can be seen as in Theorem 3.5.2 that there exists $\lambda_j > 0$ such that

$$g^j(u+\lambda(v-u), w+\lambda(z-w)) - g^j(u,w) \le \alpha\lambda\phi(v,z,u,w) \tag{7.2.12}$$

for $\lambda \in [0,\lambda_j]$. Let $\tilde{\lambda} \triangleq \min\{\ 1, \lambda_j \mid j \in J(u,w)\ \}$. Let $\varepsilon \triangleq \min\{\ \psi(u,w) - g^j(u,w) \mid j \notin J(u,w)\ \}$. Then $\varepsilon > 0$ and since $g^j(\cdot,\cdot)$ is continuous for $j \in \underline{p}$, there exists $\bar{\lambda} \in (0,\tilde{\lambda}]$ such that

$$g^j(u+\lambda(v-u), w+\lambda(z-w)) - \psi(u,w) \le \alpha\lambda\phi(v,z,u,w),\ \lambda \in (0,\bar{\lambda}] \tag{7.2.13}$$

for $j \in \underline{p}$, and therefore

$$\psi(u + \lambda(v - u), w + \lambda(z - w)) - \psi(u, w) < \alpha\lambda\phi(v,z,u,w) < -\alpha\lambda\delta\gamma < 0, \tag{7.2.14}$$

and $(u + \lambda(v - u), w + \lambda(z - w)) \in G \times C$ so (u, w) is not optimal, and Algorithm 7.2.1 will not jam up at (u, w).

Case II: $\psi(u, w) \leq 0$:

$$\langle \nabla g^0(u, w), \begin{bmatrix} v - u \\ z - w \end{bmatrix} \rangle \leq \phi(v,z,u,w), \tag{7.2.15}$$

$$g^j(u, w) + \langle \nabla g^j(u, w), \begin{bmatrix} v - u \\ z - w \end{bmatrix} \rangle \leq \phi(v,z,u,w), \; j \in \underline{p}. \tag{7.2.16}$$

Let

$$J_0(u, w) \triangleq \{ j \in \underline{p} \mid g^j(u, w) = 0 \} \cup \{ 0 \}. \tag{7.2.17}$$

For $j \in J_0(u, w)$, there exists $\lambda_j > 0$ such that

$$g^j(u + \lambda(v - u), w + \lambda(z - w)) - g^j(u, w) < \alpha\lambda\phi(v,z,u,w), \; \lambda \in [0, \lambda_j]. \tag{7.2.18}$$

Let $\tilde{\lambda} \triangleq \min\{ 1, \lambda_j \mid \; j \in J_0(u, w) \} \}$. Let $\varepsilon \triangleq \min\{ \psi(u, w) - g^j(u, w) \mid \; j \notin J_0(u, w) \}$. Since $g^j(\cdot, \cdot)$ is continuous and ε is strictly positive, there exists $\bar{\lambda} \in (0, \tilde{\lambda})$ such that for all $\lambda \in [0, \bar{\lambda}]$:

$$g^0(u + \lambda(v - u), w + \lambda(z - w)) - g^0(u, w) < \alpha\lambda\phi(v,z,u,w), \tag{7.2.19}$$

$$g^j(u + \lambda(v - u), w + \lambda(z - w)) < \alpha\lambda\phi(v,z,u,w), \quad \text{for all } j \in J_0(u, w)/\{ 0 \}, \tag{7.2.20}$$

$$g^j(u + \lambda(v - u), w + \lambda(z - w)) < 0, \quad \text{for all } j \notin J_0(u, w), \tag{7.2.21}$$

and $(u + \lambda(v - w), w + \lambda(z - w)) \in G \times C$ so (u, w) is not optimal and Algorithm 7.2.1 will not jam up at (u, w). ■

The requirement $\Theta(u, w) = 0$ is not a strong optimality condition for problem $\mathbf{P}_{II}$ (7.1.14). Suppose that at (u, w), all constraints are satisfied, but that (u, w) is not optimal. Since the state-space constraint is satisfied, $g^p(u, w) = 0$ and $\nabla g^p(u, w) = 0$. Since all constraints are satisfied, $g^j(u, w) \leq 0$ for $j \in \{ 1, 2, \cdots, p - 1 \}$, $\psi(u, w) = 0$. Consequently, (7.2.6) and (7.2.7) become

$$\Theta(u,w) = \min\{\ \phi(u,w,v,z) \mid v \in G, z \in C\ \}, \tag{7.2.22}$$

$$\phi(u,w,v,z) = \frac{1}{2}|v-u|^2 + \frac{1}{2}|z-w|^2 + \tag{7.2.23}$$

$$\max\{\ \langle \nabla g^0(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle ; g^j(u,w) + \langle \nabla g^j(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle, j \in \{1, \cdots, p-1\}; 0\ \}$$

because $g^p(u,w) + \langle \nabla g^p(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle = 0$ for all $v \in G$, $z \in C$. Hence, $\Theta(u,w) = 0$ and $\begin{bmatrix} v(u,w) \\ z(u,w) \end{bmatrix} = 0$. Consequently, as soon as all constraints become satisfied, Algorithm 7.2.1 will stop.

Even in the case that there are no state-space constraints, if for some $(u,w) \in G \times C$, all the constraints are satisfied and one of the constraints is of the form: $g^j(u,w) = x(1,u,w)^T Q x(1,u,w)$, such as the terminal potential energy constraint in (8.2.8), $\Theta(u,w) = 0$.

There are several solutions to this problem:

(1) Use a penalty function method to solve the problem. It is not possible to use exact penalty functions for this problem since $\nabla g^j(u,w) = 0$ when $g^j(u,w) = 0$ for some $j \in \underline{p}$.

(2) Guarantee that for each iterate some constraint is not satisfied. (Case II in Theorem 7.2.5)

(3) Introduce a tolerance vector $\varepsilon \in \mathbb{R}^p$, $\varepsilon > 0$, such that each constraint $g^j(u,w) \le 0$ becomes $g^j(u,w) - \varepsilon_j \le 0$. This allows all of the new constraints to be met, and as long as $\nabla g^j(u,w) \neq 0$ when $\psi(u,w) \le 0$, the algorithm will not jam at (u,w).

(4) Introduce a vector $\varepsilon \in \mathbb{R}^p$ as in (3) and apply an exact penalty algorithm to solve the perturbed problem.

Method (3) has been used successfully for solving the problems involving a flexible, rotating beam, $\mathbf{P}_1$ - $\mathbf{P}_4$, in Chapters 2 and 8. Because of the physics of the problem, at the optimal point (u, w), all constraints are active (i.e., $g^j(\hat{u}, \hat{w}) = \varepsilon_j$ and $\nabla g^j(\hat{u}, \hat{w}) \neq 0$, $j \in \underline{p}$ }.) For numerical results, see Chapter 8.

We now proceed to show convergence of Algorithm 7.2.1.

Lemma 7.2.3: The functions $v(\cdot,\cdot)$, $z(\cdot,\cdot)$, and $\Theta(\cdot,\cdot)$ are continuous.

Proof: Follow proof of Lemma 3.5.4 and Corollary 3.5.5. ■

Theorem 7.2.5: If $\{ (u_i, w_i) \} \subset G \times C$ is the sequence generated by Algorithm 7.2.1, then any accumulation point $(\hat{u}, \hat{w}) \in G \times C$ satisfies the optimality condition $\Theta(\hat{u}, \hat{w}) = 0$.

Proof: Assume that Algorithm 7.2.1 generates a sequence $\{ (u_i, w_i) \} \subset G \times C$ such that there exists a subsequence $K \subset \mathbb{Z}_+$ and an accumulation point $(\hat{u}, \hat{w}) \subset G \times C$ such that $u_i \to \hat{u}$ and $w_i \to \hat{w}$. on K, and $\Theta(\hat{u}, \hat{w}) = -\delta < 0$. By continuity of $\Theta(\cdot,\cdot)$, there exists i_0 such that for all $i \geq i_0$, $\phi(u_i, w_i, v_i, z_i) \leq \gamma \Theta(u_i, w_i) \leq \frac{\gamma}{2} \Theta(\hat{u}, \hat{w}) \leq -\frac{\delta\gamma}{2}$.

For each $j \in \{ 1, 2, \cdots, p \}$ and each subsequence $K_j \subset \mathbb{Z}_+$, we examine two possible cases:

(a) $\lim_{i \in K_j} g^j(u_i, w_i) - \psi(u_i, w_i) = 0$.

First, there exists $\tilde{i}_j$ such that for all $i \geq \tilde{i}_j$, $i \in K_j$,

$$\langle \nabla g^j(u_i), \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle \leq -\frac{\delta\gamma}{4}. \qquad (7.2.24)$$

Second, for all i, j,

$$g^j(u_i, w_i) - \psi(u_i, w_i) + \langle \nabla g(u_i, w_i), \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle \leq \phi(u_i, w_i, v_i, z_i) \leq -\frac{\delta\gamma}{2}. \qquad (7.2.25)$$

There exists $\bar{i}_j \geq \tilde{i}_j$ such that for all $i \geq \bar{i}_j$, $i \in K_j$,

$$\frac{1}{2}(1-\alpha)\phi(u_i, w_i, v_i, z_i) \leq -\frac{1}{4}(1-\alpha)\delta\gamma \leq g^j(u_i, w_i) - \psi(u_i, w_i). \qquad (7.2.26)$$

Consequently,

$$\langle \nabla g(u_i, w_i, \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle \leq \phi(u_i, w_i, v_i, z_i)[1 - \frac{1}{2}(1-\alpha)] = \phi(u_i, w_i, v_i, z_i)(\frac{1+\alpha}{2}) \qquad (7.2.27)$$

Using the first order expansion for g^j:

$$g^j(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) - \frac{2\alpha}{1+\alpha}\lambda \langle \nabla g^j(u_i, w_i), \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle$$

$$= g^j(u_i, w_i) - \psi(u_i, w_i) + \lambda(\frac{1-\alpha}{1+\alpha}) \langle \nabla g(u_i, w_i, \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle$$

$$+ \lambda \int_0^1 \langle \nabla g^j(u_i + s\lambda(v_i - u_i), w_i + s\lambda(z_i - w_i)) - \nabla g^j(u_i, w_i), \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle ds. \qquad (7.2.28a)$$

Combining (7.2.27) and (7.2.28a), and noting that $g^j(u_i, w_i) - \psi(u_i, w_i) \leq 0$, we obtain

$$g^j(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) - \frac{2\alpha}{1+\alpha}\lambda \langle \nabla g^j(u_i, w_i), \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle$$

$$\leq \lambda \Big[(\frac{1-\alpha}{1+\alpha})(\frac{1+\alpha}{2})(-\frac{\delta\gamma}{2})$$

$$+ \int_0^1 \langle \nabla g^j(u_i + s\lambda(v_i - u_i), w_i + s\lambda(z_i - w_i)) - \nabla g^j(u_i, w_i), \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle ds. \qquad (7.2.28b)$$

By continuity of $\nabla g^j(\cdot)$, there exists $\lambda_j > 0$ and $i_j \geq \bar{i}_j$ such that for all $\lambda \in [0, \lambda_j]$ and all $i > i_j$, $i \in K_j$, the right hand side of (7.2.28b) is less than or equal to zero. And so for all $i \geq i_j$, $i \in K_j$, and all $\lambda \in [0, \lambda_j]$,

$$g^j(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) \leq \alpha\lambda\phi(u_i, w_i, v_i, z_i). \qquad (7.2.29)$$

(b) There exists $\varepsilon > 0$ and a sequence $\bar{K}_j \subset K_j$ such that for all $i \in \bar{K}_j$,

$g^j(u_i, w_i) - \psi(u_i, w_i) \le -\varepsilon$. Since,

$$g^j(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) = g^j(u_i, w_i) - \psi(u_i, w_i) + \lambda\int_0^1 \langle \nabla g^j(u_i + s\lambda(v_i - u_i), w_i + s\lambda(z_i - w_i)), \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle ds, \tag{7.2.30}$$

There exists $\tilde{\lambda}_j$ such that for all $\lambda \in [0, \tilde{\lambda}_j]$, and all $i \in \overline{K}_j$,

$$|\lambda\int_0^1 \langle \nabla g(u_i + s\lambda(v_i - u_i), w_i + s\lambda(z_i - w_i)) \begin{bmatrix} v(u_i, w_i) \\ z(u_i, w_i) \end{bmatrix} \rangle ds| \le \frac{\varepsilon}{2} \tag{7.2.31}$$

and so

$$g^j(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) \le -\frac{\varepsilon}{2}. \tag{7.2.32}$$

Choose $\lambda_j \in (0, \tilde{\lambda}_j)$ such that $-\frac{\varepsilon}{2} \le \alpha\lambda_j\phi(u_i, w_i, v_i, z_i)$ for all $i \in \overline{K}_j$ and so for all $\lambda \in [0, \lambda_j]$,

$$g^j(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) \le \alpha\lambda\phi(u_i, w_i, v_i, z_i). \tag{7.2.33}$$

Similarly, we can show that for any subsequence $K_0 \subset \mathbb{Z}_+$ such that $\lim_{K_0} \psi_+(u_i, w_i) = 0$, then there exists i_0 and $\lambda_0 > 0$ such that for all $i \ge i_0$, $i \in K_0$ and all $\lambda \in [0, \lambda_0]$,

$$g^0(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - g^0(u_i, w_i) \le \alpha\lambda\phi(u_i, w_i, v_i, z_i). \tag{7.2.34}$$

We now consider three cases:

Case I ($\psi(\hat{u}, \hat{w}) > 0$)

We construct the proof using a step by step procedure:

Step 0: Set $j = 1$, $K_1 = \mathbb{Z}_+$, $\lambda_1^* = 1$, $i_1^* = 0$.

Step 1: If $\lim_{i \in K_j} g^j(u_i, w_i) - \psi(u_i, w_i) = 0$, set $\lambda_{j+1}^* = \min\{ \lambda_j^*, \lambda_j \}$ and $i_{j+1}^* = \min\{ i_j^*, i_j \}$ where λ_j and i_j are determined in case (a) above; Set $K_{j+1} = K_j$, go to Step 3.

Step 2: Set $K_{j+1} = \overline{K}_j$. Set $\lambda^*_{j+1} = \min\{ \lambda^*_j , \lambda_j \}$ and $i^*_{j+1} = \min\{ i^*_j , i_j \}$ where λ_j and i_j are determined in case (b) above;

Step 3: Set $j = j + 1$. If $j \le p$ go to Step 1.

Step 4: End.

Consequently for all $i \in K_p$ and $\lambda \in [0, \lambda_p]$,

$$\psi(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - \psi(u_i, w_i) \le \alpha\lambda\phi(u_i, w_i, v_i, z_i) \le -\frac{\alpha\lambda\delta\gamma}{2}, \tag{7.2.35}$$

and so $\psi(u_i, w_i) \to -\infty$ which is a contradiction.

Case II ($\psi(\hat{u}, \hat{w}) = 0$, $\psi(u_i, w_i) > 0$ for all $i \in \mathbb{Z}_+$)

The proof is Identical to Case I.

Case III ($\psi(\hat{u}, \hat{w}) \le 0$ and there exists $\hat{i}$ such that for all $i \ge \hat{i}$, $\psi(u_i, w_i) \le 0$)

There exists $i_0 > \hat{i}$ and $\lambda_0 > 0$ such that for all $i \ge i_0$ and $\lambda \in [0, \lambda_0]$

$$g^0(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - g^0(u_i, w_i) \le \alpha\lambda\phi(u_i, w_i, v_i, z_i) \le -\frac{\alpha\lambda\delta\gamma}{2}. \tag{7.2.36}$$

Consequently, $g^0(u_i, w_i) \to -\infty$ which is a contradiction. ∎

7.3 SEARCH DIRECTION SUBPROCEDURE

Step 2 of Algorithm 7.2.1 requires computation of $(v_i, z_i) \in G \times C$ such that $\phi(u_i, w_i, v_i, w_i) \le \gamma\Theta(u_i, w_i)$. We will present two methods to solve this subproblem. The first is to transcribe the calculation of $\Theta(\cdot, \cdot)$ (7.2.6) into a canonical quadratic program and solve the QP using a standard routine. The second is a special purpose iterative QP routine which is truncated when an appropriate (v_i, z_i) is found. In either case it is necessary to discretize the PDE in time and space to make the problem finite dimensional. We assume that the discretization is consistent with the requirements of Chapter 6. We assume special forms for G and C. Let G and C in (7.1.9) be defined:

$$G \triangleq \{\ u \in L_\infty^m([0,1]) \cap L_2^m([0,1]) \mid u(t) \in U, t \in [0,1]\ \}, \tag{7.3.1}$$

$$C \triangleq \{ \begin{bmatrix} x^0 \\ x^1 \\ x^2 \end{bmatrix} \in X \times \mathbb{R} \times \mathbb{R} \mid x^0(s) \in [\tilde{l}^0(s), \tilde{h}^0(s)] \text{ all } s \in S,$$

$$x^1 \in [\tilde{l}^1, \tilde{h}^1], x^2 \in [\tilde{l}^2, \tilde{h}^2]\ \}, \tag{7.3.2}$$

where

$$U \triangleq \{\ u \in \mathbb{R}^m \mid l_i \le y_i \le h_i, \text{ for } i \in \underline{m}\ \} \tag{7.3.3}$$

and X is a space indexed by $s \in S$.

METHOD I

We shall transform (7.2.6), (7.2.7) into a canonical QP. By (7.3.1) - (7.3.3),

$$\Theta(u,w) \triangleq \min_{v \in L_\infty^m([0,1]), z \in X \times \mathbb{R} \times \mathbb{R}, \alpha \in \mathbb{R}} \{ \frac{1}{2}|v-u|^2 + \frac{1}{2}|z-w|^2 + \alpha \mid \tag{7.3.4}$$

$$(a)\ \ -\rho\psi_+(u,w) + \langle \nabla g^0(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle - \alpha \le 0;$$

$$(b)\ \ g^j(u,w) - \psi(u,w) + \langle \nabla g^j(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle - \alpha \le 0,\ j \in \underline{p};$$

$$(c)\ \ \tilde{l}^0(s) - z^0(s) \le 0,\ z^0(s) - \tilde{h}^0(s) \le 0, s \in S,$$

$$(d)\ \ \tilde{l}^j - z^j \le 0,\ z^j - \tilde{h}^j \le 0,\ j \in \{\ 1,2\ \}$$

$$(e)\ \ l_i - v_i(t) \le 0,\ v_i(t) - h_i \le 0,\ i \in \underline{m},\ t \in ([0,1])\ \},$$

with $\tilde{h}^0(\cdot)$ and $\tilde{l}^0(\cdot)$ chosen so that C is compact.

We have used Stanford's quadratic programming package, LSSOL, to solve (7.3.4). LSSOL is an implementation of a two-phase (primal) quadratic active set programming method developed by Gill et. al [Gil.1] and is closely related to the method of Stoer [Sto.1]. The algorithm keeps track of an active set, those constraints which are satisfied exactly, adding a new constraint when one is encountered and deleting a constraint if (1) the current point is the minimum on the subspace defined by the active constraints and (2) its deletion provides for a direction of feasible descent. The algorithm treats constraints (7.3.4.c)-(7.3.4.e) specially, removing the associated variables from the QP calculation when the constraints are active. This speeds the calculation of the QP particularly when the solution has many values at its upper or lower limits (e.g., bang-bang solution).

METHOD II

Method II is a dual method. We transcribe $\Theta(u, w)$ (7.3.5) into dual form (7.3.8):

$$\Theta(u,w) = \min_{(v,z)\in G\times C} \Big\{ \frac{1}{2}\Big\| \begin{bmatrix} v-u \\ z-w \end{bmatrix} \Big\|^2 + \tag{7.3.5}$$

$$\max_{j\in \underline{p}} \Big\{ -\rho\psi_+(u,w) + \Big\langle \nabla g^0(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \Big\rangle ; g^j(u,w) + \Big\langle \nabla g^j(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \Big\rangle - \psi_+(u,w) \Big\}.$$

This is equivalent to:

$$\Theta(u,w) = \frac{1}{2} \min_{(v,z)\in G\times C} \max_{\mu\in \Sigma^{p+1}} \Big\{ \frac{1}{2}\Big\| \begin{bmatrix} v-u \\ z-w \end{bmatrix} \Big\|^2 - \mu^0\rho\psi_+(u,w) \tag{7.3.6}$$

$$+ \sum_{j=1}^{p} \mu^j (g^j(u,w) - \psi_+(u,w)) + \Big\langle \sum_{j=0}^{p} \mu^j \nabla g^j(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \Big\rangle \Big\},$$

where

$$\Sigma^{p+1} \triangleq \{ \mu \in \mathbb{R}^{p+1} \mid \sum_{j=0}^{p} \mu^j = 1 \text{ and } \mu^j \ge 0 \text{ for all } j \in \{0, 1, \cdots, p\} \}. \tag{7.3.7}$$

Since Σ^{p+1} is a compact, convex set and $G\times C$ is a closed, bounded, convex set and the expression between the braces in (7.3.6) is concave in μ and strictly convex in (v, z), we can switch the "min" and "max" according to Ky Fan's [Fan.1] extension of Van Neuman's Minimax Theorem to obtain:

$$\Theta(u,w) = \max_{\mu \in \Sigma^{p+1}} \{ \min_{(v,z) \in G \times C} \{ \frac{1}{2} \| \begin{bmatrix} v-u \\ z-w \end{bmatrix} \|^2 + \langle \sum_{j=0}^{p} \mu^j \nabla g^j(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle \} \tag{7.3.8}$$

$$- \mu^0 \rho \psi_+(u,w) + \sum_{j=1}^{p} \mu^j (g^j(u,w) - \psi_+(u,w)) \}.$$

For convenience, we define the following quantities,

$$\zeta(u,w,\mu) \triangleq \zeta_1(u,w,\mu) + \zeta_2(u,w,\mu) + \zeta_3(u,w,\mu), \tag{7.3.9}$$

where

$$\zeta_1(u,w,\mu) \triangleq -\min_{v \in G} \{ \frac{1}{2} \| v-u \|^2 + \langle \sum_{j=0}^{p} \mu^j \nabla_u g^j(u,w), v-u \rangle \}, \tag{7.3.10}$$

$$\zeta_2(u,w,\mu) \triangleq -\min_{z \in C} \{ \frac{1}{2} \| z-w \|^2 + \langle \sum_{j=0}^{p} \mu^j \nabla_w g^j(u,w), z-w \rangle \}, \tag{7.3.11}$$

$$\zeta_3(u,w,\mu) \triangleq \mu^0 \rho \psi_+(u,w) - \sum_{j=1}^{p} \mu^j (g^j(u,w) - \psi_+(u,w)), \tag{7.3.12}$$

where $\nabla g(u,w) = \begin{bmatrix} \nabla g_u(u,w) \\ \nabla g_w(u,w) \end{bmatrix}$. With this additional notation, we rewrite (7.3.8) as

$$\Theta(u,w) = -\min_{\mu \in \Sigma^{p+1}} \{ \zeta_1(u,w,\mu) + \zeta_2(u,w,\mu) + \zeta_3(u,w,\mu) \} \tag{7.3.13}$$

$$= -\min_{\mu \in \Sigma^{p+1}} \{ \zeta(u,w,\mu) \}.$$

We shall show that $\zeta_1(u,w,\mu)$, $\zeta_2(u,w,\mu)$ and $\zeta_3(u,w,\mu)$ are convex and twice differentiable in μ so that so we can use a constrained Newton-type algorithm to find $\Theta(u,w)$. From inspection, we see that $\zeta_3(u,w,\mu)$ is convex and twice differentiable in μ. We now proceed with $\zeta_1(u,w,\mu)$ and $\zeta_2(u,w,\mu)$.

Lemma 7.3.1: The function $\zeta_1(u,w,\mu)$ is convex in $\mu \in \Sigma^{p+1}$.

Proof: Let $\mu_1, \mu_2 \in \Sigma^{p+1}$. Then for $\lambda \in [0,1]$, there exists $v_\lambda \in G$ such that

$$\zeta_1(u,w,\lambda\mu_1 + (1-\lambda)\mu_2) = -\frac{1}{2}\|v_\lambda - u\|^2 - \langle \sum_{j=0}^{p} (\lambda\mu_1^j + (1-\lambda)\mu_2^j)\nabla_u g^j(u,w), v_\lambda - w \rangle$$

$$= -\lambda(\frac{1}{2}\|v_\lambda - u\|^2 + \sum_{j=0}^{p} \mu_1^j \nabla_u g^j(u,w), v_\lambda - u \rangle)$$

$$+ -(1-\lambda)(\frac{1}{2}\|v_\lambda - u\|^2 + \sum_{j=0}^{p} \mu_2^j \nabla_u g(u,w), v_\lambda - u \rangle)$$

$$\leq \lambda\zeta_1(u,w,\mu_1) + (1-\lambda)\zeta_1(u,w,\mu_2). \tag{7.3.14}$$

■

Corollary 7.3.2: $\zeta_2(u,w,\mu)$ and hence $\zeta(u,w,\mu)$ are convex in μ. ■

We now establish differentiability of $\zeta_1(u,w,\mu)$ in μ. We define

$$\xi(\mu) \triangleq -\zeta_1(u,w,\mu) = \min_{v \in G}\{ \frac{1}{2}\|v - u\|^2 + \langle \sum_{j=0}^{p} \mu^j \nabla_u g^j(u,w), v - u \rangle \}. \tag{7.3.15}$$

By completing the square, $\xi(\mu)$ becomes $\xi(\mu) = \xi_1(\mu) - \xi_2(\mu)$ where

$$\xi_1(\mu) = \min_{v \in G}\{ \frac{1}{2}\|v - (u - \sum_{j=0}^{p} \mu^j \nabla_u g^j(u,w)\|^2 \}, \tag{7.3.16}$$

$$\xi_2(\mu) = \frac{1}{2}\|\sum_{j=0}^{p} \mu^j \nabla_u g^j(u,w)\|^2. \tag{7.3.17}$$

We define

$$\phi(\mu) \triangleq \sum_{j=0}^{p} \mu^j \nabla_u g^j(u,w). \tag{7.3.18}$$

Then, $\xi_2(\mu) = \frac{1}{2}\|\phi(\mu)\|^2$, $\frac{\partial}{\partial\mu}\xi_2(\mu) = \phi^T(\mu)\frac{\partial}{\partial\mu}\phi(\mu)$, $\frac{\partial^2}{\partial\mu^2}\xi_2(\mu) = \frac{\partial}{\partial\mu}\phi^T(\mu)\,\frac{\partial}{\partial\mu}\phi(\mu)$.

