The Linear Quadratic Control Problem With Unconstrained Terminal Condition

Justice Sam Bansah

10247727

This Thesis Is Submitted To University Of Ghana, Legon In Partial Fulfillment Of The Requirement For The Award Of Mphil Mathematics Degree

July, 2015

Declaration

This thesis was written in the Department of Mathematics, University of Ghana, Legon from September 2014 to July 2015 in partial fulfillment of the requirement for the award of Master of Pholosophy degree in Mathematics under the supervision of Dr. Benoit Sehba and Dr. Margaret McIntyre of the University of Ghana.

I hereby declare that, except where due acknowledgement is made, this work has never been presented wholly or in part for the award of a degree at the University of Ghana or any other institution.

Signature:..
Student: Justice Sam Bansah

Signature:..
Supervisor: Dr. Benoit Sehba

Signature:..
Supervisor: Dr. Margaret McIntyre

Abstract

Linear Quadratic Control Problems are control problems with a quadratic cost function and linear dynamic sytem and a linear terminal constraint. This work looks at linear quadratic control problems without terminal conditions. We will first look at the controllability and observability of linear dynamical systems and then establish the necessary conditons for the variation in the cost criterion to be nonnegative for strong perturbations in the control. These conditions are the first order necessary conditions for optimality. We shall also consider the necessary and sufficient conditon for positivity of the quadratic cost criterion. Moreover, necessary and sufficient conditions for strong positivity are derived and we shall show that these conditions are based on the existence of solution to a Riccati matrix differential equation. The symplectic property of the Hamiltonian system helps us to derive the Riccati matrix differential equation. We also look at some of the properties of the Riccati variable. Using these ideas, as an illustration, we consider an application in the control of disease, by considering a variant of the SIR model.

Dedication

I dedicate this work to my lovely Mum and Dad and my little brother David.

Aknowledgement

I first of all thank the Lord Jesus Christ who is the embodiment of all wisdom, knowledge and understanding for helping me complete this thesis successfully.

I wish to express my sincere gratitude to Dr. Margaret McIntyre and Dr. Benoit Sehba for the keen interest they hard in this thesis and for all the support and guidance which contributed to the success of this work. I also want say a big thank you to Professor Allotey and CARNEGIE, University of Ghana, for their kind supports.

Contents

Chapter 1

Introduction

1.1 Overview

In control theory, the problem is to find the input which will cause the state or the output to behave in a desired manner. For example, the classical servomechanism problem is concerned with making certain variables associated with a dynamical system, say a satelite or telescope, follow a particular path and also the standard problem in the guidance of rocket is to make certain variables associated with a dynamical system take on particular values at a point in time. These two examples are referred to as the *servomechanism controllability* and *ballistic controllability* problems respectively [Ro].

In this work, we will consider a special control problem, *The Linear Quadratic Control Problem*, with an unconstrained terminal condition where we establish the necessary and sufficient conditions for optimality. The most used result in optimal control theory is that of the solution to the linear quadratic problem where the dynamic system and the terminal constraints are linear and the performance criterion is a quadratic function of the state and control variables. The solution to this problem produces a control variable as a linear function of the state variable. This solution forms the basis of modern control synthesis techniques because it produces controllers for multi-input/multi-output system for both time-varying and time-invariant systems [AM, BY]. This presentation makes use of the symplectic property of Hamiltonian system. We also consider conditions for which the second variation is strongly positive.

We will also see that the necessary and sufficient condition for optimality of a linear quadratic control problem without terminal constraint depends on the existence of solution to a matrix Riccati differential equation. We will discuss some of the properties of the Riccati variable and then we will consider an application in the control of disease, by considering a variant of the SIR model.

The last section of this present chapter considers an example of optimal control in a classical variational approach. Chapters two and three establish some fundamental results which will help in the development of the theoretical analysis. The last section of chapter three looks at integral cost functions and their significance. This gives a foretaste of the main work.

1.2 Statement of a General Optimal Control Problem

General control problems are described in several ways. In this work, the general control problem considered is described by the cost functional or performance index which quantitatively compares the effectiveness of the various controllers. The cost functional is an acceptable quantitative criterion for the efficiency of each controller, $u(t)$, on $t_0 \leq t \leq t_f$ in the class of all controllers. For example, if the class of controller consists of controllers on various time intervals that steers x_0 to the target, one often defines the cost of $u(t)$ to be

$$J(u) = \int_{t_0}^{t_f} f^0(t, x(t), u(t))\mathrm{d}t,$$

where $f^0(t, x(t), u(t))$ is a given continuous function. If $f^0(t, x(t), u(t)) = 1$ for all t then

$$J(u) = t_f - t_0$$

and we have the minimal time problem. Somtimes the set of admissible controllers consists of controllers on a fixed specified time duration, say $t_0 \leq t \leq T$, without the requirement of reaching the target, say the curve $\xi(t)$. Then, one often has the cost functional

$$J(u) = |x(T) - \xi(T)| + \int_{t_0}^{T} f^0(t, x(t), u(t))\mathrm{d}t.$$

The fundamental problem of optimal control theory can be formulated as a problem of the calculus of variation with the Bolza problem being a typical example. We thus describe the Bolza problem since the associated second variation with which we shall be much concerned, appears as such a problem.

1.2.1 The Bolza Problem

The Bolza problem is stated as follows;
Determine the control function, $u(\cdot)$, which minimizes the cost function

$$J = \phi(x(t_f), t_f) + \int_{t_0}^{t_f} L(x, u, t)\mathrm{d}t$$

where the system equation is

$$\dot{x} = f(x, u, t)$$

subject to the constraints

$$x(t_0) = x_0 \quad \text{and} \quad \Psi(x(t_f), t_f) = 0$$

where $t \in [t_0, t_f], x \in \mathcal{D}(\mathbb{R}^n)$,the space of differentiable functions into $\mathbb{R}^n$, i.e. $x : [t_0, t_f] \to \mathbb{R}^n$, $u \in \mathcal{P}(\mathbb{R}^m)$, the space of piecewise continuous functions into $\mathbb{R}^m$, i.e. $u : [t_0, t_f] \to \mathbb{R}^m$. More specifically, we let $u(\cdot) \in U$ where

$$U = \{u(\cdot) : |u_i(t)| < \infty, t_0 \leq t \leq t_f, i = 1, 2, \cdots, m\}$$

is the set of bounded piecewise continuous functions. L and ϕ are such that

$$L : \mathbb{R}^{m+n+1} \longrightarrow \mathbb{R} \quad \text{and} \quad \phi : \mathbb{R}^{m+1} \longrightarrow \mathbb{R}$$

and the terminal constraint function, Ψ, is an $s-$dimensional column vector function of $x(t_f)$ at t_f. We assume that L, ϕ, Ψ are smooth. The initial time, t_0, is given explicitly but the final time t_f may not be specified [DD].

1.2.2 The Adjoining System Equations

We adjoin the cost function to the constraints as follows

$$J = \phi(x(t_f), t_f) + \nu^T \Psi(x(t_f), t_f) + \int_{t_0}^{t_f} L(x, u, t) + \lambda^T (f(x, u, t) - \dot{x})\mathrm{d}t$$

where λ and ν denote an $n-$dimensional vector of Lagrange multiplier functions of time and an $s-$dimensional vector of Lagrange multipliers associated with Ψ respectively. The Hamiltonian for this problem is

$$H(x, u.\lambda, t) = L(x, u, t) + \lambda^T f(x, u, t)$$

and the necessary conditions (*Pontryagin's Principle*) for optimality holds along an optimal trajectory and the Legendre-Clebsch condition, $H_{uu} \geq 0$, also holds along any optimal path. One can go ahead to also establish the global necessary and sufficient condition for optimality which is the *Hamilton-Jacobi-Bellman Equation* (see appendix) [Sp].

1.3 Classical Example

We examine the optimal control problem from the viewpoint of the classical calculus of variation. We do this without regards to continuity or differentiability. Consider the control process in $\mathbb{R}^n$, i.e. $x(t)$ is a real n-vector

$$\dot{x}(t) = f(x(t), u(t)); \quad x(0) = x_0$$

with controllers $u(t) \in \mathbb{R}^m$ on $0 \leq t \leq 1$. For each controller, $u(t)$, there is a corresponding response, $x(t)$, with $x(0)$ at the fixed initial state x_0 and with cost

$$J(u) = \int_0^1 h(x(t), u(t))\mathrm{d}t.$$

Let $u^*(t)$ be an optimal controller minimizing $J(u)$ and $x^*(t)$ be the corresponding optimal response. Let

$$u(t, \epsilon) = u^*(t) + \epsilon\delta u(t)$$

be a one-parameter family of perturbations. Then, there will be a corrresponding perturbation in the response as

$$x(t, \epsilon) = x^*(t) + \epsilon\delta x(t) + o(\epsilon); \quad \delta x(0) = 0.$$

So

$$J(u(t, \epsilon)) = \int_0^1 h(x(t, \epsilon), u(t, \epsilon))\mathrm{d}t$$

which implies that

$$\left.\frac{\partial J}{\partial \epsilon}\right|_{\epsilon=0} = \delta J = \int_0^1 \left(\frac{\partial h(t)}{\partial x}\delta x(t) + \frac{\partial h(t)}{\partial u}\delta u(t) \right) \mathrm{d}t$$

upon the assumption that $h(\cdot)$ and its derivatives are continuous on $[0, 1]$. Now consider the first variation of the dynamical system, $\dot{x} = f(x, u)$, i.e.

$$\delta\dot{x} = \frac{\partial}{\partial x} f(t) \cdot \delta x(t) + \frac{\partial}{\partial u} f(t) \cdot \delta u(t); \quad \delta x(0) = 0.$$

Then by result of variation of parameters (to be seen later), this has the solution

$$\delta x(t) = \int_0^t \phi(t,s)\frac{\partial}{\partial u} f(s)\cdot \delta u(s)\mathrm{d}s$$

where $\phi(t,s)$ is the transition matrix of $\delta \dot{x} = \frac{\partial}{\partial x} f(t)\cdot \delta x(t)$. Then,

$$\delta J = \int_0^1 \left(\frac{\partial h(t)}{\partial x} \cdot \int_0^t \phi(t,s)\frac{\partial}{\partial u} f(s)\cdot \delta u(s)\mathrm{d}s + \frac{\partial h(t)}{\partial u}\delta u(t)\right) \mathrm{d}t.$$

We introduce the vector function

$$\eta^*(t) = -\eta_0 \phi^{-1}(t,0) + \int_0^t \frac{\partial}{\partial x} h(s)\cdot \phi^{-1}(t,s)\mathrm{d}s$$

which will be a solution to the system

$$\dot{\eta}(t) = -\eta(t)\frac{\partial f}{\partial x} + \frac{\partial h}{\partial x}; \quad \eta(0) = \eta_0$$

where η_0 is some constant vector chosen so that $\eta^*(1) = 0$. Now, if the transition matrix were to be a constant, then it must be the identity matrix since $\phi(t,t) = I$ for all t where I is the identity matrix. Hence

$$\eta^*(t) = -\eta_0 + \int_0^t \frac{\partial}{\partial x} h(s)\mathrm{d}s.$$

Define the Hamiltonian function of $2n + m$ real variables as

$$H(\eta, x, u) = \eta f(x,u) - h(x,u)$$

and thus

$$\frac{\partial H}{\partial x} = -\dot{\eta}; \quad \eta(1) = 0 \quad \text{and} \quad \frac{\partial H}{\partial \eta} = f(x,u) = \dot{x}; \quad x(0) = x_0.$$

We now integrate by parts the first term under the integral sign of δJ. Let

$$a = \int_0^t \phi(t,s)\frac{\partial}{\partial u} f(s)\cdot \delta u(s)\mathrm{d}s \quad \text{then} \quad a' = \frac{\partial f(t)}{\partial u}\delta u(t)$$

and

$$b' = \frac{\partial h(t)}{\partial x}; \quad b_0 = 0 \quad \text{then} \quad b = \int_0^t \frac{\partial h(s)}{\partial x}\mathrm{d}s.$$

(Note that the coefficient of b in the dynamic system $b' = \frac{\partial h(t)}{\partial x}$ is zero and thus $\dot{\phi} = 0$ which implies that ϕ is a constant which must be the identity matrix. See Theorem 2.0.0.5). Therefore,

$$
\begin{aligned}
\delta J &= \left(\int_0^t \frac{\partial}{\partial x} h(s)\mathrm{d}s\right)\left(\int_0^t \phi(t,s)\frac{\partial}{\partial u} f(s)\cdot\delta u(s)\mathrm{d}s\right)\Bigg|_0^1 \\
&\quad - \int_0^1 \left[\left(\int_0^t \frac{\partial}{\partial x} h(s)\mathrm{d}s\right)\left(\frac{\partial f(t)}{\partial u}\delta u(t)\right)\right]\mathrm{d}t + \int_0^1 \frac{\partial h(t)}{\partial u}\delta u(t)\mathrm{d}t \\
&= \left(\int_0^1 \frac{\partial}{\partial x} h(s)\mathrm{d}s\right)\left(\int_0^1 \frac{\partial}{\partial u} f(t)\cdot\delta u(t)\mathrm{d}t\right) \\
&\quad - \int_0^1 \left[\left(\int_0^t \frac{\partial}{\partial x} h(s)\mathrm{d}s\right)\left(\frac{\partial f(t)}{\partial u}\delta u(t)\right)\right]\mathrm{d}t + \int_0^1 \frac{\partial h(t)}{\partial u}\delta u(t)\mathrm{d}t \\
&= (\eta^*(1)+\eta_0)\left(\int_0^1 \frac{\partial}{\partial u} f(t)\cdot\delta u(t)\mathrm{d}t\right) \\
&\quad - \int_0^1 \left[\left(\int_0^t \frac{\partial}{\partial x} h(s)\mathrm{d}s\right)\left(\frac{\partial f(t)}{\partial u}\delta u(t)\right)\right]\mathrm{d}t + \int_0^1 \frac{\partial h(t)}{\partial u}\delta u(t)\mathrm{d}t \\
&= \int_0^1 \left[\eta_0 - \left(\int_0^t \frac{\partial}{\partial x} h(s)\mathrm{d}s\right)\right]\left(\frac{\partial f(t)}{\partial u}\delta u(t)\right)\mathrm{d}t + \int_0^1 \frac{\partial h(t)}{\partial u}\delta u(t)\mathrm{d}t \\
&= \int_0^1 \left(-\eta\frac{\partial f(t)}{\partial u} + \frac{\partial h(t)}{\partial u}\right)\delta u(t)\mathrm{d}t.
\end{aligned}
$$

Now for a minimum, $\delta J = 0$ which implies that

$$-\eta\frac{\partial f(t)}{\partial u} + \frac{\partial h(t)}{\partial u} = 0 \quad \forall \delta u(t)$$

which implies that $\frac{\partial H}{\partial u} = 0$ about any optimal $u^*(t)$ and $x^*(t)$. i.e.

$$H(\eta^*(t), x^*(t), u^*(t)) = \max_{u\in\mathbb{R}^m} H(\eta^*(t), x^*(t), u(t))$$

is the maximum principle, which plays an important role in the theory of optimal control. The minimum principle states that, for the hamiltonian function defined, in order for a controller, $u(t)$ and a trajectory $x(t)$ to be optimal, it is necessary that there exists a nonzero continuous function, in this case η, such that the above equation holds and at the terminal time, $t = 1$, $\eta(1) \leq 0$ and $H(\eta^*(1), x^*(1), u^*(1)) = 0$ [Br1, LS].

The system of equations

$$\dot{x} = \frac{\partial H}{\partial \eta}, \quad \dot{\eta} = -\frac{\partial H}{\partial x} \quad \text{and} \quad \frac{\partial H}{\partial u} = 0$$

are the *Euler-Lagrange equations* of this variational problem. If there are no boundary conditions on the controllers, these conditions are termed the *Weierstrass Necessary*

Conditions for an extremal [GS]. The functions $\eta^*(t)$ are the Lagrange multipliers for the variational problem.

Chapter 2

The Transition Matrix and The Peano-Baker Series

One way of finding the inverse of an $n \times n$ matrix A is to solve the n linear equations

$$Az_i = b_i \quad i = 1, \cdots, n$$

where $b_i = [0, 0, \cdots, 1, \cdots, 0, 0]^T$ is an n-column vector with 1 at the ith-position and simply arrange the solutions $z_i, \quad i = 1, \cdots, n,$ in a matrix whose columns are the solution vectors. That is $[z_1, \cdots, z_n] = Z$ so that

$$AZ = I \Longrightarrow Z = A^{-1}.$$

Consider the system of differential equations below.

$$\dot{x}(t) = A(t)x(t) \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{2.1}$$

Now assume that $x_1, x_2, \cdots, x_n$ are n linearly independent solutions of (2.1) subject to the boundary condition $x_i(t_0) = [0, 0, \cdots, 1, \cdots, 0, 0]^T$; an n-column vector with 1 at the ith-position. Then the square matrix

$$\Phi(t, t_0) = [\phi_1(t, t_0), \cdots, \phi_n(t, t_0)]$$

is the *fundamental matrix solution* of (2.1) where $\phi_i(t, t_0) = x_i(t)$ for $i = 1, 2, ..., n$. Then

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t, t_0) &= [\dot{\phi}_1, \cdots, \dot{\phi}_n] \\
&= \dot{x}(t) \\
&= A(t)x(t); \quad \text{from (2.1)} \\
&= A(t)\Phi(t, t_0); \quad \text{since } \Phi(t, t_0) = x(t)
\end{aligned}$$

$$\therefore \frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,t_0) = A(t)\Phi(t,t_0), \quad \Phi(t_0,t_0) = I.$$

Now, let $x(t) = \Phi(t,t_0)x(t_0)$. Then this will satisfy (2.1) and thus it is a solution.

Definition 2.0.0.1. If $E_{ij}(M)$ denotes the $ijth$ element of a matrix M, then we say that a sequence of matrices $\{M_k\}_{k=1}^{\infty}$ whose elements depend on time, **converges uniformly** on the interval $[t_0,t_1]$ if each scalar sequence $\{E_{ij}(M_k)\}_{k=1}^{\infty}$ converges uniformly.

The following theorem establishes the existence of such Φ.

Theorem 2.0.0.2 (Ro). *If A is a square matrix whose elements are continuous functions of time on the interval $[t_0,t_1]$ and if the sequence of matrices, M_k, is defined recursively by*

$$\begin{aligned} M_0 &= I \\ M_k &= I + \int_{t_0}^{t} A(\sigma)M_{k-1}(\sigma)\mathrm{d}\sigma. \end{aligned}$$

Then, the sequence of matrices $\{M_k\}_{k=1}^{\infty}$ converges uniformly on the given interval. Moreover, if the limit function is denoted by Φ, then for $t \in [t_0,t_1]$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,t_0) = A(t)\Phi(t,t_0), \quad \Phi(t_0,t_0) = I \tag{2.2}$$

and the solution of (2.1) which passes through x_0 at $t = t_0$ is

$$\Phi(t,t_0)x_0.$$

Proof. Define η as the maximum absolute value of any entry in $A(t)$. That is

$$\eta(t) = \max_{ij} |a_{ij}(t)|, \tag{2.3}$$

where $a_{ij} = E_{ij}(A)$. Let $\gamma(t)$ be defined as

$$\gamma(t) = \int_{t_0}^{t} \eta(\sigma)\mathrm{d}\sigma. \tag{2.4}$$

If A and B are $n \times n$ matrices, then

$$E_{ij}(AB) \leq n \max_{ij} |E_{ij}(A)| \max_{ij} |E_{ij}(B)|. \tag{2.5}$$

Given the recursion,

$$M_k = I + \int_{t_0}^{t} A(\sigma_1)M_{k-1}(\sigma_1)\mathrm{d}\sigma_1$$

we have that

$$M_{k-1} = I + \int_{t_0}^{t} A(\sigma_1)M_{k-2}(\sigma_1)\mathrm{d}\sigma_1.$$

Therefore

$$M_k(t,t_0) - M_{k-1}(t,t_0) = \int_{t_0}^{t} A(\sigma_1)[M_{k-1}(\sigma_1) - M_{k-2}(\sigma_1)]\mathrm{d}\sigma_1.$$

Now,

$$M_{k-1}(\sigma_1) - M_{k-2}(\sigma_1) = \int_{t_0}^{\sigma_1} A(\sigma_2)[M_{k-2}(\sigma_2) - M_{k-3}(\sigma_2)]\mathrm{d}\sigma_2.$$

Therefore

$$M_k(t,t_0) - M_{k-1}(t,t_0) = \int_{t_0}^{t} A(\sigma_1)\left[\int_{t_0}^{\sigma_1} A(\sigma_2)[M_{k-2}(\sigma_2) - M_{k-3}(\sigma_2)]\mathrm{d}\sigma_2\right]\mathrm{d}\sigma_1.$$

Upon successive substitution, we obtain

$$M_k(t,t_0) - M_{k-1}(t,t_0) = \int_{t_0}^{t} A(\sigma_1)\int_{t_0}^{\sigma_1} A(\sigma_2)\cdots\int_{t_0}^{\sigma_{k-1}} A(\sigma_k)\mathrm{d}\sigma_k\cdots\mathrm{d}\sigma_2\mathrm{d}\sigma_1.$$

Hence

$$\begin{aligned}
&E_{ij}(M_k(t,t_0) - M_{k-1}(t,t_0))\\
&= E_{ij}\left[\int_{t_0}^{t} A(\sigma_1)\int_{t_0}^{\sigma_1} A(\sigma_2)\cdots\int_{t_0}^{\sigma_{k-1}} A(\sigma_k)\mathrm{d}\sigma_k\cdots\mathrm{d}\sigma_2\mathrm{d}\sigma_1\right]\\
&\leq \int_{t_0}^{t}\int_{t_0}^{\sigma_1}\cdots\int_{t_0}^{\sigma_{k-1}} n^{k-1}\eta(\sigma_1)\eta(\sigma_2)\cdots\eta(\sigma_k)\mathrm{d}\sigma_k\cdots\mathrm{d}\sigma_1\\
&= \tfrac{n^{k-1}\gamma^k(t)}{k!}; \quad \text{(2.3)-(2.5) are used (appendix Lemma A.6.0.20).}
\end{aligned}$$

Therefore

$$\begin{aligned}
E_{ij}(M_0) + \sum_{k=1}^{\infty} E_{ij}(M_k(t,t_0) - M_{k-1}(t,t_0)) &\leq \sum_{k=1}^{\infty}\frac{n^{k-1}\gamma^k(t)}{k!} + 1\\
&= \frac{1}{n}\sum_{k=1}^{\infty}\frac{(n\gamma(t))^k}{k!} + 1\\
&= \frac{1}{n}\left(\mathrm{e}^{n\gamma(t)} - 1\right) + 1.
\end{aligned}$$

Thus, the matrix series converges since the sum is bounded by the *Weierstrass M-Test* (see appendix). Now, by iterating the recursion, $M_k = I + \int_{t_0}^{t} A(\sigma)M_{k-1}(\sigma)\mathrm{d}\sigma$, we end up with the following

$$\begin{aligned}
M_1 &= I + \int_{t_0}^{t} A(\sigma)\mathrm{d}\sigma\\
M_2 &= I + \int_{t_0}^{t} A(\sigma)M_1(\sigma_1)\mathrm{d}\sigma\\
&= I + \int_{t_0}^{t} A(\sigma)\left[I + \int_{t_0}^{\sigma_1} A(\sigma_2(\sigma_1)\mathrm{d}\sigma_2\right]\mathrm{d}\sigma_1.
\end{aligned}$$

$$\therefore M_k(t) = I + \sum_{k=1}^{\infty} I_k(t) \tag{2.6}$$

where $I_k(t) = \int_{t_0}^{t} A(\sigma_1) \int_{t_0}^{\sigma_1} A(\sigma_2) \cdots \int_{t_0}^{\sigma_{k-1}} A(\sigma_k) \mathrm{d}\sigma_k \cdots \mathrm{d}\sigma_2 \mathrm{d}\sigma_1$. We can show that (see appendix) $I_k(t) = \frac{1}{k!} \left(\int_{t_0}^{t} A(\sigma) \mathrm{d}\sigma \right)^k$.

$$\therefore \lim_{k \to \infty} M_k(t) = I + \sum_{k=1}^{\infty} I_k(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\int_{t_0}^{t} A(\sigma) \mathrm{d}\sigma \right)^k = \exp\left(\int_{t_0}^{t} A(\sigma) \mathrm{d}\sigma \right).$$

Hence $\Phi(t, t_0) = \exp\left(\int_{t_0}^{t} A(\sigma) \mathrm{d}\sigma \right)$ which will satisfy (2.2) and thus $\Phi(t, t_0) x_0$ satisfies (2.1) □

Remark 2.0.0.3. (2.6) is called the **Peano-Baker** series for the solution of the matrix equation (2.2). The matrix $\Phi(t, t_0)$ is called the **Transition Matrix (or Matrizant)** of (2.1). The transition matrix specifies how the state of a dynamic system evolves in time [Ro].

