Department of Mathematics and Statistics

The Minimum Energy Problem for Positive Linear Systems

Siti Chotijah

This thesis is presented for the Degree of
Doctor Philosophy
of
Curtin University

February 2017

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: ………………………………………….

Date: 3 February 2017

Abstract

In this thesis, we consider the minimum energy problem for the positive linear discrete-time system and the positive linear continuous-time system with fixed initial and final states. The problem has analytic solutions if no restrictions are imposed on the state and control variables. The nonnegativity of control in such systems gives rise to complementarity conditions in the first-order Karush-Kuhn-Tucker optimality conditions which complicates the problems and hence usually only numerical solutions instead of analytical solutions can be obtained.

The minimum energy optimal control problem that we firstly consider are for scalar control. Then, we consider the minimum energy optimal control problem for vector control. The studies include the discrete-time system and continuous-time system. Sufficient conditions to guarantee the positivity of the problem and the analytical solutions to the problem have been obtained using dynamic programming. The relationship between the problem and the geometric properties of the system is discussed. Moreover, the optimal controls are obtained analytically in both forms: an open-loop form and a closed-loop form. To illustrate the main results of this work, some numerical examples are presented.

Finally, two applications on positive linear systems that are related to energy and ecology are studied. The first application is a dynamic model of oil extraction and its optimization. It is a novel dynamic (discrete-time) model that describes the evolution of the oil extraction process from a single well or reservoir under water flooding. Because of the nature of oil extraction process, the model exhibits positive linear system behaviour. The optimal control problem turns out to be a novel problem for the theory of positive linear systems problem. The second application is a continuous-time dynamic mobile source air pollution optimal control problem. Because of the nature of mobile source air pollution process, the model exhibits positive linear system behaviour.

List of publications

The following papers have been completed and published during PhD candidature:

S. Chotijah and V. Rumchev, The minimum energy problem for single-input single-output positive continuous-time linear systems with fixed initial and final state, International Conference on Computational Mathematics, Computational Geometry, and Statistics (CMCGS), Global Science and Technology Forum, p.^pp. 8, 2015.

R. Gupta, V. Rumchev and S. Chotijah, "A dynamic model of oil extraction and its optimization", International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management **1**, 765, 2015.

Acknowledgement

I am especially indebted to my supervisors Prof Yong Hong Wu, Dr. Ventsi Rumchev and A/Prof Guanglu Zhou for their guidance, patience, understanding and concern with the progress of my research. Dr Rumchev, in particular, I am thankful for his valuable supports, encouragement, suggestions throughout my study. His knowledge of Control Theory has encouraged me to study in this field. Prof Wu, meanwhile, deserve deep gratitude for giving me many valuable suggestions. Last but not least, A/Prof Zou, I am very grateful for all the computational and technical guidance to complete this thesis.

I am also grateful to to Prof Kok Lay Theo, the chairman of my thesis committee. His support and guidance since I was a master student has encouraged me to pursue PhD study.

I thank all the staff in Department of Mathematics and Statistics for all the assistance and contributing to a friendly working environment.

Finally, I sincere thanks to everyone in my family, especially my husband Abi Askari for all the support, understanding and his patience. I am also grateful to our daughter and son, Aisyah and Abdullah, who always cheers my days up. I am truly indebted to my Mum, Siti Fatimah, who constantly loves and supports me.

Notations

$\mathbb{R}$	Field of real numbers
$\mathbb{R}_+$	set of nonnegative real numbers
$\mathbf{I}$	identity matrix
$\mathbf{A}^T$	transpose of matrix **A**
$\mathbf{A}^{-1}$	Inverse of matrix **A**
$\mathbf{A} \geq 0$	**A** is a nonnegative matrix, i.e. every entries of A are nonnegative
$\mathbf{A} > 0$	**A** is a positive matrix, i.e. every entries of A are positive
rank (**A**)	rank of matrix **A**
$\mathbb{R}_+^n$	nonnegative orthant of $\mathbb{R}^n$, the space of column vectors of size n with nonnegative real entries.
$\mathbb{R}_+^{n\times m}$	set of $n \times m$ matrices with nonnegative entries
$\mathbf{x}(t)$	vector state trajectory
$\mathbf{x}^*(t)$	Optimal trajectory
$x(t)$	scalar state trajectory
$\mathbf{u}(t)$	vector control
$\mathbf{u}^*(t)$	Optimal control for general linear system
$\mathbf{u}_+^*(t)$	Optimal control for positive system
$u(t)$	scalar control
$\boldsymbol{\Re}_n$	n-steps reachability matrix of the pair $(\mathbf{A}, \mathbf{B})$
$\mathcal{R}_t(\mathbf{x}_0)$	t-steps reachable set of positive linear discrete-time system
$R_t(\mathbf{x}_0)$	t-steps reachable set of positive linear continuous-time system

Contents

CHAPTER 1

Introduction

1.1. Motivation and background

Optimal control theory is an interdisciplinary branch of mathematics aimed to find optimal ways for dynamic systems. An optimal control problem includes an objective function subject to constraints of the system.

A positive system is a linear state-space system in which the state variables remain nonnegative for all time for any nonnegative initial conditions. It is an important class of systems that arises in many areas such as engineering (chemical, manufacturing, telecommunications and information science), management science, social sciences, compartmental system analysis, biology, ecology, pharmacology and medicine, macro- and micro-economics and many other fields [1-13]. A common property of positive linear systems is that if the initial state and the control are nonnegative then the whole trajectory (evolution) of the system is nonnegative too. Positive linear systems are defined on cones instead of linear spaces and that is why the criteria developed in linear system theory for recognizing the fundamental properties of linear systems are quite often fallible. For example, controllability is a system property that characterizes ability of the system to move in space. If a dynamical system possesses the controllability property then for any pair of terminal (initial and final) states there exists a control sequence that transfers the system from the given initial to the given final state. Controllability property is a fundamental system property with direct implications in a number of control problems including optimal control. The controllability criteria for positive linear systems are different from those developed for linear systems [14-16]. The non-negativity of control in such systems gives rise to complementarity conditions in the first-order Karush-Kuhn-Tucker optimality conditions [17], and hence it usually only numerical solutions rather than analytic solutions can be derived. At the same time the appeal and the advantages of analytic solutions are well appreciated.

Linear quadratic optimal control problem is one of the most common optimal control problems. The objective function of this problem is a quadratic function subject to linear dynamic system constraints described by a set of linear differential equations. The studies of linear quadratic problems in general is quite well known, as studied in [18, 19]. One of the classical problem of linear quadratic optimal control is the minimum energy problem for time invariant linear systems. It has nice analytic solutions if no restrictions are imposed on the state and control variables [20, 18]. To the best of our knowledge the first analytical solution to the minimum energy problem for positive linear systems with fixed final state is obtained by Kaczorek [14] under strict conditions. This thesis is concerned with the minimum energy problem for positive linear (discrete-time and continuous-time) systems. Related work for linear quadratic optimal control for positive systems with free final state is published in [21-29] but the positivity of the system is not exploited in these papers, except in [21] and [25]. Conditions that guarantee the positivity of the closed–loop linear quadratic optimal system with free final state are developed for continuous-time systems by Labissi et al [25] and for discrete-time systems by Beauthier and Winkin [21]. Positivity is an intrinsic property of positive systems and it helps to simplify the analysis and the results. The optimality conditions and the solution for the linear quadratic problem for positive linear system in this thesis are obtained using dynamic programming.

Dynamic programming is an optimization method, which is widely used to solve a large-scale optimization problem. This method is developed by Richard Bellman in the late 1950s [30]. It is based on Bellman's *principle of optimality:*

> "An optimal property has the property that whatever the initial state and initial decision are, the remaining decision must constitute an optimal policy with regard to the state resulting from the first decision."

Dynamic programming basically is an algorithm that are used for optimization by reducing a complex problem into smaller sub problems.

Let the system be

$$x(t+1) = f\big(t, x(t), u(t)\big), \tag{1.1}$$

where f is vector of n functions, $u(t)$ is the control at time t and is a vector of size n, $x(t)$ is the state at time t and is a vector of size m.

Suppose the objective is to minimize the function

$$J_t(x(t)) = \phi(N, x(N)) + \sum_{t=i}^{N-1} L(t, x(t), u(t)), \tag{1.2}$$

where $[i, N]$ is the time interval of the problem, $\phi(N, x(N))$ is a function of final time N and $L(t, x(t), u(t))$ is generally, a time-varying function of u and x at each time t in $[i, N]$. Then, the Bellman equation is

$$J_t^{*}(x(t)) = \min_{u(t)}\{L(t, x(t), u(t)) + J_{t+1}^{*}(x(t+1))\}, \tag{1.3}$$

where $J_t^{*}(x(t))$ is the optimal cost from time t, and $u^{*}(t)$ is the optimal control at time t that achieves the minimum. The final cost is obtained by working backward from the final time N.

1.2. Objective

The main objective of this thesis is to establish sufficient that guarantee the nonnegativity of the state and the control variables conditions of the linear quadratic problem for positive systems. The particular positive linear quadratic problem discussed in this thesis is the minimum energy problem for the positive linear discrete-time system and the positive linear continuous-time system with fixed initial and final state. Moreover, analytical solutions to the optimal control and trajectory are to be obtained. Furthermore, investigation will be carried out to reveal the geometry and properties of the reachability sets.

1.3. Scope of the thesis

Some positive system properties related to positive systems will be discussed in chapter 2 including some useful matrix analysis. The properties discussed in this chapter includes reachability, controllability, reachable sets and stability for both positive linear discrete-time systems and continuous-time systems. In this chapter, some new preliminary results related to the positive linear system obtained during this work are presented.

Chapter 3 discusses the minimum energy problem for positive linear systems with fixed initial and final state in the case of scalar control. The studies include the discrete-time system and continuous-time system. Sufficient conditions are established to achieve the optimality conditions when the final state is given. The optimal controls are obtained analytically in an open-loop form and a feedback form. Then, numerical examples confirming the finding are given.

Chapter 4 discusses the minimum energy problem for positive linear system with vector control when the initial and the final states are given. The problems discussed include the discrete-time system and continuous-time system. Similar to the previous chapter, the optimal control is obtained analytically. Moreover, sufficient conditions are established to achieve the optimality conditions. Some numerical examples are given to confirm the analytic results. The numerical examples for the continuous-time systems are solved using both the results obtained in the thesis and ICLOCS, an optimal control software that can be used to solve the optimal control problem with some constraints [31], to show the effectiveness of the analytic results.

In chapter 5, we consider two applications on positive system that related to energy and ecology. The applications include a dynamic model of oil extraction and a dynamic mobile source air pollution optimal control problem. The study includes the optimization of the problem.

Finally, in chapter 6 we summarize the main contribution in this thesis and give some suggestions for future work.

CHAPTER 2

Positive Linear Systems: literature review and preliminary results

2.1. General

Positive linear systems are linear state-space systems in which the state variables remain nonnegative for all time for any nonnegative initial conditions. In this chapter we define various concepts and results that will be used in the following chapters. First, we recall some matrix properties that play important roles in positive system theory. Then we will discuss some basic knowledges and preliminary results in positive system theory. In this chapter, a number of preliminary results on the properties of positive linear system particularly on the reachable set properties are introduced.

2.2 Basic concepts in matrix analysis for positive systems

Definition 2.1.

a) A matrix $\mathbf{A} = \left(a_{ij}\right)_{1 \le i.j \le n} \in \mathbb{R}^{n \times n}$ is said to be nonnegative, denoted $\mathbf{A} \ge 0$ (in this thesis it is also denoted $\mathbf{A} \in \mathbb{R}_{+}^{n \times n}$), if for all $i, j = 1, \cdots, n, a_{ij} \ge 0$, i.e. every entries of **A** are nonnegative.
b) A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be a Metzler matrix, if every off-diagonal entries of **A** are non-negative, in other words, for all $i, j = 1, \cdots, n,\ i \ne j,\ a_{ij} \ge 0$.

Definition 2.2. Let **A** is an $n \times n$ matrix, the *matrix exponential* $e^{\mathbf{A}}$ is defined as

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} +$$

Proposition 2.1. [14] A matrix **A** is called a Metzler matrix if and only if

$$e^{\mathbf{A}t} \geq 0 \text{ for all } t \geq 0.$$

Definition 2.3. A square matrix is said to be a stable matrix if every eigenvalue of the matrix has negative real part.

Proposition 2.2. [32] If **A** is stable Metzler matrix, then $-\mathbf{A}^{-1}$ is a nonnegative.

Theorem 2.1. [33] **(Perron-Frobenius for Metzler matrices)** Let **A** be a Metzler matrix. Then there exists a real eigenvalue λ_F of **A**, called the Frobenius eigenvalue, such that there exists a positive eigenvector v,associated to λ_F, which is called the Frobenius eigenvector, such that $Av = \lambda_F v$ and $\forall\lambda \in \sigma(A), Re(\lambda) \leq \lambda_F$, where $\sigma(A)$ is the spectrum of matrix **A**.

Definition 2.4. A monomial matrix is an $n \times n$ matrix which has precisely one nonzero entry in each row and each column, and the remaining entries are zero.

Let **M** be a monomial matrix. It can be decomposed as $\mathbf{M} = \mathbf{DP}$, where **D** is a diagonal matrix and **P** is a permutation matrix.

Lemma 2.1. The inverse of the nonnegative monomial matrix is a nonnegative matrix [14].

Definition 2.5. [34] A non-singular square matrix **A** is said to be a M-matrix if the off diagonal elements of **A** are non-positive and its inverse is a nonnegative matrix.

The M-matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ can be expressed as

$$A = \begin{bmatrix} a_{11} & -a_{12} & -a_{13} & -a_{14} & \cdots & -a_{1n} \\ -a_{21} & a_{22} & -a_{23} & -a_{24} & \cdots & -a_{2n} \\ -a_{31} & -a_{32} & a_{33} & -a_{34} & \cdots & -a_{3n} \\ -a_{41} & -a_{42} & -a_{43} & a_{44} & \cdots & -a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & -a_{n3} & -a_{n4} & \cdots & a_{nn} \end{bmatrix}$$

The M-matrix A can also be expressed in the form

$$\mathbf{A} = s\mathbf{I} - \mathbf{B} \text{ for } s > 0 \text{ and } \mathbf{B} \geq 0 \text{ such that } s \geq \rho(\mathbf{B}), \text{ the spectral radius of } \mathbf{B} \text{ [6].}$$

Theorem 2.2. [6] Let $\mathbf{A}$ be a M-matrix. Then

a) All the principle minors of $\mathbf{A}$ are positive;
b) For each $x \neq 0$ there exists a nonnegative diagonal matrix $\mathbf{D}$ such that
$$\mathbf{x}^T\mathbf{ADx} > 0.$$
c) The inverse exists and $\mathbf{A}^{-1} \in \mathbb{R}_+^{n\times n}$.

Definition 2.6. Let $\mathbf{A} \in \mathbb{R}_{n\times n}$. $\mathbf{A}$ is called a Stieltjes matrix if $a_{i,j} \leq 0$ for $i \neq j$ and $\mathbf{A}$ is symmetric and positive definite.

Definition 2.7. The matrix $\mathbf{Q}$ is called positive semi-definite, if

$$\mathbf{x}^T\mathbf{Qx} \geq 0 \text{ for all nonzero (nontrivial) } \mathbf{x} \in \mathbb{R}^n \quad (2.1)$$

It is called positive definite if the inequality (2.1) is strictly positive, i.e. $\mathbf{Q}(\mathbf{x},\mathbf{x}) = \mathbf{x}^T\mathbf{Qx} > 0$.

Theorem 2.3. [35, 36]

a) The matrix $\mathbf{Q}$ is positive definite if and only if all principal minors of $\mathbf{Q}$ are strictly positive

$$\mathbf{D}_1 > 0, \mathbf{D}_2 > 0, \cdots, \mathbf{D}_n > 0.$$

This property is called Sylvester criterion.

b) The matrix $\mathbf{Q}$ is positive semi-definite if and only if
$$\mathbf{D}_1 \geq 0, \mathbf{D}_2 \geq 0, \cdots, \mathbf{D}_n \geq 0.$$
c) Let the $n \times n$ real matrix $\mathbf{Q}$ be positive semi-definite. Then $\mathbf{C}^T\mathbf{QC}$ is positive semi-definite for any real $n \times m$ matrix $\mathbf{C}$.
d) Let the $n \times n$ real matrix $\mathbf{Q}$ be positive definite and $\text{rank}(\mathbf{C}) = m$ then the matrix $\mathbf{C}^T\mathbf{QC}$ is positive definite too since $\mathbf{Cx} = 0$ for $\mathbf{x} = 0$ only. If $\text{rank}(\mathbf{C}) < m$, then $\mathbf{C}^T\mathbf{QC}$ is positive semi-definite even when $\mathbf{Q}$ is positive definite matrix.
e) If $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices of the same size, then $\mathbf{A} + \mathbf{B}$ is also a positive definite matrix.
f) If $\mathbf{A}$ and $\mathbf{B}$ are a positive definite matrix and a positive semi-definite of the same size respectively, then $\mathbf{A} + \mathbf{B}$ is also a positive definite matrix.

g) If $\mathbf{A}$ is a positive definite matrix, then $\mathbf{A}$ is invertible and $\mathbf{A}^{-1}$ is also a positive definite matrix.

Lemma 2.2. Let $\mathbf{A} \in \mathbb{R}_{m\times m}$ be a positive definite matrix. Define $\mathbf{Q}_n = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix}$, $\mathbf{Q}_n \in \mathbb{R}_{nm\times nm}$. Then, $\mathbf{Q}_n$ is a positive definite matrix.

Proof. $\mathbf{A}$ is positive definite if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for any vector $\mathbf{x} \in \mathbb{R}_m$. Define

$$\mathbf{Q}_n = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix}, \mathbf{Q}_n \in \mathbb{R}_{nm\times nm}$$

and

vector $\mathbf{y} \in \mathbb{R}_{nm}$, where $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$, $\mathbf{y}_i \in \mathbb{R}_m$, $i = 1, 2, \cdots, n$.

Hence,

$$\mathbf{y}^T\mathbf{Q}_n\mathbf{y} = [\mathbf{y}_1^T \quad \mathbf{y}_2^T \quad \cdots \quad \mathbf{y}_n^T] \begin{bmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \mathbf{y}_1^T\mathbf{A}\mathbf{y}_1 + \mathbf{y}_2^T\mathbf{A}\mathbf{y}_2 + \cdots + \mathbf{y}_n^T\mathbf{A}\mathbf{y}_n.$$

Since $\mathbf{A}$ is positive definite, $\mathbf{y}_i^T\mathbf{A}\mathbf{y}_i > 0$ and $\sum_{i=1}^{n}\mathbf{y}_i^T\mathbf{A}\mathbf{y}_i > 0$ for $i = 1, 2, \cdots, n$.

Therefore, $\mathbf{Q}_n$ is a positive definite matrix.

2.3 Positive Linear Discrete-time System

Consider a positive discrete-time linear system

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{2.2}$$

where $\mathbf{A} \in \mathbb{R}_+^{n\times n}$, $\mathbf{B} \in \mathbb{R}_+^{n\times m}$, $\mathbf{x}(t) \in \mathbb{R}_+^n$ is the state vector, $\mathbf{u}(t) \in \mathbb{R}_+^m$ is the control vector at time $t = 0,1,2, \cdots$

2.3.1 Controllability and Reachability

Controllability and reachability for positive linear discrete-time system has been studied by [37-40, 14, 41, 15]

Definition 2.8. [15]

a) The positive system (2.2) is *reachable* if for any state $\mathbf{x} \in \mathbb{R}_+^n$, $\mathbf{x} \neq \mathbf{0}$, and some finite t there exists a non-negative control sequence $\{\mathbf{u}(s)\}$, for $s = 0,1,2,\cdots,t-1$ that steers the initial state $\mathbf{x}(0)$ to the state $\mathbf{x} = \mathbf{x}(t)$.
b) The positive system (2.2) is *null controllable* if for any state $\mathbf{x} \in \mathbb{R}_+^n$ and some finite t there exists a nonnegative control sequence $\{\mathbf{u}(s)\}$, for $s = 0,1,2,\cdots,t-1$ that steers the state $\mathbf{x}(0)$ to the origin $\mathbf{x}(t) = 0$.
c) The positive system (2.2) is *controllable* if for any nonnegative pair $\{\mathbf{x}_0, \mathbf{x}\} \in \mathbb{R}_+^n$ and some finite t there exists a nonnegative control sequence $\{\mathbf{u}(s)\}$, for $s = 0,1,2,\cdots,t-1$ that steers the initial state $\mathbf{x}_0 = \mathbf{x}(0)$ to the state $\mathbf{x} = \mathbf{x}(t)$.

Define the n-step reachability matrix of the system

$$\mathfrak{R}_n = [\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{n-1}\mathbf{B}] \tag{2.3}$$

Theorem 2.4. [39] The positive system (2.2) is called

a) *reachable* if and only if the n-step reachability matrix of the system $\mathfrak{R}_n$ contains an $n \times n$ monomial submatrix. If the nonnegativity constraint is not considered, the reachability conditions is simply the Kalman criterion, i.e., rank($\mathfrak{R}_n$) $= n$ for not necessarily positive linear discrete-time systems.
b) *null controllable* if and only if $\mathbf{A}$ is a nilpotent matrix. A non-zero square matrix $\mathbf{A}$ is a nilpotent matrix if there exists a positive integer p such that $\mathbf{A}^p = \mathbf{0}$
c) *controllable* (in finite time) if and only if the system is reachable and null-controllable.

2.3.2 Stability

Stability properties for homogenous positive linear discrete-time system, i.e., $\mathbf{B} = 0$ has been discussed in [42] and [43].

Definition 2.9. If $\lambda_1, \lambda_2, \cdots \lambda_n$ are the eigenvalues of the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ then $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \cdots \lambda_n\}$ is called the spectrum of **A**. if

$$\rho(\mathbf{A}) = \max_i \{|\lambda_i|, \lambda_i \in \sigma(\mathbf{A})\}$$

then $\rho(\mathbf{A})$ is called the dominant or maximal eigenvalue of the matrix **A**.

The positive linear discrete-time system (2.2) is *asymptotically stable* if and only if $\rho(\mathbf{A}) < 1$. It is *critically stable* if $\rho(\mathbf{A}) = 1$ and *unstable* if $\rho(\mathbf{A}) > 1$ [43].

In the case of scalar control, the positive system (2.2) is *asymptotically stable* if and only if $0 \leq a < 1$, where *a* is a scalar weight of state when the state and control in the system (2.2) are scalar [44]. It is *stable*, if $0 \leq a \leq 1$.

2.3.3 Reachable sets

The reachable sets in positive systems is a convex cone. A subset C is called a *cone* if it is closed under positive scalar multiplication that is $\lambda \boldsymbol{x} \in C$ when $\boldsymbol{x} \in C$ and $\lambda > 0$ [14]. The studies of the reachable set in positive linear discrete-time systems has been considered earlier by [45-49].

Definition 2.10. [15] The set of all states $\mathcal{R}_t(\mathbf{0})$ of positive linear discrete-time system (2.2) reachable in *t*-steps by admissible (i.e. nonnegative) control sequences $\{\mathbf{u}(0), \mathbf{u}(1), \cdots, \mathbf{u}(N-1)\}$ is called a *t*-steps reachable set. It is defined as

$$\mathcal{R}_t(\mathbf{0}) = \{\mathbf{x} | \mathbf{x} = \textstyle\sum_{k=0}^{t-1} \mathbf{A}^{t-1-k} \mathbf{B} \mathbf{u}(k)\ ;\ \mathbf{A} \in \mathbb{R}_+^{n \times n}, \mathbf{B} \in \mathbb{R}_+^{n \times m},\ \mathbf{x}_0 \in \mathbb{R}_+^n, \text{and } \mathbf{u}(k) \in \mathbb{R}_+^m \text{ for } k = 0,1,\cdots,t-1\ \} \tag{2.4}$$

with $$\mathcal{R}_0(\mathbf{0}) \equiv \mathbf{0}.$$ □

Definition 2.11. [15] The set of all states $\mathcal{R}_t(\mathbf{x}_0)$ of positive linear discrete-time system (2.2) reachable in *t*-steps by admissible (i.e. nonnegative) control sequences $\{\mathbf{u}(0), \mathbf{u}(1), \cdots, \mathbf{u}(N-1)\}$ is called a *t*-steps reachable set. It is defined as

$$\mathcal{R}_t(\mathbf{x}_0) = \{\mathbf{x}|\mathbf{x} = \mathbf{A}^t\mathbf{x}_0 + \sum_{k=0}^{t-1}\mathbf{A}^{t-1-k}\mathbf{B}\mathbf{u}(k)\ ;\ \mathbf{A} \in \mathbb{R}_+^{n\times n}, \quad (2.5)$$

$$\mathbf{B} \in \mathbb{R}_+^{n\times m},\ \mathbf{x}_0 \in \mathbb{R}_+^n, \text{and } \mathbf{u}(k) \in \mathbb{R}_+^m \text{ for } k = 0,1,\cdots,t-1\}$$

with $$\mathcal{R}_0(\mathbf{x}_0) \equiv \mathbf{x}_0.$$ □

From (2.4) and (2.5) it is clear that

$$\mathcal{R}_t(\mathbf{x}_0) = \mathbf{A}^t\mathbf{x}_0 + \mathcal{R}_0(\mathbf{0})$$

It is not difficult to see that the reachable sets $\mathcal{R}_t(\mathbf{0})$ is a convex cone. However, $\mathcal{R}_t(\mathbf{x}_0)$ is a shifted convex cone.

In the case of *positive linear discrete-time system with scalar state and control*, the reachable set is formulated as

$$\mathcal{R}_t(x_0) = \begin{cases} a^t x_0 + \mathcal{R}_0(0), \text{for } b > 0 \\ \{a^t x_0\}, \text{for } b = 0 \end{cases} \quad (2.6)$$

Let $\mathcal{R}_t^s(\mathbf{x}_0)$ denotes the reachable sets of a *stable* positive linear discrete-time system (2.2) with scalar state and control. For *b* > 0 the reachable sets (2.5) of a *stable system* possess the nested property

$$\mathcal{R}_{t-1}^s(x_0) \subseteq \mathcal{R}_t^s(x_0), t = 1,2,\cdots,T \quad (2.7)$$

where the inclusion is strict if the system is asymptotically stable, i.e. $0 < a < 1$. The inclusion property (2.7) is in the opposite direction

$$\mathcal{R}_t^{us}(x_0) \subset \mathcal{R}_{t-1}^{us}(x_0), t = 1,2,\cdots,T \quad (2.8)$$

if the positive linear discrete-time system (2.2) with scalar state and control is *unstable* (*a* > 1) and *b* > 0. In (2.8), $R_t^{us}(x_0)$ denotes the *t*-step reachable set of an unstable positive linear discrete-time system.

Lemma 2.3. Let $\mathbf{x} = \{\mathbf{x}_0, \mathbf{x}(1), \dots, \mathbf{x}(N)\}$ be a feasible trajectory of positive linear discrete-time system (2.2) that is $\mathbf{x} \geq 0$ for some admissible (that is nonnegative) control sequence $\{\mathbf{u}(0), \mathbf{u}(1), \dots, \mathbf{u}(N-1)\} \geq 0$. Then,

$$\mathbf{x}(t+1) - \mathbf{A}\mathbf{x}(t) \geq 0 \text{ for } t = 0,1,\dots.N-1. \quad (2.9)$$

Proof. The inequality (2.9) follows immediately from (2.2) since the control $\mathbf{u}(t)$, $t = 0,1,\dots.N-1$, as well the control matrix **B** are nonnegative. Proved. □

Lemma 2.4. Let the given final state $\mathbf{x}_N$ is in the *N*-step reachable set $\mathcal{R}_N(\mathbf{x}_0)$ that is $\mathbf{x}(N) = \mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$. Then, the inequality

$$\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}(t) \geq 0 \quad (2.10)$$

holds true for $t = 0,1,\cdots,N-1$ on any (feasible) trajectory $\{\mathbf{x}(t)\} \geq 0$ for $t = 0,\cdots,N$ that ends at the final state $\mathbf{x}(N) = \mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$ and starts at some initial state $\mathbf{x}_0 \geq 0$.

Proof. Let $\mathbf{x}(N) = \mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$. Then, there exist an admissible (that is non-negative) control sequence $\mathbf{u} = \{\mathbf{u}(0), \mathbf{u}(1), \dots, \mathbf{u}(N-1)\} \geq 0$ such that the corresponding trajectory $\mathbf{x} = \{\mathbf{x}0, \mathbf{x}(1), \dots, \mathbf{x}(N-1), \mathbf{x}(N)\} \geq 0$ ends at $\mathbf{x}(N) = \mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$ for some initial state $\mathbf{x}_0 \geq 0$ and

$$\mathbf{x}(N) = \mathbf{x}_N = \mathbf{A}^N\mathbf{x}_0 + \textstyle\sum_{k=0}^{N-1} \mathbf{A}^{N-1-k}\mathbf{B}\mathbf{u}(k) \quad (2.11)$$

Let now t be any time instant from the set $\{0, 1, \dots, N-1\}$. The state $\mathbf{x}(t)$. Can be expressed as

$$\mathbf{x}(t) = \mathbf{A}^t\mathbf{x}_0 + \textstyle\sum_{k=0}^{t-1} \mathbf{A}^{t-1-k}\mathbf{B}\mathbf{u}(k)\,. \quad (2.12)$$

Pre-multiplication of both sides of (2.12) by $\boldsymbol{A}^{N-t}$ leads to

$$\mathbf{A}^{N-t}\mathbf{x}(t) = \mathbf{A}^N\mathbf{x}_0 + \textstyle\sum_{k=0}^{t-1} \mathbf{A}^{N-1-k}\mathbf{B}\mathbf{u}(k). \quad (2.13)$$

The sum in the right side of (2.13) can be rewritten as

$$\textstyle\sum_{k=0}^{t-1} \mathbf{A}^{N-1-k}\mathbf{B}\mathbf{u}(k) = \sum_{k=0}^{N-1} \mathbf{A}^{N-1-k} - \sum_{k=t}^{N-1} \mathbf{A}^{N-1-k}, \quad (2.14)$$

Then, taking into account (2.11) and (2.14), the expression (2.13) becomes

$$\mathbf{A}^{N-t}\mathbf{x}(t) = \mathbf{x}_N - \textstyle\sum_{k=t}^{N-1} \mathbf{A}^{N-1-k}\mathbf{B}\mathbf{u}(k). \quad (2.15)$$

It can easily be seen from (2.15) that the inequality (2.10) is nonnegative that

$$\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}(t) \geq 0 \text{ for } t = 0,1,\cdots,N-1,$$

since $\mathbf{A}, \mathbf{B}, \mathbf{u}(t) \geq 0$ and t is any time instant from the set $\{0, 1, \dots, N-1\}$. The Lemma thus is proved. □

2.4 Positive linear continuous-time system

Consider a positive linear continuous-time system

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \tag{2.16}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is an $n \times n$ Metzler matrix,

$\mathbf{B} \in \mathbb{R}_+^{n\times m}$,

$\mathbf{x}(t) \in \mathbb{R}^n$ is the state variables at time t and $\mathbf{u}(t) \in \mathbb{R}_+^m$ is the control at time t [5].