Define $v : \Sigma^{p+1} \times G \to L_2 \cap L_\infty$:

$$v(\mu,u) \triangleq \arg\min_{v' \in G}\{ \frac{1}{2}\|v' - (u - \phi(\mu))\|^2 \}. \tag{7.3.19}$$

Lemma 7.3.3: The function $v(\mu)$ is continuous.

Proof: See Theorem 3.5.4. ■

Since G is defined by (7.3.1) and (7.3.3), for $t \in [0,1]$ and $i \in \{1, 2, \cdots, m\}$,

$$v(\mu)(t)_i = \begin{cases} h_i & \text{if } u(t)_i - \phi(\mu)(t)_i > h_i, \\ l_i & \text{if } u(t)_i - \phi(\mu)(t)_i < l_i, \\ u(t)_i - \phi(\mu)(t)_i & \text{otherwise.} \end{cases} \tag{7.3.20}$$

We define

$$S(\mu, u) \triangleq \{ t \in [0,1] \mid u(t)_i - \phi(\mu)(t)_i = h_i \text{ or } u(t)_i - \phi(\mu)(t)_i = l_i \text{ for some } i \in \underline{m} \} \tag{7.3.21}$$

We make the following assumption.

Assumption 7.3.2: $S(\mu, u)$ as defined in (7.3.21) is of zero measure. ■

Then $v(\cdot, u)$ is differentiable at any $\mu \in \Sigma^{p+1}$ and for $t \notin S(\mu, u)$,

$$\frac{\partial}{\partial \mu} v(\mu)(t)_i = \begin{cases} (\nabla_u g^0(u, w)(t)_i \ \cdots \ \nabla_u g^p(u, w)(t)_i), & \text{if } |u(t)_i - \phi(\mu)(t)_i| \in [l_i, u_i] \\ (0 \ \cdots \ 0,) & \text{otherwise.} \end{cases} \tag{7.3.22}$$

Therefore since,

$$\xi_1(\mu) = \frac{1}{2} \| v(\mu) - (u - \phi(\mu)) \|^2, \tag{7.3.23}$$

$$\frac{\partial}{\partial \mu} \xi_1(\mu) = [(v(\mu) - u) + \phi(\mu)]^T [\frac{\partial}{\partial \mu} v(\mu) + \frac{\partial}{\partial \mu} \phi(\mu)], \tag{7.3.24}$$

and since $\xi(\mu) = \xi_1(\mu) + \xi_2(\mu)$,

$$\frac{\partial}{\partial \mu} \xi(\mu) = (v(\mu) - (u - \phi(\mu))^T \frac{\partial}{\partial \mu} v(\mu) + (v(\mu) - u))^T \frac{\partial}{\partial \mu} \phi(\mu). \tag{7.3.25}$$

Since if $|v(\mu)(t)_i - (u(t)_i - \phi(\mu)(t)_i)| > 0$ (i.e., $u(t)_i - \phi(\mu)(t)_i \notin [l_i, h_i]$) then $\frac{\partial}{\partial \mu} v(\mu)(t)_i = 0$, the first term in the right hand side of (7.3.25) is zero and

$$\nabla\xi(\mu) = \left[\frac{\partial}{\partial\mu}\xi(\mu)\right]^T = [(v(\mu) - u)^T\frac{\partial}{\partial\mu}\phi(\mu)]^T = \begin{bmatrix} \langle \nabla_u g^0(u, w), v(\mu) - u \rangle \\ \vdots \\ \langle \nabla_u g^p(u, w), v(\mu) - u \rangle \end{bmatrix}. \tag{7.3.26}$$

Therefore, the second derivative of $\xi(\cdot)$ is

$$\frac{\partial^2}{\partial\mu^2}\xi(\mu) = \frac{\partial}{\partial\mu}v(\mu)^T\frac{\partial}{\partial\mu}\phi(\mu) + (v(\mu) - u)\frac{\partial^2}{\partial\mu^2}\phi(\mu). \tag{7.3.27}$$

Since $\frac{\partial^2}{\partial\mu^2}\phi(\mu) = 0$,

$$\frac{\partial^2}{\partial\mu^2}\xi(\mu) = \frac{\partial}{\partial\mu}v(\mu)^T\frac{\partial}{\partial\mu}\phi(\mu) = \begin{bmatrix} \nabla_u g^0(u, w)^T \\ \vdots \\ \nabla_u g^p(u, w)^T \end{bmatrix}\frac{\partial}{\partial\mu}v(\mu). \tag{7.3.28}$$

Therefore, we have proved the following theorem:

Theorem 7.3.3: $\xi(\cdot)$ is twice differentiable and its first and second derivatives are given by (7.3.26) and (7.3.28). ■

Finally, we show that $\frac{\partial^2}{\partial\mu^2}\xi(\cdot)$ is continuous. We first prove a lemma.

Lemma 7.3.4: Let $\{ v_i \}_{i \in \{0, 1, \cdots, p\}} \subset L_\infty([0, 1])$ and $\bar{v} \in L_\infty([0, 1])$. We define

$$T(\mu) \triangleq \{ t \in [0, 1] \mid \bar{v} + \sum_{i=0}^{p} \mu^i v_i \geq 0 \} \tag{7.3.29}$$

$$S(\mu) \triangleq \{ t \in [0, 1] \mid \bar{v} + \sum_{i=0}^{p} \mu^i v_i = 0 \} \tag{7.3.30}$$

$$R(\mu) \triangleq \{ t \in [0, 1] \mid \bar{v} + \sum_{i=0}^{p} \mu^i v_i \leq 0 \} \tag{7.3.31}$$

and we assume that $m(S(\mu)) = 0$ for all $\mu \in \mathbb{R}^{p+1}$ where $m(\cdot)$ denotes the Lesbegue measure on $[0, 1]$. Then for all $\hat{\mu} \in \mathbb{R}^{p+1}$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $m(T(\mu) \,\Delta\, T(\hat{\mu})) \leq \varepsilon$ for all $\mu \in B(\hat{\mu}, \delta)$ where

$$T(\mu)\ \Delta\ T(\hat{\mu}) \triangleq \{\ t \in [0,1] \mid t \in T(\mu) \text{ and } t \notin T(\hat{\mu})\ \}$$
$$\cup \{\ t \in [0,1] \mid t \notin T(\mu) \text{ and } t \in T(\hat{\mu})\ \}. \qquad (7.3.32)$$

Proof: We define $\rho_t \triangleq \overline{v}(t) + \sum_{i=0}^{p} \mu^i v_i(t)$, and $Q(\hat{\mu},\rho) \triangleq \{\ t \in T(\hat{\mu}) \mid \rho_t < \rho\ \}$. Then there exists $\tilde{\rho} > 0$ such that $m(Q(\hat{\mu},\tilde{\rho})) \le \frac{\varepsilon}{2}$. If such at $\tilde{\rho}$ does not exist, then for all $\rho > 0$, $m(Q(\hat{\mu},\rho)) > \frac{\varepsilon}{2}$. Since $Q(\hat{\mu},\rho_1) \subset Q(\hat{\mu},\rho_2)$ if $\rho_1 < \rho_2$, $m(\bigcap_{\rho>0} Q(\hat{\mu},\rho)) \ge \frac{\varepsilon}{2}$ and so for $t \in \bigcap_{\rho>0} Q(\hat{\mu},\rho)$, $t \in S(\hat{\mu})$, and therefore $m(S(\hat{\mu})) \ge \frac{\varepsilon}{2}$ which is a contradiction. Similarly, there exists $\hat{\rho} \in (0,\tilde{\rho})$ such that $m(P(\hat{\mu},\hat{\rho})) \le \frac{\varepsilon}{2}$ where $P(\hat{\mu},\rho) \triangleq \{\ t \in R(\hat{\mu}) \mid \rho_t > -\rho\ \}$. Consequently, since $\{\ v_i\ \}$ is bounded and $m(S(\hat{\mu})) = 0$, there exists $\delta > 0$ such that $m(T(\mu)\ \Delta\ T(\hat{\mu})) \le \varepsilon$ for all $\mu \in B(\hat{\mu},\delta)$. ■

Theorem 7.3.5: The Hessian $\frac{\partial^2}{\partial\mu^2}\xi(\cdot)$ is continuous.

Proof: Since for $S(\mu,u)$ defined in (7.3.21), $m(S(\mu,u)) = 0$, and $u(\cdot)$, and $\nabla g_u^j(\cdot)$ are bounded measurable functions, the proof follows from an extension on Lemma 7.3.4. ■

Therefore, $\zeta_1(u,w,\mu)$ is twice continuously differentiable in μ. We can similarly show that $\zeta_2(u,w,\mu)$ is twice continuously differentiable in μ. Hence,

Corollary 7.3.6: The function $\zeta(u,w,\mu)$ is twice continuously differentiable in μ. ■

The Levitin-Polyak algorithm [Lev.1] is essentially a constrained Newton method. At each step, a quadratic approximation is minimized on the feasible set, and this becomes the new point from which a new quadratic approximation is derived. Levitin and Polyak have shown that if the method converges, then the convergence is quadratic. For numerical computations, we use an implementation by J.E. Higgins [Hig.1]. This implementation has a stabilizing step-size procedure. We have slightly modified Higgins' implementation by adding a stopping criterion. Since this is the search direction finding subproblem, the algorithm must stop in a finite number of iterations.

Algorithm 7.3.7: (To solve Step 2 of Algorithm 7.2.1.)

Data: $\mu_0 \in \Sigma^{p+1}$, $u \in G$, $w \in C$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1]$.

Step 0: $i = 0$, $v_0 = u$, $z_0 = w$.

Step 1: Calculate $\nabla\zeta(u,w,\mu_i)$ and $\dfrac{\partial^2}{\partial\mu^2}\zeta(u,w,\mu_i)$.

Step 2: Calculate $\Theta(\mu_i)$ and $\sigma(\mu_i)$ using

$$\sigma(\mu_i) \triangleq \arg\min_{\sigma \in \Sigma^{p+1}} \{ \langle \nabla\zeta(u,w,\mu_i), \sigma - \mu_i \rangle + (\sigma - \mu_i)^T \frac{\partial^2}{\partial^2\mu}\zeta(u,w,\mu_i)(\sigma - \mu_i) \}. \quad (7.3.33)$$

$$\Theta(\mu_i) \triangleq \min_{\sigma \in \Sigma^{p+1}} \{ \langle \nabla\zeta(u,w,\mu_i), \sigma - \mu_i \rangle +$$

$$(\sigma - \mu_i)^T \frac{\partial^2}{\partial^2\mu}\zeta(u,w,\mu_i)(\sigma - \mu_i) \}. \quad (7.3.34)$$

Step 3: $\lambda_i \triangleq \max\{ \lambda \in \{ 1, \beta, \beta^2, \cdots \} \mid$

$$\zeta(u,w,\mu_i\, \lambda(\sigma(\mu_i) - \mu_i)) - \zeta(u,w,\mu_i) \le \alpha\lambda\Theta(\mu_i).$$

Step 4: Set $\mu_{i+1} = \mu_i + \lambda_i(\sigma(\mu_i) - \mu_i)$, and calculate v_{i+1} and z_{i+1} using

$$v_{i+1} = v(\mu_{i+1}) \triangleq \arg\min_{v \in G} \{ \frac{1}{2}|v - u|^2 + \langle \sum_{j=0}^{p} \mu_{i+1}^j \nabla_u g^j(u,w), v - u \rangle \}, \quad (7.3.35)$$

$$z_{i+1} = z(\mu_{i+1}) \triangleq \arg\min_{z \in C} \{ \frac{1}{2}|z - w|^2 + \langle \sum_{j=0}^{p} \mu_{i+1}^j \nabla_w g^j(u,w), z - w \rangle \}. \quad (7.3.36)$$

Step 5: Calculate $\phi(u,w,v_{i+1},z_{i+1})$ by (7.2.7).

If $\phi(u,w,v_{i+1},z_{i+1}) \le -\gamma\zeta(u,w,\mu_{i+1})$, STOP.

Step 6: $i = i + 1$; go to Step 1. ■

Lemma 7.3.8: The functions $v(\cdot)$ and $z(\cdot)$ are continuous.

Proof: The continuity of $v(\cdot)$ is proven in Lemma 7.3.3, and the continuity of $z(\cdot)$ follows analogously. ∎

Higgins has proven convergence of Algorithm 7.3.7 for the case with no stopping rule:

Theorem 7.3.9 [Hig.1]: If the infinite sequence $\{ \mu_i \}$ is generated by Algorithm 7.3.7 with no stopping condition and the function $\zeta(u,w,\cdot)$ is convex and twice continuously differentiable, then any accumulation point of $\{ \mu_i \}$, $\hat{\mu}$, is a minimizer for $\zeta(u,w,\cdot)$. ∎

From (7.2.6), (7.2.7), and (7.3.13) it is true that for any $u, v \in G$, $w, z \in C$ and $\mu \in \Sigma^{p+1}$,

$$-\zeta(u,w,\mu) \leq \Theta(u,w) \leq \phi(u,w,v,z). \tag{7.3.37}$$

When the stopping criteria is met, there exists v_{i+1}, z_{i+1} and μ_{i+1} such that

$$\phi(u,w,v_{i+1},z_{i+1}) \leq -\gamma\zeta(u,w,\mu_{i+1}) \leq \gamma\Theta(u,w). \tag{7.3.38}$$

The pair (v_{i+1}, z_{i+1}) can then be used by Step 2 in Algorithm 7.2.1.

Theorem 7.3.10: Algorithm 7.3.7 terminates in a finite number of steps.

Proof: Assume that Algorithm 7.3.7 does not terminate in a finite number of steps. Then, Algorithm 7.3.5 produces an infinite sequence $\{ \mu_i \}$. Since Σ^{p+1} is compact, $\{ \mu_i \}$ has a subsequence $K \subset \mathbb{Z}_+$ and an accumulation point $\hat{\mu} \in \Sigma^{p+1}$ such that $\mu_i \xrightarrow{K} \hat{\mu}$ By Theorem 7.3.8, $\hat{\mu}$ is a minimizer of $\zeta(u,w,\cdot)$ and so

$$\lim_{i \xrightarrow{K} \infty} \zeta(u,w,\mu_i) = \zeta(u,w,\hat{\mu}) = -\Theta(u,w). \tag{7.3.39}$$

Since $\phi(u,w,v(\hat{\mu}),z(\hat{\mu})) = \Theta(u,w)$, and $\phi(u,w,\cdot,\cdot)$, $v(\cdot)$ and $z(\cdot)$ are continuous, $\phi(u,w,v_i,z_i) \xrightarrow{K} \Theta(u,w)$. Consequently, $|\phi(u,w,v_i,z_i) + \zeta(u,w,\mu_i)| \xrightarrow{K} 0$ and Algorithm 7.3.7 terminates in a finite number of steps. ∎

CHAPTER 8

NUMERICAL RESULTS

This chapter reports on numerical experiments on the problem of moving a flexible beam. An optimal control problem is formulated and transcribed into a form which can be solved using semi-infinite optimization techniques. All experiments were carried out on a SUN 3 microcomputer with a Floating Point Accelerator.

8.1 PROBLEM STATEMENT

We consider the hollow aluminum tube depicted in figure 8.1 (page 153). The tube has a length of one meter, a cross sectional radius of 1.0 cm, and a thickness of 1.6 mm. Attached to one end of the tube is a mass of 1 kg; attached to the other end is a shaft connected to a motor. For simplicity, we assume that the torque produced by the motor can be directly controlled. Our aim is to determine the torque necessary to rotate the tube and bring it to rest. The maximum torque produced by the motor is 5 newton-meters. The equations of motion determined by application of the standard Euler-Bernoulli tube with Kelvin-Voigt visco-elastic damping are:

$$mw_{tt}(t,x) + CIw_{txxxx}(t,x) + EIw_{xxxx}(t,x) - m\Omega^2(t)w(t,x)$$

$$= \frac{m}{Ml^2 + \frac{1}{3}ml^3} u(t)x, x \in [0,l] \tag{8.1.1}$$

with boundary conditions:

$$w(t,0) = 0, w_x(t,0) = 0, CIw_{txx}(t,1) + EIw_{xx}(t,1) = 0. \tag{8.1.2}$$

$$M[\Omega^2(t)w(t,1) - w_{tt}(t,1) - u(t)l] + CIw_{txxx}(t,1) + EIw_{xxx}(t,1) = 0, \tag{8.1.3}$$

and rigid body dynamics:

$$\frac{d}{dt}\Theta(t) = \Omega(t), \quad \frac{d}{dt}\Omega(t) = \frac{1}{Ml^2 + \frac{1}{3}ml^3}u(t), \tag{8.1.4}$$

where $w(t,x)$ is the displacement of the tube from the *shadow tube* (which remains undeformed during the motion) due to bending as a function of time and distance along the tube; $u(t)$ is the torque applied by the motor, and $\Omega(t)$ is the resulting angular velocity (in radians per second). We shall denote by $\Theta(t)$ the angular displacement of the rigid body (in radians). The values for the parameters in (8.1.1) - (8.1.3) are: $l = 1.0$ m, $m = .257$ kg/m, $C = 6.30\times10^7$ pascals/sec., $E = 6.30\times10^9$ pascals, $I = 1.005\times10^{-8}m^4$, $M = 0.914$ kg. The tube is very lightly damped (0.1 percent).

We assume that the tube is initially at rest with no deformations, and so the following initial conditions hold:

$$w(0,x) = w_t(0,x) = 0, \quad x \in [0,1]. \tag{8.1.5a}$$

$$\Theta(0) = \Omega(0) = 0. \tag{8.1.5b}$$

We consider four problems:

P$_1$: Minimize the time required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint.

P$_2$: Minimize the total energy required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint and the maneuver time not exceeding a given bound.

P$_3$: Minimize the time required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint and an upper bound on the potential energy due to deformation of the tube throughout the entire maneuver.

P$_4$: Minimize the total energy required to rotate the tube 45 degrees, from rest to rest, subject to the given torque constraint, the maneuver time not exceeding a given bound, and an upper bound on the potential energy due to deformation of the tube throughout the entire maneuver.

8.2 MATHEMATICAL FORMULATION OF THE FOUR PROBLEMS

We will formulate the above problems P_1, P_2, P_3, and P_4 in the form of the following canonical optimization problem:

$$\mathbf{P}_0 : \min_{T \in [\tau_{min}, \tau_{max}], u \in G_T} \{ g^0(u,T) \mid g^j(u,T) \leq 0, j \in \underline{p} \}, \quad (8.2.1)$$

where $\tau_{min} > 0$ and $\tau_{max} < \infty$, $\underline{p} \triangleq \{ 1, 2, \cdots, p \}$,

$$G_T \triangleq \{ u \in L_\infty[0,T] \mid |u(t)| \leq 5, t \in [0,T] \}, \quad (8.2.2)$$

and $g^j : G_T \times T \to \mathbb{R}$ is continuously differentiable for $j \in \{ 0, 1, \cdots, p \}$. We define $\psi(u,T) \triangleq \max_{j \in \underline{p}} \{ g^j(u,T) \}$ and $\psi_+(u,T) \triangleq \max\{ 0, \psi(u,T) \}$.

For theoretical purposes, see Chapter 4, we constrain the final time, T, to be in an interval $[\tau_{min}, \tau_{max}]$ where $\tau_{min} > 0$ and $\tau_{max} < \infty$. These values τ_{min} and τ_{max} can be chosen so that these constraints are not active at the solutions to $\mathbf{P}_1$ through $\mathbf{P}_4$. We shall be making use of the following functions. First, noting that T denotes the final time, we define

$$g^1(u,T) \triangleq T. \quad (8.2.3)$$

The input energy is defined as the integral of the square of the input; hence we define

$$g^2(u,T) \triangleq \int_0^T u(t)^2 dt. \quad (8.2.4)$$

Next we define

$$g^3(u,T) \triangleq (\Theta(T) - \pi/4)^2 \quad (8.2.5)$$

to be the square of the angular error at the final time. We say that *the tube is at rest* when the total energy of the tube is zero. This energy is composed of the energy due to rigid body motion and energy due to vibration and deformation. Rigid body energy at final time is proportional to the square of the angular velocity. Hence we define

$$g^4(u,T) \triangleq \Omega(T)^2. \tag{8.2.6}$$

The kinetic energy due to vibration of the tube at time t is given by

$$K(t,u) \triangleq \frac{m}{2}\int_0^1 w_t(t,x)^2 dx, \tag{8.2.7}$$

and the potential energy due to deformation of the tube at time t is given by

$$P(t,u) \triangleq \frac{EI}{2}\int_0^1 w_{xx}(t,x)^2 dx. \tag{8.2.8}$$

We now define the values of the kinetic and potential energies at the final time:

$$g^5(u,T) \triangleq K(T,u),\ g^6(u,T) \triangleq P(T,u). \tag{8.2.9}$$

The tube is at rest if $g^4(u,T) = g^5(u,T) = g^6(u,T) = 0$.

For problems $\mathbf{P}_3$ and $\mathbf{P}_4$, we require that the potential energy due to the tube deformation be within a specified range throughout the entire maneuver. This constraint has the form $P(t,u) \le f(t)$ for all $t \in [0,T]$, where $f(\cdot)$ is a given positive bound function with a finite number of discontinuities. This is a *state-space constraint*, and does not fit the canonical form $\mathbf{P}_0$. However, we can replace it by an equivalent form which requires that we define

$$g^7(u,T) \triangleq \int_0^T [\max\{\ P(t,u) - f(t), 0\ \}]^2, \tag{8.2.10}$$

then since $P(t,u)$ is continuous, $g^7(u,T) = 0$ if and only if $P(t,u) \le f(t)$ for all $t \in [0,T]$.

The functions $g^j : G_T \times [\tau_{\min}, \tau_{\max}] \to \mathbb{R}$ are continuously differentiable for all $j \in \{\ 1,2, \cdots 7\ \}$. To improve conditioning of the problems $\mathbf{P}_1$ - $\mathbf{P}_4$, we relax each of the equality constraints by a small amount (Section 7.2). The relaxation can be be chosen to be sufficiently small so as not to matter from a practical point of view. The four problems now acquire the following mathematical form

$$\begin{aligned}\mathbf{P}_1 : \min\{\, & g^1(u,T) \mid g^3(u,T) - \varepsilon \le 0, g^4(u,T) - \varepsilon \le 0, g^5(u,T) - \varepsilon \le 0, \\ & g^6(u,T) - \varepsilon \le 0, u \in G_T \,\};\end{aligned} \tag{8.2.11}$$

$$\begin{aligned}\mathbf{P}_2 : \min\{\, & g^2(u,T) \mid g^1(u,T) - T_f \le 0, g^3(u,T) - \varepsilon \le 0, g^4(u,T) - \varepsilon \le 0, \\ & g^5(u,T) - \varepsilon \le 0, g^6(u,T) - \varepsilon \le 0, u \in G_T \,\};\end{aligned} \tag{8.2.12}$$

$$\begin{aligned}\mathbf{P}_3 : \min\{\, & g^1(u,T) \mid g^3(u,T) - \varepsilon \le 0, g^4(u,T) - \varepsilon \le 0, g^5(u,T) - \varepsilon \le 0, \\ & g^6(u,T) - \varepsilon \le 0, g^7(u,T) - \varepsilon \le 0, u \in G_T \,\};\end{aligned} \tag{8.2.13}$$

$$\begin{aligned}\mathbf{P}_4 : \min\{\, & g^2(u,T) \mid g^1(u,T) - T_f \le 0, g^3(u,T) - \varepsilon \le 0, g^4(u,T) - \varepsilon \le 0, \\ & g^5(u,T) - \varepsilon \le 0, g^6(u,T) - \varepsilon \le 0, g^7(u,T) - \varepsilon \le 0, u \in G_T \,\}.\end{aligned} \tag{8.2.14}$$

In our experiments, we set $\varepsilon = 10^{-4}$. Thus, with this relaxation, we are requiring that the final value of the angle Θ be in the interval $[45 - 0.5, 45 + 0.5]$ degrees.