Theorem 2.0.0.4 (Ro)**.** *If Φ is a transition matrix, then it satisfies the functional equation*

$$\Phi(t, t_0) = \Phi(t, t_1)\Phi(t_1, t_0).$$

Proof. Let $\dot{x}(t) = A(t)x(t)$. Given an initial condition, $x(t_0) = x_0$, then by Theorem 2.0.0.5, this will have a unique solution which is given by $x(t) = \Phi(t, t_0)x_0$. Now, at any other time, t_1, the state vector is given by $x(t_1) = \Phi(t_1, t_0)x_0$. We can consequently write

$$x(t) = \Phi(t, t_1)x(t_1) = \Phi(t, t_1)\Phi(t_1, t_0)x_0$$

which implies that $\Phi(t, t_0)x_0 = \Phi(t, t_1)\Phi(t_1, t_0)x_0$. Since the initial condition is arbitrary, it follows that

$$\Phi(t, t_0) = \Phi(t, t_1)\Phi(t_1, t_0).$$

□

Theorem 2.0.0.5 (**Variation of Constant Formula** (Ro))**.** *If $\Phi(t, t_0)$ is the transition matrix for $\dot{x}(t) = A(t)x(t)$, then the unique solution of*
$\dot{x}(t) = A(t)x(t) + f(t); \quad x(t_0) = x_0$ is given by

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, \sigma) f(\sigma) \mathrm{d}\sigma$$

for all $x(t) \in \mathbb{R}^n$.

Proof. Suppose Φ is the transition matrix of $\dot{x}(t) = A(t)x(t)$ and let

$$\dot{x}(t) = A(t)x(t) + f(t). \tag{2.7}$$

Suppose also that $A(t) = 0, \forall t$. Then $\dot{x}(t) = f(t)$ which implies that

$$x(t) = x_0 + \int_{t_0}^{t} f(\sigma)\mathrm{d}\sigma.$$

The idea now is to transform (2.7) by using the substitution $z(t) = \Phi^{-1}(t, t_0)x(t)$. Then,

$$x(t) = \Phi(t, t_0)z(t). \tag{2.8}$$

Hence, upon using (2.2), we obtain

$$\dot{x} = \Phi\dot{z} + \dot{\Phi}z = A(t)x(t) + f(t).$$

Therefore

$$\Phi\dot{z} + (A(t)\Phi)\Phi^{-1}x(t) = A(t)x(t) + f(t)$$

and so

$$\Phi(t, t_0)\dot{z}(t) = f(t) \quad \text{thus} \quad \dot{z}(t) = \Phi^{-1}(t, t_0)f(t).$$

Similarly, by making the substitution $z(t) = \Phi(t, t_0)x(t)$ we obtain $\dot{z}(t) = \Phi(t, t_0)f(t)$. Hence

$$z(t) = z(t_0) + \int_{t_0}^{t} \Phi(t_0, \sigma)f(\sigma)\mathrm{d}\sigma$$

and since $z(t_0) = \Phi(t_0, t_0)x(t_0) = x_0$,

$$z(t) = x_0 + \int_{t_0}^{t} \Phi(t_0, \sigma)f(\sigma)\mathrm{d}\sigma$$

and thus (2.8) becomes

$$\begin{aligned} x(t) &= \Phi(t, t_0)\left[x_0 + \int_{t_0}^{t} \Phi(t_0, \sigma)f(\sigma)\mathrm{d}\sigma\right] \\ &= \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, \sigma)f(\sigma)\mathrm{d}\sigma; \quad \text{using Theorem 2.0.0.4.} \end{aligned}$$

To prove the uniqueness, suppose the system has two solutions, $x(t), y(t)$, then

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, \sigma)f(\sigma)\mathrm{d}\sigma \quad \text{and} \quad y(t) = \Phi(t, t_0)y_0 + \int_{t_0}^{t} \Phi(t, \sigma)f(\sigma)\mathrm{d}\sigma.$$

Therefore,

$$\dot{x}(t) = A(t)x(t) + f(t); \quad \text{and} \quad \dot{y}(t) = A(t)y(t) + f(t); \quad x(t_0) = x_0, y(t_0) = y_0.$$

Now, $\dot{x}(t) = \dot{y}(t)$, since they both satisfy the same differential equation. Therefore $A(t)x(t) + f(t) = A(t)y(t) + f(t)$ and hence $x(t) = y(t)$. Due to the uniqueness of solutions to the physical system, the state transition matrix must be nonsingular. $\square$

Chapter 3

Optimal Control for Linear Systems and Integral Cost Functions

3.1 The Set of Attainability

Consider the linear dynamic system below

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$

where the coefficient $A(t), B(t)$ are known matrices such that

- $A(t)$ and $B(t)$ are $n \times n$ and $n \times m$ matrices respectively each of which is real and piecewise continuous.
- $A(t), B(t)$ are integrable on each compact subinterval of the time t.
- A controller $u(t)$ is a real bounded piecewise continuous m-vector on some interval $t_0 \leq t \leq t_1$ with values in a non-empty restraint set $U \subset \mathbb{R}^m$ (compact and convex)

and we choose the controller $u(t)$ to steer the response $x(t)$ from an initial state x_0 to some desired target state in $\mathbb{R}^n$.

Definition 3.1.0.6. Consider the control system $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ with restraint set U, initial state x_0, and all controllers $u(t) \in U$ on $t_0 \leq t \leq t_1$. Let $x(t)$ denote the corresponding response initiating at $x(t_0) = x_0$. The **Set of Attainability**, $K(t_1)$, is the set of all endpoints $x(t_1) \in \mathbb{R}^n$.

Theorem 3.1.0.7 (Le)**.** *Consider the linear control process in $\mathbb{R}^m$ $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ with compact convex restraint set U, initial state x_0 and controllers $u(t)$ on $t_0 \leq t \leq t_1$. Then the set of attainability is compact and convex.*

Proof. Let U be compact and consider the sequence, $\{u_r(t)\}_{r=1}^{\infty}$, of controllers in U.

Then, by Theorem 2.0.0.5, there is a corresponding sequence, $\{x_r(t)\}_{r=1}^{\infty}$ of responses such that

$$x_r(t) = \phi(t,t_0)x_0 + \int_{t_0}^{t} \phi(t,s)B(s)u_r(s)\mathrm{d}s$$

with the appropriate fundamental matrix, $\phi(t,t_0)$. Now the set of all controllers is weakly compact, by the assumption that U is compact, hence there is a subsequence, $\{u_{r_i}(t)\}_{r=1}^{\infty}$, which converges to a controller $\bar{u}(t) \in U$ on $t_0 \leq t \leq t_1$. i.e.

$$\lim_{i\to\infty} \int_{t_0}^{t} \phi(t,s)B(s)u_{r_i}(s)\mathrm{d}s = \int_{t_0}^{t} \phi(t,s)B(s)\bar{u}(s)\mathrm{d}s.$$

Let $\bar{x}(t)$ be the corresponding response to $\bar{u}(t)$ on $t_0 \leq t \leq t_1$. Then

$$\begin{aligned}
\bar{x}(t) &= \phi(t,t_0)x_0 + \int_{t_0}^{t} \phi(t,s)B(s)\bar{u}(s)\mathrm{d}s \\
&= \phi(t,t_0)x_0 + \lim_{i\to\infty} \int_{t_0}^{t} \phi(t,s)B(s)u_{r_i}(s)\mathrm{d}s \\
&= \lim_{i\to\infty} \left[\phi(t,t_0)x_0 + \int_{t_0}^{t} \phi(t,s)B(s)u_{r_i}(s)\mathrm{d}s\right] \\
&= \lim_{i\to\infty} x_{r_i}(t).
\end{aligned}$$

Therefore,

$$\lim_{i\to\infty} x_{r_i}(t_1) = \bar{x}(t_1) \in K(t_1).$$

Hence, every sequence,$\{x_r(t_1)\}_{r=1}^{\infty} \in K(t_1)$ has a subsequence $\{x_{r_i}(t_1)\}_{i=1}^{\infty}$ which also converges in $K(t_1)$ and thus $K(t_1)$ is compact.
Now suppose that U is convex. Then for any $u_0(t)$ and $u_1(t)$ in U, the line

$$u_\lambda(t) = (1-\lambda)u_0(t) + \lambda u_1(t)$$

also lies in U for $0 \leq \lambda \leq 1$. Then the corresponding response, $x_\lambda(t)$, is given as

$$\begin{aligned}
x_\lambda(t) &= \phi(t,t_0)x_0 + \int_{t_0}^{t} \phi(t,s)B(s)u_\lambda(s)\mathrm{d}s \\
&= \phi(t,t_0)x_0 + \int_{t_0}^{t} \phi(t,s)B(s)[(1-\lambda)u_0(t) + \lambda u_1(t)]\mathrm{d}s \\
&= \phi(t,t_0)x_0 + \int_{t_0}^{t} \phi(t,s)B(s)[(1-\lambda)u_0(t) + \lambda u_1(t)]\mathrm{d}s \\
&\quad + \lambda\phi(t,t_0)x_0 - \lambda\phi(t,t_0)x_0 \\
&= (1-\lambda)x_0(t) + \lambda x_1(t).
\end{aligned}$$

$\therefore x_\lambda(t_1) = (1-\lambda)x_0(t_1) + \lambda x_1(t_1) \in K(t_1)$ and thus $K(t_1)$ is convex. $\square$

3.2 Controllability and Observability

In order to be able to do whatever we want with a given dynamic system under control input, the system must be controllable. In order to see what is going on inside the system under observation, the system should be observable. Controllability is concerned with whether one can design control input to steer the state to arbitrary values. Observability is concerned with whether, without knowing the initial state, one can determine the state of a system given the input and the output. In this section, we develop some of the conditions for the observability and the controllability of linear dynamic systems.

3.2.1 The Controllability Gramian

Definition 3.2.1.1. If G is an $n \times m$ matrix whose elements are continuous functions of time defined on $[t_0, t_1]$, then the mapping $L : \mathcal{P}(\mathbb{R}^m) \longrightarrow \mathbb{R}^n$ defined by $L(u) = \int_{t_0}^{t_1} G(t)u(t)\mathrm{d}t$ is linear. We say that x_1 belongs to the **Range Space** of this mapping if there exists $u(t) \in \mathcal{P}(\mathbb{R}^m)$ such that $x_1 = \int_{t_0}^{t_1} G(t)u(t)\mathrm{d}t$.

Let the dynamic system be given by

$$\dot{z}(t) = B(t)u(t), \quad z(t_0) = z_0 \quad \text{for all} \quad u(t) \in \mathbb{R}^m, z(t) \in \mathbb{R}^n.$$

This then gives

$$z(t_1) = z_0 + \int_{t_0}^{t_1} B(t)u(t)\mathrm{d}t$$

where we have assumed that $B(t)$ and z_0 are known and the aim then is to find $u(t)$ such that at $t = t_1 > t_0$, $z(t)$ takes a certain value; $z(t_1) = z_1$ such that $z_1 - z_0$ will lie in the range space of the linear mapping

$$L(u) = \int_{t_0}^{t_1} B(t)u(t)\mathrm{d}t$$

so that the transfer from z_0 to z_1 (i.e. the desired controllability of the dynamic system) is possible.

Lemma 3.2.1.2 (Ro). *An n-tuple x_1 lies in the range space of $L(u) = \int_{t_0}^{t_1} G(t)u(t)\mathrm{d}t$ if and only if it lies in the range space of the matrix $W(t_0, t_1) = \int_{t_0}^{t_1} G(t)G^T(t)\mathrm{d}t$.*

Proof. Let x_1 be in the range space of $W(t_0, t_1)$, then by definition, there exists η_1 such that

$$W(t_0, t_1)\eta_1 = \int_{t_0}^{t_1} G(t)G^T(t)\eta_1(t)\mathrm{d}t = x_1$$

(the dimension of η_1 is chosen to be $n \times 1$). Let $u(t) = G^T(t)\eta_1(t)$. Then

$$W(t_0, t_1)\eta_1 = \int_{t_0}^{t_1} G(t)G^T(t)\eta_1(t)\mathrm{d}t = x_1 = \int_{t_0}^{t_1} G(t)u(t)\mathrm{d}t = L(u).$$

Hence x_1 is in the range space of L. Conversely, let x_1 be in the range space of L i.e.

$$L(u) = \int_{t_0}^{t_1} G(t)u(t)\mathrm{d}t = x_1. \tag{3.1}$$

We proceed by contradiction. Suppose that x_1 is not in the range space of $W(t_0, t_1)$. Then, its null space will not be empty. Hence there exists x_2 such that

$$W(t_0, t_1)x_2 = 0 \quad \text{and} \quad x_2^T x_1 \neq 0.$$

Now, from (3.1),

$$x_2^T \int_{t_0}^{t_1} G(t)u(t)\mathrm{d}t = x_2^T x_1 \neq 0. \tag{3.2}$$

But $x_2^T W(t_0, t_1)x_2 = 0$ and since $G(t)$ is assumed continuous, $x_2^T W(t_0, t_1)x_2 = 0$ implies that $G^T(t)x_2 = 0$ and thus (3.2) must be zero. Hence a contradiction. Therefore, x_1 must be in the range space $W(t_0, t_1)$. □

Corollary 3.2.1.3 (Ro). *There exists a control u which transfers the state of the system $\dot{z}(t) = B(t)u(t)$ from z_0 at $t = t_0$ to z_1 at $t = t_1$ if and only if $z_1 - z_0$ lies in the range space of*

$$W(t_0, t_1) = \int_{t_0}^{t_1} B(t)B^T(t)\mathrm{d}t.$$

If the transfer is possible, then one particular control which actually drives the state to z_1 at $t = t_1$ is

$$u(t) = B^T(t)\eta$$

where η is the solution of

$$\left(\int_{t_0}^{t_1} B(t)B^T(t)\mathrm{d}t\right) \cdot \eta = W(t_0, t_1) \cdot \eta = z_1 - z_0.$$

Proof. Suppose there exists a u which transfers the state system $\dot{z}(t) = B(t)u(t)$. Then,

$$z_1 - z_0 = \int_{t_0}^{t_1} B(t)u(t)\mathrm{d}t.$$

Hence by the above lemma, $z_1 - z_0$ lies in the range space of $\int_{t_0}^{t_1} B(t)B^T(t)\mathrm{d}t$. Conversely, suppose $z_1 - z_0$ lies in the range space of $\int_{t_0}^{t_1} B(t)B^T(t)\mathrm{d}t$. Then there exists η such that

$$\int_{t_0}^{t_1} B(t)B^T(t)\eta \mathrm{d}t = z_1 - z_0.$$

Choose $u(t) = B^T(t)\eta$. Then,

$$\int_{t_0}^{t_1} B(t)u(t)\mathrm{d}t = z_1 - z_0.$$

Hence,

$$\int_{t_0}^{t_1} B(t)B^T(t)\mathrm{d}t + z(t_0) = z(t)$$

and thus $\dot{z}(t) = B(t)u(t)$. □

We now extend these controllability results to dynamics of the form $\dot{x} = A(t)x(t) + B(t)u(t)$ by the following theorem.

Theorem 3.2.1.4 (Ro). *There exists a $u(t)$ which drives the state of the system*

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$

from the value x_0 at $t = t_0$ to the value x_1 at $t = t_1$ if and only if $x_0 - \Phi(t_0, t_1)x_1$ belongs to the range space of

$$W(t_0, t_1) = \int_{t_0}^{t_1} \Phi(t_0, t)B(t)B^T(t)\Phi^T(t_0, t)\mathrm{d}t. \tag{3.3}$$

Moreover, if η_0 is any solution of $W(t_0, t_1)\eta_0 = x_0 - \Phi(t_0, t_1)x_1$ then $u(t)$ given by

$$u(t) = -B^T(t)\Phi^T(t_0, t)\eta_0$$

is one control which accomplishes the desired transfer.

Proof. Suppose there is u which drives the system and let $z(t) = \Phi(t_0, t)x(t)$ and $z(t) = \Phi(t_0, t)^{-1}x(t)$. Then, as before, by the proof of Theorem 2.0.0.5, $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ will be transformed into

$$\dot{z}(t) = \Phi(t, t_0)B(t)u(t); \quad x(t) = \Phi(t, t_0)z(t).$$

which implies that

$$z(t_1) = z(t_0) + \int_{t_0}^{t_1} \Phi(t, t_0)B(t)u(t)\mathrm{d}t.$$

Thus $z(t_1) - z(t_0)$ is in the range space of

$$L(u) = \int_{t_0}^{t_1} \Phi(t, t_0) B(t) u(t) \mathrm{d}t.$$

Hence, by the lemma, $z(t) - z(t_0)$ will be in the range space of

$$W(t_0, t_1) = \int_{t_0}^{t_1} \Phi(t_0, t) B(t) B^T(t) \Phi^T(t_0, t) \mathrm{d}t.$$

That is

$$\Phi(t_0, t_1)x_1 - x_0 = z(t) - z(t_0).$$

belongs to the range space of (3.3). The transfer, then, is possible if and only if $[x_0 - \Phi(t_0, t_1)x_1]$ lies in the range space of $W(t_0, t_1)$ by Corollary 3.2.1.3. Also, from Corollary 3.2.1.3, we see that a particular control which accomplishes this transfer is defined by

$$u(t) = -B^T(t)\Phi^T(t_0, t)\eta_0$$

where η satisfies

$$W(t_0, t_1)\eta_0 = x_0 - \Phi(t_0, t_1)x_1.$$

□

Definition 3.2.1.5 (Controllability Gramian). We define the *controllability Gramian* of the dynamic system $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ to be the matrix

$$W(t_0, t_1) = \int_{t_0}^{t_1} \Phi(t_0, t) B(t) B^T(t) \Phi^T(t_0, t) \mathrm{d}t.$$

Theorem 3.2.1.6 (Properties of the Controllability Gramian (Ro)**).** *The controllability gramian possesses the following properties:*

1. *It is symmetric.*
2. *It is nonnegative definite for $t_1 \geq t_0$.*
3. *It satisfies the linear matrix differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} W(t, t_1) = A(t)W(t, t_1) + W(t, t_1)A^T(t) - B(t)B^T(t); \quad W(t_1, t_1) = 0.$$

Proof. 1. By definition,

$$\begin{aligned} W^T(t_0, t_1) &= \left[\int_{t_0}^{t_1} \Phi(t_0, t) B(t) B^T(t) \Phi^T(t_0, t) \mathrm{d}t\right]^T \\ &= \int_{t_0}^{t_1} \left[\Phi(t_0, t) B(t) B^T(t) \Phi^T(t_0, t)\right]^T \mathrm{d}t \\ &= W(t_0, t_1). \end{aligned}$$

2. For any constant vector η

$$\begin{aligned}\eta^T W(t_0,t_1)\eta &= \int_{t_0}^{t_1} \eta^T \Phi(t_0,t)B(t)B^T(t)\Phi^T(t_0,t)\eta \mathrm{d}t \\ &= \int_{t_0}^{t_1} \left\| B^T(t)\Phi^T(t_0,t)\eta \right\|^2 \mathrm{d}t \\ &\geq 0.\end{aligned}$$

3. Let

$$W(t,t_1) = \int_t^{t_1} \Phi(t,\sigma)B(\sigma)B^T(\sigma)\Phi^T(t,\sigma)\mathrm{d}\sigma.$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t}W(t,t_1) = \frac{\mathrm{d}}{\mathrm{d}t}\int_t^{t_1} \Phi(t,\sigma)B(\sigma)B^T(\sigma)\Phi^T(t,\sigma)\mathrm{d}\sigma.$$

Hence by Liebniz's rule (see appendix),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}W(t,t_1) &= \Phi(t,t_1)B(t_1)B^T(t_1)\Phi^T(t,t_1)\frac{\mathrm{d}t_1}{\mathrm{d}t} \\ &\quad -\Phi(t,t)B(t)B^T(t)\Phi^T(t,t)\frac{\mathrm{d}t}{\mathrm{d}t} \\ &\quad +\int_t^{t_1} \frac{\mathrm{d}}{\mathrm{d}t}\left[\Phi(t,\sigma)B(\sigma)B^T(\sigma)\Phi^T(t,\sigma)\mathrm{d}\sigma\right] \\ &= -B(t)B^T(t) + \int_t^{t_1} \frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,\sigma)[B(\sigma)B^T(\sigma)\Phi^T(t,\sigma)]\mathrm{d}\sigma \\ &\quad +\int_t^{t_1} \Phi(t,\sigma)B(\sigma)B^T(\sigma)\frac{\mathrm{d}}{\mathrm{d}t}\Phi^T(t,\sigma)\mathrm{d}\sigma \\ &\quad [\text{since } t_1 \text{ is a constant}] \\ &= -B(t)B^T(t) + \int_t^{t_1} A(t)\Phi(t,\sigma)[B(\sigma)B^T(\sigma)\Phi^T(t,\sigma)]\mathrm{d}\sigma \\ &\quad +\int_t^{t_1} [\Phi(t,\sigma)B(\sigma)B^T(\sigma)]\Phi^T(t,\sigma)A^T(t)\mathrm{d}\sigma \\ &= -B(t)B^T(t) + A(t)W(t,t_1) + W(t,t_1)A^T(t).\end{aligned}$$

□

Theorem 3.2.1.7 (Le)**.** *Let A, B be constant matrices. Then the autonomous linear process in $\mathbb{R}^n$;*

$$\dot{x}(t) = Ax(t) + Bu(t); \quad x_0$$

is controllable if and only if the matrix

$$[B, AB, A^2B, \cdots, A^{n-1}B]$$

has rank n.