2.4.1 Controllability and Reachability

Definition 2.12. Consider the positive linear continuous-time system (2.16)

a) A *state* $\mathbf{x}_N \in \mathbb{R}_+$ is *reachable* at time N, if there exists a nonnegative control $\mathbf{u}(t) \in \mathbb{R}_+$ for $t \in [0, N]$ that steers the state trajectory from initial state $\mathbf{x}_0 = 0$ to the state $\mathbf{x}_N$ [14].

b) The *system* (2.16) is called *reachable* if every state $\mathbf{x}_N \in \mathbb{R}_+$ is reachable at some time instant $N > 0$ [14].

c) The system (2.16) is called *strongly reachable* if for every $N > 0$ and every state $\mathbf{x}_N \in \mathbb{R}_+$ the state $\mathbf{x}_N$ is reachable at time N [50].

d) The positive system (2.16) is said to be *strongly reachable* if and only if for any $N > 0$ and any $\mathbf{x}(N) \in \mathbb{R}_+^n$ there exists a nonnegative piecewise continuous function $\mathbf{u}(t)$ that steers the state of the system from $\mathbf{x}(0)$ to $\mathbf{x}(N) = \mathbf{x}_N$ [51].

e) The positive system (2.16) is essentially reachable if every strictly positive state $\mathbf{x}_N \in \mathbb{R}_+^n$ is reachable at some time.

Scalar positive systems are strongly reachable [51].

Definition 2.13. The reachability gramian at time N of the positive system (2.16) is defined as

$$\mathbf{R}_N := \int_0^N e^{\mathbf{A}\tau}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T\tau}\, d\tau. \tag{2.17}$$

It is clear that $\mathbf{R}_N$ is nonnegative for all $N > 0$.

Theorem 2.5. [14] The positive system (2.16) is reachable if the matrix

$$\mathbf{R}_N := \int_0^N e^{\mathbf{A}\tau}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T\tau}\, d\tau.$$

is a monomial matrix. The input that steers the state of the system in time N from $\mathbf{x}_0 = 0$ to the state $\mathbf{x}_N \in \mathbb{R}_+^n$ is given by formula

$$\mathbf{u}(t) = \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T (\mathbf{R}_N)^{-1}\mathbf{x}_N, t \in [0, N). \tag{2.18}$$

Theorem 2.6. [39] The positive system in (2.16) is *reachable* in time t if **A** is a diagonal matrix and $\mathbf{B} \geq 0$ is a monomial matrix.

Theorem 2.7. [52] The system (2.16) is reachable if and only if **A** is diagonal and **B** contains an $n \times n$ monomial submatrix (so $m \geq n$).

Theorem 2.8. [51] The system (2.16) is *strongly reachable* if and only if after a possible reordering of the inputs the matrix A is diagonal, and matrix **B** can be written as $\mathbf{B} = (\mathbf{D}, \mathbf{B}_1)$, where **D** is an order n diagonal positive matrix with positive diagonal entries and $\mathbf{B}_1$, which exists only if $m > n$, is an arbitrary $n \times (m - n)$ positive matrix.

In other words, there exists $m \times n$ selection matrix (a submatrix of some permutation matrix) S such that

$$\mathbf{A} = \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{bmatrix} \mathbf{BS} = \begin{bmatrix} \beta_1 & & & \\ & \beta_2 & & \\ & & \ddots & \\ & & & \beta_n \end{bmatrix} \tag{2.19}$$

where $\alpha_i \in \mathbb{R}$, while $\beta_i > 0$ for every index $i = 1,2,\cdots,n$. Hence, it is clear that $m \geq n$.

Theorem 2.9. [50] For the positive continuous-time system (2.16), the following facts are equivalent:

i. The system is *reachable*;

ii. The system is *strongly reachable*;

iii. $m \geq n$ and there exists an $m \times n$ selection matrix (a submatrix of some permutation matrix) $\mathbf{S}$ described in (2.19) for some $\alpha_i \in \mathbb{R}$, while $\beta_i > 0$.

Theorem 2.10. [50] For the positive continuous-time system (2.16), the following facts are equivalent:

i. The system is *reachable*;

ii. The system is *monomially reachable*, which amounts to saying that every vector $\mathbf{e}_1, i \in \langle n \rangle$, of the canonical basis is reachable at some time instant $N_i > 0$;

iii. The system is *strongly reachable*.

Definition 2.14. [14] The positive system in (2.16) is called *null controllable* (*controllable to the origin*) in time *N* if for any nonzero initial state $\mathbf{x}_0 \in \mathbb{R}_+^n$ there exist an input $\mathbf{u}(t) \in \mathbb{R}_+^m, t \in [0, N)$ that steers the state of the system from $\mathbf{x}_0$ to origin ($\mathbf{x}_N = 0$).

Definition 2.15. [14] The positive systems in (2.16) is called *null controllable* (*controllable to the origin*) if for any nonzero initial state $\mathbf{x}_0 \in \mathbb{R}_+^n$, there exists a time instant $N > 0$ and an input $\mathbf{u}(t) \in \mathbb{R}_+^m,\ t \in [0, N)$ that steers the state of the system from $\mathbf{x}_0$ to origin ($\mathbf{x}_N = 0$).

Definition 2.16. The positive system (2.16) is called *asymptotically null-controllable* if for any nonnegative initial state $\mathbf{x}_0 \in \mathbb{R}_+^n$ there exists an input $\mathbf{u}(t) \in \mathbb{R}_+^m$ that steers the state from $\mathbf{x}_0$ to $\mathbf{x}_N = 0$ as $N \to +\infty$.

Theorem 2.11. [14] The positive system in (2.16) is not null controllable in finite time.

Theorem 2.12. [14] The positive systems in (2.16) is null controllable in infinite time if it is asymptotically stable.

Theorem 2.13. [53] If the positive systems in (2.16) is asymptotically null-controllable, then the reachability gramian (2.17) satisfies the Lyapunov equation

$$\mathbf{AP} + \mathbf{AP}^T = -\mathbf{BB}^T,$$

where $\mathbf{P} := \lim_{N\to\infty} \mathbf{R}_N = \int_0^\infty e^{\mathbf{A}\tau}\mathbf{BB}^T e^{\mathbf{A}^T\tau}\, d\tau$.

Definition 2.17. [14] The positive system in (2.16) is called controllable in time N if for any nonzero initial state $\mathbf{x}_0 \in \mathbb{R}_+^n$ and any final state $\mathbf{x}_N \in \mathbb{R}_+^n$, if there exists an input that steers the state of the system from $\mathbf{x}_0$ to $\mathbf{x}_N$.

Definition 2.18. [14] The positive system in (2.16) is called controllable for any nonzero initial state $\mathbf{x}_0 \in \mathbb{R}_+^n$ and any final state $\mathbf{x}_N \in \mathbb{R}_+^n$, if there exists a time instant $N > 0$ and an input $\mathbf{u}(t) \in \mathbb{R}_+^m$, $t \in [0, N)$ that steers the state of the system from $\mathbf{x}_0$ to $\mathbf{x}_N$.

Theorem 2.14. [14] The positive system in (2.16) is controllable if the matrix in equation (2.17) is a monomial matrix for $N > 0$ and

$$\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0 \in \mathbb{R}_+^n. \tag{2.20}$$

The input that steers the state of the system in time N from $\mathbf{x}_0 \in \mathbb{R}_+^n$ to $\mathbf{x}_N \in \mathbb{R}_+^n$ is given by formula

$$\mathbf{u}(t) = \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T (\mathbf{R}_N)^{-1}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0), t \in [0, N) \tag{2.21}$$

where the matrix $\mathbf{R}_N$ is defined in (2.17).

Define a controllability matrix

$$\mathbf{C}_n = [\mathbf{B} \quad \mathbf{AB} \quad \mathbf{A}^2\mathbf{B} \quad \cdots \quad \mathbf{A}^{n-1}\mathbf{B}].$$

Theorem 2.15. [53] For all $N > 0$, $rank(\mathbf{C}_n) = rank(\mathbf{R}_N)$.

Theorem 2.16. [53] If the system (2.16) is controllable, then the reachability gramian matrix $\mathbf{R}_N$ is positive definite for all $N > 0$.

2.4.2 Stability

To the best of our knowledge, the stability of positive linear continuous-time system is only observed when the system is homogenous, i.e. $\mathbf{B} = \mathbf{0}$

A homogenous positive linear system is formulated as

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t). \tag{2.22}$$

Definition 2.19. A homogenous linear system (2.22) is said to be positively stable if for all $t \geq 0$ and all nonnegative initial state, i.e., $\mathbf{x}_0 \geq 0$, $\mathbf{x}(t) \geq 0$ and $\mathbf{x}(t) \to 0$ as $t \to \infty$.

Theorem 2.17. A homogenous positive linear system (2.22) is exponentially stable if and only if the Frobenious eigenvalue λ_F of A is negative [33].

Theorem 2.18. [33]A homogenous positive linear system is stable if and only if there exists a diagonal positive definite matrix **P** such that the matrix **Q**, defined by

$$-\mathbf{Q} = \mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A}.$$

Theorem 2.19. [33] A homogenous linear system (2.22) is positively stable if and only if **A** is a stable Metzler matrix.

Theorem 2.20. [33] A homogenous linear system (2.22) is positively stable if and only if **A** is a Metzler matrix and there exists a diagonal positive definite matrix **P** such that $\mathbf{P}\mathbf{A}^T + \mathbf{A}\mathbf{P}$ is negative definite.

2.4.3 Reachable sets

Definition 2.20. The set of all states $\mathbf{x}(t)$ at any time $t \geq 0$ reachable from an initial state $\mathbf{x}_0 = 0$ by nonnegative input $\mathbf{u}(\tau)$, $\tau \in [0,\ t)$ is called a reachable set $R_t(0)$ and defined as

$$R_t(0) \equiv \left\{ \mathbf{x} | \mathbf{x} = \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau \right\}, \tag{2.23}$$

$$\mathbf{A} \in \mathbb{R}^{n \times n} \text{ is a metzler matrix, } \mathbf{B} \in \mathbb{R}_+^{n \times m}, \text{and } t \in [0, N].$$

with $$R_0(\mathbf{x}_0) \equiv \{0\}.$$

Definition 2.21. The set of all states $\mathbf{x}(t)$ at any time $t \geq 0$ reachable from an initial state $\mathbf{x}_0 \geq 0$ by nonnegative input $\mathbf{u}(\tau)$, $\tau \in [0, t)$ is called a reachable set $R_t(\mathbf{x}_0)$ and defined as

$$R_t(\mathbf{x}_0) \equiv \left\{ \mathbf{x} | \mathbf{x} = e^{\mathbf{A}t}\mathbf{x}_0 + \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau \right\}, \tag{2.24}$$

$$\mathbf{A} \in \mathbb{R}^{n \times n} \text{ is a metzler matrix, and } \mathbf{B} \in \mathbb{R}_+^{n \times m},\ t \in [0, N].$$

with $$R_0(\mathbf{x}_0) \equiv \{\mathbf{x}_0\}.$$

It is clear that $$R_t(\mathbf{x}_0) = e^{\mathbf{A}t}\mathbf{x}_0 + R_t(0)$$

It is not difficult to see that reachable set $R_t(0)$ is a positive cone. However, the reachable set $R_t(\mathbf{x}_0)$, is a *shifted positive cone.*

Lemma 2.5. If the final state $\mathbf{x}_N$ is in the reachable set $R_N(\mathbf{x}_0)$, then $\mathbf{x}_N - e^{\mathbf{A}N}x_0 \geq 0$.

Proof. From (2.24), we have $\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0 = \int_0^N e^{\mathbf{A}(N-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau$. Since $\mathbf{x}_N$ is in the reachable set, $\mathbf{x}_N$ is reachable. Hence, there exists a non-negative control $\mathbf{u}(\tau) \geq 0$, $\tau \in [0, N)$. Therefore $\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0 = \int_0^N e^{\mathbf{A}(N-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau \geq 0$. □

Lemma 2.6. If the final state $\mathbf{x}_N$ is in the reachable set $R_N(\mathbf{x}_0)$ and $\mathbf{x}^*(t)$ for $t \in [0, N)$ is the state on the optimal trajectory of positive continuous-time system (2) – (3), then $\mathbf{x}_N - e^{\mathbf{A}(N-t)}\mathbf{x}^*(t) \geq 0$.

Proof : from (2.24), we have $\mathbf{x}_N = e^{\mathbf{A}N}x_0 + \int_0^N e^{\mathbf{A}(N-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau$ and $\mathbf{x}^*(t) = e^{\mathbf{A}t}x_0 + \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau$. Multiplying $\mathbf{x}^*(t)$ by $e^{\mathbf{A}(N-t)}$gives us $e^{\mathbf{A}(N-t)}\mathbf{x}^*(t) = e^{\mathbf{A}N}x_0 + \int_0^t e^{\mathbf{A}(N-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau$. Therefore, $x_N - e^{\mathbf{A}(N-t)}x^*(t) = \int_t^N e^{\mathbf{A}(N-\tau)} \mathbf{B}\,\mathbf{u}(\tau) d\tau \geq 0$ as $\mathbf{u}(\tau) \geq 0$ for $\tau \in [t, N)$.

CHAPTER 3

The minimum energy problem for positive linear system with scalar control

3.1 General

Minimum energy problem is a classic problem in optimal control. This chapter concerns the minimum energy problem with scalar control for both positive linear discrete-time systems and positive linear continuous-time systems. The optimal control of the minimum energy problem for both positive linear discrete-time systems and continuous-time systems are established using dynamic programming. The optimal control formula obtained are represented in an open loop and a feedback forms. The correlation between the optimal control and the geometry of the reachable set is also revealed. This chapter is based on my publications [54-56] during my study in this field.

3.2 The minimum energy problem for positive linear discrete-time system

3.2.1 Problem formulation

The minimum energy problem for scalar positive discrete-time systems with fixed final state is formulated as follows [14]

Minimize $$J = \frac{1}{2} r \sum_{t=0}^{N-1} u^2(t), \tag{3.1}$$

subject to $$x(t+1) = ax(t) + bu(t), t = 1, 2, \cdots, N-1, \tag{3.2}$$

$$a, b \geq 0, r, u(t) \in \mathbb{R}_+,$$

where $x(t)$ is the state at time $t = 0, 1, 2, \cdots, N$, $u(t) \in \mathbb{R}_+$ is the control sequence, N is a finite-time horizon, and the initial and final state are given by

$$x(0) = x_0 \geq 0 \text{ and } x(T) = x_T \geq 0. \tag{3.3}$$

The state variables $x(t)$, $t = 0, 1, 2, \cdots, N$, are, clearly, nonnegative for any nonnegative initial state $x_0 \geq 0$, and any (nonnegative) control sequences $u(t)$, $t = 0, 1, 2, \cdots, N-1$.

3.2.2 Main results

Theorem 3.1. Let $x_N \in \mathcal{R}_t(x_0)$. Then, the optimal control sequence that minimizes the cost function (3.1) in the minimum energy problem (3.1) – (3.3) with final state is given by

$$u^*(t) = \begin{cases} \frac{a^{N-(t+1)}\left(x_T - a^{N-t}x^*(t)\right)}{b\sum_{i=t}^{N-1} a^{2(N-(i+1))}} & , b > 0 \\ 0 & , b = 0 \end{cases}, t = 0, 1, 2, \cdots, N-1 \tag{3.4}$$

where $x^*(t)$ is the corresponding optimal trajectory, and the optimal value of the cost function (3.1) is

$$J_0^* = \begin{cases} \frac{1}{2} r \frac{(x_N - a^N x_0)}{b^2 \sum_{i=0}^{N-1} a^{2(N-i-1)}} & , b > 0 \\ 0 & , b = 0 \end{cases} \tag{3.5}$$

Proof. The hypothesis $x_N \in \mathcal{R}_t(x_0)$ implies that there exists a solution to the two-point boundary value problem (3.1) – (3.3). In other words, there exists an admissible (nonnegative) control sequence $\{u(t) \geq 0, t = 0, 1, \cdots, N-1\}$ such that the corresponding trajectory $\{x_0, x(1), \cdots, x(N-1), x_N\}$ is feasible (that is nonnegative).

When $b = 0$, the positive discrete-time linear system (3.1) – (3.3) is not reachable and the reachable set $\mathcal{R}_t(x_0)$ consists of the point $a^N x_0$ only. Then the only solution to the minimum energy problem (3.1) – (3.3) is the trivial one.

$$\{u^*(t) = 0, 1, \cdots, N-1\}, \{x_0, ax_0, \cdots, a^{N-1}x_0, x_N\}, \text{ and } J_0^* = 0.$$

Let $b > 0$. By hypothesis $x_N \in \mathcal{R}_t(x_0)$, the two point boundary-value problem (3.1) – (3.3) is consistent and there exists at least one solution. To find the solution that minimizes the cost function (3.1), we apply the dynamic programming procedure [20].

The Bellman equation for the minimum energy problem (3.1) – (3.3) can be written as

$$J_t(x(t)) = \min_{u\geq 0}\left\{\frac{1}{2}ru^2(t) + J_{t+1}(x(t+1))\right\}, t = 0,1,\cdots,N-1, \tag{3.6}$$

with $J_N(x) = 0.$

Moving backwards we try for $t = N-1, t = N-2$ and formulate the induction hypothesis:

$$J_t(x) = \frac{1}{2}r\frac{(x_N - a^{N-t}x)}{b^2\sum_{i=t}^{N-1} a^{2(N-i-1)}} \tag{3.7}$$

and

$$u(t) = \frac{a^{N-(t+1)}(x_N - a^{N-t}x)}{b\sum_{i=t}^{N-1} a^{2(N-(i+1))}} \geq 0. \tag{3.8}$$

Suppose that the expressions (3.7) and (3.8) are true for $t = k+1$, that is

$$J_{k+1}(x) = \frac{1}{2}r\frac{(x_N - a^{N-(k+1)}x)}{b^2\sum_{i=k+1}^{N-1} a^{2(N-i-1)}} \tag{3.9}$$

and, respectively,

$$u(k+1) = \frac{a^{N-(k+2)}(x_N - a^{N-(k+1)}x)}{b\sum_{i=k+1}^{N-1} a^{2(N-(i+1))}} \geq 0. \tag{3.10}$$

we prove that (3.7) and (3.8) are true for $t = k$, that is

$$J_k(x) = \frac{1}{2}r\frac{(x_N - a^{N-k}x)}{b^2\sum_{i=k}^{N-1} a^{2(N-i-1)}} \tag{3.11}$$

and

$$u(k) = \frac{a^{N-(k+1)}(x_N - a^{N-k}x)}{b\sum_{i=k}^{N-1} a^{2(N-(i+1))}} \geq 0. \tag{3.12}$$

For $t = k$ the Bellman equation is specified as

$$J_k(x) = \min_{u\geq 0}\left\{\frac{1}{2}ru^2 + J_{k+1}(x)\right\}. \tag{3.13}$$

A substitution of state equations (3.2) and (3.11) into (3.13) yields

$$J_k(x) = \min_{u\geq 0}\left\{\frac{1}{2}ru^2 + \frac{1}{2}r\frac{\left(x_N - a^{N-k-1}(ax+bu)\right)^2}{b^2\sum_{i=k+1}^{N-1} a^{2(N-i-1)}}\right\}, \tag{3.14}$$

where $u = u(k)$ and $x = x(k)$ is to be specified by the initial condition $x(0) = x_0$.

From the first order necessary conditions for minimum, by setting the first order derivative of $\frac{1}{2}r\left(u^2 + \frac{\left(x_N - a^{N-k-1}(ax+bu)\right)^2}{b^2\sum_{i=k+1}^{N-1} a^{2(N-i-1)}}\right)$ with respect to *u* to zero, we get

$$u(k) = \frac{a^{N-(k+1)}(x_N - a^{N-k}x)}{b\sum_{i=k}^{N-1} a^{2(N-(i+1))}} \geq 0 \tag{3.15}$$

It is clear that the nonnegativity control constraints, i.e., $u(t) \in \mathbb{R}_+$ can be omitted in the derivative, as the final state x_N is restricted to be in the reachable set $\mathcal{R}_t(x_0)$ that guarantees the nonnegativity of control $u(k)$.

The substitution of (3.14) in (3.13) leads to

$$J_k(x) = \frac{1}{2}r\frac{(x_N - a^{N-k}x)}{b^2\sum_{i=k}^{N-1} a^{2(N-i-1)}}, \tag{3.16}$$

Thus the assumptions (3.7) and (3.8) are true for $t = k$, and, therefore, they are true by induction for any *t*.

For $t = 0$ we have

$$u(0) = \frac{a^{N-1}(x_N - a^N x_0)}{b\sum_{i=0}^{N-1} a^{2(N-(i+1))}} \geq 0.$$

This concludes the proof of the theorem. □

Remarks 3.1.

1. If $x_N \notin \mathcal{R}_N(x_0)$ then the two point boundary value problem (3.1) – (3.3) is inconsistent and, therefore, the minimum energy problem with fixed final state (3.1) – (3.3) has no solution.
2. If $b = 0$ and x_N belongs to the boundary of $\mathcal{R}_N(x_0)$, that is $x_N = a^N x_0$, the optimal control sequence (3.4) is a zero sequence and the corresponding optimal trajectory becomes $x_0, ax_0, \cdots, a^{N-1}x_0, x_N$.
3. Let the system (3.1) – (3.3) be controllable, that is let $a = 0$ and $b > 0$. Then, it is not difficult to see using limits that expression (3.4) is reduced to
   $u^*(t) = 0$ for $t = 0,1, \cdots, N-2$ and $u^*(N-1) = \frac{x_N}{b}$,
   and, consequently, the minimal value of the cost function becomes
   $J_0^x(x) = \frac{1}{2}\frac{x_N^2}{b^2}$.
4. For $b > 0$ and $t = 0$ expression (3.4) becomes

$$u^*(0) = \frac{a^{N-1}(x_T - a^N x_0)}{b\sum_{i=0}^{N-1} a^{2(N-(i+1))}} \tag{3.17}$$

   the above expression for $u^*(0)$ for $x_0 = 0$ agrees with the results in [14]. The expression (3.4), however, is obtained for any nonnegative pair $\{x_0, x_N\}$ such that $x_N \in \mathcal{R}_N(x_0)$ and, therefore, is more general than that in [14], where (among the other assumptions) the reachability of positive discrete-time linear system and a zero initial state are required.
   The expression
   $b^2 \sum_{i=0}^{N-1} a^{2(N-(i+1))} = (b, ab, a^2b, \cdots, a^{N-1}b) \cdot (b, ab, a^2b, \cdots, a^{N-1}b)^T \geq 0$
   is the gramian of positive discrete-time linear system (3.1) – (3.3).
5. The optimal control law (3.4) for $b > 0$ can be treated as a feedback control since it depends on the current state. As a matter of fact, it can also be represented as an open-loop control that depends on the initial and final states as the corollary below shows.

Corollary 3.1. Under the assumptions of Theorem 3.1.

1. For $a \neq 1$ and $a \geq 0$, the optimal control can be represented as an open-loop control namely

$$u^*(t) = \begin{cases} \frac{a^{N-(t+1)}(x_N - a^N x_0)(1-a^2)}{b(1-a^{2N})} & , b > 0 \\ 0 & , b = 0 \end{cases}, \ t = 0,1,2,\cdots, N-1 \tag{3.18}$$

The optimal trajectory to (3.2), then, becomes

$$x^*(t) = \begin{cases} a^t x_0 + \frac{a^{N-t}(x_N - a^N x_0)(1-a^{2t})}{(1-a^{2N})} & , b > 0 \\ a^t x_0 & , b = 0 \end{cases}, \ t = 0,1,2,\cdots, N \tag{3.19}$$

and the cost function (3.1) is given by

$$J_0^*(x) = \begin{cases} \frac{1}{2} r \frac{(x_N - a^N x_0)^2 (1-a^2)}{b^2 (1-a^{2N})} & , b > 0 \\ 0 & , b = 0 \end{cases}. \tag{3.20}$$

2. For $a = 1$ the optimal control can be represented as an open-loop control namely

$$u^*(t) = \begin{cases} \frac{(x_N - x_0)}{bN} & , b > 0 \\ 0 & , b = 0 \end{cases}, \ t = 0,1,2,\cdots, N-1. \tag{3.21}$$

The optimal trajectory to (3.2), then, becomes

$$x^*(t) = \begin{cases} x_0 + \frac{t(x_N - x_0)}{N} & , b > 0 \\ x_0 & , b = 0 \end{cases}, \ t = 0,1,2,\cdots, N \tag{3.22}$$

and the cost function (3.1) is given by

$$J_0^*(x) = \begin{cases} \frac{1}{2} r \frac{(x_N - x_0)^2}{(bN)^2} & , b > 0 \\ 0 & , b = 0 \end{cases}. \tag{3.23}$$

Remarks 3.2.

1. Expressions (3.18), (3.19), (3.21) and (3.22) clearly show that the optimal control sequence and the optimal trajectory are nonnegative.

2. It is easy to see from (3.19) and (3.22) that the optimal trajectory ends at the desired final state that is $x^*(N) = x_N$.
3. The optimal control sequence (3.18) is easy to calculate since $u^*(t+1) = \frac{u^*(t)}{a}$.
4. If the initial state x_0 is zero, expressions (3.18), (3.19) and (3.20) become even simpler

$$u^*(t) = \frac{a^{N-(t+1)}(1-a^2)x_N}{b(1-a^{2N})}, t = 0,1,2,\cdots,N-1$$

$$x^*(t) = \frac{a^{N-t}(1-a^{2t})x_N}{(1-a^2)}, t = 0,1,2,\cdots,N,$$

and
$$J_0^*(x) = \frac{1}{2}r\frac{(1-a^2)x_N^2}{b^2(1-a^{2N})}.$$

the meaning of the above expression is transparent.

5. It is worth noting that the expression (3.18) for the optimal control sequence and the expression (3.20) for the minimal value of the cost function are the same as those when no restrictions are imposed on the controls. This is because the minimum of J_0 as a function of $u(0), u(1), \cdots, u(N-1)$ is achieved at an interior point $u^*(k) > 0$, $k = 0,1,2,\cdots,N-1$ for $b > 0$ and $x_N \in \mathcal{R}_N(x_0)$.

3.2.3 Numerical Examples

To illustrate the results obtained in section 3.2.2, we consider the following *the minimum energy problem for scalar positive discrete-time systems with fixed final state* examples and solve the problems using the formula obtained in the previous section.

Example 3.1. (*Asymptotically stable* and $x_N \in \mathcal{R}_N(x_0)$)

Consider the following minimum energy problem

Minimize
$$J = 2\sum_{t=0}^{2} u^2(t),$$

subject to

$$x(t+1) = \frac{1}{2}x(t) + u(t), t = 0,1,2,3,$$

$$x(t), u(t) \in \mathbb{R}_+,$$

$x(0) = x_0 = 1$ and $x(4) = x_4 = 0.2$ are given.

The optimal control is formulated as

$$u^*(t) = \frac{0.5^{3-(t+1)}(0.2-0.5^3)(1-0.5^3)}{(1-0.5^6)}, t = 0,1,2.$$

Using the above formula, the optimal control sequence is obtained as follows

$$u = \{0.01428571429, 0.02857142857, 0.05714285714\}.$$

The optimal trajectory is formulated as

$$x^*(t) = 0.5^t + \frac{0.5^{2-t}(0.2-0.5^2)(1-0.5^{2t})}{(1-0.5^4)}, t = 0,1,2.$$

Graphically, the optimal control and the optimal trajectory are described as follows

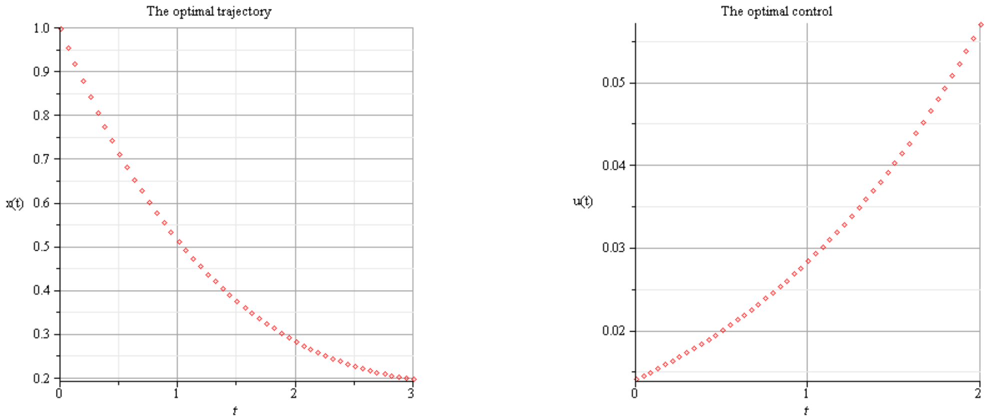

Figure 3.1. The optimal control and correseponding trajectory for Example 3.1

Example 3.2. (*Marginally stable* and $x_N \in \mathcal{R}_N(x_0)$)

Consider the following minimum energy problem

Minimize $$J = 2\sum_{t=0}^{3} u^2(t),$$

subject to

$$x(t+1) = x(t) + u(t), t = 0,1,2,3,$$

$$x(t), u(t) \in \mathbb{R}_+,$$

$x(0) = x_0 = 1$ and $x(4) = x_4 = 5$ are given.

The optimal control is formulated as

$$u^*(t) = 1 \, , t = 0,1,2,3.$$

From (3.22) the optimal trajectory is

$$x^*(t) = 1 + t \, , t = 0,1,2,3.$$

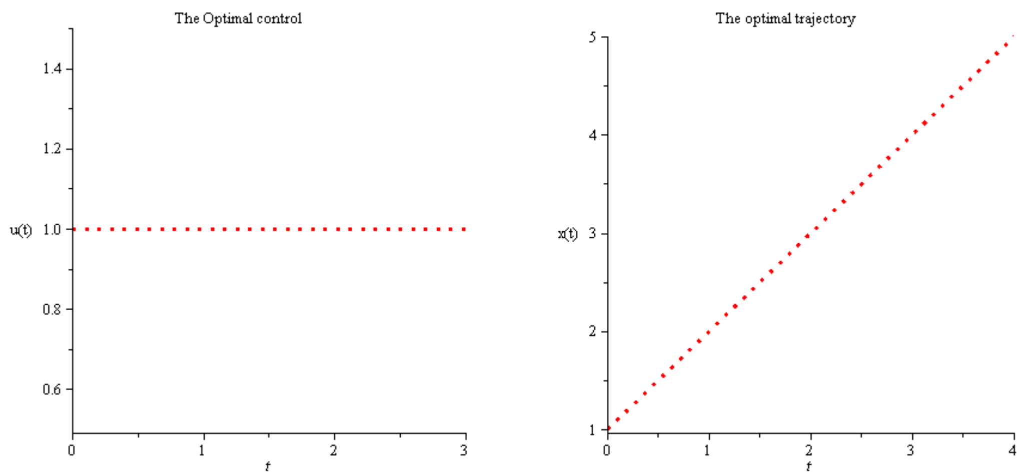

Figure 3.2. The optimal control and corresponding trajectory for Example 3.2.