8.3 SPATIAL DISCRETIZATION

In this section using spatial discretization (Chapter 6) and time scaling (Chapter 4), we transcribe problems $\mathbf{P}_1$ - $\mathbf{P}_4$ into a sequence of problems $\overline{\mathbf{P}}_1^q$ - $\overline{\mathbf{P}}_4^q$. In Section 8.4, we shall transcribe problems $\overline{\mathbf{P}}_1^q$ - $\overline{\mathbf{P}}_4^q$ into $\hat{\mathbf{P}}_1^{n,q}$ - $\hat{\mathbf{P}}_4^{n,q}$ using Newmark's method. In Section 8.5, we shall state a refinement criterion that is used to determine when n and q are to be increased. In Section 8.6, we give solutions to $\mathbf{P}_1$ - $\mathbf{P}_4$ which are solved using the implementable versions of the algorithms described in Chapter 7. Spatial discretization is accomplished by applying Galerkin's method. A basis of Hermite cubics is substituted in the weak form of (8.1.1) to derive an ordinary differential equation. The resulting ODE is discretized by Newmark's method.

We begin by deriving the weak form of (8.1.1) - (8.1.3). For $\eta \in H_E^2$ where H_E^2 is the completion of

$$H_B^2 \triangleq \{ \eta \in H^2([0,1]) \mid \eta(0) = \eta'(0) = 0 \}, \tag{8.3.0}$$

$$\int_0^1 \eta(x)[mw_{tt}(t,x) + CIw_{txxxx}(t,x) + EIw_{xxxx}(t,x) - m\Omega^2(t)w(t,x)]dx$$
$$+ \eta(1)[M(w_{tt}(t,1) - \Omega^2(t)w(t,1)) - CIw_{txxx}(t,1) - EIw_{xxx}(t,1)]$$
$$= (-\mu m\int_0^1 \eta(x)xdx - M\eta(1))u(t), \tag{8.3.1}$$

where

$$\mu \triangleq \frac{1}{M + \frac{m}{3}}. \tag{8.3.2}$$

Performing integration by parts and applying the boundary conditions (8.1.2) and (8.3.0), (8.3.1) becomes:

$$\int_0^1 (mw_{tt}(t,x)\eta(x) + CIw_{txx}(t,x)\eta_{xx}(x) + EIw_{xx}(t,x)\eta_{xx}(x) - m\Omega^2(t)w(t,x)\eta(x))dx$$
$$+ M\eta(1)[w_{tt}(t,1) - \Omega^2(t)w(t,1)] = (-\mu m\int_0^1\eta(x)xdx - M\eta(1))u(t) \tag{8.3.3}$$

for all $\eta \in H_E^2$.

Galerkin's Method

Galerkin's method consists of choosing a subspace S_q of H_E^2 and solving (8.3.3) restricted to that subspace. Let $\{ N_j(x) \}_{j \in \{1,2,\cdots,Q\}}$, where $Q \triangleq 2^{q+1}$, be a basis for S_q. (There are Q basis elements.) Then for $w \in C([0,1],S_q) \subset C([0,1],H_E^2)$, $w(t,x) = \sum_{j=1}^{Q} w^j(t)N^j(x)$ with $\{ w^j \}_{j \in \{1,2,\cdots,Q\}} \subset C([0,1])$. Restricting (8.3.3) to the subspace $S_q \subset H_E^2$:

$$m\sum_{j=1}^{Q} w_{tt}^{j}(t)\int_0^1 N^j(x)N^k(x)dx + CI\sum_{j=1}^{Q} w_t^j(t)\int_0^1 N_{xx}^j(x)N_{xx}^k(x)dx$$

$$+ EI\sum_{j=1}^{Q} w^j(t)\int_0^1 N_{xx}^j(x)N_{xx}^k(x)dx - m\Omega^2(t)\sum_{j=1}^{Q}\int_0^1 w^j(t)N^j(x)N^k(x)dx$$

$$+ M\sum_{j=1}^{Q} w_{tt}^j(t)N^j(1)N^k(1) - M\Omega^2(t)\sum_{j=1}^{Q} w^j(t)N^j(1)N^k(1)$$

$$= (-\mu m\int_0^1 N^k(x)xdx + MN^k(1))u(t) \tag{8.3.4}$$

for all $k \in \{1,2,\cdots,Q\}$, with initial conditions $w^k(0) = w_t^k(0) = 0$ for all $k \in \{1,2,\cdots,Q\}$. Define matrices $\mathbf{M} \in \mathbb{R}^{Q\times Q}$, $\mathbf{K} \in \mathbb{R}^{Q\times Q}$, $\mathbf{V} \in \mathbb{R}^{Q\times Q}$, $\mathbf{G} \in \mathbb{R}^{Q}$:

$$\mathbf{M}_{ij} \triangleq \int_0^1 N^i(x)N^k(x)dx, \qquad \mathbf{V}_{ij} \triangleq N^i(1)N^j(1). \tag{8.3.5a}$$

$$\mathbf{K}_{ij} \triangleq \int_0^1 N_{xx}^i(x)N_{xx}^k(x)dx, \qquad \mathbf{G}_i \triangleq \mu\int_0^1 N^i(x)xdx + \frac{M}{m}N^i(1) \tag{8.3.5b}$$

for $i \in \{1,2,\cdots,Q\}$, $j \in \{1,2,\cdots,Q\}$. If we define $W(\cdot) \in C^Q([0,1])$,

$$W(t) = (w^1(t) \;\; w^2(t) \;\; \cdots \;\; w^Q(t))^T, \tag{8.3.6}$$

then (8.3.4) can be written in matrix form:

$$(m\mathbf{M} + M\mathbf{V})W_{tt}(t) + CI\mathbf{K}W_t(t) + EI\mathbf{K}W(t) - \Omega^2(t)(m\mathbf{M} + M\mathbf{V})W(t) = -m\mathbf{G}u(t). \tag{8.3.7}$$

For $w \in C([0,1],S_q)$, $w(t,x) = \sum_{j=1}^{Q} w^j(t)N^j(x)$ and so $K(t,u)$ and $P(t,u)$ defined in (8.2.7) and (8.2.8) are

$$K(t,u) = \frac{m}{2}\int_0^1 (\frac{\partial}{\partial t}\sum_{j=1}^{Q} w^j(t)N^j(x))^2 dx = \frac{m}{2}W_t(t)^T\mathbf{M}W_t(t). \tag{8.3.8}$$

$$P(t,u) = \frac{EI}{2}\int_0^1 (\frac{\partial^2}{\partial x^2}\sum_{j=1}^{Q} w^j(t)N^j(x))^2 dx = \frac{EI}{2}W(t)^T\mathbf{K}W(t). \tag{8.3.9}$$

Hermite Splines:

The subspace S_q consists of functions which are twice differentiable and satisfy the boundary conditions (8.3.1). Linear elements (see equation (6.5.0)) are only once differentiable and hence are not an acceptable choice of basis functions for S_q. The standard choice for a basis for S_q are the Hermite cubics. By using Hermite cubics as a basis, we guarantee existence of a second derivative. There are two types of Hermite cubics denoted $\phi(\cdot)$ and $\omega(\cdot)$. Let $h = \frac{1}{2^q}$. Then for $i \in \{ 1,2, \cdots ,2^q \}$:

$$\phi_q^i(x) \triangleq \begin{cases} (|\frac{x}{h} - i| - 1)^2(2|\frac{x}{h} - i| + 1) & \text{for } \frac{x}{h} \in [i-1, i+1] \\ 0 & \text{otherwise.} \end{cases} \tag{8.3.10}$$

$$\omega_q^i(x) \triangleq \begin{cases} h(\frac{x}{h} - i)(|\frac{x}{h} - i| - 1)^2 & \text{for } \frac{x}{h} \in [i-1, i+1] \\ 0 & \text{otherwise.} \end{cases} \tag{8.3.11}$$

These cubics are chosen so that such that for all $i \in \{ 1,2, \cdots ,2^q \}$,

$$\phi_q^i(ih) = \frac{\partial}{\partial x}\omega_q^i(ih) = 1, \tag{8.3.12a}$$

$$\phi_q^i((i\pm1)h) = \frac{\partial}{\partial x}\phi_q^i((i\pm1)h) = \omega_q^i((i\pm1)h) = \frac{\partial}{\partial x}\omega_q^i((i\pm1)h) = 0, \tag{8.3.12b}$$

thereby guaranteeing continuity of the first derivative and existence of a second derivative. We define $\{ N_q^i(\cdot) \}_{i \in \{1,2,\cdots,Q\}}$

$$N_q^{2i-1} \triangleq \phi_q^i(x), \quad N_q^{2i} \triangleq \omega_q^i(x), \tag{8.3.13}$$

for $i \in \{ 1,2, \cdots ,2^q \}$ and $x \in [0,1]$. Given the basis functions $\{ N_q^i(\cdot) \}_{i \in \{1,2,\cdots,Q\}}$, the matrices **M**, **K**, **V**, and **G** can be calculated. There derivation of the matrices **M** and **K** are found in Strang[Str.1]. We first define symmetric matrices $\tilde{m}, \tilde{k} \in \mathbb{R}^{4\times4}$: (only the upper triangular part is given)

$$\bar{m} = \begin{bmatrix} m^{11} & m^{12} \\ m^{21} & m^{22} \end{bmatrix} \triangleq \frac{h}{420} \begin{bmatrix} 156 & 22h & 54 & -13h \\ & 4h^2 & 13h & -3h^2 \\ & & 156 & -22h \\ & & & 4h^2 \end{bmatrix}, \tag{8.3.14}$$

$$\bar{k} = \begin{bmatrix} k^{11} & k^{12} \\ k^{21} & k^{22} \end{bmatrix} \triangleq \frac{1}{h^3} \begin{bmatrix} 12 & 6h & 12 & 6h \\ & 4h^2 & -6h & 2h^2 \\ & & 12 & -6h \\ & & & 4h^2 \end{bmatrix}, \tag{8.3.15}$$

then **M**, **K**, and **V** are:

$$\mathbf{K} = \begin{bmatrix} k^{11}+k^{22} & k^{12} & & & & \\ k^{21} & k^{11}+k^{22} & k^{12} & & & \\ & k^{21} & k^{11}+k^{22} & \cdots & & \\ & & \vdots & & & \\ & & & & k^{11}+k^{22} & k^{12} \\ & & & & k^{21} & k^{22} \end{bmatrix}, \tag{8.3.16}$$

$$\mathbf{M} = \begin{bmatrix} m^{11}+m^{22} & m^{12} & & & & \\ m^{21} & m^{11}+m^{22} & m^{12} & & & \\ & m^{21} & m^{11}+m^{22} & \cdots & & \\ & & \vdots & & & \\ & & & & m^{11}+m^{22} & m^{12} \\ & & & & m^{21} & m^{22} \end{bmatrix}, \tag{8.3.17}$$

$$\mathbf{V} = \begin{bmatrix} 0 & 0 & & & \\ 0 & 0 & 0 & & \\ & 0 & 0 & \cdots & \\ & & \vdots & & \\ & & & 1 & 0 \\ & & & 0 & 0 \end{bmatrix}. \tag{8.3.18}$$

Calculation of **G**:

For $i \in \{1, 2, \cdots, 2^q - 1\}$,

$$\mathbf{G}_{2i-1} = \mu \int_0^1 \phi_q^i(x) x dx, \quad \text{and} \quad \mathbf{G}_{2i} = \mu \int_0^1 \omega_q^i(x) x dx. \tag{8.3.19}$$

By combining (8.3.10) and (8.3.19), we obtain for $i \in \{1, 2, \cdots, 2^q - 1\}$,

$$G_{2i-1} = \mu \int_{(i-1)h}^{(i+1)h} \left[|\frac{x}{h} - i| - 1 \right]^2 \left[2|\frac{x}{h} - i| + 1 \right] x dx. \tag{8.3.20}$$

Define $y = \frac{x}{h} - i$ so that $x = (y + i)h$ and $dx = hdy$.

$$\frac{1}{\mu} G_{2i-1} = h^2 \int_{-1}^{0} (y + 1)^2 (1 - 2y)(y + i) dy + h^2 \int_{0}^{1} (y - 1)^2 (1 + 2y)(y + i) dy$$

$$= h^2 \int_{-1}^{0} (y^2 + 2y + 1)(-2y^2 + (1 - 2i)y + i) dy + h^2 \int_{0}^{1} (y^2 - 2y + 1)(2y^2 + (1 + 2i)y + i) dy$$

$$= h^2 \int_{-1}^{0} (-2y^4 + (1 - 2i - 4)y^3 + (i + 2 - 4i - 2)y^2 + (2i + 1 - 2i)y + i) dy$$

$$+ h^2 \int_{0}^{1} (2y^4 + (1 + 2i - 4)y^3 + (2 - 2 - 4i + i)y^2 + (1 + 2i - 2i)y + i) dy$$

$$= h^2 [-\frac{2}{5} y^5 + \frac{-2i - 3}{4} y^4 + -\frac{3}{3} y^3 + \frac{1}{2} y^2 + yi] \Big|_{-1}^{0}$$

$$+ h^2 [\frac{2}{5} y^5 + \frac{2i - 3}{4} y^4 + -\frac{3}{3} y^3 + \frac{1}{2} y^2 + yi] \Big|_{0}^{1} \tag{8.3.21}$$

$$G_{2i-1} = \mu h^2 \left[\frac{10i - 3}{20} \right] + \mu h^2 \left[\frac{10i + 3}{20} \right] = \mu h^2 i. \tag{8.3.21a}$$

For $i = 2^q$,

$$G_{2i-1} = \mu h^2 \left[\frac{10i - 3}{20} \right] + \frac{M}{m}. \tag{8.3.22}$$

By combining (8.3.11) and (8.3.19), we obtain for $i \in \{ 1, 2, \cdots, 2^q - 1 \}$,

$$G_{2i} = \mu \int_{(i-1)h}^{(i+1)h} h \left[\frac{x}{h} - i \right] \left[|\frac{x}{h} - i| - 1 \right]^2 x dx$$

$$= \mu h^3 \int_{-1}^{0} y (y + 1)^2 (y + i) dy + \mu h^3 \int_{0}^{1} y (y - 1)^2 (y + i) dy$$

$$= \mu h^3 \int_0^1 (y(y-1)^2(y-i) + y(y-1)^2(y+i))dy$$

$$= 2\mu h^3 \left[\frac{1}{5}y^5 - \frac{2}{4}y^4 + \frac{1}{3}y^3 \Big|_0^1 \right] = \frac{\mu h^3}{15}. \qquad (8.3.23)$$

For $i = 2^q$,

$$G_{2i} = \mu h^3 \left[\int_0^1 y^2(y-1)^2 dy - i \int_0^1 y(y-1)^2 dy \right]$$

$$= \frac{\mu h^3}{60}(2 - 5i). \qquad (8.3.24)$$

Time Scaling:

By appropriate choice of A, B, C, D, and F, (8.3.7) and (8.1.4) can be rewritten:

$$AW_{tt}(t) + BW_t(t) + CW(t) + D\Omega^2(t)W(t) = Fu(t),\ t \in [0,T], \qquad (8.3.25)$$

$$\Omega_t(t) = \mu u(t),\ t \in [0,T] \qquad (8.3.26a)$$

$$\Theta_t(t) = \Omega(t),\ t \in [0,T] \qquad (8.3.26b)$$

with initial conditions:

$$W(0) = W_t(0) = 0, \quad \Theta(0) = \Omega(0) = 0. \qquad (8.3.27)$$

We introduce an additional state $z(t)$ such that

$$z_t(t) = 0, \quad z(0) = T. \qquad (8.3.28)$$

We define $\overline{W}(t) \triangleq W(tT)$, $\overline{u}(t) \triangleq u(tT)$, $\overline{\Omega} \triangleq \Omega(tT)$, $\overline{\Theta}(t) \triangleq \Theta(tT)$ and $z^0(t) \triangleq z(tT)$ for $t \in [0,1]$ as in Section 4.2. Then (8.3.25) - (8.3.28) can be rewritten:

$$A\overline{W}_{tt}(t) + Bz^0(t)\overline{W}_t(t) + Cz^0(t)^2\overline{W}(t) + Dz^0(t)^2\Omega^2(t)\overline{W}(t) = Fz^0(t)^2\overline{u}(t),\ t \in [0,1], \qquad (8.3.29)$$

$$\overline{\Omega}_t(t) = z^0(t)\overline{u}(t), t \in [0,1], \qquad (8.3.30a)$$

$$\overline{\Theta}_t(t) = z^0(t)\overline{\Omega}(t), t \in [0,1], \quad (8.3.30b)$$

$$z_t^0(t) = 0, t \in [0,1]. \quad (8.3.31)$$

We introduce additional states $z^1(\cdot)$, $z^2(\cdot)$:

$$z_t^1(t) = z^0(t)u(t)^2, \; z^1(0) = 0. \quad (8.3.32)$$

$$z_t^2(t) = \max\{ P(t,u) - f(t), 0 \}^2$$

$$= \max\{ \frac{EI}{2}\overline{W}(t)^T\mathbf{K}\overline{W}(t) - f(t), 0 \}^2, \; z^2(0) = 0. \quad (8.3.33)$$

We define the state $y(t) \in \mathbb{R}^Q \times \mathbb{R}^Q \times \mathbb{R}^5$:

$$y(t) = \begin{bmatrix} \overline{W}(t) \\ \overline{W}_t(t) \\ \overline{\Theta}(t) \\ \overline{\Omega}(t) \\ z^0(t) \\ z^1(t) \\ z^2(t) \end{bmatrix}, \quad (8.3.34)$$

and denoting the dependence of $y(t)$ on the control and initial state ω, y(t) is written $y(t,u,\omega)$. We define the functions $\overline{g}^j(u,\omega)$, $j \in \{ 1,2,...7 \}$ in terms of the state variables defined in (8.3.34):

$$\overline{g}_q^1(\overline{u},\omega) \triangleq z^0(1). \quad (8.3.35)$$

$$\overline{g}_q^2(\overline{u},\omega) \triangleq z^1(1). \quad (8.3.36)$$

$$\overline{g}_q^3(\overline{u},\omega) \triangleq (\overline{\Theta}(1) - \frac{\pi}{4})^2. \quad (8.3.37)$$

$$\overline{g}_q^4(\overline{u},\omega) \triangleq \frac{1}{z^0(1)^2}\overline{\Omega}(1)^2. \quad (8.3.38)$$

$$\overline{g}_q^5(\overline{u},\omega) \triangleq \frac{m}{2z^0(1)^2}\overline{W}_t(1)^T\mathbf{M}\overline{W}_t(1). \quad (8.3.39)$$

$$\overline{g}_q^6(\overline{u},\omega) \triangleq \frac{EI}{2}\overline{W}(1)^T\mathbf{K}\overline{W}(1). \quad (8.3.40)$$

$$\overline{g}_q^7(\overline{u},\omega) \triangleq z^2(1). \quad (8.3.41)$$

We define the set $C_q \subset \mathbb{R}^{\varrho} \times \mathbb{R}^{\varrho} \times \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^2$:

$$C_q \triangleq 0 \times 0 \times 0 \times [\tau_{\min}, \tau_{\max}] \times 0. \tag{8.3.42}$$

With these definitions, $\overline{\mathbf{P}}_1^q$ - $\overline{\mathbf{P}}_4^q$ are of the form:

$$P: \inf_{\overline{u} \in G, \omega \in C_q} \{ \overline{g}_q^k(\overline{u}, \omega) \mid g_q^j(\overline{u}, \omega) \le 0, j \in J \} \tag{8.3.43}$$

where $J \subset \underline{p}$, $k \in \underline{p}$ and

$$\frac{d}{dt} y(t,u,\omega) = f(y(t,u,\omega), u(t)),\ y(0,u,\omega) = \omega. \tag{8.3.44}$$

8.4 TEMPORAL DISCRETIZATION AND CALCULATION OF GRADIENTS

In Section 8.3, we derived problems $\overline{\mathbf{P}}_1^q$ - $\overline{\mathbf{P}}_4^q$ from $\mathbf{P}_1$ - $\mathbf{P}_4$ by discretizing the system dynamics (8.1.1)-(8.1.3) in space using Galerkin's method with Hermite splines as basis functions. In this section, we will discretize the ordinary differential equations (8.3.29)-(8.3.33) in time to obtain a set of finite difference equations. Let the number of temporal discretization points be a function of an integer n. Then, problems $\hat{\mathbf{P}}_1^{n,q}$ - $\hat{\mathbf{P}}_4^{n,q}$ are derived from $\overline{\mathbf{P}}_1^q$ - $\overline{\mathbf{P}}_1^q$ by replacing the final state $y(1,u,\omega)$ in $\overline{g}_q^j(\overline{u},\omega)$, $j \in \{1,2,\cdots,7\}$ (8.3.35)-(8.3.41) by the final state of the finite difference equation (8.4.1)-(8.4.6) resulting in (8.4.7)-(8.4.13). Finally, we derive expressions for the gradients, $\nabla\hat{g}_{n,q}^j(\overline{u},\omega)$, $j \in \{1,2,\cdots,7\}$.

For $n \in \mathbb{Z}_+$, we discretize time into $N \triangleq 2^n$ intervals of equal size and consider as inputs $\overline{u} \in G_n$ (6.5.3). Let $u_j \triangleq \overline{u}(jh)$ and $f_j \triangleq f(jh)$ for $j \in \{0,1,\cdots,N\}$ and $u_{N+1} = 0$. To solve (8.3.29), we use Newmark's method[New.1] with $\beta = \frac{1}{4}$. Newmark's method is an implicit method, and is ideally suited for systems described by second order dynamics. We approximate $\overline{w}(jh)$ by d_j, $\overline{\Omega}(jh)$ by Ω_j, $\overline{\Theta}(jh)$ by Θ_j, $z^0(jh)$ by T_j, $z^1(jh)$ by z_j^1, $z^2(jh)$ by z_j^2, for $j \in \{0,1,\cdots,N\}$ where $\{d_j,\Omega_j,\Theta_j,T_j,z_j^1,z_j^2\}$ satisfies

$$\frac{A}{T_j^2}D_tD_{\bar{t}}d_{j+1} + \frac{B}{2T_j}(D_t + D_{\bar{t}})d_{j+1} + (I + \frac{h^2}{4}D_tD_{\bar{t}})Cd_{j+1}$$

$$+ \Omega_j^2(I + \frac{h^2}{4}D_tD_{\bar{t}})Dd_{j+1} = F(I + \frac{h^2}{4}D_tD_{\bar{t}})u_{j+1}, \tag{8.4.1}$$

$$\Omega_{j+1} = \Omega_j + \mu\, hT_ju_{j+1}, \tag{8.4.2}$$

$$\Theta_{j+1} = \Theta_j + hT_j\Omega_j, \tag{8.4.3}$$

$$T_{j+1} = T_j, \tag{8.4.4}$$

We recall that $\mu = \dfrac{1}{M + \frac{m}{3}}$.