Proof. Let the linear system be controllable. Then x_0 can be steered to any arbitrary point $x_1 \in \mathbb{R}^n$. Suppose that rank$[B, AB, A^2B, \cdots, A^{n-1}B] < n$. Then there exists a non-zero constant vector, v, such that

$$v[B, AB, A^2B, \cdots, A^{n-1}B] = 0$$

since the rows of the matrix will be linearly dependent. Therefore

$$vB = vAB = \cdots = vA^{n-1}B = 0.$$

Now, since every matrix satisfies its characteristic equation,*Cayley-Hamilton Theorem* (see appendix) we have

$$a_1A^n + a_2A^{n-1} + \cdots + a_{n-1}A + a_nI = 0.$$

Which implies that

$$A^n = c_1A^{n-1} + \cdots + c_{n-1}A + c_nI$$

where $c_i = -\frac{a_i}{a_1}$ for $i = 2, 3, ..., n$ are constants for all i. Hence,

$$\begin{aligned} vA^nB &= v[c_1A^{n-1} + \cdots + c_{n-1}A + c_nI]B \\ &= c_1vA^{n-1}B + \cdots + c_{n-1}vAB + c_nvIB \\ &= 0 \qquad \text{since } vB = vAB = \cdots = vA^{n-1}B = 0. \end{aligned}$$

Thus, we can establish that $vA^{n+k}B = 0$ for $k = 0, 1, 2, 3, ...$ Now,

$$\begin{aligned} v\mathrm{e}^{At}B &= v\left[I + tA + \frac{t^2}{2!}A^2 + \cdots\right]B \\ &= vIB + tvAB + \frac{t^2}{2!}vA^2B + \cdots \\ &= 0. \end{aligned}$$

The autonomous linear system with initial condition $x_0 = 0$ has a solution

$$x(t) = \int_0^t \mathrm{e}^{A(t-s)}Bu(s)\mathrm{d}s \quad \text{and thus} \quad vx(t) = \int_0^t v\mathrm{e}^{A(t-s)}Bu(s)\mathrm{d}s = 0$$

since $v\mathrm{e}^{At}B$ is zero for all $u(t) \in U$. (see remark in appendix). Thus v and $x(t)$ are perpendicular. This then contradicts the controllability assumption of the linear system since every response is confined to a hyperplane that is orthogonal to v (we cannot steer the initial point to any desired state due to this restriction). Hence the rank must be n. Conversely, let rank$[B, AB, A^2B, \cdots, A^{n-1}B] = n$. Let K_0' be the attainable set for which the origin can be steered on the time interval $0 \leq t \leq 1$. Let $|u_i| \leq 1$ for

$i = 1, 2, \cdots, m$. Suppose dimension of $K_0' < n$, then by definition of hyperplane, see appendix, there exists a unit vector, v, such that

$$vx(1) = \int_0^1 v\mathrm{e}^{A(1-s)}Bu(s)\mathrm{d}s = 0.$$

This implies that $f(s) = v\mathrm{e}^{A(1-s)}B = 0$ for all $u \in U$. This function and all its derivatives evaluated at $s = 1$ will vanish and thus $vB = vAB = \cdots = vA^{n-1}B = 0$ which means that $\mathrm{rank}[B, AB, A^2B, \cdots, A^{n-1}B] < n$. Therefore the dimension of K_0' must be n. By Theorem 3.1.0.7, K_0' is compact and convex in $\mathbb{R}^n$. Since this set contains an open set and it is convex, it must contain the origin in its interior. It is also well to note that this set is symmetric since $u(t)$ can be replaced by $-u(t)$. Let the restriction on the controls be relaxed such that $|u_i(t)| \leq l$ for $l = 1, 2, 3, ...$ Then, the set of attainability of the origin, K_0, for all controllers $u(t)$ on $0 \leq t \leq 1$ must be the entire space $\mathbb{R}^n$. For an arbitrary point x_0, the set K of attainability for controllers, $u(t)$, on $0 \leq t \leq 1$ is $K = \mathrm{e}^A x_0 + K_0 = \mathbb{R}^n$. Therefore the system is controllable. □

3.2.2 The Observability Gramian

Consider the dynamic system $\dot{x}(t) = A(t)x(t) + B(t)u(t)$; $\quad y(t) = C(t)x(t)$ where $y(t)$ is the output system and $A(t), B(t), C(t), u(t)$ are known. Then for some initial condition, x_0, the output can be expressed as $y(t) = C(t)\left(\Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, t_0)B(t)u(t)\mathrm{d}t\right)$. Thus if we want to determine x_0, then we can consider the homogenuous system, $\dot{x}(t) = A(t)x(t)$, such that $y(t) = C(t)\Phi(t, t_0)x_0$. Hence we need a linear transformation that will associate a unique state with any output. Therefore define the linear transformation $L : \mathbb{R}^n \longrightarrow \mathcal{D}(\mathbb{R}^n)$ for all $t \in [t_0, t_1]$ by, for H continuous with respect to t,

$$L(x)(t) = C(t)\Phi(t, t_0)x(t) = H(t)x(t).$$

So the *null space* of L is the set of all $x(t) \in \mathbb{R}^n$ such that $H(t)x(t)$ is identically zero $\forall t$. **Lemma 3.2.2.1** (Ro). *Let $H(t)$ be an $m \times n$ matrix whose elements are continuous on $[t_0, t_1]$. Then, the null space of L coincides with the null space of $M(t_0, t_1) = \int_{t_0}^{t_1} H^T(t)H(t)\mathrm{d}t$.*

Proof. Let x_0 be in the null space of $M(t_0, t_1)$. Then $M(t_0, t_1)x_0 = 0$ which implies that

$$x_0^T M(t_0, t_1)x_0 = \int_{t_0}^{t_1} x_0^T H^T(t)H(t)x_0\mathrm{d}t = \int_{t_0}^{t_1} \|H(t)x_0\|^2 \,\mathrm{d}t = 0.$$

The continuity of $H(t)$ ensures that $H(t)x_0 = 0$ for all t. Hence x_0 will be in the null space of L. Conversly, suppose x_0 is in the null space of L. So $H^T(t)H(t)x_0 = 0$. Hence

$\int_{t_0}^{t_1} H^T(t)H(t)x_0 \mathrm{d}t = M(t_0,t_1)x_0 = 0$. Thus x_0 is in the null space of $M(t_0,t_1)$. □

The following theorem tells us how the starting state variable, x_0, can be found given the output, y.

Theorem 3.2.2.2 (Ro). *Suppose $A(t), C(t), y(t)$ are given on the interval $[t_0,t_1]$ such that*

$$\dot{x}(t) = A(t)x(t); \quad y(t) = C(t)x(t).$$

Then it it possible to determine $x(t_0)$ which lies in the null space of $M(t_0,t_1)$ where

$$M(t_0,t_1) = \int_{t_0}^{t_1} \Phi^T(t,t_0)C^T(t)C(t)\Phi(t,t_0)\mathrm{d}t.$$

In particular, it is possible to determine $x(t_0)$ uniquely if $M(t_0,t_1)$ is invertible. Moreover, it is not possible to distinguish with the knowledge of $y(t)$, the starting state x_1 from the starting state x_2, if $x_1 - x_2$ lies in the null space of $M(t_0,t_1)$.

Proof. From

$$\dot{x}(t) = A(t)x(t); \quad y(t) = C(t)x(t)$$

we have that $x(t) = \Phi(t,t_0)x_0$ by Theorem 2.0.0.5. Hence,

$$y(t) = C(t)x(t) = C(t)\Phi(t,t_0)x_0.$$

Premultiplying both sides by $\Phi^T(t,t_0)C^T(t)$ and integrating over the interval $[t_0,t_1]$ gives

$$\begin{aligned} \int_{t_0}^{t_1} \Phi^T(t,t_0)C^T(t)y(t)\mathrm{d}t &= \int_{t_0}^{t_1} \Phi^T(t,t_0)C^T(t)C(t)\Phi(t,t_0)x_0\mathrm{d}t \\ &= M(t_0,t_1)x_0. \end{aligned} \tag{3.4}$$

Thus $x_0 = x(t_0)$ can be determined and lies in the null space of $M(t_0,t_1)$. From (3.4), if $M(t_0,t_1)$ is invertible, then

$$x_0 = M^{-1}(t_0,t_1)\int_{t_0}^{t_1} \Phi^T(t,t_0)C^T(t)y(t)\mathrm{d}t.$$

Further, let $x_1 - x_2$ be in the null space of $M(t_0,t_1)$, i.e. $M(t_0,t_1)(x_1 - x_2) = 0$, and let $y_1(t)$ and $y_2(t)$ be such that

$$y_1(t) = C(t)\Phi(t,t_0)x_1 \quad \text{and} \quad y_2(t) = C(t)\Phi(t,t_0)x_2.$$

Then $y_1(t) - y_2(t) = C(t)\Phi(t,t_0)(x_1 - x_2)$ which implies that

$$\|y_1(t) - y_2(t)\|^2 = \|C(t)\Phi(t,t_0)(x_1 - x_2)\|^2.$$

Consequently,

$$\begin{aligned}
\int_{t_0}^{t_1} \|y_1(t) - y_2(t)\|^2 \, \mathrm{d}t &= \int_{t_0}^{t_1} \|C(t)\Phi(t,t_0)(x_1 - x_2)\|^2 \, \mathrm{d}t \\
&= \int_{t_0}^{t_1} (x_1 - x_2)^T \Phi^T(t,t_0) C^T(t) C(t) \Phi(t,t_0)(x_1 - x_2)\mathrm{d}t \\
&= (x_1 - x_2)^T \int_{t_0}^{t_1} \Phi^T(t,t_0) C^T(t) C(t) \Phi(t,t_0)(x_1 - x_2)\mathrm{d}t \\
&= (x_1 - x_2)^T M(t_0,t_1)(x_1 - x_2) \\
&= (x_1 - x_2)^T.0 \\
&= 0 \qquad (3.5)
\end{aligned}$$

$$\text{thus} \quad \|y_1(t) - y_2(t)\|^2 = 0 \Longleftrightarrow y_1(t) - y_2(t) = 0$$

which implies that $y_1(t) = y_2(t)$. Thus x_1, x_2 give the same output and hence they cannot be distinguished with the knowledge of $y(t)$. □

Thus the system is observable if the inverse $M^{-1}(t_0,t_1)$ exists since with this, x_0 can be determined.

Definition 3.2.2.3 (Observability Gramian). We define the *Observability Gramian* of the dynamic system $x(t) = A(t)x(t) + B(t)u(t); \quad y(t) = C(t)x(t)$ to be the matrix

$$M(t_0,t_1) = \int_{t_0}^{t_1} \Phi^T(t,t_0) C^T(t) C(t) \Phi(t,t_0)\mathrm{d}t.$$

Theorem 3.2.2.4 (Properties of the Observability Gramian (Ro)**).** *The observability gramian possesses the following properties:*

1. *It is symmetric.* 2. *It is nonnegative definite for $t_1 \geq t_0$.*
3. *It satisfies the linear matrix differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} M(t,t_1) = -A(t)M(t,t_1) - M(t,t_1)A^T(t) - C(t)C^T(t); \quad M(t_1,t_1) = 0.$$

Proof. The proof is similar to the proof of Theorem 3.2.1.6 □

3.3 Integral Cost Functions

Most often the choice of a performance criterion is subjective. However the choice, even though subjective, should take into consideration the goal of the optimalization hence, the need to study some specific cost functions, the integral cost functions.

3.3.1 Significance of Integral Cost Functions

The selection of the cost functional, the specified integral criterion, is not easily determined. The best policy often indicates the selection of a cost functional that leads to an easily constructed optimal controller and seems reasonable to approximate the ideal goal. It appears then that the motivation for studying certain specific optimal control systems is that such investigations can be formulated as a definite mathematical problem that leads to useful techniques and methods for control synthesis. For instance, an integral cost criterion is used whenever the average performance over the chosen time duration is of primary importance. A frequently used criterion in the design of controllers is the integral of the error squared, $\int_{t_0}^{t_1}(u(t)-u^*(t))^2 \mathrm{d}t$, where $u(t)-u^*(t)$ is the deviation from the optimal controller.

3.3.2 Integral Quadratic Cost Functions

Consider the linear control process

$$\dot{x}(t) = A(t)x(t) + B(t)u(t); \quad x(t_0) = x_0$$

where A, B are matrices of size $n \times n$ and $n \times m$ respectively on a given time interval $t_0 \leq t \leq t_f$ such that $x(t)$ is steered by $u(t)$ from x_0 to the end point $x(t_f)$. Let the cost functional be given by

$$J(u) = g(x(t_f)) + \int_0^{t_f} \left(x^T(t)Q(t)x(t) + u^T(t)R(t)u(t)\right) \mathrm{d}t$$

where we assume that $g(x)$ is a given real continuous function of $\mathbb{R}^n$ and Q and R are square matrices which are continuous and symmetric on $t_0 \leq t \leq t_f$ such that Q and R are positive semidefinite and positive definite respectively. The control problem will be to minimize the cost function among all measurable controllers such that, in order for $J(u)$ to be finite,

$$\int_0^{t_f} u^T(t)R(t)u(t)\mathrm{d}t < \infty$$

and we see that this is possible if and only if $u(t)$ lies in the Hilbert space $L_2(t_0 = 0, t_f)$, see appendix A.6, since we have assumed that $R > 0$, continuous and bounded.The existence of such controllers can be achieved by exploring the geometric properties of the set of attainability. For instance, it is established that controllers are extremals satisfying the maximal principle and corresponding to the boundary of the set of attainability. Also because of the nonnegative nature of x^TQx and u^TRu, we can expect that the minimum for the cost can be realized if $g(x(t_f))$ is bounded below or is a convex function [Le].

Chapter 4

The Second Variation

The linear quadratic control problem is a control problem where the dynamic equation and the terminal constraints are linear and the performance criterion is a quadratic function of the state and control variable. In this chapter, we derive the second variation of a general control problem and establish some basic results which serves as a preparatory ground for the next chapter where we formulate the linear quadratic problem.

4.1 The Variation of a Functional

Definition 4.1.0.1 (Functional)**.** By a functional, we mean a correspondence which assigns a definite(real) number to each function(or curve) belonging to some class. For instance, let $F(x, y, z)$ be a continuous function of three variables. Then the expression

$$J[y] = \int_a^b F(t, y(t), y'(t))\mathrm{d}t$$

where $y(t)$ ranges over the set of all continuously differentiable functions defined on the interval $[a, b]$ defines a functional.

Definition 4.1.0.2 (Continuous Linear Functional)**.** Given a normed linear space, $\mathcal{R}$, let $\phi(h)$ be a functional on $\mathcal{R}$. Then $\phi(h)$ is said to be a continuous linear functional if;

1. $\phi(\alpha h) = \alpha\phi(h), \forall h \in \mathcal{R}$ and $\alpha \in \mathbb{R}$,
2. $\phi(h_1 + h_2) = \phi(h_1) + \phi(h_2), \forall h_1, h_2 \in \mathcal{R}$ and
3. $\phi(h)$ is continuous for all $h \in \mathcal{R}$.

A functional $B[x, y]$ depending on two elements x and y belonging to some normed linear space, $\mathcal{R}$, is said to be bilinear if it is a linear functional of y for any fixed x and a linear functional of x for any fixed y i.e.

$$B[x+y,z]=B[x,z]+B[y,z]$$
$$B[\alpha x,y]=\alpha B[x,y]$$

and

$$B[x,y+z]=B[x,y]+B[x,z]$$
$$B[x,\alpha y]=\alpha B[x,y]$$

for $x,y,z\in\mathcal{R}$ and $\alpha\in\mathbb{R}$. If we set $y=x$ in a bilinear functional, we obtain an expression called a *quadratic functional.* For example, the expression

$$B[x,y]=\int_a^b \alpha(t)x(t)y(t)\mathrm{d}t \quad t\in[a,b]$$

is bilinear and its corresponding quadratic functional is

$$A[x]=\int_a^b \alpha(t)x^2(t)\mathrm{d}t \quad t\in[a,b]$$

which is positive definite if $\alpha(t)>0$.

4.1.1 The Variation (or Differential)

We now introduce the concept of variation of a functional, analogous to the concept of the differential of a function of n variables. Let $J[y]$ be a functional defined on some normed linear space and let

$$\triangle J_y[h]=J[h+y]-J[y]$$

be its increment corresponding to the increment $h=h(t)$ of the independent variable $y=y(t)$. If y is fixed, $\triangle J_y[h]$ is a functional of h. Suppose that

$$\triangle J_y[h]=\phi(h)+\epsilon\|h\|$$

where $\phi(h)$ is a linear functional and $\epsilon\to 0$ as $\|h\|\to 0$. Then the functional $J[y]$ is said to be differentiable, see appendix, and the principal linear part of the increment $\triangle J_y[h]$, i.e, the linear functional $\phi(h)$ is called the *Variation (or differential)* of $J[h]$ and is denoted by $\delta J[h]$.

Theorem 4.1.1.1 (GS). *The differential (or variation) of a differentiable functional is unique.*

Proof. We note that if $\phi(h)$ is a linear functional and if

$$\frac{\phi(h)}{\|h\|} \to 0 \quad \text{as} \quad \|h\| \to 0 \quad \text{then} \quad \phi(h) = 0 \quad \forall h.$$

Suppose that $\phi(h_0) \neq 0$ for some $h_0 \neq 0$. Then setting

$$h_n = \frac{h_0}{n}, \qquad \lambda = \frac{\phi(h_0)}{\|h_0\|}.$$

Then, $\|h_n\| \to 0$ as $n \to \infty$ but

$$\lim_{n\to\infty} \frac{\phi(h_n)}{\|h_n\|} = \lim_{n\to\infty} \frac{n\phi(h_0)}{n\|h_0\|} = \lambda \neq 0;$$

contrary to the hypothesis. Now suppose the differential of the functional $J[y]$ is not uniquely defined so that

$$\triangle J_y[h] = \phi_1(h) + \epsilon_1\|h\| \quad \text{and} \quad \triangle J_y[h] = \phi_2(h) + \epsilon_2\|h\|$$

where $\phi_1(h)$ and $\phi_2(h)$ are linear functionals and $\epsilon_1, \epsilon_2 \to 0$ as $\|h\| \to 0$. Therefore

$$\phi_1(h) - \phi_2(h) = (\epsilon_1 - \epsilon_2)\|h\|.$$

But since $\phi_1(h) - \phi_2(h)$ is a linear functional such that

$$\frac{\phi_1(h) - \phi_2(h)}{\|h\|} \to 0 \quad \text{as} \quad \|h\| \to 0,$$

it follows that it vanishes identically and thus $\phi_1(h) = \phi_2(h)$ and $\epsilon_1\|h\| = \epsilon_2\|h\|$ hence the differentials are identical. □

Theorem 4.1.1.2 (GS). *A necessary condition for the differentiable functional, $J[y]$, to have a minimum for $y = \hat{y}$ is that its variation vanishes, i.e $\delta J[h] = 0$, for $y = \hat{y}$ for all h.*

Proof. Suppose $J[y]$ has a minimum for $y = \hat{y}$, thus $\triangle J_y[h]$ is nonnegative. According to the definition of the variation, $\delta J[h]$, we have

$$\triangle J_y[h] = \delta J[h] + \epsilon\|h\|$$

where $\epsilon \to 0$ as $\|h\| \to 0$. Thus for sufficiently small $\|h\|$, the sign of $\triangle J[h]$ will be the same as the sign of $\delta J[h]$. Suppose that $\delta J[h_0] \neq 0$ for some h_0. Then, for any $\alpha > 0$, no matter how small, we have $\delta J[-\alpha h_0] = -\delta J[\alpha h_0]$, by its linearity property. Therefore, $\delta J[h] + \epsilon\|h\|$ can be made to have either sign for arbitrarily small $\|h\|$. But that is impossible since by hypothesis $\triangle J_y[h] \geq 0$ for sufficiently small $\|h\|$. Thus a contradiction and hence $\delta J[h]$ must be zero. $\square$

An optimal control problem can be solved either by the method of Lagrange multipliers or a dynamic programming method. Here, we use the method of Lagrange multipliers. Suppose we have a problem;

$$\text{minimize}\tilde{f}(x) \quad \text{subject to} \quad G(x) = 0.$$

Then, this complication can be handled by the method of Lagrange. This method reduces the constrained problem to a new unconstrained minimization problem with additional variable. The additional variable is known as a *Lagrange Multiplier* λ. That is

$$f(x, \lambda) = \tilde{f}(x) + \lambda G(x).$$

The problem then becomes

$$\text{minimize} f(x, \lambda).$$

In this work, the equality constraint will be a linear dynamic system of the form

$$\dot{x}(t) = A(t)x(t) + B(t)u(t); \quad x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m.$$

Hence

$$A(t)x(t) + B(t)u(t) - \dot{x}(t) = 0$$

and a cost function will be minimized subject to this equality constraint. We note that solving the unconstrained problem with the Lagrange multipliers is equivalent to solving the constrained problem. To see this, consider the problem

$$\text{minimize} F(x, y) \quad \text{subject to} \quad G(x, y) = 0,$$

and suppose that, the constraint equation can be solved so that $y = h(x)$. Then y can be eliminated from $F(x, y)$ which reduces the problem to

$$\text{minimize} F(x, h(x)).$$

Suppose $F(x, y)$ is diifferentiable, then at its extremum, the Euler-Lagrange equation must be satisfied i.e.

$$F_x + F_y \frac{\mathrm{d}y}{\mathrm{d}x} = 0 \quad \text{(Euler-Lagrange Equation)}.$$

Consider moving along the curve $G(x, y) = 0$. Let s be the arc length parameter such that $\frac{\mathrm{d}G}{\mathrm{d}s} = 0$. Then,

$$G_x \frac{\mathrm{d}x}{\mathrm{d}s} + G_y \frac{\mathrm{d}y}{\mathrm{d}s} = 0 \Longrightarrow \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{G_x}{G_y}.$$

Substituting this into the Euler-Lagrange equation gives

$$F_x G_y = F_y G_x. \tag{4.1}$$

Consider the unconstrained problem with the Lagrange multiplier

$$\tilde{F}(x, y, \lambda) = F(x, y) + \lambda G(x, y).$$

Then,

$$\tilde{F}(x, y, \lambda)_x = F_x + \lambda G_x \quad \text{and} \quad \tilde{F}(x, y, \lambda)_y = F_y + \lambda G_y.$$

Since at the extremum the partial derivatives must be zero, we equate these to zero and eliminate λ to give

$$F_x G_y = F_y G_x. \tag{4.2}$$

Thus, the equation that defines the extremum in the unconstrained problem using the Lagrange Multipliers is the same as (4.1) which is obtained by solving for the extrema in the constrained problem.

4.2 Motivation for the Second Variation

Let the state, $x : I \subseteq \mathbb{R} \longrightarrow \mathbb{R}^n$ be propagated by the dynamic system $\dot{x}(t) = f(x(t), u(t), t)$ with initial condition $x(t_0) = x_0$ where $u : I \subseteq \mathbb{R} \longrightarrow \mathbb{R}^m$ is the control which is assumed to be drawn from the set of piecewise continuous m-vector functions of $t \in I = [t_0, t_f]$ and $f : S \subseteq \mathbb{R}^{n+m+1} \longrightarrow \mathbb{R}^n$ is an n-vector function. Suppose also that $\phi : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ and $L : \mathbb{R}^{n+m+1} \longrightarrow \mathbb{R}$ are some scalar functions and $\lambda : \mathbb{R} \longrightarrow \mathbb{R}^n$. Consider the problem

$$\min_{u(t) \in U_T} J \quad \text{subject to} \quad \begin{cases} \dot{x}(t) = f(x(t), u(t), t), \quad x(t_0) = x_0 \\ \Psi(x(t_f), t_f) = 0 \end{cases} \tag{4.3}$$

where

$$J = \phi(x(t_f), t_f) + \int_{t_0}^{t_f} L(x(t), u(t), t)\mathrm{d}t$$

and $\Psi(x(t_f), t_f) = 0$ is a general terminal constraint. $\Psi(\cdot)$ is a p-dimensional vector function i.e. $\Psi : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^p$ and U_T is the set of piecewise continuous m-vector functions of t on the interval $[t_0, t_f]$ such that the terminal constraint is satisfied. We proceed as follows. The trick is to add additional terms to J that sum to zero. We start by appending to J a continuously differential vector $\lambda(t)$ of Lagrange multiplier functions and the final state constraints a p-vector ν of Lagrange multipliers. This will convert the constrained problem to an unconstrained problem and the optimalization problem will not change just as we have seen. Then, the augmented performance index becomes

$$\hat{J} = \phi(x(t_f), t_f) + \nu^T \Psi(x(t_f), t_f) + \int_{t_0}^{t_f} \big(H(x(t), u(t), \lambda(t), t) - \lambda^T(t)\dot{x}(t)\big)\,\mathrm{d}t$$

where

$$H(x(t), u(t), \lambda(t), t) = L(x(t), u(t), t) + \lambda^T(t) f(x(t), u(t), t)$$

is referred to as the Hamiltonian (see appendix). We make the following assumptions.