Example 3.3. (*Unstable* and $x_N \in \mathcal{R}_N(x_0)$)

Consider the following minimum energy problem

Minimize $$J = 2\sum_{t=0}^{3} u^2(t),$$

Subject to

$$x(t+1) = 2x(t) + u(t), t = 0,1,2,3,$$

$$x(t), u(t) \in \mathbb{R}_+,$$

$x(0) = x_0 = 1$ and $x(4) = x_4 = 17$ are given.

The optimal control is formulated as

$$u^*(t) = \frac{2^{4-(t+1)}(17-2^4)(1-2^2)}{(1-2^8)} \, , t = 0,1,2,3.$$

The optimal trajectory is

$$x^*(t) = 2^t + \frac{2^{4-t}(17-2^4)(1-2^{2t})}{(1-2^8)}, \; t = 0,1,2,3.$$

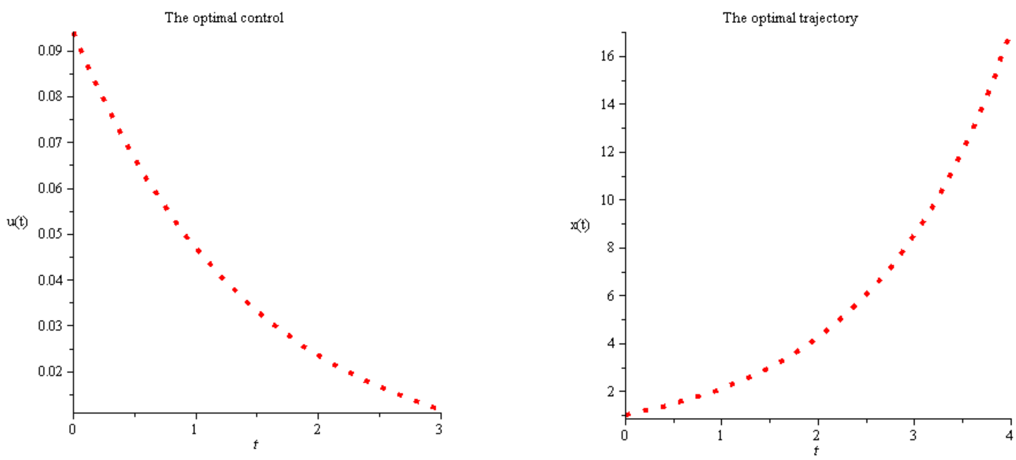

Figure 3.3. The optimal control and corresponding trajectory for Example 3.3

3.3. The minimum energy problem for positive linear discrete-time system with free final state

3.3.1 Problem Formulation

Consider *the minimum energy problem for scalar positive discrete-time systems* (3.1) – (3.3). By relaxing the boundary condition $x(N) = x_N$, problem (3.1) – (3.3) can be reduced to the following *minimum energy problem with free final state.*

Minimize

$$J = \frac{1}{2}\left(x_N - x(N)\right)^2 + \frac{1}{2} r \sum_{t=0}^{N-1} u^2(t), \tag{3.24}$$

subject to

$$x(t+1) = ax(t) + bu(t), t = 1, 2, \cdots, N-1, \quad (3.25)$$

$$a, b \geq 0, u(t) \in \mathbb{R}_+, \quad (3.26)$$

$$x(0) = x_0 \geq 0, \quad (3.27)$$

where $r \in \mathbb{R}_+$; r are the scalar weighting factor. The new term in the cost function (3.1) reflects how close the given terminal state to the terminal state in the reduced problem is. The solution to the minimum energy problem with free terminal state then gives a control sequence that minimizes the energy (3.1) of the input and at the same time the corresponding (to that control sequence) trajectory ends at the point which is in a closed proximity of the targeted terminal point x_N. In other words, the optimum control sequence resolves the 'trade-off' between minimizing the energy of the input signal and the deviation from the given terminal point x_N.

3.3.2 Main Results

Theorem 3.2. Let $x_N \in \mathcal{R}_N(x_0)$, $t = 0, 1, 2, \cdots, N-1$. Then the optimal control sequence that minimizes the cost function (3.24) in the minimum energy problem with free final state (3.24)-(3.27) is given by

$$u^*(t) = \frac{a^{N-t-1} b \left(x_N - a^{N-t} x^*(t)\right)}{r + b^2 \sum_{i=t}^{N-1} a^{2(N-t-1)}}, t = 0, 1, \cdots, N-1, \quad (3.28)$$

where $x^*(t)$ is the corresponding optimal trajectory, and the optimal value of the cost function (3.24) is

$$J_0^* = \frac{1}{2} r \frac{(x_N - a^N x_0)^2}{r + b^2 \sum_{i=t}^{N-1} a^{2(N-i-1)}}. \quad (3.29)$$

Proof.

To find an analytic solution to the optimal control problem (3.24) – (3.27), we use the dynamic programming approach [20].

The Bellman equation can be written as

$$J_t(x) = \min_{u\geq 0}\left\{\frac{1}{2}ru^2 + J_{t+1}(x)\right\}, u = u(t), x = x(t), t = 0,1,\cdots,N-1$$

with

$$J_N(x) = \frac{1}{2}\left(x_N - x(N)\right)^2 \tag{3.30}$$

We try for $t = N-1$. We have

$$J_{N-1}(x) = \min_{u\geq 0}\left\{\frac{1}{2}ru^2 + \frac{1}{2}\left(x_N - (ax+bu)\right)^2\right\}$$

where $x = x(N-1)$ and $u = u(N-1)$ is to be determined by the initial condition (3.30). A differentiation of the above expression with respect to *u* leads to

$$u^*(N-1) = \frac{b(x_N - ax)}{r+b^2},$$

and for u to satisfy $u(t) \in \mathbb{R}_+$ we impose the condition $x_N \in \mathcal{R}_N(x_0)$. Then from Lemma 2.4, $ax(N-1) \leq x_N$, and therefore

$$J_{N-1}(x) = \frac{1}{2}r\frac{(x_N - ax)^2}{r+b^2}.$$

Similarly, for $t = N-2$, we have:

$$u^*(N-2) = \frac{ab(x_N - a^2x)}{r+b^2+a^2b^2} \text{ with } a^2x(N-1) \leq x_N,$$

and, respectively,

$$J_{N-2}(x) = \frac{1}{2}r\frac{\left(x_N - a^2x\right)^2}{r+b^2+a^2b^2}.$$

We form, now, the induction hypothesis:

$$J_t(x) = \frac{1}{2}r\frac{(x_N - a^{N-1}x)^2}{r + b^2\sum_{i=t}^{N-1} a^{2(N-(i+1))}}, \tag{3.31}$$

and

$$u^*(t) = \frac{a^{N-t-1}b(x_N - a^{N-t}x)}{r + b^2\sum_{i=t}^{N-1} a^{2(N-i-1)}}, a^{N-t}x(t) \leq x_N. \tag{3.32}$$

Suppose the expression (3.29) and (3.30) are true for $t = k+1$, that is

$$J_{k+1}(x) = \frac{1}{2} r \frac{(x_N - a^{N-k-1}x)^2}{r + b^2 \sum_{i=k+1}^{N-1} a^{2(N-(i+1))}} \tag{3.33}$$

and, respectively,

$$u^*(k+1) = \frac{a^{N-k-2} b (x_N - a^{N-k-1}x)}{r + b^2 \sum_{i=k+1}^{N-1} a^{2(N-i-1)}}, a^{N-k-1}x(k+1) \le x_N \tag{3.34}$$

We then prove that (3.31) and (3.32) hold also for $t = k$ namely

$$J_k(x) = \frac{1}{2} r \frac{(x_N - a^{N-k}x)^2}{r + b^2 \sum_{i=k}^{N-1} a^{2(N-(i+1))}},$$

and, respectively

$$u^*(k) = \frac{a^{N-k-1} b (x_N - a^{N-k}x)}{r + b^2 \sum_{i=k}^{N-1} a^{2(N-i-1)}}, a^{N-k}x(k) \le x_N.$$

For $t = k$ the Bellman equation is

$$J_k(x) = \min_{u \ge 0} \left\{ \frac{1}{2} r u^2 + J_{k+1}(x) \right\}. \tag{3.35}$$

Substitution of the state equation (3.25) into (3.33) yields

$$J_k(x) = \min_{u \ge 0} \left\{ \frac{1}{2} r u^2 + \frac{1}{2} r \frac{\left(x_N - a^{N-k-1}(ax + bu)\right)^2}{r + b^2 \sum_{i=k+1}^{N-1} a^{2(N-i-1)}} \right\}, \tag{3.36}$$

where $u = u(k)$ and the state $x(k) = x$ is a parameter that is to be specified by the initial condition $x(0) = x_0$. From the first necessary condition from minimum, by setting the first order derivatives of $\frac{1}{2} r u^2 + \frac{1}{2} r \frac{\left(x_N - a^{N-k-1}(ax+bu)\right)^2}{r + b^2 \sum_{i=k+1}^{N-1} a^{2(N-i-1)}}$ with respect to u to zero, we get

$$u^*(k) = \frac{a^{N-k-1} b (x_N - a^{N-k}x)}{r + b^2 \sum_{i=k}^{N-1} a^{2(N-i-1)}} \tag{3.37}$$

where the condition $a^{N-k}x(k) \le x_N$ is imposed in order to satisfy $u(t) \in \mathbb{R}_+$. Furthermore, substitution of (3.37) into (3.36) leads to

$$J_k(x) = \frac{1}{2} r \frac{(x_N - a^{N-k}x)^2}{r + b^2 \sum_{i=k}^{N-1} a^{2(N-(i+1))}}$$

So, the assumption (3.31) and (3.32) hold for $t = k$, and therefore they are true by induction for any k. For $t = 0$ we have

$$u^*(0) = \frac{a^{N-1}b(x_N - a^N x_0)}{r + b^2 \sum_{i=k}^{N-1} a^{2(N-(i+1))}}.$$

This concludes the proof. □

Remarks 3.3.

Assume that $x^*(N) = x_N$. Then, from (3.28) we have

$$u^*(N-1) = \frac{b(x_N - ax^*(N-1))}{r+b^2}.$$

Substituting the above expression for $u^*(N-1)$ and taking into account $x^*(N) = x_N$ we obtain that $x_N = ax^*(N-1)$ and hence,

$$x^*(N-1) = \frac{x_N}{a},$$

which implies that the optimal control

$$u^*(N-1) = 0.$$

Continue the process for $t = N-2$ and $t = N-3$, and form the induction hypothesis for $k = t$, we get

$$x^*(N-t) = \frac{x_N}{a^t}, \tag{3.38}$$

and

$$u^*(N-t) = 0. \tag{3.39}$$

Now, assume that expression (3.38) and (3.39) are true for $k = t$ and we proceed to show that they hold for $k = t+1$, namely

$$x^*\big(N-(t+1)\big) = \frac{x_N}{a^{t+1}}, \tag{3.40}$$

and

$$u^*\big(N-(t+1)\big)=0.$$

For $k=t+1$, expression (3.32) for the optimal control becomes

$$u^*\big(N-(t+1)\big)=\frac{a^t b(x_N-a^{t+1}x)}{r+b^2\sum_{i=N-(t+1)}^{N-1}a^{2(N-(i+1))}}, \tag{3.41}$$

where $x=x\big(N-(t+1)\big)$. Then, substitution of (3.40) and (3.41) into the state equation (3.25) yields

$$\frac{x_N}{a^t}=ax+\frac{a^t b^2(x_N-a^{t+1}x)}{r+b^2\sum_{i=N-(t+1)}^{N-1}a^{2(N-(i+1))}},$$

which results in

$$x\big(N-(t+1)\big)=\tfrac{x_N}{a^{t+1}}, \tag{3.42}$$

so that (3.38) is true. Taking into account (3.42) it is not difficult to see that the optimal control (3.41) becomes $u^*\big(N-(t+1)\big)=0$. The hypothesis is thus proved and this concludes the proof. The optimal control sequence is

$$u^*(t)=0 \text{ for } t=0,1,2,\cdots,N-1,$$

the corresponding optimal trajectory is

$$\{x_0,ax_0,a^2x_0,\cdots,a^Nx_0\},$$

and the optimal cost function

$$J_0^*(x_0)=0.$$

Corollary 3.2. Under the assumptions of Theorem 3.2.

i. For $a\neq 1$, the optimal control can be represented as an open-loop control namely

$$u^*(t)=\begin{cases}\dfrac{a^{N-(t+1)}b(x_N-a^Nx_0)(1-a^2)}{r(1-a^2)+b^2(1-a^{2N})}, & b>0\\ 0 & ,b=0\end{cases}, \tag{3.43}$$

$$t=0,1,2,\cdots,N-1$$

The optimal trajectory to (3.25), then, becomes

$$x^*(t)=\begin{cases}a^tx_0+\dfrac{a^{N-t}b^2(x_N-a^Nx_0)(1-a^{2t})}{r(1-a^2)+b^2(1-a^{2N})}, & b>0\\ a^tx_0 & ,b=0\end{cases}, \tag{3.44}$$

$$t=0,1,2,\cdots,N$$

and the cost function (3.30) is given by

$$J_0^*(x) = \begin{cases} J_0^* = \frac{1}{2} r \frac{(x_N - a^N x_0)^2 (1 - a^2)}{r(1 - a^2) + b^2(1 - a^{2N})}, b > 0 \\ 0 \qquad\qquad\qquad\qquad\qquad , b = 0 \end{cases} \tag{3.45}$$

ii. For $a = 1$ the optimal control can be represented as an open-loop control namely

$$u^*(t) = \begin{cases} \frac{b(x_N - x_0)}{r + b^2 N}, b > 0 \\ 0 \qquad\quad , b = 0 \end{cases}, t = 0,1,2,\cdots, N-1 \tag{3.46}$$

The optimal trajectory to (3.25), then, becomes

$$x^*(t) = \begin{cases} x_0 + \frac{b^2 N(x_N - x_0)}{r + b^2 N}, b > 0 \\ x_0 \qquad\qquad\qquad , b = 0 \end{cases}, t = 0,1,2,\cdots, N \tag{3.47}$$

and the cost function (3.30) is given by

$$J_0^*(x) = \begin{cases} \frac{1}{2} r \frac{(x_N - x_0)^2}{r + b^2 N}, b > 0 \\ 0 \qquad\qquad\quad , b = 0 \end{cases} \tag{3.48}$$

3.3.3 Numerical example

Example 3.4. To illustrate the approach adopted in this paper we consider the following simple minimum energy problem with fixed final state.

Minimize

$$J = \frac{1}{2} \sum_{t=0}^{3} u^2(t), \tag{3.49}$$

subject to

$$x(t+1) = x(t) + u(t), t = 1, 2, \cdots, 3, \tag{3.50}$$

$$u(t) \in \mathbb{R}_+, \tag{3.51}$$

$$x(0) = x_0 = 1, \text{ and } x(4) = x_4 = 5 \tag{3.52}$$

The 4-steps reachable set for positive discrete-time linear system (3.50) - (3.52) is $\mathcal{R}_4(1) = (1,\ \infty]$ and the final state $x_4 = 5$ is an interior point of $\mathcal{R}_4(1)$. Note also that the system is reachable and stable but not asymptotically [44, 15].

Using expression (3.21) and (3.22) and the state equation (3.2) in Section 3.2.2 we obtain the optimal control sequence

$$u^*(0) = 1, u^*(1) = 2, u^*(3) = 3,$$

the corresponding optimal trajectory

$$x_0 = 1, x^*(1) = 2, x^*(2) = 3, x^*(3) = 4, x(4) = x_4 = 5,$$

and the optimal cost function

$$J_0^* = 2.$$

By relaxing the boundary conditions $x_4 = 5$ we reduce the problem (3.49) - (3.52) to the following minimum energy problem with free final state.

Minimize

$$J = \frac{1}{2}\left(5 - x(4)\right)^2 + \frac{1}{2}\sum_{t=0}^{3} u^2(t), \qquad (3.53)$$

subject to

$$x(t+1) = x(t) + u(t), t = 1, 2, \cdots, 3, \qquad (3.54)$$

$$u(t), \qquad (3.55)$$

$$x(0) = x_0 = 1 \qquad (3.56)$$

To determine the optimal control sequence we use expression (3.28). For $t = 0$ we have

$$u^*(0) = \frac{4}{5},$$

and using the state equation (3.54) we find the next state $x^*(1) = \frac{9}{5}$. Consequently, we obtain

$$u^*(1) = \frac{4}{5} \text{ and } x^*(2) = \frac{13}{5},$$

$$u^*(2) = \frac{4}{5} \text{ and } x^*(3) = \frac{17}{5},$$

and finally,

$$u^*(3) = \frac{4}{5} \text{ and } x^*(4) = \frac{21}{5}.$$

We see that the condition $a^{N-t}x^*(t) \leq x_N$ is satisfied on the optimal trajectory. So, the optimal control sequence is $u^* = \left\{\frac{4}{5}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}\right\}$, the corresponding optimal trajectory $x^*(t) = \left\{1, \frac{9}{5}, \frac{13}{5}, \frac{17}{5}, \frac{21}{5}\right\}$, and the optimal cost function $J_0^* = 1.6$.

The above results tells us that by relaxing the minimum energy problem (3.49) - (3.52) with fixed final state to the minimum energy problem (3.53) - (3.56) with free final state, we decrease the energy of the input (control) from 2 to 1.6 at the expense of not reaching the final state – the deviation from the desired final state 5 is $\frac{4}{5}$.

3.4 The minimum energy problem for positive linear continuous-time system

3.4.1 Problem formulation

The *minimum energy problem for scalar positive continuous-time linear systems with fixed final state* is formulated as follows [14]:

Minimize
$$J = \frac{1}{2}\int_{t=0}^{N} ru^2(t)dt \tag{3.57}$$

subject to
$$\dot{x}(t) = ax(t) + bu(t),\ t \in [0, N] \tag{3.58}$$

$$a \in \mathbb{R}, \text{and } b, r, u(t) \in \mathbb{R}_+,$$

where $x(t)$ is the state variable at time t, $u(t)$ is the control at time t, N is a finite-time horizon. The initial and the final states are given by

$$x(0) = x_0 \geq 0 \text{ and } x(N) = x_N \geq 0 \quad (3.59)$$

The solution of positive linear continuous-time systems (3.58)

$$\dot{x}(t) = ax(t) + bu(t),\ t \in [0, N],$$

has the form $x(t) = e^{at}x_0 + \int_0^t e^{a(t-\tau)} b\, u(\tau) d\tau$.

The exponential e^{at} will be positive for any $a \in \mathbb{R}$ and $t \in [0, N]$. Therefore, the system will remain nonnegative if $a \in \mathbb{R}$, and $b, u(t) \in \mathbb{R}_+$.

3.4.2 Main results

The solution of the minimum energy problem (3.57) – (3.59) will be obtained using the dynamic programming procedure. According to Lewis *et al* [18], there are two methods to solve continuous-time optimal control problems using dynamic programming - the discretization approach and the Hamilton-Jacobi-Bellman equation. We will use the discretization approach used to solve the minimum energy problem (3.57) – (3.59). To apply the dynamic programming for continuous-time system using the discretization approach, the control and trajectory must be quantised to some finite set of admissible values (admissible control and feasible trajectory). Since a finer quantisation is required to obtain more accurate results, the increasing of the number of calculation to find the accurate admissible controls and feasible trajectory is inevitable. Therefore, dynamic programming for discretised continuous-time systems is not quite often used to avoid the curse of dimensionality [18]. However, as an analytical solution for the discrete time has been obtained in previous section, the curse of dimensionality can be avoided.

To discretise the cost function (3.57), we can write

$$J = \frac{1}{2}\sum_{k=0}^{T-1} \int_{ks}^{(k+1)s} ru^2(t)dt, \quad (3.60)$$

where s is a time sampler and

$$T = \frac{N}{s}\,. \quad (3.61)$$

Using a first order approximation to each integral results in

$$J = \sum_{k=0}^{T-1} s\ ru^2(ks). \tag{3.62}$$

The discrete-time representation of (3.58) takes the form

$$x\big((k+1)s\big) = a_s x(ks) + b_s u(ks) \tag{3.63}$$

To discretise system (3.58), we will use a time sampler s and zero order hold [18, 57]. Lewis *et al* claim that this method is better than any other approximation method [18].

From (3.63) we have,

$$\begin{aligned} x\big((k+1)s\big) &= e^{a(k+1)s}x_0 + \int_0^{(k+1)s} e^{a((k+1)s-\tau)}b\,u(\tau) \\ &= e^{a(k+1)s}x_0 + \int_0^{ks} e^{a((k+1)s-\tau)}b\,u(\tau)d\tau \\ &+ \int_{ks}^{(k+1)s} e^{a((k+1)s-\tau)}b\,u(\tau)d\tau, \end{aligned} \tag{3.64}$$

and

$$x(ks) = e^{aks}x_0 + \int_0^{ks} e^{a(ks-\tau)}b\,u(\tau)d\tau. \tag{3.65}$$

Multiplying (3.65) by e^{as} results in

$$e^{as}x(ks) = e^{a(k+1)s}x_0 + \int_0^{ks} e^{a((k+1)s-\tau)}b\,u(\tau)d\tau. \tag{3.66}$$

Substitution of (3.66) into (3.64) leads to

$$x\big((k+1)s\big) = e^{as}x(ks) + \int_{ks}^{(k+1)s} e^{a((k+1)s-\tau)}b\,u(\tau)d\tau. \tag{3.67}$$

Since we use a sampler and zero order hold, the control $u(t)$ is a constant over the interval between any two consecutive sampling instants, i.e.

$$u(t) = u(ks), \text{ for } ks \le t < (k+1)s. \tag{3.68}$$

Taking into account (3.68) and using the substitution $v = s - \tau$, expression (3.67) can be rewritten as

$$x\big((k+1)s\big) = e^{as}x(ks) + \left(\int_0^{s} e^{av}b\,dv\right)u(ks). \tag{3.69}$$

To obtain more accurate results, the smaller time sampler should be used.

Let us define

$$a_s = e^{as}, \tag{3.70}$$

$$b_s = \left(\int_0^s e^{av} b\, dv\right) = \frac{b}{a}(e^{as} - 1), \tag{3.71}$$

$$r_s = rs, \tag{3.72}$$

$$x(k) \triangleq x(ks), \text{ and } u(k) \triangleq u(ks), \tag{3.73}$$

then the *discretised minimum energy problem* (3.57) – (3.59) becomes

Minimize
$$J = \frac{1}{2}\sum_{k=0}^{T-1}(r_s)u^2(k) \tag{3.74}$$

subject to
$$x(k+1) = a_s x(k) + b_s u(k),\ k = 0,1,\cdots,T, \tag{3.75}$$

$$a_s, b_s, r_s, u(k) \in \mathbb{R}_+,$$

$$x(0) = x_0 \geq 0 \text{ and } x(T) = x(N) = x_N \geq 0. \tag{3.76}$$

Theorem 3.3. Let $x_N \in R_N(x_0)$. Then the optimal control that minimizes the cost function (3.57) in the minimum energy problem (3.57) – (3.59) with fixed final state is given by

$$u^*(t) = \frac{2a}{b}\frac{e^{a(N-t)}\left(x_N - e^{a(N-t)}x^*(t)\right)}{(e^{2a(N-t)} - 1)}, t \in [0, N), \tag{3.77}$$

where $x^*(t)$ is the corresponding optimal trajectory, and the optimal value of the cost function (3.57) is

$$J_0^* = \frac{1}{2}r\frac{2a\left(x_N - e^{aN}x_0\right)^2}{b^2(e^{2aN} - 1)}. \tag{3.78}$$

Proof. Consider the discretised minimum energy problem (3.74) – (3.76). Since $x_N \in R_N(x_0)$, there exists a nonnegative control function that steers the system trajectory from x_0 to x_N. Therefore, there exists a nonnegative control sequence $\{u(0), u(1), \cdots, u(T-1)\}$ that steers the system (3.75) from the given initial state x_0 to the given final state x_N

(3.76). Hence, using the results in Corollary 3.1, the optimal control sequence $u^*(t) = u^*(ks) \triangleq u(k)$ is formulated as

$$u^*(k) = \frac{{a_s}^{T-(k+1)}\left(x_T - {a_s}^{T-k}x^*(k)\right)}{b_s \sum_{i=k}^{T-1} {a_s}^{T-(i+1)}}. \quad (3.79)$$

Substituting (3.70) – (3.71) into the above equation (3.79) results in

$$u^*(k) = \frac{e^{as(T-(k+1))}\left(x_T - e^{as(T-k)}x^*(k)\right)}{\frac{b}{a}(e^{as}-1)\sum_{i=k}^{T-1} e^{as(T-(i+1))}}.$$

The finite geometric series $\sum_{i=k}^{T-1} e^{as(T-(i+1))}$ can be represented as $\frac{\left(e^{2as(T-k)}-1\right)}{(e^{2as}-1)}$, and so the optimal control becomes

$$u^*(k) = \frac{e^{as(T-(k+1))}\left(x_N - e^{as(T-k)}x^*(k)\right)(e^{2as}-1)}{\frac{b}{a}(e^{as}-1)\left(e^{2as(T-k)}-1\right)}.$$

Hence, $$u^*(k) = \frac{e^{as(T-k)}\left(x_N - e^{as(T-k)}x^*(k)\right)}{\frac{b}{a}\left(e^{2as(T-k)}-1\right)}\frac{(e^{2as}+1)}{e^{as}}.$$

From (3.61) and (3.68), $u^*(t) = \frac{ae^{a(N-t)}\left(x_N - e^{a(N-t)}x^*(t)\right)}{b\left(e^{2a(N-t)}-1\right)}\frac{(e^{2as}+1)}{e^{as}}$

As $s \to 0$, the optimal control for the minimum energy problem (3.57) – (3.59) converges to

$$u^*(t) = \frac{2ae^{a(N-t)}\left(x_N - e^{a(N-t)}x^*(t)\right)}{b\left(e^{2a(N-t)}-1\right)}.$$

It is easy to see that the optimal control $u^*(t)$ is always non-negative since $\frac{a}{e^{2a(N-t)}-1} > 0$ and from Lemma 2.6, $x_N - e^{a(N-t)}x^*(t) \geq 0$ for $t \in [0, N)$.

Using the results of Theorem 3.1, the optimal cost of the discretised positive minimum energy problem (3.74) – (3.76) is

$$J_0^* = \frac{1}{2}r_s \frac{\left(x_N - {a_s}^T x_0\right)^2}{{b_s}^2 \sum_{k=0}^{T-1} {a_s}^{2(T-(k+1))}}, \; b > 0.$$

By substituting (3.70) – (3.71) into the above equation, the following result can be obtained

$$J_0^* = \frac{1}{2} rs \frac{\left(x_N - e^{asT} x_0\right)^2}{\left(\frac{b}{a}(e^{as}-1)\right)^2 \sum_{k=0}^{T-1} e^{2as(T-(k+1))}}.$$

The series $\sum_{i=k}^{T-1} e^{as(T-(i+1))}$ can be represented as $\frac{(e^{2asT}-1)}{(e^{2as}-1)}$.

Hence, $$J_0^* = \frac{1}{2} rs \frac{a^2\left(x_N - e^{asT} x_0\right)^2 (e^{2as}-1)}{b^2 (e^{as}-1)^2 (e^{2asT}-1)}.$$

From (3.61) and (3.68), the optimal cost is

$$J_0^* = \frac{1}{2} r \frac{a^2\left(x_N - e^{aN} x_0\right)^2 s(e^{as}+1)}{b^2 (e^{2aN}-1)(e^{as}-1)}.$$

Taking $\lim_{s\to 0} \frac{s(e^{as}+1)}{(e^{as}-1)}$ and applying l'Hospital's rule results in

$$J_0^* = \frac{1}{2} r \frac{2a\left(x_N - e^{aN} x_0\right)^2}{b^2 (e^{2aN}-1)}.$$ □

Remarks 3.4.

1. If x_N is not in the set of reachable states, then the two-point boundary value problem (3.58) – (3.59) is inconsistent and therefore, the minimum energy problem with a fixed final state (3.57) – (3.59) has no solution.

2. The optimal control law (3.77) is a feedback control since it depends on the current state. As a matter of fact, the optimal control can also be represented as an open loop control that depends on the initial and final states only as in the corollary below.

Corollary 3.3. Under the assumption of Theorem 3.2, the optimal control can be represented as an open-loop control namely

$$u^*(t) = \frac{2ae^{a(N-t)}(x_N - e^{aN} x_0)}{b(e^{2aN}-1)}, \text{ for } t \in [0, N) \tag{3.80}$$

and the corresponding optimal trajectory

$$x^*(t) = e^{at}x_0 + \frac{e^{a(N-t)}(e^{2at}-1)(x_N-e^{aN}x_0)}{(e^{2aN}-1)} \text{ for } t \in [0, N). \tag{3.81}$$

Proof. Consider the discretised problem (3.74) – (3.76). Using the results in Corollary 3.2, the optimal control sequence $u^*(t) = u^*(ks) \triangleq u(k)$ can be represented as

$$u^*(k) = \frac{{a_s}^{T-(k+1)}(x_N-{a_s}^T x_0)({a_s}^2-1)}{b_s({a_s}^{2T}-1)} .$$

Substituting (3.70) – (3.71) into the above expression results in

$$u^*(k) = \frac{e^{as(T-(k+1))}(x_N-e^{asT}x_0)(e^{2as}-1)}{\frac{b}{a}(e^{as}-1)(e^{2asT}-1)} .$$

Then from (3.61) and (3.68),

$$u^*(t) = \frac{ae^{a(N-t)}(x_N-e^{aN}x_0)(e^{as}+1)}{b(e^{2aN}-1)e^{as}} ,$$

and so as $s \to 0$ the control becomes

$$u^*(t) = \frac{2ae^{a(N-t)}(x_N-e^{aN}x_0)}{b(e^{2aN}-1)}.$$

The control $u^*(t)$ is always non-negative since $\frac{a}{e^{2a(N-t)}-1} > 0$ and from Lemma 2.5, $x_N - e^N x_0 \geq 0$ for $t \in [0, N)$.