$$z^1_{j+1} = z^1_j + hT_j\mu^2_{j+1}\,, \tag{8.4.5}$$

$$z^2_{j+1} = z^2_j + 10^4\, h \max\{\ \frac{EI}{2}\langle\, d_j, \mathrm{K}d_j \,\rangle - f_j, 0\ \}^2, \tag{8.4.6}$$

for $j \in \{\ 0, 1, ...N-1\ \}$ with $D_tD_{\bar{t}}$ and $(D_t + D_{\bar{t}})$ defined as in (6.5.5) and (6.5.6). For simplicity, we drop the bar on $\overline{u} \in G_n$. Therefore, for $u \in G_n$, we define the functions $\tilde{g}^j_{n,q}(u, \omega)$, $j \in \{\ 1, 2, \cdots, 7\ \}$ in terms of the state variables from the finite difference equations (8.4.1)-(8.4.6):

$$\tilde{g}^1_{n,q}(u, \omega) \triangleq T_N. \tag{8.4.7}$$

$$\tilde{g}^2_{n,q}(u, \omega) \triangleq z^1_N. \tag{8.4.8}$$

$$\tilde{g}^3_{n,q}(u, \omega) \triangleq (\Theta_N - \frac{\pi}{4})^2. \tag{8.4.9}$$

$$\tilde{g}^4_{n,q}(u, \omega) \triangleq \frac{1}{T_N^2}\Omega_N^2. \tag{8.4.10}$$

$$\tilde{g}^5_{n,q}(u, \omega) \triangleq \frac{m}{2}(d_{N+1} - d_N)^T\mathrm{M}(d_{N+1} - d_N). \tag{8.4.11}$$

$$\tilde{g}^6_{n,q}(u, \omega) \triangleq \frac{EI}{2}d_N^T\mathrm{K}d_N. \tag{8.4.12}$$

$$\tilde{g}^7_{n,q}(u, \omega) \triangleq z^2_N. \tag{8.4.13}$$

We shall sketch the derivation of $\nabla\tilde{g}^j_{n,q}(\overline{u}, \omega)$ for $j \in \{\ 1, 2, \cdots, 7\ \}$. We first take the variation of equations (8.4.1) - (8.4.6):

$$\frac{A}{T_j^2}D_tD_{\bar{t}}\delta d_{j+1} - \frac{2A}{T_j^3}D_tD_{\bar{t}}d_{j+1}\delta T_j + \frac{B}{2T_j}(D_t + D_{\bar{t}})\delta d_{j+1} - \frac{B}{2T_j^2}(D_t + D_{\bar{t}})d_{j+1}\delta T_{j+1}$$

$$+ (I + \frac{h^2}{4}D_tD_{\bar{t}})C\delta d_{j+1} + \Omega_j^2(I + \frac{h^2}{4}D_tD_{\bar{t}})D\delta d_{j+1} + 2\Omega_j\left[(I + \frac{h^2}{4}D_tD_{\bar{t}})Dd_{j+1}\right]\delta\Omega_j$$

$$= F(I + \frac{h^2}{4}D_tD_{\bar{t}})\delta u_{j+1}\,,\ \delta d_0 = \delta d_1 = 0\,, \tag{8.4.14}$$

$$\delta\Omega_{j+1} = \delta\Omega_j + \mu\, hT_j\delta u_{j+1} + \mu\, hu_{j+1}\delta T_j\,,\ \delta\Omega_0 = 0\,, \tag{8.4.15}$$

$$\delta\Theta_{j+1} = \delta\Theta_j + hT_j\delta\Omega_j + h\Omega_j\delta T_j,\ \delta\Theta_0 = 0, \tag{8.4.16}$$

$$\delta T_{j+1} = \delta T_j,\ \delta T_0 = \delta T, \tag{8.4.17}$$

$$\delta z^1_{j+1} = \delta z^1_j + 2hT_j u_{j+1}\delta u_{j+1} + hu^2_{j+1}\delta T_j,\ \delta z^1_0 = 0, \tag{8.4.18}$$

$$\delta z^2_{j+1} = \delta z^2_j + 2\times 10^4 h \max\{ \frac{EI}{2} \langle d_j, \mathrm{K}d_j \rangle - f_j, 0 \} EId_j^T\mathrm{K}\delta d_j,\ \delta z^2_0 = 0, \tag{8.4.19}$$

for $j \in \{0, 1, \ldots N-1\}$ where $\delta u \in R_n$, $\delta u_j \triangleq \delta u(jh)$ for $j \in \{0, 1, \cdots, N\}$ and $\delta u_{N+1} = 0$. Taking the variation of equations (8.4.7)-(8.4.13):

$$\delta\hat{g}^1_{n,q}(u, \omega, \delta u, \delta\omega) = \delta T_N. \tag{8.4.20}$$

$$\delta\hat{g}^2_{n,q}(u, \omega, \delta u, \delta\omega) = \delta z^1_N. \tag{8.4.21}$$

$$\delta\hat{g}^3_{n,q}(u, \omega, \delta u, \delta\omega) = 2(\Theta_N - \frac{\pi}{4})\delta\Theta_n. \tag{8.4.22}$$

$$\delta\hat{g}^4_{n,q}(u, \omega, \delta u, \delta\omega) = -\frac{2}{T_N^3}\Omega_N^2\delta T_N + \frac{2}{T_N^2}\Omega_N\delta\Omega_N \tag{8.4.23}$$

$$\delta\hat{g}^5_{n,q}(u, \omega, \delta u, \delta\omega) = m(d_{N+1} - d_N)^T\mathrm{M}(\delta d_{N+1} - \delta d_N). \tag{8.4.24}$$

$$\delta\hat{g}^6_{n,q}(u, \omega, \delta u, \delta\omega) = EId_N^T\mathrm{K}\delta d_N. \tag{8.4.25}$$

$$\delta\hat{g}^7_{n,q}(u, \omega, \delta u, \delta\omega) = 2z_N\delta z_N. \tag{8.4.26}$$

Expanding (8.4.14) and noting that δT_j for all values of $j \in \{0, 1, \cdots, N\}$ are equal:

$$\delta d_{j+2} + \bar{B}_j\delta d_{j+1} + \bar{C}_j\delta d_j = \bar{D}_j d_{j+1}\delta T_j + \bar{E}_j\delta\Omega_j + \bar{F}_j\delta u_{j+1}, \tag{8.4.27}$$

where

$$\bar{A}_j \triangleq \frac{A}{T_j^2} + \frac{Bh}{2T_j} + \frac{h^2}{4}(C + D\Omega_j^2), \qquad \bar{B}_j = \bar{A}_j^{-1}(-\frac{2A}{T_j^2} + \frac{h^2}{2}(C + D\Omega_j^2)) \tag{8.4.28a}$$

$$\bar{C}_j \triangleq \bar{A}_j^{-1}(\frac{A}{T_j^2} - \frac{Bh}{2T_j} + \frac{h^2}{4}(C + D\Omega_j^2),\ \bar{D}_j \triangleq \bar{A}_j^{-1}h^2(\frac{2A}{T_j^3}D_tD_{\bar{t}} + \frac{B}{2T_j^2}(D_t + D_{\bar{t}})) \tag{8.4.28b}$$

$$\bar{E}_j \triangleq \bar{A}_j^{-1}h^2(-2\Omega_j(I + \frac{h^2}{4}D_tD_{\bar{t}})Dd_j), \qquad \bar{F}_j \triangleq \bar{A}_j^{-1}h^2F(I + \frac{h^2}{4}D_tD_{\bar{t}}) \tag{8.4.28c}$$

In order to derive $\nabla\hat{g}^j_{n,q}(\cdot,\cdot)$, we recast the dynamics (8.4.1)-(8.4.6) into first order form. We define

$$\delta d_j^1 \triangleq \delta d_j, \quad \delta d_j^2 \triangleq \delta d_{j+1}, \tag{8.4.29}$$

for $j \in \{ 0, 1, \cdots, N \}$. Then system (8.4.27), (8.4.2)-(8.4.6) can be written in matrix form:

$$\delta y_{j+1} = \mathbf{A}_j \delta y_j + \mathbf{B}_j \delta u_{j+1}, \tag{8.4.30}$$

for $j \in \{ 1, 2, \cdots, N-1 \}$ where

$$\delta y_j = \begin{bmatrix} \delta d^1 \\ \delta d^2 \\ \delta\Omega \\ \delta\Theta \\ \delta T \\ \delta z^1 \\ \delta z^2 \end{bmatrix}_j, \quad \mathbf{A}_j \triangleq \begin{bmatrix} 0 & I & 0 & 0 & 0 & 0 & 0 \\ -\overline{C}_j & -\overline{B}_j & \overline{E}_j & 0 & \overline{D}_j & 0 & 0 \\ 0 & 0 & \mu & 0 & \mu h u_{j+1} & 0 & 0 \\ 0 & 0 & hT_j & 1 & h\Omega_j & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & hu_j^2 & 1 & 0 \\ G_j & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B}_j \triangleq \begin{bmatrix} 0 \\ \overline{F}_j \\ hT_j \\ 0 \\ 0 \\ 2hT_j u_{j+1} \\ 0 \end{bmatrix} \tag{8.4.31}$$

and

$$G_j = 2 \times 10^4 h \max\{ \frac{EI}{2} \langle d_j, \mathbf{K} d_j \rangle - f_j, 0 \} EI d_j^T \mathbf{K} \tag{8.4.32}$$

With this notation, (8.4.25) and (8.4.25) become

$$\delta \hat{g}^5_{n,q}(u, \omega, \delta u, \delta\omega) = m(d_{N+1} - d_N)^T \mathbf{M}(\delta d_N^2 - \delta d_N^1), \tag{8.4.35}$$

$$\delta \hat{g}^6_{n,q}(u, \omega, \delta u, \delta\omega) = EI d_N^T \mathbf{K} \delta d_N^1. \tag{8.4.36}$$

We observe that $\delta \hat{g}^j_{n,q}(u, \omega, \delta u, \delta\omega)$ is a linear function of the final state, y_N (8.4.2). We can therefore write $\delta \hat{g}^j_{n,q}(u, \omega, \delta u, \delta\omega) = \langle \delta \tilde{g}^j_{n,q}(u, \omega), \delta y_N \rangle$ where $\delta \tilde{g}^j_{n,q}(u, \omega) \in \mathbb{R}^{2Q+5}$.

Since the matrices $\{ \mathbf{A}_j \}_{j \in \{0, 1, \ldots N\}}$ are not commutative, we shall define left and right products: For $u \geq l$, the left and right products are defined:

$$\prod_{i=l}^{u}{}^{L} \mathbf{A}_i \triangleq \mathbf{A}_u \mathbf{A}_{u-1} \cdots \mathbf{A}_{l+1}\mathbf{A}_l, \quad \prod_{i=l}^{u}{}^{R} \mathbf{A}_i \triangleq \mathbf{A}_l \mathbf{A}_{l+1} \cdots \mathbf{A}_{u-1}\mathbf{A}_u. \tag{8.4.33}$$

Since $\delta y_{j+1} = \mathbf{A}_j \delta y_j + \mathbf{B}_j u_{j+1}$ for $j \in \{ 0, 1, \cdots, N-1 \}$ and $\delta y_0 = \delta\omega$,

$$\delta y_N = \prod_{i=0}^{N-1}{}^{L} \mathbf{A}_j \delta\omega + \sum_{k=1}^{N} \left[\prod_{i=k}^{N-1}{}^{L} \mathbf{A}_i \right] \mathbf{B}_{k-1} u_k. \tag{8.4.34}$$

We derive expressions for $\nabla \hat{g}^j_{n,q}(\cdot, \cdot)$ by noting that for $(\delta u, \delta\omega) \in (G_n - u) \times (C - \omega)$,

$$\langle \nabla \hat{g}^j_{n,q}(u, \omega), \begin{bmatrix} \delta u \\ \delta\omega \end{bmatrix} \rangle = \delta \hat{g}^j_{n,q}(u, \omega, \delta u, \delta\omega) = \langle \delta \hat{g}^j_{n,q}(u, \omega), \delta y_N \rangle \tag{8.4.37}$$

where δy_N is the solution to (8.4.30). We shall drop the j superscript for readability in what follows. Combining (8.4.34) and (8.4.37),

$$\langle \nabla \hat{g}_{n,q}(u, \omega), \begin{bmatrix} \delta u \\ \delta\omega \end{bmatrix} \rangle = \langle \delta \hat{g}_{n,q}(u, \omega, \delta u, \delta\omega), \prod_{i=0}^{N-1}{}^{L} \mathbf{A}_j \delta\omega + \sum_{k=1}^{N} \left[\prod_{i=k}^{N-1}{}^{L} \mathbf{A}_i \right] \mathbf{B}_{k-1} u_k \rangle.$$

$$= \langle \left[\prod_{i=0}^{N-1}{}^{R} \mathbf{A}_i^T \right] \delta \hat{g}_{n,q}(u, \omega, \delta u, \delta\omega), \delta\omega \rangle$$

$$+ \sum_{k=1}^{N} \langle \prod_{i=k}^{N-1}{}^{R} \mathbf{A}_i^T \delta \hat{g}_{n,q}(u, \omega, \delta u, \delta\omega), \mathbf{B}_{k-1} \delta u_k \rangle. \tag{8.4.38}$$

If we define

$$p_k \triangleq \left[\prod_{i=k}^{N-1}{}^{R} \mathbf{A}_i^T \right] \delta \hat{g}_{n,q}(u, \omega, \delta u, \delta\omega) \tag{8.4.39}$$

for $k \in \{ 0, 1, \cdots N-1 \}$ then $p_k = \mathbf{A}_k^T p_{k+1}$ and $p_N = \delta \hat{g}_{n,q}(u, \omega, \delta u, \delta\omega)$. We define $\nabla_\omega \hat{g}_{n,q}(u, \omega)$ and $\nabla_u \hat{g}_{n,q}(u, \omega)$ such that

$$\langle \nabla \hat{g}_{n,q}(\cdot, \cdot), \begin{bmatrix} \delta u \\ \delta\omega \end{bmatrix} \rangle = \langle \nabla_\omega \hat{g}_{n,q}(u, \omega), \delta\omega \rangle + \langle \nabla_u \hat{g}_{n,q}(u, \omega), \delta u \rangle. \tag{8.4.40}$$

Then by (8.4.38) and (8.4.39), $\nabla_\omega \hat{g}_{n,q}(u, \omega) = p_0$ and

$$\langle \nabla_u \hat{g}_{n,q}(u,\omega), \delta u \rangle = \sum_{k=1}^{N} \langle p_k, \mathbf{B}_{k-1}\delta u_k \rangle = \sum_{k=1}^{N} \langle p_k, \begin{bmatrix} 0 \\ \bar{F}_{k-1} \\ hT_{k-1} \\ 0 \\ 0 \\ 2hT_{k-1}u_k \\ 0 \end{bmatrix} \delta u_k \rangle$$

$$= \sum_{k=1}^{N} \langle diag(I, \mathbf{A}_{k-1}^{-T}, 1, 1, 1, 1, 1)p_k, \begin{bmatrix} 0 \\ \frac{h^2}{4}\delta u_{k+1} + \frac{h^2}{2}\delta u_k + \frac{h^2}{4}\delta u_{k-1} \\ hT_{k-1}\delta u_k \\ 0 \\ 0 \\ 2hT_{k-1}u_k\delta u_k \\ 0 \end{bmatrix} \rangle. \tag{8.4.41}$$

We partition p_k for $k \in \{ 0, 1, \cdots, N \}$:

$$p_k = \begin{bmatrix} p^1 \\ p^2 \\ p^{\Omega} \\ p^{\Theta} \\ p^T \\ p^{z^1} \\ p^{z^2} \end{bmatrix}_k \tag{8.4.42}$$

where p_k^1 and $p_k^2 \in \mathbb{R}^Q$. To simplify (8.4.41), we define

$$\lambda_k^2 \triangleq \mathbf{A}_{k-1}^{-T} p_k^2. \tag{8.4.43}$$

Then

$$\langle \nabla_u \hat{g}_{n,q}(u,\omega), \delta u \rangle = \sum_{k=1}^{N} \begin{bmatrix} 0 \\ \frac{h^2}{4}\lambda_k^2\delta u_{k+1} + \frac{h^2}{2}\lambda_k^2\delta u_k + \frac{h^2}{4}\lambda_k^2\delta u_{k-1} \\ p_k^{\Omega} hT_{k-1}\delta u_k \\ 0 \\ 0 \\ p_k^{z^1} 2hT_{k-1}u_k\delta u_k \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ \frac{h^2}{2}\lambda_1^2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \delta u_0 + \begin{bmatrix} 0 \\ \frac{h^2}{2}\lambda_1^2 + \frac{h^2}{4}\lambda_2^2 \\ p_1^{\Omega} h T_0 \\ 0 \\ 0 \\ 2 p_1^{z^2} h T_0 u_1 \\ 0 \end{bmatrix} \delta u_1 +$$

$$\sum_{k=2}^{N-1} \begin{bmatrix} 0 \\ \frac{h^2}{4}\lambda_{k-1}^2 + \frac{h^2}{2}\lambda_k^2 + \frac{h^2}{4}\lambda_{k+1}^2 \\ p_k^{\Omega} h T_{k-1} \\ 0 \\ 0 \\ 2 p_k^{z^2} h T_{k-1} u_k \\ 0 \end{bmatrix} \delta u_k + \begin{bmatrix} 0 \\ \frac{h^2}{2}\lambda_{N-1}^2 + \frac{h^2}{4}\lambda_N^2 \\ p_N^{\Omega} h T_{N-1} \\ 0 \\ 0 \\ 2 p_N^{z^2} h T_{N-1} u_N \\ 0 \end{bmatrix} \delta u_N . \tag{8.4.44}$$

Since

$$\langle \nabla_u \hat{g}_{n,q}(u, \omega), \delta u \rangle = \sum_{k=0}^{N} \langle \nabla_u \hat{g}_{n,q}(u, \omega)_k, \delta u_k \rangle , \tag{8.4.45}$$

where $\nabla_u \hat{g}_{n,q}(u, \omega)(t) = \nabla_u \hat{g}_{n,q}(u, \omega)_j$ if $t \in [jh, (j+1)h)$, we can readily determine $\nabla_u \hat{g}_{n,q}(u, \omega)$by (8.4.44).

Calculation of p_k:

By (8.4.39), we can write a recurrence relation:

$$p_k = \begin{bmatrix} p^1 \\ p^2 \\ p^{\Omega} \\ p^{\Theta} \\ p^T \\ p^{z^1} \\ p^{z^2} \end{bmatrix}_k = \begin{bmatrix} 0 & -\overline{C}_k & 0 & 0 & 0 & 0 & G_k^T \\ I & -\overline{B}_k & 0 & 0 & 0 & 0 & 0 \\ 0 & \overline{E}_k & \mu & hT_k & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \overline{D}_k & \mu\, h u_{k+1} & h\Omega_k & 1 & h u_k^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p^1 \\ p^2 \\ p^{\Omega} \\ p^{\Theta} \\ p^T \\ p^{z^1} \\ p^{z^2} \end{bmatrix}_{k+1} . \tag{8.4.46}$$

Substituting $p_{k+1}^2 = A_k^T \lambda_{k+1}^2$ into (8.4.46), we obtain:

$$\begin{bmatrix} p_k^1 \\ A_{k-1}^T \lambda_k^2 \\ p_k^{\Omega} \\ p_k^{\Theta} \\ p_k^T \\ p_k^{z^1} \\ p_k^{z^2} \end{bmatrix} = \mathbf{C} \begin{bmatrix} p_{k+1}^1 \\ \lambda_{k+1}^2 \\ p_{k+1}^{\Omega} \\ p_{k+1}^{\Theta} \\ p_{k+1}^T \\ p_{k+1}^{z^1} \\ p_{k+1}^{z^2} \end{bmatrix}, \tag{8.4.47}$$

where

$$\mathbf{C} \triangleq \begin{bmatrix} 0 & -(\frac{A}{T_k^2} - \frac{Bh}{2T_k} + \frac{h^2}{4}(C + \Omega_k^2 D))^T & 0 & 0 & 0 & 0 & G_k^T \\ I & -(-\frac{2A}{T_k^2} + \frac{h^2}{2}(C + \Omega_k^2 D))^T & 0 & 0 & 0 & 0 & 0 \\ 0 & h^2(-2\Omega_k(I + \frac{h^2}{4}D_t D_t^T)D d_k^1)^T & \mu & \mu h T_k & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & h^2(\frac{2A}{T_k^3}D_t D_t^T + \frac{B}{2T_k^2}D_t D_t^T)d_{k+1}^1 & h u_{k+1} & h\Omega_k & 1 & h u_k^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8.4.48}$$

Equations (8.4.47) and (8.4.44) can be used to calculate $\nabla_u \hat{g}_{n,q}(u, \omega)(t)$.

8.5. REFINEMENT CRITERION

In sections 8.3 and 8.4, we have presented methods to discretize the PDE dynamics in time and space. The resulting cost and constraint functions are denoted $\hat{g}^j_{n,q} : G_n \times C_q \to \mathbb{R}$ (8.4.7)-(8.4.13)[1] . To present an implementable algorithm, we make the following definitions:

$$\hat{\Theta}_{n,q} : G_n \times C_q \to \mathbb{R}, \quad \hat{\Theta}_{n,q}(u, w) \triangleq \min\{ \hat{\theta}_{n,q}(v, z, u, w) \mid v \in \bar{G}_{n,q}, z \in C_{n,q} \} \quad (8$$

[1] G_n and C_q are defined in (6.5.3) and (8.3.42) respectively.

$$\hat{\phi}_{n,q} : G_n \times C_q \times G_n \times C_q \rightarrow \mathbb{R}, \tag{8.5.2}$$

$$\hat{\phi}_{n,q}(v,z,u,w) \triangleq \frac{1}{2}\|v-u\|^2 + \frac{1}{2}\|z-w\|^2 + \max_{j \in \underline{p}} \{ -\rho\hat{\psi}_{n,q}(u,w)_+ + \langle \nabla\hat{g}^0_{n,q}(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle;$$

$$\hat{g}^j_{n,q}(u,w) + \langle \nabla\hat{g}^j_{n,q}(u,w), \begin{bmatrix} v-u \\ z-w \end{bmatrix} \rangle - \psi_+(u,w) \},$$

$$\hat{\psi}_{n,q}(u,w) \triangleq \max_{j \in \underline{p}} \{ \hat{g}^j_{n,q}(u,w) \} \tag{8.5.3}$$

$$\hat{\psi}_{n,q}(u,w)_+ \triangleq \max\{ 0, \hat{\psi}_{n,q}(u,w) \} \tag{8.5.4}$$

The refinement criterion in this chapter is similar to the refinement criterion in Chapter 6. Namely, we refine the discretization if insufficient progress is being made. If the algorithm is in Phase I, this means that $\hat{\psi}_{n,q}$ is not decreasing sufficiently fast. In Phase II, this means that $\hat{g}^0_{n,q}$ is not decreasing sufficiently fast.

We present the *Implementable Algorithm* we used to solve problems $\mathbf{P}_1$ - $\mathbf{P}_4$.

Algorithm 8.5.1:

Data: $n \in \mathbb{Z}_+$, $q \le n$ such that $2^q \in \mathbb{Z}_+$. $u_0 \in G_n$, $w_0 \in C_q$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1]$, $\mu \in (0,1)$, $\rho > 0$, $\varepsilon_0 > 0$.

Step 0: Set $i = 0$, $\varepsilon = \varepsilon_0$.

Step 1: Calculate $\nabla\hat{g}^j_{n,q}(u_i)$ for $j \in \{ 0, 1, ...p \}$.

Step 2: Calculate $v_i \in G_n$ and $z_i \in C_q$ such that $\phi_n(v_i, z_i, u_i, w_i) \le \gamma\hat{\Theta}_{n,q}(u_i, w_i)$.

Step 3: If $\hat{\Theta}_{n,q}(u_i, w_i) = 0$, $\lambda_i \triangleq 0$.

Else if $\hat{\psi}_{n,q}(u_i, w_i) > 0$ (Phase I),

$\lambda_i \triangleq \max\{ \lambda \in \{ 1, \beta, \beta^2, \cdots \} \mid$

$\hat{\psi}_{n,q}(u_i + \lambda(v_i - u_i),$ $w_i + \lambda(z_i - w_i)) - \hat{\psi}_{n,q}(u_i, w_i)$

$< \alpha\lambda\phi_n(v_i, z_i, u_i, w_i)$

Else ($\hat{\psi}_{n,q}(u_i, w_i) \le 0$, (Phase II)

$\lambda_i \triangleq \max\{ \lambda \in \{ 1, \beta, \beta^2, \cdots \} \mid$

$$g^0(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) - g^0(u_i, w_i) < \alpha\lambda\phi(v_i, z_i, u_i, w_i) \text{ and}$$

$$\hat{\psi}_{n,q}(u_i + \lambda(v_i - u_i), w_i + \lambda(z_i - w_i)) \le 0.$$

Step 4: Set $u_{i+1} = u_i + \lambda_i(v_i - u_i)$, $w_{i+1} = w_i + \lambda_i(z_i - w_i)$, $n_{i+1} = n$, $q_{i+1} = q$.

Step 5: Refinement Criterion

If $\hat{\psi}_{n,q}(u_i, w_i) > 0$ (Phase I),

If $\hat{\psi}_{n,q}(u_{i+1}, w_{i+1}) - \hat{\psi}_{n,q}(u_i) > -\varepsilon$, { Set $n = n + 1, q = q + 1, \varepsilon = \mu\varepsilon$ }.

Else ($\hat{\psi}(u_i, w_i) \le 0$, (Phase II)

If $\hat{g}^0_{n,q}(u_{i+1}, w_{i+1}) - \hat{g}^0_{n,q}(u_i) > -\varepsilon$, { Set $n = n + 1, q = q + 1, \varepsilon = \mu\varepsilon$ }.

Step 6: Set $i = i + 1$; go to Step 1. ■

We have implemented Algorithm 8.5.1 on a computer and the results are displayed in the next section. We have not proven that any accumulation point $\hat{u} \in \overline{G}$ of Algorithm 8.5.1 satisfies the optimality condition $\overline{\Theta}(\hat{u}) = 0$. However, if we assume that the discretization scheme satisfies Assumption 6.2.2 (vi), the uniform approximation property, then a convergence proof can be obtained by combining Theorems 6.3.8 and 7.2.5.

8.6. COMPUTATIONAL RESULTS

This section is divided into two parts. We shall first discuss the solutions to P_1 - P_4 without reference to the specific algorithm used. In the second part, we compare the relative efficiency of two different implementations of Algorithm 8.5.1 which use Method I and

Method II in Section 7.3 to find the search direction for step 2 of Algorithm 8.5.1.

The results presented here are for the case in which the $\Omega^2(t)$ terms are neglected in equations (8.1.1) - (8.1.3). Similar results have been obtained by performing experiments for the case in which the $\Omega^2(t)$ terms are included.

PROBLEM SOLUTIONS

For all problems, we choose the zero function as initial control and 2 for an initial value for the maneuver time. These results are from [Bak.1].

Problem P_1:

Figure 8.2 is a graph of the control after 150 iterations. The number of time steps is 256 and the number of finite elements is 48.

Figure 8.3a is a graph of $\hat{\psi}_{n,q}(u,\omega) \triangleq \max \{ \tilde{g}^j_{n,q}(u,\omega) \mid j \in \{3,4,5,6\} \}$ as a function of the iteration number. Figure 8.3b shows $\hat{\psi}_{n,q}(u,\omega)$ for the first 15 iterations. The initial discretization is 32 time steps ($n = 5$) and 3 finite elements ($q = \ln_2 3$). The discretization is refined at iterations 67, 99, and 123. After precision refinement, the algorithm finds a feasible value for the control and final time for the new problem $\hat{P}_1^{n,q}$ in only a few additional iterations. At each refinement the value of $\hat{\psi}_{n,q}$ increases. This is due to improvement in the accuracy of the evaluation of the partial differential equation. This increase in $\hat{\psi}_{n,q}$ decreases each time the discretization is refined and in the limit is zero.