- The first assumption is that we can generate a locally minimizing path by satisfying the first-order necessary conditions. That is we assume that we have been able to find $\bar{x}(t), \bar{u}(t)$ and $\bar{t}$ that satisfy (4.3) and minimize the performance index subject to the terminal constraints over all functions $u(t)$.

- Secondly we assume that the functions f, L, ϕ, Ψ are all twice differentiable with respect to their arguments.

Now, we define the variations in the state, control, and the multipliers as

$$\triangle x = x - \bar{x}, \quad \triangle u = u - \bar{u}, \quad \triangle \lambda = \lambda - \bar{\lambda}, \quad \triangle \nu = \nu - \bar{\nu} \quad \triangle t_f = t_f - \bar{t}_f.$$

The variation in the performance index is then

$$\begin{aligned} \triangle \hat{J} &= \hat{J}(u(\cdot), t_f; x_0, \lambda(\cdot), \nu) - \hat{J}(\bar{u}(\cdot), \bar{t}_f; x_0, \bar{\lambda}(\cdot), \bar{\nu}) \\ &= \int_{t_0}^{\bar{t}_f} \left[\triangle H - \triangle\left(\lambda^T \dot{x}\right)\right] \mathrm{d}t + \left[\triangle\phi + \triangle\nu^T \Psi\right]\Big|_{t_f = \bar{t}_f}. \end{aligned}$$

Now, using the Taylor expansion for a function of vector arguments, see appendix,

$$\begin{aligned} &\hat{J}(u(\cdot), t_f; x_0, \lambda(\cdot), \nu) \\ &= \hat{J}(\bar{u}(\cdot), \bar{t}_f; x_0, \bar{\lambda}(\cdot), \bar{\nu}) + \phi_x \triangle x + \phi_{t_f} \triangle t_f \\ &\quad + \nu^T\left(\Psi_x \triangle x + \Psi_{t_f} \triangle t_f\right) + \triangle\nu^T \Psi(\bar{x}(\bar{t}_f), \bar{t}_f) \\ &\quad + H\left(\bar{x}(\bar{t}_f), \bar{u}(\bar{t}_f), \bar{\lambda}(\bar{t}_f), \bar{t}_f\right)\triangle t_f \\ &\quad + \int_{t_0}^{\bar{t}_f} \left[H_x \triangle x + H_u \triangle u + H_\lambda \triangle \lambda - \left(\lambda^T \triangle \dot{x} + \dot{x} \triangle \lambda^T\right)\right] \mathrm{d}t + \cdots \end{aligned}$$

Therefore

$$
\begin{aligned}
&\hat{J}(u(\cdot), t_f; x_0, \lambda(\cdot), \nu) \\
&= \hat{J}(\bar{u}(\cdot), \bar{t}_f; x_0, \bar{\lambda}(\cdot), \bar{\nu}) \\
&\quad + \left(\phi_x + \nu^T \Psi_x\right) \triangle x + \Omega \triangle t_f + \triangle \nu^T \Psi(\bar{x}(\bar{t}_f), \bar{t}_f) \\
&\quad + \textstyle\int_{t_0}^{\bar{t}_f} \Big[H_x \triangle x + H_u \triangle u - \lambda^T \triangle \dot{x} + (H_\lambda - \dot{x})^T \triangle \lambda \Big] \,\mathrm{dt} \\
&\quad + \tfrac{1}{2} \textstyle\int_{t_0}^{\bar{t}_f} \big[\triangle x^T H_{xx} \triangle x + \triangle x^T H_{xu} \triangle u + \triangle u^T H_{ux} \triangle x + \triangle u^T H_{uu} \triangle u \\
&\quad + \qquad 2 \left(\triangle x^T H_{x\lambda} \triangle \lambda + \triangle \lambda^T H_{\lambda u} \triangle u - \triangle \lambda^T \triangle \dot{x} \right) \big] \,\mathrm{dt} \\
&\quad + \tfrac{1}{2} \begin{bmatrix} \triangle x^T & \triangle \nu^T & \triangle t_f \end{bmatrix} \begin{bmatrix} \phi_{xx} + \left(\nu^T \Psi\right)_{xx} & \Psi_x^T & \Omega_x^T \\ \Psi_x & 0 & \Omega_\nu^T \\ \Omega_x & \Omega_\nu & \frac{\mathrm{d}\Omega}{\mathrm{dt}} \end{bmatrix} \begin{bmatrix} \triangle x \\ \triangle \nu \\ \triangle t_f \end{bmatrix} \\
&\quad + \text{higher order terms}
\end{aligned}
$$

where $\Omega = \phi_{t_f} + \nu^T \Psi_{t_f} + H\left(\bar{x}(\bar{t}_f), \bar{u}(\bar{t}_f), \bar{\lambda}(\bar{t}_f), \bar{t}_f\right)$ and are all evaluated at $t_f = \bar{t}_f$. Using integration by parts on $-\lambda^T \triangle \dot{x}$ and using the fact that the first variation vanishes for optimality, $\delta \hat{J} = 0$, by Theorem 4.1.1.2, and the fact that $\Psi(\bar{x}(\bar{t}_f), \bar{t}_f) = 0$, all the first order derivatives vanish leaving only the second order partial derivatives i.e.

$$
\begin{aligned}
\triangle \hat{J} \;=\;& \lambda^T(t_0) \triangle x(t_0) \\
&+ \frac{1}{2} \int_{t_0}^{\bar{t}_f} \big[\triangle x^T H_{xx} \triangle x + \triangle x^T H_{xu} \triangle u + \triangle u^T H_{ux} \triangle x + \triangle u^T H_{uu} \triangle u \\
&+ \qquad 2 \left(\triangle x^T H_{x\lambda} \triangle \lambda + \triangle \lambda^T H_{\lambda u} \triangle u - \triangle \lambda^T \triangle \dot{x} \right) \big] \,\mathrm{dt} \\
&+ \frac{1}{2} \begin{bmatrix} \triangle x^T & \triangle \nu^T & \triangle t_f \end{bmatrix} \begin{bmatrix} \phi_{xx} + \left(\nu^T \Psi\right)_{xx} & \Psi_x^T & \Omega_x^T \\ \Psi_x & 0 & \Omega_\nu^T \\ \Omega_x & \Omega_\nu & \frac{\mathrm{d}\Omega}{\mathrm{dt}} \end{bmatrix} \begin{bmatrix} \triangle x \\ \triangle \nu \\ \triangle t_f \end{bmatrix} \\
&+ \text{higher order terms}
\end{aligned}
$$

Claim 1. Let the control $u^0(\cdot)$ generate the trajectory $x^0(\cdot)$. Then linearly perturbing $u^0(\cdot)$ by $\epsilon\eta(\cdot), \epsilon > 0$, perturbs $x^0(\cdot)$ by a function of exactly the same form, viz; $\epsilon z(t; \eta(\cdot))$ where $\eta(\cdot)$ is also piecewise continuous.

Proof. Consider the dynamic system $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ and suppose that the perturbation of u^0 linearly by $\epsilon\eta$ results in the perturbation of the trajectory x^0 by $\xi(\cdot; \epsilon)$. Then

$$
x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, \sigma)B(\sigma)u(\sigma)\mathrm{d}\sigma
$$

and consequently

$$\begin{aligned}\xi(\cdot;\epsilon)+x^0(t) &= \Phi(t,t_0)x_0+\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)[u^0(\cdot)+\epsilon\eta(\cdot)]\mathrm{d}\sigma \\ &= \Phi(t,t_0)x_0+\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)u^0(\cdot)\mathrm{d}\sigma+\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)\epsilon\eta(\cdot)\mathrm{d}\sigma.\end{aligned}$$

Therefore

$$\xi(\cdot;\epsilon)=\epsilon\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)\eta(\cdot)\mathrm{d}\sigma$$

since

$$x^0(t)=\Phi(t,t_0)x_0+\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)u^0(\cdot)\mathrm{d}\sigma.$$

Let

$$z(t;\eta(\cdot))=\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)\eta(\cdot)\mathrm{d}\sigma.$$

Then,

$$\xi(\cdot;\epsilon)=\epsilon z(t;\eta(\cdot))$$

as required. □

Thus we have shown that for any change in the control, there exists a corresponding change in the state which can be expanded i.e.

$$\triangle u(\cdot)=\epsilon\eta(\cdot)\Longleftrightarrow\triangle x(\cdot)=\epsilon z(\cdot)+\mathcal{O}(\cdot,\epsilon).$$

Now, let

$$\triangle\lambda(\cdot)=\epsilon\tilde{\lambda}(\cdot)+\mathcal{O}(\cdot,\epsilon);\quad\triangle\nu(\cdot)=\epsilon\tilde{\nu}(\cdot)+\mathcal{O}(\cdot,\epsilon);\quad\triangle t_f(\cdot)=\epsilon\triangle.$$

Upon substitution of this into $\hat{J}$ and noting that since the initial time is given, $\triangle x(t_0)=z(t_0)=0$, $\triangle\hat{J}$ reduces to

$$\begin{aligned}\triangle\hat{J} &= \frac{\epsilon^2}{2}\int_{t_0}^{\bar{t}_f}\left\{\begin{bmatrix} z^T & \eta^T\end{bmatrix}\begin{bmatrix} H_{xx} & H_{xu}\\ H_{ux} & H_{uu}\end{bmatrix}\begin{bmatrix} z\\ \eta\end{bmatrix}+\cdots\right\}\mathrm{d}t \\ &+\frac{\epsilon^2}{2}\begin{bmatrix} z^T & \tilde{\nu}^T & \triangle\end{bmatrix}\begin{bmatrix}\phi_{xx}+\left(\nu^T\Psi\right)_{xx} & \Psi_x^T & \Omega_x^T\\ \Psi_x & 0 & \Omega_\nu^T\\ \Omega_x & \Omega_\nu & \frac{\mathrm{d}\Omega}{\mathrm{d}t}\end{bmatrix}\begin{bmatrix} z\\ \tilde{\nu}\\ \triangle\end{bmatrix} \\ &+\int_{t_0}^{\bar{t}_f}\mathcal{O}(t,\epsilon^2)\mathrm{d}t+\mathcal{O}(\epsilon^2)\end{aligned}$$

where

$$\lim_{\epsilon\longrightarrow 0}\frac{\mathcal{O}(t,\epsilon^2)}{\epsilon^2}=0\quad\text{and}\quad\lim_{\epsilon\longrightarrow 0}\frac{\mathcal{O}(\epsilon^2)}{\epsilon^2}=0$$

Definition 4.2.0.3 (Second Variation). We define the second variation, $\delta^2\hat{J}$, as

$$\delta^2\hat{J} = \lim_{\epsilon \longrightarrow 0} \frac{\triangle \hat{J}}{\epsilon^2}$$

where it is assumed to be positive definite. That is

$$\begin{aligned}
\delta^2\hat{J} &= \frac{1}{2}\int_{t_0}^{\bar{t}_f}\left\{\begin{bmatrix} z^T & \eta^T \end{bmatrix}\begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix}\begin{bmatrix} z \\ \eta \end{bmatrix} + \cdots\right\}\mathrm{d}t \\
&+\frac{1}{2}\begin{bmatrix} z^T & \tilde{\nu}^T & \triangle \end{bmatrix}\begin{bmatrix} \phi_{xx} + \left(\nu^T\Psi\right)_{xx} & \Psi_x^T & \Omega_x^T \\ \Psi_x & 0 & \Omega_\nu^T \\ \Omega_x & \Omega_\nu & \frac{\mathrm{d}\Omega}{\mathrm{d}t} \end{bmatrix}\begin{bmatrix} z \\ \tilde{\nu} \\ \triangle \end{bmatrix}
\end{aligned}$$

Remark 4.2.0.4. Linear quadratic problems are often fixed time problems. Hence

$$\triangle t_f = \epsilon\triangle = 0.$$

Now, given $\dot{x} = f(x, u, \cdot)$, upon expansion, we have that

$$\begin{aligned}
&\dot{x}(t_0) + \epsilon\dot{z} + \mathcal{O}(\cdot;\epsilon) \\
&= f(x(t_0), u(t_0), \cdot) + f_x(x(t_0), u(t_0), \cdot)\triangle x + f_u(x(t_0), u(t_0), \cdot)\triangle u + \cdots \\
&= f(x(t_0), u(t_0), \cdot) + f_x(x(t_0), u(t_0), \cdot)(\epsilon z + \mathcal{O}(\cdot, \epsilon)) + f_u(x(t_0), u(t_0), \cdot)\epsilon\eta \\
&\quad + \cdots \\
&= f(x(t_0), u(t_0), \cdot) + \epsilon f_x(x(t_0), u(t_0), \cdot)z + \epsilon f_u(x(t_0), u(t_0), \cdot)\eta + \mathcal{O}(\cdot, \epsilon^2).
\end{aligned}$$

Hence

$$\epsilon\dot{z}(\cdot) + \mathcal{O}(\cdot, \epsilon^2) = \epsilon f_x(x(t_0), u(t_0), \cdot)z + \epsilon f_u(x(t_0), u(t_0), \cdot)\eta.$$

Thus we obtain the differential equation

$$\dot{z}(\cdot) = f_x(x(t_0), u(t_0), \cdot)z + f_u(x(t_0), u(t_0), \cdot)\eta; \quad z(t_0) = 0.$$

Consequently, the $(\cdots)$ term, which is

$$z^T H_{x\lambda}\tilde{\lambda} + \tilde{\lambda}^T H_{\lambda u}\eta - \tilde{\lambda}^T\dot{z}$$

in the integral of the second variation vanishes. Thus $\delta^2\hat{J}$ becomes

$$\begin{aligned}
\delta^2\hat{J} &= \frac{1}{2}\int_{t_0}^{\bar{t}_f}\left\{\begin{bmatrix} z^T & \eta^T \end{bmatrix}\begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix}\begin{bmatrix} z \\ \eta \end{bmatrix}\right\}\mathrm{d}t \\
&+\frac{1}{2}\begin{bmatrix} z^T & \tilde{\nu}^T & 0 \end{bmatrix}\begin{bmatrix} \phi_{xx} + \left(\nu^T\Psi\right)_{xx} & \Psi_x^T & \Omega_x^T \\ \Psi_x & 0 & \Omega_\nu^T \\ \Omega_x & \Omega_\nu & \frac{\mathrm{d}\Omega}{\mathrm{d}t} \end{bmatrix}\begin{bmatrix} z \\ \tilde{\nu} \\ 0 \end{bmatrix}\Bigg|_{t=t_f}.
\end{aligned}$$

The second term on the right hand side of this equation reduces to

$$\frac{1}{2}\left[z^T\left(\phi_{xx} + (\nu^T\Psi)_{xx}\right)z + \tilde{\nu}^T\Psi_x z + z^T\Psi_x^T\nu\right]$$

and setting $\Psi_x z = 0$, it further reduces to

$$\frac{1}{2}\left[z^T\left(\phi_{xx} + (\nu^T\Psi)_{xx}\right)z\right].$$

Thus,

$$\begin{aligned}\delta^2\hat{J} &= \frac{1}{2}\int_{t_0}^{t_f}\left\{\left[\begin{array}{cc} z^T & \eta^T \end{array}\right]\left[\begin{array}{cc} H_{xx} & H_{xu} \\ H_{ux} & H_{uu}\end{array}\right]\left[\begin{array}{c} z \\ \eta \end{array}\right]\right\}\mathrm{d}t \\ &\quad + \frac{1}{2}\left[z^T\left(\phi_{xx} + (\nu^T\Psi)_{xx}\right)z\right]\Big|_{t=t_f}\end{aligned}$$

Hence the problem becomes

$$\min_{\eta(\cdot)} J(\eta(\cdot)) \quad \text{subject to} \quad \Psi_x(\bar{x}(t_f))z(t_f) = 0$$

where $J(\eta(\cdot)) = \delta^2\hat{J}$.

Theorem 4.2.0.5 (GS). *A necessary condition for a functional, $J[y]$, to have a minimum for $y = \hat{y}$ is that the second variation is nonnegative. i.e. $\delta^2 J[h] \geq 0$ for $y = \hat{y}$ and for all h.*

Proof. By definition, we have that

$$\triangle J_y[h] = \delta J[h] + \delta^2 J[h] + \epsilon\|h\|^2$$

where $\epsilon \to 0$ as $\|h\| \to 0$. By Theorem 4.1.1.2, $\delta J[h] = 0$ for all h hence

$$\triangle J_y[h] = \delta^2 J[h] + \epsilon\|h\|^2.$$

Thus for sufficiently small $\|h\|$, the sign of $\triangle J_y[h]$ will be the same as the sign of $\delta^2 J[h]$. Now suppose $\delta^2 J[h_0] < 0$ for some h_0. Then for $\alpha \neq 0$, no matter how small, we have

$$\delta^2 J[\alpha h_0] = \alpha^2\delta^2 J[h_0] < 0.$$

Therefore $\delta^2 J[h] + \epsilon\|h\|^2$ can be made negative for arbitrary small $\|h\|$. But this is impossible since $\triangle J_y[h] \geq 0$. $\square$

Chapter 5

Formulation of the Linear Quadratic Control Problem

In this chapter, we will derive the quadratic cost criterion from the second variation obtained from the previous chapter. We shall also discuss the first order necessary condition for optimality and the general optimal control rule for the linear optimal control problem with unconstrained terminal condition.

We have seen what the second variation is in the previous chapter. For simplicity, suppose that

$$S_f = \left(\phi_{xx} + (\nu^T\Psi)_{xx}\right)\big|_{t=t_f}; \quad z(t) = x(t); \quad \eta(t) = u(t).$$

Then,

$$\begin{aligned}&\delta^2\hat{J}\\ &= \tfrac{1}{2}\int_{t_0}^{t_f}\left\{\begin{bmatrix} x^T(t) & u^T(t)\end{bmatrix}\begin{bmatrix} H_{xx} & H_{xu}\\ H_{ux} & H_{uu}\end{bmatrix}\begin{bmatrix} x(t)\\ u(t)\end{bmatrix}\right\}\mathrm{d}t + \tfrac{1}{2}\left[x^T(t_f)S_f x(t_f)\right]\end{aligned}$$

which implies that

$$\begin{aligned}&\delta^2\hat{J}\\ &= \tfrac{1}{2}\left[x^T(t_f)S_f x(t_f)\right]\\ &\quad +\tfrac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)H_{xx}x(t) + u^T(t)H_{ux}x(t) + x^T H_{xu}u(t) + u^T(t)H_{uu}u(t)\right]\mathrm{d}t\\ &= \tfrac{1}{2}\left[x^T(t_f)S_f x(t_f)\right]\\ &\quad +\tfrac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)H_{xx}x(t) + 2u^T(t)H_{ux}x(t) + u^T(t)H_{uu}u(t)\right]\mathrm{d}t.\end{aligned}$$

(assuming that H_{ux} is symmetric)

Thus, letting

$$Q(t) = H_{xx}; \quad C(t) = H_{ux}; \quad \text{and} \quad R(t) = H_{uu}$$

we obtain a cost criterion

$J(u(\cdot), x_0, t_0)$

$$\begin{aligned} &= \frac{1}{2}\left[x^T(t_f)S_f x(t_f)\right] \\ &+\frac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right]\mathrm{d}t \end{aligned} \tag{5.1}$$

which is a quadratic form. Thus, the Linear Quadratic Problem is the minimization of (5.1) with respect to $u(\cdot)$ subject to

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \tag{5.2}$$

with initial condition $x(t_0) = x_0$ and a linear terminal constraint

$$Dx(t_f) = 0 \tag{5.3}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ and D a $p \times n$ matrix $\forall t \in [t_0, t_f]$.

We make the following assumptions.

- First we assume that Q, C, R, A, B are piecewise continuous functions of time and we assume also that $Q(t), R(t), S_f$ are symmetric matrices.
- Secondly, we assume that $Q(t) - C(t)R^{-1}(t)C(t) \geq 0$.
- Thirdly, we assume that $R(t) > 0$, is positive definite, and bounded for all $t \in [t_0, t_f]$ and also $u(\cdot)$ belongs to the class, U, of piecewise continuous m-vector functions of time in the interval $[t_0, t_f]$.

Theorem 5.0.0.6 (First Order Necessary Condition for Optimality (Sp)**).** *Suppose the second assumption is satisfied and the matrix*

$$\bar{W}(t_0, t_f) = \int_{t_0}^{t_f} D\Phi(t_0, t)B(t)B^T(t)\Phi^T(t_0, t)D^T\mathrm{d}t$$

is positive definite (controllability assumption). Then, necessary conditions for $\triangle J$ to be nonnegative to first order for strong perturbation in the control, are that

$$\begin{aligned} \dot{x}^0(t) &= A(t)x^0(t) + B(t)u^0(t); \quad x^0(t_0) = x_0 \\ -\dot{\lambda}(t) &= A^T(t)\lambda(t) + Q(t)x^0(t) + C^T(t)u^0(t) \\ \lambda(t_f) &= S_f x^0(t_f) + D^T\nu \end{aligned}$$

$$\begin{aligned} Dx(t_f) &= 0 \\ 0 &= R(t)u^0(t) + C(t)x^0(t) + B^T(t)\lambda(t) \end{aligned}$$

where $\lambda(t)$, an n-vector function, which is the Lagrange multiplier is assumed continuous and differentiable and ν is a constant p-vector. $x^0(t)$ and $u^0(t)$ are the candidates for optimality.