The trajectory of system (3.58) can be represented as

$$x(t) = e^{at}x_0 + \int_0^t e^{a(t-\tau)} b\, u(\tau) d\tau .$$

Substituting the optimal control (3.80) into the above expression yields

$$x^*(t) = e^{at}x_0 + \int_0^t \frac{2ae^{a(t-\tau)}\ e^{a(N-\tau)}(x_N-e^{aN}x_0)}{(e^{2aN}-1)} d\tau.$$

Hence,

$$x^*(t) = e^{at}x_0 + \frac{e^{a(N-t)}(e^{2at}-1)(x_N-e^{aN}x_0)}{(e^{2aN}-1)}.$$

Lemma 2.5 tells us that $x_N - e^{aN}x_0 \geq 0$, and so it is clear that the optimal trajectory $x^*(t)$ is non-negative. □

Remarks 3.5.

1. If $b > 0$ and $x_N = e^{aN}x_0$, the optimal control (3.80) is zero and the corresponding optimal trajectory becomes $x(t) = e^{at}x_0$ for $t \in [0, N)$.
2. It is clear from (3.81) that the optimal trajectory ends at the desired final state $x^*(N) = x_N$.
3. If the initial state x_0 is zero, formula (3.80) and (3.81) become even simpler

$$u^*(t) = \frac{2ae^{a(N-t)}}{b(e^{2aN}-1)}x_N \, , \tag{3.82}$$

$$u^*(t) = \frac{2ae^{a(N-t)}}{b(e^{2aN}-1)}x_N, \tag{3.83}$$

$$x(t) = \frac{e^{a(N-t)}(e^{2at}-1)}{(e^{2aN}-1)}x_N, \tag{3.84}$$

and the optimal cost (3.78) becomes

$$J_0^* = \frac{1}{2}r\frac{2ax_N{}^2}{b^2(e^{2aN}-1)}$$

3.4.3 Numerical examples

To illustrate the results obtained in section 3.3.2, we consider the following minimum energy problem for scalar positive continuous-time systems with fixed final state and solve the problems using the formula obtained in the previous section.

Problem 3.5. Consider the following *minimum energy problem for positive linear continuous-time system with scalar control*.

Minimize $J = \frac{1}{2}\int_{t=0}^{N} 2u^2(t)dt$

subject to $\dot{x}(t) = 2x(t) + u(t), t \in [0, N]$

$u(t) \in \mathbb{R}_+,$

$x(0) = 1 \geq 0$ and $x(N) = 55 \geq 0$

From Corollary 3.3, the optimal control is formulated as

$$u^*(t) = 0.000539e^{4-2t},$$

and the optimal trajectory is

$$x^*(t) = e^{2t} + 0.0001348e^{4-2t}(e^{4t} - 1).$$

Graphically, the optimal control and the optimal trajectory are show in Figure 3.4.

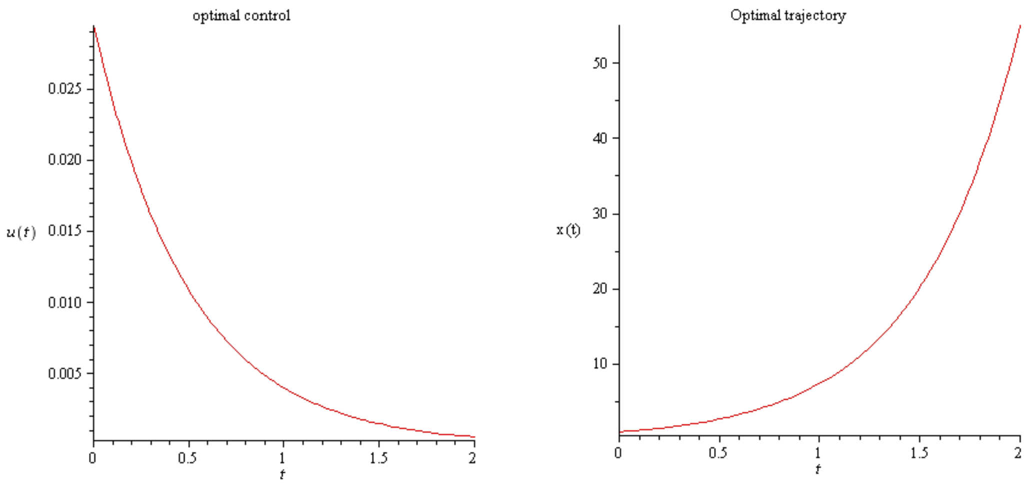

Figure 3.4. The optimal control and corresponding trajectory for Problem 3.5.

Problem 3.6. Consider the following *minimum energy problem for a positive linear continuous-time system with scalar control*.

Minimize $J = \frac{1}{2}\int_{t=0}^{N} 2u^2(t)dt$

subject to $\dot{x}(t) = \frac{1}{2}x(t) + u(t), t \in [0, N]$

$u(t) \in \mathbb{R}_+,$

$$x(0) = 1 \geq 0 \text{ and } x(N) = 55 \geq 0$$

Using the result from Corollary 3.3, the optimal control is

$$u^*(t) = \frac{e^{1-0.5t}(55-e)}{2(e^2-1)}, t \in [0, 2),$$

and the optimal trajectory is

$$x^*(t) = e^{0.5t} + \frac{e^{(1-0.5t)}(e^t - 1)(55 - e)}{(e^2 - 1)}.$$

Graphically, the optimal control and the optimal trajectory are as shown in Figure 3.5.

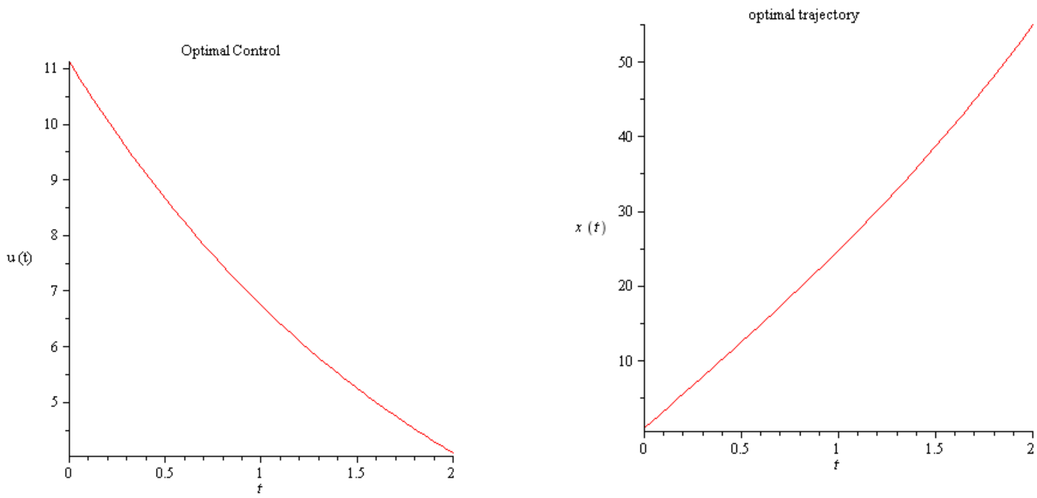

Figure 3.5. The optimal control and corresponding trajectory for Problem 3.6.

Problem 3.7. Consider the following *minimum energy problem for a positive linear continuous-time system with scalar control*.

Minimize $J = \frac{1}{2}\int_{t=0}^{N} 2u^2(t)dt$

subject to $\dot{x}(t) = -2x(t) + u(t), t \in [0, N]$

$u(t) \in \mathbb{R}_+,$

$x(0) = 1 \geq 0$ and $x(N) = 0.02 \geq 0$

The optimal control is

$$u^*(t) = \frac{4e^{-4+2t}(0.02-e)}{e^{-8}-1},$$

and the corresponding optimal trajectory is

$$x^*(t) = e^{at}x_0 + \frac{e^{a(N-t)}(e^{2at}-1)(x_N - e^{aN}x_0)}{e^{2aN}-1}$$

The optimal control and the corresponding trajectory is shown graphically in Figure 3.6.

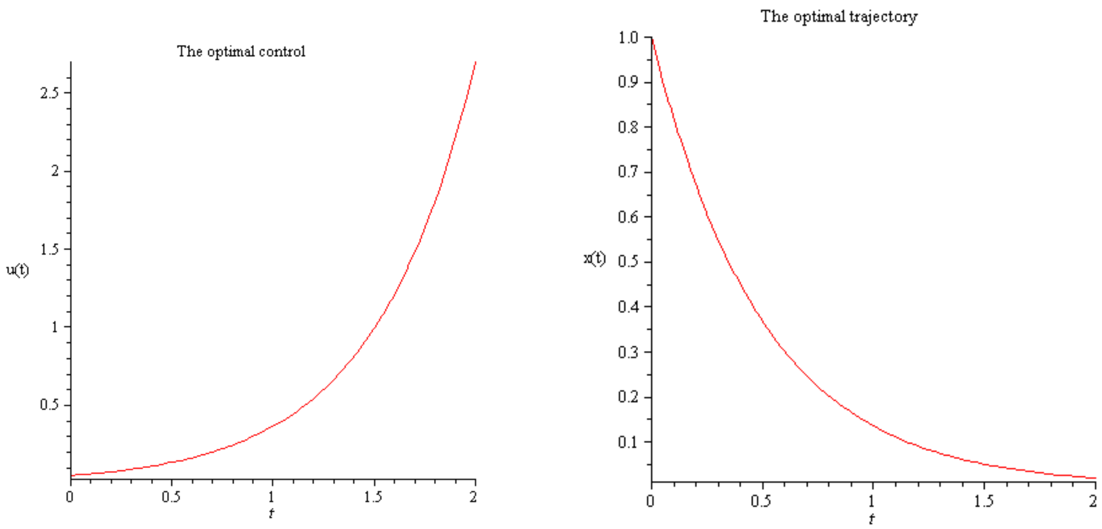

Figure 3.6. The optimal control and corresponding trajectory for Problem 3.7.

3.5 Concluding Remarks

In this chapter, the criteria for the existence of solutions to the minimum energy problem for scalar control have been established in continuous and discrete time. The sufficient conditions to guarantee the positivity of the problem and the analytical solutions to the problem have been obtained using dynamic programming. The relationship between the problem and the geometric properties of the system is well exploited. The optimal control of the problem is formulated in both open loop control and feedback control forms. The minimum energy problem does not have a solution if the final state does not belong to the *N-step reachable set*. The optimal solution becomes very simple if the system is controllable or the initial state is zero. The minimum energy problem for positive linear discrete-time systems with fixed final state can be reduced to a minimum energy problem with free final state by including in the cost function a term that reflects the deviation of the final state in the reduced problem from the targeted final state. Using dynamic programming approach, an analytical solution of the reduced minimum energy problem with free final state is obtained and analysed. It is shown that the relaxation of the problem leads to a decrease of the consumed energy of the input but at the expense of not reaching the desired final state. Such a "trade-off" might be quite appealing in a number of real-life problems.

CHAPTER 4

The minimum energy problem for positive linear system with vector control

4.1 General

In this chapter, we study the minimum energy problem for positive linear discrete-time and continuous-time systems with vector control and fixed initial and final states. The main objective of the study is to obtain the sufficient conditions to guarantee the nonnegativity of the control that steers the initial state to the final state. The minimum energy problem for positive linear systems has been studied previously by [22] and [14]. Beauthier [22] established the sufficient condition on the penalty matrix to guarantee the positivity. Kaczorek [14] established the nonnegativity of the problem with a monomial gramian. Moreover, the problem discussed in [14] is restricted to zero initial state. Hence, this chapter aims to extent the previous work to less restrictive conditions. The optimality conditions are established using dynamic programming. The results obtained are represented as feedback and open-loop forms.

4.2 The minimum energy problem for positive linear discrete-time system

4.2.1 Problem formulation

The minimum energy problem for vector positive discrete-time linear systems (PDLS) with fixed initial and final state is formulated as follows [14]:

Minimize $$J = \frac{1}{2}\sum_{t=0}^{N-1}\left(\mathbf{u}^T(t)\mathbf{R}\mathbf{u}(t)\right), \tag{4.1}$$

subject to

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), t = 0, \cdots, N-1, \tag{4.2}$$

$$\mathbf{A} \in \mathbb{R}_+^{n\times n}, \mathbf{B} \in \mathbb{R}_+^{n\times m}, \text{ and } \mathbf{u}(t) \in \mathbb{R}_+^m, \tag{4.3}$$

where $\mathbf{x}(t) \in \mathbb{R}_+^n$ is the state vector at time $t = 0,1,2,\cdots,N$, $\mathbf{u}(t) \in \mathbb{R}_+^m$ is the control vector sequence, the weighting matrix $\mathbf{R}$ is a symmetric positive definite matrix, N is a finite time horizon, and the initial and terminal (final) states are given by

$$\mathbf{x}(0) = \mathbf{x}_0 \text{ and } \mathbf{x}(N) = \mathbf{x}_N. \tag{4.4}$$

The optimal control problem (4.1) and (4.4) is named *the minimum energy problem for positive linear discrete-time systems with fixed initial and final states.*

The solution (in open-loop form) to *the minimum energy problem without the constraints* (4.3) (derived by Lagrange-multipliers approach) is well known (see, for example, [18]).

Kaczorek [14] has considered the minimum energy problem for positive linear discrete-time systems with fixed terminal state under the following assumptions.

a) The PLDS (Positive Linear Discrete-time System) (4.2) – (4.3) is reachable, that is, the reachability matrix $\mathfrak{R}_n$ contains a monomial sub matrix [44];
b) The initial state $\mathbf{x}(0) = 0$;
c) The inverse of the matrix

$$\mathbf{W} = [\mathbf{B} \vdots \mathbf{AB} \vdots \cdots \vdots \mathbf{A}^{N-1}\mathbf{B}] \begin{bmatrix} \mathbf{R}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{R}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}^T \\ \mathbf{B}^{\mathbf{T}}\mathbf{A}^T \\ \vdots \\ \mathbf{B}^T(\mathbf{A}^T)^{N-1} \end{bmatrix}$$

$$= \mathfrak{R}_N \, diag\, [\mathbf{R}^{-1}, \cdots, \mathbf{R}^{-1}] \mathfrak{R}_N{}^T \tag{4.5}$$

is a nonnegative matrix that is $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$;

d) $\mathbf{R}^{-1} \in \mathbb{R}_+^{n\times n}$.

Under the assumption (a) – (d), the optimal control is given by

$$\mathbf{u}_k = \begin{bmatrix} \mathbf{u}(k-1) \\ \mathbf{u}(k-2) \\ \vdots \\ \mathbf{u}(0) \end{bmatrix} = diag[\mathbf{R}^{-1}, \cdots, \mathbf{R}^{-1}]\mathfrak{R}_k{}^T \mathbf{W}^{-1}\mathbf{x}_N \tag{4.6}$$

for $k = 1,2, \cdots, N-1$.

Formula (4.6) can be written as

$$\mathbf{u}(k) = \mathbf{R}^{-1}\mathbf{B}^T(\mathbf{A}^T)^{N-1-k}\mathbf{W}^{-1}\mathbf{x}_N \text{ for } k = 1,2, \cdots, N-1. \tag{4.7}$$

While assumption (a) is needed for the minimum energy problem for (not positive) linear discrete-time systems, the assumptions (b), (c), and (d) above seem to be more restrictive. The initial state more often than not is nonzero so that (b) is quite restrictive. The control weighting matrix **R** is a symmetric positive definite matrix and therefore its inverse, i.e., $\mathbf{R}^{-1}$ and is also positive definite but $\mathbf{R}^{-1}$ is not necessarily a nonnegative matrix. Indeed, the only class of matrices, which have a nonnegative inverse, is the class of monomial matrices [58]. Furthermore, since **R** is a symmetric matrix the only class of matrices with a nonnegative inverse is the class of non-singular non-negative diagonal matrices, so that the assumption (d) is also very restrictive. The inverse $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ if and only if **W** is once again a monomial matrix, so that the assumption (c) is also satisfied in quite a few case.

4.2.2 Main results

Consider the minimum energy problem defined in (4.1) – (4.4). Using the dynamic programming approach, we will solve the problem analytically. Initially, we will not consider the nonnegativity constraints on the procedure. But, we will impose some conditions to guarantee the nonnegativity when the final results are derived.

Theorem 4.1. If the linear discrete-time system (4.2) (not necessarily positive system) is reachable and $\text{rank}(\mathbf{B}) = n$ then the optimal feedback control sequence that minimizes the cost function (4.1) in the minimum energy problem (4.1) – (4.4) with fixed final state is given by

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T \left(\sum_{i=t}^{N-1}(\mathbf{A}^{N-i-1}\mathbf{B})\, \mathbf{R}^{-1}(\mathbf{A}^{N-i-1}\mathbf{B})^T\right)^{-1} (\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}^*(t)), \tag{4.8}$$

for $t = 0,1,\cdots,N-1$,

and the optimal value of the cost function (4.1) is

$$J_0^*(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)^T, \tag{4.9}$$

where **W is** defined in (4.5), $\mathbf{x}^*(t)$ is the corresponding optimal trajectory.

Proof. To find the solution that minimises the cost function (4.1) we apply the dynamic programming procedure [20].

The Bellman equation for the minimum energy problem (4.1) – (4.4) can be written as

$$J_t^*(\mathbf{x}(t)) = \min_{\mathbf{u}(t)\in\mathbb{R}_m}\left\{\frac{1}{2}\mathbf{u}(t)^T\mathbf{R}\mathbf{u}(t) + J_{t+1}^*(\mathbf{x}(t+1))\right\}, t = 0,1,\cdots,N-1,$$

with $J_N^*(\mathbf{x}_N) = 0$ and $\mathbf{x}_N$ is fixed.

Moving backwards, we try for $t = N-1$.

The Bellman equation can be written as

$$J_{N-1}^*(x) = \min_{\mathbf{u}(N-1)\in\mathbb{R}_m}\left\{\frac{1}{2}\mathbf{u}^T(N-1)\mathbf{R}\mathbf{u}(N-1) + J_N^*(\mathbf{x})\right\},$$

where $J_N^*(\mathbf{x}) = 0$.

Since $\mathbf{x}_N$ is given and $\mathbf{x}_N = \mathbf{A}\mathbf{x}(N-1) + \mathbf{B}\mathbf{u}(N-1)$, then

$\mathbf{B}\mathbf{u}(N-1) = \boldsymbol{x}_N - \mathbf{A}\mathbf{x}(N-1)$.

Next the control $\mathbf{u}(N-1)$ will be formulated using Lagrange multiplier.

Let $L(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u} + \lambda(\mathbf{B}\mathbf{u} - \mathbf{x}_N + \mathbf{A}\mathbf{x})$, $\mathbf{x} = \mathbf{x}(N-1)$ and $\mathbf{u} = \mathbf{u}(N-1)$.

$\frac{\partial L}{\partial \mathbf{u}} = \mathbf{R}\mathbf{u} + \mathbf{B}^T\lambda = 0$ results in $\mathbf{u} = -\mathbf{R}^{-1}\mathbf{B}^T\lambda$.

$\frac{\partial L}{\partial \lambda} = \mathbf{B}\mathbf{u} - \mathbf{x}_N + \mathbf{A}\mathbf{x} = 0.$

Substituting $\mathbf{u} = -\mathbf{R}^{-1}\mathbf{B}^T\lambda$ into the above equation leads to

$$-\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\lambda - \mathbf{x}_N + \mathbf{A}\mathbf{x} = 0.$$

Thus,

$\lambda = -(\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T)^{-1}(\mathbf{x}_N - \mathbf{A}\mathbf{x}).$

Therefore, $\mathbf{u}(N-1) = \mathbf{R}^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T)^{-1}\big(\mathbf{x}_N - \mathbf{A}\mathbf{x}(N-1)\big)$

Then, the optimal control can be formulated as

$\mathbf{R}\mathbf{u}(N-1) = \mathbf{B}^T(\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T)^{-1}\big(\boldsymbol{x}_N - \mathbf{A}\mathbf{x}(N-1)\big)$.

From Theorem 2.3, $\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T$ is a positive definite matrix, therefore, $(\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T)^{-1}$ exists. Moreover, since $J_{N-1}^*(x)$ is quadratic, there exists a unique control $\mathbf{u}(N-1)$. Therefore, the optimal control $\mathbf{u}^*(N-1)$ is

$\mathbf{u}^*(N-1) = \mathbf{R}^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T)^{-1}\big(\mathbf{x}_N - \mathbf{A}\mathbf{x}(N-1)\big)$,

and the optimal cost is

$$J_{N-1}^*(\mathbf{x}) = \frac{1}{2}\big(\boldsymbol{x}_N - \mathbf{A}\mathbf{x}(N-1)\big)^T(\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T)^{-1}\big(\mathbf{x}_N - \mathbf{A}\mathbf{x}(N-1)\big).$$

After doing some iterations for $t = N-2$ and $N-3$, the induction hypothesis can be formulated as

$J_t^*(\mathbf{x}) = \frac{1}{2}\big(\boldsymbol{x}_N - \mathbf{A}^{N-t}\mathbf{x}(t)\big)^T\left(\sum_{i=t}^{N-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)\,\mathbf{R}^{-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)^T\right)^{-1}\big(\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}(t)\big)$,

and

$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T\left(\sum_{i=t}^{N-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)\,\mathbf{R}^{-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)^T\right)^{-1}\big(\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}(t)\big)$.

Assume that the hypothesis is true for $t = k+1$ that is

$J_{k+1}^*(\mathbf{x}) = \frac{1}{2}\Big(\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x}(k+1)\Big)^T\left(\sum_{i=k+1}^{N-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)\,\mathbf{R}^{-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)^T\right)^{-1}$

$\Big(\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x}(k+1)\Big)$,

and

$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^T\left(\sum_{i=k+1}^{N-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)\,\mathbf{R}^{-1}\big(\mathbf{A}^{N-1-i}\mathbf{B}\big)^T\right)^{-1}\Big(\mathbf{x}_N -$

$\mathbf{A}^{N-k-1}\mathbf{x}(k+1)\Big)$.

We will prove that the hypothesis is true for $t = k$.

The Bellman equation can be written as

$$J_k^*(\mathbf{x}) = \min_{\mathrm{u}(k)\in\mathbb{R}_m}\left\{\frac{1}{2}\mathbf{u}(k)^T\mathbf{R}\mathbf{u}(k) + J_{k+1}^*(\mathbf{x})\right\}.$$

Minimizing the function $J_k^*(\mathbf{x})$ results in

$\mathbf{Ru}(k) = (\mathbf{A}^{N-k-1}\mathbf{B})^T \left(\sum_{i=k}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T\right)^{-1}\left(\boldsymbol{x}_N - \mathbf{A}^{N-k}\mathbf{x}(k)\right).$

Since $J_k^*(\mathbf{x})$ is quadratic, there exists a unique control $\mathbf{u}(k)$.

Therefore, the optimal control $\mathbf{u}^*(k)$ is formulated as,

$\mathbf{u}^*(k) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-1}\mathbf{B})^T \left(\sum_{i=k}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T\right)^{-1}(\mathbf{x}_N - \mathbf{A}^{N-k}\mathbf{x}),$

and the optimal cost is

$J_k^*(\mathbf{x}) = \frac{1}{2}(\boldsymbol{x}_N - \mathbf{A}^{N-k}\mathbf{x})^T \left(\sum_{i=k}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T\right)^{-1}(\mathbf{x}_N - \mathbf{A}^{N-k}\mathbf{x}),$

where $\mathbf{x} = \mathbf{x}(k)$.

Hence, the hypothesis formulas are true for $t = k$ and they are true by induction for any t,

$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T \left(\sum_{i=t}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T\right)^{-1}(\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}),\ \mathbf{u}^*(t) \in \mathbb{R}_m,$

and

$J_t^*(\mathbf{x}) = \frac{1}{2}\left(\boldsymbol{x}_N - \mathbf{A}^{N-t}\mathbf{x}(t)\right)^T \left(\sum_{i=t}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T\right)^{-1}\left(\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}(t)\right).$

Finally, for $t = 0$, the optimal cost is formulated as

$J_0^*(\mathbf{x}) = \frac{1}{2}(\boldsymbol{x}_N - \mathbf{A}^{N}\mathbf{x}_0)^T \left(\sum_{i=0}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T\right)^{-1}(\mathbf{x}_N - \mathbf{A}^{N}\mathbf{x}_0),$

and the optimal control is

$\mathbf{u}^*(0) = \mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^T \left(\sum_{i=0}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T\right)^{-1}(\mathbf{x}_N - \mathbf{A}^{N}\mathbf{x}_0).$

From (4.5), the expression of the gramian $\sum_{i=0}^{N-1}(\mathbf{A}^{N-1-i}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-1-i}\mathbf{B})^T$ is defined as $\mathbf{W}$. Hence, $J_0^*(\mathbf{x}) = \frac{1}{2}(\boldsymbol{x}_N - \mathbf{A}^{N}\mathbf{x}_0)^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^{N}\mathbf{x}_0)$.

Theorem 4.2. If the linear discrete-time system (4.2) (not necessarily positive system) is reachable and rank $(\mathbf{B}) = m$, for $n > m$, then the optimal feedback control sequence that minimizes the cost function (4.1) in the minimum energy problem (4.1) – (4.4) with fixed final state is given by

i.
$$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T \left(\sum_{i=t}^{N-1} (\mathbf{A}^{N-i-1}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-i-1}\mathbf{B})^T\right)^{-1} (\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}^*(t)), \quad (4.10)$$

$t = 0,1,\cdots,k-1$, such that $(N-t)m \geq n$

and

ii.
$$[\mathbf{x}_N - \mathbf{A}^{N-k}\mathbf{x}^*(k)] = [\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-k-1}\mathbf{B}] \begin{bmatrix} \mathbf{u}^*(N-1) \\ \mathbf{u}^*(N-2) \\ \vdots \\ \mathbf{u}^*(k) \end{bmatrix} \quad (4.11)$$

such that $(N-k)m < n \leq (N-(k-1))m$,

where $\mathbf{x}^*(t)$ is the corresponding optimal trajectory.

***Proof*.** i. In the case of $n > m$ and rank $(\mathbf{B}) = m$, the optimal control of the minimum energy problem (4.1) – (4.4) cannot be determined using the formula (4.8). Recall the formula (4.8),

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T \left(\textstyle\sum_{i=t}^{N-1} (\mathbf{A}^{N-i-1}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-i-1}\mathbf{B})^T\right)^{-1} (\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}^*(t)).$$

For $t = N-1$, the gramian matrix can be represented as

$$\mathbf{BR}^{-1}\mathbf{B}^T.$$

From Theorem 2.3, the matrix $\mathbf{BR}^{-1}\mathbf{B}^T$ is positive semidefinite as rank $(\mathbf{B}) = m$. Hence, the inverse does not exist.

In general, the inverse of the grammian $\sum_{i=t}^{N-1} (\mathbf{A}^{N-i-1}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-i-1}\mathbf{B})^T$ exists if and only if the gramian is a positive definite matrix. The gramian can be written as

$$\sum_{i=t}^{N-1} (\mathbf{A}^{N-i-1}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-i-1}\mathbf{B})^T =$$

$$[\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-t-1}\mathbf{B}] \begin{bmatrix} \mathbf{R}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}^T \\ (\mathbf{AB})^T \\ \vdots \\ (\mathbf{A}^{N-t-1}\mathbf{B})^T \end{bmatrix}$$

It is clear that rank$\left(\begin{bmatrix} \mathbf{R}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}^{-1} \end{bmatrix}\right) = (N-t)m$. Therefore, the above gramian is positive definite if $(N-t)m \geq n$ and rank($[\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-t-1}\mathbf{B}]) = n$ (i.e., $[\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-t-1}\mathbf{B}]$ has full row rank).

Define

$$\mathfrak{R}_l = [\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{l-1}\mathbf{B}] \text{ for } l = 1, \cdots, N. \tag{4.12}$$

As l increases by one unit, the rank of matrix $\mathfrak{R}_l$ either increases or remain constant. The $\mathfrak{R}_l$ will remain constant if rank($\mathfrak{R}_l$) reaches the maximum rank [5].

From (4.12), $\mathfrak{R}_N = [\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-k}\mathbf{B} \quad \mathbf{A}^{N-k+1}\mathbf{B} \quad \cdots \quad \mathbf{A}^{N-1}\mathbf{B}]$ can be written as

$\mathfrak{R}_N = [\mathfrak{R}_{N-(k-1)} \quad \mathbf{A}^{N-k+1}\mathbf{B} \quad \cdots \quad \mathbf{A}^{N-1}\mathbf{B}]$, where $(N-k)m < n \leq \big(N - (k-1)\big)m$.

Rank($\mathfrak{R}_N$) $= n$ for any $N \geq n$ if rank($\mathfrak{R}_{N-k+1}$) $= n$. Hence, the maximum rank of the reachability matrix is reached at $t = k-1$, where $\big(N-(k-1)\big)m \geq n$ and the rank of the gramian will remain constant for any $t = 0, 1, \cdots, k-1$ such that the optimal control (4.8) can be determined.

Therefore, the optimal control sequence $\{\mathbf{u}^*(0), \mathbf{u}^*(1), \cdots, \mathbf{u}^*(k-1)\}$ and the optimal trajectory $\{\mathbf{x}^*(0), \mathbf{x}^*(1), \cdots, \mathbf{x}^*(k)\ \}$ can be obtained using formula (4.8).

ii. Since $\mathbf{x}_N$ belongs to the reachable sets $\mathcal{R}_N(\mathbf{x}_0)$, the admissible controls $\mathbf{u}(k), \mathbf{u}(k+1), \cdots, \mathbf{u}(N-1)$ exist and can be represented by the following equation

$$[\mathbf{x}_N - \mathbf{A}^{N-k}\mathbf{x}^*(k)] = [\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-k-1}\mathbf{B}] \begin{bmatrix} \mathbf{u}(N-1) \\ \mathbf{u}(N-2) \\ \vdots \\ \mathbf{u}(k) \end{bmatrix}, \tag{4.13}$$

where $\mathbf{x}^*(k)$ *is the corresponding optimal trajectory* at $t = k$.

If the rank of the matrix $\mathfrak{R}_l$ increases by m as l increases by one unit for $l < n$, in other words $\text{rank}(\mathfrak{R}_l) = lm$ for $l < n$, then $\text{rank}([\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-k-1}\mathbf{B}]) = m(N-k)$ (i.e. full column rank) for $(N-k)m < n$.

As the linear system (4.13) is consistent and $\text{rank}([\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-k-1}\mathbf{B}]) = m(N-k)$, $\begin{bmatrix} \mathbf{u}(N-1) \\ \mathbf{u}(N-2) \\ \vdots \\ \mathbf{u}(k) \end{bmatrix}$ is unique.

Since $\mathbf{x}^*(k)$ is the optimal and optimal trajectory, the admissible control $\begin{bmatrix} \mathbf{u}(N-1) \\ \mathbf{u}(N-2) \\ \vdots \\ \mathbf{u}(k) \end{bmatrix}$ is optimal.