Figure 8.4 is the graph of the cost as a function of iteration number.

Figure 8.5 is the graph of $w(t,1)$, the displacement of the tip of the tube, from the *shadow tube*, as a function of time. There is a maximum displacement of the tip of about 5 mm. This is within the range of validity of the Euler-Bernoulli model. The tip displacement is large between 0.36 seconds and 0.437 seconds.

Figure 8.6 is a profile of the tube deformation, $w(t,x)$ (see figure 8.1), during this interval. The total time for the entire maneuver is 0.7883 seconds.

Problem P_2:

Formulating the slewing problem as a minimum time problem has two drawbacks. First, the solution to the problem is a bang-bang control (figure 8.2). Bang-bang controls may be undesirable because they may cause premature aging of the equipment. Furthermore, bang-bang controls tend to excite the high frequency modes of the system. High frequency modes are less well modeled by system (8.1.1) - (8.1.3), and it is therefore best not to excite them. Second, the simple minimum time formulation does not take into account the amount of input energy expended in performing the maneuver. In certain applications, the total input energy available may be limited, while the total time of the slewing motion is less critical. Fortunately, both of the problems arising from minimum time control can be mitigated by reformulating the problem. We minimize the total input energy while constraining the final time to be less that a specified amount.

Figure 8.7 is the graph of the control produced by minimizing the total input energy while constraining the final time to be less than 0.800 seconds. The resulting final time is 0.800 seconds. This is an increase of only 1.4 percent in the final time. The control has become much smoother and the total input energy is reduced from 19.15 to 15.72, a reduction of 18 percent.

Figure 8.8a is the graph of the control for the final time being 0.90 seconds. This is an increase of 14 percent in the time over the minimum time case, but the total input energy is reduced to 9.87, a decrease of 48 percent.

Figure 8.8b is the graph of the control for final time being 1.00 second. This is an increase of 27 percent in time over the minimum time case, but the total input energy is reduced to 7.27, a decrease of 62 percent.

Problem P_3:

In Figure 8.9, curve A is the graph of the potential energy of the tube as a function of time for the control generated in solving the minimum time problem P_1. In problem P_3, we have the additional requirement to keep the potential energy, which is a measure of the total tube deformation, below the parabola (B) for all time.

Figure 8.10 shows the optimal bang-bang control for problem P_3. The optimal final time for this case is 0.8250 seconds, an increase of 4.6 percent over the solution of problem P_1. The total input energy is 16.55. Figure 8.11 shows the potential energy curve for the optimal control (Figure 8.10).

Problem P_4:

The final time is restricted to be less than 0.90 seconds. The minimum input energy is 10.49, a decrease of 57.8 percent from P_3 Figure 8.12 shows the optimal control for P_4.

The above results are summarized in Table 8.1.