Proof. The proof of this theorem is based on the fact that we try to look for conditions for which the first variation of the cost function is zero or $\triangle\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0)$ will be nonnegative. Suppose the hypothesis of the theorem and adjoin (5.1),(5.2) and (5.3) as follows

$$\begin{aligned} \hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0) &= J(u(\cdot),x_0,t_0) \\ &+ \int_{t_0}^{t_f} \lambda^T(t)[A(t)x(t) + B(t)u(t) - \dot{x}(t)]\mathrm{d}t + \nu^T Dx(t_f). \end{aligned}$$

Upon integration by parts, we obtain the following

$$\begin{aligned} &\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0) \\ &= \textstyle\int_{t_0}^{t_f}(H(t) + \dot{\lambda}^T(t)x(t))\mathrm{d}t + \lambda^T(t_0)x_0 - \lambda^T(t_f)x(t_f) + \frac{1}{2}\left[x^T(t_f)S_f x(t_f)\right] \\ &\quad +\nu^T Dx(t_f) \end{aligned}$$

where

$$\begin{aligned} H(t) &= \frac{1}{2}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right] \\ &\quad +\lambda^T(t)[A(t)x(t) + B(t)u(t)]. \end{aligned}$$

Suppose there exists $u^0(t) \in U$ that minimizes the cost criterion, (5.1), and causes $Dx^0(t_f) = 0$. We now determine the change in $\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0)$ brought about by changing $u(t)$ to $u^0(t) \in U$. Let the change in the control be $\delta u(t) = u(t) - u^0(t)$ and the resulting change in the state be $\delta x(t) = x(t) - x^0(t)$. Hence, we seek conditions that guarantee that the change in $\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0)$ is nonnegative for all variations. Consider variations for which $\delta x(t)$ remains small i.e.

$$\|\delta x(t)\| \leq \bar{\epsilon} \quad \forall t \in [t_0,t_f], \bar{\epsilon} > 0$$

and choose $\delta u(t) = \xi(t)\eta(t)$ where $\eta(t) \in U$ and for $\epsilon > 0$,

$$\xi(t) = \begin{cases} 1 & t_i \leq t \leq t_i + \epsilon\delta_i \quad i = 1,2,\cdots,n \\ \epsilon & t \in I = [t_0,t_f] - \bigcup_i^n [t_i, t_i + \epsilon\delta_i]. \end{cases} \tag{5.4}$$

where δ_i is some small positive number. Now

$$H_x = x^T(t)Q(t) + u^T(t)C(t) + \lambda^T(t)A(t)$$

and

$$H_u = x^T(t)C^T(t) + u^T(t)R(t) + \lambda^T(t)B(t).$$

Hence the variation in $\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0)$, as a result of changing $u(t)$ to $u^0(t)$, and remembering that change in $u(t)$ causes the state vector to change, is given as

$$\begin{aligned}
&\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0)\\
&= \hat{J}(u^0(\cdot);\lambda(\cdot),\nu,x_0,t_0)\\
&\quad + \int_{t_0}^{t_f}\Big[H_x\delta x(t) + H_u\delta u(t) + \dot{\lambda}^T(t)\delta x(t)\Big]\,\mathrm{d}t\\
&\quad + \int_{t_0}^{t_f}\mathcal{O}(t;\epsilon)\mathrm{d}t - \lambda^T(t_f)\delta x(t_f) + x^T(t_f)S_f\delta x(t_f) + \nu^T D\delta x(t_f) + \mathcal{O}(\epsilon)\\
&= \hat{J}(u^0(\cdot);\lambda(\cdot),\nu,x_0,t_0)\\
&\quad + \tfrac{1}{2}\int_{t_0}^{t_f}\Big[2\left(H_x\delta x(t) + H_u\delta u(t)\right) + 2\dot{\lambda}^T(t)\delta x(t)\Big]\,\mathrm{d}t\\
&\quad + \int_{t_0}^{t_f}\mathcal{O}(t;\epsilon)\mathrm{d}t - \lambda^T(t_f)\delta x(t_f) + x^T(t_f)S_f\delta x(t_f) + \nu^T D\delta x(t_f) + \mathcal{O}(\epsilon).
\end{aligned}$$

Now, we choose $\lambda(t)$ such that

$$H_x + \dot{\lambda}^T = 0$$

and

$$-\lambda^T(t_f) + x^T(t_f)S_f + \nu^T D = 0.$$

Hence

$$\triangle\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0) = \frac{1}{2}\int_{t_0}^{t_f} 2H_u\delta u(t)\mathrm{d}t + \int_{t_0}^{t_f}\mathcal{O}(t;\epsilon)\mathrm{d}t + \mathcal{O}(\epsilon).$$

We choose

$$\eta(t) = -R^{-1}(t)H_u^T = -R^{-1}(t)\left[R(t)u^0(t) + C(t)x^0(t) + B^T(t)\lambda(t)\right]$$

which implies that $H_u = -\eta^T(t)R(t)$, (since $R(t)$ is symmetric). This choice is possible because of the controllability assumption. Substituting $\delta u(t) = \xi(t)\eta(t)$ into the integral, we obtain

$$\triangle\hat{J}(u(\cdot);\lambda(\cdot),\nu,x_0,t_0) = \int_{t_0}^{t_f} H_u(t)\xi(t)\eta(t)\mathrm{d}t + \int_{t_0}^{t_f}\mathcal{O}(t;\epsilon)\mathrm{d}t + \mathcal{O}(\epsilon).$$

Thus for ϵ small,

$$\triangle \hat{J} = -\sum_{i=1}^{n} \int_{t_i}^{t_i+\epsilon\delta_i} H_u^T R^{-1}(t) H_u \mathrm{d}t - \epsilon \int_I H_u^T R^{-1}(t) H_u \mathrm{d}t + \int_{t_0}^{t_f} \mathcal{O}(t;\epsilon)\mathrm{d}t + \mathcal{O}(\epsilon),$$

where

$$\frac{\mathcal{O}(t;\epsilon)}{\epsilon} \longrightarrow 0 \quad \text{as} \quad \epsilon \longrightarrow 0$$

will be negative since $H_u R^{-1} H_u^T$ is positive and thus for it to be nonnegative, $\eta(t)$ must be zero which implies that $H_u(t)$ is zero, where

$$H_u(t) = [R(t)u^0(t) + C(t)x^0(t) + B^T(t)\lambda(t)]^T.$$

This choice of $\eta(t)$, also helps to determine ν. To see this, we proceed as follows. By the linearity of the dynamic system, $z(t, \eta(\cdot))$ satisfies

$$\dot{z}(t, \eta(\cdot)) = A(t)z(t, \eta(\cdot)) + B(t)\eta(t); \quad z_0 = 0.$$

Therefore,

$$z(t_f, \eta(\cdot)) = \int_{t_0}^{t_f} \Phi(t_f, t)B(t)\eta(t)\mathrm{d}t \quad \text{(by Theorem 2.0.0.5)}.$$

Hence, by substituting $\eta(t)$ into this equation, we obtain

$$z(t_f, \eta(\cdot)) = -\int_{t_0}^{t_f} \Phi(t_f, t)B(t)R^{-1}(t)\left[R(t)u^0(t) + C(t)x^0(t) + B^T(t)\lambda(t)\right].$$

Now from

$$-\dot{\lambda}(t) = A^T(t)\lambda(t) + Q(t)x^0(t) + C^T(t)u^0(t) \quad \text{(from the necessary condition)},$$

we can solve for $\lambda(t)$ by applying Theorem 2.0.0.5 once again to get

$$\lambda(t) = \Phi(t, t_f)S_f x^0(t_f) + \Phi(t, t_f)D^T\nu + \int_{t_0}^{t_f} \Phi(t, \tau)[Q(\tau)x^0(\tau) + C^T(t)u^0(t)]\mathrm{d}t.$$

Substituting this into $z(t_f, \eta(\cdot))$ and using the fact that $Dz(t_f, \eta(\cdot)) = 0$, (from the necessary condition, i.e. since $Dx(t_f) = 0$), we obtain an equation which we can solve for ν explicitly. □

5.1 Unconstrained Terminal Linear Quadratic Problem

From the above theorem, the necessary conditions for the unconstrained terminal problem are obtained by putting $D = 0$. In this section we shall derive the general optimal control rule and consider some important properties of the transition matrix of the Hamiltonian system.

5.1.1 The General Optimal Control Rule

In Theorem 5.0.0.6, one of the conditions for optimality is

$$H_u^T = R(t)u^0(t) + C(t)x^0(t) + B^T(t)\lambda(t) = 0$$

and thus we can solve for $u^0(t)$ in terms of $\lambda(t)$ and $x^0(t)$ since we have assumed that $R(t) > 0$ and bounded. That is

$$u^0(t) = -R^{-1}(t)\left[C(t)x^0(t) + B^T(t)\lambda(t)\right]$$

if the inverse, $R^{-1}(t)$, exists. We substitute this into

$$\dot{x}^0(t) = A(t)x^0(t) + B(t)u^0(t)$$

and

$$-\dot{\lambda}(t) = H_x^T = A^T(t)\lambda(t) + Q(t)x^0(t) + C^T(t)u^0(t)$$

to get

$$\begin{aligned} \dot{x}^0(t) &= \left(A(t) - B(t)R^{-1}(t)C(t)\right)x^0(t) - B(t)R^{-1}(t)B^T(t)\lambda(t) \\ \dot{\lambda}(t) &= \left(C^T(t)R^{-1}(t)C(t) - Q(t)\right)x^0(t) + \left(C^T(t)R^{-1}(t)B^T(t) - A^T(t)\right)\lambda(t). \end{aligned}$$

Hence, we have

$$\begin{bmatrix} \dot{x}^0(t) \\ \dot{\lambda}(t) \end{bmatrix} =$$

$$\begin{bmatrix} A(t) - B(t)R^{-1}(t)C(t) & -B(t)R^{-1}(t)B^T(t) \\ C^T(t)R^{-1}(t)C(t) - Q(t) & -\left(A(t) - B(t)R^{-1}(t)C(t)\right)^T \end{bmatrix} \begin{bmatrix} x^0(t) \\ \lambda(t) \end{bmatrix} \tag{5.5}$$

which is a $2n$-vector system of differential equation with initial conditions

$$x^0(t) = x_0 \quad \text{and} \quad \lambda(t_f) = S_f x^0(t_0) \quad \text{since} \quad D = 0.$$

For

$$\boldsymbol{H}(t) = \begin{bmatrix} A(t) - B(t)R^{-1}(t)C(t) & -B(t)R^{-1}(t)B^T(t) \\ C^T(t)R^{-1}(t)C(t) - Q(t) & -\left(A(t) - B(t)R^{-1}(t)C(t)\right)^T \end{bmatrix},$$

we write (5.5) in compact form as

$$\dot{\boldsymbol{M}}(t) = \boldsymbol{H}(t)\boldsymbol{M}(t) \tag{5.6}$$

where $\boldsymbol{H}(t)$ is referred to as the Hamiltonian Matrix. Thus there exists a transition matrix $\Phi_{\boldsymbol{H}(t)}(t, \tau)$ of $\boldsymbol{H}(t)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t, \tau) = \boldsymbol{H}(t)\Phi_{\boldsymbol{H}(t)}(t, \tau); \quad \Phi_{\boldsymbol{H}(t)}(t, t) = I \tag{5.7}$$

where we represent the Hamiltonian transition matrix in block-partition form as

$$\Phi_{\boldsymbol{H}(t)}(t, \tau) = \begin{bmatrix} \Phi_{11}(t, \tau) & \Phi_{12}(t, \tau) \\ \Phi_{21}(t, \tau) & \Phi_{22}(t, \tau) \end{bmatrix}. \tag{5.8}$$

Hence, by Theorem 2.0.0.5 the dynamic system (5.6) has solution

$$\boldsymbol{M}(t_f) = \Phi_{\boldsymbol{H}(t)}(t_f, t_0)\boldsymbol{M}(t_0)$$

i.e.

$$\begin{bmatrix} x^0(t_f) \\ \lambda(t_f) = S_f x^0(t_f) \end{bmatrix} = \begin{bmatrix} \Phi_{11}(t_f, t_0) & \Phi_{12}(t_f, t_0) \\ \Phi_{21}(t_f, t_0) & \Phi_{22}(t_f, t_0) \end{bmatrix} \begin{bmatrix} x^0(t_0) \\ \lambda(t_0) \end{bmatrix} \tag{5.9}$$

which implies that

$$x^0(t_f) = \Phi_{11}(t_f, t_0)x^0(t_0) + \Phi_{12}(t_f, t_0)\lambda(t_0) \tag{5.10}$$

$$\lambda(t_f) = S_f x^0(t_f) = \Phi_{21}(t_f, t_0)x^0(t_0) + \Phi_{22}(t_f, t_0)\lambda(t_0). \tag{5.11}$$

Substituting (5.10) into (5.11), we obtain

$$S_f\left[\Phi_{11}(t_f, t_0)x^0(t_0) + \Phi_{12}(t_f, t_0)\lambda(t_0)\right] = \Phi_{21}(t_f, t_0)x^0(t_0) + \Phi_{22}(t_f, t_0)\lambda(t_0).$$

Rearranging this and solving for $\lambda(t_0)$ we obtain

$$\lambda(t_0) = \left[S_f\Phi_{12}(t_f,t_0) - \Phi_{22}(t_f,t_0)\right]^{-1}\left[\Phi_{21}(t_f,t_0) - S_f\Phi_{11}(t_f,t_0)\right]x^0(t_0)$$

assuming the inverse, $\left[S_f\Phi_{12}(t_f,t) - \Phi_{22}(t_f,t)\right]^{-1}$, exists. For convenience, let

$$S(t_f,t,S_f) = \left[S_f\Phi_{12}(t_f,t) - \Phi_{22}(t_f,t)\right]^{-1}\left[\Phi_{21}(t_f,t) - S_f\Phi_{11}(t_f,t)\right]$$

then

$$\lambda(t_0) = S(t_f,t_0,S_f)x^0(t_0).$$

Now, we evaluate (5.10) and (5.11) over the interval $[t_0,t]$ where $t_0 \leq t \leq t_f$ to get

$$\begin{aligned} x^0(t) &= \Phi_{11}(t,t_0)x^0(t_0) + \Phi_{12}(t,t_0)\lambda(t_0) \\ \lambda(t) &= \Phi_{21}(t,t_0)x^0(t_0) + \Phi_{22}(t,t_0)\lambda(t_0). \end{aligned}$$

Therefore, substituting $\lambda(t_0) = S(t_f,t_0,S_f)x^0(t_0)$ into these equations gives

$$x^0(t) = \Phi_{11}(t,t_0)x^0(t_0) + \Phi_{12}(t,t_0)S(t_f,t_0,S_f)x^0(t_0) \tag{5.12}$$

$$\lambda(t) = \Phi_{21}(t,t_0)x^0(t_0) + \Phi_{22}(t,t_0)S(t_f,t_0,S_f)x^0(t_0). \tag{5.13}$$

Hence,

$$\begin{aligned} \begin{bmatrix} x^0(t) \\ \lambda(t) \end{bmatrix} &= \begin{bmatrix} \Phi_{11}(t,t_0) & \Phi_{12}(t,t_0) \\ \Phi_{21}(t,t_0) & \Phi_{22}(t,t_0) \end{bmatrix} \begin{bmatrix} x^0(t_0) \\ S(t_f,t_0,S_f)x^0(t_0) \end{bmatrix} \\ &= \Phi_{\boldsymbol{H}(t)}(t,t_0) \begin{bmatrix} 1 \\ S(t_f,t_0,S_f) \end{bmatrix} x^0(t_0). \end{aligned} \tag{5.14}$$

Consequently,

$$\begin{aligned} u^0(t) &= -R^{-1}(t)\left[C(t)x^0(t) + B^T(t)\lambda(t)\right] \\ &= -R^{-1}(t)\begin{bmatrix} C(t) & B^T(t) \end{bmatrix}\begin{bmatrix} x^0(t) \\ \lambda(t) \end{bmatrix} \\ &= -R^{-1}(t)\begin{bmatrix} C(t) & B^T(t) \end{bmatrix}\Phi_{\boldsymbol{H}(t)}(t,t_0)\begin{bmatrix} 1 \\ S(t_f,t_0,S_f) \end{bmatrix} x^0(t_0). \end{aligned}$$

Thus we have been able to express the control in terms of the transition matrix.
Definition 5.1.1.1 (The General Optimal Control Rule). The General optimal control rule of the Linear quadratic problems is defined as

$$u^0(t) = -R^{-1}(t)\begin{bmatrix} C(t) & B^T(t) \end{bmatrix}\Phi_{\boldsymbol{H}(t)}(t,t_0)\begin{bmatrix} 1 \\ S(t_f,t_0,S_f) \end{bmatrix} x^0(t_0).$$

We see that if $t = t_0$ then, $\Phi_{\boldsymbol{H}(t)}(t_0, t_0) = I$ and thus

$$u^0(t_0) = -R^{-1}(t_0) \left[\begin{array}{cc} C(t_0) & B^T(t_0) \end{array} \right] \left[\begin{array}{c} 1 \\ S(t_f, t_0, S_f) \end{array} \right] x^0(t_0).$$

5.1.2 Some Properties of the Transition Matrix of the Hamiltonian System

Here, we consider two properties of the transition matrix of the Hamiltonian system which will be helpful in the sequel.

The Symplectic Property

Definition 5.1.2.1 (Symplectic Matrix)**.** A symplectic matrix is a square matrix, M, with real entries that satisfies the condition

$$M^T \Omega M = \Omega$$

where Ω is a fixed square nonsingular skew-symetric matrix.

The transition matrix of the Hamiltonian system is symplectic i.e. it satisfies the following;

$$\Phi_{\boldsymbol{H}(t)}(t_f, t) \underline{J} \Phi^T_{\boldsymbol{H}(t)}(t_f, t) = \underline{J}$$

where

$$\underline{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

is known as the *fundamental symplectic matrix*. To see this we first note that the Hamiltonian matrix satisfies the following;

$$\boldsymbol{H}(t)\underline{J} + \underline{J}\boldsymbol{H}^T(t) = 0.$$

We then transpose this and premultiply by $\Phi_{\boldsymbol{H}(t)}(t_f, t)$ and postmultiply by $\Phi^T_{\boldsymbol{H}(t)}(t_f, t)$ to obtain

$$\Phi_{\boldsymbol{H}(t)}(t_f, t) \left[\boldsymbol{H}(t)\underline{J} + \underline{J}\boldsymbol{H}^T(t)\right]^T \Phi^T_{\boldsymbol{H}(t)}(t_f, t) = 0.$$

We expand this and use the fact that $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f, t) = -\Phi_{\boldsymbol{H}(t)}(t_f, t)\boldsymbol{H}(t)$, (see (6.5)), to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f, t)\underline{J}^T \Phi^T_{\boldsymbol{H}(t)}(t_f, t) = 0.$$

$\Phi_{\boldsymbol{H}(t)}(t_f,t)\underline{J}^T\Phi^T_{\boldsymbol{H}(t)}(t_f,t) = \text{const} \quad \forall t \in [t_0,t_f]$. Using the initial condition of (5.7),

$$\Phi_{\boldsymbol{H}(t)}(t,t)\underline{J}^T\Phi^T_{\boldsymbol{H}(t)}(t,t) = \underline{J}^T$$

and hence

$$\Phi_{\boldsymbol{H}(t)}(t_f,t)\underline{J}^T\Phi^T_{\boldsymbol{H}(t)}(t_f,t) = \underline{J}^T.$$

The result then follows after we transpose this matrix.

The Spectral Property

This property describes the eigenvalues of the transition matrix of the Hamiltonian matrix. Given $\underline{J}$, we see that $\underline{J}^T = \underline{J}^{-1}$. From the symplectic property,

$$\Phi^{-1}_{\boldsymbol{H}(t)}(t_f,t) = \underline{J}\Phi^T_{\boldsymbol{H}(t)}(t_f,t)\underline{J}^{-1}$$

which implies that

$$\Phi^{-1}_{\boldsymbol{H}(t)}(t_f,t) - \lambda I = \underline{J}\Phi^T_{\boldsymbol{H}(t)}(t_f,t)\underline{J}^{-1} - \lambda\underline{J}\cdot\underline{J}^{-1}.$$

Thus

$$\begin{aligned} \det\left(\Phi^{-1}_{\boldsymbol{H}(t)}(t_f,t) - \lambda I\right) &= \det\left(\underline{J}\Phi^T_{\boldsymbol{H}(t)}(t_f,t)\underline{J}^{-1} - \lambda\underline{J}\cdot\underline{J}^{-1}\right) \\ &= \det\left(\underline{J}\Phi_{\boldsymbol{H}(t)}(t_f,t)\underline{J}^{-1} - \lambda I\right) \\ &= \det\left(\Phi_{\boldsymbol{H}(t)}(t_f,t) - \lambda I\right). \end{aligned}$$

Therefore

$$\det\left(\Phi_{H(t)}(t_f,t)\right) = \pm 1.$$

This means that since $\Phi_{\boldsymbol{H}(t)}(t_f,t)$ is a $2n \times 2n$ nonsingular matrix, if μ_i for $i = 1,\cdots,n$ are n eigenvalues of $\Phi_{\boldsymbol{H}(t)}(t_f,t)$ with $\mu_i \neq 0$ for all i, then, the remaining n eigenvalues are

$$\mu_{i+n} = \frac{1}{\mu_i} \quad i = 1,\cdots,n$$

since the product of all the $2n$ eigenvalues must be the determinant which in this case is one (see appendix- Theorem A.5.0.17). Now, let

$$\Phi_{\boldsymbol{H}(t)}(t_f,t) = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}.$$

Then, the symplectic property becomes

$$\begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \Phi_{11}^T & \Phi_{21}^T \\ \Phi_{12}^T & \Phi_{22}^T \end{pmatrix} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

which implies that

$$\begin{aligned} \Phi_{11}\Phi_{12}^T &= \Phi_{12}\Phi_{11}^T \\ \Phi_{21}\Phi_{22}^T &= \Phi_{22}\Phi_{21}^T \\ \Phi_{11}\Phi_{22}^T &= I + \Phi_{12}\Phi_{21}^T. \end{aligned}$$

These relations are going to be useful in the future.

Chapter 6

Riccati Matrix Differential Equation

Instead of obtaining $S(t_f, t; S_f)$ from the transition matrix of the Hamiltonian system, we show here that $S(t_f, t; S_f)$ is propagated directly by a quadratic matrix differential equation, the matrix Riccati equation, and thus $S(t_f, t, S_f)$ can be obtained by solving this differential equation. Therefore in this chapter, we derive the Riccati matrix differential equation and some of its properties.

Given a linear homogeneous differential equation, in x defined on an inner product space, we will say that a linear homogeneous differential equation in y defined on the same inner product space is the *Adjoint Equation* associated with the given equation in x provided that for any initial data, the inner product of the solution x and solution y is constant.