Therefore there exists a unique optimal control sequence $\{\mathbf{u}^*(k), \mathbf{u}^*(k+1), \cdots, \mathbf{u}^*(N-1)\}$ such that $[\mathbf{x}_N - \mathbf{A}^{N-k}\mathbf{x}^*(k)] = [\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{N-k-1}\mathbf{B}] \begin{bmatrix} \mathbf{u}^*(N-1) \\ \mathbf{u}^*(N-2) \\ \vdots \\ \mathbf{u}^*(k) \end{bmatrix}$,

where $\mathbf{x}^*(k)$ *is the corresponding optimal trajectory* at $t = k$. □

Next we will derive the open-loop optimal control from the feedback optimal control in result (4.8).

Theorem 4.3 . If the linear discrete-time system (4.2) (not necessarily positive system) is reachable then the optimal feedback control sequence (4.8) that minimizes the cost function (4.1) in the minimum energy problem (4.1) – (4.4) with fixed final state can be represented as an open loop control is

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0), \text{ for } t = 0,1,\cdots,N-1, \tag{4.14}$$

Proof. Consider the feedback control (4.8) in Theorem 4.1. Moving forwards the formula (4.8), we try for $t = 0$. The optimal control

$\mathbf{u}^*(0) = \mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^{\mathrm{T}} \left(\sum_{i=0}^{N-1} (\mathbf{A}^i\mathbf{B})\, \mathbf{R}^{-1} (\mathbf{A}^i\mathbf{B})^{\mathrm{T}} \right)^{-1} (\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$. Then by (4.2) $\mathbf{x}^*(1) = \mathbf{A}\mathbf{x}_0 + \mathbf{B}\mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^{\mathrm{T}} \left(\sum_{i=0}^{N-1} (\mathbf{A}^i\mathbf{B})\, \mathbf{R}^{-1} (\mathbf{A}^i\mathbf{B})^{\mathrm{T}} \right)^{-1} (\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$. Hence, for $t = 1$, $\mathbf{u}^*(1) = \mathbf{R}^{-1}(\mathbf{A}^{N-2}\mathbf{B})^{\mathrm{T}} \left(\sum_{i=0}^{N-2} (\mathbf{A}^i\mathbf{B})\, \mathbf{R}^{-1} (\mathbf{A}^i\mathbf{B})^{\mathrm{T}} \right)^{-1} \left(\mathbf{x}_N - \mathbf{A}^{N-1}\mathbf{x}^*(1)\right)$ results in

$\mathbf{u}^*(1) = \mathbf{R}^{-1}(\mathbf{A}^{N-2}\mathbf{B})^{\mathrm{T}} \left(\sum_{i=0}^{N-1} (\mathbf{A}^i\mathbf{B})\, \mathbf{R}^{-1} (\mathbf{A}^i\mathbf{B})^{\mathrm{T}} \right)^{-1} (\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$. Therefore, we formulate the following induction hypothesis

$$\mathbf{u}^*(\mathrm{t}) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0). \qquad (4.15)$$

where $\mathbf{W}$ is described in (4.5) and can be represented as

$$\mathbf{W} = \sum_{i=0}^{N-1} (\mathbf{A}^i\mathbf{B})\, \mathbf{R}^{-1} (\mathbf{A}^i\mathbf{B})^{\mathrm{T}}.$$

Since system (4.2) is reachable, rank($\mathfrak{R}_N$) $= n$. Hence, from Theorem 2.3, it is clear that $\mathbf{W}$ is a positive definite matrix and its inverse, i.e. $\mathbf{W}^{-1}$, exists.

Assume that the formula (4.15) is true for time $t = k$, i.e.

$$\mathbf{u}^*(k) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-1}\mathbf{B})^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0). \qquad (4.16)$$

Then we will show that the formula (4.16) is also true for time $t = k + 1$, i.e.

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

From (4.8), the optimal control at time $t = k + 1$ is

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}} \left(\sum_{i=0}^{N-k-2} (\mathbf{A}^i\mathbf{B})\, \mathbf{R}^{-1} (\mathbf{A}^i\mathbf{B})^{\mathrm{T}} \right)^{-1} (\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x})$$

for $\mathbf{x} = \mathbf{x}(k+1)$.

If the optimal control at time *k* is $\mathbf{u}^*(k)$ then the optimal state at time *k* is

$$\mathbf{x}^*(k) = \mathbf{A}^k x_0 + \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j}\mathbf{B}\mathbf{u}^*(j)\,.$$

Therefore $\mathbf{x}^*(k+1) = \mathbf{A}^{k+1}\mathbf{x}_0 + \sum_{j=0}^{k} \mathbf{A}^{k-1-j}\mathbf{B}\mathbf{u}^*(j)$ and

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\left(\sum_{i=0}^{N-k-2}\left(\mathbf{A}^i\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^i\mathbf{B}\right)^{\mathrm{T}}\right)^{-1}\left(\mathbf{x}_N - \mathbf{A}^{N-k-1}\left(\boldsymbol{A}^{k+1}\mathbf{x}_0 + \sum_{j=0}^{k}\mathbf{A}^{k-j}\mathbf{B}\mathbf{u}^*(j)\right)\right). \tag{4.17}$$

From (4.16) $\mathbf{u}^*(j) = \mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$, then substituting it to (4.17) leads to

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\left(\sum_{i=0}^{N-k-2}\left(\mathbf{A}^i\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^i\mathbf{B}\right)^{\mathrm{T}}\right)^{-1}\left(\mathbf{x}_N - \mathbf{A}^{N-k-1}\left(\boldsymbol{A}^{k+1}\mathbf{x}_0 + \sum_{j=0}^{k}\mathbf{A}^{k-j}\mathbf{B}\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)\right)\right). \tag{4.18}$$

Simplifying (4.18) results in

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\left(\sum_{i=0}^{N-k-2}\left(\mathbf{A}^i\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^i\mathbf{B}\right)^{\mathrm{T}}\right)^{-1}\left(\mathbf{x}_N - \boldsymbol{A}^N\mathbf{x}_0 - \sum_{j=0}^{k}\mathbf{A}^{N-j-1}\mathbf{B}\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)\right).$$

The above formula can be represented as

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\left(\sum_{i=0}^{N-k-2}\left(\mathbf{A}^i\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^i\mathbf{B}\right)^{\mathrm{T}}\right)^{-1}\left(\mathbf{I} - \sum_{j=0}^{k}\mathbf{A}^{N-j-1}\mathbf{B}\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^{\mathrm{T}}\mathbf{W}^{-1}\right)(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

Since $\sum_{j=0}^{k} \mathbf{A}^{N-1-j}\mathbf{B}\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T = \sum_{i=N-k-1}^{N-1}\mathbf{A}^i\mathbf{B}\mathbf{R}^{-1}\left(\mathbf{A}^i\mathbf{B}\right)^T$ and $\mathbf{I} = \mathbf{W}\mathbf{W}^{-1}$,

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\left(\sum_{i=0}^{N-k-2}(\mathbf{A}^i\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^i\mathbf{B})^{\mathrm{T}}\right)^{-1}\left(\mathbf{W} - \sum_{j=0}^{k}\mathbf{A}^{N-j-1}\mathbf{B}\mathbf{Q}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^{\mathrm{T}}\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

Hence,

$$\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\left(\sum_{i=0}^{N-k-2}(\mathbf{A}^i\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^i\mathbf{B})^{\mathrm{T}}\right)^{-1}\left(\sum_{i=0}^{N-k-2}(\mathbf{A}^i\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^i\mathbf{B})^{\mathrm{T}}\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

It is clear that $\left(\sum_{i=0}^{N-k-2}(\mathbf{A}^i\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^i\mathbf{B})^{\mathrm{T}}\right)^{-1}\left(\sum_{i=0}^{N-k-2}(\mathbf{A}^i\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^i\mathbf{B})^{\mathrm{T}}\right) = \mathbf{I}$.

Therefore $\mathbf{u}^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-2}\mathbf{B})^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$.

Hence, the assumption (4.15) is true. □

Next, we will show that the optimal control obtained in Theorem 4.3 is nonnegative under certain conditions.

Theorem 4.4. Let $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$. If the positive linear discrete-time system (4.2) is reachable then the optimal control that minimizes the cost function (4.1) in the minimum energy problem (4.1) – (4.4) with fixed final state is

$$\mathbf{u}_+^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0), \tag{4.19}$$

In particular, $\mathbf{u}_+^*(t) \in \mathbb{R}_+^m$ for all $t = 0,1,\cdots,N-1$.

Proof. From Lemma 2.4, it is clear that if $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, then $\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0 \geq 0$.

If $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$, then from (4.14) we have

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$$

be nonnegative for all $t = 0,1,\cdots,N-1$. Therefore, the optimal control for the positive linear discrete-time system (4.2) is formulated as

$$\mathbf{u}_+^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)\text{ , for all } t = 0,1,\cdots,N-1 \qquad \square$$

Remarks 4.1.

1. The optimal trajectory corresponding to (4.19) is represented as

$$\mathbf{x}^*(t) = \mathbf{A}^t\mathbf{x}_0 + \mathfrak{R}_t diag[\mathbf{R}^{-1},\cdots,\mathbf{R}^{-1}]\begin{bmatrix}(\mathbf{A}^{N-t-1}\mathbf{B})^T\\ \vdots \\ (\mathbf{A}^{N-1}\mathbf{B})^T\end{bmatrix}\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0) \qquad (4.20)$$

   Hence,

$$\mathbf{A}^{N-t}\mathbf{x}^*(t) = \mathbf{A}^N\mathbf{x}_0 + \left(\sum_{j=0}^{t-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0) \qquad (4.21)$$

   $t = 0,1,\cdots,N.$

   It is easy to see that from (4.21) the optimal trajectory ends at the desired final state $\mathbf{x}^*(N) = \mathbf{x}_N$.

2. If the initial state $\mathbf{x}_0 = \mathbf{0}$, then the expression (4.21) becomes

   $\mathbf{u}_+^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T\mathbf{W}^{-1}\mathbf{x}_N,\ t = 0,1,\cdots,N-1.$

   which agrees with the result that obtained by Kaczorek [14].

3. The sufficient conditions in Theorem 4.4 for guaranteeing the nonnegativity of controls are clearly weaker than Kaczorek's sufficient conditions [14] because they do not require neither a monomial submatrix in the reachability matrix $\mathfrak{R}_n$ nor a zero initial.

Theorem 4.5. Let $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, rank$(\mathbf{B}) = n$, $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$. If the positive linear discrete-time system (4.2) is reachable then the optimal control that minimizes the cost function (4.1) in the minimum energy problem (4.1) – (4.4) with fixed final state can be formulated as a feedback control

$$\mathbf{u}_+^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T \left(\sum_{i=t}^{N-1}(\mathbf{A}^{N-i-1}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-i-1}\mathbf{B})^T\right)^{-1}(\mathbf{x}_N - \mathbf{A}^{N-t}\mathbf{x}^*(t)), \tag{4.22}$$

$\mathbf{u}_+^*(t) \in \mathbb{R}_+^m$ for all $t = 0,1,\cdots,N-1$.

Proof. From Lemma 2.4, it is clear that if $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, then $\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0 \geq 0$.

Consider the optimal control (4.22). If $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$, then it is clear that

$$\mathbf{u}_+^*(0) = \mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0) \tag{4.23}$$

be nonnegative.

For $t = 1$, the optimal control is formulated as

$$\mathbf{u}_+^*(1) = \mathbf{R}^{-1}(\mathbf{A}^{N-2}\mathbf{B})^T \left(\sum_{i=1}^{N-1}(\mathbf{A}^{N-i-1}\mathbf{B})\,\mathbf{R}^{-1}(\mathbf{A}^{N-i-1}\mathbf{B})^T\right)^{-1}(\mathbf{x}_N - \mathbf{A}^{N-1}\mathbf{x}^*(1)), \tag{4.24}$$

where $\mathbf{x}^*(1) = \mathbf{A}\mathbf{x}_0 + \mathbf{B}\mathbf{u}_+^*(0)$. Substituting (4.23) into (4.24) results in

$$\mathbf{x}^*(1) = \mathbf{A}\mathbf{x}_0 + \mathbf{B}\mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0). \tag{4.25}$$

Then multiplying (4.25) with $\mathbf{A}^{N-1}$ on both sides leads to

$$\mathbf{A}^{N-1}\mathbf{x}^*(1) = \mathbf{A}^N\mathbf{x}_0 + \mathbf{A}^{N-1}\mathbf{B}\mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

Hence, $\mathbf{x}_N - \mathbf{A}^{N-1}\mathbf{x}^*(1) = \mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0 - \mathbf{A}^{N-1}\mathbf{B}\mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$.

Consequently, $\mathbf{x}_N - \mathbf{A}^{N-1}\mathbf{x}^*(1) = (\mathbf{I} - \mathbf{A}^{N-1}\mathbf{B}\mathbf{R}^{-1}(\mathbf{A}^{N-1}\mathbf{B})^T\mathbf{W}^{-1})(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$, where $\mathbf{W}$ is defined in (4.5).

Therefore,

$$\mathbf{x}_N - \mathbf{A}^{N-1}\mathbf{x}^*(1) = \left(\sum_{i=1}^{N-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0) \tag{4.26}$$

Substituting (4.26) into (4.24) results in

$$\mathbf{u}_+^*(1) = \mathbf{R}^{-1}(\mathbf{A}^{N-2}\mathbf{B})^T\left(\sum_{i=1}^{N-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)^T\right)^{-1}$$

$$\left(\sum_{i=1}^{N-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

Since $\mathrm{rank}(\mathbf{B}) = n$, the matrix $\sum_{i=1}^{N-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)^T$ is positive definite.

Consequently,

$$\mathbf{u}_+^*(1) = \mathbf{R}^{-1}(\mathbf{A}^{N-2}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

If $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$, then it is clear that $\mathbf{u}_+^*(1) \in \mathbb{R}_+^m$.

Now we will prove that the optimal control (4.22) is nonnegative for all $= 0,1,\cdots,N-1$.

Assume that $\mathbf{u}_+^*(t) \in \mathbb{R}_+^m$ is true for $t = k$. Then we will show that $\mathbf{u}_+^*(t) \in \mathbb{R}_+^m$ is also true for $t = k+1$.

Let $\mathbf{u}_+^*(t) \in \mathbb{R}_+^m$ is true for $t = k$, then $\mathbf{u}_+^*(k) \in \mathbb{R}_+^m$ and $\mathbf{u}_+^*(k)$ can be represented as

$$\mathbf{u}_+^*(k) = \mathbf{R}^{-1}(\mathbf{A}^{N-k-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

From (4.22), the optimal control for $t = k+1$ is

$$\mathbf{u}_+^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k}\mathbf{B})^T\left(\sum_{i=k+1}^{N-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)^T\right)^{-1}\left(\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x}^*(k+1)\right), \tag{4.27}$$

From (4.21), the optimal trajectory at time $t = k + 1$ can be represented as

$$\mathbf{A}^{N-k-1}\mathbf{x}^*(k+1) = \mathbf{A}^N\mathbf{x}_0 + \left(\sum_{j=0}^{k}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0),$$

Therefore,

$$\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x}^*(k+1) = \mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0 - \left(\sum_{j=0}^{k}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

Hence,

$$\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x}^*(k+1) = \left(\mathbf{I} - \left(\sum_{j=0}^{k}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}\right)(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)\ .$$

Since $\mathbf{I} = \mathbf{W}\mathbf{W}^{-1}$,

$$\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x}^*(k+1) = \left(\mathbf{W} - \left(\sum_{j=0}^{k}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T\right)\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0)$$

and consequently

$$\mathbf{x}_N - \mathbf{A}^{N-k-1}\mathbf{x}^*(k+1) = \left(\sum_{j=k+1}^{N-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

Hence, equation (4.27) becomes

$$\mathbf{u}_+^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k}\mathbf{B})^T\left(\sum_{i=k+1}^{N-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)^T\right)^{-1}\left(\sum_{j=k+1}^{N-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-j-1}\mathbf{B}\right)^T\right)\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0),$$

Since $\text{rank}(\mathbf{B}) = n$, the matrix $\sum_{i=k+1}^{N-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{N-i-1}\mathbf{B}\right)^T$ is positive definite.

Therefore, $\mathbf{u}_+^*(k+1)$ can be formulated as

$$\mathbf{u}_+^*(k+1) = \mathbf{R}^{-1}(\mathbf{A}^{N-k}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0),$$

and it is clear that the assumption $\mathbf{u}_+^*(t) \in \mathbb{R}_+^m$ is true for $t = k + 1$ if $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$.

Hence, the optimal control (4.22) is nonnegative for all $t = 0,1,\cdots,N-1$. □

If the control of the the minimum energy problem (4.1) – (4.4) with fixed final state is scalar, then the optimal control is formulated in the following theorem.

Theorem 4.6. Let $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$. If the linear discrete-time system (4.2) is reachable and the finite horizon $N = n$ then the optimal control sequence that minimizes the cost function (4.1) in the minimum energy problem (4.1) – (4.4) with fixed final state is given by

$$u_+^*(t) = r^{-1}(\mathbf{A}^{N-t-1}b)^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0), t = 0,1,\cdots,N-1, \quad (4.28)$$

which is nonnegative for all $t = 0,1,\cdots,N-1$.

Proof. Consider the optimal control (4.28)

$$u_+^*(t) = r^{-1}(\mathbf{A}^{N-t-1}b)^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0), t = 0,1,\cdots,N-1.$$

From (4.5), the optimal control (4.28) can be represented as

$$\begin{bmatrix} u_+^*(N-1) \\ \vdots \\ u_+^*(0) \end{bmatrix} = \begin{bmatrix} r^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r^{-1} \end{bmatrix} \mathfrak{R}_N{}^T \left(\mathfrak{R}_N \begin{bmatrix} r^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r^{-1} \end{bmatrix} \mathfrak{R}_N{}^T \right)^{-1} (\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0) \quad (4.29)$$

It is clear that if the reachability matrix $\mathfrak{R}_N$ is a square matrix with full rank, i.e., rank $(\mathfrak{R}_N) = n$, then the inverse of $\mathfrak{R}_N$ exists. Therefore,

$$\begin{bmatrix} u_+^*(N-1) \\ \vdots \\ u_+^*(0) \end{bmatrix} = \mathfrak{R}_N{}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0),$$

and $\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0 = \mathfrak{R}_N \begin{bmatrix} u_+^*(N-1) \\ \vdots \\ u_+^*(0) \end{bmatrix}$.

Since $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, there exists nonnegative control $u(t) \in \mathbb{R}_+$ for $t = 0,1,\cdots,N-1$ such that

$$\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0 = \mathfrak{R}_N \begin{bmatrix} u(N-1) \\ \vdots \\ u(0) \end{bmatrix}. \quad (4.30)$$

It is mentioned in the reachability properties that rank $(\mathfrak{R}_N) = N$. Therefore, equation (4.30) has a unique solution $\begin{bmatrix} u(N-1) \\ \vdots \\ u(0) \end{bmatrix}$.

Hence, $\begin{bmatrix} u(N-1) \\ \vdots \\ u(0) \end{bmatrix} = \begin{bmatrix} u_+^*(N-1) \\ \vdots \\ u_+^*(0) \end{bmatrix}$ is nonnegative. □

If the final state is in the minimal generator of the reachability matrix positive linear discrete-time system (4.2), the optimal control of the minimum energy problem for positive linear discrete-time system (4.1) – (4.4) with fixed final state is become simpler.

Theorem 4.7 . Let $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, the reachability matrix $[\mathbf{B} \vdots \mathbf{AB} \vdots \cdots \vdots \mathbf{A}^{N-1}\mathbf{B}] = [\beta_1 \vdots \beta_2 \vdots \cdots \vdots \beta_{Nm}]$, and the control sequence $\begin{bmatrix} \mathbf{u}^*(N-1) \\ \mathbf{u}^*(N-2) \\ \vdots \\ \mathbf{u}^*(0) \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{Nm} \end{bmatrix}$ where β_i is the column vector of the reachability matrix and $\beta_i \in \mathbb{R}_+^n$, $i = 1,2,\cdots,Nm$. If the positive linear discrete-time system (4.2) is reachable in N-steps and $\mathbf{x}_N$ is on the minimal generator β, then $u_j = 0$ for any index j which $\beta_j \neq \beta$.

*Proof***.** From Theorem 4.3 the optimal control is

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{N-t-1}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0),\ \text{for } t = 0,1,\cdots,N-1.$$

The optimal control can be expressed as

$$\begin{bmatrix} \mathbf{u}^*(N-1) \\ \mathbf{u}^*(N-2) \\ \vdots \\ \mathbf{u}^*(0) \end{bmatrix} = \begin{bmatrix} \mathbf{R}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{R}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}^T \\ \mathbf{B}^T\mathbf{A}^T \\ \vdots \\ \mathbf{B}^T(\mathbf{A}^T)^{N-1} \end{bmatrix} \mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

Multiplying both sides by $[\mathbf{B} \vdots \mathbf{AB} \vdots \cdots \vdots \mathbf{A}^{N-1}\mathbf{B}]$ results in

$$[\mathbf{B} \vdots \mathbf{AB} \vdots \cdots \vdots \mathbf{A}^{N-1}\mathbf{B}] \begin{bmatrix} \mathbf{u}^*(N-1) \\ \mathbf{u}^*(N-2) \\ \vdots \\ \mathbf{u}^*(0) \end{bmatrix}$$
$$= [\mathbf{B} \vdots \mathbf{AB} \vdots \cdots \vdots \mathbf{A}^{N-1}\mathbf{B}] \begin{bmatrix} \mathbf{R}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{R}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}^T \\ \mathbf{B}^T\mathbf{A}^T \\ \vdots \\ \mathbf{B}^T(\mathbf{A}^T)^{N-1} \end{bmatrix} \mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0).$$

From (4.5),

$$[\mathbf{B} \vdots \mathbf{AB} \vdots \cdots \vdots \mathbf{A}^{N-1}\mathbf{B}] \begin{bmatrix} \mathbf{u}^*(N-1) \\ \mathbf{u}^*(N-2) \\ \vdots \\ \mathbf{u}^*(0) \end{bmatrix} = (\mathbf{x}_N - \mathbf{A}^N \mathbf{x}_0).$$

Since $\mathbf{x}(N) = \mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$ and $\mathbf{x}_N$ is on the minimal generator β, we get $\mathbf{x}_N - \mathbf{A}^N \mathbf{x}_0 = c\beta$ for any nonnegative constant *c*.

Since β is the minimal generator (i.e., β cannot be expressed as a positive linear combination of $\beta_i \in \mathbb{R}_+^n$, $i = 1,2,\cdots,Nm$) and $\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{Nm} \end{bmatrix} \in \mathbb{R}_+^{Nm}$,

$c\beta = \sum_{\beta_j \neq \beta} \beta_j u_j + \sum_{\beta_i = \beta} \beta_i u_i$ if and only if $u_j = 0$. □

4.2.3 Numerical Examples

To illustrate the approach adopted in this chapter, we consider the following examples on the minimum energy problem with fixed final state. All the following numerical example problems are solved using the formula obtained in this chapter.

Example 4.1. Consider the following minimum energy problem for positive discrete-time system with fixed final state:

Minimize $J = \frac{1}{2}\sum_{t=0}^{2}\left(\mathbf{u}^T(t)\begin{bmatrix}1 & 0 \\ 0 & 3\end{bmatrix}\mathbf{u}(t)\right)$,

subject to

$$\mathbf{x}(t+1) = \begin{bmatrix}1 & 0 \\ 0 & 2\end{bmatrix}\mathbf{x}(t) + \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\mathbf{u}(t),\ t = 0,1,2$$

$$\mathbf{x}(0) = \mathbf{x}_0 = \begin{bmatrix}1 \\ 1\end{bmatrix} \text{ and } \mathbf{x}(3) = \mathbf{x}_3 = \begin{bmatrix}2 \\ 9\end{bmatrix}$$

$$\mathbf{u}(t) \in \mathbb{R}_+$$

The reachability matrix of the problem is

$$\mathfrak{R}_2 = [\mathbf{B} \quad \mathbf{AB}] = \begin{bmatrix}1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 2\end{bmatrix} \tag{4.31}$$

and $rank(\mathfrak{R}_2) = 2$.

The *3-steps* reachable sets of the above problem is defined as

$$\mathcal{R}_3(\mathbf{x}_0) = \left\{\mathbf{x} \,\middle|\, \mathbf{x} = \begin{bmatrix}1\\8\end{bmatrix} + \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\mathbf{u}(2) + \begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}\mathbf{u}(1) + \begin{bmatrix}1 & 0\\0 & 4\end{bmatrix}\mathbf{u}(0)\,;\ \mathbf{u}(k) \in \mathbb{R}_+^m \text{ for } k = 0,1,2\right\}$$

$$\mathbf{x}_3 = \begin{bmatrix}2\\9\end{bmatrix} = \begin{bmatrix}1\\8\end{bmatrix} + \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\mathbf{u}(2) + \begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}\mathbf{u}(1) + \begin{bmatrix}1 & 0\\0 & 4\end{bmatrix}\mathbf{u}(0)$$

The above expression can be formulated as

$$\begin{bmatrix}1\\1\end{bmatrix} = \left[\begin{bmatrix}1\\0\end{bmatrix}u_1(2) + \begin{bmatrix}0\\1\end{bmatrix}u_2(2) + \begin{bmatrix}1\\0\end{bmatrix}u_1(1) + \begin{bmatrix}0\\2\end{bmatrix}u_2(1) + \begin{bmatrix}1\\0\end{bmatrix}u_1(0) + \begin{bmatrix}0\\4\end{bmatrix}u_2(0)\right]$$

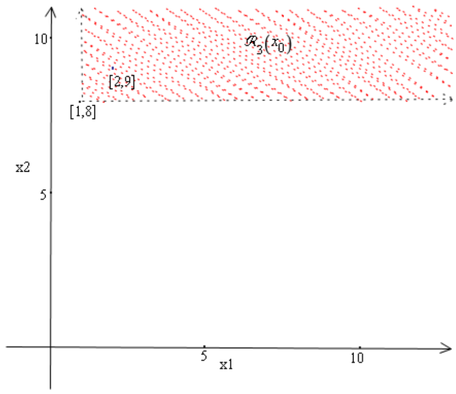

Figure 4.1. The reachable sets of Example 4.1.

It is clear from the Figure 4.1, that the final state $\mathbf{x}_3 = \begin{bmatrix}2\\9\end{bmatrix}$ is in the reachable set $\mathcal{R}_3(\mathbf{x}_0)$. Therefore, from Theorem 4.4, the optimal control is

$$\mathbf{u}_+^* = \mathbf{R}^{-1}(\mathbf{A}^{2-t}\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_N - \mathbf{A}^3\mathbf{x}_0),$$

for $t = 0,1,2$

The weighting matrix, $\mathbf{R}$, is clearly a monomial matrix and

$$\mathbf{W} = (\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T + (\mathbf{A}\mathbf{B})\mathbf{R}^{-1}(\mathbf{A}\mathbf{B})^T + (\mathbf{A}^2\mathbf{B})\mathbf{R}^{-1}(\mathbf{A}^2\mathbf{B})^T)^{-1} = \begin{bmatrix}3 & 0\\0 & 7\end{bmatrix}.$$

Hence, its inverse $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$, i.e., $\mathbf{W}^{-1} = \begin{bmatrix}0.333 & 0\\0 & 0.1429\end{bmatrix}$.