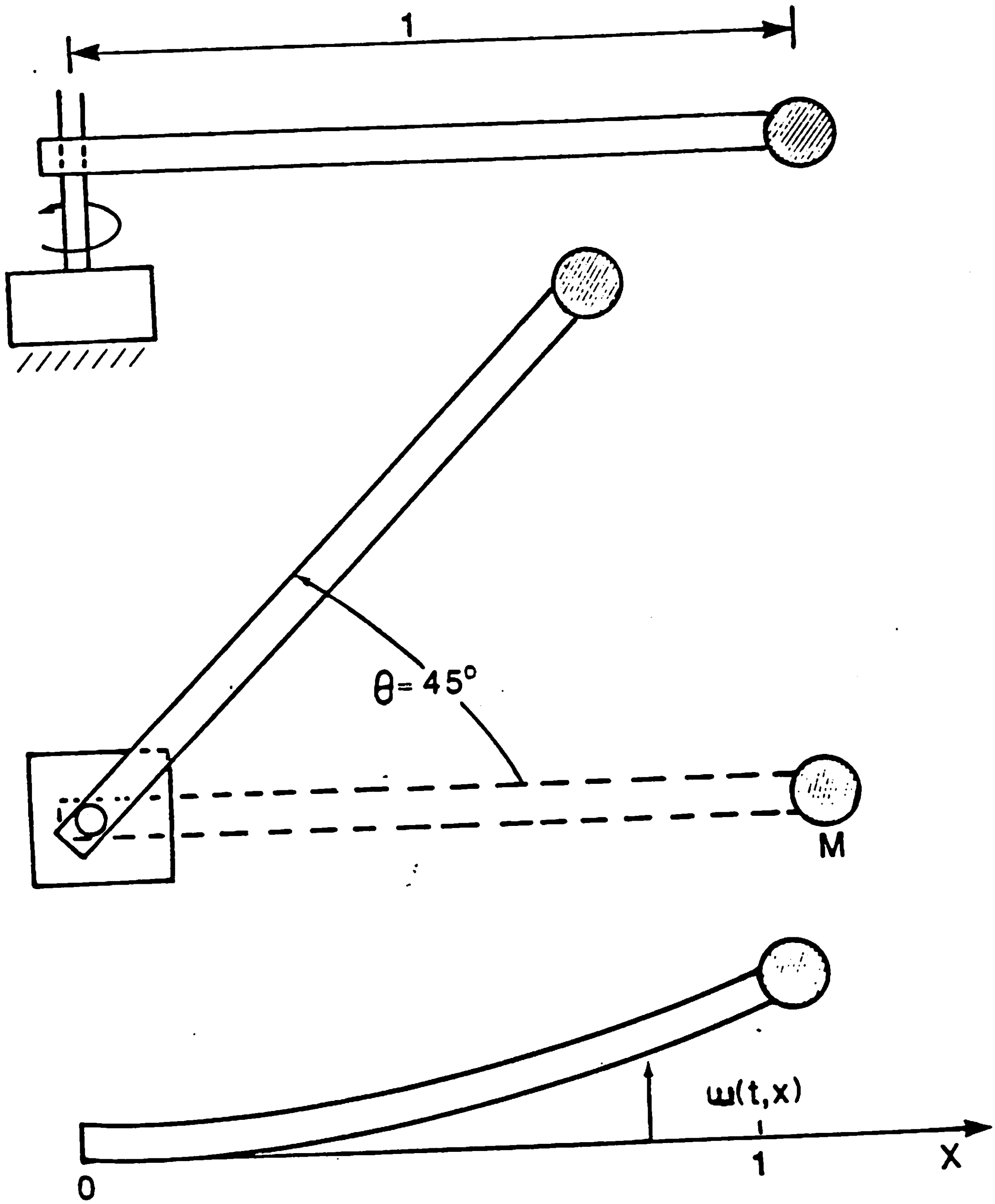

Figure 8.1 - Configuration of Slewing Experiment

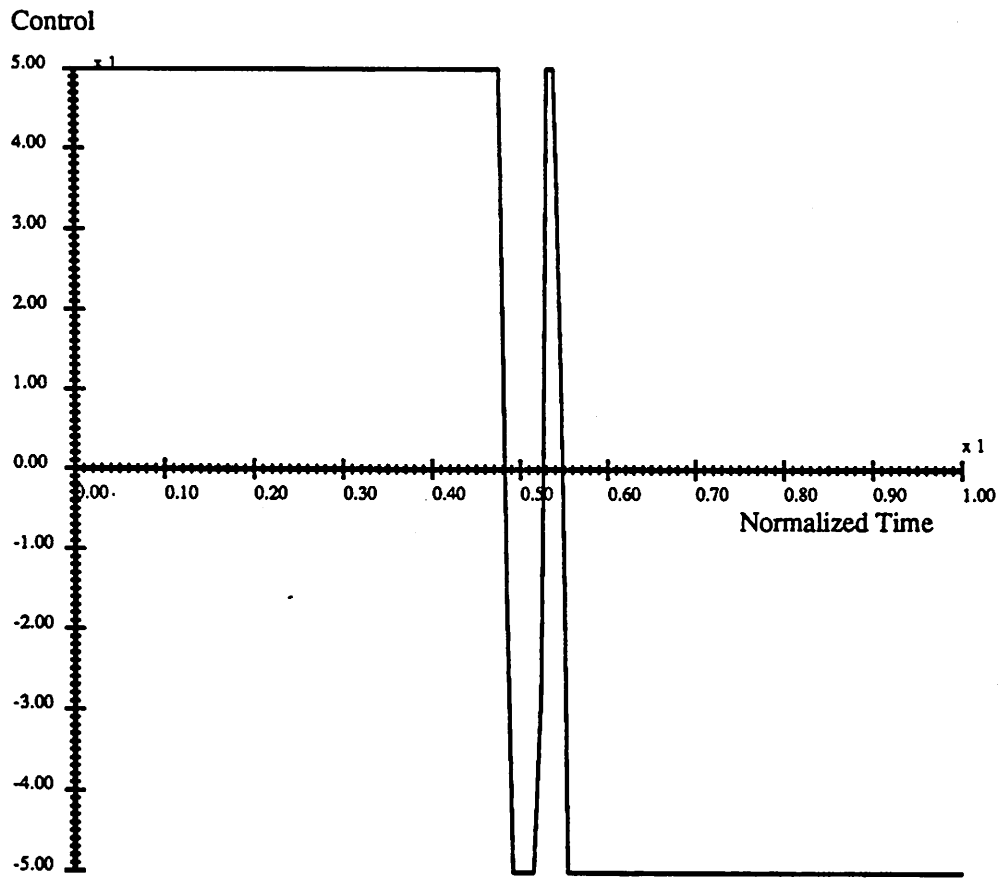

Figure 8.2 - Problem 1 Control

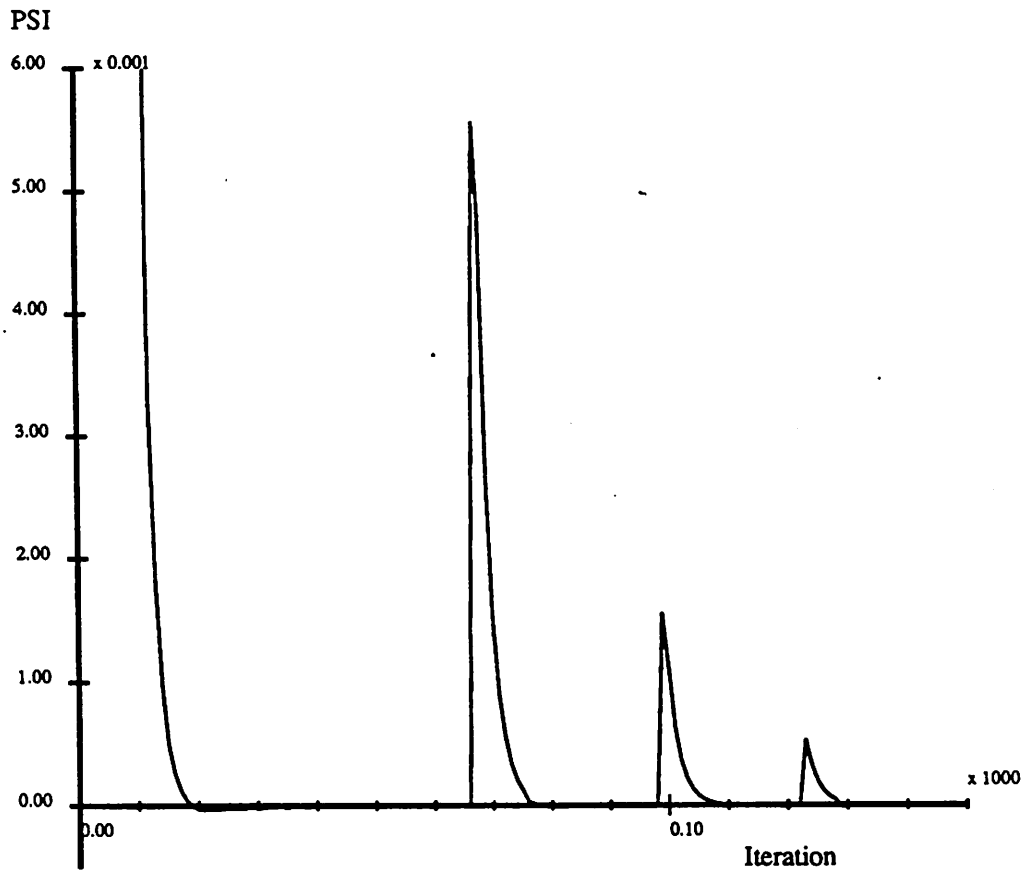

Figure 8.3a - Problem 1 PSI

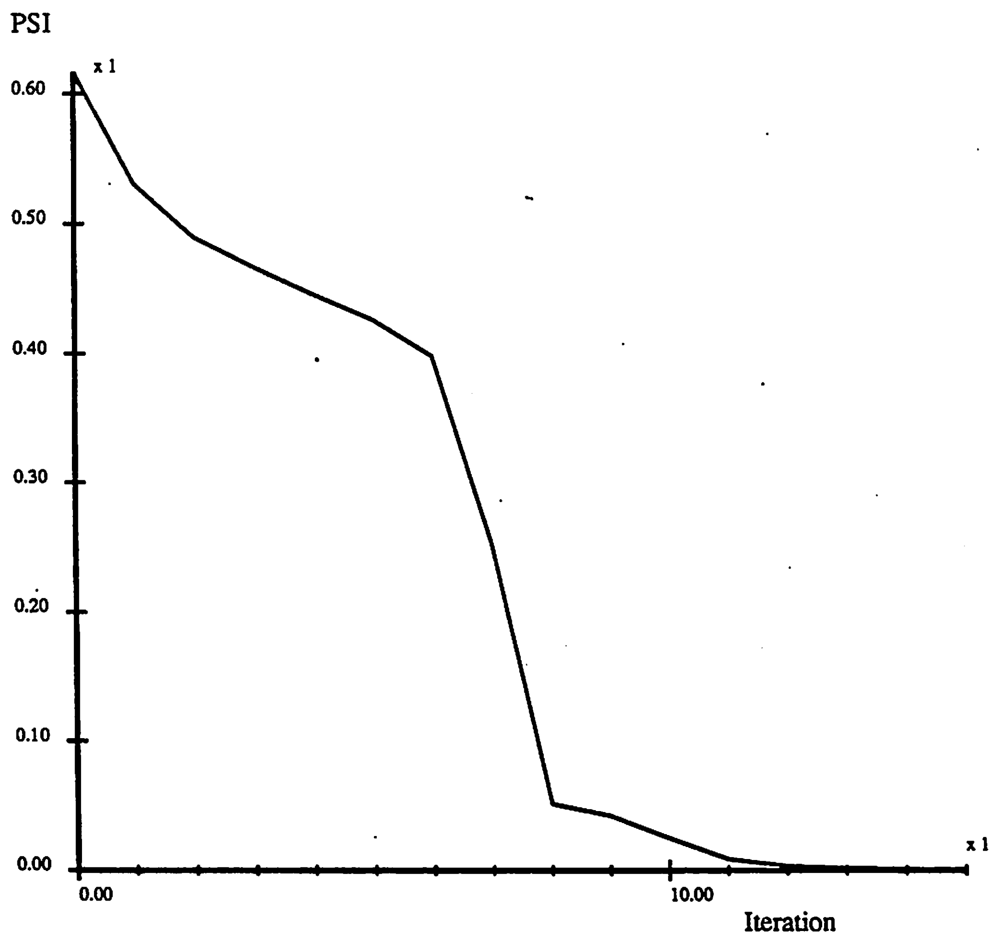

Figure 8.3b - Problem 1 PSI

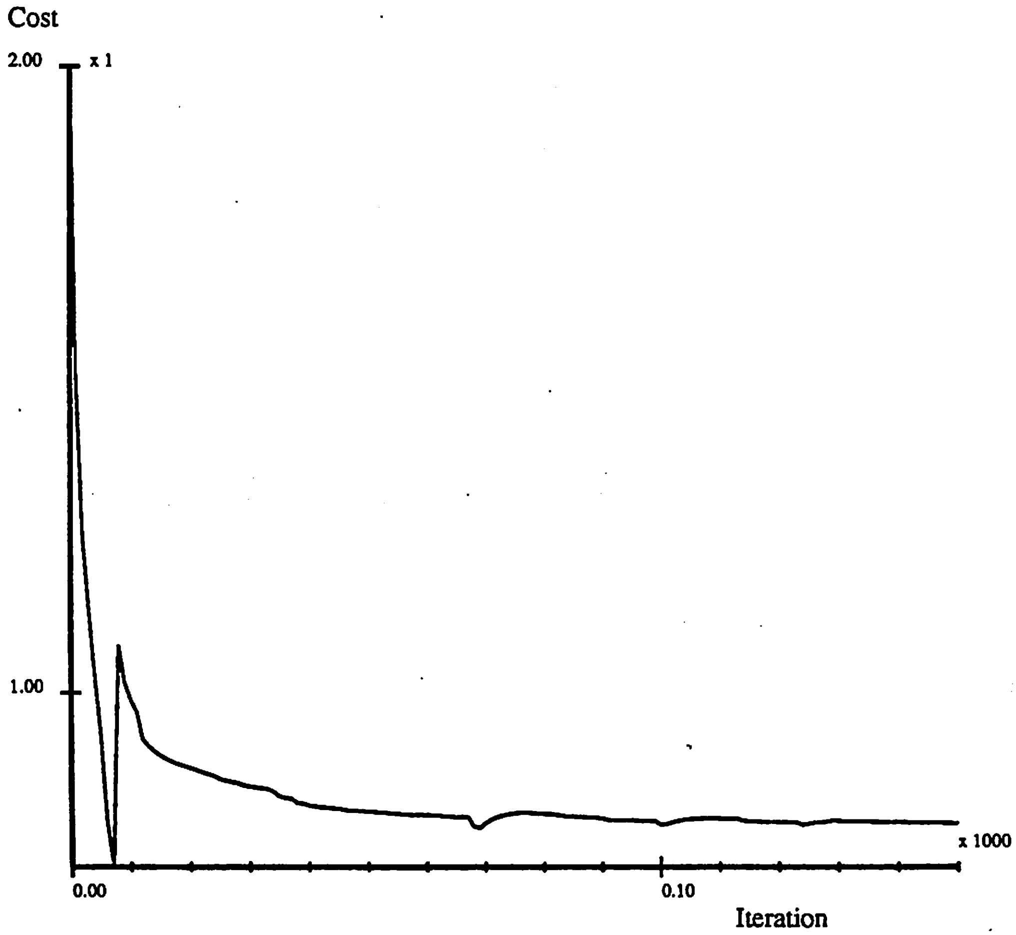

Figure 8.4 - Problem 1 Cost

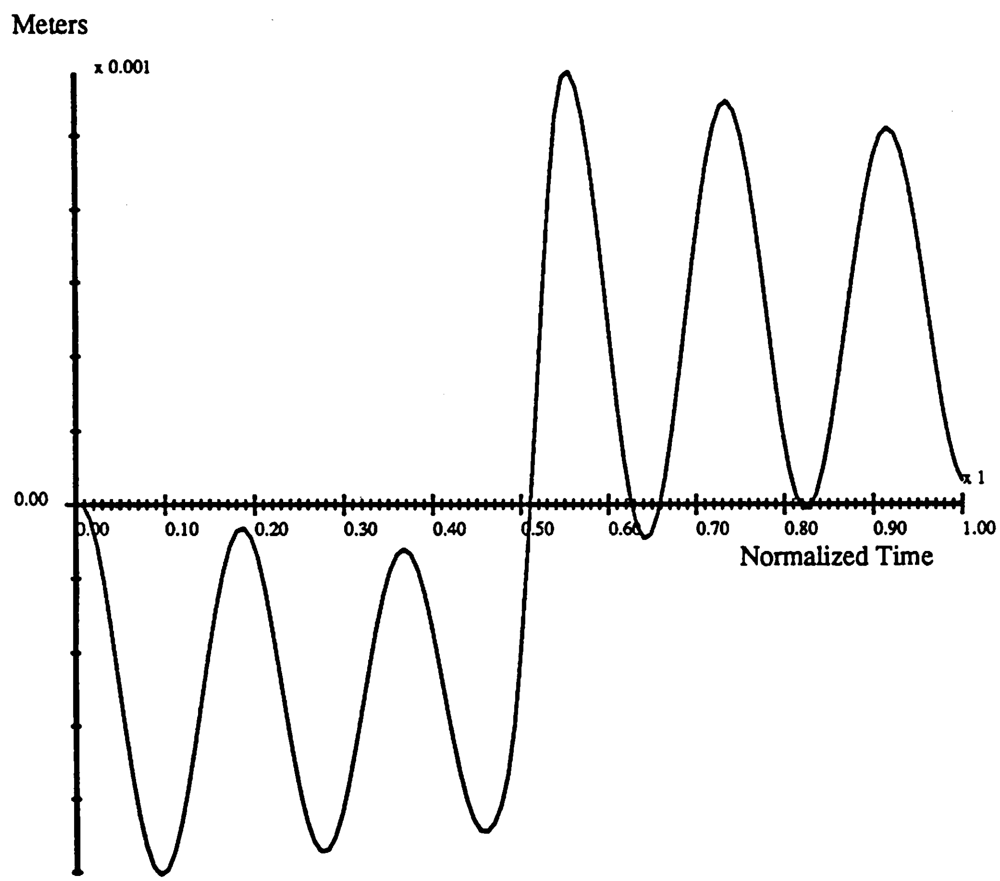

Figure 8.5 - Displacement of Tip of Tube

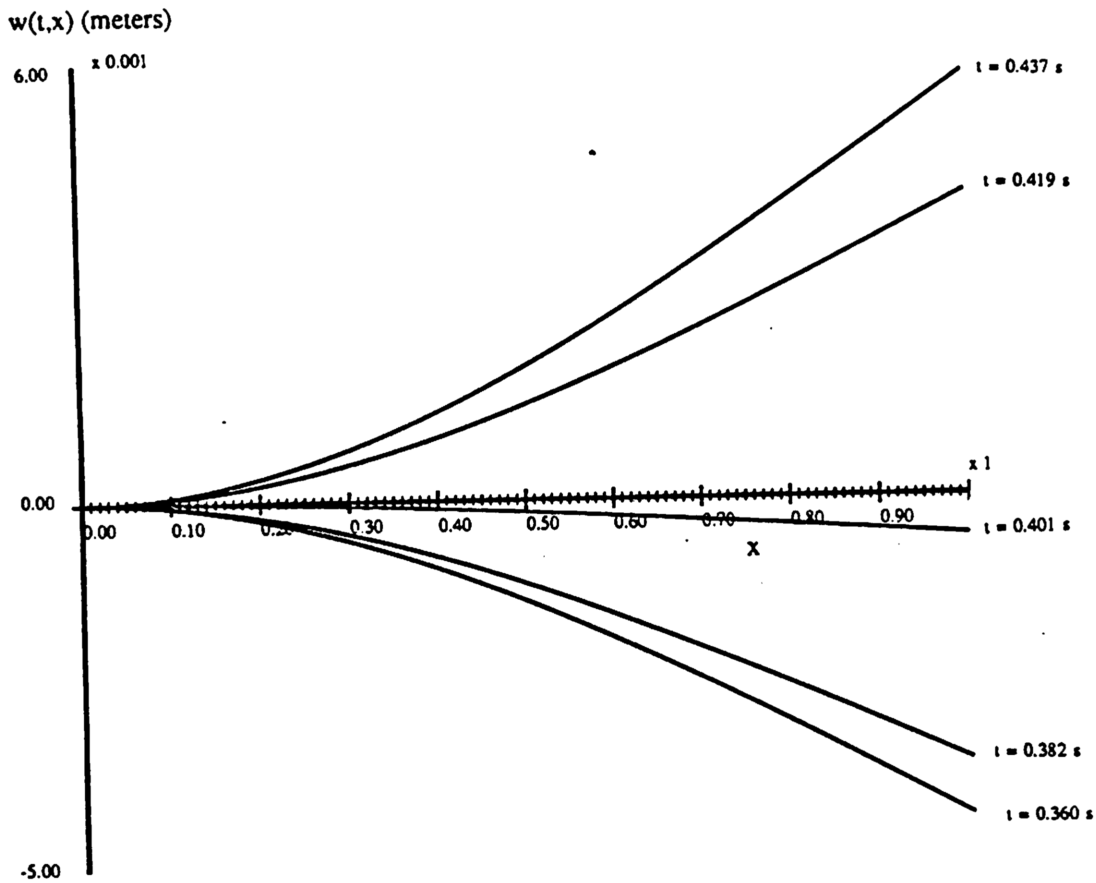

Figure 8.6 - Beam Profile

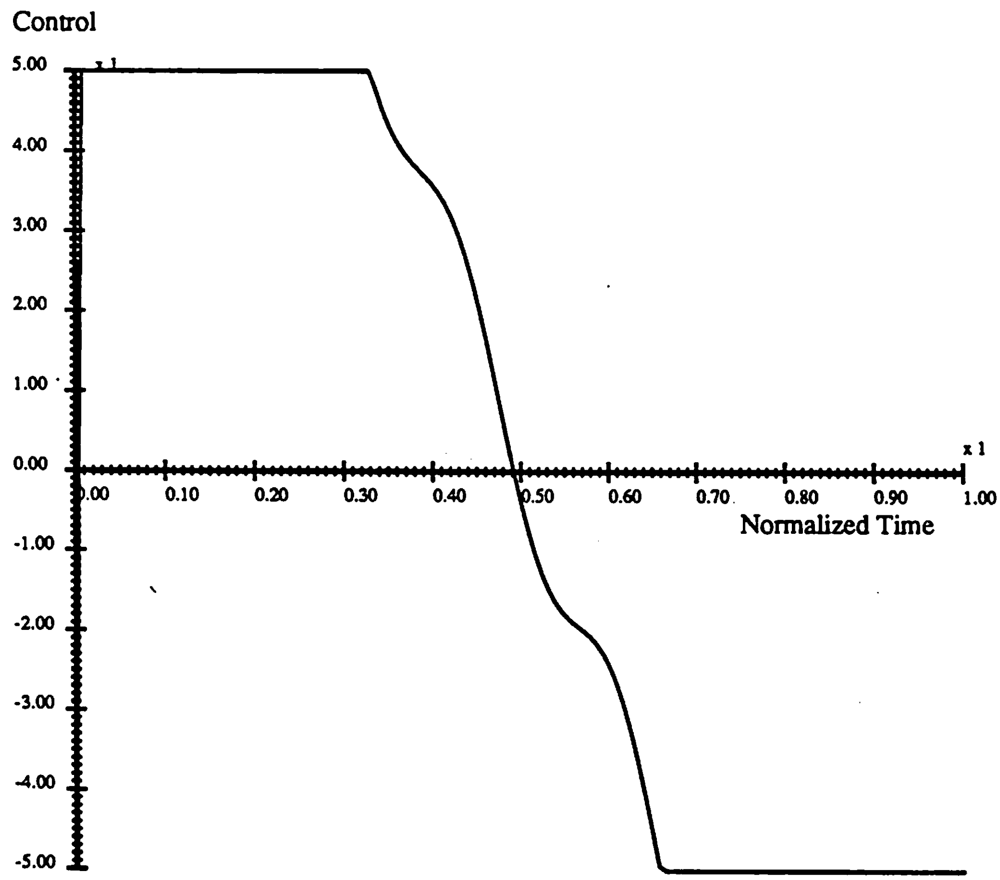

Figure 8.7 - Problem 2 Time of Maneuver = 0.800 seconds

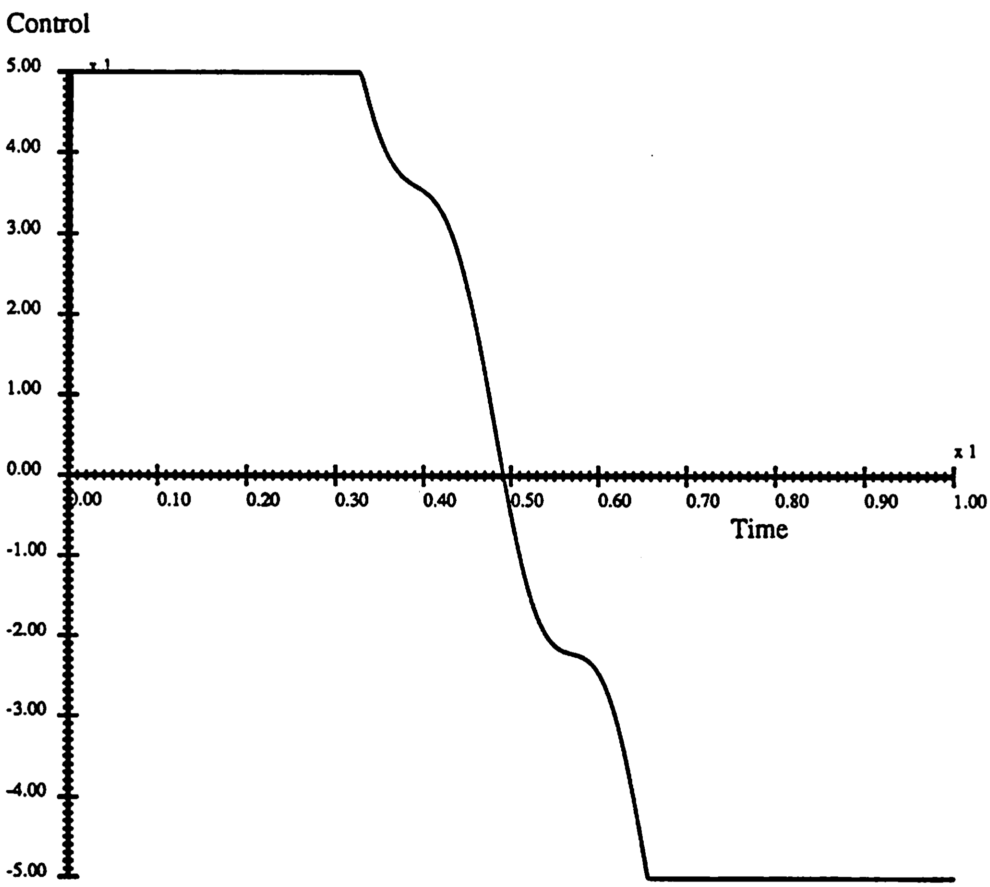

Figure 8.8a - Problem 2 Time of Maneuver = 0.900 seconds

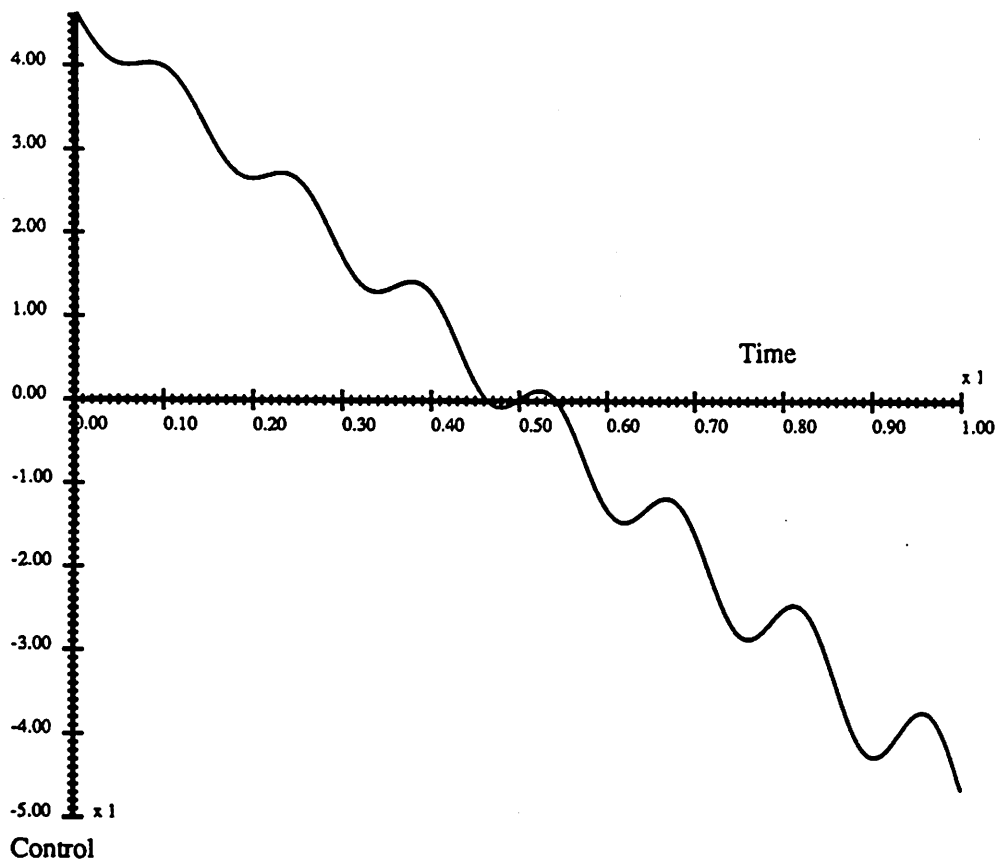

Figure 8.8b - Problem 2 Time of Maneuver = 1.000 seconds

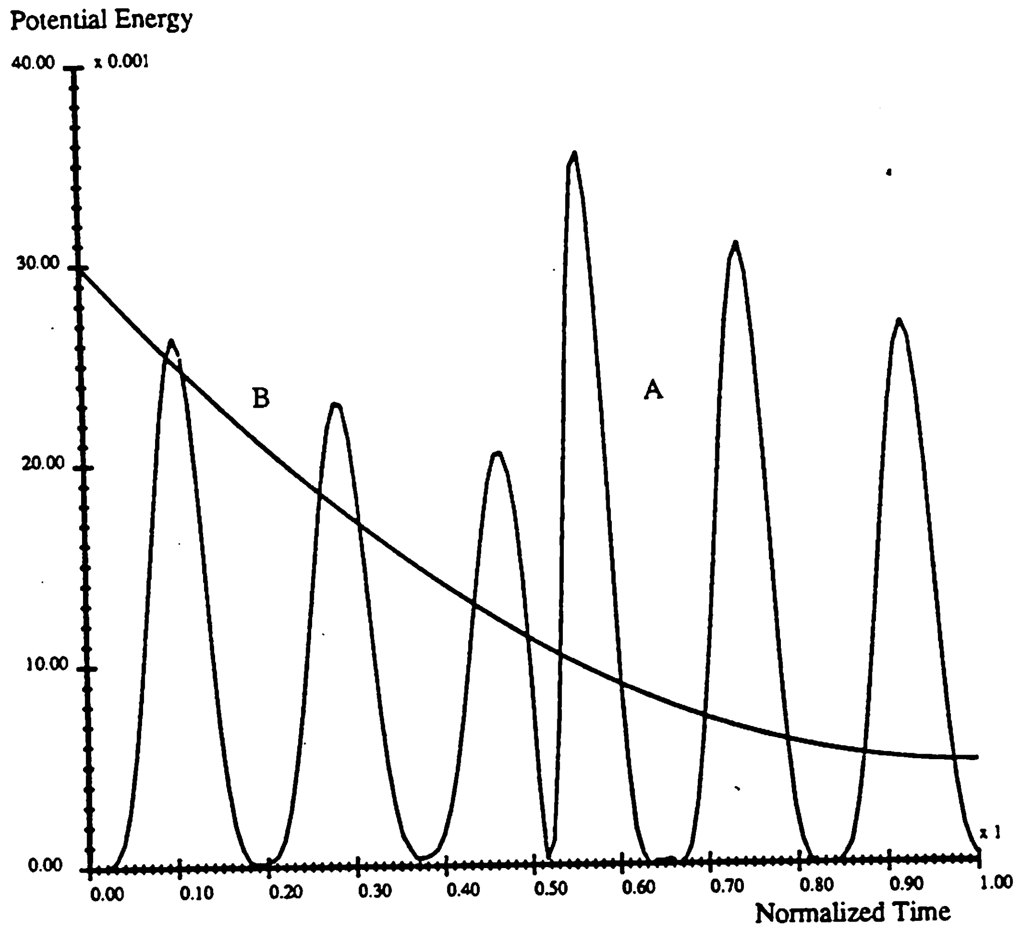

Figure 8.9 - Problem 1 Potential Energy

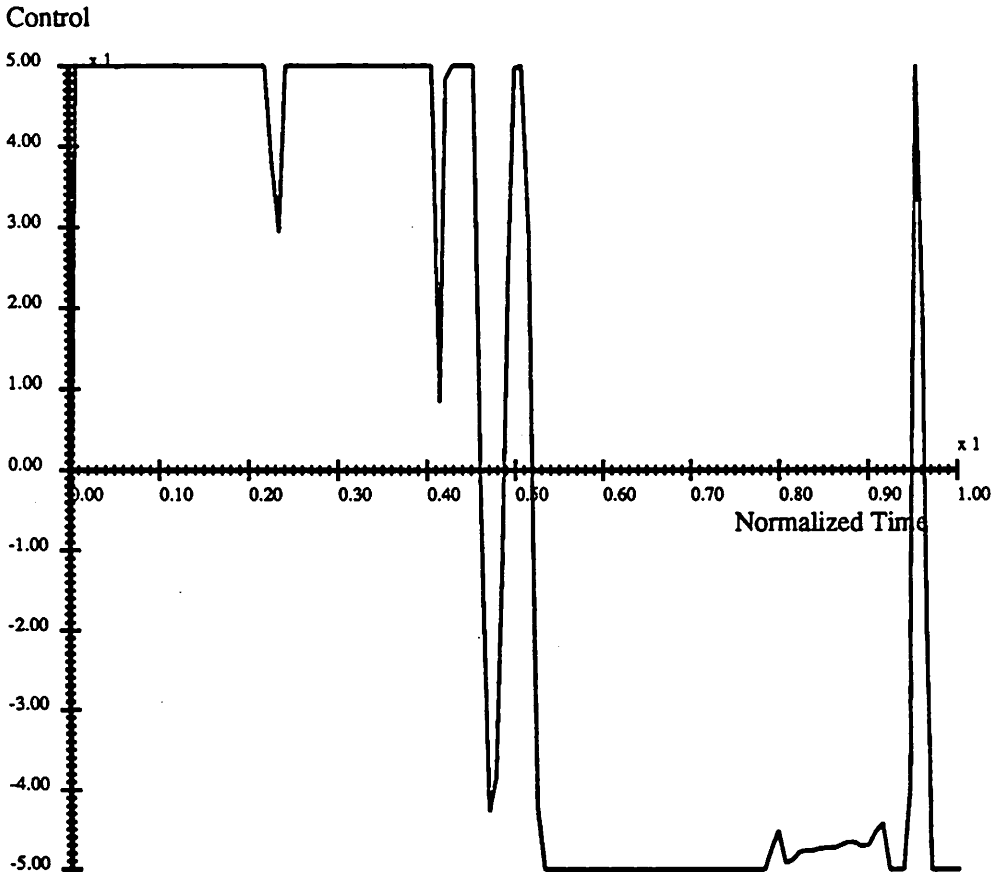

Figure 8.10 - Problem 3

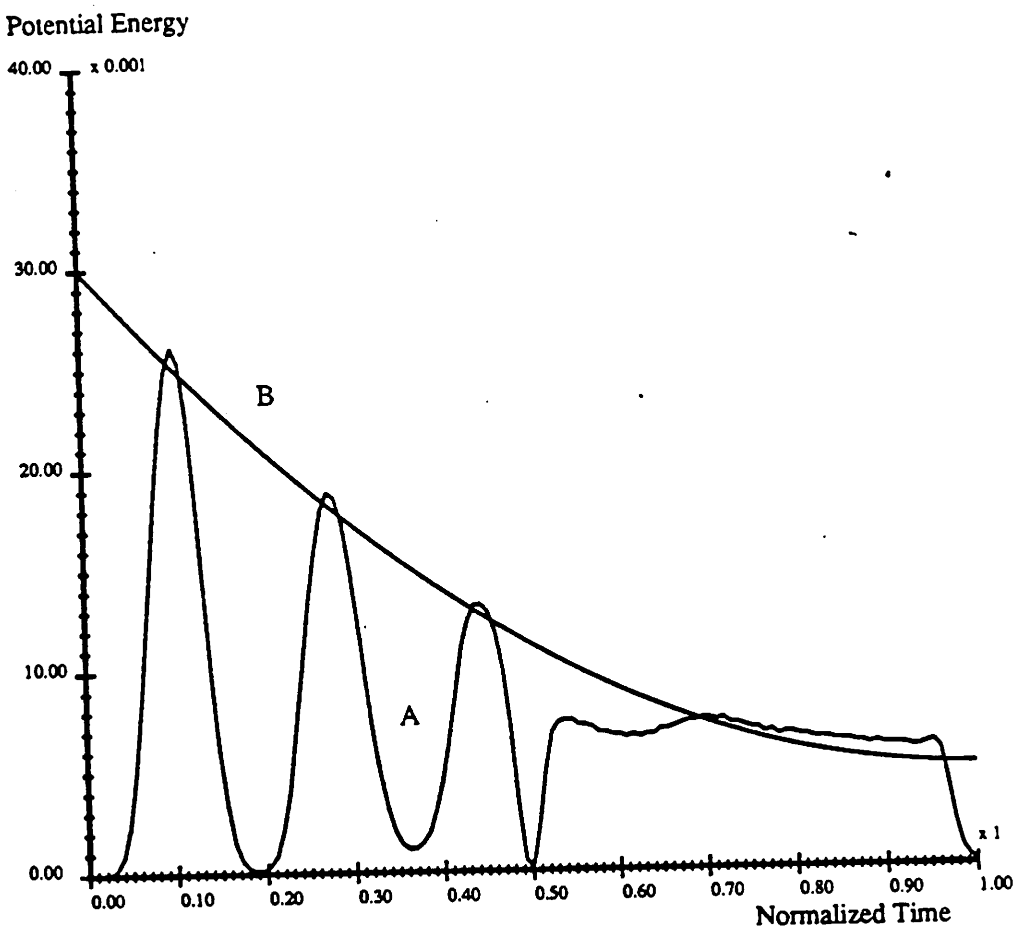

Figure 8.11 - Problem 3 Potential Energy

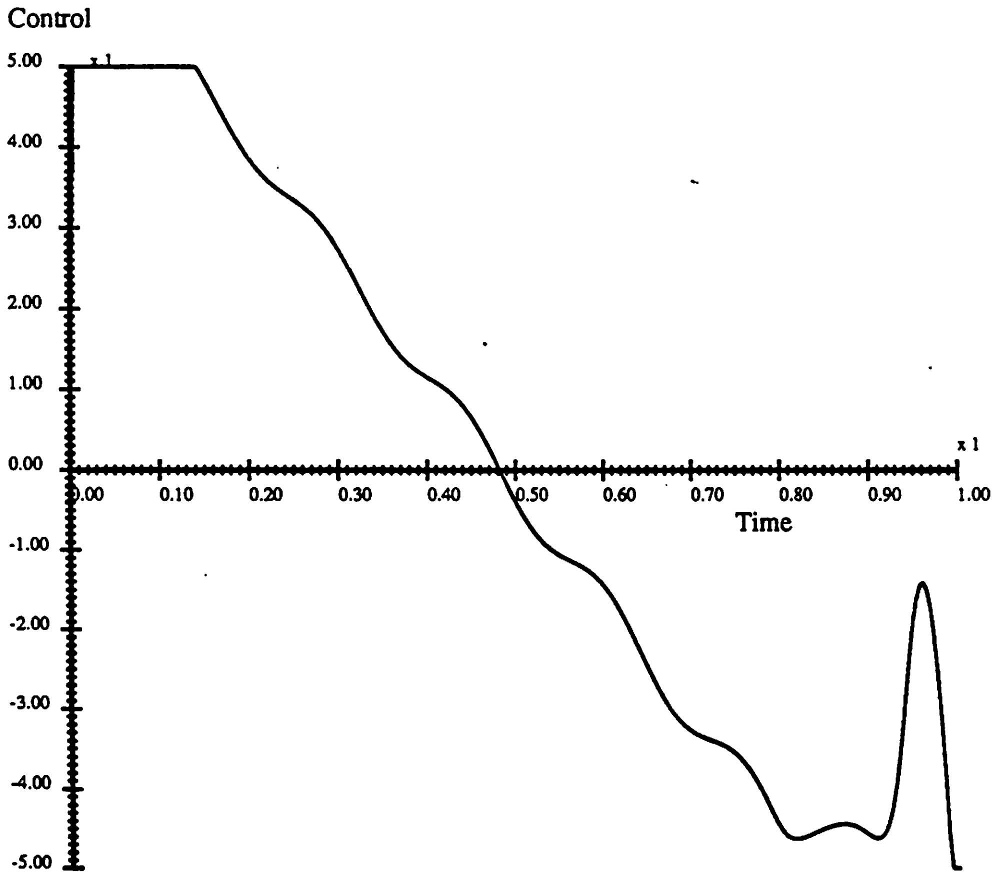

Figure 8.12 - Problem 4 Time of Maneuver = 0.900 seconds

Table 8.1 - Summary of Problem Results		
Problem	Final Time	Input Energy
$\mathbf{P}_1$	0.7883	19.15
$\mathbf{P}_2(0.80)$	0.8000	15.72
$\mathbf{P}_2(0.90)$	0.9000	9.87
$\mathbf{P}_2(1.00)$	1.0000	7.27
$\mathbf{P}_3$	0.8250	16.55
$\mathbf{P}_4(0.90)$	0.9000	10.49

IMPLEMENTATION COMPARISONS

We shall compare the performance of the two implementations of Algorithm 8.5.1. In Method I (Section 7.3), the search direction is calculated using the general purpose quadratic program LSSOL [Gil.2]. In Method II (Section 7.3), the search direction is calculated using a special purpose quadratic program. By experimentation, we have determined that the performance of Algorithm 8.5.1 is sensitive to two parameters, $\rho > 0$ and $\gamma \in (0, 1]$. The first parameter, ρ, determines the effect of the cost function in determining a search direction when the algorithm is at an infeasible point (Phase I). If ρ is large, then the effect of the cost function is negligible during Phase I. If ρ is large, then the algorithm finds a feasible point quickly; however, if ρ is too large, then the this feasible point is far from a local minimum. A value of ρ which is too small results in a very large times to find the first feasible point. In the limit that $\rho = 0$, the algorithm may never reach a feasible point. We have solved problems $\mathbf{P}_1$ - $\mathbf{P}_4$ using Method I with various values of ρ.

The other parameter, γ, can only be set in Method II. It is set equal to one in Method I by the producers of the QP package. The parameter $\gamma \in (0, 1]$ determines how accurately the QP (7.2.6) is solved. A large value of γ, (γ near one) forces precise solution of the QP while a small value (γ near zero) requires only an approximate solution of the QP. This solution will also be a direction of descent for Algorithm 8.5.1, but is probably not as good a direction as if the QP had been solved exactly. Since a more exact solution of the QP is more time consuming, we have found that there exists a tradeoff between the quality of the search direction and the time to find the direction. We have solved problems P_1 - P_4 using Method II with various values of ρ and γ.

We examine timings for the solutions to P_1 - P_4 using Method I, Table 8.2. In these experiments, we have fixed the number of time steps to be 256 and the number of finite elements to be 6. We fix these values to facilitate comparisons. *Time to Feasible* is the time in cpu seconds the algorithm required to find a feasible point. *Time to 1 percent* is the time in cpu seconds the algorithm required to find a feasible point whose cost is less that one percent higher than the optimal cost. For problems P_2 and P_4 the first feasible point has a cost within this one percent tolerance.

Table 8.2 - Timings with Standard Parameters (Method I)

Problem	ρ	Final Value of Cost	Time to Feasibility	Time to 1%
P_1	0.5	0.78837	2539	12500
P_2	0.5	15.711	15804	15804
P_3	0.5	0.82501	11692	74701
P_4	2.0	10.492	33300	33300

We examine in greater depth timings for the solutions to Problem P_4, Table 8.3, and P_1, Table 8.4. Blanks in Tables 8.3 and 8.4 correspond to values of parameters for which the experiments were not carried out. It is immediately apparent that proper selection of ρ is imperative. For problem P_4, when ρ is too small (ρ = 0.5), Method I does not determine a feasible point after 100,000 cpu seconds. The performance of Method II with γ = 1.10 is similarly poor. As ρ increases, the time to feasibility decreases and the difference between the time to 1 percent and time to feasibility usually increases since the first feasible point is usually farther from the optimal point. It is clear that there is an optimal value of ρ which causes the time to 1 percent to be minimized. However, this value is problem dependent, and we currently do not know how to determine it apriori. The best values of ρ which we tried are 8.0 for $\mathbf{P}_4$ and 0.2 for $\mathbf{P}_1$

It is also evident that the performance of Method II surpasses Method I by a factor between 6 to 10. We believe this is due to the ability of Method II to take advantage of the structure of the search direction finding problem. In particular, the matrix in the quadratic term of the QP is the Identity. However, the algorithm on which LSSOL is based does not make use of this information. We also note that the storage requirements for Method II is much smaller than for Method I. This is significant for computers which have little core memory and even more significant for computers which do not support virtual address memory.

From the available data, it is difficult to determine an optimal value or $\gamma \in (0,1]$, however, 1.10 appears to be the best value of γ for $\mathbf{P}_4$ that we have tried.

Table 8.3 - Timings for $\mathbf{P}_4$

		γ / ρ	0.5	2.0	8.0	16.0	32.0
Method I	Time to Feas.		>100,000	33,300	6452	5800	
	Time to 1%		>100,000	33,300	30957	35468	
Method II	Time to Feas.	1.02			848		
	Time to 1%				4082		
Method II	Time to Feas.	1.10	>100,000	4262	933	1271	1470
	Time to 1%		>100,000	4262	2972	5568	5964
Method II	Time to Feas.	1.30			1264		
	Time to 1%				3900		
Method II	Time to Feas.	2.00			975		
	Time to 1%				5945		

Table 8.3 - Timings for P_1								
		γ / ρ	0.025	0.1	0.2	0.3	0.5	2.0
Method I	Time to Feas.			10459			2539	1058
	Time to 1%			10459			12,500	31,526
Method II	Time to Feas.	1.02					373	
	Time to 1%						1877	
Method II	Time to Feas.	1.10	>100,000	1823	907	568	425	201
	Time to 1%		>100,000	1823	1695	1819	1913	5146
Method II	Time to Feas.	1.30					425	
	Time to 1%						1956	
Method II	Time to Feas.	2.00					354	
	Time to 1%						2068	
Method II	Time to Feas.	4.00					355	
	Time to 1%						2074	

CHAPTER 9

CONCLUSION

In this thesis, we have examined a class of optimal control problems with control, terminal inequality and state-space constraints in which the dynamics can be described by a canonical abstract differential equation. We have presented conceptual algorithms to solve these optimal control problems and shown convergence of these algorithms in the space of both ordinary controls and relaxed controls. We have also presented implementable algorithms and shown that for discretization schemes which satisfy certain reasonable requirements, the algorithms find stationary points. Finally& we have used these algorithms to solve several optimal slewing problems rapidly (i.e., less than one hour cpu time on a SUN workstation) .