We first find the adjoint form for propagating the transition matrix of the Hamiltonian system and proceed with the analysis. Recall from the previous chapter that the Hamiltonian system is symplectic. Then the matrix defined as

$$\Theta_{\boldsymbol{H}(t)}(t, t_0) = L\Phi_{\boldsymbol{H}(t)}(t, t_0) \quad \text{where} \quad L = \begin{pmatrix} I & 0 \\ -S_f & I \end{pmatrix}$$

is also symplectic since L is symplectic with respect to the fundamental symplectic matrix. Now suppose that $\Theta_{\boldsymbol{H}}$ can be written in block partition form as

$$\Theta_{\boldsymbol{H}(t)}(t, t_0) = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix},$$

then we have that

$$\Theta_{\boldsymbol{H}(t)}(t, t_0) = \begin{pmatrix} I & 0 \\ -S_f & I \end{pmatrix} \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}$$

$$\Theta_{\boldsymbol{H}(t)}(t,t_0) = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ -S_f\Phi_{11}+\Phi_{21} & -S_f\Phi_{12}+\Phi_{22} \end{pmatrix}$$

and this implies that

$$\begin{aligned} \theta_{11} &= \Phi_{11} && (6.1)\\ \theta_{12} &= \Phi_{12} && (6.2)\\ \theta_{21} &= -S_f\Phi_{11}+\Phi_{21} && (6.3)\\ \theta_{22} &= -S_f\Phi_{12}+\Phi_{22}. && (6.4) \end{aligned}$$

Now, by Theorem 2.0.0.4, the transition matrix $\Phi_{\boldsymbol{H}(t)}(t_f,t_0)$ satisfies

$$\Phi_{\boldsymbol{H}(t)}(t_f,t_0) = \Phi_{\boldsymbol{H}(t)}(t_f,t)\Phi_{\boldsymbol{H}(t)}(t,t_0).$$

Then, differentiating both sides with respect to t, where t_f and t_0 are fixed times, gives

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f,t_0) = \frac{\mathrm{d}}{\mathrm{d}t}\left[\Phi_{\boldsymbol{H}(t)}(t_f,t)\Phi_{\boldsymbol{H}(t)}(t,t_0)\right].$$

Therefore,

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f,t)\right)\Phi_{\boldsymbol{H}(t)}(t,t_0) + \Phi_{\boldsymbol{H}(t)}(t_f,t)\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t,t_0) = 0.$$

Using (5.7), this can be writtten as

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f,t)\right)\Phi_{\boldsymbol{H}(t)}(t,t_0) + \Phi_{\boldsymbol{H}(t)}(t_f,t)\boldsymbol{H}(t)\Phi_{\boldsymbol{H}(t)}(t,t_0) = 0$$

and hence

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f,t) + \Phi_{\boldsymbol{H}(t)}(t_f,t)\boldsymbol{H}(t)\right]\Phi_{\boldsymbol{H}(t)}(t,t_0) = 0.$$

Since the transition matrix of the Hamiltonian system is not zero, it implies that the term in the bracket must be zero i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f,t) + \Phi_{\boldsymbol{H}(t)}(t_f,t)\boldsymbol{H}(t) = 0.$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f,t) = -\Phi_{\boldsymbol{H}(t)}(t_f,t)\boldsymbol{H}(t); \quad \Phi_{\boldsymbol{H}(t)}(t_f,t_f) = I \tag{6.5}$$

is the adjoint form for propagating the transition matrix of the Hamiltonian system where the input time t_0 is the independent variable and the output time t_f, is fixed. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Theta_{\boldsymbol{H}(t)}(t_f,t) = \frac{\mathrm{d}}{\mathrm{d}t}L\Phi_{\boldsymbol{H}(t)}(t_f,t)$$

implies that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Theta_{\boldsymbol{H}(t)}(t_f,t) &= L\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\boldsymbol{H}(t)}(t_f,t); \quad \text{since } t_f \text{ is given, L is a constant}\\
&= -L\Phi_{\boldsymbol{H}(t)}(t_f,t)\boldsymbol{H}(t)\\
&= -\Theta_{\boldsymbol{H}(t)}(t_f,t)\boldsymbol{H}(t)
\end{aligned}$$

with the initial condition

$$\Theta_{\boldsymbol{H}(t)}(t_f,t_f) = L\Phi_{\boldsymbol{H}(t)}(t_f,t_f) = LI = L.$$

Now, we recall from the previous chapter that

$$S(t_f,t,S_f) = \left[S_f\Phi_{12}(t_f,t) - \Phi_{22}(t_f,t)\right]^{-1}\left[\Phi_{21}(t_f,t) - S_f\Phi_{11}(t_f,t)\right].$$

Using (6.1)-(6.4) and assuming the inverse, $\left[S_f\Phi_{12}(t_f,t_0) - \Phi_{22}(t_0,t_f)\right]^{-1}$, exists, this equation reduces to

$$S(t_f,t,S_f) = -\theta_{22}^{-1}(t_f,t)\theta_{21}(t_f,t)$$

and thus

$$-\theta_{21}(t_f,t) = \theta_{22}(t_f,t)S(t_f,t,S_f).$$

Differentiating this equation with respect to t, we obtain

$$-\frac{\mathrm{d}}{\mathrm{d}t}\theta_{21}(t_f,t) = \left(\frac{\mathrm{d}}{\mathrm{d}t}\theta_{22}(t_f,t)\right)S(t_f,t,S_f) + \theta_{22}(t_f,t)\frac{\mathrm{d}}{\mathrm{d}t}S(t_f,t,S_f). \tag{6.6}$$

Now, from

$$\frac{\mathrm{d}}{\mathrm{d}t}\Theta_{\boldsymbol{H}(t)}(t_f,t) = -\Theta_{\boldsymbol{H}(t)}(t_f,t)\boldsymbol{H}(t),$$

we get that

$$\begin{pmatrix}\dot{\theta}_{11} & \dot{\theta}_{12}\\ \dot{\theta}_{21} & \dot{\theta}_{22}\end{pmatrix} = -\begin{pmatrix}\theta_{11} & \theta_{12}\\ \theta_{21} & \theta_{22}\end{pmatrix}\begin{pmatrix}K & P\\ N & -K^T\end{pmatrix}$$

where, $K = A(t) - B(t)R^{-1}(t)C(t)$, $P = -B(t)R^{-1}(t)B^T(t)$ and $N = C^T(t)R^{-1}(t)C(t) - Q(t)$. Therefore,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\theta_{21}(t_f,t) &= -\left(\theta_{21}(t_f,t)K + \theta_{22}(t_f,t)N\right) \quad \text{and}\\
\frac{\mathrm{d}}{\mathrm{d}t}\theta_{22}(t_f,t) &= \theta_{22}(t_f,t)K^T - \theta_{21}(t_f,t)P.
\end{aligned}$$

We substitute this into (6.6) to get

$$(\theta_{21}K + \theta_{22}N) = \left(\theta_{22}K^T - \theta_{21}P\right) S(t_f, t, S_f) + \theta_{22}\frac{\mathrm{d}}{\mathrm{d}t}S(t_f, t, S_f).$$

With the assumption that the inverse of θ_{22} exists, we can multiply through by this inverse to get

$$\theta_{22}^{-1}\theta_{21}K + N = (K^T - \theta_{22}^{-1}\theta_{21}P)S(t_f, t, S_f) + \frac{\mathrm{d}}{\mathrm{d}t}S(t_f, t, S_f).$$

Using $S(t_f, t, S_f) = -\theta_{22}^{-1}(t_f, t)\theta_{21}(t_f, t)$, the equation becomes

$$-S(t_f, t, S_f)K + N = (K^T + S(t_f, t, S_f)P)S(t_f, t, S_f) + \frac{\mathrm{d}}{\mathrm{d}t}S(t_f, t, S_f).$$

Therfore

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t_f, t, S_f) = N - S(t_f, t, S_f)K - [K^T + S(t_f, t, S_f)P]S(t_f, t, S_f) \tag{6.7}$$

with boundary condition $S(t_f, t_f, S_f) = S_f$. This equation with its boundary condition is referred to as the *Riccati matrix differential equation.* Thus we have proved the following theorem.

Theorem 6.0.2.2 (Sp). *Let $\Phi_{\boldsymbol{H}(t)}(t_f, t)$ be the transition matrix for the Hamiltonian system defined in* (5.6) *with block partitioning defined by* (5.8)*. Then the symmetric matrix $S(t_f, t; S_f)$ satisfies* (6.7) *and its boundary condition if S_f is symmetric and $\Phi_{22}(t_f, t)$ is invertible.*

The following theorem shows that the necessary and sufficient condition for the quadratic cost function to be positive definite depends on the existence of the Riccati variable, $S(t_f, t; S_f)$, and it also justifies that our control rule is actually the minimizing controller. We assume that the dynamic system, (5.2), is completely controllable. That is to say it is controllable on any interval $[t_0, t']$ where $t_0 \le t \le t' \le t_f$. Hence by Theorem 3.2.1.4, there exists a bounded control

$$\bar{u}(t) = -B^T(t)\Phi_A^T(t', t)W_A^{-1}(t', t)x(t')$$

which transfers the state $x(t_0) = 0$ to any desired state $x(t')$ at $t = t'$ where $W_A(t', t)$ is the controllability Gramian of the dynamic system and $\Phi_A(t', t)$ is the transition matrix corresponding to the matrix $A(t)$ from the dynamic equation.

Definition 6.0.2.3. For initial condition $x_0 = 0$, $J(u(\cdot); x_0, t_0)$ is said to be **positive definite** if for each $u(\cdot) \in U$ such that $u(\cdot) \neq 0$, $J(u(\cdot); 0, t_0) > 0$.

Theorem 6.0.2.4 (**Necessary and Sufficient condition for the Positivity of the Quadratic cost function** (Sp)). *Suppose that the dynamic system* (5.2) *is completely* *controllable. A necessary and sufficient condition for the quadratic cost criterion* (5.1) *to be positive definite, is that there exists a Riccati variable $S(t_f, t; S_f)$ for all $t \in (t_0, t_f]$ that satisfies the Riccati equation* (6.7).

Proof. Let the dynamic system, (5.2), be controllable and consider the zero quantity

$$\begin{aligned}&\tfrac{1}{2}\int_{t_0}^{t_f}\tfrac{\mathrm{d}}{\mathrm{d}t}\left(x^T(t)S(t_f,t;S_f)x(t)\right)\mathrm{d}t\\ &\quad+\tfrac{1}{2}x^T(t_0)S(t_f,t_0;S_f)x(t_0)-\tfrac{1}{2}x^T(t_f)S(t_f,t_f;S_f)x(t_f)=0.\end{aligned}$$

Then, we have that

$$\begin{aligned}&\tfrac{1}{2}\int_{t_0}^{t_f}\left[\dot{x}^T(t)S(t_f,t;S_f)x(t)+x^T(t)S(t_f,t;S_f)\dot{x}(t)+x^T(t)\tfrac{\mathrm{d}}{\mathrm{d}t}S(t_f,t;S_f)x(t)\right]\mathrm{d}t\\ &\quad+\tfrac{1}{2}x^T(t_0)S(t_f,t_0;S_f)x(t_0)-\tfrac{1}{2}x^T(t_f)S(t_f,t_f;S_f)x(t_f)=0.\end{aligned}$$

Now, using the dynamic sytem, (5.2), and the Riccati equation, (6.7), this equation becomes

$$\begin{aligned}0 \;=\;&\\ &+\frac{1}{2}\int_{t_0}^{t_f}\Big\{\left(A(t)x(t)+B(t)u(t)\right)^T S(t_f,t;S_f)x(t)\\ &\qquad+x^T(t)S(t_f,t;S_f)\left[\left(A(t)x(t)+B(t)u(t)\right)\right]\\ &+x^T(t)(N-S(t_f,t;S_f)K-[K^T-S(t_f,t;S_f)P]S(t_f,t;S_f))x(t)\Big\}\,\mathrm{d}t\\ &+\frac{1}{2}x^T(t_0)S(t_f,t_0;S_f)x(t_0)-\frac{1}{2}x^T(t_f)S(t_f,t_f;S_f)x(t_f)\end{aligned}$$

where K, N, P are defined as previously. Now, adding this zero quantity to the cost criterion, (5.1), and using the boundary condition of the Riccati equation, we will have that

$$\begin{aligned}&J(u(\cdot);x(t_0),t_0)\\ &\quad=\tfrac{1}{2}x^T(t_0)S(t_f,t_0;S_f)x(t_0)\\ &\qquad+\tfrac{1}{2}\int_{t_0}^{t_f}\left[u(t)+R^{-1}(t)\left(C(t)+B^T(t)S(t_f,t;S_f)\right)x(t)\right]^T\\ &\qquad\quad\times R(t)\left[u(t)+R^{-1}(t)\left(C(t)+B^T(t)S(t_f,t;S_f)\right)x(t)\right]\mathrm{d}t.\end{aligned}$$

Consequently, from the general control rule, choosing

$$u(t)=-R^{-1}(t)\left(C(t)+B^T(t)S(t_f,t;S_f)\right)x(t),$$

reduces the cost criterion to

$$J(u(\cdot);x(t_0),t_0)=\frac{1}{2}x^T(t_0)S(t_f,t_0;S_f)x(t_0)\geq 0$$

since $S(t_f, t_0, S_f) \geq 0$ regardless of t_0 (refer to Theorem 6.1.0.10). This is the minimum value of the cost function since from the general optimal control, this choice of u is the optimal control. Thus, the cost takes on its minimum value when $u(t)$ takes the form of the general optimal control where for $x(t_0) = 0$, the optimal control is the null control and any other control $u(\cdot) \neq u^0(\cdot)$ will give a positive value to $J(u(\cdot); x(t_0), t_0)$. Hence the Riccati variable, $S(t_f, t_0; S_f)$, must exist for this to be true. Thus on the interval $t' \leq t \leq t_f$, the minimizing cost will be

$$J(u^0(\cdot); x(t'), t_0) = \frac{1}{2} x^T(t') S(t_f, t'; S_f) x(t')$$

upon using the optimal control from t' to t_f for some $x(t')$. The existence of the Riccati variable, $S(t_f, t; S_f)$, is indispensable and to see this, we proceed by contradiction. Suppose there exists an escape time, $t = t_e$,(time where the solution to the Riccati differential equation goes to infinity), such that the Riccati variable ceases to exist. Now, by the controllability assumption, we can get a bounded control, $\bar{u}(t) \in U$, which transfers $x(t_0) = 0$ to any $x(t')$ at $t' > t_0$ such that

$$u(t) = \bar{u}(t), \quad t_0 \leq t \leq t'$$

and

$$u(t) = u^0(t), \quad t' < t \leq t_f.$$

Thus the cost criterion of (5.1) can be written as

$J(u(\cdot), 0, t_0)$

$$\begin{aligned} &= \frac{1}{2}\left[x^T(t') S(t_f, t'; S_f) x(t')\right] \\ &\quad + \frac{1}{2}\int_{t_0}^{t'} \left[x^T(t) Q(t) x(t) + 2\bar{u}^T(t) C(t) x(t) + \bar{u}^T(t) R(t) \bar{u}(t)\right] \mathrm{d}t \end{aligned} \tag{6.8}$$

for all $t \in [t_0, t']$ since the first term on the right side of the equation will be the minimum value of the cost criterion in this interval. Since $\bar{u}(t)$ is bounded, by the controllablity assumption, the integral in (6.8) is also bounded. Consequently, $J(u(\cdot), 0, t_0) \longrightarrow -\infty$ if $\frac{1}{2}\left[x^T(t') S(t_f, t'; S_f) x(t')\right] \longrightarrow -\infty$ as $t' \longrightarrow t_e$, $(t_e < t')$ for some $x(t')$ which contradicts the fact that $J(u(\cdot), 0, t_0) > 0$. Similarly, $x^T(t') S(t_f, t'; S_f) x(t')$ cannot go to infinity since $J(u(\cdot), 0, t_0)$ is finite. Hence the Riccati variable must exist and be finite. □

To ensure that the second variation dominates the expansion, it is required to be strongly positive.

Definition 6.0.2.5. $J(u(\cdot); 0, t_0)$ is said to be strongly postive if for each $u(\cdot) \in U$ and some $k > 0$,

$$J(u(\cdot); 0, t_0) \geq k \|u(\cdot)\|^2$$

where $\| \cdot \|$ is some suitable norm, specifically the L_2-norm, defined on U. (See example

2 in the appendix.)

Theorem 6.0.2.6 (GS). *Suppose the first variation δJ of a functional $J[y]$ vanishes for $y = \hat{y}$. If the second variation $\delta^2 J$ is strongly positive for $y = \hat{y}$ then $J[y]$ has a minimum.*

Proof. Suppose that the first variation vanishes. Let the second variation be strongly positive, i.e $\delta^2 J[h] \geq k\|h\|^2$ for $k > 0$. Now, by definition, $\triangle J_y[h] = \delta^2 J[h] + \epsilon\|h\|^2$, since $\delta J[h] = 0$ where $\epsilon \to 0$ as $\|h\| \to 0$. Therefore

$$\triangle J_y[h] \geq (k + \epsilon)\|h\|^2 \geq 0,$$

so $J[y]$ has a minimum. □

Hence we state and prove the necessary and sufficient condition for strong positivity of our cost criterion. We establish that the necessary and sufficient condition for the second variation, as represented by $J(u(\cdot); 0, t_0)$, to be strongly positive depends on the existence of a solution to the Riccati differential equation (6.7).

Theorem 6.0.2.7 (**Necessary and Sufficient Condition for Strong Positivity** (Sp)). *A necessary and sufficient condition for $J(u(\cdot); 0, t_0)$ to be strongly positive is that for all $t \in [t_0, t_f]$, there exists a function $S(t_f, t; S_f)$ which satisfies the Riccati differential equation* (6.7).

Proof. Suppose there exists a solution to (6.7) then $J(u(\cdot); 0, t_0)$ is positive definite regardless of t_0 by Theorem 6.0.2.4. Now, consider the linear quadratic problem whose cost function, $J(u(\cdot); \epsilon; 0, t_0)$, is given as

$$\begin{aligned} &\frac{1}{2}\left[x^T(t_f)S_f x(t_f)\right] \\ &+\frac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + \frac{2}{2-\epsilon}u^T(t)R(t)u(t)\right]\mathrm{d}t. \end{aligned}$$

Then with the same analogy from Theorem 6.0.2.4, we can say that $J(u(\cdot); \epsilon; 0, t_0)$ is positive definite and finite if and only if

$$\begin{aligned} -\frac{\mathrm{d}}{\mathrm{d}t}S(t_f, t; S_f, \epsilon) &= Q(t) + S(t_f, t; S_f, \epsilon)A(t) + A^T(t)S(t_f, t; S_f, \epsilon) \\ &\quad -\frac{2-\epsilon}{2}\left[C(t) + B^T(t)S(t_f, t; S_f, \epsilon)\right]^T \\ &\quad \times R^{-1}(t)\left[C(t) + B^T(t)S(t_f, t; S_f, \epsilon)\right]; \end{aligned} \tag{6.9}$$

$$S(t_f, t_f; S_f, \epsilon) = S_f$$

has a solution $S(t_f, t; S_f, \epsilon)$ defined for all $t \in [t_0, t_f]$ and $2-\epsilon > 0$. This is true because for $J(u(\cdot); \epsilon; 0, t_0)$ to be positive definite and finite, the solution, $S(t_f, t; S_f, \epsilon)$, of (6.9) must exist for sufficiently small ϵ. We can see also that for $\epsilon = 0$, $J(u(\cdot); \epsilon; 0, t_0) = J(u(\cdot); 0, t)$

and thus $S(t_f, t; S_f, \epsilon = 0)$ exists. Hence for $\epsilon < 0$ and sufficiently small, $J(u(\cdot); \epsilon; 0, t_0)$ is positive definite. Now,

$$J(u(\cdot); \epsilon; 0, t_0) - J(u(\cdot); 0, t_0) = \int_{t_0}^{t_f} \left(\frac{1}{2-\epsilon} - \frac{1}{2} \right) u^T(t) R(t) u(t) \mathrm{d}t.$$

Hence,

$$J(u(\cdot); \epsilon; 0, t_0) = J(u(\cdot); 0, t_0) + \frac{\epsilon}{2(2-\epsilon)} \int_{t_0}^{t_f} u^T(t) R(t) u(t) \mathrm{d}t.$$

Since we have shown that $J(u(\cdot); \epsilon; 0, t_0)$ is positive definite it implies that

$$J(u(\cdot); 0, t_0) + \frac{\epsilon}{2(2-\epsilon)} \int_{t_0}^{t_f} u^T(t) R(t) u(t) \mathrm{d}t \geq 0$$

and thus

$$J(u(\cdot); 0, t_0) \geq -\frac{\epsilon}{2(2-\epsilon)} \int_{t_0}^{t_f} u^T(t) R(t) u(t) \mathrm{d}t.$$

Since we have assumed that $R(t) > 0$ and bounded, we can have that $0 < k_1 \leq \|R(t)\| \leq k_2$ (with the appropriate norm). Therefore,

$$J(u(\cdot); 0, t_0) \geq -\frac{\epsilon k_1}{2(2-\epsilon)} \int_{t_0}^{t_f} u^T(t) u(t) \mathrm{d}t.$$

Then,

$$J(u(\cdot); 0, t_0) \geq k \|u(\cdot)\|_{L_2}^2$$

where

$$k = -\frac{\epsilon k_1}{2(2-\epsilon)} > 0 \quad \text{since} \quad \epsilon < 0.$$

Hence $J(u(\cdot); 0, t_0)$ is strongly positive definite by definition. Conversely, suppose that $J(u(\cdot); 0, t_0)$ is strongly positive then by definition, $J(u(\cdot); 0, t_0) \geq k\|u(\cdot)\|^2$ for $k > 0$. Now for all $u(\cdot)$, $J(u(\cdot); 0, t_0) \geq 0$. Thus $J(u(\cdot); 0, t_0)$ is positive definite and hence the Riccati variable must exist as established by Theorem 6.0.2.4. $\square$

6.1 Some Properties of the Riccati Variable

In the following, we discuss some important properties of solution to the matrix Riccati differential equation. We first show that if the terminal weight, S_f, in the cost criterion is increased, then there is a corresponding increase in the solution of matrix Riccati differential equation. Secondly, we show that upon restricting certain matrices in the cost criterion to be positive semidefinite, the Riccati variable also remains nonnegative and bounded. Moreover, we will show that if the terminal weight is zero, $S_f = 0$, and upon imposing some restriction, then $S(t_f, t; 0)$ is monotonically increasing as t_f increases.

Theorem 6.1.0.8 (Sp). *If S_f^1 and S_f^2 are two terminal weights in the cost criterion,* (5.1), *such that $S_f^1 - S_f^2 \geq 0$, then the difference $S(t_f, t; S_f^1) - S(t_f, t; S_f^2) \geq 0$ for all $t \in [t_0, t_f]$ where $S(t_f, t; S_f^2)$ exists and $S(t_f, t; S_f^1)$ and $S(t_f, t; S_f^2)$ are solutions of* (6.7) *for S_f^1 and S_f^2 respectively.*

Proof. Consider the cost criterion

$$J(u(\cdot), x_0, t_0)$$

$$\begin{aligned} &= \frac{1}{2}\left[x^T(t_f)S_f x(t_f)\right] \\ &+\frac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right]\mathrm{d}t. \end{aligned}$$

We have also seen that

$$\begin{aligned} &\tfrac{1}{2}\int_{t_0}^{t_f}\tfrac{\mathrm{d}}{\mathrm{d}t}\left(x^T(t)S(t_f, t; S_f)x(t)\right)\mathrm{d}t \\ &+\tfrac{1}{2}x^T(t_0)S(t_f, t_0; S_f)x(t_0) - \tfrac{1}{2}x^T(t_f)S(t_f, t_f; S_f)x(t_f) = 0. \end{aligned}$$

Adding these two equations, we will arrive at

$$J(u(\cdot); x_0, t_0)$$

$$\begin{aligned} &= \frac{1}{2}\left[x^T(t_f)S_f x(t_f)\right] - \frac{1}{2}\left[x^T(t_f)S(t_f, t_f; S_f)x(t_f)\right] \\ &+\frac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right]\mathrm{d}t \\ &+\frac{1}{2}\int_{t_0}^{t_f}\frac{\mathrm{d}}{\mathrm{d}t}\left(x^T(t)S(t_f, t; S_f)x(t)\right)\mathrm{d}t \\ &+\frac{1}{2}x^T(t_0)S(t_f, t_0; S_f)x(t_0). \end{aligned}$$

Simplifying this by using the Riccati equation and the dynamic system, as in the proof of Theorem 6.0.2.4, and using the fact that $S(t_f, t_f; S_f) = S_f$, we obtain

$$\begin{aligned} &J(u(\cdot); x(t_0), t_0) \\ &= \tfrac{1}{2}x^T(t_0)S(t_f, t_0; S_f)x(t_0) \\ &+\tfrac{1}{2}\int_{t_0}^{t_f}\left[u(t) + R^{-1}(t)\left(C(t) + B^T(t)S(t_f, t; S_f)\right)x(t)\right]^T \\ &\quad \times R(t)\left[u(t) + R^{-1}(t)\left(C(t) + B^T(t)S(t_f, t; S_f)\right)x(t)\right]\mathrm{d}t. \end{aligned}$$

Since the general control rule can be taken to be

$$u(t) = -R^{-1}(t) \begin{bmatrix} C(t) & B^T(t) \end{bmatrix} \begin{bmatrix} 1 \\ S(t_f, t, S_f) \end{bmatrix} x(t), \forall t \in [t_0, t_f],$$

this equation will reduce to

$$J(u(\cdot); x_0, t_0) = \frac{1}{2} x^T(t_0) S(t_f, t_0; S_f) x(t_0). \tag{6.10}$$

This equation is the optimal cost. Given S_f^1 and S_f^2 as the two terminal weights, we can get

$J_1(u(\cdot), x_0, t_0)$

$$\begin{aligned} &= \frac{1}{2}\left[x^T(t_f) S_f^1 x(t_f)\right] \\ &\quad + \frac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right] \mathrm{d}t \end{aligned}$$

and

$$\begin{aligned} &\tfrac{1}{2}\textstyle\int_{t_0}^{t_f} \frac{\mathrm{d}}{\mathrm{d}t}\left(x^T(t)S(t_f, t; S_f^2)x(t)\right) \mathrm{d}t \\ &\quad + \tfrac{1}{2} x^T(t_0) S(t_f, t_0; S_f^2) x(t_0) - \tfrac{1}{2} x^T(t_f) S(t_f, t_f; S_f^2) x(t_f) = 0. \end{aligned}$$

Adding this zero quantity to $J_1(u(\cdot), x_0, t_0)$ and observing that $S(t_f, t_f; S_f^2) = S_f^2$, we obtain