Therefore, for $\mathbf{x}_3 \in \mathcal{R}_3(\mathbf{x}_0)$, where $\mathbf{x}_3 = \begin{bmatrix}2\\9\end{bmatrix}$, the optimal controls at time $t = 0,1,2$ are

$$\mathbf{u}_+^*(2) = \mathbf{R}^{-1}(\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0) = \begin{bmatrix}0.3333\\0.0476\end{bmatrix},$$

$$\mathbf{u}_+^*(1) = \mathbf{R}^{-1}(\mathbf{AB})^T\mathbf{W}^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0) = \begin{bmatrix} 0.3333 \\ 0.0952 \end{bmatrix},$$

$$\mathbf{u}_+^*(0) = \mathbf{R}^{-1}(\mathbf{A}^2\mathbf{B})^T\mathbf{W}^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0) = \begin{bmatrix} 0.3333 \\ 0.1905 \end{bmatrix}.$$

The corresponding optimal trajectories are

$\mathbf{x}^*(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{x}^*(1) = \begin{bmatrix} 0.3333 \\ 0.0476 \end{bmatrix}, \mathbf{x}^*(2) = \begin{bmatrix} 1.6667 \\ 4.4762 \end{bmatrix}, \mathbf{x}^*(3) = \begin{bmatrix} 2 \\ 9 \end{bmatrix},$

and the optimal cost is

$J_0^*(\mathbf{x}_0) = 0.8095.$

It is clear from (4.31) and Figure 4.1 that the system is reachable and the final state is the reachable set $\mathcal{R}_3(\mathbf{x}_0)$. Since $rank(\mathbf{B}) = 2$, Theorem 4.5 also applies and the optimal control is

$$\mathbf{u}_+^*(t) = \mathbf{R}^{-1}(\mathbf{A}^{2-t}\mathbf{B})^T\left(\sum_{i=t}^{2}\left(\mathbf{A}^{2-i}\mathbf{B}\right)\mathbf{R}^{-1}\left(\mathbf{A}^{2-i}\mathbf{B}\right)^T\right)^{-1}\left(\mathbf{x}_3 - \mathbf{A}^{3-t}\mathbf{x}^*(t)\right)$$

for $t = 0,1,2$

- For $t = 0$,

  $(\mathbf{BR}^{-1}\mathbf{B}^T + (\mathbf{AB})\mathbf{R}^{-1}(\mathbf{AB})^T + (\mathbf{A}^2\mathbf{B})\mathbf{R}^{-1}(\mathbf{A}^2\mathbf{B})^T)^{-1} = \begin{bmatrix} 0.333 & 0 \\ 0 & 0.1429 \end{bmatrix}.$

  Hence,

  $$\mathbf{u}_+^*(0) = \mathbf{R}^{-1}(\mathbf{A}^2\mathbf{B})^T(\mathbf{BR}^{-1}\mathbf{B}^T + (\mathbf{AB})\mathbf{R}^{-1}(\mathbf{AB})^T + (\mathbf{A}^2\mathbf{B})\mathbf{R}^{-1}(\mathbf{A}^2\mathbf{B})^T)^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0) = \begin{bmatrix} 0.3333 \\ 0.1905 \end{bmatrix}$$

- For $t = 1$, $(\mathbf{BR}^{-1}\mathbf{B}^T + (\mathbf{AB})\mathbf{R}^{-1}(\mathbf{AB})^T)^{-1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.6 \end{bmatrix},$

  $\mathbf{x}^*(1) = \mathbf{Ax}_0 + \mathbf{Bu}_+^*(0) = \begin{bmatrix} 0.3333 \\ 0.0476 \end{bmatrix}$. Therefore,

  $\mathbf{u}_+^*(1) = \mathbf{R}^{-1}(\mathbf{AB})^T(\mathbf{BR}^{-1}\mathbf{B}^T + (\mathbf{AB})\mathbf{R}^{-1}(\mathbf{AB})^T)^{-1}\left(\mathbf{x}_3 - \mathbf{A}^2\mathbf{x}^*(1)\right) = \begin{bmatrix} 0.3333 \\ 0.0952 \end{bmatrix}$

- For $t = 2$, $(\mathbf{BR}^{-1}\mathbf{B}^T)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix},$

  $\mathbf{x}^*(2) = \mathbf{Ax}^*(1) + \mathbf{Bu}_+^*(1) = \begin{bmatrix} 1.6667 \\ 4.4762 \end{bmatrix}$. Therefore,

  $\mathbf{u}_+^*(2) = \mathbf{R}^{-1}\mathbf{B}^T(\mathbf{BR}^{-1}\mathbf{B}^T)^{-1}\left(\mathbf{x}_3 - \mathbf{Ax}^*(2)\right) = \begin{bmatrix} 0.3333 \\ 0.0476 \end{bmatrix}.$

The optimal cost is

$$J_0^*(\mathbf{x}_0) = 0.8095.$$

Example 4.2. Consider the following minimum energy problem with fixed final state:

Minimize $\quad J = \frac{1}{2}\sum_{t=0}^{N-1} ru^2(t),$

subject to

$$\mathbf{x}(t+1) = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} u(t),\ t = 0,1,2,$$

$$\mathbf{x}(0) = \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } \mathbf{x}(3) = \mathbf{x}_3 = \begin{bmatrix} 12 \\ 16 \\ 5 \end{bmatrix},$$

$$\mathbf{u}(t) \in \mathbb{R}_+.$$

From Theorem 4.6, the optimal control is

$$u_+^*(t) = r^{-1}(\mathbf{A}^{2-t}b)^T \mathbf{W}^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0),$$

for $t = 0,1,2$

$$\mathbf{W}^{-1} = (r^{-1}b^2 + (\mathbf{A}b)r^{-1}(\mathbf{A}b)^T + (\mathbf{A}^2b)r^{-1}(\mathbf{A}^2b)^T)^{-1} = \begin{bmatrix} 14 & -9 & -7 \\ -9 & 6 & 4 \\ -7 & 4 & 5 \end{bmatrix},$$

$$\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0 = \begin{bmatrix} 8 \\ 9 \\ 4 \end{bmatrix},$$

$$u_+^*(2) = r^{-1}b\mathbf{W}^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0)$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 14 & -9 & -7 \\ -9 & 6 & 4 \\ -7 & 4 & 5 \end{bmatrix} \begin{bmatrix} 8 \\ 9 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 8 \\ 9 \\ 4 \end{bmatrix} = 1.$$

$$u_+^*(1) = r^{-1}(\mathbf{A}b)^T \mathbf{W}^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0)$$

$$= \begin{bmatrix} 3 & -2 & -1 \end{bmatrix} \begin{bmatrix} 8 \\ 9 \\ 4 \end{bmatrix} = 2.$$

$$u_+^*(0) = r^{-1}(\mathbf{A}^2b)^T \mathbf{W}^{-1}(\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0)$$

$$= \begin{bmatrix} -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 8 \\ 9 \\ 4 \end{bmatrix} = 1.$$

The corresponding optimal trajectory

$$\mathbf{x}^*(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}^*(1) = \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}, \mathbf{x}^*(2) = \begin{bmatrix} 7 \\ 8 \\ 4 \end{bmatrix}, \mathbf{x}^*(3) = \begin{bmatrix} 12 \\ 16 \\ 5 \end{bmatrix}$$

Example 4.3. Consider the following minimum energy problem with fixed final state and $\mathbf{x}_N$ is on the minimal generator of $\boldsymbol{\mathcal{R}}_3(\mathbf{x}_0)$.

Minimize the objective function

$$J = \frac{1}{2}\sum_{t=0}^{2} \mathbf{u}^T(t) \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \mathbf{u}\,(t),$$

subject to

$$\mathbf{x}(t+1) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{u}(t),\ t = 0,1,2$$

$$\mathbf{x}(0) = \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{and } \mathbf{x}(3) = \mathbf{x}_3 = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

$$\mathbf{u}(t) \in \mathbb{R}_+$$

The *3-steps* reachable set for the PDLS above is

$$\boldsymbol{\mathcal{R}}_3(\mathbf{x}_0) = \left\{ \mathbf{x} \,\middle|\, \mathbf{x} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} + \left[\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{u}(2) + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \mathbf{u}(1) + \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \mathbf{u}(0) \right], \mathbf{u}(t) \in \mathbb{R}_+, t = 0,1,2 \right\}$$

The final state is fixed, i.e.,

$$\mathbf{x}_3 = \begin{bmatrix} 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} + \left[\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{u}(2) + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \mathbf{u}(1) + \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \mathbf{u}(0) \right].$$

$$\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \left[\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{u}(2) + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \mathbf{u}(1) + \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \mathbf{u}(0) \right].$$

The above expression can be formulated as

$$\begin{bmatrix}4\\4\end{bmatrix} = \left[\begin{bmatrix}1\\1\end{bmatrix}u_1(2) + \begin{bmatrix}0\\1\end{bmatrix}u_2(2) + \begin{bmatrix}1\\2\end{bmatrix}u_1(1) + \begin{bmatrix}0\\1\end{bmatrix}u_2(1) + \begin{bmatrix}1\\3\end{bmatrix}u_1(0) + \begin{bmatrix}0\\1\end{bmatrix}u_2(0)\right].$$

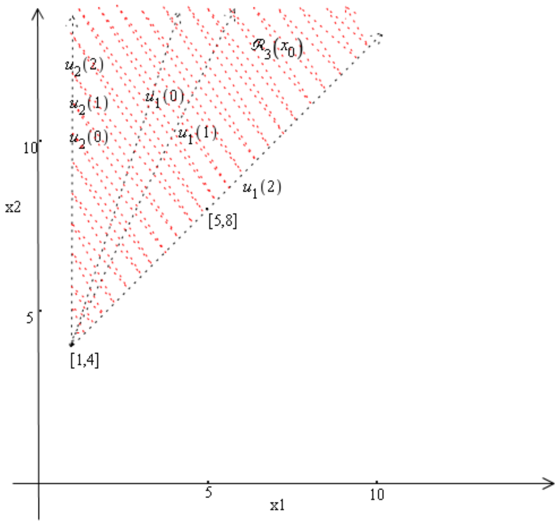

Figure 4.2. The reachable set in Example 4.3

From Theorem 4.7 , it is clear that the optimal control entries other than $u_1(2)$ are zero, i.e., $u_2(2) = u_1(1) = u_2(1) = u_1(0) = u_2(0) = 0$.

Therefore, $\mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0 = \begin{bmatrix}4\\4\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}u_1(2)$, and $u_1(2) = 4$.

Hence, the optimal control is

$\mathbf{u}^*(0) = \begin{bmatrix}0\\0\end{bmatrix}$,

and $\mathbf{x}^*(1) = \mathbf{A}\mathbf{x}_0 + \mathbf{B}\mathbf{u}^*(0) = \begin{bmatrix}1\\2\end{bmatrix}$

$\mathbf{u}^*(1) = \begin{bmatrix}0\\0\end{bmatrix}$,

$\mathbf{x}^*(2) = \mathbf{A}\mathbf{x}^*(1) + \mathbf{B}\mathbf{u}^*(1) = \begin{bmatrix}1\\3\end{bmatrix}$

$\mathbf{u}^*(2) = \begin{bmatrix}4\\0\end{bmatrix}$.

Example 4.4. Consider the following minimum energy problem with fixed final state and

$\mathbf{x}_N$ is on the boundary of $\boldsymbol{\mathcal{R}}_N(\mathbf{x}_0)$.

Minimize the objective function

$$J = \frac{1}{2}\sum_{t=0}^{2} \mathbf{u}^T(t) \begin{bmatrix} 3 & -1 & 1 \\ -1 & 2 & 2 \\ 1 & 2 & 9 \end{bmatrix} \mathbf{u}(t),$$

Subject to

$$\mathbf{x}(t+1) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \mathbf{u}(t),\ t = 0,1,2$$

$\mathbf{x}(0) = \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{x}(3) = \mathbf{x}_3 =$

$\begin{bmatrix} 5 \\ 8 \end{bmatrix}$

$\mathbf{u}(t) \in \mathbb{R}_+$

The *3-steps* reachable sets for the PDLS above is

$$\boldsymbol{\mathcal{R}}_3(\mathbf{x}_0) = \left\{ \mathbf{x} \middle| \mathbf{x} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} + \left[\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \mathbf{u}(2) + \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 2 \end{bmatrix} \mathbf{u}(1) + \begin{bmatrix} 1 & 0 & 1 \\ 3 & 1 & 3 \end{bmatrix} \mathbf{u}(0) \right], \mathbf{u}(t) \in \mathbb{R}_+, t = 0,1,2 \right\}$$

The final state is fixed, i.e.,

$$\mathbf{x}_3 = \begin{bmatrix} 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} + \left[\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \mathbf{u}(2) + \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 2 \end{bmatrix} \mathbf{u}(1) + \begin{bmatrix} 1 & 0 & 1 \\ 3 & 1 & 3 \end{bmatrix} \mathbf{u}(0) \right].$$

$$\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \left[\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \mathbf{u}(2) + \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 2 \end{bmatrix} \mathbf{u}(1) + \begin{bmatrix} 1 & 0 & 1 \\ 3 & 1 & 3 \end{bmatrix} \mathbf{u}(0) \right]$$

$$\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix} u_1(2) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_2(2) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u_3(2) + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u_1(1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_2(1) + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u_3(1) + \begin{bmatrix} 1 \\ 3 \end{bmatrix} u_1(0) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_2(0) + \begin{bmatrix} 1 \\ 3 \end{bmatrix} u_3(0) \right]$$

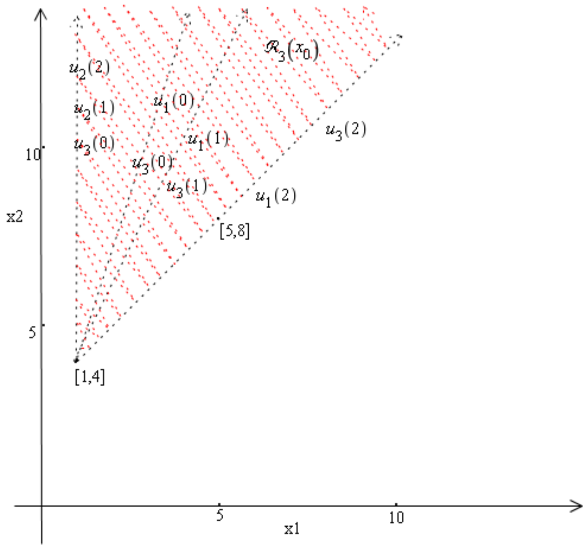

Figure 4.3 The reachable set in Example 4.4

From Theorem 4.7 , it is clear that the optimal control entries other than $u_1(2)$ and $u_3(2)$

are zero, i.e.,$u_2(2) = u_1(1) = u_2(1) = u_3(1) = u_1(0) = u_2(0) = u_3(0) = 0$.

Therefore, the optimal control is formulated as

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u_1(2) \\ u_3(2) \end{bmatrix} = \mathbf{x}_3 - \mathbf{A}^3\mathbf{x}_0 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

which results in $u_1(2) + u_3(2) = 4$.

As $u_2(2) = u_1(1) = u_2(1) = u_3(1) = u_1(0) = u_2(0) = u_3(0) = 0$, the objective function can be formulated as

$$J = \begin{bmatrix} u_1(2) \\ u_3(2) \end{bmatrix}^T \begin{bmatrix} 3 & 1 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} u_1(2) \\ u_3(2) \end{bmatrix}$$

Since $u_1(2) + u_3(2) = 4$, $u_3(2) = 4 - u_1(2)$.

$J = \begin{bmatrix} u_1(2) \\ 4 - u_1(2) \end{bmatrix}^T \begin{bmatrix} 3 & 1 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} u_1(2) \\ 4 - u_1(2) \end{bmatrix}$. Consequently, we get $J = 10u_1(2)^2 - 64u_1(2) + 144$

and the optimal control $\begin{bmatrix} u_1(2) \\ u_3(2) \end{bmatrix}$ that minimise the problem are $\begin{bmatrix} 3.2 \\ 0.8 \end{bmatrix}$.

Hence,

$$\mathbf{u}^*(0) = \begin{bmatrix}0\\0\\0\end{bmatrix},$$

$$\mathbf{x}^*(1) = \mathbf{A}\mathbf{x}_0 + \mathbf{B}\mathbf{u}^*(0) = \begin{bmatrix}1\\2\end{bmatrix}$$

$$\mathbf{u}^*(1) = \begin{bmatrix}0\\0\\0\end{bmatrix},$$

$$\mathbf{x}^*(2) = \mathbf{A}\mathbf{x}^*(1) + \mathbf{B}\mathbf{u}^*(1) = \begin{bmatrix}1\\3\end{bmatrix}$$

$$\mathbf{u}^*(2) = \begin{bmatrix}3.2\\0\\0.8\end{bmatrix}.$$

4.3 The minimum energy problem for positive linear continuous-time system

4.3.1 Problem formulation

The minimum energy problem for vector positive continuous-time linear system with fixed final state is formulated as follows [14]:

Minimize the objective function

$$J = \int_0^N \mathbf{u}^T(t)\mathbf{R}\mathbf{u}(t)dt \tag{4.32}$$

Subject to

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t),\ t \in [0, N] \tag{4.33}$$

where

$\mathbf{A} \in \mathbb{R}^{n\times n}$ is an $n \times n$ Metzler matrix, (4.34)

$\mathbf{B} \in \mathbb{R}_+^{n\times m}$,

$\mathbf{R} \in \mathbb{R}^{m\times m}$ is an $m \times m$ positive definite matrix,

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state variables at time t, $\mathbf{u}(t) \in \mathbb{R}_+^m$ is the control at time t, N is a finite-time horizon. The initial and the final states are given by

$$\mathbf{x}(0) = \mathbf{x}_0 \geq 0 \text{ and } \mathbf{x}(N) = \mathbf{x}_N \geq 0. \tag{4.35}$$

4.3.2 Main Results

In this section, a solution of the minimum energy problem (4.32) – (4.35) is obtained using the dynamic programming procedure. According to Lewis *et al* [18], there are two methods to solve continuous-time optimal control problems using the dynamic programming - the discretization approach and the Hamilton-Jacobi-Bellman equation. We will use the discretization approach to solve the minimum energy problem (4.32) – (4.35). To apply the dynamic programming approach for continuous-time systems using the discretization approach, the control and trajectory must be quantised to some finite set of admissible values (admissible control and feasible trajectory). Since a finer quantisation is required to obtain more accurate results, the increasing of number of calculation to find the accurate admissible controls and feasible trajectory is inevitable. Therefore, dynamic programming for discretised continuous-time systems is not often used to avoid the curse of dimensionality [18]. However, as an analytical solution for the discrete time has been obtained in previous section, the curse of dimensionality can be avoided.

To discretise the cost function (4.32), we can write

$$J = \frac{1}{2}\sum_{k=0}^{N_s-1} \int_{ks}^{(k+1)s} \mathbf{u}^T(t)\mathbf{R}\mathbf{u}(t)dt, \tag{4.36}$$

4where s is a time sampler and

$$N_s = \frac{N}{s}\,. \tag{4.37}$$

Using a first order approximation to each integral results in

$$J = \sum_{k=0}^{N_s-1} s\ \mathbf{u}^T(ks)\mathbf{R}\mathbf{u}(ks). \tag{4.38}$$

The discrete-time representation of (4.33) takes the form

$$\mathbf{x}\big((k+1)s\big) = \mathbf{A}_s\mathbf{x}(ks) + \mathbf{B}_s\mathbf{u}(ks). \tag{4.39}$$

To discretise system (4.33), we use a time sampler s and zero order hold [18, 59]. Lewis et al [18] claim that this method is better than any other approximation method.

From (2.24) we have,

$$\mathbf{x}\big((k+1)s\big) = e^{\mathbf{A}(k+1)s}\mathbf{x}_0 + \int_0^{(k+1)s} e^{\mathbf{A}\left((k+1)s-\tau\right)}\mathbf{B}\,\mathbf{u}(\tau)d\tau \tag{4.40}$$

$$= e^{\mathbf{A}(k+1)s}\mathbf{x}_0 + \int_0^{ks} e^{\mathbf{A}\left((k+1)s-\tau\right)}\mathbf{B}\,\mathbf{u}(\tau)d\tau + \int_{ks}^{(k+1)s} e^{\mathbf{A}\left((k+1)s-\tau\right)}\mathbf{B}\,\mathbf{u}(\tau)d\tau$$

and

$$\mathbf{x}(ks) = e^{\mathbf{A}ks}\mathbf{x}_0 + \int_0^{ks} e^{\mathbf{A}(ks-\tau)}\mathbf{B}\,\mathbf{u}(\tau)d\tau. \tag{4.41}$$

Multiplying (4.41) by $e^{\mathbf{A}s}$ results in

$$e^{\mathbf{A}s}\mathbf{x}(ks) = e^{\mathbf{A}(k+1)s}\mathbf{x}_0 + \int_0^{ks} e^{\mathbf{A}\left((k+1)s-\tau\right)}\mathbf{B}\,\mathbf{u}(\tau)d\tau. \tag{4.42}$$

Substituting (4.42) into (4.40) leads to

$$\mathbf{x}\big((k+1)s\big) = e^{\mathbf{A}s}\mathbf{x}(ks) + \int_{ks}^{(k+1)s} e^{\mathbf{A}\left((k+1)s-\tau\right)}\mathbf{B}\,\mathbf{u}(\tau)d\tau. \tag{4.43}$$

Since we use a sampler and zero order hold, the control $\mathbf{u}(t)$ is constant over the interval between any two consecutive sampling instants, i.e.,

$$\mathbf{u}(t) = \mathbf{u}(ks), \text{ for } ks \le t \le (k+1)s \tag{4.44}$$

Taking into account (4.44) and using $v = s - \tau$, the expression (4.32) results in

$$\mathbf{x}\big((k+1)s\big) = e^{\mathbf{A}s}\mathbf{x}(ks) + \left(\int_0^{s} e^{\mathbf{A}v}\mathbf{B}\,dv\right)\mathbf{u}(ks). \tag{4.45}$$

To obtain more accurate results, the smaller time sampler should be used.

Let us define

$$\mathbf{A}_s = e^{\mathbf{A}s}, \tag{4.46}$$

$$\mathbf{B}_s = \left(\int_0^s e^{\mathbf{A}v}\mathbf{B}\,dv\right), \tag{4.47}$$

$$\mathbf{R}_s = \mathbf{R}s, \tag{4.48}$$

$$\mathbf{x}(k) \triangleq \mathbf{x}(ks), \text{ and } \mathbf{u}(k) \triangleq \mathbf{u}(ks), \tag{4.49}$$

then the *discretised minimum energy problem* (4.32) – (4.35) becomes

Minimize

$$J = \frac{1}{2}\sum_{k=0}^{N_s-1} \mathbf{u}^T(k)\mathbf{R}_s\mathbf{u}(k) \tag{4.50}$$

s.t

$$\mathbf{x}(k+1) = \mathbf{A}_s\mathbf{x}(k) + \mathbf{B}_s\mathbf{u}(k),\, k = 0,1,\cdots,N_s \tag{4.51}$$

$$\mathbf{u}(k) \in \mathbb{R}_+^m, \mathbf{A}_s \in \mathbb{R}_+^{n\times n}, \mathbf{B}_s \in \mathbb{R}_+^{n\times m}, \mathbf{x}(k) \in \mathbb{R}^n \tag{4.52}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}_+^n \text{ and } \mathbf{x}(N_s) = \mathbf{x}(N) = \mathbf{x}_N \in \mathbb{R}_+^n, \tag{4.53}$$

Theorem 4.8. If the linear continuous-time system (4.33) (not necessarily positive system) is reachable then the optimal control that minimizes the cost function (4.32) in the minimum energy problem (4.32) – (4.35) with fixed final state can be represented as an open loop control

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T\mathbf{W}_c^{-1}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0), t \in [0, N), \tag{4.54}$$

and the optimal value of the cost function (4.32) is

$$J_0^* = \frac{1}{2}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0)^T\mathbf{W}_c^{-1}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0). \tag{4.55}$$

where

$$\mathbf{W}_c = \int_0^N \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T dt.$$

Proof. Consider the discretised problem (4.50) – (4.53). Using the results on Theorem 4.3 , the admissible control $\mathbf{u}^*(t) = \mathbf{u}^*(ks) \triangleq \mathbf{u}(k)$ is formulated as

$$\mathbf{u}^*(k) = \mathbf{R}_s^{-1}\left(\mathbf{A}_s^{N_s-k-1}\mathbf{B}_s\right)^T \left(\sum_{i=0}^{N_s-1}\left(\mathbf{A}_s^{N_s-i-1}\mathbf{B}_s\right)\mathbf{R}_s^{-1}\left(\mathbf{A}_s^{N_s-i-1}\mathbf{B}_s\right)^T\right)^{-1} \left(\mathbf{x}_N - \mathbf{A}_s^{N_s}\mathbf{x}_0\right). \tag{4.56}$$

Using the definition of matrix exponential, the formula $\mathbf{B}_s = \left(\int_0^s e^{\mathbf{A}v}\mathbf{B}\,dv\right)$ can be defined as

$$\mathbf{B}_s = \left(\int_0^s \left(\mathbf{I} + \mathbf{A}v + \frac{(\mathbf{A}v)^2}{2!} + \frac{(\mathbf{A}v)^3}{3!} + \cdots\right) dv\, \mathbf{B}\right).$$

Therefore,

$$\mathbf{B}_s = s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2s^2}{3!} + \frac{\mathbf{A}^3s^3}{4!} + \cdots\right)\mathbf{B}. \tag{4.57}$$

Let us define the gramian of the optimal control (4.56)

$$\mathbf{W}(s) = \sum_{i=0}^{N_s-1}\left(\mathbf{A}_s^{N_s-i-1}\mathbf{B}_s\right)\mathbf{R}_s^{-1}\left(\mathbf{A}_s^{N_s-i-1}\mathbf{B}_s\right)^T. \tag{4.58}$$

Substituting (4.46) – (4.48) into (4.58) leads to

$$\mathbf{W}(s) = \sum_{i=0}^{N_s-1}\left(e^{\mathbf{A}s(N_s-i-1)}s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2s^2}{3!} + \frac{\mathbf{A}^3s^3}{4!} + \cdots\right)\mathbf{B}\right)\frac{\mathbf{R}^{-1}}{s}\left(e^{\mathbf{A}s(N_s-i-1)}s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2s^2}{3!} + \frac{\mathbf{A}^3s^3}{4!} + \cdots\right)\mathbf{B}\right)^T.$$

From (4.37), the above expression can be rewritten as

$$\mathbf{W}(s) = \sum_{i=0}^{N_s-1}\left(e^{\mathbf{A}(N-si-s)}s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2s^2}{3!} + \frac{\mathbf{A}^3s^3}{4!} + \cdots\right)\mathbf{B}\right)\frac{\mathbf{R}^{-1}}{s}\left(e^{\mathbf{A}(N-si-s)}s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2s^2}{3!} + \frac{\mathbf{A}^3s^3}{4!} + \cdots\right)\mathbf{B}\right)^T.$$

Let $s\mathbf{F}(si) = \left(e^{\mathbf{A}(N_s-si-s)}s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2s^2}{3!} + \frac{\mathbf{A}^3s^3}{4!} + \cdots\right)\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-si-s)}\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2s^2}{3!} + \frac{\mathbf{A}^3s^3}{4!} + \cdots\right)\mathbf{B}\right)^T$. Using the first order approximation results in $s\mathbf{F}(si) = \int_{is}^{(i+1)s}\mathbf{F}(t)dt$. Therefore, $\mathbf{W}(s) = \sum_{i=0}^{N_s-1}\int_{is}^{(i+1)s}\mathbf{F}(t)dt = \int_0^N \mathbf{F}(t)dt$.

Hence, the gramian $\mathbf{W}(s)$ can be represented as

$$\mathbf{W}(s) = \int_0^N \left(e^{\mathbf{A}(N-t-s)} \left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2 s^2}{3!} + \frac{\mathbf{A}^3 s^3}{4!} + \cdots \right) \mathbf{B} \right) \mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t-s)} \left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2 s^2}{3!} + \frac{\mathbf{A}^3 s^3}{4!} + \cdots \right) \mathbf{B} \right)^T dt \tag{4.59}$$

Then, substituting (4.46) – (4.48) into the optimal control (4.56) leads to

$$\mathbf{u}^*(k) = \mathbf{R}^{-1} \left(e^{\mathbf{A}s(N_s-k-1)} \left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2 s^2}{3!} + \frac{\mathbf{A}^3 s^3}{4!} + \cdots \right) \mathbf{B} \right)^T \left(\mathbf{W}(s) \right)^{-1} (\mathbf{x}_N - e^{\mathbf{A}sN_s} \mathbf{x}_0).$$

From (4.37) the above equation becomes

$$\mathbf{u}^*(k) = \mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t-s)} \left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2 s^2}{3!} + \frac{\mathbf{A}^3 s^3}{4!} + \cdots \right) \mathbf{B} \right)^T \left(\mathbf{W}(s) \right)^{-1} (\mathbf{x}_N - e^{\mathbf{A}N} \mathbf{x}_0).$$

To obtain more accurate results, the smaller time sampler should be used. Therefore, taking limit, one has

$$\lim_{s\to 0} \mathbf{u}^*(k) = \lim_{s\to 0} \mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t-s)} \left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2 s^2}{3!} + \frac{\mathbf{A}^3 s^3}{4!} + \cdots \right) \mathbf{B} \right)^T \left(\lim_{s\to 0} \mathbf{W}(s) \right)^{-1} (\mathbf{x}_N - e^{\mathbf{A}N} \mathbf{x}_0).$$

The expression $\lim_{s\to 0} \mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t-s)} \left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^2 s^2}{3!} + \frac{\mathbf{A}^3 s^3}{4!} + \cdots \right) \mathbf{B} \right)^T$ can be written as

$$\mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t)} \mathbf{B} \right)^T,$$

and

$$\lim_{s\to 0} \mathbf{W}(s) \coloneqq \mathbf{W}_c = \int_0^N \left(e^{\mathbf{A}(N-t)} \mathbf{B} \right) \mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t)} \mathbf{B} \right)^T dt. \tag{4.60}$$

As $\mathbf{u}^*(t)$=$\mathbf{u}^*(k)$, the optimal control (4.56) becomes

$$\mathbf{u}^*(t) = \mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t)} \mathbf{B} \right)^T \left(\int_0^N \left(e^{\mathbf{A}(N-t)} \mathbf{B} \right) \mathbf{R}^{-1} \left(e^{\mathbf{A}(N-t)} \mathbf{B} \right)^T dt \right)^{-1} (\mathbf{x}_N - e^{\mathbf{A}N} \mathbf{x}_0), \tag{4.61}$$

for $t \in [0, N)$.

We will show that the gramian is positive definite. Define $\mathbf{y} = \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \mathbf{z}$ where $\mathbf{y} \in \mathbb{R}^m, \mathbf{y} \neq 0$, and $\mathbf{z} \in \mathbb{R}^n$. Since R is positive definite, $\mathbf{R}^{-1}$ is positive definite and $\mathbf{y}^T\mathbf{R}^{-1}\mathbf{y} > \mathbf{0}$ for any nonzero $\mathbf{y} \in \mathbb{R}^m$. From Theorem 4.3 , the gramian $\mathbf{R}_N :=$ $\int_0^N e^{\mathbf{A}(N-\tau)}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T(N-\tau)}\, d\tau$ is positive definite. Then, $\int_0^N \mathbf{z}^T e^{\mathbf{A}(N-\tau)}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T(N-\tau)}\, \mathbf{z}\, d\tau = 0$ if and only if $\mathbf{z} = \mathbf{0}$. Therefore, $\mathbf{y} = \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \mathbf{z} = \mathbf{0}$ if and only if $\mathbf{z} = \mathbf{0}$. Hence, $\int_0^N \mathbf{z}^T \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \mathbf{z}\, dt > \mathbf{0}$ for any nonzero $\mathbf{z} \in \mathbb{R}^n$ and the inverse of the gramian $\mathbf{W_c}$ exists.

Using the result of Theorem 4.3 , the optimal control of the discretised positive minimum energy problem (4.50) – (4.53) is

$$J_0^*(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_N - \mathbf{A}_s{}^{N_s}\mathbf{x}_0\right)^T \left(\sum_{i=0}^{N_s-1}\left(\mathbf{A}_s{}^{N_s-i-1}\mathbf{B}_s\right)\mathbf{R}_s{}^{-1}\left(\mathbf{A}_s{}^{N_s-i-1}\mathbf{B}_s\right)^T\right)^{-1} \left(\mathbf{x}_N - \mathbf{A}_s{}^{N_s}\mathbf{x}_0\right). \quad (4.62)$$

By substituting (4.46) – (4.48) into (4.62), the following result can be obtained

$$J_0^*(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_N - e^{\mathbf{A}sN_s}\mathbf{x}_0\right)^T\left(\mathbf{W}(s)\right)^{-1}\left(\mathbf{x}_N - e^{\mathbf{A}sN_s}\mathbf{x}_0\right),$$

where $\mathbf{W}(s)$ is as defined by (4.59).

From (4.37), the optimal cost is

$$J_0^*(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0\right)^T\left(\mathbf{W}(s)\right)^{-1}\left(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0\right).$$

Taking $\lim_{s\to 0} J_0^*(\mathbf{x}_0)$ and from (4.60), the optimal cost can be represented as

$$J_0^*(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0\right)^T \left(\int_0^N \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T dt\right)^{-1} \left(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0\right). \quad \square$$

Theorem 4.9. Consider the positive linear continuous-time system (4.33). If $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, $\mathbf{W_c}^{-1}$ exist, $\mathbf{W_c}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$ then the optimal control that minimizes the cost function (4.32) in the minimum energy problem (4.32) – (4.35) with a fixed final state is

$$\boldsymbol{u}_+^*(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \mathbf{W}_{\boldsymbol{c}}{}^{-1}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0), t \in [0, N), \tag{4.63}$$

and the optimal value of the cost function (4.32) is

$$J_0^* = \frac{1}{2}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0)^T \mathbf{W}_{\boldsymbol{c}}{}^{-1}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0). \tag{4.64}$$

Proof. Consider the optimal control (4.54) in Theorem 4.8.

From Lemma 2.5, it is clear that if $\mathbf{x}_N \in \mathcal{R}_N(\mathbf{x}_0)$, then $\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0 \geq 0$.