We see future work in the areas of

(1) Generalization of the abstract differential equation to admit a larger class of PDEs;

(2) The use of refined models to obtain more precise answers;

(3) More efficient algorithms that utilize recent advances in scaling.

Open loop optimal control is only a partial answer. Work is being done at Berkeley to integrate these open loop optimal controls with finite dimensional compensators for robust control of systems described by partial differential equations.

REFERENCES

[Ara.1] Araya-Schulz, Roberto, "Control of a Large Space Structure Using a Distributed Parameter Model," *Ph.D. Dissertation*, U.C.L.A., 1986.

[Ath.1] Athans, M. and Falb, P., *Optimal Control*, McGraw-Hill, Inc., New York, 1966 .

[Bak.1] Baker, T. E. and Polak, E., *Computational Experiments in the Optimal Slewing of Flexible Structures*, UCB/ERL M87/72, Electronics Research Laboratory, U. C. Berkeley, September, 1987.

[Ber.1] Bertsekas, D. P., "On the Goldstein-Levitin-Polyak Gradient Projection Method," *IEEE Trans. AC*, vol. AC-21, pp. 174-184, 1976.

[Bla.1] Black, F. and Scholes, M., "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, pp. 637-653, 1973.

[Bre.1] Breakwell, J. A., "Optimal Feedback Slewing of Flexible Spacecraft," *Journal of Guidance and Control*, vol. 4, pp. 472-479, September-October 1981.

Lockhead Missles and Space Co, inc Palo Alto Ca Bending Modes, Feedback Compatible Experimental results, sensitivity to modelling error results

[Can.1] Canon, M., Cullum, C., and Polak, E., "Constrained Minimization Problems in Finite-dimensional Spaces," *SIAM Journal of Control*, vol. 4, pp. 528-548, 1966.

[Che.1] Chen, G. and Russell, D. L., "A Mathematical Model for Linear Elastic Systems with Structural Damping," *Quarterly of Applied Mathematics*, pp. 433-454, January, 1982.

[Cho.1] Chow, N. S., Paret-Mallet, J., and Yorke, J. A., "Finding Zeros of Maps Homotopy Methods that are Constructive with Pobability One," *Math. Comp.*, vol. 32, 1978.

[Chu.1] Chun, H. M., "Large–Angle Slewing Maneuvers for Flexible Spacecraft," *Ph.D. Dissertation*, M.I.T., 1986.

[Cul.1] Cullum, Jane, "Discrete Approximations to Continuous Optimal Control Problems," *Siam J. Control*, vol. 7, no. 1, February, 1969.

[Die.1] Dieudonne, J., *Foundations of Modern Analysis*, Academic Press, New York, 1960.

[Fan.1] Fan, Ky, "Minimax Theorems," *Proc. National Academy of Sciences*, vol. 39, pp. 42-47, 1953. Q11.N26

[Flo.1] Floyd, M. A., Brown, M. E., Turner, J. D., and Vandervelde, W. E., "Implementation of a Minimum Time and Fuel On/Off Thruster Control System for Flexible Spacecraft," *Journal of Astronautical Sciences*, To Appear.

[Fuj.1] Fujii, H., "Finite Element Schemes: Stability and Convergence," *Advances in Computational Methods in Structural Mechanics and Design*, pp. 201-218., 1972.

[Gib.1] Gibson, J. S., "An Anallysis of Optimal Nodal Regulation: Convergence and Stability," *Siam Journal for Control and Optimization*, vol. 19, no. 5, pp. 686-707, September, 1981.

[Gil.1] Gill, P. E., Murray, W., Saunders, M., and Wright, M. H., "Procedures for Optimization Problems with a Mixture of Bounds and General Constraints," *ACM Transactions on Mathematical Software*, vol. 10, 1984.

[Gil.2] Gill, P. E. et. al., *User's Guide for LSSOL*, SOL 86-1, Systems Optimization Laboratory, Department of Operations Research, Stanford University.

[Gol.1] Goldstein, A. A. and Price, J. F., "An Effective Algorithm for Minimization," *Numer. Math*, vol. 10, pp. 184-189, 1967.

[Gra.1] Graff, Karl F., *Wave Motions in Elastic Solids*, Ohio State University Press, 1975.

[Hal.1] Hale, Jack K., *Ordinary Differential Equations*, Robert E. Krieger Publishing Co., Malabar, Fl., 1980.

[Hig.1] Higgins, J. E., Electronics Research Laboratory, U. C. Berkeley, In Preparation.

[Jun.1] Junkins, John L. and Turner, James D., *Optimal Spacecraft Rotational Maneuvers*, Studies in Astronautics, 3, Elsevier, Amsterdam, 1986.

[Kle.1] Klessig, R. and Polak, E., "An adaptive algorithm for unconstrained optimization with applications to optimal control," *SIAM J. Control*, vol. 11, no. 1, pp. pp 80-94, 1973. QA402.3 A1 S13 v.11

[Kle.2] Klessig, R. and Polak, E., "A Method of Feasible Directions Using Function Approximations with Applications to Min Max Problems," *J. Math. Analysis and Applications*, vol. 41, no. 3, pp. 583-602, 1973.

[Lev.1] Levitin, E. S. and Polyak, B. T., "Constrained Minimization Methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 6, no. 5, pp. 27-29, 1966. QA1.U14

[Lue.1] Luenberger, David G., *Linear and Nonlinear Programming, second edition*, Addison-Wesley, Menlo Park, 1984.

[May.2] Mayne, D. Q. and , E. Polak, "An Exact Penalty Function Algorithm for Optimal Control Problems with Control and Terminal Equality Constraints, Parts 1 and 2," *JOTA*, vol. 32, pp. 211-246 and 345-364, 1980.

[May.3] Mayne, D. Q. and , E. Polak, "An Exact Penalty Function Algorithm for Optimal Control Problems with State and Control Constraints," *IEEE Transactions on Automatic Control*, vol. AC-32, no. 5, pp. 380-388, 1987.

[Mit.1] Mitter, S. K., "Successive Approximation Methods for the Solution of Optimal Control Problems," *Automatica*, vol. 3, pp. 135-149, 1966.

[Mun.1] Munkres, J. R., *Topology*, Prentice Hall, Englewood Cliffs, N. J., 1975.

[Nas.1] NASA,, *First NASA / DOD CSI Technology Conference, Proceedings*, OMNI International Hotel, Norfolk, Virginia, November, 1986.

[New.1] Newmark, Nathan M., "A Method of Computation for Structural Dynamics," *American Society of Civil Engineers*, vol. 127, pp. 1406-1435, 1962. TA 1 A5 v.127

[Ode.1] Oden, J. T. and Fost, R. B., "Convergence, Accuracy and Stability of Finite Element Approximations of a Class of Non-linear Hyperbolic Equations," *Intl. Journal for Numerical Methods in Engineering*, vol. 6, pp. 357-365.

[Paz.1] Pazy, A., *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer-Verlag, New York, 1983.

[Pir.1] Pironneau, O. and Polak, E., *A dual Method for Optimal Control Problems with Initial and Final Boundary Constraints*, ERL-M299, Electronics Research Laboratory, U. C. Berkeley, May, 1971.

[Pir.2] Pironneau, O. and Polak, E., "On the Rate of Convergence of Certain Methods of Centers," *Mathematical Programming*, vol. 2, pp. 230-258, 1972.

[Pol.1] Polak, E., *Computational Methods in Optimization: A Unified Approach*, Academic Press, New York, 1971.

[Pol.2] Polak, E. and Mayne, D. Q., "A Feasible Directions Algorithm for Control Problems with Control and Terminal Inequality Constraints," *IEEE Trans. AC*, vol. AC-22, pp. 741-751, 1977.

[Pol.3] Polak, E. and Mayne, D. Q., "First-Order Strong Variation Algorithms for Optimal Control Problems with Termianl Inequality Constraints," *JOTA*, vol. 16, no. 3/4, pp. 303-325, August, 1975.

[Pol.4] Polak, E., "An Historical Survey of Computational Methods in Optimal Control," *SIAM Review*, vol. 15, no. 2, pp. 553-584., April, 1973.

[Pol.5] Polak, E., Trahan, R., and Mayne, D. Q., "Combined Phase I - Phase II Methods of Feasible Directions," *Mathematical Programming*, vol. 17, pp. 61-73, 1979.

[Pol.6] Polak, E. and Wiest, J., *Domain Rescaling Algorithms for Composite Minimax Problems*, Electronics Research Laboratory, U. C. Berkeley, In Preparation.

[Ros.1] Rosen, J. B., "Iterative Solution of Nonlinear Optimal Control Problems," *SIAM Journal of Control*, vol. 4, 1966.

[Rud.1] Rudin, W., *Real and Complex Analysis*, McGraw–Hill, 1974.

[Sac.1] Sackman, J., *Personal Communication.*

[Sch.1] Schmidt, W. F., "Adaptive Step Size Selection for Use with the Continuation Method," *International Journal for Numerical Methods in Engineering*, vol. 12, 1978.

[Sho.1] Showalter, R. E., *Hilbert Space Methods for Partial Differential Equations*, Pitman, London, 1977.

[Sim.1] Simo, J. C. and Vu-Quoc, L., "The Role of Nonlinear Theories in Transient Dynamics Analysis of Flexible Structures," *Journal of Sound and Vibration*, 1988 (to appear).

[Sto.1] Stoer, Josef, "On the Numerical Solution of Constrained Least-Squares Problems," *Siam J. Numer. Anal.*, vol. 8, no. 2, pp. 382-411, June, 1971.

[Str.1] Strang, Gilbert and Fix, George, *An Analysis of the Finite Element Method*, Prentice Hall, Englewood Cliffs, N. J., 1973.

[Tay.1] Taylor, Larry, ed., *SCOLE: Space Control Laboratory Experiment Workshop*, NASA Langly Research Center, Hampton, Virginia 23665, 1985.

[Tay.2] Taylor, Larry, Leary, Terry, and Stweart, Eric, "On Incorporating Damping and Gravity Effects in Models of Structural Dynamics of the SCOLE Configuration," *SCOLE: Space Control Laboratory Experiment Workshop*, NASA Langly Research Center, Hampton, Virginia 23665, 1985.

[Vuq.1] Vu-Quoc, L., "Dynamics of Flexible Structures Performing Large Overall Motions: A Geometrically-Nonlinear Approach," *Doctoral Dissertation*, vol. UCB/ERL M86/36, Electronics Research Laboratory, University of California, May, 1986.

[War.1] Warga, J., *Optimal Control of Differential Equations and Functional Equations*, 1972. QA402.3 W371

[War.2] Warga, J., "Iterative Procedures for Constrained and Unilateral Optimization Problems," *SIAM J. Control and Optimization*, vol. 20, pp. 360-376, May, 1982. No. 3

[Wil.1] Williamson, L. V. and Polak, E., "Relaxed Controls and the Convergence of Optimal Control Algorithms," *Siam J. Control and Optimization*, vol. 14, no. 4, July, 1976.

APPENDIX 1

EVOLUTION SYSTEMS

A two parameter family of bounded linear operators, $U(t,s)$, $0 \le s \le t \le 1$ on X is called an evolution system if the following two conditions are satisfied:

(i) $U(s,s) = I$, $U(t,r)U(r,s) = U(t,s)$ for $0 \le s \le t \le 1$,

(ii) $(t,s) \to U(t,s)$ is strongly continuous for $0 \le s \le t \le 1$.

Evolution Systems are the generalization of state transition matrices. We give a version of an existence result from Pazy [Paz.1] for evolution systems which is sufficiently general for our purposes.

Theorem A1.1: Given $A{:}D(A) \to X$, the infinitesimal generator of a semigroup $\{ T(t) \}_{t \ge 0}$ such that $\|T(t)\| \le Me^{\bar{\omega} t}$ for $M < \infty$ and $\bar{\omega} \ge 0$, and $B(\cdot) \in L_\infty([0,1],B(X,X))$ there exists a unique evolution system $U(t,s)$ for $A + B(\cdot)$ such that:

(i) $\|U(t,s)\| \le Me^{\omega(t-s)}$, for $0 \le s \le t \le 1$, $\omega \ge 0$,

(ii) $\frac{\partial}{\partial t} U(t,s)v \,\big|_{t=s} = (A + B(s))v$, for $v \in D(A)$, a.e. on $0 \le s \le t \le 1$ for $v \in D(A)$,

(iii) $\frac{\partial}{\partial s} U(t,s)v = -U(t,s)(A + B(s))v$, for $v \in D(A)$, a.e. $0 \le s \le t \le 1$.

Proof: Since $B(\cdot) \in L_\infty([0,1],B(X,X))$, $\|B\| < \infty$. For almost all $t \in [0,1]$, $A + B(t)$ is the infinitesimal generator of a semigroup $\{ S_t(s) \}_{s \ge 0}$ satisfying $\|S_t(s)\| \le Me^{\omega s}$ where $\omega \triangleq \bar{\omega} + M\|B\|_\infty$.

Since the step functions are dense in L_2, we can construct a sequence of partitions $\{ t_k^n \}_{k=1}^{N(n)}$, for $n \in \mathbb{Z}+ \triangleq \{ 1,2,3, \cdots \}$ with $0 = t_0^n \le t_1^n \le t_2^n \le \cdots \le t_{N(n)}^n = 1$, and

$$B_n(t) \triangleq \begin{cases} B(t_k^n) \text{ for } t_k^n \le t < t_{k+1}^n, \ k \in \{0,1,\cdots,n-1\} \\ B(t_{N(n)-1}^n) \text{ for } t = 1 \end{cases}$$

such that

(i) $\delta_n \triangleq \max\{ t_{k+1}^n - t_k^n \} \to 0$ as $n \to \infty$,

(ii) $|B_n(t)| \le |B|_\infty$, for all $t \in [0,1]$,

(iii) $\lim_{n \to \infty} \int_0^1 |B_n(r) - B(r)|^2 dr \to 0.$

Next, we define a sequence of two parameter operators $U_n(t,s)$, $0 \le s \le t \le 1$, $n \in \mathbb{Z}+$ by:

$$U_n(t,s) \triangleq \begin{cases} S_{t_j^n}(t-s), \text{ for } t_j^n \le s \le t \le t_{j+1}^n \\ S_{t_k^n}(t - t_k^n)\left[\prod_{j=l+1}^{k-1} S_{t_j^n}(t_{j+1}^n - t_j^n)\right] S_{t_l^n}(t_{l+1}^n - s) \text{ for } k < l, \ t_k^n \le t \le t_{k+1}^n, \ t_l^n \le s \le t_{l+1}^n \end{cases}$$

The evolution operators $U_n(\cdot,\cdot)$ have the following properties:

(i) $U_n(\cdot,\cdot)$ is an evolution system,

(ii) $|U_n(t,s)| \le M^{\omega(t-s)}$, for $0 \le s \le t \le 1$,

(iii) $\frac{\partial}{\partial t} U_n(t,s)v = (A + B_n(t))U_n(t,s)v$, for $v \in D(A)$, for $t \ne t_k^n$, $k \in \{0,1,2,...N(n)\}$

(iv) $\frac{\partial}{\partial s} U_n(t,s)v = -U_n(t,s)(A + B_n(t))v$, for $v \in D(A)$, for $t \ne t_k^n$, $k \in \{0,1,2,...N(n)\}$.

Let $v \in D(A)$ and consider the map: $r \to U_n(t,r)U_m(r,s)v$. From (iii) and (iv), it follows that except for a finite number of values of r, the map is differentiable in r, $s \le r \le t$ and

$$\|U_n(t,s)v - U_m(t,s)v\| = \|\int_s^t \frac{\partial}{\partial r} U_n(t,r)U_m(r,s)v dr\| = \|\int_s^t U_n(t,r)(B_n(r) - B_m(r))U_m(r,s)v ds\|$$

$$\leq M^2 e^{\omega(t-s)}\|v\|\int_s^t \|B_n(r) - B_m(r)\|dr \leq M^2 e^{\omega(t-s)}\|B_n - B_m\|_2\|v\|. \qquad (A1.1)$$

Since $D(A)$ is dense in X, and (ii) from above, it follows that $\{ U_n(t,s) \}$ is a Cauchy sequence, and therefore, it converges in the operator topology uniformly on $0 \leq s \leq t \leq 1$ as $n \to \infty$. We denote the limit $U(t,s)$. $U(t,s)$ is an evolution system and property (i) of the theorem is satisfied.

To prove (ii) and (iii), we consider the map $r \to U_n(t,r)S_\tau(r-s)v$ for $v \in D(A)$. This function is differentiable except for a finite number of values of r and

$$\|U_n(t,s)v - S_\tau(t-s)v\| = \|\int_s^t \frac{\partial}{\partial r} U_n(t,r)S_\tau(r-s)v dr\| = \|\int_s^t U_n(r,t)(B_n(r) - B(\tau))U_m(r,s)v ds\|$$

$$\leq M^2 e^{\omega(t-s)}\|v\|\int_s^t \|B_n(r) - B_m(\tau)\|dr \leq M^2 e^{\omega(t-s)}\|v\|\int_s^t \|B_n(r) - B(\tau)\|dr. \qquad (A1.2)$$

Passing to the limit as $n \to \infty$ (A1.2) yields

$$\|U(t,s)v - S_\tau(t-s)v\| \leq M^2 e^{\omega(t-s)}\|v\|\int_s^t \|B(r) - B(\tau)\|dr. \qquad (A1.3)$$

■

CONTINUITY OF EVOLUTION SYSTEMS

Let $\{ B^i \} \subset C^1([0,1], B(X,X))$ be a sequence of linear operators converging to $B \in L_2([0,1], B(X,X)) \cap L_\infty([0,1], B(X,X))$ in the L_2 norm. As previously shown, by choosing a sequence of partitions, we can construct a sequence of evolution systems $U_n(t,s)$ converging to $U(t,s)$ satisfying properties (i),(ii), and (iii) of Theorem A2.1. For each B^i, using the same partition as for B, we can construct a sequence of evolution systems $U_n^i(t,s)$ converging to an evolution system $U^i(t,s)$. We state the following lemma.

Lemma A1.2: For $\{ B^i \} \subset C^1([0,1],B(X,X))$ such that $B^i \to B \in L_2([0,1],B(X,X)) \cap L_\infty([0,1],B(X,X))$ in the L_2 norm, $U^i(t,s) \to U(t,s)$ uniformly in $0 \le s \le t \le 1$ in the operator topology.

Proof:

$$U^i(t,s)v - U(t,s)v = (U^i(t,s)v - U^i_n(t,s)v)$$
$$+ (U^i_n(t,s)v - U_n(t,s)v) + (U_n(t,s)v - U(t,s)v). \tag{A1.4}$$

By Theorem A1.1, for $\varepsilon > 0$, there exists n_0 such that for all $n > n_0$, $|U_n(t,s) - U(t,s)| \le \varepsilon$, and for all i, there exists n_i such that for all $n > n_i$, $|U^i_n(t,s) - U^i(t,s)| \le \varepsilon$, for $0 \le s \le t \le 1$. Applying the argument in Theorem A1.1, we see that except for a finite number of values of $r \in [0,1]$, the map $r \to U_n(t,r)U^i_n(r,s)v$ is differentiable for all $v \in D(A)$,

$$|U_n(t,s)v - U^i_n(t,s)v| = \left|\int_s^t \frac{\partial}{\partial r} U_n(t,r)U^i_n(r,s)v dr\right| \le \left|\int_s^t U_n(t,r)[B_n(r) - B^i(r)]U^i_n(r,s)v dr\right|$$
$$\le M^2 e^{\omega(t-s)}|v| \int_s^t |B_n(r) - B^i_n(r)| dr, \tag{A1.5}$$

and [1]

$$\int_s^t \|B_n(r) - B^i_n(r)\| dr \le |B_n - B^i_n|_1 \le |B_n - B|_1 + |B - B^i|_1 + |B^i - B^i_n|_1$$
$$\le |B_n - B|_2 + |B - B^i|_2 + |B^i - B^i_n|_2. \tag{A1.6}$$

For $\varepsilon > 0$, there exists $\tilde{n}_0 > n_0$ such that for all $n > \tilde{n}_0$, $|B_n - B|_2 \le \varepsilon$, there exists i_0 such that for all $i > i_0$, $|B - B^i|_2 \le \varepsilon$; for all i there exists $\tilde{n}_i > \max\{ \tilde{n}_0, n_i \}$ such that for all $n > \tilde{n}_i$, $|B^i - B^i_n|_2 \le \varepsilon$.

---

[1] $|B_n - B|_1 \triangleq \int_0^1 |B_n(r) - B(r)| dr$; $|B_n - B|_2 \triangleq (\int_0^1 |B_n(r) - B(r)|^2 dr)^{1/2}$.

Consequently, for all $i > i_0$, there exists $\bar{n}_i$ such that for all $n > \bar{n}_i$, $|B_n - B_n^i|_1 \leq 3\varepsilon$ and

$$|U_n(t,s)x - U_n^i(t,s)x| \leq M^2 e^{\omega}|x|3\varepsilon, \tag{A1.7}$$

for all $x \in D(A)$. Since $D(A)$ is dense in X,

$$|U_n(t,s) - U_n^i(t,s)| \leq 3M^2 e^{\omega}\varepsilon, \tag{A1.8}$$

for all $i > i_0$, $n > \bar{n}_i$, and

$$|U^i(t,s) - U(t,s)| \leq 3M^2 e^{w}\varepsilon, \tag{A1.9}$$

for $i > i_0$. Therefore, $U^i(t,s) \to U(t,s)$ in the operator topology, uniformly on $0 \leq s \leq t \leq 1$.

■

Lemma A1.3: Consider the Cauchy initial value problem:

$$\frac{d}{dt}x(t) = (A + B(t))x(t) + f(t), \quad x(0) = x, \tag{A1.10}$$

where $A{:}D(A) \to X$ is the infinitesimal generator of a continuous semigroup $\{ T(t) \}_{t \geq 0}$; and $B \in L_2([0,1],\mathrm{B}(X,X)) \cap L_\infty([0,1],\mathrm{B}(X,X))$, and $f \in L_2([0,1],X) \cap L_\infty([0,1],X)$. The mild solution exists and satisfies:

$$x(t) = U(t,0)x + \int_0^t U(t,s)f(s)ds, \tag{A1.11}$$

where $U(t,s), 0 \leq s \leq t \leq 1$ is the evolution system generated by $\{ A + B(t) \}_{t \in [0,1]}$.

Proof: Since $x \in X$, $B \in L_2([0,1],B(X,X))$ and $f \in L_2([0,1],X)$, there exists sequences $\{ x^i \} \subset D(A)$, $\{ B^i \} \subset C^1([0,1]),B(X,X))$ and $\{ f^i \} \subset C^1([0,1],X)$ such that $x^i \to x$, $B^i \to B$ and $f^i \to f$ in the appropriate topologies. For each i, the system:

$$\frac{d}{dt}x^i(t) = (A + B^i(t))x^i(t) + f^i(t), \quad x^i(0) = x^i \tag{A1.12}$$

has a classical solution

$$x^i(t) = U^i(t,0)x^i + \int_0^t U^i(t,s)f^i(s)ds \tag{A1.13}$$

which is also a weak solution i.e.,

$$x^i(t) = T(t)x^i + \int_0^t T(t-s)[B^i(s)x^i(s) + f^i(s)]ds. \tag{A1.14}$$

While (A1.10) does not necessarily have a classical solution, it has a weak solution:

$$x(t) = T(t)x + \int_0^t T(t-s)[B(s)x(s) + f(s)]ds. \tag{A1.15}$$

Therefore,

$$|x^i(t) - x(t)| = |T(t)(x^i - x) + \int_0^t T(t-s)[B^i(s)x^i(s) - B(s)x(s) + f^i(s) - f(s)]ds| \tag{A1.16}$$

$$\leq M[|x^i - x| + |f^i - f|_2 + \sup_i|B^i|\int_0^t |x^i(s) - x(s)|ds + |x|_\infty \|B^i - B|_2].$$

Applying the Bellman-Gronwall Lemma, and noting $\sup_i|B^i| < \infty$,

$$|x^i(t) - x(t)| \leq k_0[|x^i - x| + |f^i - f|_2 + |B^i - B|]. \tag{A1.17}$$

Therefore, $x^i(t) \to x(t)$ uniformly in $t \in [0,1]$. Since $x^i(\cdot) \to x(\cdot)$, $U^i(\cdot,\cdot) \to U(\cdot,\cdot)$, $f^i(\cdot) \to f(\cdot)$, $B^i(\cdot) \to B^i(\cdot)$. It follows that $x(t) = U(t,0)x + \int_0^t U(t,s)f(s)ds$. ■

Lemma A1.4: Consider the Cauchy final value problem:

$$\frac{d}{dt}x(t) = -(A^* + B^*(t))x(t), \quad t \in [0,1], \quad x(1) = x, \tag{A1.18}$$

where $A^*:D(A^*) \to X$ and $B^*(\cdot):X \to X$ are the adjoints of A and $B(\cdot)$ defined in Theorem A1.1.