$J_1(u(\cdot), x_0, t_0)$

$$\begin{aligned} &= \frac{1}{2}\left[x^T(t_f) S_f^1 x(t_f)\right] - \frac{1}{2} x^T(t_f) S(t_f, t_f; S_f^2) x(t_f) \\ &\quad + \frac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right] \mathrm{d}t \\ &\quad + \frac{1}{2}\int_{t_0}^{t_f} \frac{\mathrm{d}}{\mathrm{d}t}\left(x^T(t)S(t_f, t; S_f^2)x(t)\right) \mathrm{d}t \\ &\quad + \frac{1}{2} x^T(t_0) S(t_f, t_0; S_f^2) x(t_0) \quad \text{which implies that} \end{aligned}$$

$J_1(u(\cdot), x_0, t_0)$

$$\begin{aligned} &= \frac{1}{2} x^T(t_0) S(t_f, t_0; S_f^2) x(t_0) + \frac{1}{2}\left[x^T(t_f)(S_f^1 - S_f^2) x(t_f)\right] \\ &\quad + \frac{1}{2}\int_{t_0}^{t_f}\left[u(t) + R^{-1}(t)\left(C(t) + B^T(t)S(t_f, t; S_f^2)\right) x(t)\right]^T \\ &\qquad \times R(t)\left[u(t) + R^{-1}(t)\left(C(t) + B^T(t)S(t_f, t; S_f^2)\right) x(t)\right] \mathrm{d}t. \end{aligned}$$

Now, let

$$v(t) = u(t) + R^{-1}(t)\left(C(t) + B^T(t)S(t_f, t; S_f^2)\right)x(t).$$

Then,

$J_1(u(\cdot), x_0, t_0)$

$$\begin{aligned} &= \frac{1}{2}x^T(t_0)S(t_f, t_0; S_f^2)x(t_0) + \frac{1}{2}\left[x^T(t_f)(S_f^1 - S_f^2)x(t_f)\right] \\ &\quad + \frac{1}{2}\int_{t_0}^{t_f} v^T(t)R(t)v(t)\mathrm{d}t. \end{aligned}$$

From (6.10), $J_1(u(\cdot), x_0, t_0)$ will also have an optimal cost of

$$\frac{1}{2}x^T(t_0)S(t_f, t_0; S_f^1)x(t_0).$$

i.e.

$$J_1(u(\cdot), x_0, t_0) = \frac{1}{2}x^T(t_0)S(t_f, t_0; S_f^1)x(t_0).$$

Let

$$J_2(u(\cdot), x_0, t_0) = \frac{1}{2}\left[x^T(t_f)(S_f^1 - S_f^2)x(t_f)\right] + \frac{1}{2}\int_{t_0}^{t_f} v^T(t)R(t)v(t)\mathrm{d}t$$

which is a new cost criterion and since $S_f^1 - S_f^2 \geq 0$, it implies that $J_2(u(\cdot), x_0, t)) \geq 0$ and thus

$$\min_{u(\cdot)} J_2(u(\cdot), x_0, t_0) \geq 0.$$

Therefore,

$$\frac{1}{2}x^T(t_0)S(t_f, t_0; S_f^1)x(t_0) = \frac{1}{2}x^T(t_0)S(t_f, t_0; S_f^2)x(t_0) + \min_{u(\cdot)} J_2(u(\cdot), x_0, t_0)$$

which implies that

$$\frac{1}{2}x^T(t_0)S(t_f, t_0; S_f^1)x(t_0) - \frac{1}{2}x^T(t_0)S(t_f, t_0; S_f^2)x(t_0) = \min_{u(\cdot)} J_2(u(\cdot), x_0, t_0) \geq 0.$$

Hence,

$$\frac{1}{2}x^T(t_0)\left(S(t_f, t_0; S_f^1) - S(t_f, t_0; S_f^2)\right)x(t_0) \geq 0$$

and therefore,

$$S(t_f, t_0; S_f^1) - S(t_f, t_0; S_f^2) \geq 0, \quad \forall t_0 \leq t_f$$

as required. □

The next property of the Riccati variable is based on the following theorem.

Theorem 6.1.0.9 (Sp). *If $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ is controllable on the interval*

$[t_0, t_f]$ *and if* $u(t) = v(t) + \Lambda(t)x(t)$, *then*

$$\begin{aligned} \dot{x}(t) &= (A(t) + B(t)\Lambda(t))\, x(t) + B(t)v(t) \\ &= \underline{A}(t)x(t) + B(t)v(t) \end{aligned}$$

is controllable with respect to $v(t)$ *for any finite piecewise continuous* $\Lambda(t)$ *on the interval* $[t_0, t_f]$ *where* $\underline{A}(t) = A(t) + B(t)\Lambda(t)$.

Proof. Let the state x with dynamic system $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ be controllable. Then by Theorem 3.2.1.4, there is a control $u(t)$ which transfers the state from the value, say x_0 at $t = t_0$, to the value say, x_f at $t = t_f > t_0$ such that $x_0 - \Phi_A(t_0, t_f)x_f$ is in the range space of

$$W_A(t_0, t_f) = \int_{t_0}^{t_f} \Phi_A(t_0, t)B(t)B^T(t)\Phi_A^T(t_0, t)\mathrm{d}t \tag{6.11}$$

where $\Phi_A(t_0, t)$ is the transition matrix corresponding to the matrix $A(t)$. Suppose that $W_A(t_0, t_f) = 0$. Then $\Phi_A(t_0, t)B(t)$ must be zero and thus $x_0^T\Phi_A(t_0, t)B(t) = 0$. $W_A(t_0, t_f) = 0$ also means that $x_0 - \Phi_A(t_0, t_f)x_f = 0$ since it is in the range space of $W_A(t_0, t_f)$. Now, $x_0 - \Phi_A(t_0, t_f)x_f = 0$ means that the state needs no control to accomplish the transfer because the free motion passes through x_f [Ro]. Thus $x_0^T\Phi_A(t_0, t)B(t)$ must not be zero. It then sufices to show that $x_0^T\Phi_{\underline{A}}(t_0, t)B(t) = 0$ implies that $x_0^T\Phi_A(t_0, t)B(t) = 0$ which will contradict the fact that $x_0^T\Phi_A(t_0, t)B(t)$ must not be zero where $\Phi_{\underline{A}}(t_0, t)$ is the transition matrix of $\underline{A}(t)$. Now by Theorem 3.2.1.6, (6.11) satisfies the linear system

$$\frac{\mathrm{d}}{\mathrm{dt}}W_A(t, t_f) = A(t)W_A(t, t_f) + W_A(t, t_f)A^T(t) - B(t)B^T(t); \quad W_A(t_f, t_f) = 0.$$

Similarly, the controllability Gramian of $\underline{A}(t)$ also satisfies the linear system

$$\frac{\mathrm{d}}{\mathrm{dt}}W_{\underline{A}}(t, t_f) = \underline{A}(t)W_{\underline{A}}(t, t_f) + W_{\underline{A}}(t, t_f)\underline{A}^T(t) - B(t)B^T(t); \quad W_{\underline{A}}(t_f, t_f) = 0.$$

Now, define the matrix

$$E(t) = W_{\underline{A}}(t, t_f) - W_A(t, t_f). \tag{6.12}$$

Then

$$\begin{aligned} \dot{E}(t) &= \dot{W}_{\underline{A}}(t, t_f) - \dot{W}_A(t, t_f) \\ &= \underline{A}(t)E(t) + E(t)\underline{A}^T(t) + B(t)\Lambda(t)W_A(t, t_f) + W_A(t, t_f)\Lambda^T(t)B(t) \end{aligned}$$

where we have made use of the above two linear systems and the fact that $A(t) = \underline{A}(t) - B(t)\Lambda(t)$. Therefore,

$$\dot{E}(t) = \underline{A}(t)E(t) + E(t)\underline{A}^T(t) - \left[-B(t)\Lambda(t)W_A(t,t_f) - W_A(t,t_f)\Lambda^T(t)B(t)\right]$$

with boundary condition $E(t_f) = 0$. Thus

$$E(t_0) = -\int_{t_0}^{t_f} \Phi_{\underline{A}}(t_0,t)\left[B(t)\Lambda(t)W_A(t,t_f) + W_A(t,t_f)\Lambda^T(t)B(t)\right]\Phi_{\underline{A}}^T(t_0,t)\mathrm{d}t.$$

Therefore

$x_0^T E(t_0)x_0 =$

$$-\int_{t_0}^{t_f} x_0^T\Phi_{\underline{A}}(t_0,t)\left[B(t)\Lambda(t)W_A(t,t_f) + W_A(t,t_f)\Lambda^T(t)B(t)\right]\Phi_{\underline{A}}^T(t_0,t)x_0\mathrm{d}t$$

and since we have assumed that $x_0^T\Phi_{\underline{A}}(t_0,t)B(t) = 0$, it implies that $x_0^T E(t_0)x_0 = 0$. Hence (6.12) becomes

$$x_0^T E(t_0)x_0 = x_0^T\left[W_{\underline{A}}(t_0,t_f) - W_A(t_0,t_f)\right]x_0 = 0.$$

Now, $x_0^T W_{\underline{A}}(t_0,t_f)x_0$ will go to zero given that $x_0^T\Phi_{\underline{A}}(t_0,t)B(t) = 0$ implying that

$$x_0^T E(t_0)x_0 = x_0^T\left[-W_A(t_0,t_f)\right]x_0 = 0.$$

Thus $x_0^T W_A(t_0,t_f)x_0 = 0$ implying that $x_0^T\Phi_A(t_0,t)B(t) = 0$ which is a contradiction hence $x_0^T W_{\underline{A}}(t_0,t_f)x_0 \neq 0$ which ensures the controllability of the state with dynamic system $\dot{x}(t) = \underline{A}(t)x(t) + B(t)v(t)$ with respect to $v(t)$. $\square$

Theorem 6.1.0.10 (Sp). *Let $R(t) > 0$ and bounded for all $t \in [t_0,t_f]$ and let the dynamic system of (5.2) be controllable. Assume further that $S_f \geq 0$ and $Q(t) - C^T(t)R^{-1}(t)C(t) \geq 0$ for all $t \in [t_0,t_f]$. Then, if the Riccati variable, $S(t_f,t;S_f)$, exists it is nonnegative definite for all $t \in [t_0,t_f]$ regardless of t_0.*

Proof. Suppose that the Riccati variable, $S(t_f,t;S_f)$, exists. Then we have seen from the proof of Theorem 6.0.2.7 that the minimum cost is related to the Riccati variable for some initial state as

$\frac{1}{2}x_0^T S(t_f,t_0;S_f)x_0$

$$= \min_{u(\cdot)}\left\{\frac{1}{2}\left[x^T(t_f)S_f x(t_f)\right] + \frac{1}{2}\int_{t_0}^{t_f}\left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right]\mathrm{d}t\right\}.$$

Let us make the substitution $u(t) = v(t) - R^{-1}(t)C(t)x(t)$. Then,

$\frac{1}{2}x_0^T S(t_f, t_0; S_f)x_0$

$$\begin{aligned} &= \min_{u(\cdot)} \left\{ \frac{1}{2} \left[x^T(t_f)S_f x(t_f)\right] \right. \\ &\qquad \left. + \frac{1}{2}\int_{t_0}^{t_f} \left[x^T(t)\left[Q(t) - C^T(t)R^{-1}(t)C(t)\right]x(t) + v^T(t)R(t)v(t)\right] \mathrm{d}t \right\} \\ &\geq 0 \end{aligned}$$

since $S_f \geq 0$ and $Q(t) - C^T(t)R^{-1}(t)C(t) \geq 0$ and $R(t) > 0$ by the hypothesis of the theorem. Therefore

$$\frac{1}{2}x_0^T S(t_f, t_0; S_f)x_0 \geq 0 \Longrightarrow S(t_f, t_0; S_f) \geq 0.$$

Now, since the state x is controllable with respect to $u(t)$, it implies then by Theorem 6.1.0.9 that the system $\dot{x}(t) = (A(t) - B(t)R^{-1}(t)C(t))x(t) + B(t)v(t)$ must be controllable with respect to $v(t)$. Hence the cost is bounded from above and below even as $t_0 \longrightarrow -\infty$. □

Theorem 6.1.0.11 (**Monotonicity of the Riccati Variable** (Sp)). *Given the hypothesis of Theorem 6.0.2.2 and letting $S_f = 0$ then, $S(t_f, t_0; 0) \geq S(t_1, t_0; 0), \quad t_0 \leq t_1 \leq t_f$.*

Proof. Let $S_f = 0$. Then, $\frac{1}{2}\left[x^T(t_f)S_f x(t_f)\right] = 0$. Therefore $\frac{1}{2}x_0^T S(t_f, t_0; S_f)x_0$ will be

$$\begin{aligned} &\min_{u(\cdot)} \left\{ \frac{1}{2}\int_{t_0}^{t_1} \left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right] \mathrm{d}t \right. \\ &\qquad \left. + \frac{1}{2}\int_{t_1}^{t_f} \left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right] \mathrm{d}t \right\}. \end{aligned}$$

Hence, $\frac{1}{2}x_0^T S(t_f, t_0; S_f)x_0$

$$\begin{aligned} &= \min_{u(\cdot)} \left\{ \frac{1}{2}\int_{t_0}^{t_1} \left[x^T(t)Q(t)x(t) + 2u^T(t)C(t)x(t) + u^T(t)R(t)u(t)\right] \mathrm{d}t \right. \\ &\qquad \left. + \frac{1}{2}x^T(t_1)S(t_f, t_1; 0)x(t_1) \right\} \\ &= \min_{u(\cdot)} \left\{ \frac{1}{2}x^T(t_0)S(t_1, t_0; 0)x(t_0) + \frac{1}{2}x^T(t_1)S(t_f, t_1; 0)x(t_1) \right\} \\ &= \frac{1}{2}x^T(t_0)S(t_1, t_0; S(t_f, t_1; 0))x(t_0) \end{aligned}$$

i.e. the solution from t_0 to t_f is the composition from t_0 to t_1 and t_1 to t_f. Therefore $\frac{1}{2}x_0^T S(t_f, t_0; S_f)x_0 \geq \frac{1}{2}x_0^T S(t_1, t_0; S_f)x_0$ and thus the result follows. □

Chapter 7

Application to Disease Model

As an illustration, we consider an SIR model with variable size population and we formulate an optimal control problem subject to these dynamics with vaccination and treatment, $u_1(t)$ and $u_2(t)$ respectively as controls and the state variables, $x(t), y(t)$, and $z(t)$ being the susceptibles (S), the infected (I), and the recovered (R) respectively such that

$$x(t) + y(t) + z(t) = N(t)$$

where $N(t)$ is the total population at time t. Our aim is to find the optimal combinations of vaccination and treatment strategies that will minimize the cost function of the two control measures as well as the number of recoveries.

Consider a variant of the SIR model below (it must be emphasised that this variant of the SIR is chosen for illustration sake).

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \alpha x - bu_1\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \beta x - au_2\\ \frac{\mathrm{d}z}{\mathrm{d}t} &= \gamma y + cu_2\end{aligned}$$

where x is the number of susceptible, y is the number of infected and z is the number of recovered and thus in matrix form we have

$$\begin{pmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{pmatrix} = \begin{pmatrix}\alpha & 0 & 0\\ \beta & 0 & 0\\ 0 & \gamma & 0\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix} + \begin{pmatrix}-b & 0\\ 0 & -a\\ 0 & c\end{pmatrix}\begin{pmatrix}u_1\\ \\ u_2\end{pmatrix}.$$

We can write this in compact form as

$$\dot{X}(t) = A(t)X(t) + B(t)u(t)$$

where $A(t)$ and $B(t)$ are the constant matrices

$$\begin{pmatrix} \alpha & 0 & 0 \\ \beta & 0 & 0 \\ 0 & \gamma & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -b & 0 \\ 0 & -a \\ 0 & c \end{pmatrix}$$

respectively and $X(t) \in \mathbb{R}^3$ and $u(t) \in \mathbb{R}^2$ are $(x, y, z)^T$ and $(u_1, u_2)^T$ respectively. By Theorem 3.2.1.7, this linear system is controllable since the rank of the following 3×6 matrix is 3.

$$\begin{bmatrix} -b & 0 & -\alpha b & 0 & -\alpha^2 b & 0 \\ 0 & -a & -\beta b & 0 & -\alpha\beta b & 0 \\ 0 & c & 0 & -\gamma a & -\gamma\beta b & 0 \end{bmatrix}.$$

We define the cost, objective, function as

$$J(\cdot) = \frac{1}{2} z^2(t_f) + \frac{1}{2} \int_0^{t_f} c_1 u_1^2 + c_2 u_2^2 \mathrm{d}t$$

such that $u_1, u_2 \in U$, where

$$U = \{(u_1(t), u_2(t)) : 0 \leq u_i(t) \leq 1, i = 1, 2 \quad \text{and} \quad t \in [0, t_f]\}$$

and c_1, c_2 are the relative weights attached to the cost of vaccination and treatment respectively.

The goal, then, is to characterize an optimal control $(u_1^*, u_2^*) \in U$ which minimizes the cost of the vaccination and the cost of the treatment over the specified time interval as well as the number of recovered at terminal time (constant terminal time). Hence, the control problem is;

minimize

$$J(\cdot) = \frac{1}{2} z^2(t_f) + \frac{1}{2} \int_0^{t_f} c_1 u_1^2 + c_2 u_2^2 \mathrm{d}t$$

subject to

$$\dot{X}(t) = A(t)X(t) + B(t)u(t)$$

with an appropriate state initial condition which in this example we choose to be $(x(0), y(0), z(0))^T = (\tau N_0, \tau_1 N_0, 0)^T$ where $0 < \tau, \tau_1 \leq 1$ such that $\tau + \tau_1 = 1$ and $N_0 = N(0)$. We proceed as follows.

The cost function can be rewritten as

$$J(\cdot) = \frac{1}{2}\begin{pmatrix} x & y & z \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}\Bigg|_{t=t_f} + \frac{1}{2}\int_0^{t_f}\begin{pmatrix} u_1 & u_2 \end{pmatrix}\begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\mathrm{d}t.$$

Comparing this with (5.1), we see that $Q(t) = 0, C(t) = 0$ and S_f and $R(t)$ are

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$$

respectively which are constant symmetric matrices. Now, from the first order necessary condition for optimality, Theorem 5.0.0.6, we have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha x - bu_1$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = \beta x - au_2$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = \gamma y + cu_2$$

subject to $(x(0), y(0), z(0))^T = (\tau N_0, \tau_1 N_0, 0)^T$.

$$-\begin{pmatrix} \dot{\lambda}_1 \\ \dot{\lambda}_2 \\ \dot{\lambda}_3 \end{pmatrix} = \begin{pmatrix} \alpha & 0 & 0 \\ \beta & 0 & 0 \\ 0 & \gamma & 0 \end{pmatrix}^T \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}$$

subject to

$$\begin{pmatrix} \lambda_1(t_f) \\ \lambda_2(t_f) \\ \lambda_3(t_f) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x(t_f) \\ y(t_f) \\ z(t_f) \end{pmatrix}$$

and since there are no terminal constraints, $D = 0$, and so

$$\begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} -b & 0 \\ 0 & -a \\ 0 & c \end{pmatrix}^T \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = 0.$$

Multiplying these matrices out results in the system of equations

$$
\begin{aligned}
\dot{x} &= \alpha x - bu_1 \\
\dot{y} &= \beta x - au_2 \\
\dot{z} &= \gamma y + cu_2 \\
-\dot{\lambda}_1 &= \alpha\lambda_1 + \beta\lambda_2 \\
-\dot{\lambda}_2 &= \gamma\lambda_3 \\
-\dot{\lambda}_3 &= 0 \\
0 &= c_1u_1 - b\lambda_1 \\
0 &= c_2u_2 - a\lambda_2 + c\lambda_3
\end{aligned}
$$

subject to

$$x(0) = \tau N_0, y(0) = \tau_1 N_0, z(0) = 0, \lambda_1(t_f) = 0, \lambda_2(t_f) = 0, \lambda_3(t_f) = z(t_f)$$

and the Hamiltonian is given as

$$
\begin{aligned}
H(\cdot) &= \frac{1}{2}u^T R(t)u + \lambda^T[A(t)X(t) + B(t)u(t)] \\
&= \frac{1}{2}(c_1u_1^2 + c_2u_2^2) + \alpha\lambda_1 x - b\lambda_1 u_1 + \beta\lambda_2 x - a\lambda_2 u_2 + \gamma\lambda_3 y + \lambda_3 c u_2.
\end{aligned}
$$

Solving for the unknown quantities in this system, we obtain

$$
\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} \frac{z(t_f)\beta\gamma}{\alpha}(t - t_f) \\ z(t_f)\gamma(t_f - t) \\ z(t_f) \end{pmatrix}
$$

$$
\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \frac{z(t_f)\beta\gamma b}{\alpha c_1}(t - t_f) \\ \frac{z(t_f)}{c_2}\left[a\gamma(t_f - t) - c\right] \end{pmatrix}
$$

$$
\begin{pmatrix} x \\ y \\ z \end{pmatrix} =
$$

$$\begin{pmatrix} \frac{kt}{\alpha} + \tau N_0 \\ \left[\frac{\beta k}{2\alpha} + \frac{z(t_f)a^2\gamma}{c_2}\right]\frac{t^2}{2} + \left[\beta\tau N_0 - \frac{z(t_f)a^2\gamma t_f}{c_2} + \frac{z(t_f)ac}{c_2}\right]t + \tau_1 N_0 \\ \left[\frac{\beta k}{2\alpha} + \frac{z(t)a^2\gamma}{c_2}\right]\frac{\gamma t^3}{6} + \left[\beta\tau N_0 - \frac{z(t_f)a^2\gamma t_f}{c_2}\right]\frac{\gamma t^2}{2} + (a\gamma t_f - c)\frac{z(t_f)ct}{c_2} + \gamma\tau_1 N_0 t \end{pmatrix},$$

where

$z(t_f) =$

$$\frac{6\beta\tau\alpha^2 c_1 c_2 N_0 t_f^2 + 12\tau_1\gamma c_1 c_2 \alpha^2 N_0 t_f}{12c_1c_2\alpha^2 + 6c_1\alpha^2 a^2\gamma^2 t_f^3 - \beta^2\gamma^2 b^2 c_2 t_f^3 - 2a^2\gamma^2 c_1\alpha^2 t_f^3 - 12c_1\alpha^2(a\gamma t_f - c)t_f}$$

and

$$k = \frac{z(t_f)\beta\gamma b^2}{c_1\alpha}.$$

To check whether (u_1, u_2) gives the optimal solution, we consider the Riccati matrix differential equation. Let the Riccati variable be

$$S(t) = \begin{pmatrix} s_1 & s_2 & s_3 \\ s_4 & s_5 & s_6 \\ s_7 & s_8 & s_9 \end{pmatrix}.$$

Using (6.7), identifying the various matrices and multiplying them out results in the the following nonlinear system of differential equations (a quadratic matrix differential equation) which is the Riccati matrix differential equation.