If $\mathbf{W}^{-1} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{R}^{-1} \in \mathbb{R}_+^{m\times m}$, then from (4.54) we have

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \mathbf{W}_{\boldsymbol{c}}{}^{-1}(\mathbf{x}_N - \mathbf{A}^N\mathbf{x}_0),$$

which is nonnegative for all $t \in [0, N)$. Therefore, the optimal control for the positive linear discrete-time system (4.33) is formulated as

$$\mathbf{u}_+^*(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \mathbf{W}_{\boldsymbol{c}}{}^{-1}(\mathbf{x}_N - e^{\mathbf{A}N}\mathbf{x}_0), t \in [0, N). \qquad \square$$

Remarks 4.2.

i. If $\mathbf{x}_N$ is not in the set of reachable states, then the two-point boundary value problem (4.33) – (4.35) is inconsistent and therefore, the minimum energy problem with a fixed final state (4.32) – (4.35) has no solution.

ii. If the initial state $\mathbf{x}_0$ is zero, formula (4.63 and (4.64) become even simpler

$$\mathbf{u}_+^* = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \mathbf{W}_{\boldsymbol{c}}{}^{-1}\mathbf{x}_N, \text{ for } t \in [0, N), \text{ and}$$

$$J_0^*(0) = \frac{1}{2}\mathbf{x}_N{}^T \mathbf{W}_{\boldsymbol{c}}{}^{-1}\mathbf{x}_N.$$

iii. $\mathbf{W}_{\boldsymbol{c}}{}^{-1} \in \mathbb{R}_+^{n\times n}$ if matrix $\mathbf{W}_{\boldsymbol{c}}$ is a monomial matrix

Theorem 4.10. If the linear continuous-time system (4.33) (not necessarily positive system) is reachable and $\text{rank}(\mathbf{B}) = n$, then the optimal feedback control sequence that minimizes the cost function (1) in the minimum energy problem (4.32) – (4.35) with fixed final state is given by

$$u^{*}(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^{T}\left(\int_{t}^{N}\left(e^{A(N-t)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{A(N-t)}\mathbf{B}\right)^{T}dt\right)^{-1}\Big(\mathbf{x}_{N} - e^{\mathbf{A}N}\mathbf{x}^{*}(t)\Big), t \in [0, N). \tag{4.65}$$

where $\mathbf{x}^{*}(t)$ is the corresponding optimal trajectory.

Proof. Consider the discretised continuous-time system (4.50) – (4.53). Using the results on Theorem 4.1, the admissible control $\mathbf{u}^{*}(t) = \mathbf{u}^{*}(ks) \triangleq \mathbf{u}(k)$ is formulated as

$$\mathbf{u}^{*}(k) = \mathbf{R}_{s}^{-1}\left(\mathbf{A}_{s}^{N_s-k-1}\mathbf{B}_{s}\right)^{T}\left(\sum_{i=k}^{N_s-1}\left(\mathbf{A}_{s}^{N_s-i-1}\mathbf{B}_{s}\right)\mathbf{R}_{s}^{-1}\left(\mathbf{A}_{s}^{N_s-i-1}\mathbf{B}_{s}\right)^{T}\right)^{-1}\Big(\mathbf{x}_{N} - \mathbf{A}_{s}^{N_s-k}\mathbf{x}^{*}(k)\Big) \tag{4.66}$$

From (4.57), $\mathbf{B}_{s}$ is defined as follows

$$\mathbf{B}_{s} = s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^{2}s^{2}}{3!} + \frac{\mathbf{A}^{3}s^{3}}{4!} + \cdots\right)\mathbf{B}$$

Let us define the gramian of the optimal control (4.66) as follows

$$\mathbf{Q}(s) = \sum_{i=k}^{N_s-1}\left(\mathbf{A}_{s}^{N_s-i-1}\mathbf{B}_{s}\right)\mathbf{R}_{s}^{-1}\left(\mathbf{A}_{s}^{N_s-i-1}\mathbf{B}_{s}\right)^{T}. \tag{4.67}$$

Substituting (4.46) – (4.48) into (4.67) leads to

$$\mathbf{Q}(s) = \sum_{i=k}^{N_s-1}\left(e^{\mathbf{A}s(N_s-i-1)}s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^{2}s^{2}}{3!} + \frac{\mathbf{A}^{3}s^{3}}{4!} + \cdots\right)\mathbf{B}\right)\frac{\mathbf{R}^{-1}}{s}\left(e^{\mathbf{A}s(N_s-i-1)}s\left(\mathbf{I} + \frac{\mathbf{A}s}{2!} + \frac{\mathbf{A}^{2}s^{2}}{3!} + \frac{\mathbf{A}^{3}s^{3}}{4!} + \cdots\right)\mathbf{B}\right)^{T}.$$

From (4.37), the above expression can be rewritten as

$$\mathbf{Q}(s) = \sum_{i=k}^{N_s-1} \left(e^{\mathbf{A}(N-si-s)}s\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)\frac{\mathbf{R}^{-1}}{s}\left(e^{\mathbf{A}(N-si-s)}s\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)^T.$$

Define $s\mathbf{G}(si) = \left(e^{\mathbf{A}(N_s-si-s)}s\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-si-s)}\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)^T$. Using the first order approximation results in $s\mathbf{G}(si) = \int_{is}^{(i+1)s}\mathbf{G}(\tau)d\tau$. Therefore, $\mathbf{Q}(s) = \sum_{i=k}^{N_s-1}\int_{is}^{(i+1)s}\mathbf{G}(\tau)d\tau = \int_t^N \mathbf{G}(\tau)d\tau$.

Hence, the gramian $\mathbf{Q}(s)$ can be represented as

$$\mathbf{Q}(s) = \int_t^N \left(e^{\mathbf{A}(N-\tau-s)}\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-\tau-s)}\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)^T d\tau.$$

Then, substituting (4.46) – (4.48) into the optimal control (4.66) leads to

$$\mathbf{u}^*(k) = \mathbf{R}^{-1}\left(e^{\mathbf{A}s(N_s-k-1)}\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)^T\left(\mathbf{Q}(s)\right)^{-1}\left(\mathbf{x}_N - e^{\mathbf{A}s(N_s-k)}\mathbf{x}^*(k)\right).$$

From (4.37) and $\mathbf{u}^*(t)$=$\mathbf{u}^*(k)$, the above equation becomes

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t-s)}\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)^T\left(\mathbf{Q}(s)\right)^{-1}\left(\mathbf{x}_N - e^{\mathbf{A}(N-t)}\mathbf{x}^*(t)\right).$$

To obtain more accurate results, the smaller time sampler should be used. Therefore, taking limit yields

$$\lim_{s\to 0}\mathbf{u}^*(t) = \lim_{s\to 0}\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t-s)}\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)^T\left(\lim_{s\to 0}\mathbf{Q}(s)\right)^{-1}\left(\mathbf{x}_N - e^{\mathbf{A}(N-t)}\mathbf{x}^*(t)\right).$$

The expression $\lim_{s\to 0}\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t-s)}\left(\mathbf{I}+\frac{\mathbf{A}s}{2!}+\frac{\mathbf{A}^2s^2}{3!}+\frac{\mathbf{A}^3s^3}{4!}+\cdots\right)\mathbf{B}\right)^T$ can be written as

$$\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T,$$

and

$$\lim_{s\to 0}\mathbf{Q}(s) \coloneqq \mathbf{Q} = \int_t^N \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T dt. \tag{4.68}$$

Therefore, the optimal control (4.66) becomes

$$\mathbf{u}^*(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \left(\int_t^N \left(e^{\mathbf{A}(N-\tau)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-\tau)}\mathbf{B}\right)^T d\tau\right)^{-1} \left(\mathbf{x}_N - e^{\mathbf{A}(N-t)}\mathbf{x}^*(t)\right), \text{ for } t \in [0, N). \tag{4.69}$$

We will show that the gramian $\mathbf{Q}$ is positive definite for all $t \in [0, N)$. Define $\mathbf{y} = \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T\mathbf{z}$ where $\mathbf{y} \in \mathbb{R}^m, \mathbf{y} \neq 0$, and $\mathbf{z} \in \mathbb{R}^n$. Since $\mathbf{R}$ is positive definite, $\mathbf{R}^{-1}$ is positive definite and $\mathbf{y}^T\mathbf{R}^{-1}\mathbf{y} > \mathbf{0}$ for any nonzero $\mathbf{y} \in \mathbb{R}^m$. From Theorem 4.3 , the gramian $\mathbf{R}_N := \int_t^N e^{\mathbf{A}(N-\tau)}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T(N-\tau)}\, d\tau$ is positive definite. Then, $\int_t^N \mathbf{z}^T e^{\mathbf{A}(N-\tau)}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T(N-\tau)}\,\mathbf{z}\, d\tau = 0$ if and only if $\mathbf{z} = \mathbf{0}$. Therefore, $\mathbf{y} = \left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T\mathbf{z} = \mathbf{0}$ if and only if $\mathbf{z} = \mathbf{0}$. Hence, $\int_t^N \mathbf{z}^T\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T\mathbf{z}\, dt > \mathbf{0}$ for any nonzero $\mathbf{z} \in \mathbb{R}^n$ and the inverse of the gramian $\mathbf{Q}$ exists.

Theorem 4.11. Consider the positive linear continuous-time system (4.33). If $\mathbf{x}_N \in R_t(\mathbf{x}_0)$ and $\text{rank}(\mathbf{B}) = n$ and $\mathbf{W^{-1}} \in \mathbb{R}_+^{n\times n}$ then the optimal control that minimizes the cost function (4.32) in the minimum energy problem (4.32) – (4.35) with fixed final state is

$$\boldsymbol{u}_+^*(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^T \left(\int_t^N \left(e^{\mathbf{A}(N-\tau)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-\tau)}\mathbf{B}\right)^T d\tau\right)^{-1} \left(\mathbf{x}_N - e^{\mathbf{A}(N-t)}\mathbf{x}^*(t)\right), \text{ for } t \in [0, N). \tag{4.70}$$

Proof. The open loop optimal control (4.54) is derived from the feedback optimal control (4.65). From Theorem 4.9, if $\mathbf{x}_N \in R_t(\mathbf{x}_0)$ and $\mathbf{W^{-1}} \in \mathbb{R}_+^{n\times n}$, then the open loop (4.54) is nonnegative. Therefore, it is clear that the feedback optimal control (4.65) is nonnegative under the same conditions, i.e., $\mathbf{x}_N \in R_t(\mathbf{x}_0)$ and the positive linear continuous-time system (4.33) is reachable.

Thus, the optimal control is formulated as

$$\boldsymbol{u}_{+}^{*}(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^{T}\left(\int_{t}^{N}\left(e^{\mathbf{A}(N-\tau)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{\mathbf{A}(N-\tau)}\mathbf{B}\right)^{T}d\tau\right)^{-1}\Big(\mathbf{x}_{N} - e^{\mathbf{A}(N-t)}\mathbf{x}^{*}(t)\Big)$$

for $t \in [0, N)$.

4.3.3 Numerical Example

To illustrate the approach adopted in this chapter, we consider the following examples on the minimum energy problem with fixed final state. The following numerical example problems are solved using the formula obtained in this chapter. We will show the effectiveness of the analytic formula by comparing the results solved using Imperial College London Optimal Control Software (ICLOCS) [31].

Example 4.5. Minimize the objective function

$$J = \frac{1}{2}\int_{0}^{2} \mathbf{u}^{T}(t)\begin{bmatrix}1 & 0\\0 & 3\end{bmatrix}\mathbf{u}(t)\,dt$$

subject to

$$\dot{\mathbf{x}}(t) = \begin{bmatrix}3 & 0\\0 & 4\end{bmatrix}\mathbf{x}(t) + \begin{bmatrix}4 & 0\\0 & 5\end{bmatrix}\mathrm{u}(t), t \in [0, 2),$$

$$\mathbf{x}(0) = \mathbf{x}_{0} = \begin{bmatrix}1\\1\end{bmatrix},$$

$$\mathbf{x}(2) = \mathbf{x}_{2} = \begin{bmatrix}405\\2999\end{bmatrix},$$

$$\mathbf{u}(t) \in \mathbb{R}_{2}^{+}.$$

From the result (4.63), the optimal control is

$$\mathbf{u}_{+}^{*}(t) = \mathbf{R}^{-1}\left(e^{\mathbf{A}(N-t)}\mathbf{B}\right)^{T}\mathbf{W}_{c}{}^{-1}(\mathbf{x}_{N} - e^{\mathbf{A}N}\mathbf{x}_{0}), t \in [0, N).$$

where

$$\mathbf{W}_{c} = \int_{0}^{N}\left(e^{A(N-t)}\mathbf{B}\right)\mathbf{R}^{-1}\left(e^{A(N-t)}\mathbf{B}\right)^{T}dt$$

Substituting the corresponding matrices, A, B and R into the above gramian results in

$$\mathbf{W}_{c} = \begin{bmatrix}434010.11025 & 0\\ 0 & 92526364.084\end{bmatrix}.$$

Therefore, it is clear that $\mathbf{W_c}^{-1} = \begin{bmatrix} 0.2304E-5 & 0 \\ 0 & 0.108034E-6 \end{bmatrix}$ is nonnegative.

Hence, the optimal control of the problem is

$$\mathbf{u}_+^*(t) = \begin{bmatrix} 1 & 0 \\ 0 & 0.33 \end{bmatrix} \begin{bmatrix} 4e^{6-3t} & 0 \\ 0 & 5e^{8-4t} \end{bmatrix} \begin{bmatrix} 0.2304E-5 & 0 \\ 0 & 0.108034E-6 \end{bmatrix} \left(\begin{bmatrix} 405 \\ 2999 \end{bmatrix} - \begin{bmatrix} 403.4 \\ 2981 \end{bmatrix} \right)$$

and leads to

$$\mathbf{u}_+^*(t) = \begin{bmatrix} 0.000014480e^{6-3t} \\ 0.000003248578e^{8-4t} \end{bmatrix}.$$

Graphically, the optimal control is shown in Figure 4.4.

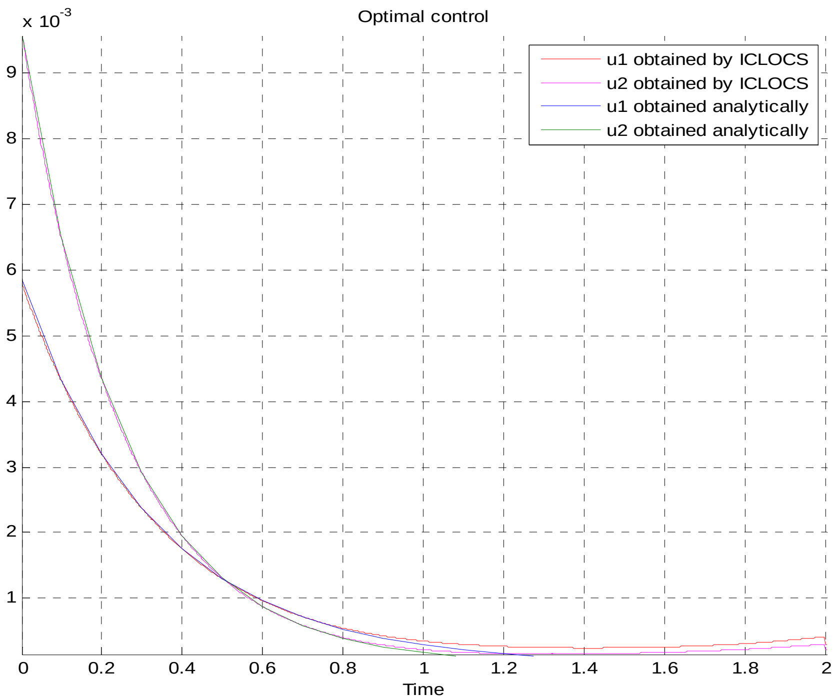

Figure 4.4. The Optimal control for Example 4.5.

The optimal trajectory (4.64) of the problem is

$$\mathbf{x}^*(t) = \begin{bmatrix} e^{3t} + 0.0038947e^{-3t}(e^{6t}-1) \\ e^{4t} + 0.0060524e^{-4t}(e^{8t}-1) \end{bmatrix},$$

which is shown in Figure 4.5.

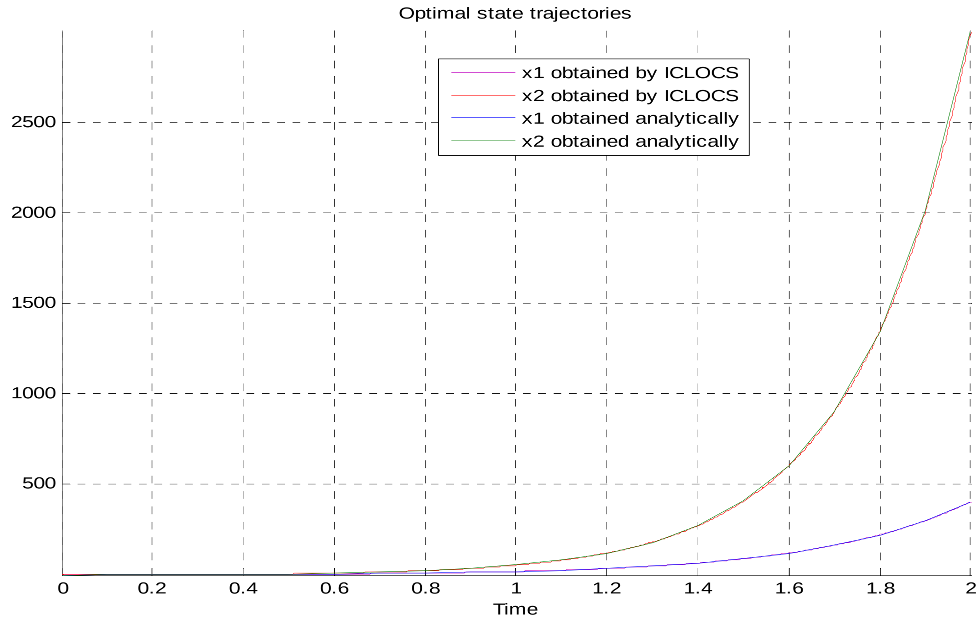

Figure 4.5. The corresponding optimal trajectory for Example 4.5

The optimal cost is 2.0217×10^5.

4.4 Concluding remarks

The solution and the criteria for the existence of the solutions of the minimum energy problem for positive discrete-time linear systems and positive linear continuous-time systems with any pair of initial and fixed final state have been established in this chapter by using dynamic programming approach. The relationship between the problem and the geometric properties of the systems is revealed and well exploited. The optimal control sequence is presented in two different (equivalent) forms – a feedback form and an open-loop form. The minimum energy problem has a trivial (zero) solution if the positive discrete-time linear system does not possess the reachability property. It does not have solution if the final state does not belong to the *T-steps* reachable set $\boldsymbol{\mathcal{R}}_N(\mathbf{x}_0)$. The optimal solution becomes very simple if the system is controllable or the initial state is zero.

CHAPTER 5

Applications

5.1 General

In this chapter, two applications related to energy and ecology are exposed. The first application is a dynamic model of oil extraction and its optimization. It is a novel dynamic (discrete-time) model that describes the evolution of the oil extraction process from a single well or reservoir under water flooding. The model incorporates some parameters important for the production planning and control as well as restrictions on the used water resource. It is assumed that the mixture in the well consists of oil and water only, that is the quantity of all other components is negligible, and that the extracted volume of mixture, water and oil, in a given time period is proportional to the amount of mixture in the reservoir at the beginning of the time period. Because of the nature of oil extraction process, the model exhibits positive linear system behaviour. On the basis of the proposed model an optimal control problem is formulated in this chapter. The optimal control problem turns out to be a novel problem for the theory of positive linear system problems. It is discussed in the chapter in details and a method for its solution is proposed. The obtained results could have important applications in improving the production methodology and supporting the managerial decisions in the process of oil extraction under water flooding. The results in this application is based on my publication [60].

The second application is a continuous-time dynamic mobile source air pollution optimal control problem. The model to predict and control the emission level from a total vehicle population is developed using a comprehensive set of input data for the range of variables that can influence the vehicle emissions including manufacture year (age of vehicles), vehicle populations, total annual kilometres travelled, and emission factors for a number of selected pollutant. The vehicle population is structured into different groups (cohorts) on the basis of the attribute “age”. It is assumed that vehicles of the same group have similar physical characteristics and behaviour. Therefore, the model is suitable for large cities, metropolitan areas or regions, where small group characteristics do not affect the aggregate behaviour of the vehicle population. Because of the nature of mobile source air pollution process, the model exhibits positive linear system behaviour. Utilising the model developed by Rumchev, et al [61], the optimal control problem is developed and analysed in this chapter.

5.2 A dynamic model of oil extraction and its optimization

5.2.1 The model

Let us denote by $y(t)$ the amount of mixture, measured in barrels, of crude oil and water in the well/reservoir at the beginning of time period t of oil extraction under water flooding. It is assumed that the mixture consists of crude oil and water only and the quantity of all other components is negligible – quite a realistic assumption. The time period t is usually measured in days or weeks so $t = 0, 1, 2, \ldots, T$; T is named a planning horizon. The mixture of crude oil and water is obviously nonnegative: $y(t) \geq 0$.The amount [in barrels] of crude oil $x(t)$ in the well at the beginning of time period t is nonnegative too: $x(t) \geq 0$.

One of the important parameters characterizing the mixture in the reservoir is the water to oil ratio (WOR) [62, 63] defined as

$$\gamma(t) = \frac{y(t) - x(t)}{x(t)} = \frac{water}{oil}$$

The WOR $\gamma(t)$is a monotonically increasing function of t, see for example [64, 65], which is clearly positive, $\gamma(t) > 0$. It can be predicted sufficiently well [64, 65] and we assume that it is known. Related to WOR is another technological characteristic of oil extraction process - the oil to mixture ratio (OMR):

$$\beta(t) = \frac{1}{1 + \gamma(t)} = \frac{x(t)}{y(t)} = \frac{oil}{mixture}.$$

The OMR $\beta(t)$ is a monotonically decreasing function, which is also positive, since there is always some amount of crude oil left in the well, and less than one: $0 < \beta(t) < 1$.

Let $u(t)$ denote the amount [in barrels] of water injected into the well during the time period t. The injected amount of water during the time period t, which is clearly nonnegative, is restricted from below by the necessity to maintain or even slightly increase the pressure in the well, and from above by some technological constraints so that

$$0 \leq \underline{u}(t) \leq u(t) \leq \bar{u}(t).$$

It is reasonable to assume that the lower $\underline{u}(t)$and upper $\bar{u}(t)$ bounds on the injected water are constant, that is $\underline{u}(t) = \underline{u}$ and $\bar{u}(t) = \bar{u}$, since the technology usually does not change during the planning horizon. However, the lower bound $\underline{u}(t)$ might slightly vary in time and because of that as well as for the sake of generality we assume in the further considerations that the bounds on the injected water may vary in time.

Let us denote by $v(t)$ the amount [in barrels] of mixture of crude oil and water extracted from the well during the time period t. A standard assumption is that the extracted volume of mixture in the time period t is proportional to the amount of mixture in the well at the beginning of the time period, i.e. $v(t) = \mu(t)y(t)$, where $\mu(t)$ is the extraction rate. For a given technology the extraction rate does not really change in time so that $\mu(t) \approx \mu = const$, and the volume of mixture extracted from the well during the time period t can be rewritten as

$$v(t) = \mu y(t),$$

The extraction rate μ is, obviously, a fraction between zero and one, that is $0 < \mu < 1$.

There are two alternative ways of describing the process of oil extraction on the basis of material balance, depending on the choice of state variables. If the mixture of oil and water in the well is chosen as a state variable, the process of oil extraction is described by the simple balance equation

$$y(t+1) = y(t) + u(t) - v(t),$$

for $t = 0, 1, 2, \dots, T-1$ which can be rewritten as $y(t+1) = y(t) + u(t) - \mu y(t)$, or

$$y(t+1) = (1-\mu)y(t) + u(t), \text{ for } t = 0, 1, 2, \dots, T-1 \tag{5.1}$$

Alternatively, if the amount of oil in the well is adopted as a state variable, then using the water to oil ratio $\gamma(t)$, the balance equation (5.1) reduce to

$$\big(1+\gamma(t+1)\big)x(t+1) = (1-\mu)\big(1+\gamma(t)\big)x(t) + u(t).$$

The above equation can be rearranged as

$$x(t+1) = (1-\mu)\frac{1+\gamma(t)}{1+\gamma(t+1)}x(t) + \frac{1}{1+\gamma(t+1)}u(t)$$

or $$x(t+1) = \alpha(t)x(t) + \beta(t+1)u(t), \text{ for } t = 0, 1, 2, \dots, T-1 \quad (5.2)$$

where

$$\alpha(t) = (1-\mu)\,\frac{\beta(t+1)}{\beta(t)} \text{ and } 0 < \alpha(t) < 1 \quad (5.3)$$

In (5.3), the fractional coefficient $(1-\mu)$ reflects the retention rate, since μ is the extraction rate of the mixture of crude oil and water from the well, and $\beta(t)$ is the oil to mixture ratio. Since the OMR $\beta(t)$ is a monotonically decreasing function, $\alpha(t)$ is also a monotonically decreasing function in time.

The state equations (5.1) and (5.2) give exogenous descriptions of the process of oil extraction under water flooding from a single well / reservoir suitable for the purposes of scheduling and planning, reservoir-management optimization and decision-making. They are discrete-time dynamic positive linear systems [14] since the system parameters $(1-\mu)$ and $\alpha(t)$, the control parameter $\beta(t)$ and the input (control) $u(t)$ are nonnegative.

The positive linear system (5.2) is a system with time-varying coefficients. Controllability properties of time-varying positive linear systems are studied in [16], where criteria for identifying reachability, null-controllability and controllability properties of such systems are developed. In the rest of this chapter the exogenous dynamic representation (5.2) of the oil extraction from a single well under water flooding will be used. On the bases of that description two optimal control problems will be formulated and discussed in the next section.

5.2.2 Optimal Control Problems

Let the cost of extraction of crude oil from the mixture in time period t is denoted as p [$/barrel] and its selling price as $s(t)$ [$/barrel]. For a given technology the extraction cost does not usually change in time and it does not include the cost of injected water. Let r [$/barrel] be the cost of injected water. Then the following two optimal control problem arise

OCP I. Minimizing the production expenditure

$$\min_{u(t)} \sum_{t=0}^{T-1} (p(x(t+1) - x(t)) + ru(t))$$

subject to

$$x(t+1) = \alpha(t)x(t) + \beta(t+1)u(t), t \ = \ 0, 1, 2, \dots, T-1$$

$$0 \leq x(0) = x_0,$$

$$0 \leq \underline{u}(t) \leq u(t) \leq \bar{u}(t).$$

OCP II. Maximizing the profit

$$\max_{u(t)} \sum_{t=0}^{T-1} (s(t) - p)(x(t+1) - x(t)) - ru(t) = \max_{u(t)} \sum_{t=0}^{T-1} (c(t)(x(t+1) - x(t)) - ru(t))$$

subject to

$$x(t+1) = \alpha(t)x(t) + \beta(t+1)u(t), t \ = \ 0, 1, 2, \dots, T-1 \ 0 \leq x(0) = x_0,$$

$$0 \leq \underline{u}(t) \leq u(t) \leq \bar{u}(t).$$

Note that in the optimality criteria (cost functions) of both OCP I and OCP II the difference $x(t+1) - x(t)$ represents the amount of crude oil extracted during time period t. The initial amount of crude oil x_0 in the well can be estimated sufficiently well. We assume that it is given.

OCP I and OCP II are optimal control problems of time-varying positive linear discrete-time systems with constraints on control, fixed initial state and linear criteria but the criteria are not in the standard for optimal control theory form. A substitution of equation (5.2) in the criteria reduces the problems to a standard form. This technique is used below to reduce OCP I to the equivalent OCP Ia. Further on in the section we focus on OCP I only since both problems belong to the same class of optimal control problems and the solution procedure for OCP II is similar to that presented for OCP I below.

OCP Ia. Minimizing the production expenditure

$$\min_{u(t)} \sum_{t=0}^{T-1} (c_x(t)x(t) + c_u(t)u(t))$$

subject to

$$x(t+1) = \alpha(t)x(t) + \beta(t+1)u(t), t = 0, 1, 2, \dots, T-1$$

$$0 \leq x(0) = x_0,$$

$$0 \leq \underline{u}(t) \leq u(t) \leq \bar{u}(t).$$

In the problem OCP Ia above the cost coefficients $c_x(t)$ and $c_u(t)$ in the criterion F are, respectively, $c_x(t) = p(\alpha(t) - 1) < 0$ and $c_u(t) = \beta(t+1) + r$. Note, please, that $c_u(t)$is a monotonically decreasing function in time, since is $\beta(t)$ monotonically decreasing, and so is $c_x(t)$, because $\alpha(t)$ is also monotonically decreasing. That is: $c_x(t+1) < c_x(t)$ and $c_u(t+1) < c_x(t)$.

OCP I and OCP II can be solved numerically given numerical values of parameters as static linear programming (LP) problems by standard LP software but in this work we aim to go deeper into the problems by finding an analytic solution. As noted above further on in this section we concentrate on the solution of OCP Ia. We use the dynamic programming approach [20].

We write down the Bellman equation for the problem OCP Ia.

$$F_t^*(x) = \min_{u(t)}\{c_x(t)x(t) + c_u(t)u(t) + F_{t+1}^*(x)\} \text{ with } F_T^*(x) = 0 \qquad (5.4)$$

Further on we use the general scheme of mathematical induction. We first solve (5.4) for $t = T-1, t = T-2$, $t = T-3$, and formulate the induction hypothesis:

$$F_t^*(x) = d_x(t)x(t) + c_u(T-1)\underline{u}(T-1) + \sum_{s=t}^{T-2}[c_u(s) + d_x(s+1)\beta(s+1)]u^*(s), \quad (5.5)$$

where

$$u^*(t) = \begin{cases} \underline{u}(t) & \text{if} \quad c_u(t) + d_x(t+1)\beta(t+1) \geq 0 \\ \overline{u}(t) & \text{if} \quad c_u(t) + d_x(t+1)\beta(t+1) \leq 0 \end{cases} \quad (5.6)$$

and

$d_x(t) = c_x(t) + d_x(t+1)\alpha(t)$ for $t = 0,1,\cdots,T-2$ with $d_x(T-1) = c_x(T-1)$.