Then the mild solution to (A1.18) exists and satisfies $x(t) = U(1,t)^* x$ for $t \in [0,1]$.

Proof: We first define $s \triangleq 1 - t$, $y(s) \triangleq x(1 - s)$. (A.2.2) becomes:

$$\frac{d}{ds} x(1 - s) = (A^* + B^*(1 - s))x(1 - s), \; s \in [0,1], \tag{A1.19}$$

$$\frac{d}{ds} y(s) = (A^* + B^*(1 - s))y(s), \; s \in [0,1], \; y(0) = x. \tag{A1.20}$$

We construct a sequence of partitions $\{ s_k^n \}_{k=1}^{N(n)}$ for $n \in \mathbb{Z}$ with $0 = s_{N(n)}^n \leq \cdots \leq s_1^n \leq s_0^n = 1$ by setting $s_k^n = 1 - t_k^n$ where $\{ t_k^n \}$ is the sequence of partitions in Lemma A1.1. Next we define a sequence of two parameter operators $V_n(t,s)$, $0 \leq s \leq t \leq 1$, $n \in \mathbb{Z}_+$

$$V_n(t,s) = \begin{cases} S_{t_j^n}^*(t - s), \; s_{j+1} \leq s \leq t \leq s_j^n \\ S_{t_l^n}(t - s_{l+1}^n)\left[\prod_{j=k-1}^{l+1} S_{t_j^n}^*(s_j^n - {}_{j+1}^n)\right] S_{t_k^n}^*(s_k^n - s), \; s_{l+1}^n \leq t \leq s_l^n, \; s_{k+1}^n \leq s \leq s_k^n \end{cases}$$

$$V_n^*(1 - s, 1 - t) = \begin{cases} S_{t_j^n}(t - s), \; t_j^n \leq s \leq t \leq t_{j+1}^n, \\ S_{t_k^n}(t - t_k^n)\left[\prod_{j=l+1}^{k-1} S_{t_j^n}(t_{j+1}^n - t_j^n)\right] S_{t_l^n}(t_l^n - s), \; t_l^n \leq s \leq t_{l+1}^n, \; t_k^n \leq t \leq t_{k+1}^n. \end{cases} \tag{A1.21}$$

Therefore $V_n^*(1 - s, 1 - t) = U_n(t,s)$ or $V_n(t,s) = U_n^*(1 - s, 1 - t)$, $0 \leq s \leq t \leq 1$. Following A1.1, there exists an evolution operator $V(t,s)$ such that $V_n(t,s)$ converges to $V(t,s)$ uniformly in $0 \leq s \leq t \leq 1$, and $V(t,s) = U^*(1 - s, 1 - t)$ with $U^*(\cdot,\cdot)$, the adjoint of $U(\cdot,\cdot)$ defined in A1.1.

By A1.20, $y(s) = V(s,0)x$; $x(s) = y(1 - s) = V(1 - s,0)x = U^*(1,s)x$. ∎

APPENDIX 2

TRANSCRIPTIONS

A2.1 TRANSCRIPTION I

Lemma A2.1.1: A problem of the PDE-FORM II can be transcribed into a problem of PDE-FORM I.

Proof: Define $X \triangleq W \times \mathbb{R}^n$, $x(t,u) \triangleq w(t,u) \oplus z(t,u)$,

$$A \triangleq \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix}, \quad F(x(t,u),u(t)) \triangleq \begin{bmatrix} B(z(t,u))w(t,u) + C(z(t,u)) + E(u(t)) \\ f(z(t,u),u(t)) \end{bmatrix}.$$

Then $D(A) = D(\hat{A}) \times \mathbb{R}^n$ and

$$\frac{\partial F}{\partial x}(x,u) = \begin{bmatrix} B(z) & B_z(z)w + C_z(z) \\ 0 & 0 \end{bmatrix}, \quad \frac{\partial F}{\partial u} = \begin{bmatrix} E_u(u) \\ f_u(z,u) \end{bmatrix}.$$

Since B, C, and f are continuously differentiable, it follows that $F \in C^1(X \times \mathbb{R}^n, X)$. $F \in C^2(X \times \mathbb{R}^n, X)$ follows similarly.

For $\hat{x} = \hat{w} \oplus \hat{z}$ and $x = w \oplus z$,

$$\|F(\hat{x},\hat{u}) - F(x,u)\| = \left\| \begin{matrix} B(\hat{z})\hat{w} + C(\hat{z}) + E(\hat{z}) - B(z)w - C(z) - E(z) \\ f(\hat{z},\hat{u}) - f(z,u) \end{matrix} \right\| \tag{A2.1.1}$$

$$= \left\| \begin{matrix} (B(\hat{z}) - B(z))\hat{w} + B(z)(\hat{w} - w) + C(\hat{z}) - C(z) \\ f(\hat{z},\hat{u}) - f(z,h) \end{matrix} \right\|.$$

Choose $S = S_w \oplus S_z \subset W \times \mathbb{R}^n$ such that S is bounded, i.e. there exists b such that $\|x\| < b$ for all $x \in S$. By Assumption 3.3.1(ii) there exists K, L and $\hat{b} < \infty$ such that

$$\|F(\hat{x},\hat{u}) - F(x,u)\| \leq Lb\|\hat{z} - z\| + \hat{b}\|\hat{w} - w\| + L\|\hat{z} - z\| + K[\|\hat{z} - z\| + \|\hat{u} - u\|] \tag{A2.1.2}$$

$$\leq M[\|\hat{x} - x\| + \|\hat{u} - u\|]$$

for some $M < \infty$, showing that Assumption 3.3.1(iva) is valid. Assumptions 3.3.1(ivb) and

3.3.1(vc) can be similarly shown to be valid.

A2.2 TRANSCRIPTION II

We now show how the rotating beam equations (2.2.5)-(2.2.9) can be transcribed into PDE-FORM II. The equations of motion are:

$$\mathrm{m}w_{tt}(t,x) + \mathrm{CI}w_{txxxx}(t,x) + \mathrm{EI}w_{xxxx}(t,x) - m\Omega(t)w(t,x) = -\mathrm{m}\alpha(t)x, x \in [0,1]. \quad \text{(A2.2.1)}$$

The boundary conditions are:

$$w(t,0) = 0, w_x(t,0) = 0, w_{xx}(t,1)=0, \quad \text{(A2.2.2)}$$

$$\mathrm{M}(\Omega^2(t)w(t,1) - w_{tt}(t,1) - \alpha(t)) + \mathrm{CI}w_{txxx}(t,1) + \mathrm{EI}w_{xxx}(t,1) = 0.$$

We define $u(t) \in X \triangleq L_2([0,1]) \times \mathbb{R}$, and $F : X \times \mathbb{R}^2 \to X$:

$$u(t) \triangleq \begin{bmatrix} w(t,x) \\ w(1,x) \end{bmatrix}, \quad F(u(t),\alpha(t),\Omega(t)) \triangleq \begin{bmatrix} \Omega^2(t)w(t,x) - \alpha(t)x \\ \Omega^2(t)w(t,1) - \alpha(t) \end{bmatrix}. \quad \text{(A2.2.3)}$$

We define A with domain of A, $D(A)$:

$$D(A) \triangleq \{ \overline{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \frac{\partial^4}{\partial x^4}w_1 \in L_2([0,1]),\ w_1(0) = \frac{\partial}{\partial x}w_1(0) = \frac{\partial^2}{\partial x^2}w_1(1) = 0, w_1(1) = w_2 \},$$

$$A{:}D(A) \to X \text{ is such that } A\begin{bmatrix} w(t,x) \\ w(t,1) \end{bmatrix} = \begin{bmatrix} \dfrac{\mathrm{EI}}{\mathrm{m}}\dfrac{\partial^4}{\partial x^4}w(t,x) \\ \dfrac{\mathrm{EI}}{\mathrm{M}}\dfrac{\partial^3}{\partial x^3}w(t,1) \end{bmatrix}. \quad \text{(A2.2.4)}$$

We define the operator $Q : D(Q) \to X$:

$$D(Q) = D(A), \quad Q = \frac{C}{E}A, \quad \text{(A2.2.5)}$$

then

$$u_{tt} + Qu_t + Au = F(u,\alpha,\Omega). \tag{A2.2.6}$$

Gibson [Gib.1] has developed a recipe to derive an infinitesimal generator of a semigroup from A and Q if they satisfy the following properties:

(1) $D(A)$ is dense in X.

(2) A is invertible, and A^{-1} is compact.

(3) A is self-adjoint.

(4) A is coercive, i.e. there exists $\rho>0$ such that

$$\langle Ax,x \rangle \geq \rho^2 |x|_X^2. \tag{A2.2.7}$$

(5) A is a closed operator.

(6) Q is a nonnegative, symmetric linear operator.

(7) $D(A) \subset D(Q)$, and there exists $\gamma > 0$ such that

$$|Qx|_X \leq \gamma^2 |Ax|_X, \; x \in D(A). \tag{A2.2.8}$$

Assumptions (6) and (7) are trivially satisfied. The proofs in this section were derived with help from Ywh-Pyng Harn.

Lemma A2.2.1: $D(A)$ is dense in X.

Proof: Choose any $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in X$. Let

$$\overline{z}_n(x) \triangleq \begin{cases} 0 & x \in [-1/n, 1/n) \\ w_1(x) & x \in [1/n, 1 - 1/n) \\ w_2 & x \in [1 - 1/n, 1 + 1/n] \end{cases}. \tag{A2.2.9}$$

Let $\phi_\varepsilon \in C^\infty$ be such that: (1) $\phi_\varepsilon(-x) = \phi_\varepsilon(x)$. (2) $\int_{-\infty}^{\infty} \phi_\varepsilon(x)dx = 1$. (3) $\phi_\varepsilon(x) = 0$, $x \notin (-\varepsilon,\varepsilon)$.

(4) $\phi_\varepsilon(x) \geq 0$. We define

$$z_n \triangleq \bar{z}_n * \phi_{\frac{1}{4n}}, \quad x \in [0,1], \tag{A2.2.10}$$

$$= \int_{-\infty}^{\infty} \phi_{\frac{1}{4n}}(x-y)\bar{z}_n(y)dy.$$

Therefore, $z_n \in C^{\infty}$. It is easily seen that $z_n(0) = \frac{\partial}{\partial x} z_n(0) = \frac{\partial^2}{\partial x^2} z_n(1) = 0$ and $z_n(1) = w_2$, and so, $\begin{bmatrix} z_n \\ w_2 \end{bmatrix} \in D(A)$. By showing that $z_n \to w_1$ we show that $D(A)$ is dense in X.

$$\|z_n - w_1\|_2 \le |z_n - \bar{z}_n|_2 + \|z_n - w_1\|_2. \tag{A2.2.11}$$

We consider the two terms on the right hand side of (A2.2.11) individually.

$$|z_n - \bar{z}_n|_2^2 = \int_0^1 [\int_{\frac{-1}{4n}}^{\frac{1}{4n}} |\bar{z}_n(x-y)\phi_{\frac{1}{4n}}(y)dy - \bar{z}_n(x)|^2]dx \tag{A2.2.12}$$

$$\le \int_0^1 |\int_{\frac{-1}{4n}}^{\frac{1}{4n}} ((\bar{z}_n(x-y) - \bar{z}_n(x))\left[\phi_{\frac{1}{4n}}(y)\right]^{1/2})\left[\phi_{\frac{1}{4n}}(y)\right]^{1/2} dy|^2 dx.$$

Applying the Schwartz Inequality,

$$|z_n - \bar{z}_n|_2^2 \le \int_0^1 \int_{\frac{-1}{4n}}^{\frac{1}{4n}} |\bar{z}_n(x-y) - \bar{z}_n(x)|^2 \phi_{\frac{1}{4n}}(y)dydx, \tag{A2.2.13}$$

$$\le \int_{\frac{-1}{4n}}^{\frac{1}{4n}} dy\left[\phi_{\frac{1}{4n}}(y)\right]\int_0^1 |\bar{z}_n(x-y) - \bar{z}_n(x)|^2 dx.$$

Since the map $y \to \int_0^1 |\bar{z}_n(x-y) - \bar{z}_n(x)|^2 dx$ is equi-continuous in y for all $n \in \mathbb{Z}_+$ (see Theorem 9.5 in Rudin [Rud.1]), for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|\int_0^1 |\bar{z}_n(x-y) - \bar{z}_n(x)|^2 dx \le \varepsilon^2$ if $|y| \le \delta$. Choose n_0 such that $\frac{1}{4n_0} < \delta$. Then for all $n > n_0$,

$\|z_n - \bar{z}_n\|_2 < \varepsilon$.

We consider the second term on the right hand side of (A2.2.11).

$$\|\bar{z}_n - w_1\|_2^2 = \int_0^{\frac{1}{n}} w_1(x)^2 dx + \int_{1-\frac{1}{n}}^{1} (w_2 - w_1(x))^2 dx. \tag{A2.2.14}$$

For all $\varepsilon > 0$, there exists a continuous function $g \in C([0,1])$ such that $\|g - w_1\|_2^2 < \frac{\varepsilon^2}{2}$ (See Theorem 3.14 in Rudin).

$$\|z_n - w_1\|_2^2 = \int_0^{\frac{1}{n}} (w_1(x) - g(x) + g(x))^2 dx + \int_{1-\frac{1}{n}}^{1} (w_2 - g(x) + g(x) - w_1(x))^2 dx \tag{A2.2.15}$$

$$\leq \int_0^{\frac{1}{n}} g(x)^2 dx + \int_0^{\frac{1}{n}} (g(x) - w_1(x))^2 dx + \int_{1-\frac{1}{n}}^{1} (w_2 - g(x))^2 dx + \int_{1-\frac{1}{n}}^{1} (g(x) - w_1(x))^2 dx$$

Since $g(\cdot)$ is continuous, there exists $n_1 > n_0$ such that for all $n > n_1$

$$\int_0^{\frac{1}{n}} g(x)^2 dx + \int_{1-\frac{1}{n}}^{1} (w_2 - g(x))^2 dx \leq \frac{\varepsilon^2}{2}, \tag{A2.2.16}$$

so that $\|z_n - w_1\|_2^2 \leq \varepsilon^2$ and $\|\bar{z}_n - w_1\|_2 < 2\varepsilon$ for all $n > n_0$. ■

Lemma A2.2.2: The operator A is invertible, and A^{-1} is compact.

Proof: We first give the inverse of A, and then show that it is compact. For $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in L_2([0,1]) \times \mathbb{R}$,

$$A^{-1}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \int_0^x\int_0^y\int_1^z\int_1^u v_1(t)dtdudzdy + v_2(\frac{x^2}{2} - \frac{x^3}{6}) \\ \int_0^1\int_0^y\int_1^z\int_1^u v_1(t)dtdudzdy + \frac{v_2}{3} \end{bmatrix}. \quad \text{(A2.2.17)}$$

A^{-1} is compact if and only if $A^{-1}B_0$ is compact where $B_0 \triangleq \{ v \mid \|v\| \leq 1, v \in L_2([0,1]) \times \mathbb{R} \}$. We show that $A^{-1}B_0$ is compact in $C([0,1]) \times \mathbb{R}$. This implies that $A^{-1}B_0$ is compact in the coarser $L_2([0,1]) \times \mathbb{R}$ topology. Consider $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in B_0$. Then $\|v_1\|_2 \leq 1$ and $|v_2| \leq 1$.

$$\left|\frac{d}{dx}w_1(x)\right| \leq \int_0^x\int_z^1\int_u^1 |v_1(t)|dtdudz + |v_2|(x - \frac{x^2}{2})^2 \leq 1, \quad |w_2| \leq 2. \quad \text{(A2.2.14)}$$

Therefore $A^{-1}B_0$ is equi-Lipschitz continuous and for each $x \in [0,1]$, $|w_1(x)| \leq 1$. By the Arzela-Ascoli Theorem, $A^{-1}B_0$ is pre-compact in the $C([0,1]) \times \mathbb{R}$ topology. Since A^{-1} is a bounded linear operator and B_0 is closed, it follows that $A^{-1}B_0$ is closed and therefore compact in $C([0,1]) \times \mathbb{R}$. ■

Lemma A2.2.3: The operator A is self-adjoint.

Proof: Consider $u, v \in D(A)$. Then

$$u(0) = v(0) = \frac{\partial}{\partial x}u(0) = \frac{\partial}{\partial x}v(0) = \frac{\partial^2}{\partial x^2}u(1) = \frac{\partial^2}{\partial x^2}v(1) = 0.$$

$$\langle u, Av \rangle = \langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, A\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \rangle = \langle \begin{bmatrix} u(\cdot) \\ u(1) \end{bmatrix}, \begin{bmatrix} \frac{\partial^4}{\partial x^4}v(\cdot) \\ -\frac{\partial^3}{\partial x^3}v(1) \end{bmatrix} \rangle \quad \text{(A2.2.15)}$$

$$= \int_0^1 u(x)\frac{\partial^4}{\partial x^4}v(x)dx - u(1)\frac{\partial^3}{\partial x^3}v(1).$$

Applying integration by parts,

$$\langle u,Av \rangle = \int_0^1 \frac{\partial^4}{\partial x^4}u(x)v(x)dx - \frac{\partial^3}{\partial x^3}u(1)v(1) = \langle Au,v \rangle. \tag{A2.2.16}$$

Therefore $u \in D(A^*)$ and $D(A) \subset D(A^*)$. Now suppose $y \in D(A^*)$ and $p = A^*y$. Then, $\langle y,Au \rangle = \langle p,u \rangle$ for all $u \in D(A)$. Since $p \in X$ and A is invertible, there exists $v \in D(A)$ such that $Av = p$. $\langle y,Au \rangle = \langle Av,u \rangle = \langle v,Au \rangle$ for all $u \in D(A)$. Since $R(A) = X$, $y = v$ so $y \in D(A)$. A is self-adjoint. ∎

Lemma A2.2.4: The operator A is coercive.

Proof:

$$\langle Au,u \rangle = \int_0^1 u(x)\frac{\partial^4}{\partial x^4}u(x)dx - u(1)\frac{\partial^3}{\partial x^3}u(1) = \int_0^1 \left[\frac{\partial^2}{\partial x^2}u(x)\right]^2 dx. \tag{A2.2.17}$$

Since $u(0) = 0$,

$$|u(x)|^2 = |u(0) + \int_0^x \frac{\partial}{\partial x}u(y)dy|^2 \le \|\frac{\partial}{\partial x}u\|_2^2. \tag{A2.2.18}$$

Consequently $\|u\|_2 \le \|\frac{\partial u}{\partial x}\|_2$. Similarly, $\|\frac{\partial u}{\partial x}\|_2 \le \|\frac{\partial^2 u}{\partial x^2}\|_2$. Therefore,

$$\|u\|_X^2 = \|u\|_2^2 + u(1)^2 \le 2\|\frac{\partial^2 u}{\partial x^2}\|_2^2 \le 2\langle Au,u \rangle. \tag{A2.2.19}$$

∎

Lemma A2.2.5: The operator A is closed.

The operator A is closed if for any sequence $\{ u_i \} \subset D(A)$ such that $u_i \to u$ and $Au_i \to v$, then $u \in D(A)$ and $Au = v$. Given a sequence $\{ u_i \} \subset D(A)$, such that $u_i \to u$, we define $v_i \triangleq Au_i$. Since A is invertible, there exists $\hat{u} = A^{-1}v$. If $u = \hat{u}$ then A is closed. Since A^{-1} is compact, there exists $K < \infty$ such that

$$\|u_i - \hat{u}\| = \|A^{-1}v_i - A^{-1}v\| = \|A^{-1}(v_i - v)\| \le k\|v_i - v\|. \tag{A2.2.20}$$

Since $v_i \to v$, $u_i \to \hat{u}$. Since $u_i \to u$, $\hat{u} = u$. ■

With A and Q satisfying the necessary properties we can employ Gibson's method to derive the infinitesimal generator of a contraction semigroup. We give a brief outline of this derivation. See Gibson for the details. We define the space W:

$$W \triangleq D(A^{1/2}) \times X, \tag{A2.2.21}$$

so that if $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in V$, then

$$\left\| \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\| = \langle v_1, Av_1 \rangle + \langle v_2, v_2 \rangle, \tag{A2.2.22}$$

where $\langle \cdot, \cdot \rangle$ is the L_2 inner product. Define $v(t) \in V$:

$$v(t) \triangleq \begin{bmatrix} w_1(t,x) \\ w_1(t,1) \\ w_2(t,x) \\ w_2(t,1) \end{bmatrix} \triangleq \begin{bmatrix} w(t,x) \\ w(t,1) \\ \dot{w}(t,x) \\ \dot{w}(t,1) \end{bmatrix}. \tag{A2.2.23}$$

Define the operator $\bar{A} : D(\bar{A}) \to V$ where $D(\bar{A}) = D(A) \times D(A) \subset V$:

$$\bar{A}v(t) \triangleq \begin{bmatrix} 0 & I \\ -A & -Q \end{bmatrix} v(t) = \begin{bmatrix} w_2(t,x) \\ w_2(t,1) \\ -\frac{\mathrm{CI}}{\mathrm{m}}\frac{\partial^4}{\partial x^4}w_2(t,x) - \frac{\mathrm{EI}}{\mathrm{m}}\frac{\partial^4}{\partial x^4}w_1(t,x) \\ \frac{\mathrm{CI}}{\mathrm{M}}\frac{\partial^3}{\partial x^3}w_2(t,1) + \frac{\mathrm{EI}}{\mathrm{M}}\frac{\partial^3}{\partial x^3}w_1(t,1) \end{bmatrix}. \tag{A2.2.24}$$

By Theorem 2.2 in Gibson, there exists an extension of $\bar{A}$ to $\hat{A}$ where $\hat{A}$ is the generator of a contraction semigroup. Showalter[Sho.1] has shown that $\hat{A}$ generates an analytic semigroup.

To complete this section, we define: $B:\mathbb{R} \to \mathrm{B}(V,V)$ and $E:\mathbb{R} \to X$:

$$B(\Omega(t))v(t) = \begin{bmatrix} 0 \\ 0 \\ \Omega(t)^2 w_1(t,x) \\ \Omega(t)^2 w_1(t,1) \end{bmatrix}, \quad E(\alpha(t)) = \begin{bmatrix} 0 \\ 0 \\ -\alpha(t)x \\ -\alpha(t) \end{bmatrix}. \tag{A2.2.25}$$

Then from (A2.2.1) and (A2.2.2), we obtain the system of coupled partial and ordinary equations:

$$\frac{d}{dt}v(t) = \hat{A}v(t) + B(\Omega(t))v(t) + E(\alpha(t)), \tag{A2.2.26}$$

$$\frac{d}{dt}\Omega(t) = \alpha(t), \tag{A2.2.27}$$

and therefore all assumptions of PDE-FORM II are satisfied.

A2.3 EXISTENCE OF MILD SOLUTIONS

Lemma A2.3.1: (Existence of Mild Solutions for PDE-FORM I with Condition 3.2.1.)

Given the system:

$$\frac{d}{dt}x(t) = Ax(t) + f(x(t),u(t)),\ x(0) = x_0,\ u \in G \triangleq \{\ u \in L_\infty^m([0,1]) \mid u(t) \in U\ \} \tag{A2.3.1}$$

for all $t \in [0,1]$, where X is a Hilbert space, $A:D(A) \to X$ is the infinitesimal generator of a continuous semigroup, $T(t)$, and $f:X \times U \to X$ is a nonlinear operator and U is a compact convex subset of $\mathbb{R}^m$. If f is Lipschitz continuous with constant $K_f < \infty$ over $X \times U$, then a mild solution to (A2.3.1) exists.

Proof: Define a map $F:C([0,1];X) \to C([0,1];X)$ by

$$(Fx)(t) = T(t)x_0 + \int_0^t T(t-s)f(x(s),u(s))ds. \tag{A2.3.1}$$

Therefore, for all $t \in [0,1]$,

$$\|(Fx)(t) - (Fy)(t)\| \le \int_0^t \|T(t-s)\|K_f\|x(s) - y(s)\|ds \le M_A e^{\omega_A} K_f \|x - y\|_\infty. \tag{A2.3.2}$$

$$(F^2x)(t) = T(t)x_0 + \int_0^t T(t-s)f(T(t)x_0 + \int_0^s T(s-r)f(x(r),u(r))dr, u(s))ds. \tag{A2.3.3}$$

$$|(F^2x)(t) - (F^2y)(t)| \le (M_A e^{\omega_A})^2 \int_0^t K_f \int_0^s |f(x(r),u(r)) - f(y(r),u(r))| dr ds \tag{A2.3.4}$$

$$\le (M_A e^{\omega_A})^2 K_f^2 \int_0^t \int_0^s |x(r) - y(r)| dr ds \le \frac{(M_A e^{\omega_A} K_f)^2}{2} |x - y|_\infty.$$

By induction,

$$|F^n(x) - F^n(y)|_\infty \le \frac{(M_A e^{\omega_A} K_f)^n}{n!} |x - y|_\infty. \tag{A2.3.5}$$

There exists $n < \infty$ such that $\frac{(M_A e^{\omega_A} K_f)^n}{n!} < 1$ and so F is a contraction mapping. There exists a unique fixed point $x^* \in C([0,1];X)$ such that $x^* = Fx^*$ or

$$x^*(t) = T(t)x_0 + \int_0^t T(t-s)f(x^*(s),u(s))ds \tag{A2.3.6}$$

for all $t \in [0,1]$.