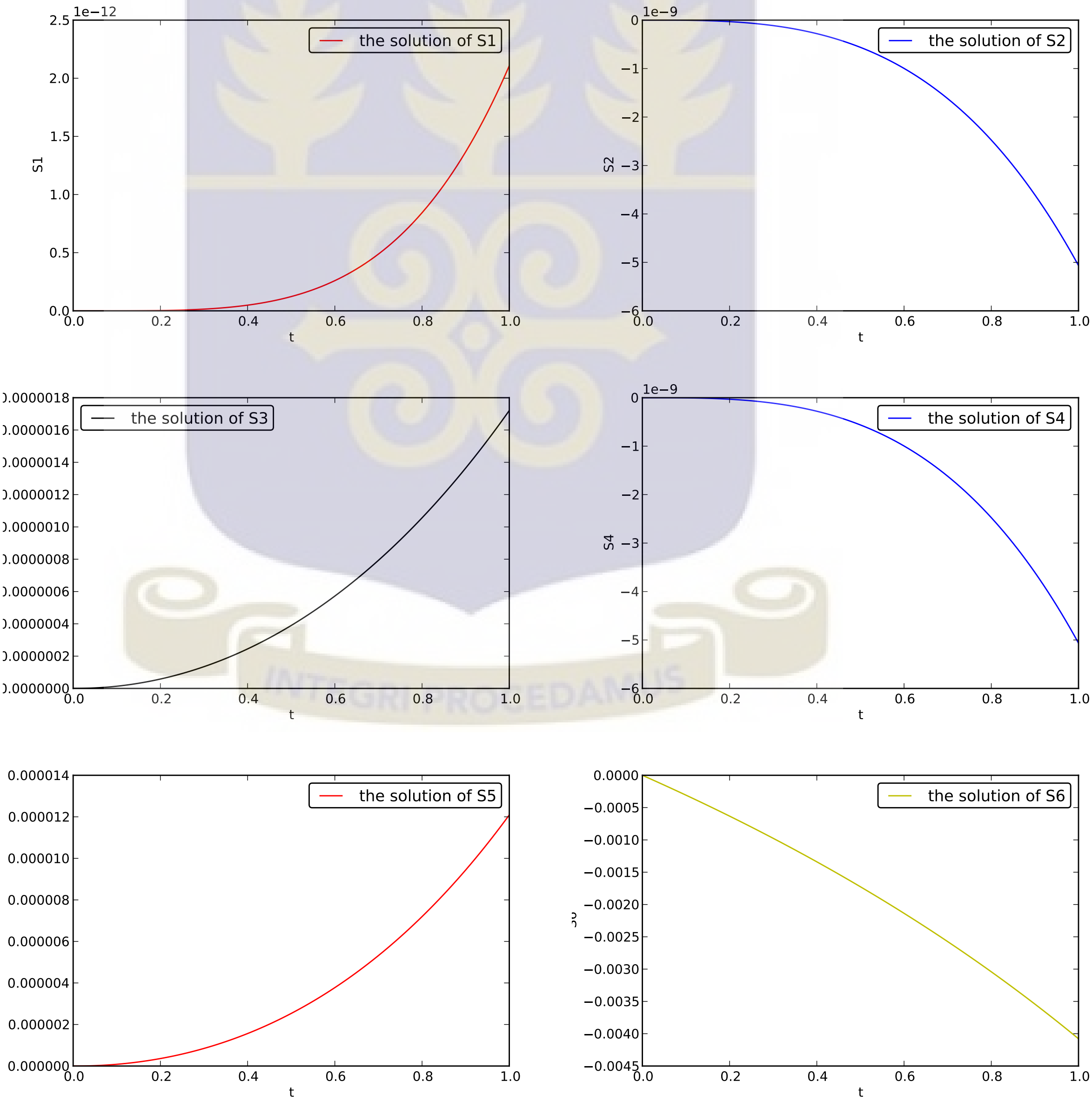
$$\begin{aligned}
\dot{s_1} &= \frac{1}{c_1c_2}(c_2b^2s_1^2 + c_1a^2s_2s_4 - c_1cas_3s_4 + c_1c^2s_3s_7 - c_1cas_2s_7) \\
&\quad -2\alpha s_1 - \beta s_4 - \beta s_2 \\
\dot{s_2} &= \frac{1}{c_1c_2}(c_2b^2s_1s_2 + c_1a^2s_2s_5 - c_1cas_3s_5 + c_1c^2s_3s_8 - c_1cas_2s_8) \\
&\quad -\gamma s_3 - \alpha s_2 - \beta s_5 \\
\dot{s_3} &= \frac{1}{c_1c_2}(c_2b^2s_1s_3 + c_1a^2s_2s_6 - c_1cas_3s_6 + c_1c^2s_3s_9 - c_1cas_2s_9) \\
&\quad -\alpha s_3 - \beta s_6 \\
\dot{s_4} &= \frac{1}{c_1c_2}(c_2b^2s_4s_1 + c_1a^2s_5s_4 - c_1cas_6s_4 + c_1c^2s_6s_7 - c_1cas_5s_7) \\
&\quad -\alpha s_4 - \beta s_5 - \gamma s_7 \\
\dot{s_5} &= \frac{1}{c_1c_2}(c_2b^2s_4s_2 + c_1a^2s_5^2 - c_1cas_6s_5 + c_1c^2s_6s_8 - c_1cas_5s_8) \\
&\quad -\gamma s_6 - \gamma s_8 \\
\dot{s_6} &= \frac{1}{c_1c_2}(c_2b^2s_4s_3 + c_1a^2s_5s_6 - c_1cas_6^2 + c_1c^2s_6s_9 - c_1cas_5s_9) \\
&\quad -\gamma s_9
\end{aligned}$$

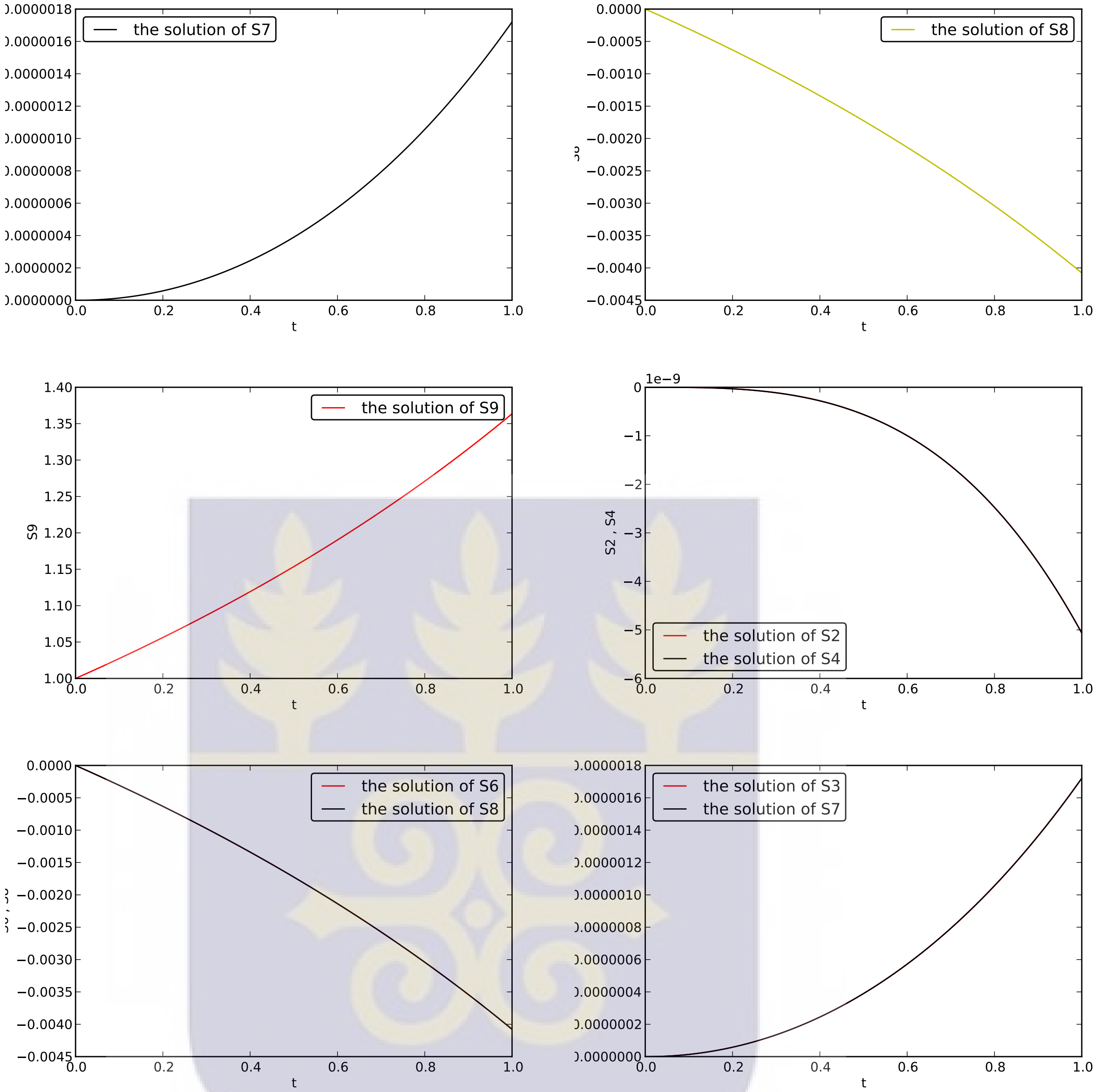
$$
\begin{aligned}
\dot{s_7} &= \frac{1}{c_1c_2}(c_2b^2s_7s_1 + c_1a^2s_8s_4 - c_1cas_9s_4 + c_1c^2s_9s_7 - c_1cas_8s_7) \\
&\quad -\alpha s_7 - \beta s_8 \\
\dot{s_8} &= \frac{1}{c_1c_2}(c_2b^2s_7s_2 + c_1a^2s_8s_5 - c_1cas_9s_5 + c_1c^2s_9s_8 - c_1cas_8^2) \\
&\quad -\gamma s_9 \\
\dot{s_9} &= \frac{1}{c_1c_1}(c_2b^2s_7s_3 + c_1a^2s_8s_6 - c_1cas_9s_6 + c_1c^2s_9^2 - c_1cas_8s_9)
\end{aligned}
$$

subject to the boundary conditions

$$
S_f = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
$$

We solve this numerically to obtain;





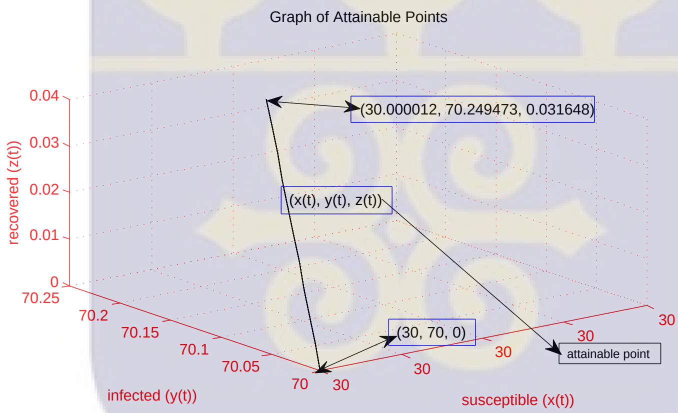
These solutions are obtained for $t_f = 1, c_1 = 0.4, c_2 = 0.6, b = 0.8, \beta = 0.001, \gamma = 0.003, \alpha = 0.5, a = 0.4, \tau_1 = 0.7, \tau = 0.3, N_0 = 100, c = 0.4$. The solution is symmetric since $s_2 = s_4$, $s_6 = s_8$ and $s_3 = s_7$ (evident from the diagrams) and is also justified by the fact that S_f and $R(t)$ are symmetric matrices. Since the solution exists for the Riccati matrix differential equation, the necessary and sufficient condition are satisfied, by Theorem 6.0.2.7, hence the cost criterion will be positive and strongly positive and

$$\begin{pmatrix} u_1 \\ \\ u_2 \end{pmatrix} = \begin{pmatrix} \frac{z(t_f)\beta\gamma b}{\alpha c_1}(t - t_f) \\ \\ \frac{z(t_f)}{c_2}\left[a\gamma(t_f - t) - c\right] \end{pmatrix},$$

is the optimal solution for the control problem i.e. for the cost function to be minimized (by Theorem 6.0.2.7) where u_1 and u_2 are the respective vaccination and treatment functions that minimizes the cost criterion. We see that u_1, u_2 depends on the entries of the matrices A and B. Substituting u_1, u_2 into the cost function, and performing the integration, we obtain

$$J(\cdot) = 0.159z^2(t_f) + 0.5z(t_f)$$

to be the minimum cost. The following diagram indicates the path in which the optimal controller, $u(t) = (u_1, u_2)$, steers the response, (x, y, z) from the initial point $(x_0, y_0, z_0) = (30, 70, 0)$ to the final point $(x(1), y(1), z(1)) = (30.000012, 70.249473, 0.031648)$ where the final point is an attainable point.



7.1 Conclusion

Two main theorems have been considered in this work; first order necessary condition for optimality and the strong positive definiteness of the second variation for the linear control problem with unconstrained terminal condition. Theorem 5.0.0.6 explains the first order necessary condition. The positive definiteness of the second variation, Theorem 6.0.2.7, explains the sufficient condition for optimality where we saw that the sufficient condition for optimality depends on the existence of solution to a Riccati matrix differential equation, (6.7). Using these ideas, we have been able to find the optimal solution to the linear quadratic control problem considered in this chapter.

Throughout this work, we have made a crucial assumption that the general Legendre - Clebsch condition, $H_{uu} = R(t) > 0$, holds. However, there are situations whereby this assumption fails. For instance, suppose $H_{uu} = 0$. Then, $R^{-1}(t)$ does not exist and thus the general optimal control rule can not be formulated as it has been done here. Such a situation comes under the umbrella of *Singular Problems* in control theory [DD]. They thus require a different strategy to that used in this work. Singular problems in control theory are thus worth studying in the future.

References

[AM] Anderson B. D. O. , Moore J. B., (1989), *Linear Quadratic Methods*, Prentice-Hall International, Inc.

[DD] Bell J. David and Jaccobson H. David H., (1975), *Singular Optimal Control Problems*, Academic Press, New York.

[Br1] Bellman R., (1967),*Introduction to Mathematical Theory of Control Processes*, Volume I, Linear Equations and Quadratic Criteria, New York and London, Accademic Press.

[Ro] Brocket R. W., (1970), *Finite Dimensional Linear Systems*, New York, John Wiley and Sons, Inc.

[BY] Bryson A. E. Jr., Yu-Chi Ho, (1975), Applied Optimal Control, London, Taylor and Francis.

[GS] Gelfand I. M. , Fomin V. S. (1963), *Calculus of Variation*, New Jersey, Prentice-Hall, Inc, Englewood Cliffs.

[GM] Gernett B., Saunders M. (1977), *A Survey of Modern Algebra*, New York, Macmillan Publishing Co. Inc.

[Le] Markus L. , Lee E.B., (1986), *Foundations of Optimal Control Theory*, Malabar, Florida, John Wily and Sons, Inc.

[LS] Pontryagin L. S., Boltyanski V. G., Gamkrelidze R. V. and Mishchenko E. F., (1962),*The Mathematical Theory of Optimal Processes*, Translated and Edited by K. N. Trirogoff and L. W. Neustadt, New York, Interscience Publisher.

[Sp] Speyer J. L. , Jacobson D. H. (2010), *Primer on Optimal Control Theory*, Philadelphia, Society of Industrial and Applied Mathematics.

[MU] Michael Baake and Ulrike Schlägel, (2012), *The Peano-Baker Series*, arXiv:math.CA/1011.1775v2.

Appendix A

Background

A.1 Liebniz's Rule

Liebniz's integral rule gives a formula for differentiation of a definite integral whose limits are functions of differentiable functions. Let $f(x(t), t)$ be continuous and suppose $f_t(x(t), t)$ exists and is continuous. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{a(t)}^{b(t)} f(x,t)\mathrm{d}t = \int_{a(t)}^{b(t)} \frac{\partial f}{\partial t}\mathrm{d}t + f(b(t),t)\frac{\mathrm{d}b}{\mathrm{d}t} - f(a(t),t)\frac{\mathrm{d}a}{\mathrm{d}t}.$$

A.2 Hamilton's Principles

Hamiltonian system is a system of variables which can be written in the form of Hamilton's equations;

$$\begin{aligned}\frac{\mathrm{d}q}{\mathrm{d}t} &= \frac{\partial H}{\partial p}\\ \frac{\mathrm{d}p}{\mathrm{d}t} &= -\frac{\partial H}{\partial q}\end{aligned}$$

where $H = H(q, p, t)$ is the Hamiltonian.

Example 1. Recall that $H(x, u, \lambda, t) = L(x, u, t) + \lambda^T f(x, u, t)$ where $\dot{x}(t) = f(x, u, t)$. Now, $H_\lambda(x, u, \lambda, t) = f(x, u, t) = \dot{x}$. Also, recall, from Theorem 5.0.0.6, that $H_x(x, u, \lambda, t) = -\dot{\lambda}^T$ thus $H(x, u, \lambda, t)$ defined above qualifies to be Hamiltonian where in this case, $q = x$ and $p = \lambda$.

A.3 Hamilton-Jacobi-Bellman Equation

Let $V : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be continuously differentiable. Then the equation

$$-\frac{\partial V}{\partial t} = \min_{u \in U} \left[L(x, u, t) + \frac{\partial V}{\partial x} \cdot f(x, u, t) \right]$$

where $V(x(t_f), t_f) = \phi(x(t_f), t_f)$ is the Hamilton-Jacobi-Bellman Equation. V is called the *optimal value function.* See [Sp] for details.

A.4 Cayley-Hamilton Theorem

Let A be a square matrix and let $p(t) = \det(A - tI)$ be the characteristic polynomial of A. Then $p(A) = 0$. (GM p. 341)

A.5 Some Basic Definitions

Definition A.5.0.12 (Hyperplane)**.** Let E be a vector space. A Hyperplane in E is any set of the form

$$H = \{x \in E : f(x) = \alpha, \quad \alpha \in \mathbb{R}\}$$

where $f \neq 0$ is a linear functional on E.

Definition A.5.0.13 (Positive Definite and Semi-definite Matrices)**.** A positive definite matrix is a matrix with all its eigenvalues positive. It is semi-definite if at least one of its eigenvalues is zero.

Theorem A.5.0.14 (Taylor Theorem for Function of Vector Argument)**.** *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be n-times continuously differentiable in all of its arguments at a point x_0. Consider $f(x_0 + \epsilon h)$ where $\|h\| = 1$ (it is not important which specific norm is used). Then,*

$$f(x_0 + \epsilon h) = f(x_0) + \epsilon \frac{\partial f}{\partial x}\bigg|_{x_0} h + \frac{\epsilon^2}{2!} h^T \frac{\partial^2 f}{\partial x^2}\bigg|_{x_0} h + \cdots + R_n$$

where

$$\lim_{\epsilon \to 0} \frac{R_n}{\epsilon^n} \longrightarrow 0.$$

$\frac{\partial f}{\partial x}$ *and* $\frac{\partial^2 f}{\partial x^2}$ *are the gradient vector and the Hessian matrix respectively [Sp].*

Definition A.5.0.15 (Dynamical System)**.** A smooth dynamical system on $\mathbb{R}^n$ is a continuously differentiable function $\phi : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$ where $\phi(t, x) = \phi_t(x)$ satisfies

1. $\phi_0 : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is the identity function; $\phi_0(x_0) = x_0$

2. The composition $\phi_t \circ \phi_s = \phi_{t+s}$ for all $t, s \in \mathbb{R}$.

Definition A.5.0.16. 1. L_2-space is the set of square integrable functions.

2. Hilbert space is a complete inner product space.

3. If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$, there is a unique linear map $\alpha : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{\|f(a+h) - f(a) - \alpha(h)\|}{\|h\|} = 0.$$

Theorem A.5.0.17. *The product of all the eigenvalues of an $n \times n$ matrix, A, is the determinant of A.*

Proof. Let $t_1, t_2, ..., t_n$ be the eigenvalues of A. Then the t_i's are the roots of the characteristic polynomial

$$p(t) = \det(A - tI) = (t_1 - t)(t_2 - t) \cdots (t_n - t).$$

Since t is arbitrary, we can let $t = 0$. then

$$p(t) = \det(A) = t_1 \cdot t_2 \cdots t_n.$$

□

Example 2 (Strongly Positive Functional)**.** Consider the functional $J(u(\cdot)) = \int_0^{t_f} x^2 \mathrm{d}t$ where $\dot{x} = u; x_0 = 0$ and $u(\cdot) \in U$. Let $u = \cos \omega t, \forall t \in [0, t_f]$. Then, $x(t) = \frac{\sin \omega t}{\omega}$ and $\|u(\cdot)\|^2 = \int_0^{t_f} \cos^2 \omega \mathrm{d}t$. Then,

$$\begin{aligned} J(u(t)) - k\|u(\cdot)\| &= \left(\frac{1}{\omega^2} + k\right) \int_0^{t_f} \sin^2 \omega t \mathrm{d}t - k \int_0^{t_f} 1 \mathrm{d}t \\ &= \left(\frac{1}{\omega^2} + k\right) \int_0^{t_f} \sin^2 \omega t \mathrm{d}t - k t_f \\ &= \frac{1}{2}\left(\frac{1}{\omega^2} + k\right)\left(\frac{-\sin 2\omega t_f}{2\omega}\right) + \frac{t_f}{2\omega^2} - \frac{1}{2} k t_f. \end{aligned}$$

Taking the limit as $\omega \to \infty$, we have that $-\frac{1}{2} k t_f \leq 0$ for $k > 0$. Hence $J(u(\cdot))$ is not strongly positive.

Remark A.5.0.18. Consider $\dot{x} = Ax(t)$. Since A is a constant, the transition matrix for this system is given as $\Phi(t, t_0) = \mathrm{e}^{A(t-t_0)}$ and we can establish this using the Peano-Baker series.

Definition A.5.0.19. A sequence of functions, $\{f_n\}_{n=1}^{\infty}$, is said to be uniformly convergent to f of for a set E of values of x, if for each $\epsilon > 0$, an integer N can be found such that $|f_n(x) - f(x)| < \epsilon$ for $n \geq N$ for all $x \in E$.

A.6 Weierstrass M-Test

Let $\sum_{n=1}^{\infty} u_n(x)$ be a series of functions defined on set E of values of x. If there is a convergent series of constants, $\sum_{n=1}^{\infty} M_n$, such that $|u_n(x)| \leq M_n$ for all $x \in E$, then the series exhibits absolute convergence for all $x \in E$ as well as uniforn convergence in E.

Lemma A.6.0.20 (MU). *If A is continuous on an interval, say I, such that the matrix $A(t)$ commute with the intergral $\int_{t_0}^{t} A(\sigma)\mathrm{d}\sigma$, for all $t \in I$, then*

$$I_k(t) := \int_{t_0}^{t} A(\sigma_1) \int_{t_0}^{\sigma_1} A(\sigma_2) \cdots \int_{t_0}^{\sigma_{k-1}} A(\sigma_k)\mathrm{d}\sigma_k \cdots \mathrm{d}\sigma_2 \mathrm{d}\sigma_1 = \frac{\gamma^k(t)}{k!}$$

where $\gamma(t) = \int_{t_0}^{t} A(\sigma)\mathrm{d}\sigma$ for all $t \in I$ and $k \in \mathbb{N}_0$.

Proof. W will prove this by induction. Let $I_k(t)$ be defined as above then by construction, we will have that

$$I_{k+1} = \int_{t_0}^{t} A(\sigma) I_k(\sigma)\mathrm{d}\sigma.$$

Now for $k = 0$, I_0 coincides with $M_0 = I$ and for $k = 1$, $I_1 = \int_{t_0}^{t} A(\sigma)\mathrm{d}\sigma$. Thus it is true for $k = 0, 1$. Assuming this is true for all k, then from the above equation,

$$I_{k+1} = \int_{t_0}^{t} A(\sigma)\frac{\gamma^k(\sigma)}{k!}\mathrm{d}\sigma.$$

Now, by the chain rule and the Fundamental Theorem of Calculus,

$$\frac{1}{(k+1)!}\frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma^{k+1}(\sigma) = \frac{A(\sigma)}{k!}\gamma^k(\sigma).$$

Therefore,

$$I_{k+1} = \frac{1}{(k+1)!}\int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma^{k+1}(\sigma)\mathrm{d}\sigma$$

which implies that $I_{k+1}(t) = \frac{\gamma^{k+1}(t)}{(k+1)!}$ as required. □