Assume now that the induction hypothesis (5.5) is true for some $t = k+1$ that is

$$F_{k+1}^*(x) = d_x(k+1)x(k+1) + c_u(T-1)\underline{u}(T-1) + \sum_{s=k+1}^{T-2}[c_u(s) + d_x(s+1)\beta(s+1)]u^*(s) \quad (5.7)$$

We prove that the expression (5.5) holds true for $t = k$. Indeed,

$$F_k^*(x) = \min_{u(k)}\{c_x(k)x(k) + c_u(k)u(k) + F_{k+1}^*(x)\}$$

$$= \min_{u(k)}\left\{c_x(k)x(k) + c_u(k)u(k) + d_x(k+1)x(k+1) + c_u(T-1)\underline{u}(T-1) + \sum_{s=k+1}^{T-2}[c_u(s) + d_x(s+1)\beta(s+1)]u^*(s)\right\}$$

$$= \min_{u(k)}\left\{c_x(k)x(k) + c_u(k)u(k) + d_x(k+1)[\alpha(k)x(k) + \beta(k+1)u(k)] + c_u(T-1)\underline{u}(T-1) + \sum_{s=k+1}^{T-2}[c_u(s) + d_x(s+1)\beta(s+1)]u^*(s)\right\}$$

$$= \min_{u(k)}\left\{\big(c_x(k) + d_x(k+1)\alpha(k)\big)x(k) + \big(c_u(k) + d_x(k+1)\beta(k+1)\big)u(k) + \sum_{s=k+1}^{T-2}[c_u(s) + d_x(s+1)\beta(s+1)]u^*(s)\right\}$$

Hence,

$$F_k^*(x) = \min_{u(k)} \Big\{ d_x(k)x(k) + \big(c_u(k) + d_x(k+1)\beta(k+1)\big)u(k) \qquad (5.8)$$
$$+ \sum_{s=k+1}^{T-2} [c_u(s) + d_x(s+1)\beta(s+1)]u^*(s) \Big\}$$

The minimum in (5.8) with respect to $u(k)$, where $0 \le \underline{u}(t) \le u(t) \le \bar{u}(t)$, is achieved at

$$u^*(k) = \begin{cases} \underline{u}(k) & \text{if} \quad c_u(k) + d_x(k+1)\beta(k+1) \ge 0 \\ \bar{u}(k) & \text{if} \quad c_u(k) + d_x(k+1)\beta(k+1) < 0 \end{cases}$$

and its value is

$$F_k^*(x) = d_x(k)x(k) + c_u(T-1)\underline{u}(T-1) \qquad (5.9)$$
$$+ \sum_{s=k}^{T-2} [c_u(s) + d_x(s+1)\beta(s+1)]u^*(s)$$

It can be seen from (5.9) that the assumption (5.7) holds true and hence the induction hypothesis (5.5) - (5.6) is proved**.** Thus, we summarize, the optimal control for OCP Ia (and for OCP I) is given by (5.6) and the optimal value of the cost function is $F_{min} = F_0^*(x_0)$ or, alternatively, it can be evaluated as $F_{min} = \sum_{i=0}^{T-1}(c_x(t)x(t) + c_u(t)u(t))$, where the optimal trajectory $\{x^*(t)\}$ and optimal control sequence $\{u^*(t)\}$ satisfy the equations (5.2)

5.3 A continuous-time dynamic mobile source air pollution optimal control problem

5.3.1 The Model

A continuous-time dynamic mobile source air pollutant model [61] to predict and control the emission level from a total vehicle population is developed using a comprehensive set of input data for the range of variables that can influence the vehicle emissions including manufacture year (age of vehicles), vehicle populations, total annual kilometres travelled, and emission factors for a number of selected pollutant.

Vehicle population evolution

Consider a vehicle population on the road in a given area (state, city, etc). The total vehicle population can be divided into n different cohorts (groups, categories, grades) $G_i, i = 1,2,\ldots,n$, on the basis of the attribute "age", the cohort G_1 being the "youngest" and the cohort G_n being the "oldest".

Let $x_i(t), i = 1,2,\ldots,n$, denote the size (the number of vehicles) of cohort G_1 at time t; $x_i(0)$ is clearly, the initial number of vehicles in the ith cohort. Let, now, Δt be a sufficiently small period of time. It is assumed in general that during the time Δt a fraction β_i of vehicles within cohort G_i will progress to the $(i+1)$th cohort, whilst another fraction γ_i of the vehicles will be scrapped due to accidents or breakdowns or because of long lifetime, or simply leave the vehicle population for other systems. It is assumed also that the progression β_i and withdrawal γ_i fractions are nonnegative constant independent of time that is $\beta_i > 0$ and $\gamma_i > 0$. They, clearly, satisfy the inequalities $0 \le \beta_i \le 1$ and $0 \le \gamma_i \le 1, i = 1,2,\ldots,n$. Under realistic conditions, the progression and the withdrawal frictions satisfy the conditions

$$0 \le \beta_i + \gamma_i \le 1 \tag{5.10}$$

where the lower inequality assures that not all of the vehicles within a cohort suddenly leave, and the upper inequality arises sine some attenuation is assumed. Note that by adopting a strict upper inequality it is not necessary to incorporate the attenuation explicitly within the system dynamics. It will, however, affect the influence the evolution through the values of is β_i and γ_i.

Taking into account the continuous inflow and outflow of individuals into the cohort and out of it, a standard balance equation takes place

$$x_i(t+\Delta t) = x_i(t) + \beta_{i-1}x_{i-1}(t)\Delta t + u_i(t)\Delta t - \gamma_i x_i(t)\Delta t - \beta_i x_i(t)\Delta t \tag{5.11}$$

where β_i and γ_i are respectively the progression and the withdrawal rates and the $u_i(t)$ is the number vehicles entering cohort G_i from other vehicle population outside the system (the total vehicle population under consideration). All vehicles $u_i(t)$ – new and used, purchased from outside the system can be used as decision (control) variables. They are, obviously, non-negative, $u_i(t) \geq 0$ for $t \in [0, T)$, and $i = ,2, \dots, p$, where $p \leq n$, assuming that vehicles are delivered into the first p cohort only.

According to Rumchev, et al [61], the continuous-time model of vehicle population dynamic systems can be expressed as

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \text{ for } t \in [0, T) \tag{5.12}$$

with the system (cohort) matrix **A** and control matrix **B** given, respectively, by

$$\mathbf{A} = \begin{bmatrix} -(\beta_1 + \gamma_1) & 0 & 0 & \cdots & 0 \\ \beta_1 & -(\beta_2 + \gamma_2) & 0 & \cdots & 0 \\ 0 & \beta_2 & -(\beta_3 + \gamma_3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \beta_{n-1} & -\gamma_n \end{bmatrix}, \tag{5.13}$$

and

$$\mathbf{B} = \left[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p\right] \geq 0, p \leq n \tag{5.14}$$

In the continuous-time system (5.12), the vector $\mathbf{x}(t) \geq 0$ is the state vector (whose components are the numbers of vehicles in different cohorts at time t, i.e., $\mathbf{x}(t) = (\, x_1(t), x_2(t), \dots, x_n(t)\,)'$), $\mathbf{u}(t) \geq 0$ is the control vectors (which components denote the numbers of vehicles purchased from other systems into the individual cohorts at time t, i.e., $\mathbf{u}(t) = \left(u_1(t), u_2(t), \dots, u_p(t)\,\right)'$) and $\mathbf{e}_j$ is the unit basis vector (with 1 in the i_s position and all the other entries zeros).

Matrix **A** has all its off-diagonal elements nonnegative, that is $a_{ij} \geq 0$, $i \neq j$. Matrices with nonnegative off-diagonal entries are named Metzler matrix [5, 14]. Therefore, the system (5.12) is a positive linear system.

Emission level

Let e_{ij}(called emission level of factor) be the average emission (in *gram*) of pollutant type j, $j = ,2, \dots, k$, from one vehicle of cohort $G_i, i = 1,2, \dots, n$. The emission levels e_{ij} are, clearly, positive: $e_{ij} > 0$ for all i and j. Then the total emission level $\varepsilon_j(t)$ for the jth type of pollutant at time t can be calculated by the following formula.

$$\varepsilon_j(t) = \sum_{i-1}^{n} e_{ij}\, x_i(t), \tag{5.15}$$

and the total emission for the jth type of pollutant over the planning horizon *T* will be

$$\varepsilon_j = \int_0^T \varepsilon_j(t)dt = \int_0^T \sum_{i-1}^{n} e_{ij}\, x_i(t)dt. \tag{5.16}$$

Let the emission vector at time t for the total vehicle population be denoted as $\boldsymbol{\varepsilon}(t) = [\varepsilon_1(t), \varepsilon_2(t), \dots, \varepsilon_k(t)]'$, and the matrix of emission levels as $E = [e_{ij}]_{k\times n}$. Then the total emission vector at time t can be represented as $\boldsymbol{\varepsilon}(t) = \mathbf{Ex}(t)$, and the total emission vector for the planning horizon T will be

$$\boldsymbol{\varepsilon} = \int_0^T \boldsymbol{\varepsilon}(t)dt = \int_0^T \mathbf{EMx}(t)dt. \tag{5.17}$$

The emission levels and the average distances are usually known from demographic and engineering studies, and there are (at least in the developed countries) established standards and these standards are, indeed, upper bounds $\overline{\boldsymbol{\varepsilon}_j}(t)$ and $\overline{\boldsymbol{\varepsilon}_j}$ for the emission levels at time t and respectively the emission levels over the planning horizon T. Therefore, the emissions are subject to the following restrictions

$$\varepsilon_j(t) \le \overline{\varepsilon_j}(t),$$

or, in vector form,

$$\mathbf{Ex}(t) \le \overline{\boldsymbol{\varepsilon}_j}(t),$$

and respectively

$$\varepsilon_j = \int_0^T \varepsilon_j(t)dt \le \overline{\varepsilon_j},$$

or, in vector form,

$$\int_0^T \mathbf{EMx}(t)dt \le \bar{\boldsymbol{\varepsilon}}, \tag{5.18}$$

where $\bar{\boldsymbol{\varepsilon}}(t) = \left(\varepsilon_1(t), \varepsilon_2(t), \dots, \varepsilon_k(t)\right)'$ and $\bar{\boldsymbol{\varepsilon}} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k)'$ are, respectively, the vector of standards at time t (an hour, a day) and the vector of standards over the planning horizon.

5.3.2 Optimal control problem

In this section, an optimal control problem of the continuous-time dynamic mobile source air pollution is developed. Consider the model discussed earlier, we will discuss the optimal control problem to optimize the number of new or used purchased in cohorts i

such that the minimum total pollutant emission cost in an area over planning horizon T is achieved.
Hence, the objective of the problem is formulated as

Minimizing
$$J = \int_0^T c\mathbf{EMx}(t)dt, \tag{5.19}$$

where $c = [c_1, c_2, \ldots, c_n]$, c_i is a nonnegative scalar to measure the different types of pollutant to optimize the total emission cost. In practice, the pollutant emissions have different cost level to the environment. Hence, the value of c_i can be chosen accordingly depends on the harm level of the pollutant to the environment.
The optimal control is subject to the following restrictions:

- The vehicle population evolution is described by the continuous-time model of vehicle population dynamic systems (5.12)
$$\dot{\mathbf{x}}(t) = \mathbf{Ax}(t) + \mathbf{Bu}(t) \text{ for } t \in [0, T], \text{ where } \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \tag{5.20}$$
- with initial and desired final vehicle population
$$\mathbf{x}(0) = \mathbf{x}_0 \geq 0 \text{ and } \mathbf{x}(T) = \mathbf{x}_T \geq 0. \tag{5.21}$$
- nonnegative control $\mathbf{u}(t)$ as it represents the numbers of vehicles purchased from other systems into the individual cohorts at time t.
$$\mathbf{u}(t) \geq 0. \tag{5.22}$$
- emission restriction vector (5.18) as a standard on the total emission for planning horizon T established by some countries (at least in the developed countries)
$$\int_0^T \mathbf{EMx}(t)dt \leq \bar{\boldsymbol{\varepsilon}}. \tag{5.23}$$
where $\bar{\boldsymbol{\varepsilon}} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)'$ and ε_i is a scalar for $i = 1,2, \ldots, k$.

Solution Method

The optimal control problem (5.19) – (5.23) is called an isoperimetric problem as it contains an integral constraint (5.23). The classical method to solve the problem is to introduce the Lagrange multiplier and apply the Euler-Lagrange equations to the augmented integrand. The value of the Lagrange multiplier must be selected such that the integrand constraint satisfied which requires a trial and error procedure. The integral constraint (5.23) contains k scalar integral constraints. Let the scalar integral constraint be formulated as

$$\int_{t_0}^{t_f} F(t,x,u)dt = s \tag{5.24}$$

where s is a constant. According to [66], there are two equivalent ways to apply integral constraint (5.24) .

i. The original method for handling integral constraints as discussed by Kalaba and Spingarn [67]. The method is performed by introducing the Lagrange multiplier as the state variable and evaluated simultaneously with the optimum solution.
ii. The integral constraint (5.24) can be converted into a differential constraint and two given boundary conditions. If a new state, say $x_{n+1}(t)$, be defined as
$$\dot{x_{n+1}}(t) = F(t,x,u),\ x_{n+1}(t_0) = 0, \tag{5.25}$$
then the integral constraint becomes the final condition
$$x_{n+1}(t_f) = s. \tag{5.26}$$

In this method, the integral constraint (5.24) corresponds directly into an optimal control problem. The Lagrange multiplier associated with the differential equation (5.25) is constant because the state $x_{n+1}(t)$ does not occur on the right side, i.e., the function $F(t,x,u)$.

Thus, the optimal control problem (5.19) – (5.23) can be solved by introducing new variables $x_{n+1}(t), x_{n+2}(t), \dots, x_{2n}(t)$ such that

$$\begin{bmatrix} \dot{x_{n+1}}(t) \\ \dot{x_{n+2}}(t) \\ \vdots \\ \dot{x_{2n}}(t) \end{bmatrix} = \mathbf{EMx}(t),$$

that satisfy

$$\begin{bmatrix} x_{n+1}(T) \\ x_{n+2}(T) \\ \vdots \\ x_{2n}(T) \end{bmatrix} = \int_0^T \mathbf{EMx}(t)dt \leq \bar{\boldsymbol{\varepsilon}},$$

and $x_{n+1}(0) = x_{n+2}(0) = \ \dots, = x_{2n}(0) = 0.$

Therefore, the objective function (5.19) of the optimal control problem (5.19) – (5.23) becomes minimizing the modified cost function

$$J = c \begin{bmatrix} x_{n+1}(T) \\ x_{n+2}(T) \\ \vdots \\ x_{2n}(T) \end{bmatrix}. \tag{5.27}$$

Subject to

$$\dot{\mathbf{x}}(t) = \mathbf{Ax}(t) + \mathbf{Bu}(t) \text{ for } t \in [0, T], \text{ where } \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \tag{5.28}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \geq 0 \text{ and } \mathbf{x}(T) = \mathbf{x}_T \geq 0, \tag{5.29}$$

$$u(t) \geq 0, \tag{5.30}$$

$$\begin{bmatrix} x_{n+1}(T) \\ x_{n+2}(T) \\ \vdots \\ x_{2n}(T) \end{bmatrix} \leq \bar{\boldsymbol{\varepsilon}}, \tag{5.31}$$

$$\text{and } x_i(0) = 0, i = n+1, n+2, \dots, 2n. \tag{5.32}$$

The computation of the modified optimal control problem (5.27) – (5.32) can be solved using optimization software such as ICLOCS [31] or Miser 3.1. [68]. An example of solving an optimization problem with integral constraint using Miser 3.1. is described by Teo et al [69].

5.4 Concluding remarks

The first application is a dynamic model of oil extraction and its optimization. A discrete-time dynamic model that describes the evolution of oil extraction process from a single well or reservoir under water flooding is developed and analysed. On the basis of the model, an optimal control problem is formulated. The obtained results could have important applications in improving the production methodology and supporting the managerial decisions in the process of oil extraction under water flooding.

The second application is a continuous-time dynamic mobile source air pollution optimal control problem. The optimal control on the basis of the continuous-time dynamic mobile source air pollution model is developed and analysed. The model can be used to compute and predict the emission level of the mobile source air pollution level in the future or to control the number of the new or used vehicle in certain period of time.

CHAPTER 6

Summary and future research directions

6.1 Main contributions of the thesis

In this thesis, we consider the minimum energy problem, one of the classical problems of linear control theory, for positive linear systems. We have developed analytical solutions for this problem using dynamic programming approach. The main contributions are summarised below.

In Chapter 3, criteria for existence of solutions to the minimum energy problem for positive linear systems with scalar control are established both in the continuous-time and the discrete-time case. The relationship between the problem and the geometric properties of the system is well exploited. We have successfully derived analytic solutions to the minimum energy problem for positive linear discrete-time systems with scalar control and fixed initial and final states in both forms: a feedback form and an open-loop form. Such solutions have not been known in the literature previously. The minimum energy problem for positive linear discrete-time systems with scalar control and fixed initial and final states can be reduced to a minimum energy problem with free final state by including in the cost function a term that reflects the deviation of the final state in the reduced problem from the targeted final state. Using dynamic programming approach, an analytical solution of the reduced minimum energy problem with free final state has been obtained and analysed. It has also been shown that the relaxation of the problem leads to a decrease of the consumed energy of the input but at the expense of not reaching the desired final state. Such a "trade-off" might be quite appealing in a number of real-life problems. In Section 3.3, we have derived analytical solutions to the minimum energy problem for positive linear continuous-time systems with scalar control and fixed initial and final states in both form: a feedback form and an open loop form. Such solutions have not been known in the literature previously. In Chapter 4, the solution and the criteria for the existence of the solution of the minimum energy problem for positive discrete-time linear systems and positive linear continuous-time systems with vector control and any pair of initial and fixed final state have been established in by using dynamic programming approach. Initially, the problems are solved without nonnegativity constraints, then some conditions that guarantee the nonnegativity of the systems are imposed. To the best of my knowledge, the solution to the minimum energy problem with fixed initial and final state without the nonnegativity constraints (derived by Lagrange-multipliers approach) is well known only in an open-loop form. In this chapter, solutions derived using dynamic programming are developed in both feedback and open-loop forms. The relationship between the problem and the geometric properties of the systems is revealed and well exploited. We have obtained analytical solutions to the minimum energy problem for positive linear discrete-time systems with vector control and fixed initial and final states in an open-loop form and a feedback form under less restrictive than in the previous studies (i.e. Kaczorek [14]) conditions. We do not require a zero initial state and the positive reachability of the positive linear discrete-time system. Furthermore, we have also obtained analytical solutions to the minimum energy problem for positive linear continuous-time systems with vector control and fixed initial and final state in open-loop forms and feedback form under less restrictive than the previous studies (i.e. Kaczorek [14]) conditions. We require neither a zero initial state nor the positive reachability of the positive linear continuous-time systems.

In Chapter 5, two important applications of positive linear systems related to energy and ecology are addressed. The first application is a dynamic model of oil extraction and its optimization. A discrete-time dynamic model that describes the evolution of the oil extraction process from a single well or reservoir under water flooding has been developed and analysed. On the basis of the model, an optimal control problem is formulated and solved. The obtained results could have important applications in improving the production methodology and supporting the managerial decisions in the process of oil extraction under water flooding. The second application is a continuous-time dynamic mobile-source air pollution optimal control problem. Based on the continuous-time dynamic mobile source air pollution model discussed in previous study [61], the optimal control has been developed and analysed. Solution to the optimal control discussed is also provided. The model can be used to compute and predict the emission level of the mobile source air pollution level in the future or to control the number of the new or used vehicle in certain period of time.

6.2 Future research directions

The work in this thesis has opened several research topics in the future. We discuss some of them below.

In Chapters 3 and 4 we have discussed the minimum energy problem which is one of the classical problems of linear quadratic problem. Analytical solutions and sufficient conditions to guarantee the nonnegativity for this particular problem are addressed in this thesis. An advanced study for general linear quadratic problem for positive linear system can be considered. Utilising the geometry analysis of reachable set and nonnegative matrix properties one can also establish sufficient conditions that guarantee the nonnegativity. Furthermore, utilising the dynamic programming approach, the nonnegativity criteria of the linear quadratic problem for positive linear system can be observed in every stage of the procedure.

In Chapter 5, two applications related to positive linear systems and energy topics are addressed. In the first application, the proposed optimization approach to production planning and support to the managerial decisions in the process of oil extraction under water flooding can be extended to include a desirable final state x_T in the model and minimize the time for achieving it. This leads to a different type of optimal control problem - a minimum-time optimal control problem in which controllability property mentioned in the second section has an important role to play.

Bibliography

[1] L. Benvenuti and L. Farina, Positive and compartmental systems, Automatic Control, IEEE Transactions on **47**, no. 2, 370-373, (2002).

[2] L. Farina and S. Rinaldi, Positive linear systems: Theory and applications, Wiley-Interscience, 2011.

[3] W. M. Haddad and V. Chellaboina, Stability and dissipativity theory for nonnegative dynamical systems: A unified analysis framework for biological and physiological systems, Nonlinear Analysis: Real World Applications **6**, no. 1, 35-65, (2005).

[4] R. Shorten, F. Wirth and D. Leith, A positive systems model of tcp-like congestion control: Asymptotic results, Networking, IEEE/ACM Transactions on **14**, no. 3, 616-629, (2006).

[5] D. G. Luenberger, Introduction to dynamic systems : Theory, models, and applications, Wiley, New York, 1979.

[6] A. Berman and R. J. Plemmons, Nonnegative matrices, The Mathematical Sciences, Classics in Applied Mathematics **9**, (1994).

[7] E. D. Sontag, Molecular systems biology and control, European journal of control **11**, no. 4, 396-435, (2005).

[8] A. J. Lotka, Elements of physical biology, Williams & Wilkins Baltimore, 1925.

[9] V. Volterra, Fluctuations in the abundance of a species considered mathematically, Nature **118**, 558-560, (1926).

[10] J. Hofbauer and K. Sigmund, Evolutionary games and population dynamics, Cambridge University Press, 1998.

[11] W. W. Leontief, Quantitative input and output relations in the economic systems of the united states, The Review of economics and Statistics **18**, no. 3, 105-125, (1936).

[12] J. v. Neumann, A model of general economic equilibrium, The Review of Economic Studies **13**, no. 1, 1-9, (1945).

[13] J. A. Jacquez and C. P. Simon, Qualitative theory of compartmental systems, Siam Review **35**, no. 1, 43-79, (1993).

[14] T. Kaczorek, Positive 1d and 2d systems, Springer, London, 2002.

[15] V. Rumchev and D. James, Controllability of positive linear discrete-time systems, International Journal of Control **50**, no. 3, 845-857, (1989).

[16] V. Rumchev and J. Adeane, Reachability and controllability of time-variant discrete-time positive linear systems, Control and Cybernetics **33**, no. 1, 85-94, (2004).

[17] F. Facchinei and J.-S. Pang, Finite-dimensional variational inequalities and complementarity problems, vol. 1, Springer, 2003.

[18] F. L. Lewis, D. Vrabie and V. L. Syrmos, Optimal control, John Wiley & Sons, 2012.

[19] B. D. Anderson and J. B. Moore, Optimal control: Linear quadratic methods, Courier Corporation, 2007.

[20] D. P. Bertsekas, Dynamic programming and optimal control, Athena Scientific, Belmont, Mass., 2005.

[21] C. Beauthier and J. Winkin, On the positive LQ-problem for linear discrete time systems, Positive Systems, 45-53, (2009).

[22] C. Beauthier and J. J. Winkin, LQ□optimal control of positive linear systems, Optimal Control Applications and Methods **31**, no. 6, 547-566, (2010).

[23] W. P. M. H. Heemels, s. J. L. V. Eijndoven and A. A. Stoorvogel, Linear quadratic regulator problem with positive control, International Journal of Control **70**, 551-578, (1998).

[24] Y. Hu and X. Y. Zhou, Constrained stochastic LQ control with random coefficients, and application to portfolio selection, SIAM Journal on Control and Optimization **44**, no. 2, 444-466, (2005).

[25] M. Labissi, J. J. Winkin, and C. Beauthier, "On the positive LQ-problem for linear continuous-time system," *Lecture Notes in Control and Information Sciences*, vol. 341, Heidelberg: Springer, Berlin, 2006.

[26] M. Pachter, The linear-quadratic optimal control problem with positive controllers, International Journal of Control **32**, no. 4, 589-608, (1980).

[27] S. G. Nersesov, W. M. Haddad and V. Chellaboina, Optimal fixed□structure control for linear non□negative dynamical systems, International Journal of Robust and Nonlinear Control **14**, no. 5, 487-511, (2004).

[28] S. L. CAMPBELL, On positive controllers and linear quadratic optimal control problems†, International Journal of Control **36**, no. 5, 885-888, (1982).

[29] J.-P. Quadrat, The discrete state linear quadratic problem, Decision and Control, 1989., Proceedings of the 28th IEEE Conference on, IEEE, p.^pp. 736-741, 1989.

[30] R. E. Bellman, Dynamic programming, Princeton Univ. Pr., Princeton, N.J., 1957.

[31] P. Falugi, E. Kerrigan and E. Van Wyk, Imperial college london optimal control software user guide (iclocs), Department of Electrical and Electronic Engineering, Imperial College London, London, England, UK, (2010).

[32] K. J. Arrow, Mathematical methods in the social sciences, 1959: Proceedings, vol. 4, Stanford University Press, 1960.

[33] C. Beauthier, The linear quadratic (LQ)-optimal control problem for invariant linear systems, (2011).

[34] C. Dellacherie, S. Martinez and J. San Martin, Inverse m-matrices and ultrametric matrices, vol. 2118, Springer, 2014.

[35] R. Bhatia, Positive definite matrices, Princeton University Press, 2009.

[36] J. E. Gentle, "Matrix algebra theory, computations, and applications in statistics," *Springer texts in statistics*, Springer,, New York, N.Y. ; London, pp. xxii, 528 p. ill. 524 cm., 2007.

[37] M. P. FANTI, B. Maione and B. Turchiano, Controllability of linear single-input positive discrete-time systems, International Journal of Control **50**, no. 6, 2523-2542, (1989).

[38] M. Fanti, B. Maione and B. Turchiano, Controllability of multi-input positive discrete-time systems, International Journal of Control **51**, no. 6, 1295-1308, (1990).

[39] R. Bru, L. Caccetta, S. Romero, V. Rumchev and E. Sanchez, Recent developments in reachability and controllability of positive linear systems, Proceedings of the 15th IFAC World Congress, p.^pp., 2002.

[40] R. Bru, C. Coll, S. Romero and E. Sanchez, Reachability indices of positive linear systems, Electronic Journal of Linear Algebra **11**, 88-102, (2004).

[41] M. A. Rami, F. Tadeo and A. Benzaouia, Control of constrained positive discrete systems, American Control Conference, 2007. ACC'07, IEEE, p.^pp. 5851-5856, 2007.

[42] C. Burgat and A. Benzaouia, Stability properties of positively invariant linear discrete time systems, Journal of Mathematical Analysis and Applications **143**, no. 2, 587-596, (1989).

[43] G. James and V. Rumchev, Stability of positive linear discrete-time systems, TECHNICAL SCIENCES **53**, no. 1, (2005).

[44] L. Caccetta and V. Rumchev, A survey of reachability and controllability for positive linear systems, Annals of Operations Research **98**, no. 1-4, 101-122, (2000).

[45] M. Evans and D. Murthy, Controllability of discrete-time systems with positive controls, Automatic Control, IEEE Transactions on **22**, no. 6, 942-945, (1977).

[46] D. Murthy, Controllability of a linear positive dynamic system, International journal of systems science **17**, no. 1, 49-54, (1986).

[47] L. Farina and L. Benvenuti, Polyhedral reachable set with positive controls, Mathematics of Control, Signals, and Systems (MCSS) **10**, no. 4, 364-380, (1997).

[48] L. Benvenuti, On the reachable set for third-order linear discrete-time systems with positive control, Systems & Control Letters **60**, no. 9, 690-698, (2011).

[49] ---, On the reachable set for third-order linear discrete-time systems with positive control: The case of complex eigenvalues, Systems & Control Letters **60**, no. 12, 1000-1008, (2011).

[50] M. E. Valcher, Reachability properties of continuous-time positive systems, Automatic Control, IEEE Transactions on **54**, no. 7, 1586-1590, (2009).

[51] C. Commault and M. Alamir, On the reachability in any fixed time for positive continuous-time linear systems, Systems & Control Letters **56**, no. 4, 272-276, (2007).

[52] Z. Bartosiewicz, Linear positive control systems on time scales; controllability, Mathematics of Control, Signals, and Systems **25**, no. 3, 327-343, (2013).

[53] D. S. Bernstein, Matrix mathematics: Theory, facts, and formulas, Princeton University Press, 2009.

[54] V. Rumchev and S. Chotijah, "The minimum energy problem for positive discrete-time linear systems with fixed final state," *Positive systems*, Springer, pp. 141-149, 2009.

[55] ---, On the minimum-energy problem for positive discrete-time linear systems, International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE., p.^pp. 1381-1391, 2009.

[56] S. Chotijah and V. Rumchev, The minimum energy problem for single-input single-output positive continuous-time linear systems with fixed initial and final state, International Conference on Computational Mathematics, Computational Geometry, and Statistics (CMCGS), Global Science and Technology Forum, p.^pp. 8, 2015.

[57] K. Ogata, Discrete-time control systems, vol. 8, Prentice-Hall Englewood Cliffs, NJ, 1995.

[58] A. Berman and R. J. Plemmons, Nonnegative matrices in the mathematical sciences, vol. 9, SIAM, 1994.

[59] Y. Ohta, H. Maeda and S. Kodama, Reachability, observability, and realizability of continuous-time positive systems, SIAM Journal on Control and Optimization **22**, no. 2, 171-180, (1984).

[60] R. Gupta, V. Rumchev and S. Chotijah, A dynamic model of oil extraction and its optimization, International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management **1**, 765, (2015).

[61] V. Rumchev, K. Rumchev and Y. Higashiyama, A continuous-time dynamic mobile source air pollution model and its mainainability property, International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management, 511, (2013).

[62] T. Ahmed, Reservoir engineering handbook, Gulf Professional Publishing, 2010.

[63] G. P. Willhite and Society of Petroleum Engineers of AIME., Waterflooding, Society of Petroleum Engineers, Richardson, TX, 1986.

[64] R. Baker, Reservoir management for waterfloods-part ii, Journal of Canadian Petroleum Technology **37**, no. 01, (1998).

[65] R. Gupta, M. Rabiei and D. V. Rumchev, Decision models for water production diagnosis based on water oil ratio via soft computing techniques, Petrotech 2012, p.^pp., 2012.

[66] D. G. Hull, Optimal control theory for applications, Springer Science & Business Media, 2013.

[67] R. Kalaba and K. Spingarn, Optimization of functionals subject to integral constraints, Journal of optimization theory and applications **24**, no. 2, 325-335, (1978).

[68] L. Jennings, M. Fisher, K. Teo and C. Goh, Miser3: Solving optimal control problems—an update, Advances in Engineering Software and Workstations **13**, no. 4, 190-196, (1991).

[69] A. Lim, Y. Lieu, K. Teo and J. Moore, Linear-quadratic optimal control with integral quadratic constraints, Optimal Control Applications and Methods **20**, no. 2, 79-92, (1999).

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledge.